IMT School for Advanced Studies Lucca

Lucca, Italy

Stratification of first episode psychosis based on clinical and neurobiological features: from single-center studies to big data

PhD in Cognitive and Cultural Systems - Track in Cognitive, Computational and Social Neurosciences

XXXV Cycle

Candidate Alessandro Pigoni

The dissertation of Alessandro Pigoni

Programme Coordinator: Emiliano Ricciardi, IMT School for Advanced Studies Lucca.

Advisor: Pietro Pietrini, IMT School for Advanced Studies Lucca.

Co-Advisors: Luca Cecchetti, IMT School for Advanced Studies Lucca

The dissertation of Alessandro Pigoni has been reviewed by:

Prof. Stefania Paola Corti

Prof. Hilleke Hulshoff Pol

Prof. Sevdalina Kandilarova

IMT School for Advanced Studies, Lucca, 2024

Table of Contents

Vita and publications

Abstract

1. Introduction

1.1 First Episode Psychosis –presentation and challenges

1.2 How to study First Episode Psychosis – findings in MRI studies

2. Single-center studies

2.1 Insula volumes in first-episode and chronic psychosis: A longitudinal MRI study (Pigoni et al., 2022)

2.1.1 introduction and rationale

2.1.2 Methods

2.1.3 Results

2.1.4 Discussion

2.2 Association between brain metabolism and clinical and temperamental scales in first episode psychosis (Pigoni et al., in publishing)

2.2.1 introduction and rationale

2.2.2 Methods

2.2.3 Results

2.2.4 Discussion

3. Big Data and Machine Learning – we need a bigger boat!

3.1 Classification of first-episode psychosis using cortical thickness: A large multicenter MRI study (Pigoni et al., 2021)

3.1.1 introduction and rationale

3.1.2 Methods

3.1.3 Results

3.1.4 Discussion

3.2 Clustering studies in FEP

3.2.1 Oezturk et al., 2022

3.2.2 Dwyer et al., 2023

4. Moving forward – the PHLAMES study (GR-2019-12369479)

4.1 Introduction to autoimmune encephalitis

4.2 Aims of the PHLAMES study

4.3 Methods employed

4.4 First results

4.5 Discussion

4.6 Limits

5. Discussion: where are we and where are we going

6. References

Vita and Publications

2014 Graduated in Medicine at the University Vita-Salute San Raffaele (Milan, Italy), with top marks and an experimental thesis on genetics in schizophrenia (Full title: "*Exploring the effect of adducins genetic variability on cognition in schizophrenia*").

2011 Fellow researcher During at the **Yale University Cancer Center**

2019 **specialized Psychiatrist at the University of Milan** in 2019, with top marks and an experimental thesis on machine learning applied to first episode psychosis (Full title: "*Classification of first episode psychosis using machine learning techniques*").

2018-2019 **Ludwig-Maximillan University in Munich** (Germany), collaborating with the PRONIA group directed by Professor Koutsouleris.

Publications

Pigoni A et al., Machine learning and the prediction of suicide in psychiatric populations: a systematic review. Translational Psychiatry, 2024

Colli C, **Pigoni A**, et al. Combining clozapine with cariprazine: Two case reports highlighting potential drug-drug interaction. Schizophrenia research 2024, 266, 180-182

Maggioni E, **Pigoni A**, et al., Right frontal cingulate cortex mediates the effect of prenatal complications on youth internalizing behaviors. Molecular Psychiatry, 2024 1-10

G Cereda, M Lazzaretti, C Colli, E Fontana, P Enrico, **A Pigoni**, et al., Pharmacokinetic drug-drug interaction between clozapine and brexpiprazole: A case series. Schizophrenia research 2023, 261, 15-17

Tesic I, **Pigoni A**, et al., How does feeling pain look like in depression: A review of functional neuroimaging studies. Journal of Affective Disorders, 2023 Vol 339, Pages 400-411

Ciceri T, Squarcina L, **Pigoni A** et al. Geometric Reliability of Super-Resolution Reconstructed Images from Clinical Fetal MRI in the Second Trimester. Neuroinform 2023, 21, 549–563.

Christian Theisen, Marlene Rosen, Nikolaos Koutsouleris, Stephan Ruhrmann, Joseph Kambeitz, Lana Kambeitz-Ilankovic, Anita Riecher-Rössler, Katharine Chisholm, Rachel Upthegrove, Linda A Antonucci, Alessandro Bertolino, **Alessandro Pigoni**, et al. The heterogeneity of attenuated and brief limited psychotic symptoms: association of contents with age, sex, country, religion, comorbidities, and functioning. Frontiers in psychiatry 2023, 14, 1209485

Dwyer DB, Chand GB, **Pigoni A** et al. Psychosis brain subtypes validated in first-episode cohorts and related to illness remission: results from the PHENOM consortium. Molecular Psychiatry, 28 – 2023

Enrico, G Delvecchio, N Turtulici, R Aronica, **A Pigoni** et al. A machine learning approach on whole blood immunomarkers to identify an inflammation-associated psychosis onset subgroup. P, Molecular Psychiatry 28 (3), 1190-1200 – 2023

D'ambrosio E, Abrol A, **Pigoni A**. Machine learning and psychosis: Diagnosis, prognosis and treatment. Frontiers in Psychiatry 2023, 14, 1133072

Cellini P, **Pigoni A**, et al., Machine learning in the prediction of postpartum depression: A review. Journal of Affective Disorders 2022, 309, 350-357

Pigoni A et al., Insula volumes in first-episode and chronic psychosis: A longitudinal MRI study. Schizophr Res 2022 Mar;241:14-23

Chiacchio M, **Pigoni A.** Red in tooth and claw: A review of animal antagonistic roles in movies. People and Nature 2022, 4 (3), 701-710

LA Antonucci, N Penzel, R Sanfelici, **A Pigoni** et al., Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. The British Journal of Psychiatry 2022, 220 (4), 229-245

Oeztuerk OF, **Pigoni A**, et al. The clinical relevance of formal thought disorder in the early stages of psychosis: results from

the PRONIA study European archives of psychiatry and clinical neuroscience, 2022 1-11

Oeztuerk OF, **Pigoni A**, et al. Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half-century studies European archives of psychiatry and clinical neuroscience, 2022, 1-13

Boscutti A, **Pigoni A**, et al., The influence of 5-HTTLPR, BDNF Rs6265 and COMT Rs4680 polymorphisms on impulsivity in bipolar disorder: the role of gender Genes 2022, 13 (3), 482

Enrico P, Delvecchio G, Turtulici N, Aronica R, **Pigoni A**, et al., Identification of an inflammation-associated psychosis onset subgroup by applying unsupervised machine learning to whole-blood expression levels of immune gene transcripts, Neuroscience Applied 2022, 1, 100273

Junhao Wen, Erdem Varol, Aristeidis Sotiras, Zhijian Yang, Ganesh B Chand, Guray Erus, Haochang Shou, Ahmed Abdulkadir, Gyujoon Hwang, Dominic B Dwyer, **Alessandro Pigoni**, et al. Multi-scale semi-supervised clustering of brain

images: Deriving disease subtypes Medical image analysis 2022, 75, 102304

A Boscutti, G Delvecchio, **A Pigoni**, et al., Olfactory and gustatory dysfunctions in SARS-CoV-2 infection: A systematic review. Brain, behavior, & immunity-health 2021, 15, 100268

P Enrico, G Delvecchio, N Turtulici, **A Pigoni**, et al., Classification of psychoses based on immunological features: a machine learning study in a large cohort of first-episode and chronic patients. Schizophrenia Bulletin 2021, 47 (4), 1141-1155

Pigoni A et al. Classification of first-episode psychosis using cortical thickness: A large multicenter MRI study. Eur Neuropsychopharmacol. 2021 May 3;47:34-47

LA Antonucci, N Penzel, **A Pigoni**, et al., Flexible and specific contributions of thalamic subdivisions to human cognition. Neuroscience & Biobehavioral Reviews 2021, 124, 35-53

G Delvecchio, E Maggioni, A Pigoni, et al., Sexual regional dimorphism of post-adolescent and middle age brain maturation. A multi-center 3T MRI study Frontiers in Aging Neuroscience 2021, 13, 622054

Pigoni A et al., Sex differences in brain metabolites in anxiety and mood disorders. Psychiatry Res Neuroimaging. 2020 Sep 21;305:111196.

H Walger, LA Antonucci, **A Pigoni**, et al., Basic symptoms are associated with age in patients with a clinical high-risk state for psychosis: results from the PRONIA study. Frontiers in Psychiatry 2020, 11, 552175

LA Antonucci, G Pergola, **A Pigoni**, et al., A pattern of cognitive deficits stratified for genetic and environmental risk reliably classifies patients with schizophrenia from healthy control subjects Biological psychiatry 2020, 87 (8), 697-707

Ganesh B Chand, Dominic B Dwyer, Guray Erus, Aristeidis Sotiras, Erdem Varol, Dhivya Srinivasan, Jimit Doshi, Raymond Pomponio, **Alessandro Pigoni**, Paola Dazzan, Rene S Kahn, Hugo G Schnack, Marcus V Zanetti, Eva Meisenzahl, Geraldo F Busatto, Benedicto Crespo-Facorro, Christos Pantelis, Stephen J Wood, Chuanjun Zhuo, Russell T Shinohara, Haochang Shou, Yong Fan, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, Nikolaos Koutsouleris, Daniel H Wolf, Christos Davatzikos. Two distinct neuroanatomical

subtypes of schizophrenia revealed using machine learning. Brain 2020, 143 (3), 1027-1038

Honors and Awards

12-15 September 2020: **ECNP Excellence Award**, 33rd ECNP Congress - virtual

April 2020: **Travel Award** for year 2020 Schizophrenia International Research Society (SIRS), SIRS Congress (online congress because of COVID19 pandemic)

18-22 June 2017: **WFSBP Young Investigator Award** the 13th World Congress of Biological Psychiatry in Copenhagen

2-4 November 2016 : ECNP Workshop on Clinical Research Methods Barcelona, Spain

14-18 June 2015: **Travel Grant Winner** for poster presentation at The 12th World Congress of Biological Psychiatry, Athene, Greece.

Abstract

Psychosis is a common and functionally disruptive clinical syndrome that might be present in many psychiatric, neurodevelopmental, neurologic, and medical conditions. Rather than a nosological entity, psychosis is a syndrome characterized by different symptoms and domains.

Therefore, an increasing amount of pointed out the importance of recognising and treating a first episode as soon as possible. For these reasons, first episode psychosis (FEP) rapidly became a very important population of study and assessment. More than just the first symptomatic presentation of a disease, FEP often shows already some of the features of the advanced psychiatric illnesses, although to a minor extent. On the other hand, great efforts are being made in order to establish an effective intervention, given the fact that early treatment has been proved to ameliorate the course of the disease, ranging from symptoms, relapse, and number of hospitalisations to quality-of-life measures such as involvement in school or work and global functioning.

Given the multifactorial nature of FEP and the different trajectories it can follow (e.g., affective vs. non-affective psychosis), the possibility of predicting future trajectories and

to obtain clear and valid biomarkers is becoming of paramount importance.

Prediction modelling has the potential to revolutionize medicine by predicting individual patient outcome. Early identification of those with good and poor outcomes would allow for a more personalised approach to care, matching interventions and resources to those most at need.

Through a series of studies, we explored: 1) the possibility to stratify FEP patients based on neuroimaging and biological measures; 2) the possibility of use cutting edge machine learning techniques to improve classification and cluster subtypes of FEP patients; 3) the presence of autoimmune features in FEP in a multi-site study I had the opportunity to coordinate as Co-PI (namely the PHLAMES study).

Specifically, in single-site studies we showed that neuroimaging and biological variables can be predictive of the course of the disease. Moreover, in large multi-site bid data analyses we presented how machine learning can improve the prediction of the disorder and help in stratify the risk, using both clinical and neuroimaging data. Finally, in the first results of PHLAMES study emerged that a subsample of FEP with

autoimmune characteristics might be defined. This subsample shows some unique features in terms of neurological symptoms, cognitive deficits, and brain imaging alterations.

The studies presented in this dissertation point out that it is possible to dissect the clinical and biological heterogeneity of psychosis at the beginning of its disease course, by defining meaningful groups of patients and therefore tailor personalized management. In conclusion, these data foster the research for subtypes of FEP and the definition of disease trajectories. These advances might have a great impact on patients' lives, by defining specific subgroups or progression that benefit of tailored interventions.

1. Introduction

1.1 First Episode Psychosis – presentation and challenges

Psychosis is a common and functionally disruptive clinical syndrome that might be present in many psychiatric, neurodevelopmental, neurologic, and medical conditions (Arciniegas, 2015). Rather than a nosological entity, psychosis is a syndrome characterized by different symptoms and domains. The DSM 5 (American Psychiatric Association, 2013) defined psychotic disorders as abnormalities of one, or more, of five domains: delusions, hallucinations, disorganized thinking (speech), grossly disorganized or abnormal motor behaviour (including catatonia), and negative symptoms (Gaebel and Zielasek, 2015).

Psychosis can occur in different mental disorders, ranging from Schizophrenia to Bipolar Disorder, that might share a considerable overlap in their clinical presentations. Moreover, psychosis can be caused by drugs, substances of abuse (typically psychostimulants such as cocaine and amphetamine) and their withdrawal, or organic brain disorders such as

delirium and encephalitis (Gaebel and Zielasek, 2015). Of note, psychotic disorders are listed in the top 20 leading causes of disability worldwide, since they are associated with significant societal and individual burden including years of lost productivity, higher mortality rates, and substantial cost to the health care system (Kane et al., 2015). Individuals with psychotic disorders have increased overall mortality; in addition, the 3-year period after first diagnosis of a psychotic disorder may be a high-risk time for suicide (Lally et al., 2017) highlighting the need for early recognition and management of these individuals.

Among psychiatric illnesses, the most common psychotic disorder (that was once considered prototypical) is schizophrenia. The core features of this disruptive disease are positive symptoms (psychotic symptoms like delusions and hallucinations), negative symptoms (impaired motivation, reduced speech, and social withdrawal), and cognitive impairment over a range of cognitive domains, such as attention, working memory and executive functions (Owen et al., 2016). Although the positive symptoms tend to have a remit-and-relapse course, the negative and cognitive symptoms are usually chronic and are associated with the most invalidating

long-term effects on the patients' functioning (Owen et al., 2016). Interestingly, the age of onset of bipolar disorder is similar to the one of Schizophrenia and can present psychotic features, although the most common manifestation is a depressive phase (Axelson et al., 2011). Although considered less severe compared to Schizophrenia, this mood disorder characterized by swings between euphoric (mania) and depressive episodes, can manifest psychosis in more than 70% of patients (Goodwin and Jamison, 2007) and be the cause of severe impairment of functioning and cognitive deficits (Grande et al., 2016). When the first presentation shows psychotic features, the differential diagnosis with Schizophrenia can be challenging.

First-episode psychosis (FEP) is defined as the first occurrence of delusions, hallucinations, or psychic disorganization of significant intensity, and duration of more than 7 days (Marques et al., 2016). The first episode of the disease usually occurs in late adolescence or early adulthood (usually in the age between 15 and 30 years), but it is frequently preceded by a prodromal phase or a so-called at-risk mental state (Addington and Heinssen, 2012). Although a sudden onset is possible, premorbid impairments in cognition or social functioning, or

both, are common and can manifest years before the first clear psychotic episode (Lewandowski et al., 2011).

Neurocognitive and social cognitive deficits are present early in the course of both diseases (Bora and Murray, 2014; Stouten et al., 2017) and seem not to be dependent on symptomatic status (Rund et al., 2016). Recent meta-analysis proposed that neurocognitive deficits are already established before the early stages of psychosis (Bora and Murray, 2014) and are among the predictors of functional outcome and psychosocial functioning, especially in interpersonal functioning and academic/vocational performance (Malla and Payne, 2005; Bratlien et al., 2013). By the time of their first assessment first-episode psychosis (FEP) patients present with reduced scores in verbal memory, attention and psychomotor speed domains (Nuechterlein et al., 2006; Castro-de-Araujo and Kanaan, 2017).

Therefore, an increasing amount of literature (Altamura et al., 2015; Sullivan et al., 2019) pointed out the importance of recognizing and treating a first episode as soon as possible. The Duration of Untreated Psychosis (DUP) could indeed impact on the progression of the disease (Altamura et al., 2015) and the general functioning of the individual (Sullivan et al., 2015), although the biological basis of these findings is not clear yet,

and mixed results are reported (Bora et al., 2018). Moreover, the study of FEP might help unveil the etiophatologic roots of the disease, since the cumulative effects of confounding variables, such as antipsychotic treatment and chronicity of the disease, are lower (Fusar-Poli et al., 2013; Vita et al., 2015).

For these reasons, FEP rapidly became a very important population of study and assessment. More than just the first symptomatic presentation of a disease, FEP often shows already some of the features of the advanced psychiatric illnesses, although to a minor extent (Stouten et al., 2017). On the other hand, great efforts are being made in order to establish an effective intervention, given the fact that early treatment has been proved to ameliorate the course of the disease. Indeed, any gains made in this early phase have long-term benefits on the further illness course (Albertand Weibell, 2019). Recent trials and meta-analyses highlighted the important effect of early intervention on basically all outcomes, ranging from symptoms, relapse, and number of hospitalizations to quality-of-life measures such as involvement in school or work and global functioning (Fusar-Poli et al., 2017; Correll, 2018; Albertand Weibell, 2019).

The current gold standard is a multidisciplinary approach that tackles not only the psychotic symptoms but also the cognitive aspects and the environmental issues, helping the subjects and the family (Albertand Weibell, 2019). Treatment should be therefore team-based and should deliver a multi-component intervention that includes, aside from medical treatment, family involvement, social skill training, assertive treatment, cognitive behavioral therapy, vocational and educational counselling, crises response team and management (Correll, 2018).

Given the multifactorial nature of FEP and the different trajectories it can follow (e.g., affective vs. non-affective psychosis), the possibility of predict future trajectories and to obtain clear and valid biomarkers is becoming of paramount importance.

Prediction modelling has the potential to revolutionize medicine by predicting individual patient outcome. Early identification of those with good and poor outcomes would allow for a more personalized approach to care, matching interventions and resources to those most at need. Despite the effort, recent reviews (Lee et al., 2022) have pointed out that prediction models are still lacking consistency and no clear biomarker has emerged (Dunleavy et al., 2022).

1.2 How to study First Episode Psychosis

1.2.1 Neuroanatomical Findings in First Episode Psychosis

Grey matter alterations have been consistently reported in literature involving FEP and early-in-the-course Schizophrenia patients (Shah et al., 2017). Although psychotic disorders are usually considered progressive, some abnormalities in the patients' GM seem to be present already in the beginning of the disease, when chronic course and antipsychotic treatment have less or no impact (Shah et al., 2017; Calvo et al., 2019).

A recent meta-analysis (Shah et al., 2017), that assessed more than 800 treated and drug-naïve patients, showed GM abnormalities encompassing the right anterior cingulate/paracingulate gyrus, left superior temporal gyrus, left hippocampus, right gyrus rectus and right insula. Interestingly, these patterns were overlapping between treated and drug-naïve subjects, indicating that their presence is rather specific and not really influenced by the short duration of antipsychotic medications. This finding suggests that these changes might be primarily due to a genetic risk for the disease rather than to an exposure to environmental factors (Costain et al. 2010).

Abnormalities in temporo-frontal regions and insula are well documented in both Schizophrenia (Walton et al., 2018; Alnæs et al., 2019) and bipolar disorder (Altamura et al., 2017; Hibar et al., 2018). Temporal areas have been involved in the onset of hallucinations (especially verbal) (Hugdahl et al., 2017) and have been associated with cognitive and verbal alterations in psychiatric disorders (Zhang et al., 2018). Cortical thinning in these areas is often present also in substance-induced psychotic disorders, which suggests that it might represent a biological marker of psychosis (Tosato et al., 2012; Altamura et al., 2017).

A reduction of GM in frontal regions is a common finding in both Schizophrenia and Bipolar Disorder (Yang et al., 2019). Regarding Schizophrenia, one of the most specific regions of cortical thinning is the dorsolateral prefrontal cortex (Fan et al., 2019). This area has been associated to working memory, a cognitive domain that often presents a deficit in Schizophrenia (Fan et al., 2019). Both temporal and frontal abnormalities in Schizophrenia might also represent a part of fronto-temporal dysconnectivity, which is considered the basis of the disorder (Zhou et al., 2015). Moreover, prefrontal alterations in both structural and functional imaging have been reported for bipolar as well (Altamura et al., 2017; Hibar et al., 2018).

Specifically, abnormalities in regions within the fronto-temporo-limbic network (encompassing amygdala, cingulate and prefrontal cortices), which is known to be involved in emotional processing and regulation, are considered a common finding in affective disorders (Wolfensberger et al., 2008; Miskowiak et al., 2015).

Regarding FEP, recent literature (Calvo et al., 2019) reported the presence of common and distinct patterns of GM alterations between affective and non-affective FEP. Specifically, while the GM volume deficits in fronto-temporal areas and anterior cingulate cortex seem to characterize both disorder's trajectory, amygdala abnormalities have been reported mainly in affective first-episode patients while hippocampal and insular shrinkage seems to be the neurobiological signature of non-affective psychosis in the early stages.

2. Single center studies

2.1 Insula volumes in first-episode and chronic psychosis: A longitudinal MRI study (Pigoni et al., 2022)

This paper has been published on Schizophrenia Research. Figure 1 has been selected by Schizophrenia Research as cover image for the Volume 241

2.1.1 Introduction and rationale

The diagnosis and clinical management of major mental disorders are still founded on psychopathological knowledge (Carvalho et al., 2020), causing delays in the diagnostic process and treatment definition, which in turn determines diagnostic uncertainty and lack of reliability on early clinical interventions. Identification of objective biomarkers would help differentiate more specific diagnostic groups and would facilitate the definition of therapeutic approaches, therefore delivering more tailored interventions (Dusi et al., 2017; Pigoni et al., 2019).

In particular, first episode psychoses (FEP) often undergo a long follow-up period before a stable diagnosis and treatment are defined (McGorry, 2015). In this context, magnetic resonance imaging (MRI) can offer a practical tool for a better characterization of FEP patients and a deeper investigation of the neurobiological processes determining their psychopathology (Calvo et al., 2019; Rokicki et al., 2020). In particular, structural MRI studies provided key evidence that could orient clinicians toward a differentiation between affective and non-affective psychoses, already at the early stages of illness (Satterthwaite et al., 2016).

Interestingly, when a stable diagnosis is reached, both bipolar disorder (BD) and schizophrenia (SCZ) patients show grey matter (GM) abnormalities, with SCZ usually presenting more severe alterations, especially in prefrontal and temporal areas and in subcortical regions, such as the insula, thalamus, and amygdala (Maggioni et al., 2017). Although the progression of these changes during the illness history has not yet been completely understood, GM alterations has been found to be associated with the age of onset (Pina-Camacho et al., 2016; Torres et al., 2016), the number of hospitalizations, and illness severity (Rosa et al., 2015; Vita et al., 2012). These suggest that a

remitting-relapsing course might negatively affect brain measures, as well as the duration of antipsychotic treatment (Hashimoto et al., 2018; Vita et al., 2015). However, the effect of these factors on the brain is still difficult to disentangle from the standard progression of the disorders, and therefore the employment of FEP patients with little or no pharmacological treatment might be paramount to identifying the putative brain alterations associated with psychosis, and to at least reduce the effect of the described confounders partially.

In an effort to identify common patterns, rather than differences across main psychiatric diagnoses, the insula and anterior cingulate cortices were consistently found to be altered in both the BD and SCZ (Goodkind et al., 2015), as well as in FEP patients (Díaz-Caneja et al., 2019) and subjects at high risk for psychosis (Chung et al., 2019).The patients usually show GM reductions in these regions compared to controls.

Specifically for the insula, this structure is a paralimbic multimodal cortical region that is highly connected to frontal, parietal and temporal lobes, limbic structures, and other subcortical areas, such as the caudate nucleus, the putamen, and the thalamus (Ferenczi et al., 2016; Leong et al., 2021). While the posterior insula is involved in the modulation of external

and interoceptive stimuli, the anterior insula is responsible for the integration between internal and external inputs and cognitive and emotional responses, through its connections with the anterior cingulate, the prefrontal cortex, and the amygdala (Ferenczi et al., 2016; Leong et al., 2021). Thus, this area might be a putative target for identifying early brain alterations among FEP compared to stabilized SCZ and BD patients. However, the trajectories of insular volume in patients with both affective and non-affective illnesses have not been defined yet, and to the best of our knowledge, a direct longitudinal comparison between FEP, BD and, SCZ patients and healthy controls (HC) on insular volume has not been performed yet.

Therefore, in this study, we followed up for ≃3 years, three populations of patients with FEP, SCZ, and BD and one group of HC. Insular volume was measured longitudinally with the final aim of understanding the role of this brain area in the development of psychosis, in its early and chronic stages. Specifically, we will explore a) differences in insula GM volumes at baseline between patients and HC, b) the progression of GM changes in the insular subregions from baseline to follow-up, to model specific disease trajectories

associated with diagnoses, and c) whether clinical and demographic variables are associated with GM modifications over the follow-up period, to try to disentangle the effects of such confounders from the standard progression of psychotic disorders. Our main hypothesis is that chronic BD or SCZ patients will show significant insular volume differences at baseline compared to FEP patients that will likely reduce at follow-up, given that the most prominent brain modifications tend to occur in the first years of the disease (Vita et al., 2012). Finally, we also hypothesized the presence of different and unique volumetric trajectories for affective and non-affective psychoses.

2.1.2 Methods

Participants

The sample included 68 FEP patients, 68 schizophrenia patients, 47 bipolar disorder patients and, 94 healthy controls. FEP patients were recruited as participants in the PICOS project (Psychosis Incident Cohort Outcome Study) (For more details about the project, please refer to our prior publications

(Lasalvia et al., 2012; Squarcina et al., 2015) and the Supplementary Materials).

Patients with SCZ and BD were recruited through the South-Verona Psychiatric Case Register (PCR) (please refer to our previous studies:(Ferro et al., 2017), within the FIRST project as per our prior publications (Enrico et al., 2021). For all patients, a formal ICD-10 diagnosis was obtained and confirmed by the item group checklist (IGC) of the Schedule for clinical assessment in neuropsychiatry (SCAN). Exclusion criteria were: (1) comorbid axis I psychiatric disorders, (2) alcohol or substance abuse within the 6 months prior to MRI, (3) history of traumatic head injury with loss of consciousness, (4) major neurological medical illnesses, or (5) mental retardation.

HCs were recruited through word of mouth and advertisements in the geographically defined catchment area of South Verona. They had no history of head injury or psychiatric disorder, this was determined by using a brief modified version of the Structured Clinical Interview - Nonpatient version (First et al., 2002), no history of psychiatric disorder among first-degree relatives, no history of alcohol or substance misuse, and no current major medical illness.

At baseline, all subjects underwent a 1.5T MRI evaluation and 99 subjects (10 HC, 23 BD, 29 SCZ, 37 FEP) were rescanned a second time after a mean time of 2,53 ± 1,68 years. All subjects signed informed consent and The PICOS and FIRST projects were approved by the ethics committees of the coordinating center (Azienda Ospedaliera Universitaria di Verona; PICOS [registration n: 1103], FIRST [registration n: 954]). The clinical symptomatology of all patients was evaluated using the 24 item Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham, 1962). An assessment of the day-life impact of mental diseases was performed with the Global Assessment of Functioning Scale (GAF) (Endicot et al., 1976). Data on pharmacological treatment was collected for all patients. The demographic and clinical details of all the subjects at baseline and follow-up are presented in Table 1 and Table 2, respectively.

Table 1: Descriptive statistics of the baseline sample.

	Healthy controls (N=94)	SCZ patients (N 68)	BD patients (N 47)	FEP patients (N68)	Statistics	
Age (years), range (min-max)	38.66±11.53 (20 –61)	39.36±11.92 (18 – 62)	47.78±9.53 (27 – 64)	30.64±9.19 (18 - 64)	$F_{3.273}$ =23.76	**p<0.001**
Gender: M/F	48/46	44/24	17/30	38/30	χ^2_3=9.45	**p=0.024**
Ethnicity	Caucasian (all)	Caucasian (all)	Caucasian (all)	Caucasian (65) Asian (3)		
Handedness (% right)	80.0 %	84.62 %	95.35%	80. 0 %	χ^2_3=5.8	**p=0.123**
BPRS total score		43.38± 16.89 (N 58)	31.29± 6.88 (N 17)	30.64±8.87 (N 59)	$F_{2.131}$= 15.83	**p<0.001**
Negative symptoms		11.20±4.00 (N 58)	4.78±3,58 (N 40)	8.62±2.77 (N 60)	$F_{2.155}$= 40.62	**p<0.001**
Positive symptoms		11.07±6.26 (N 58)	6.94±3.11 (N 17)	6.34±1.93 (N 59)	$F_{2.131}$= 17.68	**p<0.001**
GAF scores	77.18±5.46 (N 81)	52.46±11.83 (N 60)	58.39±11.96 (N 41)	60.52±17.20 (N 57)	$F_{3.235}$= 57.11	**p<0.001**
Atypical antipsychotics (YES/NO)	0/93	44/24	19/27	3/63	χ^2_3=111.82	**p<0.001**
Typical antipsychotics (YES/NO)	0/93	24/44	9/37	5/62	χ^2_3= 44.658	**p<0.001**
Mood stabilizers(YES/	0/93	24/44	9/37	5/62	χ^2_3= 44.658	**p<0.001**

BD, Bipolar Disorder, BPRS, Brief Psychiatry Rating Scale, FEP first-episode psychosis, GAF, Global Assessment of Functioning, SCZ, schizophrenia.

Table 2: Descriptive statistics of the follow-up sample.

	Healthy controls (N=10)	SCZ patients (N 28)	BD patients (N 23)	FEP patients (N 38)	Statistics	
Age (years) , range (min-max)	39.40±12.13 (22 - 57)	44.21±11.23 (22 – 65)	50.65±8.75 (33 –65)	33.14 ± 9.48 (20 - 54)	$F_{3.95}$= 14.47	**p<0.001**
Gender: M/F	2/8	20/9	7/16	22/15	χ^2_3= 12.539	**p=0.005**
Ethnicity	Caucasian (all)	Caucasian (all)	Caucasian (all)	Caucasian (37) Asian (1)		
Handedness (% right)	66.6%	90 %	94.74%	75.68%	χ2=5.99	**p=0.11**
BPRS total score		38.06± 9.38 (N 16)	27± 5.2 (N 3)	28.86±4.84 (N 36)	$F_{2.52}$=11.85	**p<0.001**
Negative symptoms		10.2±2.52 (N 16)	4±5.19 (N 19)	7.78 ±1.33 (N 36)	$F_{2.68}$=18.79	**p<0.001**
Positive symptoms		5.78±4.79 (N 23)	7.33±4.04 (N 3)	5.97±1.61 (N 36)	$F_{2.59}$=0.3	**p=0.742**
GAF scores	79±6.45 (4 missing)	48.11±12.89 (0 missing)	62.22±13.9 (N 22)	61.57±13.38 (N 37)	$F_{3.89}$=12.09	**p<0.001**
Atypical antipsychotics (YES/NO)	0/10	17/11	7/16	1/36	χ^2_3= 29.863	**p<0.001**
Typical antipsychotics (YES/NO)	0/10	10/18	4/19	4/33	χ^2_3= 8.0913	**p=0.042**
Mood stabilizers(YES/NO)	0/10	0/27	14/7	2/2	χ^2_3= 29.633	**p<0.001**

BD, Bipolar Disorder, BPRS, Brief Psychiatry Rating Scale, FEP first-episode psychosis, GAF, Global Assessment of Functioning, SCZ, schizophrenia.

MRI procedure

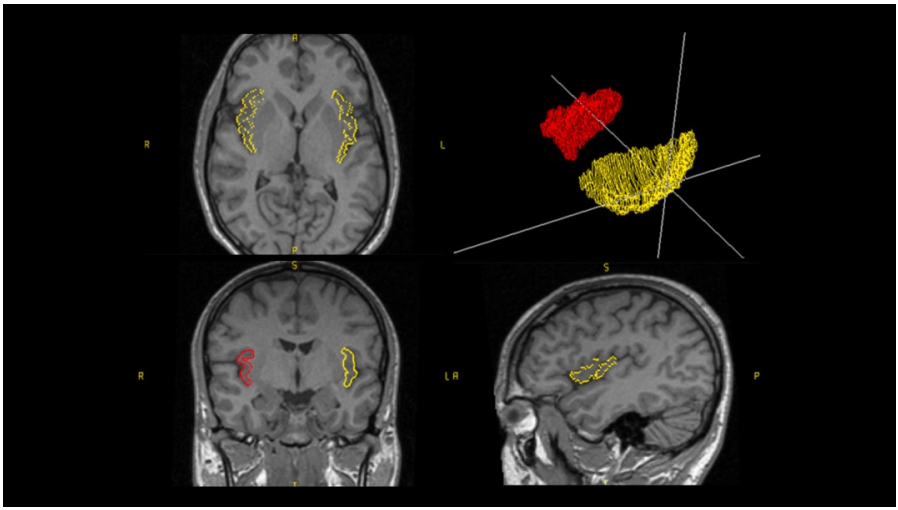
MRI scans were acquired with a 1.5-T Siemens Magnetom Symphony Maestro Class, Syngo MR 2002B, at the Borgo Roma University Hospital of Verona. All participants were provided with earplugs to reduce acoustic noise; a head holder was used to maintain patients stable, in order to minimize movement artifacts. Initially, exploratory T1-weighted images (Repetition Time (TR) = 450 ms, Echo Time (TE) = 14 ms, Flip Angle (FA) = 90°, Field Of View (FOV) = 230 × 230, slice thickness = 5 mm, matrix size = 384 × 512) were obtained to verify subject's head position and image quality. Proton density and T2-weighted images were then acquired (TR = 2500 ms, TE = 24/121 ms, FA = 180°, FOV = 230 × 230, slice thickness = 5 mm, matrix size = 410 × 512) according to an axial plane parallel to the anterior–posterior commissures, to exclude focal lesions. Subsequently, a coronal 3D magnetisation prepared rapid gradient echo (MP-RAGE) sequence was acquired (TR = 2060 ms, TE = 3.9 ms, FA = 158°, FOV = 176 × 235, slice thickness = 1.25 mm, matrix size = 270 × 512, TI = 1100 ms) to obtain 144 images covering the entire brain.

Image analysis

All imaging data were analyzed using the BRAINS2 software (https://medicine.uiowa.edu/psychiatry/research/iowa-neuroimaging-consortium/facilities/software-image-analysis). The software allows the manual tracing of a region of interest (ROI) on the coronal plane of MR images. The insular cortex was defined by the superior (SCIS) and inferior circular insular sulci (ICIS). The fusion of the SCIS and the ICIS within the fundus of the Sylvian fissure was considered the posterior landmark of the insula. Rostrally, the boundary between the anterior insula and the orbitofrontal cortex (OFC) is not clearly defined since the ICIS does not extend rostral to the limen insulae. The orbitoinsular sulcus (OIS) was considered the topographic boundary between the anterior insula and the adjacent orbitofrontal cortex. The ROI was then divided into an anterior and posterior portion, which reflect different cytoarchitectonics and connections, along the central insular sulcus (Ferenczi et al., 2016; Leong et al., 2020). Intracranial volume (ICV) was traced in the coronal plane along the border of the brain and included the cerebrospinal fluid, dura mater, sinus, optic chiasma, brainstem, and cerebral and cerebellar matter. The inferior border did not extend below the base of the

cerebellum. The first and last 10 slices including brain matter were traced; afterwards 1 slice on every 5 was traced. Tracing was performed by a rater (FC) blind to both subjects' identity and sociodemographic and clinical variables. The rater achieved a high reliability with a second skilled rater, as defined by intraclass correlation coefficients (ICCs) of 0.90 for insula and ICV, established by blindly tracing 10 randomly selected scans.

Figure 1. ROI tracing of insular volume with its 3D reconstruction.



Statistical analysis

Statistical analyses were conducted using the R software (https://www.R-project.org). Demographic and clinical variables were compared using Pearson's chi-square and one-way analysis of variance (ANOVA), where appropriate. We used a multiple testing correction method for structured hypotheses for the analysis of insular volumes, called the inheritance procedure (Goeman and Finos, 2012). Firstly, the

procedure applied multivariate analysis of variance (MANOVA) test, using insula volumes (i.e., left/right anterior/posterior) as response variables; diagnostic groups such as age, gender, handedness, and total intracranial volumes were used as covariates. If the multivariate analyses for diagnostic groups were significant at a significance level of $p = 0.05$, two MANOVAs were fitted, one for the anterior insula and one for the posterior insula. The significance level was then set at $p = 0.05/2 = 0.025$. When an area results significantly, each sub-area was used as the dependent variable in a linear model. If the diagnostic group covariate is significant at level $p = 0.05/K$, where K equals the number of MANOVAs found significant, post-hoc analysis with Holm correction was performed for the four diagnostic groups. Briefly, the Inheritance Procedure guarantees the familywise error rate (FWER) at level $\alpha = 0.05$ in multiple testing analyses, i.e., 3 MANOVAs. In addition, we adjusted for multiple comparisons using the Holm method and for multiple testing using the Bonferroni correction by setting the significance level at $p = 0.05/K$, since K regressions have been performed.

For the longitudinal changes in volumes, the same method was applied, using as a dependent variable the delta volumes (ΔV)

of each part of the insula, i.e., insula volume at T1 – insula volume at T0. Finally, correlations between sociodemographic (i.e., age) or clinical features (i.e., Global Assessment of Functioning (GAF), Brief Psychiatry Rating Scale (BPRS), length of illness, antipsychotics lifetime, pharmacological treatment, and ΔV of the four insular portions at baseline and follow-up were carried out by an ANOVA based on a linear model with age and gender as covariates; Bonferroni correction was applied. In addition, Cohen's f was computed for measuring the effect and size of the regression with the following interpretation: small (0.02), medium (0.15), and large (0.35).

For all analyses, the coefficients representing the predicted values of the insular volumes for each group after controlling for all the covariates have been shown in Figures 2-5. Finally, the distributions of the insular volumes (reported in the Supplementary Materials) were not perfectly normal, but the analysis of the residuals validated the regression model used.

2.1.3 Results

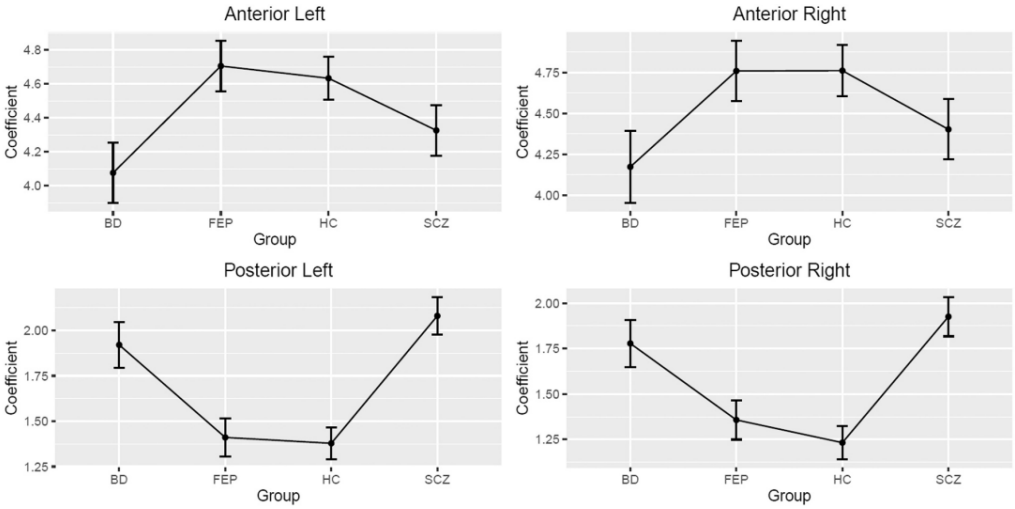
Baseline

Baseline results are shown in Figure 2. Since the first step MANOVA was significant ($F(12, 737) = 12.488$, $p < 0.001$), a second step MANOVA was performed and was also found significant for both insula portions (anterior insula GM, $F(6, 490) = 4.5.119$, $p < 0.001$; posterior insula GM, ($F(6, 490) = 26.517$, $p < 0.001$)). Subsequently, a linear regression was carried out with each subdivision of insular GM volume – anterior left, anterior right, posterior left, posterior right – as the dependent variable and the diagnostic groups, age, gender handedness, and total intracranial volumes as covariates. Then, for each significant linear model, a post-hoc analysis with Holm correction for multiple comparisons was performed. It revealed statistically significant differences in left anterior insular GM volume in HC vs SCZ ($t = 4.305$, $p < 0.001$), and HC vs BD ($t = 3.863$, $p = 0.003$) with greater volume in the patients' group, as well as in FEP vs SCZ ($t = 3.353$, $p = 0.02$) with greater volume in the FEP group.

Significant differences were found in right anterior insular GM volume between HC and SCZ ($t = 4.08$, $p = 0.001$), with patients

having greater volume. Analyzing the left posterior insular GM volume, significant differences were found between HC vs SCZ ($t= -11.263$, $p < 0.001$), HC vs BD ($t = -8.813$, $p < 0.001$), FEP and SCZ ($t = -10.27$, $p < 0.001$) and FEP and BD ($t =-7.226$, $p < 0.001$), with HC and FEP having lower GM volumes. Finally, in the right posterior insular GM volume significant differences were found between HC and SCZ ($t = -10.018$, $p < 0.001$), HC and BD ($t = -5.863$, $p < 0.001$) with patients having lower GM volumes, and between FEP and SCZ ($t = -7.162$, $p < 0.001$), FEP and BD ($t = -3.614$, $p = 0.007$), with FEP having lower GM volumes.

Figure 2: Significant mean differences in grey matter volumes of insular subportions at baseline between patients with bipolar disorder (BD), schizophrenia (SCZ), first-episode psychosis (FEP) and healthy controls (HC).

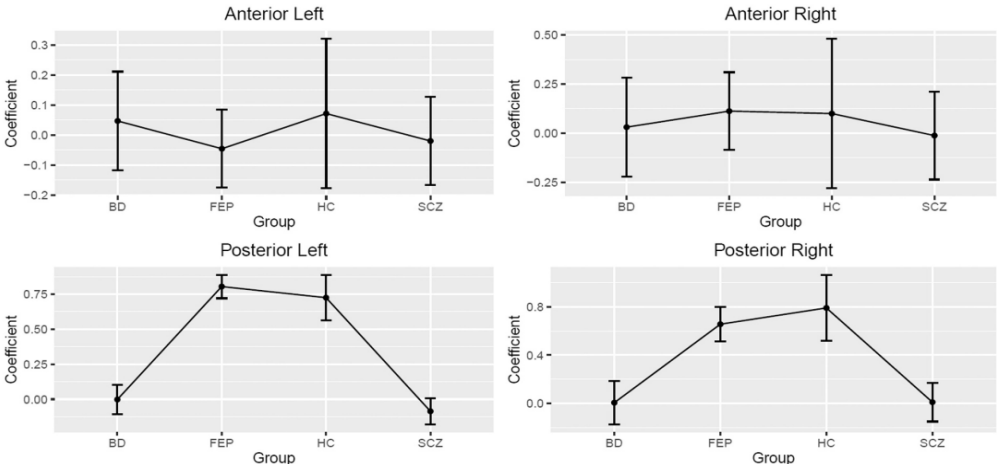


Longitudinal analysis

The longitudinal analysis replicated the multi-step procedure of the baseline analyses using as dependent variables the volumetric change (Δ) parameters. Longitudinal results are shown in Figure 3.

In addition, in this case, since the first step MANOVA was significant ($F(12, 210) = 6.721$, $p < 0.001$), a second step MANOVA was carried out. The MANOVA reached statistical significance only for the posterior insula ($F(6,142) = 15.614$, $p < 0.001$). Therefore, linear regression analyses were performed only for the two posterior insular regions (posterior left and posterior right). Then, post-hoc analysis with Holm correction for multiple comparisons was performed for each significant linear model. Post-hoc analyses revealed statistically significant differences in left posterior insula GM ΔV, with greater values in FEP compared to BD ($t= 11.013$, $p < 0.001$) and to SCZ ($t= 12.174$, $p < 0.001$) as well as between HC vs SCZ ($t = 8.017$, $p < 0.001$) and HC vs BD ($t = 6.749$, $p < 0.001$), with patients having greater ΔV. Further, statistically significant differences were observed in right posterior insula ΔV, with greater values in HC compared to both SCZ ($t= 3.526$, $p=0.007$) and BD ($t= 2.974$, $p = 0.038$) as well as between FEP vs SCZ ($t = 4.653$, $p = 0.002$) and FEP vs BD ($t = 4.108$, $p = 0.001$), with FEP having greater ΔV.

Figure 3: Significant mean differences in ΔV of insular subportions between patients with bipolar disorder (BD), schizophrenia (SCZ), first-episode psychosis (FEP) and healthy controls (HC).



Additional analyses

The multi-step procedure was applied to analyze the difference in insular GM volumes between FEP at follow-up and SCZ/BD at baseline. The results are shown in Figure 4.

Since the first step MANOVA was significant ($F(8, 268) = 3.336$, $p = 0.001$), the second step was performed. The MANOVA

reached statistical significance for the anterior insula ($F(4, 272) = 3.123$, $p = 0.031$), and for posterior insula ($F(4, 272) = 3.355$, $p = 0.02$). Therefore, the four linear regressions were carried out, one for each insular GM volume. The post-hoc analysis found a significant difference between FEP and BD ($t = -3.013$, $p = 0.034$) in the left anterior region, with FEP having greater volumes. Finally, a significant difference was found between FEP and BD in the right posterior insula volumes ($t = -3.562$, $p = 0.006$) with FEP having greater volumes. No differences have been observed in any of the other insular regions.

The same multi-step procedure was performed to study the differences in the volumetric change (ΔV) parameters between affective and non-affective FEP patients. The MANOVA with the ΔV of GM of the four insular portions (left, right, anterior, posterior) as dependent variables, and diagnostic groups, age, gender, handedness, and total intracranial volumes as covariates were not significant ($F(4,64) = 0.663$, $p = 0.619$). Moreover, we also investigated the differences in insular GM volumes, both at baseline and by using the volumetric change (ΔV) parameters, between the whole group of patients with affective (BD and affective FEP) and non-affective (SCZ and non-affective FEP) psychosis using the same multi-step

procedure (Figure 5). Analyzing the four insular GM volumes at baseline, the group covariate in the first MANOVA was significant ($F(8,488)$, $p < 0.001$), as well as considering only the left volumes ($F(4,492)$, $p = 0.005$) and the right volumes ($F(4,492$, $p < 0.001$) whereas by using the ΔV parameters as dependent variables, no significant differences were observed ($F(8, 140)$, $p = 0.09$). Therefore, the four linear regression analyses only for the baseline results were performed. The results showed a significant difference between HC and patients with affective psychoses in all four portions of the insula (left anterior ($t = 3.199$, $p = 0.01$), right anterior ($t = 3.077$, $p = 0.026$), with patients showing lower GM volumes compared to HC, and left posterior ($t = -4.923$, $p < 0.001$), and right posterior ($t = -6.697$, $p < 0.001$)), with patients showing lower GM volumes compared to HC. In addition, we found significant differences in the posterior insula bilaterally between HC and patients with non-affective psychoses (left ($t = -6.697$, $p < 0.001$), right ($t = -7.398$, $p < 0.001$)), with patients having higher GM volumes compared to HC, and in the right ($t = 3.214$, $p = 0.017$) and left ($t = 3.119$, $p = 0.023$) anterior insula.

Figure 4: Significant mean differences in insular volumes between patients with bipolar disorder (BD), schizophrenia (SCZ) and first-episode psychosis (FEP), where FEP volumes refer to T1 and SCZ/BD volumes refer to T0.

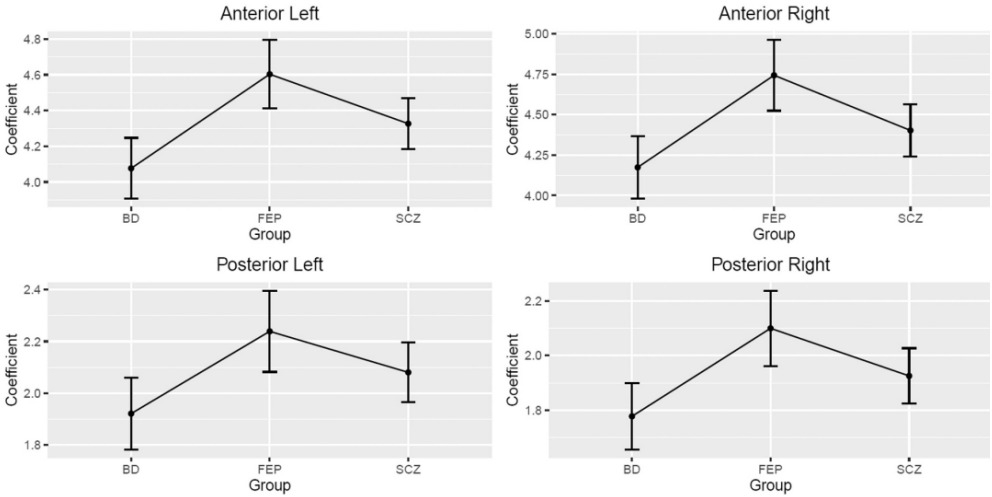
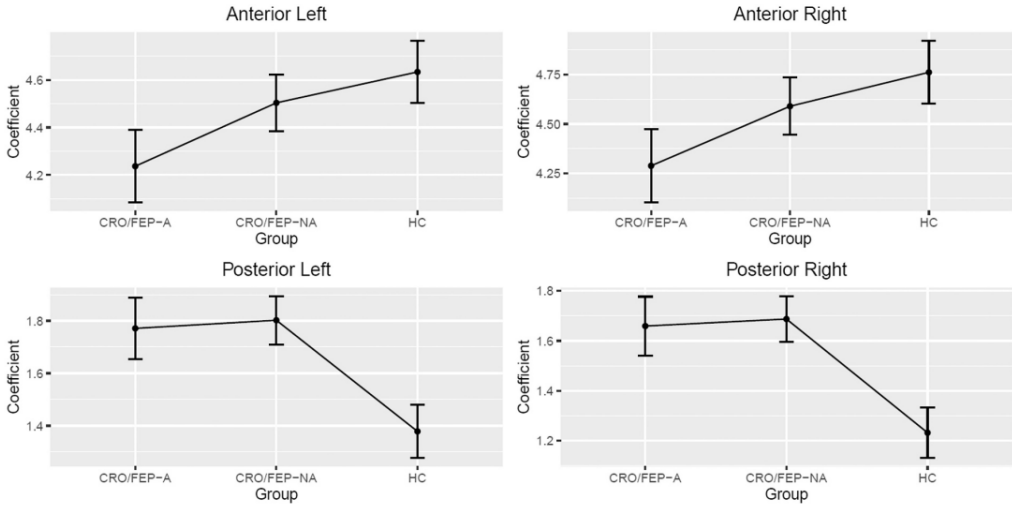


Figure 5: Significant mean differences in grey matter volumes of insular subportions at baseline between patients with affective psychoses (CRO/FEP-A), non-affective psychoses (CRO/FEP-NA) and healthy controls (HC).



Correlation analyses

Regarding medications, significant correlations were found for both first- and second-generation antipsychotics. Specifically, second generation antipsychotics showed positive correlations with the right (Effect size = 0.1477, $p = 0.01$) and left (Effect size = 0.3659, $p < 0.001$) posterior insula at baseline. Similarly, first generation antipsychotics showed positive correlations with the right (Effect size = 0.2018, $p < 0.001$) and left (Effect size = 0.4075, $p < 0.001$) posterior insula at baseline as well as negative correlations with right (Effect size = 0.707, $p = 0.012$) and left (Effect size = 0.6537, $p = 0.002$) anterior portions. Moreover, a significant negative correlation was detected between first generation antipsychotics and left posterior insula delta GM (Effect size = 0.2902, $p <0.001$).

Regarding clinical measures, GM of insula sub-portions significantly correlated with both GAF and BPRS scores at baseline. Specifically, GAF scores significantly positive correlated with the right anterior insula (Effect size = 0.6567, $p = 0.008$) and negative correlated with the right posterior (Effect size = 0.2634, $p < 0.001$) and left (Effect size = 0.3675, $p < 0.001$) posterior portions of the insula. Similarly, BPRS scores positively correlated with left posterior insula (Effect size =

0.4373, p = 0.001) at baseline. Finally, age at baseline showed a negative correlation with right anterior insula (Effect size = 0.5844, p <0.001, and left anterior insula (Effect size = 0.6103, p < 0.001).

2.1.4 Discussion

In this study, we examined insular volumes in patients with a diagnosis of SCZ, BD, FEP, and in a group of healthy controls both at baseline and at follow-up, to assess whether this region might represent a valid marker for the progression of distinct psychotic disorders. A complex pattern of GM modifications was observed both at baseline and follow-up. Specifically, at baseline, chronic patients showed increased GM in both anterior and left posterior portions of the insula, as well as a reduced GM volume in the right posterior portion compared to HC. Moreover, FEP had steeper baseline-to-follow-up modifications, reaching the baseline volumes of SCZ patients, with opposite directions for anterior and posterior portions of

the insula. Finally, significant correlations between insula GM volumes and age, antipsychotic medication, as well as a clinical and functioning measure were detected.

Baseline results

Results at baseline partially confirmed the available literature, suggesting that the insula can be regarded as one of the most affected regions in psychosis (Gupta et al., 2015; Nelson et al., 2020). Specifically, similar to our findings, recent studies also reported that the insula was found reduced in patients affected by SCZ (Gupta et al., 2015; Nelson et al., 2020) and BD (Choi et al., 2020) as well as in FEP patients (Escarti et al., 2019; Jung et al., 2019), although to a lesser extent compared to SCZ and BD.

In our sample, increased GM volumes in a sub-portion of the insula in chronic patients with both affective and non-affective psychosis were reported. Although this might be surprising at a first glance, it is now widely recognized that both SCZ and BD are not only characterized by a reduction in GM volumes but by more complex patterns of alterations (Chand et al., 2020; Voineskos et al., 2020). Indeed, recent studies (Voineskos et al., 2020) have also suggested that biological heterogeneity is a

characteristic of SCZ and psychotic disorders, which can be overcome by finer parcellations of GM or by overcoming average differences as with recent machine learning studies (Chand et al., 2020; Pigoni et al., 2021, 2019).

It is important to highlight that the majority of previous neuroimaging studies on insular volumes have treated this structure as a single, homogeneous region, or have only considered its anterior part (Choi et al., 2020; Gupta et al., 2015). In contrast, in our sample, we explored both the anterior and posterior insula, highlighting a different pattern of increased and decreased GM modification, as also reported by a recent study on SCZ (Tian et al., 2019). Interestingly, our results seem in line with a recent study suggesting a different involvement of insular subregions and a differential association of these subregions to both clinical and cognitive aspects in psychotic patients (Sheffield et al., 2021). Although the causes of this differential alteration of insular GM are not easy to draw, some possibilities may be considered in the context of the existing literature showing that anterior and posterior insula are part of different networks subserving different functions (Li et al., 2019). Specifically, the anterior insula is a key hub of the salience network, which is often found altered in psychosis

(Liang et al., 2021), possibly due to a persistent hyperdopaminergic state (Saviola et al., 2021), which leads the psychotic brain to a misleading assignment of object's salience. Therefore, it is not surprising that aberrant connectivity between the anterior portion of the insula and temporal areas was found to be associated with auditory hallucinations (Chen et al., 2020; Yang et al., 2020) and negative symptoms (Sheffield et al., 2020) in SCZ, ultimately suggesting that this region might be a key functional area in psychotic disorders.

On the other hand, the posterior region is functionally connected with primary and secondary somatomotor cortices, parieto-temporo-occipital lobe, and motor regions of the cerebellum (Tian et al., 2019). Although the literature regarding this specific subregion is sparse in psychiatry, a recent study reported a significant association between the posterior insula and measures of impulsivity in patients affected by BD (Lois et al., 2020), therefore also suggesting that this portion of the insula might play a role in psychiatric disorders. Of note, in the last few years, some studies suggested that these two subregions might be altered in different and opposite ways in psychosis and in clinical high-risk patients (Li et al., 2019; Tian et al., 2019). Specifically, Tian and colleagues (Tian et al., 2019)

reported enlargement of the posterior insula and shrinkage of the anterior in patients with SCZ, as also reported by our study. The causes and implications of these divergent alterations for psychotic disorders are not clear but could be related to the disrupted dopaminergic tone at the basis of psychoses (Saviola et al., 2021) that lead to an "overstimulation" of some areas compared to others. Additionally, a recent study (Rössler et al., 2020) showed that psychotic-like experiences are associated with functional disruption of dopamine afferents to the anterior insula, which is considered one of the possible causes of both affective and non-affective psychosis (McCutcheon et al., 2020). Therefore, the mixed picture of increased and decreased GM volume seems to further support the hypothesis that the anterior and posterior insula are differently altered in psychotic disorders, possibly because they subserve different functions.

Interestingly, in our samples, such alterations are not present in the FEP group at baseline. This finding might be related to several factors. One of the possible reasons could be related to the fact that GM differences in FEP might still be subtle and not clearly evident from group-level analyses (Calvo et al., 2019; Tordesillas-Gutierrez et al., 2015) since they are less affected by the illness duration and long-term treatments (Hashimoto et al.,

2018; Tomelleri et al., 2009). Moreover, differences in demographic and clinical features might also explain the differences between FEP and SCZ/BD at baseline. Indeed, age is known to affect GM both in healthy subjects and in the clinical population (Delvecchio et al., 2021) and similar influences have also been described for the duration of illness (Smieskova et al., 2012). These aspects have been further confirmed by our findings, which showed an overall negative influence of age on the anterior portion of the insula. Moreover, given the nature of our study, FEP patients are significantly younger and have a shorter duration of illness, and the cumulative effect of these variables on GM might not be clearly visible at baseline. Similarly, the different exposures to psychotropic drugs on GM volumes should also be considered since chronic patients have been treated for significantly longer periods. Even though evidence on the effects of antipsychotic therapy on insula GM volumes in patients with SCZ and BD are still mixed (Roiz-Santiañez et al., 2015), a significant correlation between antipsychotic treatment and insular volumes was observed in our sample (see Supplementary Materials), therefore supporting the hypothesis that different treatments' duration and dosage might partially explain the negative findings observed in FEP at baseline.

Longitudinal results

Longitudinal results showed steeper modifications in insular GM in FEP compared to chronic patients, with FEP reaching the baseline volumes of SCZ patients in most subregions.

This finding seems to be in line with previous literature reporting progressive GM changes characterizing psychotic disorders mostly occurring during the first years after the onset (Vita et al., 2012). However, there is considerable dispute as to whether such progressive brain changes are invariably present or are associated with and modulated by specific factors, such as symptoms, clinical outcome, antipsychotic treatments, and comorbidity (Zipursky et al., 2013) as well as age and age of onset (Pina-Camacho et al., 2016; Torres et al., 2016). In more details, it has been consistently reported that while chronic older patients showed on average more widespread GM and WM abnormalities in various cortical and subcortical regions (Morimoto et al., 2021; Torres et al., 2016), brain alterations in FEP patients were found to be usually subtle when compared to healthy individuals (Pigoni et al., 2021; Torres et al., 2016). Also, it has been demonstrated that the trajectories of GM

deficits seem to be influenced by an early age of onset in both affective and non-affective psychoses (Pina-Camacho et al., 2016; Torres et al., 2016).

Moreover, recent studies (Rosa et al., 2015; Saviola et al., 2021) suggest that GM modifications in the first years after a psychotic onset might be greatly influenced by disease severity and persistence of symptoms, as partially confirmed by the correlations observed between BPRS scores and insula GM in our sample (see Supplementary Materials). Among patients with a remitting course of illness, a pattern of increase in GM volumes in selective brain regions was reported, while non-remitters presented GM reductions (Rosa et al., 2015). It is also unclear whether brain volume changes are different in those who will further develop affective or non-affective disorders (Rosa et al., 2015). Indeed, although progressive brain abnormalities may occur both in affective and non-affective psychoses, different abnormalities may be detected among the two groups, since different mechanisms may underlie the progression of these disorders (Rosa et al., 2015). Although in our sample we did not find differences in GM trajectories in affective and non-affective FEP subjects, probably due to the small sample size at follow-up, some differences were detected

in both the anterior and posterior insula at baseline between non-affective FEP and chronic psychoses compared to HC, while no differences were reported for affective patients. Specifically, individuals with affective psychosis showed a significant lower GM volume in all subregions of insula when compared to HC; on the other hand, a significantly higher volume in the posterior portion is reported for non-affective patients, when compared to HC, in line with previous findings (Tian et al., 2019). Moreover, GM modifications in FEP patients observed at follow-up seem to point in the direction of a possible difference between the two groups of patients. Indeed, while GM volumes in some insular subregions of FEP patients at follow-up resembles the ones observed in SCZ patients at baseline, they still showed a statistical difference with the GM volumes of BD patients at baseline. Therefore, we can hypothesize a differential development in insula subregions in affective and non-affective psychotic disorders that could be the focus of future studies aiming at discovering the disease's trajectories.

Limitations

Our findings must be considered in light of some limitations. First, our sample size is small, even though it is in line with the numerosity of previous studies (Nelson et al., 2020; Rosa et al., 2015). Second, high rates of dropouts were observed between baseline and follow-up. Therefore, analyses based on longitudinal data might be influenced and needs to be replicated on larger samples. Third, we acknowledge that the sample size prevented us from investigating the effects of specific variables, such as the number of episodes or hospitalizations on brain changes during the follow-up. Finally, the study had a naturalistic design, with no control over the pharmacologic and non-pharmacologic treatments.

2.2 Association between brain metabolism and clinical and temperamental scales in first episode psychosis (Pigoni et al., in publishing)

2.2.1 Introduction and rationale

Understanding the molecular and structural mechanisms underlying clinical features of schizophrenia and other psychotic disorders is an important step in identifying biomarkers and potential targets for novel medications, treatment strategies, and interventions, especially for cognitive alterations, which are among the most disabling aspect of these disorders (Davison et al., 2018; Bora, 2019; Shang et al., 2022).

One of the neurological factors that has been implicated in schizophrenia is altered brain metabolism, which can be effectively measured non-invasively in vivo using magnetic resonance spectroscopy (1H-MRS) providing information about energy production and neurotransmitter levels.

Hypometabolism is thought to be an important component of the pathophysiology of psychotic disorders, potentially underlying cognitive and negative symptoms (McCutcheon et al., 2020; White-hurst et al., 2020). Evidence for this includes in vivo brain imaging studies of N-acetyl aspartate (NAA), in both chronic and first episode psychosis (FEP), whose concentration in the brain correlates positively with other measures of neuronal metabolic activity (Whitehurst et al., 2020).

A recent meta-analysis (Whitehurst et al., 2020) showed a wide hypometabolism in frontotemporal regions and anterior cingulate cortex in chronic schizophrenia and medicated FEP. How-ever, no significant differences were reported for unmedicated FEP compared to healthy individuals.

Similarly, a recent review (Squarcina et al., 2017) suggested a possible glutamatergic alteration in anterior cingulate cortex (ACC) in FEP, that seems to be associated with cognitive performances in neuropsychological tests (Dempster et al., 2015). However, the number of studies assessing metabolic alterations in FEP is still limited and often conflicting or negative results are reported (Borga et al., 2019).

The study of FEP patients is of particular interest since it allows the investigation of neurobiological processes, ruling out the effects of long-term medications and chronicity (Pigoni et al., 2021). In this regard, metabolite levels are reported to be particularly sensitive to the different phases of the disorder, as well as to several demographic and clinical variables, such as sex (Pigoni et al., 2020), medications (Musazzi et al., 2013; Whitehurst et al., 2020), and illicit drug use (Colizzi et al., 2016). Therefore, in order to understand the progression of the disease and to identify neurobiological underpinning of cognitive deficits and treatment response, it is important to evaluate FEP from the early phases of the disease and take into account potential confounders affecting brain metabolism.

In this framework, the present study aims at evaluating the possible association between brain metabolites levels and clinical and temperamental scales in a sample of FEP and healthy controls (HC). The quantified metabolites included phosphocreatine plus creatine (PCr+Cr), glycerophosphocholine plus phosphocholine (GPC+PC), myo-inositol, glutamate, and N-acetylaspartate (NAA).

Our hypothesis is that metabolic differences might be present already at early stages of the disorder, even though to a lesser extent, and might be associated with specific behavioral traits.

2.2.2 Methods

2.2.2.1 Participants

Sixty FEP patients (37 males, 29.3 ± 9.4 y.o.) and sixty matched HC (28 males, 31.8 y.o. ± 7.5 y.o.) took part to the study. FEP patients were selected within the subproject CONTRABASS (Cognitive Neuroendophenotypes for treatment and rehabilitation of psychosis; Brain imaging, inflammation and stress), part of the GETUP study (Genetics, Endophenotypes and Treatment: Understanding early Psychosis. The HC had no history of psychiatric disorders, as measured with the SCID-Non patient version. Mann-Whitney tests were used to assess differences in age, clinical variables and metabolites' levels, and the chi-square test was used to assess differences in sex. All subjects signed an informed consent prior to participating to the study. The GET UP Program, was approved by the ethics

committees of the coordinating center (Azienda Ospedaliera Universitaria di Verona).

2.2.2.2 Psycho-Cognitive Measures

All subjects had the following scales administered:

General assessment of functioning (GAF)

Buss-Durkee Hostility Inventory (BDHI, divided in the following subscales assault, indirect hostility, irritability, negativism, resentment, suspicion, verbal hostility, guilt)

Experiences in close relationships (ECR, subscales: anxiety, avoidance)

Barratt Impulsiveness Scale (BIS) (Patton et al., 1995), Italian version (Fossati et al., 2001). BIS-11 is a self-report questionnaire composed of 30 items; every single item is answered on a 4-point scale (Rarely/Never, Occasionally, Often, Almost Always/Always). Each item contributes to define six first-order factors, which combined produce the following three second-order factors: "Attentional Impulsiveness", "Motor Impulsiveness" and "Non-Planning Impulsiveness"

Temperament and character inventory (TCI).

2.2.2.3 Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopy (MRS) Acquisition

Single-voxel in vivo 1H MRS spectra were acquired using the point-resolved spectroscopy (PRESS) sequence on a 3T whole-body MR system (Philips Achieva, Philips, the Netherlands; TR = 3,000 ms, TE = 36 ms, voxel sixe 17 x 17 x 17 mm = 4.91 mm3, 2,048 complex data points, spectral bandwidth of 2,000 Hz, 128 water suppressed, and 2 water unsuppressed averages). The 1H MRS voxel was positioned in the anterior cingulate cortex. 3D T1-weighted images (190 slices, TR = 8.2 ms, TE = 3.75 ms, flip angle = 8°, FOV = 240x240 mm, pixel dimension = 1x1x1 mm3) were also acquired. The proportions of tissue content of gray matter, white matter, and cerebro-spinal fluid within the localized voxels were estimated using FSL and FreeSurfer (http://surfer.nmr.mgh.harvard.edu/). Metabolite levels (PCr-Cr ratios) were estimated using the linear combination model [LCModel, version 6.3-1 (Provencher 2001)] software with a simu-lated basis set, the unsuppressed water signal, and by incorporating the appropriate correction factors

(T1 and T2 relations) (Gasparovic et al., 2006). The quantified metabolites included phos-phocreatine plus creatine (PCr+Cr), glycerophosphocholine plus phosphocholine (GPC+PC), myo-inositol, glutamate, and N-acetylaspartate (NAA).

2.2.2.4 Statistical analysis

We used Mann-Whitney tests to compare scales' scores (significance level alpha=0.003, Bonferroni corrected) and metabolite levels (significance level alpha=0.006, Bonferroni corrected) between HC and FEP.

We also ran linear regressions with metabolite levels as independent variables and scales' scores as dependent variables. To do that, we tested linear models using f-tests, against the null hypothesis that the model didn't fit the data. We selected the significant models (FDR corrected by Benja-mini-Hochberg method, adjusting p-values starting with a significance alpha=0.05) and selected those in which the impact of the considered independent variable was statistically significant on the considered dependent one.

2.2.3 Results

HC and FEP subjects did not differ for age or sex ($p>0.05$). Demographic and clinic details are re-ported in Table 3. Table 4 reports the average values of metabolites for the two groups. The groups differed in most clinical scales (see Table 3) and for the concentrations of WM and GM rGPC+PC and GM rNAA.

Table 3: clinical variables differences between groups

Variable	FEP (mean (SD))	HC (mean (SD))	p-value* (Mann-Whitney)
Sex	23/37	32/28	0.09
Age	29.32 (9.4)	31.83(7.47)	0.11
GAF	**48.15 (14.10)**	**84.32 (5.89)**	**<0.001**
BDHI			
Assault	**2.98 (1.39)**	**1.00 (1.02)**	**<0.001**
Indirect	5.02 (2.42)	4.21 (1.69)	0.11
Irritability	5.20 (2.68)	4.21 (1.62)	0.086
Negativism	2.88 (2.25)	2.46 (1.35)	0.50
Resentment	**4.69 (3.54)**	**1.14 (1.43)**	**<0.001**
Suspicion	**5.18 (3.23)**	**1.39 (1.07)**	**<0.001**
Verbal	6.96 (1.99)	6.29 (2.46)	0.14
Guilt	**5.35 (3.66)**	**3.14 (1.99)**	**<0.001**
Total score	**38.50 (16.96)**	**23.86 (7.07)**	**<0.001**
ECR			
Anxiety	76.64 (24.21)	63.72 (15.22)	0.012
Avoidance	**78.66 (18.77)**	**34.96 (14.61)**	**<0.001**
BIS			
Attentional	**19.60 (3.79)**	**14.96 (2.38)**	**<0.001**
Motor	21.07 (4.68)	18.75 (3.70)	0.022
Nonplanning	27.62 (5.08)	24.25 (3.40)	0.0027
Total score	**68.29 (10.13)**	**57.96 (7.54)**	**<0.001**
TCI			
Novelty seeking	10.72 (3.02)	8.00 (3.16)	0.27
Harm Avoidance	10.13 (3.11)	9.22 (4.55)	0.89
Reward Dependence	7.48 (2.53)	9.78 (2.28)	0.0027
Persistence	2.70 (1.36)	2.78 (1.72)	0.55
Self-directedness	**13.76 (4.56)**	**19.56 (3.71)**	**<0.001**
Cooperativeness	**13.30 (2.63)**	**22.22 (0.97)**	**<0.001**
Self-trascendence	8.09 (3.94)	3.00 (1.87)	0.007

*Level of significance after Bonferroni correction: p=0.002.

FEP=first episode psychosis, HC=healthy controls; SD=standard deviation; GAF=General assessment of functioning; BDHI= Buss-

Durkee Hostility Inventory ; ECR= Experiences in close relatioships; BIS=Barratt Impulsiveness Scale; TCI=Temperament and charactery inventory.

Table 4: metabolites levels differences between groups

Variable	FEP (average (SD)	HC (average (SD))	p-value* (Mann-Whitney)
GM NAA	**1.63 (0.14)**	**1.55 (0.11)**	**<0.001**
GM myoInositol	1.08 (0.09)	1.07 (0.09)	0.50
GM Glu	1.06 (0.12)	1.04 (0.10)	0.22
GM GPC+PC	**0.34 (0.04)**	**0.32 (0.03)**	**<0.001**
WM NAA	1.76 (0.15)	1.73 (0.13)	0.22
WM myoInositol	0.96 (0.10)	0.97 (0.08)	0.69
WM Glu	1.04 (0.11)	1.04 (0.09)	0.73
WM GPC+PC	**0.36 (0.05)**	**0.34 (0.04)**	**<0.001**

*Level of significance after Bonferroni correction: p=0.006.

FEP=first episode psychosis; HC=healthy controls; Cr+Pcr=Creatine plus phosphocreatine; WM=white matter; GM=gray matter; GPC-PC= glycerophosphocholine plus phosphocholine; NAA= N-acetylaspartate; Glu=glutamate.

2.2.3.1 Regression analyses

Statistically significant models for linear regressions are reported in Table 5 and 6. No model resulted significant in the FEP group, while considering the whole group and the HC group separate-ly, multiple scales resulted linearly related to rGPC+PC, rNAA and myoinositol. In particular, in HC (Table 5) the following metabolites and scores resulted significantly linearly related: WM rGPC+PC with GAF and three subscales of BDHI, GM rGPC+PC with GAF, BDHI assault and TCI cooperativeness, GM rNAA with BDHI suspicion and GM r-myoinositol with GAF and two subscales of BDHI. When the considering the whole sample without distinguishing between HC and FEP (Table 6), the following metabolites and scores resulted significantly linearly related: WM rGPC+PC with GAF and BIS attentional, GM rNAA with GAF, GM rGPC+PC with BIS attentional and GM r-glutamate with BIS attentional.

Table 5: Statistically significant models for linear regressions in HC

Scale	Metabolite	T - stat	Variable p-value	Model p-value
GAF	WM GPC-PC	8,988	<0.001	<0.001
BDHI-Assault	WM GPC-PC	-3,586	0.001	<0.001
BDHI-Resentment	WM GPC-PC	-2,166	0.034	0.001
BDHI-Suspicion	WM GPC-PC	-2,161	0.035	<0.001
BDHI-Suspicion	GM NAA	-2,402	0.020	0.001
GAF	GM GPC-PC	7,370	0.000	0.004
BDHI-Assault	GM GPC-PC	-2,738	0.008	0.000
TCI-cooperativeness	GM GPC-PC	7,626	<0.001	0.006
GAF	GM myoInositol	7,025	<0.001	0.001
BDHI-Resentment	GM myoInositol	-2,142	0.036	0.002
BDHI-Suspicion	GM myoInositol	-2,553	0.013	<0.001

BDHI= Buss-Durkee Hostility Inventory ; GAF= General assessment of functioning; TCI=temperament and character inventory; WM =

white matter; GM=gray matter; GPC-PC= glycerophosphocholine plus phosphocholine; NAA= N-acetylaspartate.

Table 6: Statistically significant models for linear regressions in the whole sample.

Scale	Metabolite	T - stat	Variable p-value	Model p-value
GAF	WM GPC-PC	8.393	<0.001	0.001
BIS-Attentional	WM GPC-PC	4.545	<0.001	0.001
GAF	GM NAA	6.389	<0.001	0.002
BIS-Attentional	GM GPC-PC	3.782	<0.001	0.000
BIS-Attentional	GM Glu	3.312	0.001	0.003

GAF= General assessment of functioning; BIS= Barratt Impulsiveness Scale; WM=white matter; GM=gray matter; GPC-PC= glycerophosphocholine plus phosphocholine; NAA= N-acetylaspartate; Glu=glutamate.

To further define the role of metabolites, we implemented our analyses by including relevant co-variates in the regressions.

Namely, sex, medications, and substance of abuse were included as covariates, since previous literature suggests an effect of these aspects on brain metabolites (Pigoni et al., 2020; Musazzi et al., 2013; Whitehurst et al., 2020; Colizzi et al., 2016). No significant as-sociations were reported for the FEP group. On the contrary, significant associations were found for the whole sample and for the HC group. These results are reported in Table 7; all the present-ed results are FDR corrected.

Table 7: regression analyses including sex, medications, and substance of abuse as covariates

Scale	Metabolite	Covariate	T - stat	Variable p-value	Model p-value
Whole sample					
GAF	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	-37.181.740.913.153.300	0.000310583	0.01352813
BDHI_Assault	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	287.537.360.743.945	0.004802697	0.018378943
ECR_avoidance	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	30.900.892.300.523.100	0.002505243	0.048611387
BIS_Attentional	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	3.654.953.552.627.990	0.00038782	0.028843015
ECR_avoidance	GM_NAA.PCr_Cr.ratio_mean	Sex	2.641.631.445.319.600	0.009388574	0.04835957
BIS_Attentional	GM_NAA.PCr_Cr.ratio_mean	Sex	32.154.781.176.563.900	0.001687371	0.028843015
BDHI_Assault	GM_GPC_PC.PCr_Cr.ratio_mean	Sex	27.394.881.321.755.500	0.007126115	0.025280792
TCI_autotrascendenza	GM_GPC_PC.PCr_Cr.ratio_mean	Sex	22.679.019.847.117.900	0.025186064	0.035667485
BIS_Attentional	GM_GPC_PC.PCr_Cr.ratio_mean	Sex	31.799.154.025.203.100	0.001889634	0.042307061
GAF	WM_NAA.PCr_Cr.ratio_mean	Substances	-23.173.438.429.991.800	0.022238901	0.000139285
ECR_avoidance	WM_NAA.PCr_Cr.ratio_mean	Substances	2.436.885.806.099.420	0.016333601	7.81E+09
GAF	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	-327.951.753.321.802	0.001373122	2.23E+11

BDHI_Assault	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	4.476.574.965.529.210	1.78E+10	0.000202209
BDHI_Resentment	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	22.918.331.758.958.600	0.023719299	0.002576775
BDHI_Suspicion	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	24.563.411.061.077.200	0.015517491	0.01769134
BDHI TOT	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	2.260.528.179.686.220	0.025653761	0.039195905
ECR_avoidance	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	29.261.668.446.961.100	0.004129775	7.81E+09
TCI_cooperativitÃ	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	-25.307.005.354.427.300	0.012723131	0.001231092
TCI_autotrascendenza	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	27.980.779.769.200.700	0.006021229	0.047874287
BIS_Attentional	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	19.829.957.790.668.900	0.049731008	0.003926829
GAF	GM_NAA.PCr_Cr.ratio_mean	Substances	-33.513.174.869.032.600	0.001086174	2.06E+11
BDHI_Assault	GM_NAA.PCr_Cr.ratio_mean	Substances	19.913.891.812.954.400	0.048788144	0.023796721
BDHI_Suspicion	GM_NAA.PCr_Cr.ratio_mean	Substances	26.749.178.060.058.900	0.008554877	0.01441608
ECR_avoidance	GM_NAA.PCr_Cr.ratio_mean	Substances	30.168.231.178.762.600	0.003139828	7.81E+09
GAF	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	-28.298.601.770.358.700	0.005489555	3.70E+09
BDHI_Assault	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	3.804.443.011.400.790	0.000228433	0.001501112
BDHI_Suspicion	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	2.147.858.874.952.580	0.033805175	0.033548007
BDHI TOT	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	21.139.818.257.496.700	0.036658982	0.042307061

ECR_avoidance	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	26.953.876.424.318.600	0.008076075	1.01E+11
TCI_cooperativitÃ	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	-26.889.967.255.990.900	0.008222887	0.000386924
TCI_autotrascendenza	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	2.851.934.331.007.200	0.005145928	0.033114037
GAF	GM_myo.PCr_Cr.ratio_mean	Substances	-23.927.894.050.535.100	0.018326268	8.84E+10
BDHI_Suspicion	GM_myo.PCr_Cr.ratio_mean	Substances	2.462.140.447.499.570	0.015281355	0.016636772
BDHI TOT	GM_myo.PCr_Cr.ratio_mean	Substances	22.130.911.704.783.000	0.028848441	0.020183764
GAF	GM_Glu.PCr_Cr.ratio_mean	Substances	-20.544.769.085.772.300	0.042176225	0.000191152
BDHI_Assault	GM_Glu.PCr_Cr.ratio_mean	Substances	21.061.994.064.186.600	0.037343198	0.028701153
ECR_avoidance	GM_Glu.PCr_Cr.ratio_mean	Substances	23.273.247.475.840.400	0.021682233	1.65E+11
HC group					
GAF	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	-35.218.558.558.537.700	0.00086044	0.020183764
BDHI_Assault	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	4.477.935.295.394.410	3.76E+11	0.000720087
BDHI_Suspicion	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	2.528.314.318.516.680	0.014307498	0.040087258
TCI_persistenza	WM_GPC_PC.PCr_Cr.ratio_mean	Sex	21.717.026.726.191.600	0.034129483	0.016414282
BDHI_Assault	GM_GPC_PC.PCr_Cr.ratio_mean	Sex	35.404.685.909.056.300	0.000812398	0.008423004
TCI_persistenza	GM_GPC_PC.PCr_Cr.ratio_mean	Sex	232.234.399.197.483	0.023874684	0.028941455

BDHI_Assault	GM_myo.PCr_Cr.ratio_mean	Sex	2.621.176.329.363.640	0.01126115	0.048525939
BDHI_Suspicion	GM_myo.PCr_Cr.ratio_mean	Sex	254.964.163.874.695	0.013548271	0.048361113
GAF	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	-4.108.158.264.020.390	0.000126787	0.005407532
BDHI_Assault	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	5.407.455.610.026.010	1.26E+09	0.000139285
BDHI_Resentment	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	36.081.964.220.774.300	0.000643332	0.01352813
BDHI_Suspicion	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	39.308.995.494.824.400	0.00022816	0.007593035
BDHI TOT	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	3.178.149.740.292.080	0.00237734	0.033114037
ECR_avoidance	WM_GPC_PC.PCr_Cr.ratio_mean	Substances	33.951.608.953.924.400	0.001243203	0.020183764
BDHI_Suspicion	GM_NAA.PCr_Cr.ratio_mean	Substances	35.251.361.185.406.400	0.000833846	0.015871833
BDHI TOT	GM_NAA.PCr_Cr.ratio_mean	Substances	2.932.576.341.790.420	0.004805709	0.048525939
ECR_avoidance	GM_NAA.PCr_Cr.ratio_mean	Substances	30.841.601.577.417.900	0.003124311	0.039195905
BIS_Attentional	GM_NAA.PCr_Cr.ratio_mean	Substances	3.075.051.784.982.970	0.003207339	0.039343364
GAF	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	-3.039.539.874.438.880	0.003551149	0.042307061
BDHI_Assault	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	4.329.658.606.013.560	5.99E+11	0.002915371
BDHI_Suspicion	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	33.140.215.431.347.200	0.001588616	0.02458643
BDHI TOT	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	31.545.952.325.101.400	0.002546946	0.033741947
ECR_avoidance	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	29.461.995.782.172.000	0.004625692	0.04848194

TCI_autotrascendenza	GM_GPC_PC.PCr_Cr.ratio_mean	Substances	29.585.439.819.447.200	0.004468004	0.04835957
GAF	GM_myo.PCr_Cr.ratio_mean	Substances	-3.488.380.577.826.800	0.000934311	0.016636772
BDHI_Assault	GM_myo.PCr_Cr.ratio_mean	Substances	301.600.270.938.574	0.00379766	0.043673087
BDHI_Resentment	GM_myo.PCr_Cr.ratio_mean	Substances	31.936.356.684.566.500	0.002271674	0.032430101
BDHI_Suspicion	GM_myo.PCr_Cr.ratio_mean	Substances	3.866.518.409.059.390	0.000281566	0.008423004
BDHI TOT	GM_myo.PCr_Cr.ratio_mean	Substances	38.275.829.307.894.500	0.000319493	0.008673456
BDHI_Irritability	GM_Glu.PCr_Cr.ratio_mean	Substances	31.648.809.742.591.500	0.002471532	0.033548007
BDHI_Suspicion	GM_Glu.PCr_Cr.ratio_mean	Substances	3.503.349.604.083.430	0.000892081	0.016414282
BDHI TOT	GM_Glu.PCr_Cr.ratio_mean	Substances	33.415.795.764.880.900	0.001462225	0.023283118

2.2.4 Discussion

To the best of our knowledge, this is the first study trying to correlate temperament traits and brain metabolites in FEP patients. In line with previous evidence, we found differences in WM and GM rGPC+PC and GM rNAA concentrations between FEP and HC. Moreover, we found an interplay between brain metabolites and temperament traits, especially BDHI subscales, in the sample of HC. These associations seem

to be further modulated by covariates such as sex and substance abuse, in line with previous reports (Colizzi et al., 2016; Pigoni et al., 2020). Contrary to our hypothesis, no correlation between the brain metabolites and temperament traits were found in the FEP sample, not even taking the use of medication into account as covariate.

2.2.4.1. Metabolic alterations and their association with clinical scales in FEP patients

In our sample, we found a significant difference in the concentrations of WM and GM rGPC+PC and GM rNAA between FEP and HC. These results are in line with previous studies suggesting an early metabolic alteration in FEP (Squarcina et al., 2017; Whitehurst et al., 2020). This alteration seems to be less prominent than in chronic psychosis (Whitehurst et al., 2020).

NAA is a marker of neuronal integrity and acts as a precursor for N-acetylaspartylglutamate, which regulates the Glu and dopamine release (Xi et al., 2002). Previous studies in chronic patients have consistently reported the altered stimulation of prefrontal dopamine receptors and disruption of Glu

transmission in the N-methyl-D-aspartate (NMDA) receptors (Laurelle et al., 2003), which potentially related with the deficits of NAA. On the other hand, GPC + PC are catabolic and anabolic metabolites of membrane phospholipids, and neurotransmitters, and they have been suggested to be altered in schizophrenia (Adam and Neuhausen, 2019). Similarly, these results have been replicated in FEP, although to a lesser extent, suggesting that metabolic brain alterations occur early in the course of the disease, probably even before the clinical onset (Keshavan et al., 2009), but are sensitive to the phase of the disorder and become more evident with the chronic course (Whitehurst et al., 2020). Metabolic changes may reflect reduction in neuronal and synaptic processes that can be present already in high-risk subjects (Keshavan et al., 2009; Whitehurst et al., 2020) and therefore may represent markers of liability to the disorder, but this issue needs to be investigated further.

Contrary to our hypothesis, no correlation was found between metabolite levels and temperament and impulsivity scales in the FEP sample. Although this was the first attempt to correlate MRS metabolite levels and temperament in FEP, it was plausible to hypothesize a correlation, given the fact that both

are considered risk factors for the development of psychiatric disorders. Specifically, behaviors and temperament items, even before the clinical onset, are often regarded as risk fac-tors for later development of psychiatric disorders (); in particular, young individuals with persistent internalizing and externalizing behaviors are more likely to develop a mental health disorder in adulthood (Hofstra et al., 2002 Goodwin et al., 2004). Similarly, impulsivity and externalizing behaviors have been described as possible endophenotypes in psychiatric disorders (Porcelli et al., 2017; Boscutti et al., 2022).

Our results remained negative even after taking into account sex, substances, and treatment status, all reported to have an impact on brain metabolite levels (Musazzi et al., 2013; Colizzi et al., 2016; Pigoni et al., 2020; Whitehurst et al., 2020). Such findings are in partial agreement with previous literature (Borgan et al., 2019) showing no direct influence of medications and substances on FEP brain metabolites, but also in disagreement with other studies (Whitehurst et al., 2020). The lack of consistency in the literature may be linked to a great deal of heterogeneity in proton magnetic spectroscopy measures owing to between-study heterogeneity in sample characteristics, in terms of duration of untreated illness or

medication story, imaging acquisition parameters, and analytical techniques. The present findings therefore extend the existing literature taking into account as covariates the most important variables known to be associated with metabolite levels.

2.2.4.2. Metabolic alterations and their association with clinical scales in HC

Our results in HC showed an influence of various brain metabolites on many temperament items and also an interplay with sex and substances. Previous literature regarding healthy individuals suggested how substances might alter brain metabolites, usually reducing them, and how these reductions also correlate with temperamental aspect, especially impulsiveness (Li et al., 2020; Subramaniam et al., 2022). Specifically for our sample, the most affected temperament aspects were TCI self-transcendence, BDHI assault, ECR-avoidance, and BIS-attentional. These results are in line with previous literature and move towards the possibility of discriminating the biochemical systems underpinning personality functioning.

Specifically, in line with other studies conducted in healthy individuals (Squarcina et al., 2019), we found an association between GPC-PC levels and TCI self-transcendence, further modulated by substance use. Self-transcendence identifies subjects with higher levels of creativity, spirituality, and mysticism (Brambilla et al., 2013), brain functions that do not require the engagement of higher order cognitive regions such as the anterior cingulate cortex, suggesting that this region needs a lower metabolic availability for self-transcendence aspects.

Similarly, attentional aspects are reported to be affected by brain metabolites levels in both healthy individuals and sample affected by psychiatric disorders (Husarova et al., 2014; Bakhshinezhad et al., 2022; McGlade et al., 2022). Our results showed an influence of NAA on BIS-attentional, with also an effect of substance abuse. This result is in line with previous literature, suggesting alterations in NAA levels as a possible hallmark of increased risk of attention disorder. Levels of NAA have been proposed to be associated with energy production, which may reflect changes in brain activity (Stork et al., 2005). Alterations in the anterior cingulate cortex, that is in-volved in decision making and deployment of cognitive control (Halroyd

and Yeung, 2012), might reflect an increased risk for attentional deficits, further impacted by substance use.

2.2.4.3. Limitations

Importantly, some limitations must be taken into account while evaluating the presented findings. The major limitation to the study is the small sample size. Even though our sample is in line with other spectroscopy studies in FEP (Borgan et al., 2019) and allowed a sex and age balance with HC, the small sample size is prone to outliers effect and to be driven by specific characteristics of the sample, reducing the generalizability of the findings. Moreover, other confounding factors and co-variates might influence the levels of brain metabolites that are not considered in the present study, such as inflammatory aspects. The study is also limited by its cross-sectional framework. Given that brain metabolites are highly sensible to the phase of the disorder (Whitehurst et al., 2020), longitudinal studies are highly recommended in order to assess the modifications in metabolites levels during the course of the disease, also in relation to the progression of FEP towards an affective or non-affective disorder. Finally, methodological

limitations should be reported. MRS al-low to assess the brain biochemistry only within a restricted region of interest and not at whole brain level, reducing the possibility of findings. Moreover, each metabolite was singularly entered as predictor, instead of running multiple regressions (i.e., including all the metabolites as independent variables). However, given the high collinearity between the independent variables, including them in one model was not possible. This grew the number of models that were tested, leading to correction for multiple comparisons.

3. Big Data and Machine Learning – we need a bigger boat!

Recent years have witnessed the raise of the use of big data and innovative computational techniques such as Machine Learning in psychiatry (Dwyer et al., 2018). The availability of big data in medicine (e.g., electronic health records) have fed these new computational techniques, with interesting results in different fields.

Machine Learning (ML) is broadly defined as a computational strategy that automatically determines (i.e., learns) methods and parameters to reach an optimal solution for a problem without being specifically programmed by a human a priori. ML approaches are also especially designed for the multivariate analysis of data sets with high dimensionality (i.e., many variables). ML is considered a subfield of artificial intelligence (AI) because the learning process putatively simulates human intelligence and can be used for ostensibly intelligent ends (e.g.,

speech, writing, face recognition, self-driving cars, or medical decision aides) (Dwyer et al., 2018).

Machine Learning algorithms are now implemented into our daily life, f. e. in the form of Internet searches engines and personalized product recommendations ("you might also like" in Amazon, Netflix, Spotify etc.), translation services, speech recognition services, and self-driving cars (Jordan and Mitchell 2015).

Regarding health care, the ML approach has been already implemented in some fields, given that it proved to perform equally or better than clinicians in tasks that involve pattern recognition in images, such as the detection of skin cancer (Esteva et al., 2017), lung cancer (Yu et al., 2016), or eye disease (Long et al., 2017). Medical imaging companies have already integrated machine learning algorithms to ultrasound devices to detect breast cancer (e.g., Samsung, RS80A).

These advances are fueling the ML industry and psychiatry could benefit as well. The problems of translational clinical psychology and psychiatry that can be optimally addressed with ML fall into four main categories: diagnosis, prognosis, treatment prediction, and the detection of potential biomarkers.

However, we have to bear in mind that the aim of any translational ML approach is to produce models that are sufficiently meaningful, accurate, and generalizable to be integrated into everyday clinical practice and which might help psychiatrists and clinicians in informed decisions in order to move towards a really personalized medicine.

Regarding diagnosis, ML was used to classify and distinguish diagnostic division based mostly on structural and functional neuroimaging (Arbabshirani et al. 2017). Recent machine learning studies in chronic schizophrenia have employed multivariate algorithms (Arbabshirani et al., 2017; Dwyer et al., 2018) to derive patterns of interconnected abnormalities that distinguish patients from controls (Kambeitz et al., 2015; Guo et al., 2016; Rozycki et al., 2018) with high accuracy and generalizability in large samples (Rozycki et al., 2018). Overall, extant literature suggests that ML can provide a valid classification of chronic psychosis with accuracies above 75% (Kambeitz et al., 2015). However, similar multi-site, large-scale studies are lacking early in the course of psychotic illnesses, when the cumulative effects of confounding variables are lower (Fusar-Poli et al., 2013; Vita et al., 2015) and when it is critical to characterize abnormality patterns given the need for

biologically-targeted early intervention (Fusar-Poli et al., 2016; Correll et al., 2018). Results regarding FEP are still mixed, with some authors reporting similar brain classification accuracies above 80% (Borgwardt et al., 2013) while other studies reported accuracies between 50% and 60% (Winterburn et al., 2017; Vieira et al., 2019).

To overcome such limitations, consortia sharing data from different centers are needed, since the specific characteristic of each sample might greatly influence prediction. I present here results from 3 big consortia, namely ClassiFEP, PRONIA, and PHENOM.

ClassiFEP gathered together data from different European sites (namely, Verona, Munich, Basel, Santander, London, Jena, Rome, and Prague). The shared data consisted of raw T1 MRI scans of FEP and matched healthy controls, along with clinical and demographic data.

The PRONIA consortium (Personalised Prognostic Tools for Early Psychosis Management) is a research project funded by the European Union aiming to considerably improve the early detection of psychotic diseases. Over the course of more than 5 years, it gathered data from different sites across Europe, the

US, and Australia, regarding recent onset psychosis, recent onset depression, clinically high risks and healthy controls.

The PHENOM consortium ('Psychosis Heterogeneity Evaluated via Dimensional Neuroimaging') gathered together data spanning four continents (US and South America, EU, Australia, and Chinese cohorts), from FEP, chronic patients affected by schizophrenia and healthy controls. Raw T1 MRI scans, clinical and demographic data were collected.

3.1 Classification of first-episode psychosis using cortical thickness: A large multicenter MRI study (Pigoni et al., 2021)

This paper has been published in 2021 on European Neuropsychopharmachology

3.1.1 Introduction and Rationale

Extant univariate FEP literature suggests a pattern of abnormalities encompassing prefrontal, temporal and cingulate regions (Chang et al., 2016; Ohtani et al., 2018; Calvo et al., 2019). Such results have been supported by meta-analyses (Fusar-Poli et al., 2012; Shah et al., 2017) but machine learning results in single samples have been mixed. While some authors evidenced similar brain patterns with classification accuracies above 80% (Borgwardt et al., 2013), other studies reported accuracies between 50% and 60% (Winterburn et al., 2017; Vieira et al., 2019). This heterogeneity can be explained by MRI machine differences (Abdulkadir et al., 2011; Squarcina et al., 2017), various machine learning algorithms (Torres et al., 2016; Bugger et al., 2017; Vieira et al., 2019) and by the fact that single-

site studies might capture specific illness characteristics of the local samples.

Accuracies in the classification of first episode patients might also be due to difficulties in obtaining a subtle and distributed brain abnormality signature from the region-of-interest (ROI) summary statistics (Torres et al., 2016; Palaniyappan et al., 2016). Most studies have employed such an approach (Yun et al., 2016; Squarcina et al., 2017; Liang et al., 2018), commonly using brain atlases that derive estimates of volume or thickness from large brain areas that may capture the expected subtle patterns (Fusar-Poli et al., 2013). Only few studies have employed voxel-based imaging using volumetric data (Pettersson et al., 2013; Squarcina et al., 2015; Koutsouleris et al., 2016; Mikolas et al., 2018; Vieira et al., 2019) with results demonstrating good accuracies ranging from 65% to more than 80%. However, no studies have currently compared the whole-brain, vertex-wise, and thickness predictions with classical ROI approach.

Whole-brain analysis of cortical thickness in FEP using ML may increase accuracy and improve the interpretation of results (Tamnes et al., 2010; Rimol et al., 2012; Squarcina et al., 2017). Cortical thickness alterations are hypothesized to be caused by

disturbances in neurogenesis, myelination, and synaptic pruning (Sekar et al., 2016) that selectively affect specific cortical regions in the first decades of life (Tamnes et al., 2010; Rimol et al., 2012). These thickness measures were preferentially selected by recent analyses and proved to carry a discriminative power when classifying FEP from healthy individuals (Yun et al., 2016; Squarcina et al., 2017; Liang et al., 2018). However, all these studies employed small, single-site, samples that might limit the generalizability of the findings and cannot provide detailed cortical thickness abnormality maps.

In order to disentangle heterogeneity in the data and provide a clearer picture of thickness abnormality patterns, we had three aims:

1) to use ML to classify individuals with a first-episode from controls on the basis of whole-brain cortical thickness maps in one of the largest, multi-site sample used for this purpose to-date;

2) to systematically investigate the contributions of different demographics, sites, scanners, image quality, and brain parcellation schemes associated to the prediction;

3) to investigate the relationship between the predictions and critical clinical baseline and outcome variables. We hypothesized that vertex-wise cortical thickness maps would demonstrate the most generalizable showing pattern of decreased thickness in fronto-temporal regions that would correlate with symptom severity, and outcomes.

3.1.2 Methods

Healthy controls (HC) and FEP patients were recruited at 8 sites in 6 countries across Europe (Verona, Munich, Basel, Santander, London, Jena, Rome, and Prague) as part of independent studies, and gathered into the ClassiFEP group. Comorbidity with substance dependence disorders was excluded. A total of 448 HC (230 females, mean age 27.06, Standard Deviation [SD] 7.37) and 428 FEP patients (147 females, mean age 27.14, SD 6.32) were included. The diagnosis of the first psychotic episode was established according to the criteria of DSM-IV or ICD-10. Written informed consent was obtained from all participants. The authors assert that all procedures contributing to this work

comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures were approved by the respective ethics committees. Participants also underwent a clinical interview at the time of the MRI scan, where the Positive and Negative Symptoms Scale (PANSS), the Global Assessment of Functioning (GAF), and the Clinical Global Impression Scale (CGI), were administered.

Table 8: Total sample demographics (chi square for nominal variables and t-test for continuous variables)

	FEP	HC	Sign.
Sex (F/M) (n. 428FEP/448HC)	147/281	230/218	$\chi^2_{(1)}$=10.2 p<0.05
Age (n. 428FEP/448HC)	27.14 (±7.37)	27.06 (±6.32)	$t_{(1,875)}$ =0.38 p=0.702
Yrs of Education (n. 284FEP/257HC)	12.46 (±3.27)	14.36 (±3.0)	$t_{(1,540)}$=-6.9 p<0.05
Ethnicity (caucasian%) (n. 261FEP/300HC)	78%	95%	$\chi^2_{(1)}$=36.6 p<0.05
Handedness (r/ambi/l) (n. 245FEP/308HC)	216/6/23	285/10/13	$\chi^2_{(2)}$=6.2 p=0.045
Tobacco smoke (n. 206FEP/213HC)	68%	29%	$\chi^2_{(1)}$=28.8 p<0.05
Lifetime Cannabis use (n. 180FEP/103HC)	36%	22.5%	$\chi^2_{(1)}$=5.8 p<0.05
Age at onset (n. 160FEP)	27.15 (±8.07)		
DUI (month) (n. 146FEP)	4.8 (±8.32)		
N. prior hospitalizations (n. 152FEP)	0.8 (±0.9)		
Chlorpromazine equival. (n. 253FEP)	4347.6(±9374)		
PANNS pos (n. 223FEP)	14.33 (±6.6)		
PANNS neg (n. 223FEP)	15.31 (±7.6)		
PANNS gen (n. 223FEP)	31.25 (±12.5)		
PANNS tot (n. 223FEP)	60.83 (±22.6)		
GAF (n. 256FEP)	55.96 (±18.9)		

CGI (n. 136FEP)	3.44 (±1.38)		
PANNS pos T1 (n. 43FEP)	8.66 (±2.3)		
PANNS neg T1 (n. 43FEP)	12.86 (±6.3)		
PANNS gen T1 (n. 43FEP)	23.36 (±8.7)		
PANNS tot T1 (n. 43FEP)	44.97 (±16.2)		
GAF T1 (n. 62FEP)	76.20 (±16.8)		
CGI T1 (n. 53FEP)	2.09 (±1.3)		

3.1.2.1. Imaging acquisition and feature extraction

All enrolled participants underwent a 3T MRI brain scan. Reconstructed surfaces obtained with Freesurfer (version 5.3 http://surfer.nmr.mgh.harvard.edu/)
(Squarcina et al., 2017) were visually inspected for topological defects according to standard guidelines. Moreover, we developed a semi-automated image quality control protocol based on Rosen et al. (2018), using the Euler number (Dale et al., 1999). The Euler number is calculated separately for each hemisphere and represents a good indicator of the quality of

reconstructions (Rosen et al., 2018). We averaged it across both hemispheres to produce one value per subject (Rosen et al., 2018).

68 ROIs for brain cortical thickness defined by the Desikan–Killiany atlas (Desikan et al., 2006) were obtained from T1-weighted images using FreeSurfer. We extracted cortical thickness maps and resampled them to a common coordinate system (fsaverage5, 10,242 vertices for each hemisphere for each subject). This gave for every subject a vector composed of the value of thickness associated with vertices, by calculating the shortest distance from each point on the gray/white boundary to the pial surface at each vertex (Doan et al., 2017).

3.1.2.2. Multivariate pattern classification analysis

We established a ML pipeline using the tool NeuroMiner (version 0.999; https://www.pronia.eu/neurominer/), running under MATLAB (R2015b, The MathWorks Inc.). To enable the unbiased estimation of the prediction and to prevent information leaking between subjects used for training and validating the models, we used repeated nested cross-validation (CV) (Borgwardt et al., 2013; Dwyer et al., 2018a,

2018b). As previously described (Borgwardt et al., 2013; Dwyer et al., 2018a, 2018b), the process we used produced an unbiased estimate of the method's expected diagnostic accuracy on new individuals and ensured that the test set remained fully independent from the training set, with only the latter being used to optimize model parameters, avoiding overfitting and increasing generalizability.
After visual inspection of the images, we performed SVM classification of the cortical thickness full brain maps using 20,484 features constituted by the thickness value at each vertex for each subject across the brain, and of the ROI-based thickness (68 features) for comparison.

SVMs are multivariate statistical methods that provide optimal decision rules for classifying individuals. The algorithm maps the input data into a feature space using a set of similarity functions known as kernels. In this feature space, the model finds the optimal separating hyperplane by finding the largest margin of separation between the two classes within the training set. Once the hyperplane is determined, it can be used to predict the class of new unseen observations (Dwyer et al., 2018). Models that maximised the balanced accuracy (BAC, (Sensitivity+Specificity)/2)) were selected. This process

produced an unbiased estimate of the method's expected diagnostic accuracy on new individuals rather than merely fitting this study population. As previously described (Borgwardt et al., 2013; Dwyer et al., 2018), this process ensures that the test set remained fully independent from the training set, with only the latter being used to optimize model parameters, and therefore should avoid overfitting and increase the generalizability of the model.

We applied the following preprocessing steps:

Step 1: Prune non-informative columns from matrix [Zero Var, Nan, Inf]

Step 2: Partial correlations (Covariates: Gender, Age, Site, Mean Euler) regressed out only on Healthy Controls.

Given that HC and FEP significantly differed for gender (Table 8) and given the effect of age on the brain structures, we regressed out the variance associated with these variables. More precisely, we removed the variance associated with age and gender from the feature scores within each inner and outer CV fold through partial correlations, carried out only on HC, in order to not remove any relevant information from the FEP group in relation to the disease. The same procedure was

applied to sites and the Euler number, in order to remove the heterogeneity associated to the different MRI scan employed in different centers (Schnack and Kahn, 2016) and the variance associated with the quality of the reconstruction (Rosen et al., 2018).

Step 3: PCA. The dimensionality of the training cases' thickness values was reduced by means of Principal Component Analysis (PCA) (Hansen et al., 1999), a well-established unsupervised method for feature reduction in neuroimaging. PCA geometrically projects the data into lower dimensions called principal components (PCs) and apply an orthogonal transformation to convert a set of observations of possibly correlated features. PCA aimed at finding the best description of the data by using a limited number of PCs. PCs are then ranked according to the explained variance in the descending order (Lever et al., 2017).

Step 4: Scale featurewise [from 0 to 1], zero-out completely non-finite features.

The resulting PC scores were scaled to [0, 1]

As many ML algorithms are sensitive to scale differences between features, we scaled each variable between 0 and 1 in

order to remove these effects from each training sample matrix. The scaling parameters were then applied to both the inner and outer CV cycles.

Permutations: To assign statistical significance to the prediction performance of our classifiers, we performed 100 random permutations between the labels (HC vs. FEP). For each permutation we retrained all linear SVM models in the pooled repeated nested CV experiment, using the respective feature subsets obtained from the observed-label analyses. The significance of the observed out-of-training BAC was calculated as the number of events in which the permuted out-of-training BAC was higher or equal to the observed BAC divided by the number of permutations performed. The significance of the model was determined at $\alpha=0.05$.

3.1.2.3. Heterogeneity assessment

In order to increase the generalizability of our classification (Schnack and Kahn, 2016), the role of heterogeneity of the sample in our classifications' results was assessed in the following ways: First, the sample was divided according to sex (male vs. female) and the MRI scanner employed. Given the fact

that five centers out of eight presented a Siemens machine (namely, Jena, Prague, Verona, Rome, and Base), we divided our sample accordingly, consequently obtaining four subsamples composed by only male subjects, only female subjects, subjects scanned with a Siemens scanner, and subjects scanned with a non-Siemens scanner. Among the non-Siemens sites, two employed a Philips scanner and one a General Electric Signa HDx. Single SVM classifiers were then trained on every subsample. The obtained classifiers were tested on the remaining samples as an independent sample validation (i.e., the classifier obtained from the Siemens subsample was then tested on the non-Siemens subsample). The criteria for subsampling (namely, sex and MRI scanner) were combined and SVM classifiers were trained on combined subsamples (i.e. male subjects scanned on a Siemens scanner, male subjects scanned on a non-Siemens scanner, etc.). The obtained classifiers were tested on the remaining samples. Finally, to further test the effects of sample size (Schnack and Kahn, 2016), we performed random resampling by removing 100 randomly selected participants at a time. We performed SVM classification on these reduced samples. To compare the significant differences between models built on different samples, we computed confidence intervals of BACs.

Specifically, we calculated confidence intervals based on the BACs obtained in each CV fold for each permutation (10 × 10 BACs in each analysis). If the confidence intervals between tests were not overlapping, we considered the differences between the two BACs as significantly different.

3.1.2.4. Impact of brain regions and clinical variables

Visualization of the multivariate patterns used in classification was achieved by back-projecting the feature weight vector of each SVM model from PCA to the original space. Voxel probability maps were visualized reflecting the degree of consistency and reliability of the signatures (Dwyer et al., 2018a, 2018b; Kambeitz-Ilankovic et al., 2019). We performed regression analyses between the predictive decision scores (representing the degree of certainty with which an individual is classified in one group) generated by the classifiers for every subject and clinical variables, obtaining as an output a correlation slope, using support vector regression (SVR)

(Kambeitz-Ilankovic et al., 2019; Koutsouleris et al., 2014), chosen for its established ability to generate unbiased models that generalize well across the population (Koutsouleris et al., 2014). To train and crossvalidate our models, we wrapped our ML pipeline into a nested CV framework using NeuroMiner. The analysis was carried out on a subsample (n. of subjects: 91) that had clinical variables available. The clinical variables considered as predictors were: age at onset, duration of illness, chlorpromazine equivalents, type of antipsychotic used, lifetime use of illicit drugs, affective/non-affective psychosis, clinical diagnosis, PANSS total score, positive, negative, and general subscale scores at baseline, and GAF at baseline.

3.1.3 Results

After visual inspection of the images for topological defects within FreeSurfer reconstruction, a total of 84 subjects (9.6%, 47 FEP and, 37 HC) were classified as poor reconstruction. Interestingly, 93% of the poor reconstructed images presented an Euler number higher than 2 folds standard deviations of the

mean of the sample. This finding validates the image quality-control method based on the Euler number we applied. We decided to use all the images in the first analysis, given the fact that in a real-world environment, images with artifacts and poor reconstruction are common (Koutsouleris et al., 2018).

3.1.3.1 SVM analysis

3.1.3.1.1 Whole brain models

This prediction reached a BAC of 65.1%, with a sensitivity of 64.3% and a specificity of 65.8%. The area-under-the-curve (AUC) was 71%, with a positive predictive value (PPV) of 64.3% and a negative predictive value (NPV) of 65.8%. However, the decision scores of this classification analysis showed a significant correlation with the Euler number ($p<0.0001$). Given the fact that the Euler number is a measure of the quality of reconstruction (Rosen et al., 2018), we could not be sure if our classification was driven by the quality of the images. In order to avoid this important pitfall, we replicated this analysis using only the FreeSurfer reconstructions that were not labeled as poor. The classification was performed on 381 FEP and 411 HC. The discarded subjects (47 FEP and 37 HC) were retained and

further used as a validation sample. This new prediction carried out with the same preprocessing pipeline reached a BAC of 66.2%, with a sensitivity of 64.8%, a specificity of 67.6%, and an AUC of 72%. PPV and NPV were 65.0% and 67.5% respectively. No correlation between the decision scores and the Euler number was detected. The classifier was then applied to the sample composed of poor-reconstructed images as a test sample. This classification on the validation sample yielded a BAC of 65%, an AUC of 75%, a sensitivity of 78.7%, a specificity of 51.3%, a PPV of 67.3%, and an NPV of 65.5%. The model resulted thus generalizable to poor-reconstructed images. However, the prediction was only poorly generalizable across sites. Applying a Leave-Site-Out CV scheme we obtained a just above chance BAC of 55%, with an AUC of 64% and a sensitivity and specificity of 26.2% and 84.2%, respectively.

3.1.3.1.2 ROI-based models

This SVM analysis based on ROIs reached a BAC of 59.1%, with a sensitivity of 44.9%, a specificity of 73.3%, and an AUC of 61%. PPV and NPV were 61.7% and 58.1%, respectively.

The models obtained from vertex-wise cortical thickness maps were significantly different from ROI-based classifications, as tested with the McNemar's chi-square test (Dietterich, 1998).

Table 9: statistical measure from SVM model trained on RoI (68 features) and cortical thickness full maps (20484 features)

Preprocessing	Sample (FEP/HC)	BAC	Sensitivity	Specificity	AUC	PPV	NPV
RoI based models							
Whole sample; standard preprocessing (covariates regression, PCA, Scaling [0,1]	876 (428/448)	57.2%	47.7%	66.8%	60%	58%	81.1 %
Whole sample; Wrapping (early stop at 50%; stepping rule: best 5%)	876 (428/448)	57.9%	48.1%	67.7%	62%	58.9 %	57.9 %
Whole sample; Soft filtering using Pearson correlation	876 (428/448)	59.1%*	44.9%	73.3%	61%	61.7 %	58.1 %
Reduced sample composed by only good images; standard preprocessing (covariates regression, PCA, Scaling [0,1]	792 (381/411)	54.7%**	37.8%	71.6%	56%	55.4 %	55.3 %
Full brain maps based models							

Whole sample; standard preprocessing (covariates regression, PCA, Scaling [0,1]	876 (428/448)	65.1%*	64.3%	65.8%	71%	64.3 %	65.8 %
Whole sample; Soft filtering using Pearson correlation	876 (428/448)	65.8%	54.9%	76.8%	71%	69.3 %	64.1 %
Whole sample; Variance extraction on Sites using ANOVA	876 (428/448)	60.2%	57.7%	62.7%	65%	59.7 %	60.8 %
Reduced sample composed by only good images; standard preprocessing (covariates regression, PCA, Scaling [0,1]	792 (381/411)	66.2%** CI: 65.4-67.0	64.8%	67.6%	72%	65.0 %	67.5 %
Test on „poor quality“ images	84 (47/37)	65.0% CI: 64.8-65.2	78.7%	51.3%	75%	67.3 %	65.5 %
Reduced Sample composed by only good images; Leave site out CV	792 (381/411)	55.2%	26.2%	84.2%	64%	60.6 %	55.2 %

3.1.3.2 Heterogeneity assessment

The best performances were achieved with the model trained on the sites employing a Siemens scanner and with the model trained on male subjects scanned on a Siemens machine, with BAC of 69.5% and 70.4%, respectively. These two classifications resulted significantly different from the other models. Moreover, such models resulted different also from models trained on the single site's samples. The classifiers were then applied to the remaining sample serving as an independent validation sample.

Specifically, the first model trained on the Siemens sites and tested on the remaining sites achieved a BAC of 57.3%, while the second one based on males scanned on Siemens scanners retained an accuracy of 62.8% when tested on the remaining individuals. In both cases, we observed a drop in the BAC by more than 10%. To prove that subsampling could reduce heterogeneity and therefore increase generalizability across sites, we trained a model using a Leave-Site-Out CV scheme. The Leave-Site-Out model trained on the subsample of Siemens

scans reached a BAC of 61.2%, proving a more generalizable classifier.

For comparison, the random resampling of the subjects did not show any significant improvement in BAC, which remained between 66.8% and 62%.

Table 10: results from resampling and heterogeneity assessment

		Sample (FEP/ HC)	BAC	Sensitivity	Specificity	AUC	PPV	NPV
Resampling according to scan and sex								
Train: Only Male		441 (246/195)	65.4% CI 64.0-66.7	77.6%	53.1%	69%	67.5%	65.4%
	Test: Only Female	351 (135/216)	57.7%	60.0%	55.3%	63%	45.7%	68.8%
Train: Only Female		351 (135/216)	62.9% CI 61.1-64.0	46.7%	79.1%	71%	58.3%	70.2%
	Test: Only Male	441 (246/195)	61.1%	67.3%	54.8	64%	67.8%	54.3%
Train: Only Siemens machines (Jena, Prague, Verona, Rome and Basel)		457 (195/262)	69.5% CI 68.1-70.9	63.1%	76.0%	73%	66.1%	73.4%

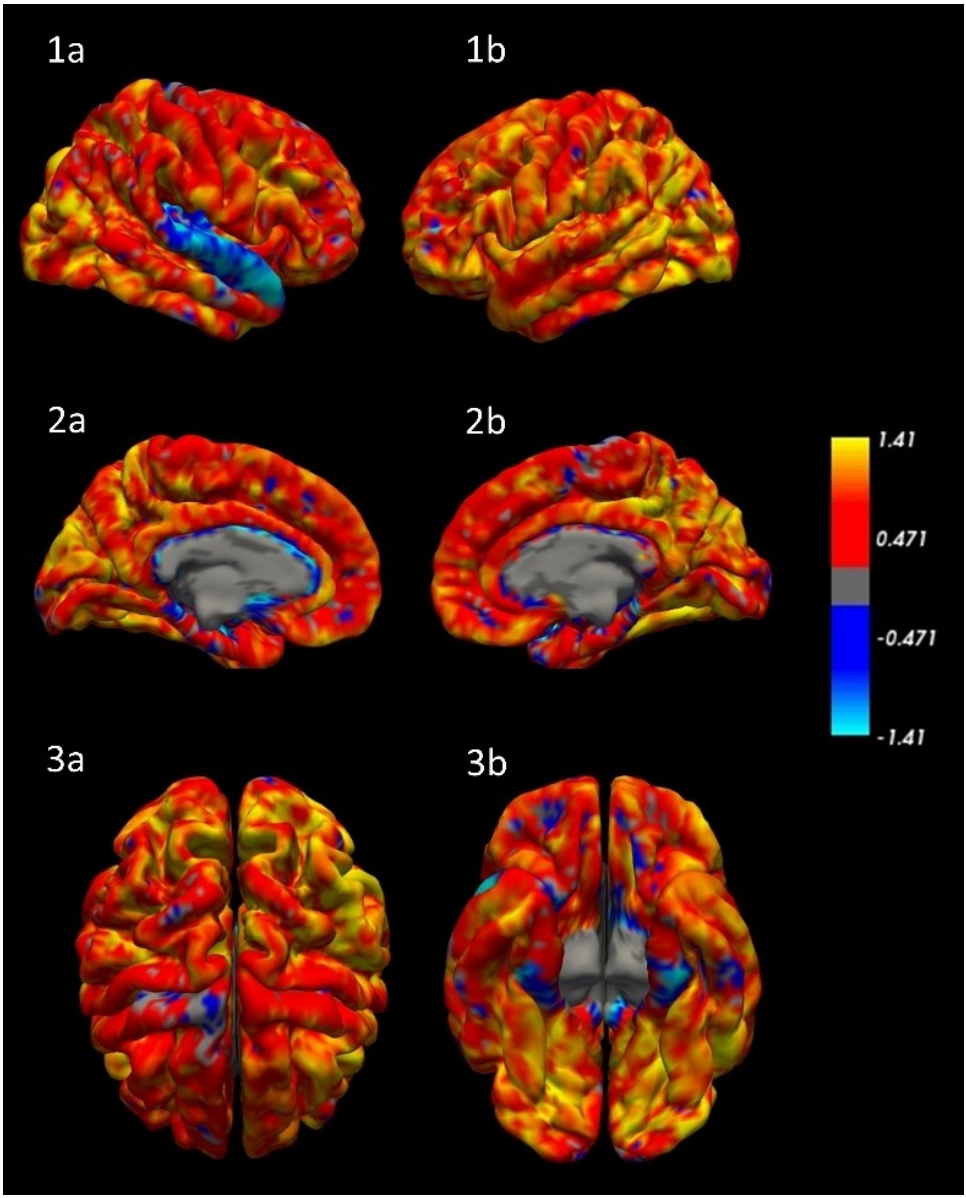
	Test: Non-Siemens machines (Munich, Santander and London)	335 (186/149)	57.3% CI: 57.2-57.4	33.9%	81.2%	64%	69.2%	49.6%
Train: Non-Siemens machines (Munich, Santander and London)		335 (186/149)	66.0% CI 63.6-66.4	68.8%	63.1%	70%	69.9%	61.8%
Train: Only Siemens + Only Male		233 (116/117)	70.4% CI 68.6-72.2	67.2%	73.5%	73%	71.6%	69.4%
	Test: Non-Siemens + Only Male	209 (130/79)	62.8% CI:62.7-62.9	44.6%	81.0%	66%	79.4%	47.0%
Train: Non-Siemens + Only Male		209 (130/79)	61.5% CI 58.2-62.4	72.3%	50.6%	66.0%	70.7%	52.6%
	Test: Only Siemens + Only Male	233 (116/117)	50.0%	0	100%	49%		50.2%
Leave-site-out performed on Only Siemens sample		457 (195/262)	61.2% CI 60.8-61.5	78.5%	43.9%	61%	51.0%	73.2%
Random resampling								
Train: Resampling 1		692 (335/257)	66.8% CI 65.5-67.6	66.0%	67.7%	72%	65.8%	67.9%
Train: Resampling 2		592 (297/295)	62.7% CI 61.4-63.9	62.8%	62.7%	68%	63.0%	62.5%

Train: Resampling 3		492 (250/242)	66.0% CI 64.8-67.1	70.0%	62.0%	70%	65.5%	66.7%
	Test: remaining subjects	300 (131/169)	63.1% CI:63.05-63.14	66.4%	59.8%	68%	56.1	69.6
Train: Resampling 4		392 (185/207)	63.5% CI 62.0-64.8	58.4%	68.6%	69%	62.4%	64.8%
Train: Resampling 5		292 (127/165)	61.7% CI 60.4-63.8	52.0%	71.5%	67%	58.4%	65.9%
Train: Resampling 6		192 (100/92)	61.8% CI 59.6-64.2	66.0%	57.6%	69%	62.9%	60.9%

3.1.3.3 Neuroanatomical findings

Fig. 6 shows the general pattern of selected features in the Leave-Site-Out CV model, as it is the most generalizable one. When setting the threshold to ± 2, which corresponded to an α level of 0.05, only a few voxels located in the right superior temporal gyrus and planum temporale remained significant.

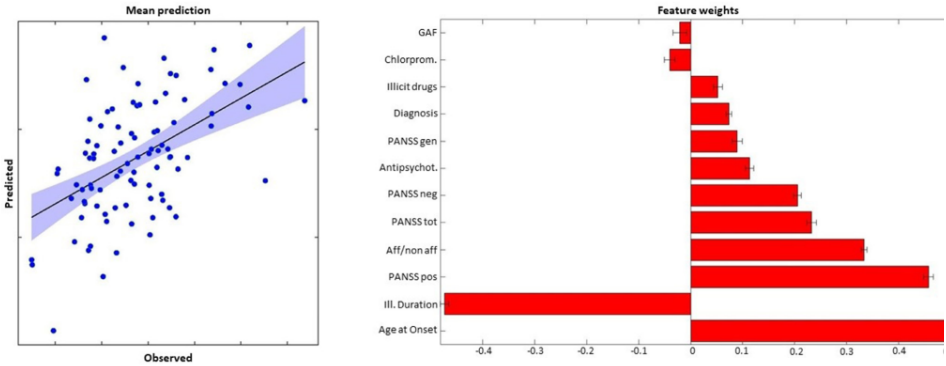
Figure 6: The reliability of predictive voxels on the Leave-Site Out analysis was plotted via a cross-validation ratio map, as presented by Koutsouleris et al. (2018) Color scales indicate increased (warm color scale) vs decreased (cool color scale) thickness in individuals with FEP compared with HC. Subpanels 1a and 1b show the external surface of the brain, subpanels 2a and 2b show the ventral surface of the brain, and subpanels 3a and 3b show the superior and inferior surface of the brain.


1a
1b
2a
2b
3a
3b
1.41
0.471
-0.471
-1.41

3.1.3.4 Impact of clinical variables

The regression model is presented in Fig. 7, panel a. The model predicted a significant correlation ($p<0.001$) with a Pearson r of 0.51 and a mean absolute error of 0.4. Fig. 7 panel b presents the weights of each clinical variable in the analysis. Age at onset, illness duration, PANSS positive score, and the presence of an affective or non-affective psychosis seems to have the highest weight in the prediction.

Figure 7 Panel a): the model illustrates the relationship between clinical and demographic variables and the predicted decision scores, obtained with the Support Vector Regression. The model predicted the decision scores using the clinical and demographic variables. The figures show a positive correlation ($p<0.001$) and the model holds a Pearson r of 0.51 and mean absolute error of 0.4. Panel b): the CV probability weights of clinical measures derived from SVR model. The weight of each feature indicates how much a single feature contributed to the prediction.



3.1.4 Discussion

This study aimed to evaluate the application of ML approach to neuroanatomical data of individuals at the early stages of psychosis by assessing some of the limitations reported in

recent ML literature. Results indicated that vertex-wise analyses outperform classic brain parcellation approaches, improving the classification to accuracy above 66% and showing generalizability in a test sample composed of images labeled as "poor reconstructions". However, the analysis on the pooled sample did not show generalizability across sites. Conversely, the assessment of site-dependent contributors not only proved to be important to enhance the accuracy of the prediction but also increased the generalizability across sites. This classification proved meaningful from a clinical point of view, showing significant correlations with clinical variables. In agreement with our hypothesis, the significant voxels were located in temporal brain areas.

3.1.4.1. SVM models

To go beyond the limitations of the existing literature, we used a large, pooled dataset of subjects enrolled at 8 different sites across Europe and employed cutting-edge methodological precautions (repeated nested-CV) to avoid overfitting. Although already used in ML studies on psychiatric populations (Ciufolini et al., 2018; Doan et al., 2017), to the best

of our knowledge this is the first time that whole-brain and ROI approaches are compared in FEP within a ML framework. Our results suggest that ROI approaches might have a limited ability to detect subtle alterations in FEP, leading to a loss of information regarding brain patterns in psychosis.

For these reasons, cortical thickness maps seem more suitable to assess these morphological differences. Compared to previous classification studies on cortical thickness in FEP (Liang et al., 2018; Squarcina et al., 2017; Vieira et al., 2019; Yun et al., 2016), the accuracy we obtained is similar to the one obtained by Liang et al. (2018) and Vieira et al. (2019), but smaller compared to Yun et al. (2016) and Squarcina et al. (2017). Notably Squarcina et al. (2017) and Liang et al. (2018) employed brain parcellation based on Desikan atlas, as in our analysis, while the Destrieux atlas was used by Yun et al. (2016). Further differences between these studies regard preprocessing pipelines and multivariate algorithms, which could partially account for the different accuracies (Zanetti et al., 2013). Moreover, previous studies' models were trained on smaller and homogenous samples recruited at the same site, which may achieve encouraging accuracies but are prone to a higher risk of overfitting (Brugger and Howes, 2017). The sample in our study

is one of the biggest FEP samples used so far for classification. A recent publication by Vieira et al. (2019) employed a similar sample size and obtained comparable accuracies. Similar to Vieira et al.'s (2019) study, our results on the pooled sample failed to achieve generalizability across sites. Site-dependent differences in scanning parameters and population differences could at least partially explain these findings. Nevertheless, such low cross-site generalizability might limit the clinical applicability of ML models.

However, it should be noted that the accuracy of our study and previous ones (Liang et al., 2018; Vieira et al., 2019) is far from the generally accepted threshold, which could be useful in clinical terms of 80% accuracy. Although beneficial from a research perspective with the aim of better explaining pathologic processes and psychosis development, these ML approaches still appear not to be feasible yet in clinical practice, due to the heterogeneity intrinsic to the disease and the samples employed by the different sites (Vieira et al., 2019). This result points out the possible limitations of ML techniques in psychiatry and the future challenge we will have to face.

3.1.4.2. Resampling and heterogeneity assessment

The tradeoff between accuracy and generalizability of results could be found in homogeneity and dimension of samples (Rimol et al., 2012; Schnack and Kahn, 2016). We demonstrated the importance of the contribution of demographic and scanner-dependent variables, whose inclusion led to an improvement in accuracy up to more than 70% in the pooled sample. The most important factor in our analyses seems to be the heterogeneity within the sample since reducing it randomly did not improve the classification. Interestingly, classifications based on single-sites samples resulted significantly less accurate ranging from chance level to 66% accuracy, compared to the analysis with reduced heterogeneity. Thus, reducing heterogeneity was also able to level out differences among sites, leading to a more generalizable prediction.

Accounting for the differences between MRI scans, we considerably improved the accuracy of the Leave-Site-Out CV scheme, thus enhancing the generalizability of the classifier across different sites and partially overcoming one of the major issues in ML literature. This should be considered among the challenges and limitations of applying ML to psychiatric disorders and could be of crucial importance in the framework

of creating large multisite cohorts (Koutsouleris et al., 2018). However, it is important to highlight that it was not feasible to assess all the possible variables that might have negatively impacted accuracies. Indeed, given their early presentation of our subjects, some variables are intrinsic to the FEP diagnosis and difficult to assess. For example, the presence of both patients with affective and non-affective psychosis that might be characterized by different brain alterations (Amoretti et al., 2021; Rodriguez-Perez et al., 2020), and the disease trajectories in terms of severity and relapses. These confounding variables might have influenced the accuracy of the classification and should be carefully considered in future longitudinal studies.

3.1.4.3. Brain thickness reduction

Fig. 6 shows a complex pattern of increased and decreased thickness in FEP compared to HC. When setting a threshold for the α-level of 0.05, only a few voxels with reduced thickness remain significant, which cluster in the superior temporal gyrus and specifically in the planum temporale.

Alterations in these areas have been previously described in schizophrenia, bipolar disorder (Akdeniz et al., 2016; Chang et

al., 2018; Delvecchio et al., 2017a, 2017b), and in FEP (Calvo et al., 2019). Our results are in line with previous literature in the field of multivariate methods. In recent analyses (Liang et al., 2018; Squarcina et al., 2017), temporal areas were at the base of multimodal classifications of FEP from HC and these regions were also selected to classify FEP with or without auditory hallucinations (Yun et al., 2016). Therefore, these results further confirm that alterations in temporal regions seem to be key factors in the pathophysiology of psychotic symptoms and might represent a hallmark from the earlier stages of the disease.

3.1.4.4. Impact of clinical variables

The decision scores of our model are predicted by meaningful clinical variables (Fig. 7). Interestingly, the most relevant features of the model (Fig. 7, panel b) are age at onset, illness duration, PANSS positive score, and the presence of affective or non-affective psychosis. Age at onset is considered a robust predictor of outcome in schizophrenia and a strong determinant of neurocognition and social cognition levels (Golay et al., 2017). This clinical aspect seems to correlate with

different patients' profiles that might need differential personalized treatment (Golay et al., 2017). Similarly, the illness duration and especially the duration of untreated psychosis seem to have a prognostic impact, as suggested by recent studies in both bipolar and schizophrenia samples (Penttilä et al., 2014; Serafini et al., 2018) and might help in defining subgroups of patients with potential different clinical courses and treatments (Serafini et al., 2018). Finally, we found a significant correlation with the presence of affective or non-affective psychosis. Although the literature about ML approaches in this field is limited to chronic populations (Suvisaari et al., 2018), the possibility to distinguish the different trajectories of affective and non-affective disorders early in the course of the illness is fascinating. The great challenge for clinicians dealing with patients at their first episode is to predict different disease courses and make the best treatment choices for individual patients (Suvisaari et al., 2018). These correlations, although carried out in a smaller sample that had this information available, are critical for the idea of developing a clinically beneficial tool (Koutsouleris et al., 2018; Pigoni et al., 2019). Even though a recent meta-analysis (Lally et al., 2017) substantially reduced the importance of clinical variables as predictors of remission and recovery in FEP, the

discovery of putative biomarkers that help in predicting outcomes would have a great impact on patients' lives and assessment.

3.1.4.5. Limitations

Our primary limitation was that at the time of scanning, the majority of FEP patients were medicated. Antipsychotic exposure, though limited in time and not significantly correlated with decision scores, may still have contributed to the classification. Studies conducted on drug-naïve individuals in large samples are warranted.

Secondly, the generalizability of our result is partial, since it indicates that the classifier does not perform well when applied to data from different sites, thus limiting the clinical applicability. As discussed, heterogeneity may be due to differences across sites but is also embedded in the diagnosis of FEP, given the fact that it includes several diagnostic categories. The multisite structure of our study increases the overall heterogeneity of the sample, given the differences in demographics reported. However, this is a limit in every multisite study and should always be considered. We tried at

least partially to assess this heterogeneity proving that ML classification could be deeply affected by it.

Third, the multi-site nature of the study posed some limitations in the assessment of associated variables such as handedness and cannabis smoke. The different selections performed in the different sites did not allow a complete evaluation of these variables, which might have an impact on the classification. One of the future challenges in multisite studies is the ability to obtain homogeneous and consistent sociodemographic variables to use for the classification or a better stratification of the sample.

Moreover, we did not evaluate subcortical brain regions. These areas might have a role in psychiatric disorders, starting from the early stages of the disease (Cuesta et al., 2017), even though they proved heavily affected by antipsychotics, leading to difficulties in disentangling their own effect from the effect of medications. Follow-ups studies are needed to better assess the role of these regions on psychosis and their possible impact on machine learning classification and predictive models.

Finally, the last limitation is related to the cross-sectional design of the study. A longitudinal design with an MRI scan at

repeated points in time would be useful in order to assess the brain changes over time.

3.2 Clustering studies in FEP

ML and big data proved useful also in defining subgroups and subpopulations. Psychiatric diagnosis are often "umbrella" diagnoses, based on association of symptoms. These diagnoses do not have an underlying biological model (e.g., endophenotype) to rely on. This is of paramount importance when we assess FEP, since their trajectories and severity can differ a lot. One of the biggest challenges of modern psychiatry is the possibility to stratify an individual's prognostic outcome in order to provide the right and necessary treatment to the correct patient.

Currently, this stratification is based on clinical symptoms and the course of the disease, but in the beginning of the illness prognoses are often inaccurate (Wunderink et al. 2009). ML research has demonstrated promising results in predicting illness trajectories in different disorders and defining clinical meaningful subtypes, using different modalities and features such as clinical features and neuroimaging to predict for example the progression to treatment-resistant depression (Pigoni et al., 2019) and the diagnosis of post-partum

depression (Cellini et al., 2022). A very important example of trajectories prediction regards suicidal behaviours. A recent systematic review from our group (Pigoni et al., 2024) suggested how ML can achieve a good prediction of suicidal behaviours in psychiatric population (with accuracies and AUC above 70%); however, many pitfalls are still present and none of these algorithms has found a clinical application in the psychiatric field so far.

I provide here three distinct examples from our research group of definition of clinical subtypes in FEP and their impact on prognosis, using different features.

3.2.1. The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study (Oezturk et al., 2021)

In the study of Oezturk and colleagues (Oezturk et al., 2021), the aim was to identify a stable and generalizable FEP patient stratification based on distinct patterns of formal thought disorders identified by means of data-driven unsupervised machine learning techniques, in a sample of 279 individuals experiencing a recent onset psychosis between 15 and 40 years of age from the PRONIA consortium (pronia.eu).

Using as features the items that assessed positive and negative formal thought disorders from the Positive and Negative Symptom Scale (PANSS) (Kay et al., 1987) and the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen, 1989), we performed a clustering in order to evaluate whether it is possible to identify robust subtypes of patients with recent-onset psychosis that are characterized by distinct patterns of formal thought disorders. To do so, we implemented a multistep pipeline to choose the best clustering algorithm and test its stability and generalizability.

First, we used the ClValid (Burock et al., 2008) package to assess three clustering algorithms: (i) k-means, (ii) hierarchical clustering, and (iii) partitioning around medoids with the number of clusters ranging from 2 to 10. ClValid compares solutions from several clustering algorithms while the numbers of clusters vary, allowing researchers to choose the optimal algorithm and number of clusters with a majority rule based on a battery of internal validity and stability measures.

Second, we retested the winner algorithms for the optimal number of clusters with the NbClust (Charrad et al., 2014) package, that provides an automatized scan through 26 validity indices and reports the optimal number of clusters for a given clustering algorithm with majority rule.

Third, we examined the stability and generalizability of the winning algorithm solution. To exclude possible package or function-specific effects, we retested the robustness of the clustering solution with the R package ClusterStability (Lord et al., 2017). This package provides a global stability (ST) index and an individual STindex based on stability measures.

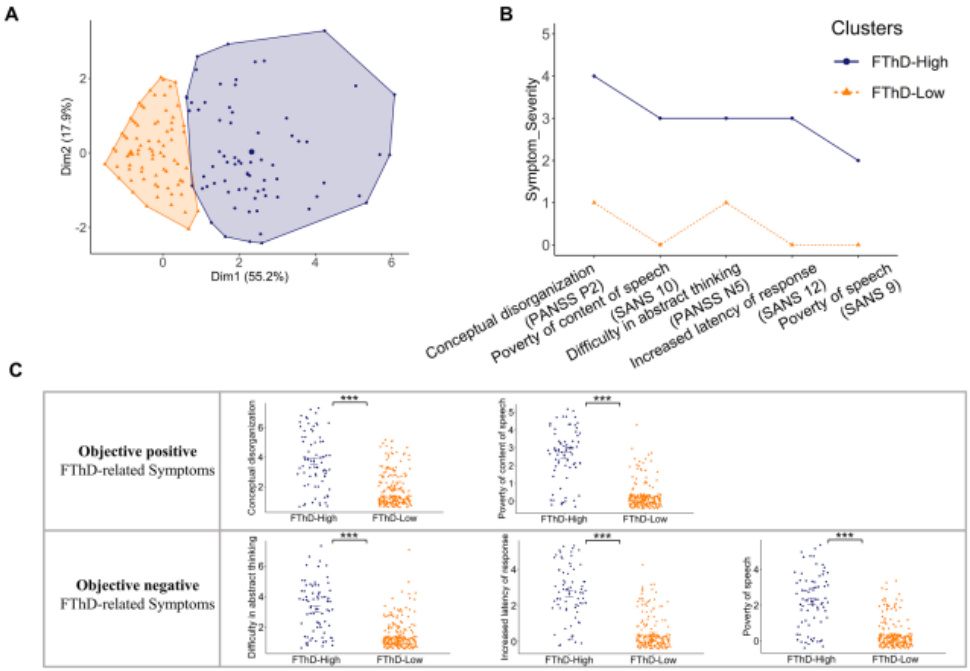
Lastly, we investigated the predicted clustering strength of the partitioning algorithms using the predict.strength package

(Tibshirani et al. 2005), which applies the principles of cross-validation (see section 3.1) to unsupervised ML by performing n-fold random resampling, partitioning the resampled population into training and test data folds, and clusters these through iterations.
For each iteration, the centroids of training data are applied to test data to compute the proportion of pair–observations falling into the same cluster with the centroids of test data.

Our multi-step clustering analyses identified a k-means algorithm-based two-clusters solution as the most stable and generalizable stratification approach. This approach delineated two subgroups, High and Low formal thought disorder (n= 75 vs. 204) in our patient cohort (Fig. N). The clustering solution was significantly informed by (1) observed positive thought disorder as measured by conceptual disorganization (PANSS P2) (median = 4 vs. 1, pfdr < 0.001, r= 0.488) and poverty of content of speech (SANS 10) (median= 3 vs. 0, pfdr< 0.001, r = 0.712), and (2) observed negative thought disorders as measured by difficulty in abstract thinking (PANSS N5) (median=3 vs. 1, pfdr<0.001, r=0.503), increased latency of response (SANS 12) (median = 3 vs. 0, pfdr < 0.001, r = 0.653)

and poverty of speech (SANS 9) (median = 2 vs. 0, pfdr < 0.001, r= 0.611).

Figure 8: The Psychopathological Comparison of FTD subgroups. Panel A Represents the results of the principal component analysis in twodimensional space, panel B the difference between medians of symptom severity, panel C the distributions of each symptom and their statistical comparisons with Wilcoxon rank-sum test. Statistical significances are shown ***pfdr< 0.001



We observed no significant interaction of the subgroups with sex (χ^2= 0.7053, pfdr = 0.401, phi = 0.050) and site (χ^2= 16.452, pfdr = 0.073, phi = 0.243). At baseline, the High group was younger than the Low group (median = 23 vs. 24, pfdr = 0.022, r= 0.153) in group level comparison. The two groups also differed in their education level: fewer years in education in High group than in Low group (median = 12 vs. 14, pfdr = 0.002, r= 0.209). Furthermore, education years and age were positively correlated in each subgroup: FTD-High ($p < 0.001$, r= 0.51) and FTD-Low ($p< 0.001$, r= 0.36).

Cluster assignment was not influenced by global disease severity as measured by PANSS total score at baseline (pfdr = 0.779, r = 0.017) but was associated with more pronounced negative symptoms as measured by SANS total score at baseline (pfdr < 0.001, r = − 0.551).

As reported in Figure 9, the two clusters significantly differed in several functioning and cognitive aspects, including social functioning, verbal and semantic fluency, verbal short-term memory and abstract reasoning.

Figure 9: Correlogram among FTD-related symptoms, functioning and neurocognition; A FTD-High B FTD-Low. Different colours: red =negative or blue= positive represent the direction of correlations, different size of the circles represents the strength of the correlations. Statistical Significances are shown; *pfdr< 0.05, **pfdr< 0.01, ***pfdr< 0.001

There correlations suggest the that the two clusters obtain via data-driven approaches contain clinically meaningful aspects, that might help care-giver informed decision processing by stratifying the subjects into clusters of severity.

Of note, some limitation should be taken into account when evaluating these results. First of all, only few variables were used to inform the algorithm, suggesting that increasing the complexity of the model could significantly alter the results. Second, given the presence of two clusters, we cannot exclude the possibility of a continuum or spectrum. The clustering algorithms are not trained to give a continuative answer, but they rather create clusters. Therefore, it is not possible to assess with these methods weather clusters based on symptoms are better depicted by a continuum spectrum of severity of the disorder. Third, the presented study did not differ affective and non-affective psychosis and assessed such patient in a cross-sectional manner, limiting the speculation regarding the future validity of the obtained clusters and their prognostic value. Lastly, a missing external validation sample restricts the generalizability of these clustering solutions.

3.2.2. Psychosis brain subtypes validated in first-episode cohorts and related to illness remission: results from the PHENOM consortium (Dwyer et al., 2023)

Biological heterogeneity underlying psychosis has prevented the identification and implementation of precision medicine approaches in the clinical care of this disorder. A long history of research into the relationship between brain morphology and diagnoses, illness stages, or symptoms (Haijma et al., 2013) has used topdown approaches, but the lack of clinical translation (Abi-Dargham et al., 2016) has motivated calls for bottom-up approaches to identify neuromorphological subgroups directly from brain scans (Dwyer et al., 2018; Chand et al., 2020). Brain clustering is one approach that aims to find naturally occurring subgroupings of individuals defined by similar brain patterns using brain volume (Chand et al., 2020). Such studies can identify subgroups that explain underlying psychosis heterogeneity in the neuroimaging field and lead to more

precise treatment approaches if related to clinical outcomes, such as symptom remission.

To date, studies have focused on determining neuroimaging subgroups within samples of patients with psychosis in advanced illness stages, reporting two (Chand et al., 2020; Zhao et al., Zao et al., 2022; Lubeiro et al., 2016), or more (Honnorat et al., 2019; Xiao et al 2022; Sugihara et al., 2016) subgroup solutions. Such studies generally find separations between groups based largely on the severity of the brain differences in the patient population related to controls. Evidence also suggests that some of the elements crucial to the subgrouping may not be specific to psychosis, as similar elements are found across diagnoses, in depression risk samples (Kaczkurkin et al., 2020), and to some degree in samples of healthy controls (Lubeiro et al., 2016; Sugihara et al., 2016). These findings point to the presence of a brain-risk phenotype that is normatively present in the healthy population but is enriched in samples of individuals with psychosis. However, a major limitation in this research has been the limited internal or external validation of the subgroups in independent samples. Furthermore, commonly used clustering algorithms used for the subgroupings are often confounded by disease-unrelated

factors that influence brain variability across individuals, such as age and sex (Varol et al., 2016).

The recent PHENOM consortium aimed to address these limitations by applying a semi-supervised machine learning technique (heterogeneity through discriminative analysis; HYDRA (Varol et al., 2016)) to cluster brain MRI volume in a large sample of schizophrenia patients with internal validation (i.e., cross, splithalf, and leave-site-out validation) (Chand et al., 2020). In this work, we found a stable and highly reproducible two-subgroup solution. One subgroup of patients (SG1) demonstrated widespread smaller cortical volume. For the first time, we also found a subgroup consisting of approximately one-third of patients that showed no cortical reductions compared to healthy controls but presented with larger striatal volumes (SG2). While these morphological alterations were not associated with baseline symptoms, individuals in SG2 exhibited higher educational attainment.

In our recent follow-up work, the PHENOM subgroups were also found to be expressed in population samples both in young adults (16–23 yrs), where greater SG1 membership was found in youth with psychosis spectrum symptoms, and in adults (44–50 years) where SG1 was associated with lower cognitive

performance and higher genetic risk for schizophrenia (Chand et al., 2022). These findings suggest that SG1 is associated with normative psychosis risk that is enriched in clinical samples while SG2 is not and may even be protective. A limitation of ours and other work thus far, however, is that we do not know how these subgroups are expressed at the time of a first-episode of psychosis (FEP) when there is a confirmed psychosis diagnosis but less influence of illness or treatment confounds. Determination of FEP subgroup membership is also critical because this is a clinical window when decisions may influence the symptom and functioning illness course (i.e., prognoses).

3.2.2.1 Methods

We previously (Chand et al., 2020) created a two-group structural MRI statistical model in a sample with established schizophrenia from USA (n = 96), Germany (n = 145), and China (n = 66) in addition to controls from the same sites (n = 364). For the current study, we included a PHENOM subsample of 572 first-episode psychosis (FEP) patients, collected from 4 sites: Sao Paulo (n = 128; from two independent studies), Santander (n = 186), London (n = 122), and Melbourne (n = 136). In

addition, 424 healthy control (HC) participants without Axis-I diagnoses were included from the same sites.

Demographic and clinical measures were collected, including age, sex, relationship status, employment status, highest education, age at illness onset, duration of illness, duration of untreated psychosis, chlorpromazine equivalent dose, Global Assessment of Functioning (GAF). Symptoms were assessed with the PANSS. Remission was defined as a binary variable. Follow-up timepoints were available at 1-year (Santander/Sao Paulo; n = 261), 3-years (Santander; n = 147), and 5-years (London/Sao Paulo; n = 135).

HYDRA (Varol et al., 2016) is a semi-supervised method that employs a supervised machine learning algorithm (support vector machine; SVM) to determine boundaries that separate controls from patients while simultaneously identifying patient-specific subgroups. Reproducibility of the subgroup solution is assessed by using an internal cross-validation cycle to identify subgroups. This technique was previously applied to volumetric ROIs of the sample with chronic schizophrenia in our original analysis and the solution was validated using split-half and leave-site-out analyses (Chand et al., 2020). Because SVM is a margin-based hyperplane method, each subject

receives a subgroup membership score quantifying the degree to which their multivariate ROI pattern matches the subgroup label based on the distance from the hyperplanes. The continuous decision scores from the two SVMs representing each original subgroup were used in further analyses to provide binary labels within a two dimensional coordinate space, such that: (1) a SG1 'lower brain volume' label was given if the SG1 decision score was positive and the SG2 score was negative; (2) an SG2 'higher striatum volume' subgroup label was given if the SG2 decision score was positive and the SG1 decision score was negative; (3) a SG1 + SG2 'mixed' label was given if the decision score was positive for both SG1 and SG2; (4) a 'None' label was given if the decision score was negative for both SG1 and SG2.

Voxel-wise analysis

Once subgroups were defined in the FEP and HC samples, descriptive whole-brain voxel-wise volumetric analyses were conducted to provide more fine-grained brain volume differences between FEP subgroups, and separately, between the HC subgroups. In each comparison, we contrasted the

'None' subgroup with SG1 and SG2. Regionally linear multivariate discriminative statistical mapping (MIDAS) (Varol et al., 2018) was used for this purpose due to its demonstrated ability to detect sensitive and specific subgroup differences compared to other multivariate methods.

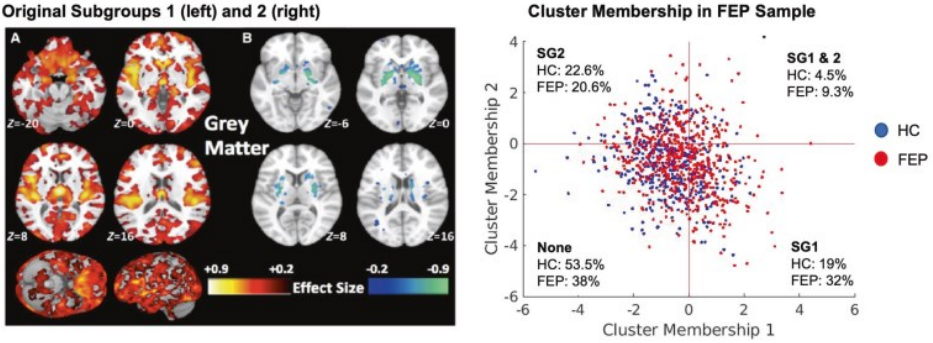
Baseline and longitudinal univariate tests of subtype differences

Baseline differences in demographic characteristics, education, symptoms, and level of functioning variables were assessed across all 4 subgroups using the F-test (continuous) and Chi-square/Kruskall–Wallis (categorical) analyses. Remission was analyzed comparing subgroups for each timepoint using chi-squared analyses. In addition, a composite measure quantifying whether each individual available had ever remitted during any follow-up period (i.e., to identify individuals who exhibit ongoing long-term clinical impairment) was created.

3.2.2.2 Results

Following application of our previously trained HYDRA model (Chand et al., 2020) we obtained the following subgroup split: FEP (n = 572): SG1, n = 184(32%); SG2, n = 118(21%); SG1 + SG2, n = 53(9%); None, n = 217(34%); HC (n = 424): SG1, n = 82(19%); SG2, n = 96(23%); SG1 + SG2, n = 19(5%); None, n = 227(54%) (Fig. XXX). The higher proportion of FEP in SG1 compared to the other subgroups was significant ($X2(8) = 36.5$, $p = 5.96e{-}08$). These findings thus confirm that the two neuromorphological subtypes we had identified in patients in the advanced illness stages are already present at illness onset.

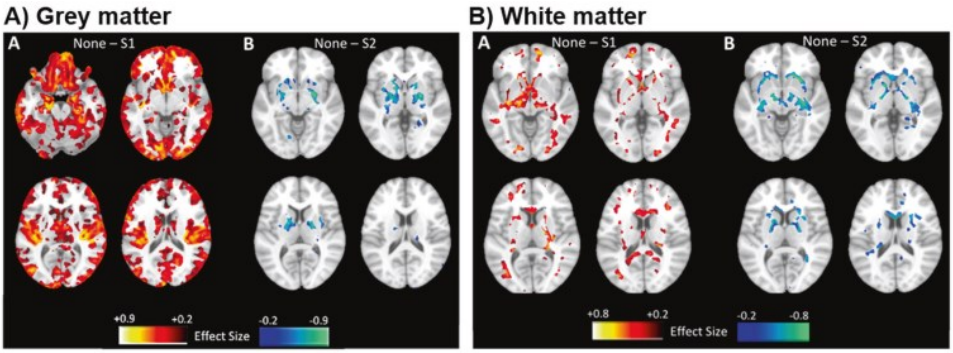
Figure 10: Model application to the FEP and HC samples. A Original subgroups in the sample of individuals with chronic schizophrenia demonstrating subgroup 1 with widespread volume reductions compared to healthy controls and subgroup 2 with no volumetric reductions and increased striatum. B Application of the models to the FEP sample and healthy controls defined the subgroup membership within four quadrants. A proportional difference was found across subgroups, indicating a higher proportion of FEP cases in subgroup 1 and a lower proportion within the 'None' category.



When whole-brain voxel-wise maps of the SG1 subgroup were compared to those of the 'None' subgroup in FEP, the SG1 subgroup showed smaller widespread cortical volumes coupled with smaller volume in some parts of caudate (Figure 11). In contrast, larger gray matter volumes in SG2 subgroup were restricted to subcortical structures including the striatum. White matter was also smaller in SG1, mainly in subcortical

regions adjoining the striatum, whereas white matter was larger in SG2 (Figure 11). Because healthy control participants were also included in the SG1 and SG2 subgroups, we tested whether these subgroups demonstrated similar patterns to the patients and our previous work in population samples.

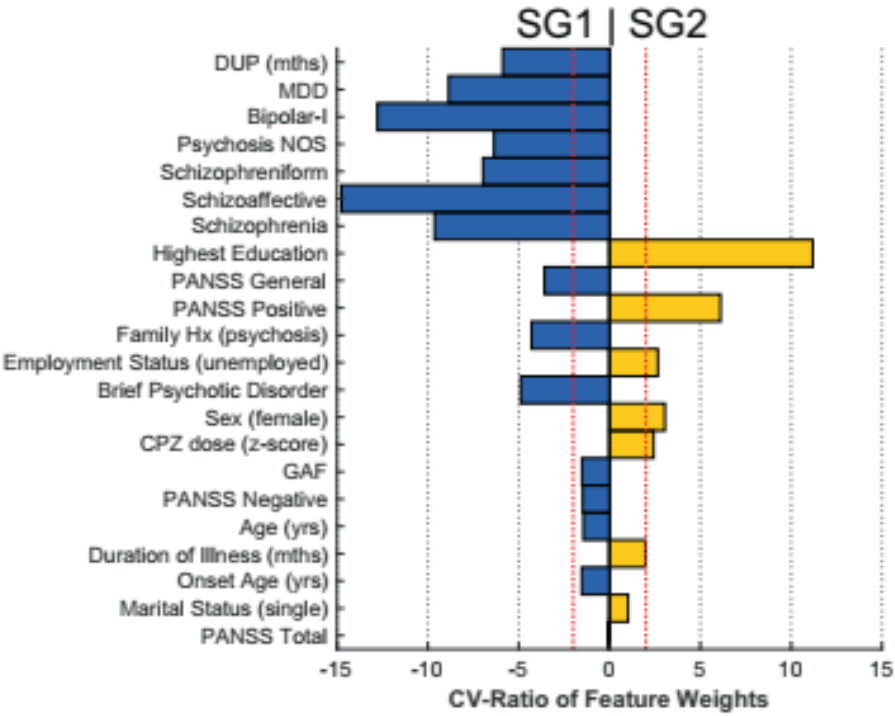
Figure 11: Subgroup comparisons in the FEP sample. Gray matter (A) and white matter (B) comparisons of subgroup SG1 (left) and SG2 (right) when compared to cases in the 'None' subgroup in the FEP sample. A Decreased gray matter volume was found in the FEP subgroup relative to the 'None' classification, including the caudate, whereas relatively increased gray matter was found in SG2 subcortical areas including the striatum. B Decreased white matter was also found in the SG1 subgroup in areas including the internal capsule.



In patients, multigroup machine learning analyses revealed a high accuracy of subgroup separation with the SG1 vs. SG2 comparison. Specifically, SG2 membership was separated from SG1 with a sensitivity of 66% (balanced accuracy (BAC) = 64.03%) by a pattern including: education, higher positive symptoms, unemployment, female sex, and higher CPZ (Figure 12); while longer duration of untreated psychosis, diagnoses (including comorbidity), general psychosis symptoms, and family history of psychosis significantly predicted SG1 membership. Other subgroup comparisons were non-significant.

Figure 12: Baseline multivariate signatures related to the classification of subgroups. Cross-validation ratio is plotted (x-axis) in and sign-based consistency significance is indicated by colored bars (yellow, positive weights; blue, negative weights). SG2 was separated from the SG1 subgroup by positive weights (commonly indicating relative increases) in highest education, positive symptoms, unemployment, female sex, and CPZ dose (z-scored within each site). Prediction of SG2 was also associated with negative weights (commonly indicating relative decreases) of duration of untreated psychosis (DUP), psychotic diagnoses, general psychosis symptoms, and family history of psychosis. PANSS positive and negative syndrome scale, GAF

global assessment of functioning, CPZ chlorpromazine equivalent dose, MDD major depressive disorder, Psychosis NOS Psychosis not otherwise specified.



Finally, regarding prognosis, individuals in SG2 were more likely to have experienced at least one period of symptom remission across all follow-up timepoints (SG1, 57%; SG2, 78%; $X2(8) = 9.8$, $p = 0.02$). For individual follow-up points, higher 1-year remission in SG2 was found at uncorrected levels (SG1, 52.7%; SG2, 73.2%; $X2(8) = 7.7$, $p = 0.05$).

Our findings show for the first time that the previously defined schizophrenia subgroups (Chand et al., 2020), SG1 with lower widespread cortical volumes and SG2 with larger striatal volume but otherwise normal brain morphology, are already visible at the first presentation of illness. Furthermore, the two subgroups present a distinct clinical signature and different prognosis, with SG2 significantly more likely to achieve remission in the following years.
These findings support the presence of reproducible neuromorphological subgroups in individuals with psychosis that may help disentangle the heterogeneity of clinical presentations. With further follow-up studies, the findings could also be crucial in informing clinician and researchers to refine stratified therapeutic approaches.

4. Moving forward – the PHLAMES study (GR-2019-12369479)

The present study, called "neuronal synaptic autoantibodies and PsycHosis: evaLuation of A systematic approach to increase autoiMmune Encephalitis diagnosis in acute adult and pediatric patients and relevance of the autoantibodieS in chronic forms. The PHLAMES Study" has been funded by the Italian Ministry of Health (Bando Ricerca Finalizzata Giovani ***GR-2019-12369479****); I am currently PI of the collaborating center 2 (Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico)*

4.1 Introduction to autoimmune encephalitis

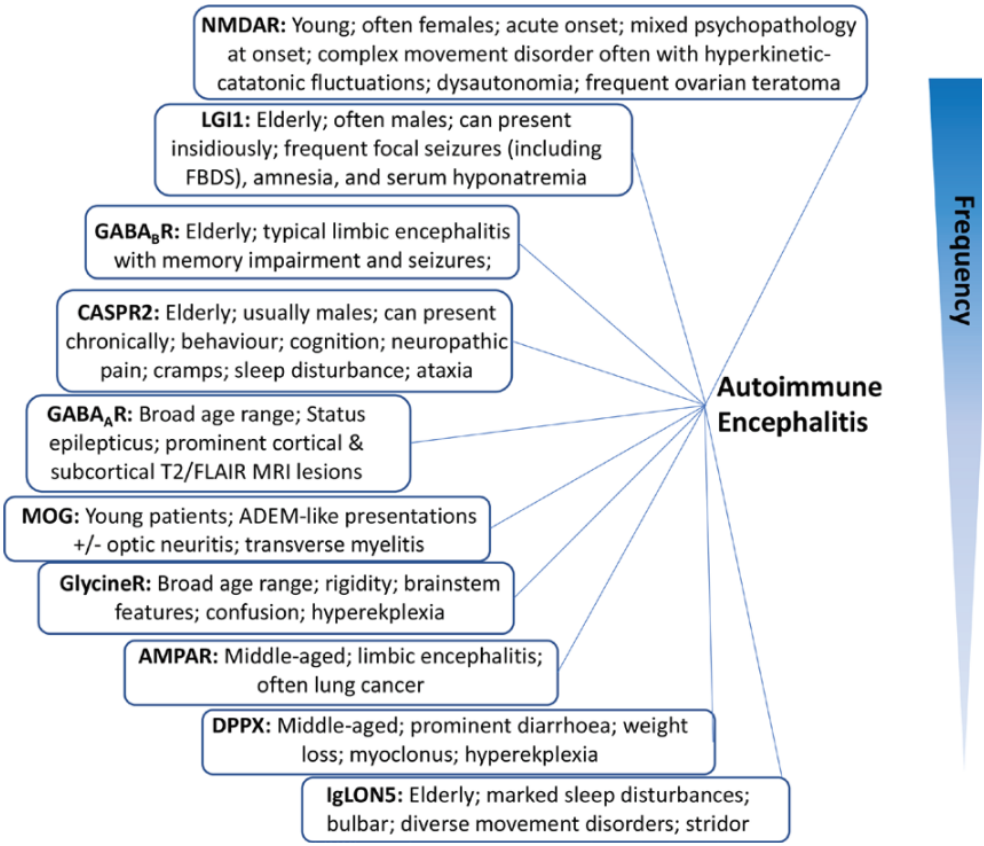
Acute encephalitis is a debilitating neurological disorder that develops as a rapidly progressive encephalopathy (usually in less than 6 weeks) caused by brain inflammation. The estimated incidence of encephalitis in high-income countries is about 5–10 per 100 000 inhabitants per year; encephalitis affects patients of all ages and represents a significant burden to patients, families, and society (Graus et al., 2016). Opposed to infectious encephalitis, in the past 15 years an increasing number of non-

infectious, mostly autoimmune, encephalitis cases have been identified, since the first discovery of encephalitis with antibodies against N-methyl-D-aspartate receptor. These newly identified forms of autoimmune encephalitis might be associated with antibodies against neuronal cell-surface or synaptic proteins and can develop with core symptoms resembling infectious encephalitis, and also with neurological and psychiatric manifestations without fever or CSF pleocytosis (Graus et al., 2016; Pollak et al., 2020). These patients have pathogenic antibodies that target surface epitopes on synaptic and related proteins, with the most common and well-described being the N-methyl-D-aspartate receptor (NMDAR, specifically the NR1 subunit), and the voltage-gated potassium channel (VGKC)-complex proteins, leucine-rich-glioma inactivated 1 (LGI1) and contactin-associated protein-like 2 (CASPR2) (Pollak et al., 2016; Varley et al., 2018; Pollak et al., 2020).

These disorders might present in different ways, also depending on the type of CNS structure involved (see Figure 13), but psychiatric symptoms present in almost 95% of patients and up to 4% patients might present with purely psychiatric

disorder (Titulauer et al., 2013; Keiser et al., 2014; Warren et al., 2018).

Figure 13: clinical presentation and frequency of the most common Autoimmune Encephalitis (from Uy et al., 2021)



In NMDAR antibody-associated encephalitis, patients present with a polymorphic psychosis with prominent affective symptoms and cognitive impairment, but negative psychotic symptoms might also feature, with depression and suicidality relatively common. The rapidly progressive course of the disease and especially of the cognitive impairment is considered quite typical.

In addition to psychiatric disturbance, patients with autoimmune encephalitis develop clear neurological features, including seizures, cognitive dysfunction, and movement disorders (Dalmau et al., 2018).

Regarding imaging, limbic encephalitis is the classical form of autoimmune encephalitis and is associated with unilateral or bilateral hippocampal MRI FLAIR-T2 hyperintensities (Heine et al., 2015; Kelley et al., 2017). However, MRI features are neither sensitive, specific (mainly non-specific white matter changes) and up to 60% of patients show no MRI alterations. Similarly, alterations at the EEG are not sensitive nor specific. However, a very typical EEG pattern—extreme delta brush (ie, δ waves with superimposed beta waves [brush])— can be observed in a small portion of patients (Schmitt et al., 2012; Gillinder et al., 2018).

To confirm the diagnosis, in autoimmune encephalitis, CSF positivity indicating intrathecal synthesis of NMDAR antibodies, is considered necessary and sufficient. Other CSF findings include CSF lymphocytes that is typically moderately increased (even though less than the high values observed in viral encephalitis), mainly during the early stages, with oligoclonal bands appearing later. Lymphocytosis (>5 white blood cells per µL) generally occurs in less than 5% of patients.

However, in some cases, behavioural and psychiatric disturbance dominate the course of autoimmune encephalitis (Yoshimura et al., 2017), without other important neurologic symptoms or imaging signs; these clinical presentations stimulate the search for these antibodies in patients with primary psychiatric disorders such as schizophrenia and first-episode psychosis.

4.1.2 Autoimmune psychosis

Clinical psychiatry is entering an era in which there is an ever-greater emphasis on the early identification and prevention of disease and the development of personalized treatment

approaches (Pollak et al., 2019). As such, biomarkers, which can be diagnostic, prognostic, or predictive, are currently being researched extensively across every area of medicine and psychotic disorders like schizophrenia are no exception. A fundamental problem in the management of psychosis is that outcomes are not predictable on clinical grounds.

Immune biomarkers are now used widely in relation to physical disease, both systemic and CNS-restricted, to improve clinical understanding and management, reflecting an increasing awareness of the involvement of the immune response in aspects of many diseases. With the recognition that psychotic disorders, too, have considerable immune involvement, it is likely that immune biomarkers will play a role in the biological psychiatry of the future.

Focusing more specifically on humoral immunity, the notion that psychosis might be caused by a pathogenic antibody has a long history. A systematic review demonstrated that among patients with established schizophrenia, 20 autoantibodies (including antinuclear antibody [ANA], anti-cardiolipin, anti-N-methyl-d-aspartate receptor [NMDAR], and anti-serotonin) were present at higher rates than among controls; rates of anticardiolipin and anti-NMDAR antibodies were also present

in patients with first-episode psychosis (Ezeoke et al., 2013). Even among the unaffected first-degree relatives of patients with schizophrenia, a higher prevalence of some autoantibodies has been observed. However, the mere presence of an antibody does not imply pathogenicity.

Moreover, the discovery of neuronal surface protein antibodies in autoimmune encephalitis has generated a great deal of interest in the possibility that some psychiatric patients, in particular those with both affective and non-affective psychoses, have a specific autoantibody-mediated disease, or so-called autoimmune psychosis (Pollak et al., 2020)

IgGs for NMDAR have the strongest association with psychiatric features, with or without signs of encephalitis. In a recent meta-analysis (Pollak et al., 2014), the prevalence of positive IgG NMDAR antibodies in serum in individuals with first-episode psychosis varied from 0% to 12%; however, the clear role of these antibodies is ye to be understood.

To date, a diagnostic algorithm to exclude or confirm the diagnosis of autoimmune psychosis is not available yet, since the findings are often mixed. A position paper by Pollak and colleagues (Pollak et al., 2020), proposed diagnostic criteria for

autoimmune psychosis (divide into possible, probable and definite criteria), that are reported in Panel ZY.

Panel 1: proposed diagnostic criteria for autoimmune psychosis, from Pollak et al. 2020.

Panel 1: Proposed diagnostic criteria for autoimmune psychosis

For a diagnosis of possible autoimmune psychosis:

The patient must have current psychotic symptoms of abrupt onset (rapid progression of <3 months) with at least one of the following:

- Currently or recently diagnosed with a tumour
- Movement disorder (catatonia or dyskinesia)
- Adverse response to antipsychotics, raising suspicion of neuroleptic malignant syndrome (rigidity, hyperthermia, or raised creatine kinase)
- Severe or disproportionate cognitive dysfunction
- A decreased level of consciousness
- The occurrence of seizures that are not explained by a previously known seizure disorder
- A clinically significant autonomic dysfunction (abnormal or unexpectedly fluctuant blood pressure, temperature, or heart rate)

If a patient has possible autoimmune psychosis, they should be investigated as per section 5 ("Consensus multimodal approach to the systematic investigation of patients with suspected autoimmune psychosis"), including electroencephalography , MRI, serum autoantibodies, and cerebrospinal fluid (CSF) analysis (including CSF autoantibodies). The results should lead to a diagnosis of non-autoimmune psychosis or probable/definite autoimmune psychosis.

For a diagnosis of probable autoimmune psychosis:

The patient must have current psychotic symptoms of abrupt onset (rapid progression of <3 months) with at least one of the seven clinical criteria listed above for possible autoimmune psychosis and at least one of the following:

- CSF pleocytosis of >5 white blood cells per µL
- Bilateral brain abnormalities on T2-weighted fluid-attenuated inversion recovery MRI highly restricted to the medial temporal lobes

Or two of the following:

- Electroencephalogram encephalopathic changes (ie, spikes, spike-wave activity, or rhythmic slowing [intermittent rhythmic delta or theta activity] focal changes, or extreme delta brush
- CSF oligoclonal bands or increased IgG index
- The presence of a serum anti-neuronal antibody detected by cell-based assay after exclusion of alternative diagnoses.

For a diagnosis of definite autoimmune psychosis:

The patient must meet the criteria for probable autoimmune psychosis with IgG class anti-neuronal antibodies in CSF.

Note that these criteria do not exclude a diagnosis being made in a patient with an acute onset (<3 months) of psychosis, even if that patient has had a previous psychotic, other psychiatric, or encephalopathic episode that resolved.

Although useful, these criteria suffer from several issues. First of all, they are very restrictive on the timescale. Given the usual duration of untreated psychosis, it is unlikely that a 3-month cut-off would represent the standard of presentation of such patients. Moreover, it would be interesting to assess whether signs of immune psychosis are present also in more chronic patients, given that antibodies have been found also in this population. Second, to achieve a definite diagnosis, invasive testing (e.g., lumbar puncture) or secondary screening is needed. In this sense, it would be helpful to reach a strategy to stratify risk and therefore implement screening and testing accordingly.

For this very reason, several authors have proposed lists of clinical red flags (and some have proposed so-called yellow flags) that should raise suspicion of CNS autoimmunity in patients presenting with psychosis (Al-Diwani et al., 2017; Herken et al., 2017) (Panel 2)

Panel 2: red flags for suspicion of autoimmune encephalitis in patients with psychosis

Panel 2: Red flags for suspicion of autoimmune encephalitis in patients with psychosis*

- Infectious prodrome
- New-onset severe headache or clinically significant change in headache pattern
- Rapid progression
- Adverse response to antipsychotics or presence of neuroleptic malignant syndrome
- Insufficient response to antipsychotics
- Movement disorder (eg, catatonia or dyskinesia)
- Focal neurological disease
- Decreased consciousness
- Autonomic disturbance
- Aphasia, mutism, or dysarthria
- Seizures
- Presence of a tumour, or history of a recent tumour
- Hyponatraemia (not explained by side-effects of medication, eg, SSRIs, carbamazepine, and others)
- Other autoimmune disorders (eg, systemic lupus erythematosus, autoimmune thyroid disease)
- Paraesthesia

*Based on Al-Diwani and colleagues[3] and Herken and Prüss.[42] See panel 1 for the criteria required for a diagnosis of autoimmune psychosis.

These red flags, although based on available evidence, should in any case be considered provisional and still require clinical validation.

4.2 the PHLAMES study: rationale and aims

The discovery of autoantibodies to synaptic neuronal antigens (NSAbs) has revolutionized clinical neuroimmunology (Graus et al., 2016). NSAbs associate with autoimmune encephalitis (AE), a heterogeneous group of diseases combining both neurological and psychiatric features, being fundamental for their diagnosis (Graus et al., 2016). Notably, NSAbs are considered pathogenic, and the administration of immunosuppressive therapies can lead to dramatic responses in most AE patients (Titulaer et al., 2013). NSAb discovery has had a great impact also in the psychiatric field, since AE is an important differential diagnosis of psychosis spectrum disorder. Indeed, psychotic symptoms are present in up to 95% of patients with N-methyl-Daspartate receptor (NMDAR) encephalitis (Dalmau et al., 2019), the most common AE, and in over 60% of patients with other AE. Up to 1/3 of AE patients is initially assessed in psychiatric facilities.

NSAb screening in all first episodes of psychosis could be useful for diagnostic purposes (Pollak et al., 2019), but this is routine practice, and patients often undergo further examinations, including brain MRI, electroencephalography, and lumbar puncture only when overt neurological signs, such as tonic-clonic seizures, manifest. In addition, a small proportion of patients with NMDAR encephalitis can show pure psychiatric presentations, further hampering the AE diagnosis (Kaiser et al., 2014; Warren et al., 2018).

Even NSAb screening can be problematic. Most laboratories use commercial fixed CBAs that can perform sub-optimally. In-house assays including live CBAs and tissue-based assays (TBAs) on rat brain optimized with light fixation procedures improve accuracy but are rarely available in routine diagnostics.

For the above-mentioned reasons, AE are likely under-recognized, but the frequency of misdiagnosis in general, and in psychiatric care in particular, is unknown.

Hypothesizing the existence of mild encephalitic forms masquerading as chronic diseases (Pollak et al., 2019), a possible role of NSAbs has also been studied in chronic psychotic

disorders, such as schizophrenia, with inconsistent results (Dahm et al., 2014). The frequency of NSAb-positive psychotic patients is 3-5% when testing for NMDAR IgG, and 7% if including the other most common NSAbs (Scott et al., 2018). In these patients, NSAbs are often found in serum, but not in the CSF, with similar, or slightly higher frequencies vs controls, thus triggering engaging debates on their clinical meaning and pathogenic relevance fueled by immune-mediate inflammation (Dahm et al., 2014; Scott et al., 2018).

Therefore, the PHLAMES study decided to focus on the following aims:

AIM 1: To evaluate the impact of a systematic diagnostic approach on the frequency of AE diagnosis in acute patients at their FEP referred to adult and pediatric psychiatric care (PHLAMES I);

AIM 2: To investigate neuronal damage and BBB impairment in chronic patients with NSAb-pos-psychoses, searching for mild AE forms and therefore to characterize novel and potentially clinically relevant NSAbs in psychotic spectrum (PHLAMES II)

4.3 Methods

4.3.1: PHLAMES I

This study aims to refine the current psychiatric management of patients that can possibly have AE. In parallel, it provides the opportunity to evaluate the performance of the recent criteria for autoimmune psychosis in a real-life setting.

The design entails the application of a systematic diagnostic approach relying on an advanced NSAbs testing (See Figure 13).

All patients referring to acute psychiatric care fulfilling the criteria will be included:

- individual with psychosis with onset and progression in the previous 6 months (lasting at least 10 days); to detect an improvement of 5% in the diagnosis of AE in the psychiatric setting through the application of our protocol, we will need to evaluate 163 FEP patients to reach a significant difference with an 80% power.

- Age between 10-60 years

Exclusion criteria: head injury, intellectual disability. Of note, abuse of alcohol or other substances is not strictly considered an exclusion criterion; single cases will be evaluated by clinicians.

Patients will undergo the following investigations:

a) serum NSAb panel: Screening for NSAbs will include: conformational TBA on rat brain, Live CBA for NMDAR, GABABR, GABAAR, LGI1, CASPR2, AMPAR1/2, Live CBAs for rare antigens according to the TBA staining pattern, Live neuronal cultures; NSAbs results will be available within 10 days, allowing timely treatment.

b) neurological evaluation;

c) psychiatric evaluation; Psychiatric scales will also be administered: structured interview according to the manual for assessment and documentation of psychopathology (AMDP); brief psychiatric rating scale (BPRS); Social and Occupational Functioning Assessment Scale (SOFAS), clinical global impression severity scale (CGI-S); Positive and Negative

Symptoms Scale (PANSS); K-SADS questionnaire (for pediatric patients) d) cognitive evaluation, using the Montreal Cognitive Assessment (MoCA); PRONIA Cognitive Battery (only for NSAb positive)

e) Evaluation for the red flags (Panel 2).

Patients with positive serum NSAbs and/or any red flag will perform further evaluation including:

- brain MRI (T1-weighted sequences)

- EEG

- lumbar puncture

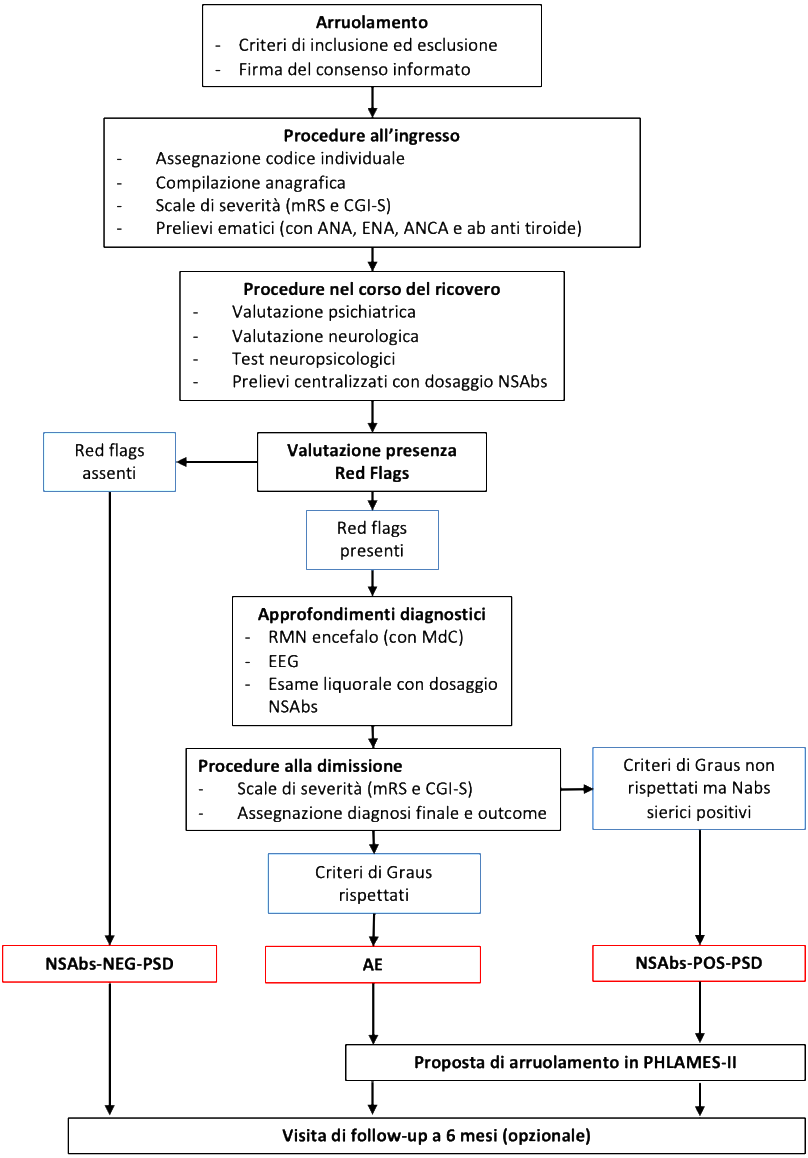
- CSF analysis and a complete cognitive assessment.

Patients will be ultimately diagnosed as AE, NSAb-pos-psychosis, or NSAb-neg-psychosis, and treated following good clinical practice. Patients will be followed-up at 6 months to evaluate disability and response to immunosuppressive treatment (if administered).

Primary endpoint: frequency of patients diagnosed with AE during the study compared with the number of AE diagnoses performed in the same psychiatric facilities in the 2 years preceding the study.

Secondary endpoints: clinical and paraclinical features predicting a final diagnosis of AE or probable/definite autoimmune psychosis, focusing on the relevance of each red flag.

Figure 14: methodologic algorithm for PHLAMES I



4.3.2: PHLAMES II

The second branch of the PHLAMES study attempts to characterize the challenging group of patients with chronic NSAb-pos-psychoses, searching for mild/subclinical forms of AE, and to support the role of NSAbs as biomarkers of specific clinical phenotypes in psychotic subjects.

Patients accessing to psychiatric outpatient clinics or residential facilities fulfilling the criteria will be included:

- psychosis onset between 6 months and 2 years but still symptomatic (CGI-S >4)

- Age between 10-80 years

At the end of the study, four population will be described: AE, NSAb-pos-psychosis, NSAb-neg-psychosis, and HC

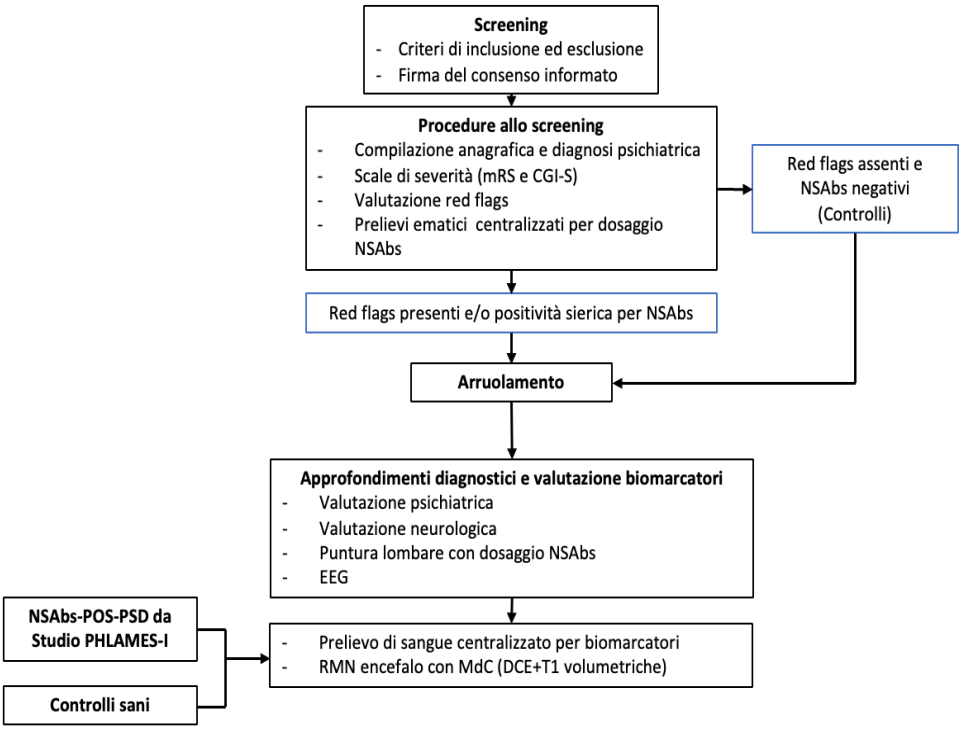
The patients will be screened for serum NSAbs and if positive, a retrospective revision of clinical data will be performed. If any red flags are detected, they will undergo the same investigations as per PHLAMES I (see algorithm in Figure 14). Patients ultimately diagnosed as NSAb-pos-psychosis, will further perform investigations to assess:

a) BBB dysfunction: Serum S100B levels; Brain Diffusion Contrast enhanced MRI (DCE), measuring changes in brain tissues over time after bolus administration of gadolinium, to evaluate BBB permeability;

b) neuronal damage: serum neurofilaments; volumetric brain 3T MRI with T1 3D-weighted images c) immune activation markers: Immune gene expression analysis on peripheral blood

The same analysis will be performed on a control group of age and sex matched healthy volunteers or disease controls with neurological syndromes without known neuronal damage or BBB dysfunction (headache, non-inflammatory peripheral neuropathies), and no history or current psychiatric disorders.

Figure 15: methodologic algorithm for PHLAMES II



4.3.3 Preliminary analyses

For these first analyses, a preliminary sample was employed. All the subjects went through the blood analyses and evaluation of serum antibodies and clinical red flags. Moreover, they were all assessed in terms of clinical and demographic data, psychiatric scales (e.g., PANSS, BPRS), neurologic evaluation, cognitive evaluation and treatment response.

In a sub-group also T1 MRI images were available.

For the analyses, subjects were divided into NSAbs-POS-psychosis and NSAbs-NEG-psychosis. All the analyses in terms of clinical and demographic variables were assessed using SPSS 22.0; IBM Corp., Armonk, NY, USA. Demographic variables were assessed using t-test for continuous and chi-square for discrete variables. Repeated multiple ANOVAs were used to assess the differences in terms of clinical variables. The variables were grouped as follow:

A. Demographic variables and inflammatory status (including BMI, family history, neurologic or inflammatory disorders, smoke, cannabis, substance use, history of neoplasia).

B. Clinical presentation, including duration of untreated psychosis, psychotic symptoms, behavioral symptoms, catatonia, suicidal behaviours.
C. Neurological signs or symptoms, including cognitive disturbances, movement disturbances, sleep alterations, speech, alterations.
D. Clinical scales (PANSS, BPRS, HAMD).
E. Neurocognitive test administered.
F. Treatment and side effects
G. Blood test
H. Red flags

These variables have been selected based on the literature, that suggest that typically AE are characterized by more evident neurological disturbances, cognitive alterations, altered treatment response, disproportioned side effects to antipsychotics treatment, when compared to prototypical psychotic patients (Uy et al., 2021).

Moreover, a subgroup presented T1 MRI images that could be analyzed. All the images have been acquired with a Time Field Echo sequence and converted and analyzed within Matalab

using Statistical Parametric Mapping (SPM) 12. Voxel based morphometry (VBM) was used to analyzed differences between the two groups of patients.

4.3.4 MRI preprocessing

A sub-sample of patients (3 NSAbs-POS-PSD and 7 NSAbs-NEG-PSD), all from the center of Milan, performed structural MRI.

The raw MRI scans were processed after acquisition using a gold-standard protocol. Firstly, images underwent a visual check to assess their quality and were converted from DICOM to NIFTI, the standard format for neuroimaging analysis software. The preprocessing pipeline was performed in Matlab R2018a (The Mathworks, Inc®) environment.

The preprocessing was performed using SPM version 12 Software (Penny et al., 2007) hiip://www.fil.ion.ucl.ac.uk/spm/software/spm12/), using the preprocessing module of the "basic VBM analysis" described in chapter 3.1 Brain Morphological Feature Extraction. The used parameters were the default ones for the

basic VBM analysis (Gaser, 2017). VBM is the study of local brain regions' sizes and shapes compared at a voxel-wise level for a given population. To do this, brain images need to be processed and tissues segmented, as the raw ones present noise, intensity-inhomogeneities, vessels, etc., all details that are not reliable or interested to study. Besides, to perform a group analysis or comparison one wants to eliminate all variations that are external and unrelated to group differences. The idea is to extract volumetric and cortical thickness measures from ROIs, as well as the Total Intracranial Volume (TIV), from each brain scan. Due to anatomy inter-subject variability, brain image scans contain unwanted spatial variations, but scans need to match spatially between each other so that the location of each region in one scan can correspond to the same location in another scan so that a direct comparison can be performed. Therefore, scans must be aligned in the same space, and shot to the same brain template, so that the remaining differences account for group differences and nothing else.

The preprocessing pipeline followed the hereunder steps:

1. Data Segmentation

a. Writing Options: Process Volume ROI – Atlases – CoBra;

2. Display Slices – Quality check after segmentation;

3. Estimation of Total Intracranial Volume (TIV)

5. Data Smoothing

a. FWHM: 6mm

The brain ROI measures included in this work regarded only Gray Matter (GM) and were extracted from 52 subcortical regions of CoBra Atlas for volumes and 68 cortical regions of the Desikan-Killiany Atlas.

4.4 First results

The recruitment is still ongoing, and it is expected to end by June 2024. I report here the first preliminary results.

We preliminary confronted FEP subjects that presented serum autoantibodies against CNS structures to those who did not show them. In this preliminary sample, 56 FEP were included.

Of those, 12.8% presented one or more circulating autoantibodies against CNS structures (See table 11).

Table 11: AB positivity observed in the preliminary sample.

Subject	**AB positivity 1**	**AB positivity**
PHL1_PV1	L-GABA$_B$	
PHL1_PV10	IHC IgG	L-GlyR
PHL1_PV12	L-NMDAR	
PHL1-VO2	L-NMDAR	
PHL1_MI2	IHC IgG	
PHL1_MI9	L-Caspr2	Pann Enc
PHL1_MI11	L-GlyR	

Table Xserum : NSAbs

Legenda: L-GABA$_B$: γ-aminobutirric acid receptor; L-GlyR: glycine receptor; IHC IgG: aspecific immunoistochimic positivity; L-NMDR: N-metil-D-aspartate receptor; L-Caspr2: contactin-associated protein-like 2; Pann Enc: apsecific positivity to PANN Euroimmune.

The two groups showed no differences in terms of age, sex, BMI, and education, as reported in Table 12.

Table 12: demographic characteristics of the NSAbs-POS and NSAbs-NEG subjects

	NSAbs-POS-PSD (n=7) mean (SD)	**NSAbs-NEG-PSD (n=48) mean (SD)**	**F/chi square**	**Sign.**
Age	39.6 (20)	33.1 (13.8)	2.02	0.161
Sex	4F/3M	24/23	0.09	0.76
BMI	21.9 (2.9)	22.6 (3.3)	0.77	0.387
School years	14 (2.5)	13 (3.4)	0.1	0.744
Site	4/3	21/27	0.44	0.51
Ethnic group	6/1	32/16	1.04	0.31

ANOVAs were then performed to evaluate differences in terms of clinical, neurocognitive, and neurologic features. Significant differences (uncorrected) were detected for cognitive disturbances (p 0.05), especially working memory (p 0.05), focal sovratentorial symptoms (p 0.008), movement disturbances (p 0.008), including parkinsonism (p 0.008) and postural tremor (p 0.008), and speech disorders with dysarthria (p 0.008). All these features resulted more prevalent in the NSAbs-POS group.

Moreover, PANSS insight items (p 0.024) and somatic implications (0.006) resulted significantly different, with the NSAbs-POS group showing a high score in both insight and somatic implication.

Regarding cognitive tests, differences were found in Raven progressive matrices (p 0.029), with the NSAbs-POS group showing a lower score compared to NSAbs-NEG patients. Even though they did not reach statistically significance, we could observe a trend also in TMT-A (p 0.079), and RAVLT (p 0.067) scores, with NSAbs-POS group showing worse performances.

Finally, the use of antidepressants, resulted significantly different in the two groups (p 0.049), with NSAbs-POS group receiving more antidepressant compared to the negative group.

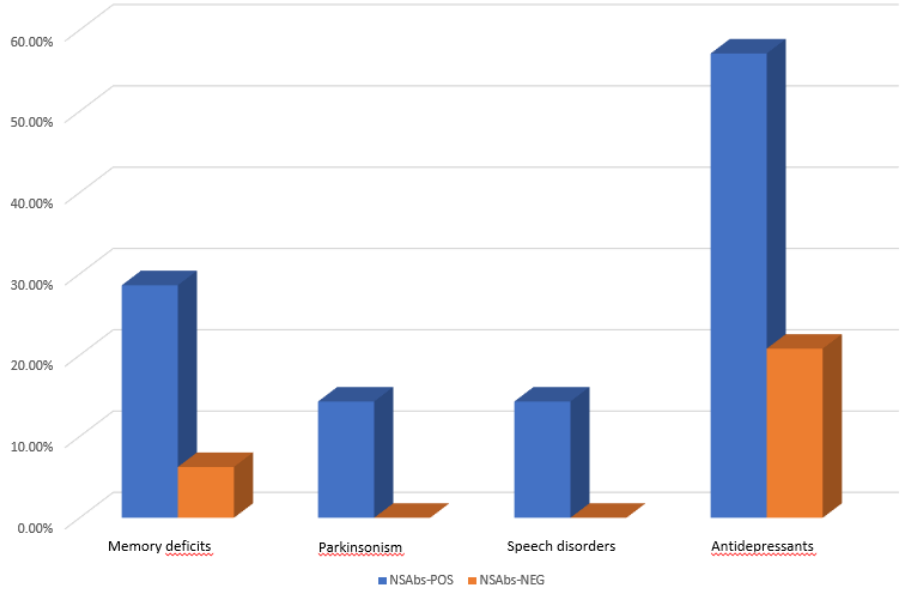
Table 13: variables that resulted significantly different between the groups (uncorrected)

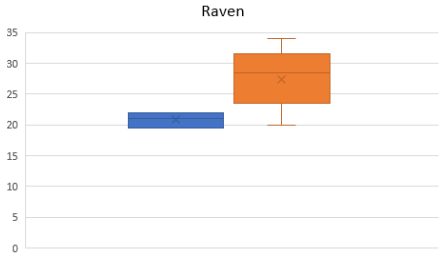
	NSAbs-POS-PSD (n=7)	**NSAbs-NEG-PSD (n=48)**	**F**	**Sign.**
Cognitive disturbances	28.6%	6.25%	3.8	**0.05**
Memory deficits	28.6%	6.25%	3.8	**0.05**
Focal sovratentorial symptoms	14.3%	0%	7.7	**0.008**
Parkinsonism	14.3%	0%	7.7	**0.008**
Postural tremor	14.3%	0%	7.7	**0.008**
Speech disorders	14.3%	0%	7.7	**0.008**
dysarthria	14.3%	0%	7.7	**0.008**
Antidepressants	57.1%	20.8%	4.1	**0.049**
Raven progressive matrices	20.8 (1.2)	27.4 (4.6)	5.72	**0.029**
RAVLT	3.0 (3.0)	6.1 (2.8)	3.86	0.067
TMT-A	68.0 (6.2)	49.3 (16.7)	3.5	0.079

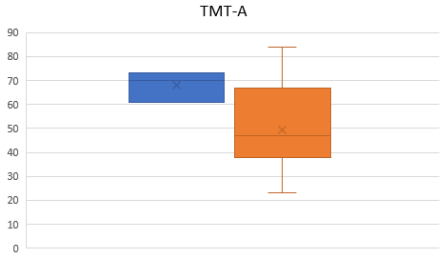
PANSS – somatic implication	3.6 (2.6)	1.9 (1.2)	8.0	**0.006**
PANSS – insight	2.7 (1.9)	4.1 (1.5)	5.45	**0.024**

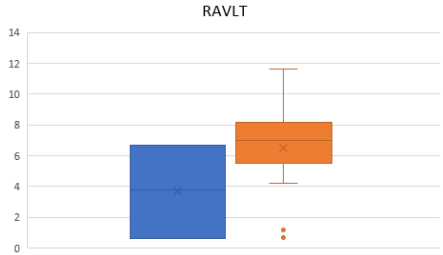
The results are also shown in the Figure 16

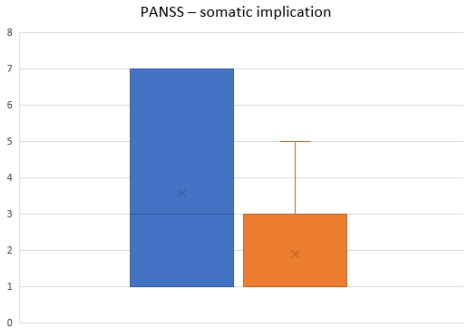
Figure 16: comparisons between significant variables among the two groups. Panel A shows significant differences in neurological aspects; Panel B presents differences in cognitive tests; Panel C focuses on PANSS scores differences.

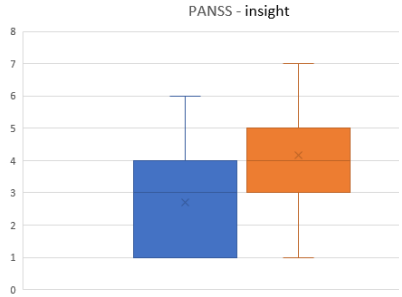



Raven
35
30
25
20
15
10
5
0


TMT-A
90
80
70
60
50
40
30
20
10
0


RAVLT
14
12
10
8
6
4
2
0





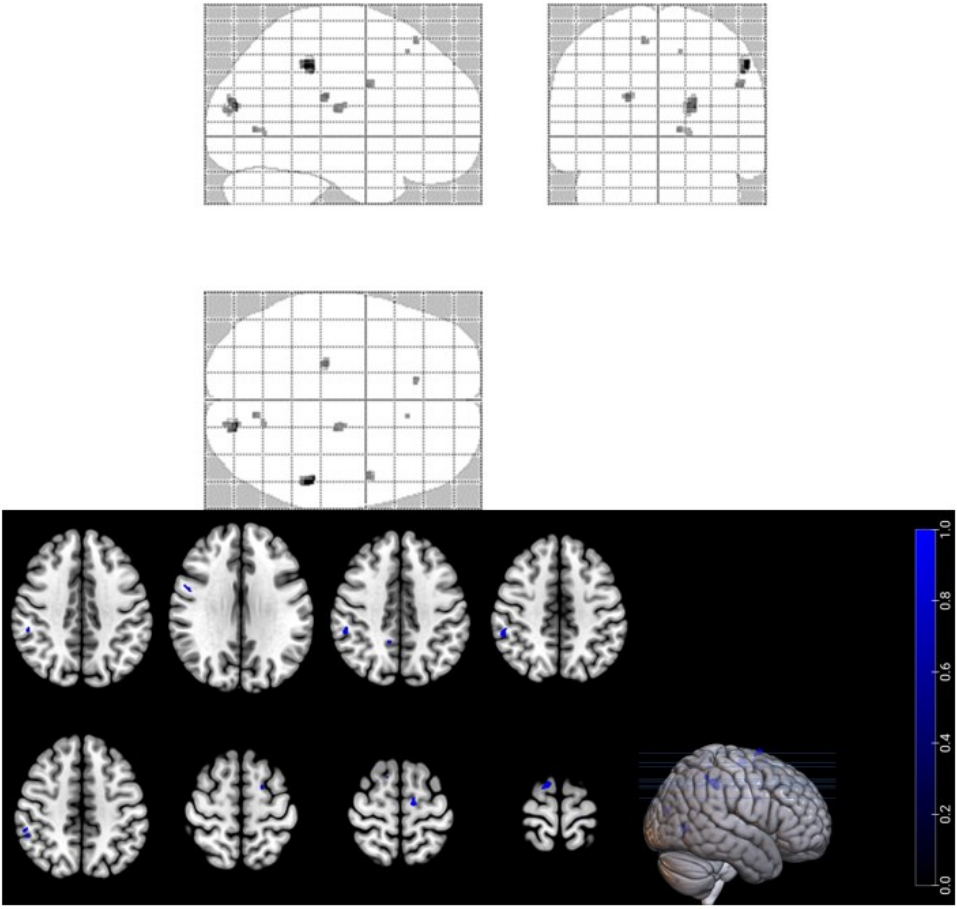
Moreover, in a subgroup of 10 patients (3 NSAbs-POS-PSD and 7 NSAbs-NEG-PSD) structural MRI scan were acquired. Also in this case, differences were observed between the 2 groups. Specifically, NSAbs-NEG-PSD showed a grey matter reduction in several areas, most prominently in frontal and temporal areas (see table YY and Figure YX). All the clusters have a p value <0.001 uncorrected.

Table 14: clusters that resulted as significantly different between the 2 groups.

T value	Cluster size	xyz	Areas
16.56	59	-12 -12 66	Left Paracentral lobule Left Precentral gyrus Left Supplementary motor area
13.15	98	54 -39 45	Right SupraMarginal gyrus Right Inferior parietal gyrus
8.76	101	16 -64 2	Right Lingual gyrus Right Calcarine fissure and surrounding cortex
7.72	11	-50 -56 15	Left Middle temporal gyrus
7.65	20	-34 -78 21	Left Middle occipital gyrus

7.21	27	20 -86 21	Right Superior occipital gyrus Right Calcarine fissure and surrounding cortex Right Cuneus
7.10	61	10 4 75	Right Superior frontal gyrus-dorsolateral Right Supplementary motor area
6.49	10	2 -48 50	Left Precuneus Right Precuneus
6.21	15	-21 2 60	Left Superior frontal gyrus-dorsolateral
5.90	14	52 4 30	Right Precentral gyrus

Figure 17: areas showing a reduction in the NSAbs-NEG group, compared to the NSAbs-POS



Although preliminary, these results point out at the possibility to distinguish early in the course of the disorder different subgroups of FEP patients. If confirmed, these results might give the possibility to define clinical subgroups and therefore apply to them a more tailored management, in terms of care and treatment.

4.5 Discussion

The PHLAMES I study aims to evaluate the prognostic significance of serum NSAbs assay for the diagnosis of AE in subjects with acute/subacute onset psychotic disorder. In fact, this test could be a rapid and minimally invasive screening test for the diagnosis of AE in these subjects. The PHLAMES I study is highly innovative because a rigorous diagnostic algorithm (Fig 3), based on "red flags," is used to screen for AE.

Preliminary results of the project, given the absence of subjects diagnosed with AEs and net of the small sample size examined, show elements of interest, particularly through the stratification of patients with first psychotic episodes according to circulating

autoantibodies and the examination of differences between the subgroups.

First, within our sample there are no subjects diagnosed with AEs based on Graus criteria (Graus et al., 2016). AEs represent a diagnostic entity that, although considered more present than in the past, remains rarely encountered with an incidence of 0.8/100,000/year (Dubey et al., 2018). It is consequently understandable that it is difficult to make an initial diagnosis of AE within a psychiatric setting, as already been highlighted by previous studies (Kelleher et al., 2020; Theorell et al., 2021).

However, it is relevant to highlight that the proportion of NSAbs-POS-PSD patients compared to NSAbs-NEG-PSD patients within our sample was 12.8 %. This result is partially in agreement with literature data, that report widely varying rates of serum anti-neuronal antibody positivity in psychotic patients, ranging from 2% (Bien et al., 2021) to almost 12% (Schou et al., 2016). However, it is important to note that studies assessing FEP patients (Bien et al., 2021; Kelleher et al., 2020; Theorell et al, 2021), comparable with our study, evaluated only the presence of anti-NMDAR, whereas in our study more types of NSAbs are evaluated at once. This likely justifies the prevalence found that is slightly higher, but comparable to

studies with patients with longer disease history that evaluated different NSAbs (Dahm et al., 2014; Steiner et al., 2013).

Another relevant finding is the lack of clinical utility of "red flags" in defining the NSAbs-POS-PSD subgroup within our sample. These clinical-anamnestic elements, which are still under development, were proposed with the aim of guiding the diagnostic process of subjects with autoimmune psychosis, i.e., subjects with AEs characterized by prevalence of psychotic symptoms.

Although the "Red flags" did not show real utility in defining the subgroup with NSAbs, some clinical variables emerged that seem to differentiate the two groups (Fig. XXX).

First, the finding of the higher prevalence of neurological symptoms in the NSAbs-POS-PSD group than in the NSAbs-NEG-PSD group. This finding emphasizes the usefulness of a thorough neurological history and neurological physical examination in patients accessing psychiatric services, particularly when presenting with recent-onset psychotic symptomatology. In our sample, NSAbs-POS-PSD patients present with a more polymorphous, neurological

symptomatology, as is also the case in individuals with AE (Uy et al, 2021).

On the other hand, neurological manifestations, particularly at onset, are an element of interest in psychiatric disorders, although they are often overlooked in psychiatric clinical practice. Ever since the development of the term neurological soft signs (NSS) (Quitkin et al., 1976; Tucker et al., 1975), there has been evidence about the importance of accurate neurological examination in the psychiatric setting. NSS include mild neurological alterations (deficits in sensory integration, motor coordination, and complex motor act sequence), which are more common in patients with schizophrenia than in healthy individuals and have been regularly observed in first-episode patients not yet treated with antipsychotic drugs, supporting the hypothesis that they are an intrinsic component of psychosis (Petrescu et al., 2023). However, studies demonstrated that NSS are not a static feature of schizophrenia, but vary over the clinical course of the disorder, and it has therefore been hypothesized that NSS could therefore be used to monitor disease progression, as well as to identify individuals at risk of developing schizophrenia (Bachmann et al., 2014). These symptoms require careful clinical

examination to detect and, as a result, are likely to be underestimated in clinical practice.

In addition to neurological symptoms, individuals with AE usually present with highly variable and polymorphous neuropsychiatric symptoms (Uy et al., 2021), and, within our sample, NSAbs-POS-PSD individuals also seem to exhibit a wider range of neuropsychiatric features. First, the finding of a statistically significant higher use of antidepressants in NSAbs-POS-PSD patients highlights that other symptomatologic components are clinically relevant to the point of influencing treatment choices. Moreover, the finding of PANSS questions exhibiting statistical significance also supports this assertion. The question for "somatic implication" presents a higher mean score in the group of NSAbs-POS-PSD patients, while the question for "absence of judgment/insight" presents a lower mean score in that population. Both elements allow us to hypothesize that symptomatology is experienced by NSAbs-POS-PSD patients in a more conscious manner, considering the question related to "absence of judgment/insight," and more egodystonic and polymorphic, in which psychological/psychiatric distress is also accompanied by

somatic symptomatology, as in the question related to "somatic implication."

Patients with FEP have global cognitive deficits, and analysis of specific cognitive domains reveals a similar pattern of impairment to that found in chronic forms. These deficits are observed even in the absence of antipsychotic therapy and before the disease becomes chronic, so they are neither a byproduct of antipsychotics nor the result of degeneration as the disease progresses (Sheffield et al., 2018). Verbal memory and processing speed were found to be the two most severely impaired domains in patients with FEP, consistent with what has been described in CHR patients and chronic psychosis (Fatouros-Bergman et al., 2014; Mesholam-Gately et al., 2009). Within our sample, there was a lower mean score in the NSAbs-POS-PSD group than in the "Raven's progressive matrices" group, indicative of greater impairment, and a trend of comparing the mean score of the RAVLT (with lower mean score in the NSAbs-POS-PSD group, a lower score reveals greater impairment) and the TMT-A (with higher mean score in the NSAbs-POS-PSD group; a higher score means greater impairment) (Fig. XXX). Thus, it could be hypothesized that patients in the NSAbs-POS-PSD group have greater cognitive

impairment than the NSAbs-NEG-PSD group, especially in verbal memory (assessed by the RAVLT) and processing speed (assessed by the TMTA. These findings are in agreement with a study of CHR subjects, in which seropositivity to NMDA receptor antibodies was associated with greater global cognitive impairment (lower IQ) and verbal memory impairment (lower RAVLT score) (Pollak et al., 2021). It is important to point out that such cognitive deficits then have important clinical fallout, such as worse functioning, less insight, and increased aggressive behaviors (Gebreegziabhere et al., 2022; Sheffield et al., 2018). Sub-typing patients with FEP based on serum anti-neuronal antibody profile could allow to identify a population with greater cognitive impairment at onset on which to apply cognitive remediation techniques, which are increasingly gaining relevance in clinical practice (Bellani et al., 2019).

Finally, NSAbs-NEG-PSD patients showed reduction in GMV relative to the NSAbs-POS-PSD group. Although counterintuitive, especially in the light of others results, it could be hypothesized that the inflammatory process may represent a defense mechanism, at least initially. Analyzing the results, we find an almost generalized reduction of GMV, with a focus

on temporal areas, which have already been associated with use to discriminate FEPs from healthy patients (Pigoni et al., 2021), and on frontal areas, whose correlation with schizophrenic patients is already described (Alnæs et al., 2019). Although it is well known that schizophrenia is associated with diffuse brain abnormalities, with the greatest evidence about enlarged ventricles, reduced thickness and area at the level of the frontotemporal cortex, and reduced volumes of the hippocampus and amygdala, it must be emphasized that the identification of reproducible subgroups of patients with psychosis on the basis of neuroanatomical, and potentially neuroinflammatory, profiling could help to understand the mechanisms of the disease process and to identify subgroups of candidates for new treatments (Alexandros Lalousis et al. , 2023; Banaj et al., 2023; van Erp et al., 2018).

4.6 Limits

Our data should be considered in light of some limitations. First, our observations need to be replicated in a larger sample to allow for more reliable results, with the difficulties associated with recruiting subjects in the acute phase of psychotic

symptomatology. Follow-up of these patients is also necessary to go on to assess the disease trajectory, an analysis planned in a second part of the PHLAMES project. Although our study is very inclusive in terms of the NSAbs sought, it must be pointed out that there may be other autoantibodies present that are not yet known, even considering that their number is increasing in recent years due to the relevance assumed by this topic. In addition, as described in the literature, the contribution of NSAbs to psychotic manifestations is unclear, considering that they have also been found in subjects without neuropsychiatric disease (Steiner et al., 2020), and subjects with psychiatric pathology may present with both CSF and serum NSAbs, exclusively CSF and non-serum NSAbs, and exclusively serum and non-serum NSAbs (Blackman et al., 2022).

5. Discussion: where are we and where are we going

Early intervention in emerging mental illnesses is essential to improve outcomes. Clinical heterogeneity is substantial, and underlying brain structural, biological, functional heterogeneity is also substantial (Addigton et al., 2020; Brugger et al., 2017). This extreme variability in the presentation of symptom suggests that people diagnosed with FEP may encompass distinct sub-populations of patients, with potentially different trajectories of disease and treatment needs (Dwyer et al., 2022). Research suggests this heterogeneity may only partially overlap with conventional categories based on symptoms and potentially be based on common biological substrates of mental illness that hinder translation (Romer et al., 2021). Thus, dissecting intersecting multimodal heterogeneity across early illness categories is critical to find clinically meaningful biological subsets of patients with similar characteristics, that might inform preventive approaches, facilitating stratification tools for personalized treatment development and care management (Keshavan et al., 2021).

Recent techniques could help in this difficult task: brain imaging, metabolic data, "omics" data, cutting-edge machine learning approaches might be of great benefit to psychiatric research, although intrinsic limitation should be acknowledged (Boudreaux et al., 2021). In the studies presented here, we aimed at dissecting the clinical and biological heterogeneity of psychosis at the beginning of its disease course, showing that it is possible to define meaningful groups of patients and therefore tailor personalized management.

First, we showed how specific brain imaging features (such as insula volume) could predict disease course progression in a longitudinal MRI study. Similarly, we showed how brain metabolites correlate with temperament traits in FEP patients and HC. We found an interplay between brain metabolites and temperament traits, especially BDHI subscales, that are modulated by covariates such as sex and substance abuse.

Second, we presented data from large datasets coming from consortia that spanned around the globe. We showed how ML techniques can deal with this huge amount of data, presenting results from both supervised and unsupervised algorithms. In Pigoni et al., 2021, we showed a supervised ML study presenting meaningful differences between FEP and HC

stemming from MRI data. Moreover, in Dwyer et al. 2023 we presented how FEP can be clustered in different subgroups by means of brain imaging data. It is important to note that the last study suggests that some features can be already present among HC, reinforcing the idea that we should move towards the use of normative model approaches (Segal et al., 2023).

Regarding ML, an open discussion is underway in the field, regarding the interpretability (e.g., "black boxes") of these algorithms, and this aspect might prove crucial in the further development of AI techniques in medicine and psychiatry. This is true especially in the field of mental health, where AI tools should assist clinicians and not introduce further complexity. For an AI to become useful in clinical practice, it should prove to be trustworthy, therefore not only valid and reliable, but also easily understandable (Joyce et al., 2023). In the last years, the concept of explainable AI ("XAI") emerged, as a possibility to close the gap between the algorithms and the clinicians, creating a human-understandable correspondence between inputs and outputs of the black-box model either through intrinsic transparency of the model or through post-hoc techniques. Given that clinical applications are high stakes, we require understandability from the prediction tools or either AI

tools will grow in distrust (Joyce et al., 2023). In the studies we presented, we applied these ideas and always gave interpretable results, in terms of features importance, back-projected in the native space, to inform clinicians.

The idea of definition of clinically-relevant subgroups is intriguing and dates back to the clinical observation of the extreme variability in the presentation of symptom in FEP. This heterogeneity may likely encompass distinct sub-populations of patients, with potentially different illness courses and, hence, different treatment needs. To date only few studies have explored the trajectories of different subgroups (Amoretti et al., 2021) and clustering algorithms emerge as a great opportunity to stratify risk and define single-patient precision treatment.

To achieve such result, however, great efforts are needed. It is likely that single-center studies with limited sample sizes could not be enough to define subtle subpopulations of patients. Bi consortia must act to overcome these limitations. Multicenter studies such as PRONIA and PHENOM, gathering data form different countries in a standardized manner, are highly needed in psychiatry, in order to define clinically-relevant subpopulations of patients. Moreover, future studies should include multimodal data, from neuroimaging, clinical, and

immunobiological, and environmental factors, in order to give the most complete description of patients. Regarding FEP, nowadays most clustering studies focused on cognitive measures (Amoretti et al., 2021). Multimodal studies have the potential to shed light on homogeneous phenotypic targets to understand the subtle neuropathologic differences among individuals with FEP, since individuals within the same cluster may share cognitive, genetic, and neurophysiologic features.

Such studies should also follow-up patients, in order to evaluate their trajectories and differential response to treatments. Only in this way we could achieve personalized psychiatry. Some of these cooperative efforts are already at work (e.g., the studies stemming on depression from the UK biobank, one of the most important source databases available at time) and others need to begin, since complex disorders need a cooperative effort. Moreover, these type of studies should also be expanded to high-risk populations, in order to prevent and deal the most effective treatment (Ramain et al., 2022).

Finally, we presented the preliminary results of an ongoing study fully funded by the Italian Ministry of Health, leaded by A. Pigoni as co-PI. In this study, we prospectively assessed in a diagnostic algorithm the overlap between autoimmune

encephalitis and presentation of FEP. Moreover, we defined a subgroup of almost 13% of patients that presents circulating autoantibodies against CNS structures, that presents clinically significant differences compared to the "standard" group. Thus, subtyping patients with FEP with regard to the presence of serum autoantibodies could be useful in both clinical and research terms. Indeed, if it is widely described in the literature that autoimmunity within psychiatric disorders show discordant results (Wittenberg et al., 2020), it is equally evident that the presence of a subpopulation of patients in which the immune component plays a primary role in the etiology and progression of the neuropsychiatric disorder cannot be excluded (Enrico et al., 2021). The definition of a subgroup of patients based on a rapid, non-invasive, test could be of paramount importance.

The presence of different subgroups it is likely on the base of recent failures in clinical trials with immunomodulant molecules in depressions (Hellmann-Regen et al., 2022; Kavakbasi et al., 2023). Although it is widely recognized that depression has an inflammatory aspect (Harsanyi et al., 2022) and this is considered to be even clearer than in psychotic disorders, different recent trials (e.g., the PREDDICT trial) with

immunomodulatory drugs failed to prove their effect in depressive symptoms. The most likely reason is that not all depressions show an important inflammatory base and diagnostic subgroups are likely present: some would benefit from immunoinflammatory modulators, some would not. Future advancements in psychiatry should be able to stratify patients that will participate to clinical trials, as is already happening in neurology, where patients are stratified before entering clinical trials randomization on the base of neurobiological features (e.g., for Alzheimer's disorder, patients could be stratified based on their PET or MRI).

Research if fundamental in the definition of such subgroups that would benefit from different treatments. PHLAMES could serve as the basis for defining different subgroups of patients to test immunomodulatory treatment. As a future step, clinical trials with immunomodulant drugs could be designed to include only patients with verified immunological alterations, in order not to jeopardize results as happened with the present trials in depression. Having a non-invasive method to stratify patients based on their immunological status would be of great impact. For instance, a future direction could be a clinical trial

with an immunomodulant on the PHLAMES patients, divided on the base of their serum immunological biomarkers.

Such improvements will dramatically impact the patient's life and the psychiatrist's management. The possibility to have biologically based subgroups might represent the first step towards a real, diffuse, precision psychiatry that is able to translate research into clinical practice.

Among the challenges for the future, however, there is the long-term impact of these classification and subgroups. Future studies are needed to define how the trajectories of these subgroups diverge from each other and how different treatments might influence such trajectories.

In conclusion, these data foster the research for subtypes of FEP and the definition of disease trajectories. These advances might have a great impact on patients' lives, by defining specific subgroups or progression that benefit of tailored interventions.

6. References

1. Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. Nat Med. 2016;22:1248–55
2. Addington J, Farris M, Devoe D, Metzak P. Progression from being at-risk to psychosis: next steps. NPJ Schizophr. 2020;6(1):27
3. Al-Diwani A, Pollak TA, Langford AE, Lennox BR. Synaptic and neuronal autoantibody-associated psychiatric syndromes: controversies and hypotheses. Front Psychiatry 2017; 8: 13
4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Publishing. 2013
5. Amoretti S, Verdolini N, Mezquida G, Rabelo-da-Ponte FD, Cuesta MJ, Pina-Camacho L, Gomez-Ramiro M, De-la-Cámara C, González-Pinto A, Díaz-Caneja CM, Corripio I, Vieta E, de la Serna E, Mané A, Solé B, Carvalho AF, Serra M, Bernardo M. Identifying clinical clusters with distinct trajectories in first-episode psychosis through an unsupervised

machine learning technique. Eur Neuropsychopharmacol. 2021 Jun;47:112-129.

6. Amoretti S, Rabelo-da-Ponte FD, Rosa AR, Mezquida G, Sánchez-Torres AM, Fraguas D, Cabrera B, Lobo A, González-Pinto A, Pina-Camacho L, Corripio I, Vieta E, Torrent C, de la Serna E, Bergé D, Bioque M, Garriga M, Serra M, Cuesta MJ, Bernardo M; PEPs Group. Cognitive clusters in first-episode psychosis. Schizophr Res. 2021 Nov;237:31-39.
7. Andreasen NC (1989) The scale for the assessment of negative symptoms (SANS). Br J Psychiatry 155:49–52
8. Arciniegas DB. Psychosis. Continuum (Minneap Minn). 2015 Jun; 21 (3 Behavioral Neurology and Neuropsychiatry): 715–36.
9. Bora E. Peripheral inflammatory and neurotrophic biomarkers of cognitive impairment in schizophrenia: a meta-analysis. Psychol Med. 2019 Sep;49(12):1971-1979.
10. Borgan FR, Jauhar S, McCutcheon RA, Pepper FS, Rogdaki M, Lythgoe DJ, Howes OD. Glutamate levels in the anterior cingulate cortex in un-

medicated first episode psychosis: a proton magnetic resonance spectroscopy study. Sci Rep. 2019 Jul 2;9(1):8685.

11. Boscutti A, Pigoni A, Delvecchio G, Lazzaretti M, Mandolini GM, Girardi P, Ferro A, Sala M, Abbiati V, Cappucciati M, Bellani M, Perlini C, Rossetti MG, Balestrieri M, Damante G, Bonivento C, Rossi R, Finos L, Serretti A, Brambilla P, The Gecobip Group. The Influence of 5-HTTLPR, BDNF Rs6265 and COMT Rs4680 Polymorphisms on Impulsivity in Bipolar Disorder: The Role of Gender. Genes (Basel). 2022 Mar 9;13(3):482. doi: 10.3390/genes13030482.
12. Boudreaux ED, et al. Applying Machine Learning Approaches to Suicide Prediction Using Healthcare Data: Overview and Future Directions. Front Psychiatry. 2021;12. doi:10.3389/FPSYT.2021.707916
13. Brambilla, P, Fagnani, C, Cecchetto, F, Medda, E, Bellani, M, Salemi, M, et al. Genetic and environmental bases of the interplay between magical ideation and personality. Psychiatry Res (2014) doi: 10.1016/j.psychres.2013.11.021
14. Brugger SP, Howes OD. Heterogeneity and homogeneity of regional brain structure in

schizophrenia: a meta-analysis. JAMA Psychiatry. 2017;74(11):1104-1111.

15. Burock G, Pihur V, Datta S, Datta S (2008) clValid: an R package for cluster validation. J Stat Softw. https://doi.org/10.18637/jss.v025.i04
16. C. Gaser, "Manual Computational Anatomy Toolbox- cat12," Struct. Brain Mapp. Gr. Dep. Psychiatry Neurol. Univ. Jena, pp. 1–53, 2017, [Online]. Available: hiip://www.neuro.uni-jena.de/cat/index.html.
17. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, Doshi J, Pomponio R, Pigoni A, Dazzan P, Kahn RS, Schnack HG, Zanetti MV, Meisenzahl E, Busatto GF, Crespo-Facorro B, Pantelis C, Wood SJ, Zhuo C, Shinohara RT, Shou H, Fan Y, Gur RC, Gur RE, Satterthwaite TD, Koutsouleris N, Wolf DH, Davatzikos C. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain. 2020;143:1027–38.
18. Chand GB, Singhal P, Dwyer DB, Wen J, Erus G, Doshi J, et al. Schizophrenia Imaging Signatures and Their Associations With Cognition,

Psychopathology, and Genetics in the General Population. Am J Psychiatry. 2022:179:650–60.

19. Charleen D Adams, Susan L Neuhausen. Bi-directional Mendelian randomization of epithelial ovar-ian cancer and schizophrenia and uni-directional Mendelian randomization of schizophrenia on circulating 1- or 2-glycerophosphocholine metabolites. Mol Genet Metab Rep. 2019 Nov 6;21:100539.
20. Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61(1):1–36. https://doi.org/10.18637/jss.
21. Chong, H. Y. et al. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr. Dis. Treat. 12, 357–73 (2016)
22. Colizzi, M., McGuire, P., Pertwee, R. G. & Bhattacharyya, S. Efect of cannabis on glutamate signal-ling in the brain: A systematic review of human and animal evidence. Neurosci. Biobehav. Rev. 64, 359–381 (2016).
23. Dahm L, Ott C, Steiner J, Stepniak B, Teegen B, Saschenbrecker S, Hammer C, Borowski K,

Begemann M, Lemke S, Rentzsch K, Probst C, Martens H, Wienands J, Spalletta G, Weissenborn K, Stöcker W, Ehrenreich H. Seroprevalence of autoantibodies against brain antigens in health and disease. Ann Neurol. 2014 Jul;76(1):82-94.

24. Dalmau J, Graus F. Antibody-mediated encephalitis. N Engl J Med 2018; 378: 840–51.
25. Davison J, O'Gorman A, Brennan L, Cotter DR. A systematic review of metabolite biomarkers of schizophrenia. Schizophr Res. 2018 May;195:32-50.
26. Dempster K, Norman R, Théberge J, Densmore M, Schaefer B, Williamson P (2015). Glutamatergic metabolite correlations with neuropsychological tests in first episode schizophrenia. Psychiatry Research 233, 180–185.
27. Dunleavy C, Elsworthy RJ, Upthegrove R, Wood SJ, Aldred S. Inflammation in first-episode psychosis: The contribution of inflammatory biomarkers to the emergence of negative symptoms, a systematic review and meta-analysis. Acta Psychiatr Scand. 2022 Jul;146(1):6-20.
28. Dwyer DB, Buciuman MO, Ruef A, Kambeitz J, Sen Dong M, Stinson C, Kambeitz-Ilankovic L,

Degenhardt F, Sanfelici R, Antonucci LA, Lalousis PA, Wenzel J, Urquijo-Castro MF, Popovic D, Oeztuerk OF, Haas SS, Weiske J, Hauke D, Neufang S, Schmidt-Kraepelin C, Ruhrmann S, Penzel N, Lichtenstein T, Rosen M, Chisholm K, Riecher-Rössler A, Egloff L, Schmidt A, Andreou C, Hietala J, Schirmer T, Romer G, Michel C, Rössler W, Maj C, Borisov O, Krawitz PM, Falkai P, Pantelis C, Lencer R, Bertolino A, Borgwardt S, Noethen M, Brambilla P, Schultze-Lutter F, Meisenzahl E, Wood SJ, Davatzikos C, Upthegrove R, Salokangas RKR, Koutsouleris N; PRONIA Consortium. Clinical, Brain, and Multilevel Clustering in Early Psychosis and Affective Stages. JAMA Psychiatry. 2022 Jul 1;79(7):677-689.

29. Dwyer DB, Cabral C, Kambeitz-Ilankovic L, Sanfelici R, Kambeitz J, Calhoun V, et al. Brain Subtyping Enhances The Neuroanatomical Discrimination of Schizophrenia. Schizophr Bull. 2018;44:1060–9.
30. Dwyer DB, Chand GB, Pigoni A, Khuntia A, Wen J, Antoniades M, Hwang G, Erus G, Doshi J, Srinivasan D, Varol E, Kahn RS, Schnack HG,

Meisenzahl E, Wood SJ, Zhuo C, Sotiras A, Shinohara RT, Shou H, Fan Y, Schaulfelberger M, Rosa P, Lalousis PA, Upthegrove R, Kaczkurkin AN, Moore TM, Nelson B, Gur RE, Gur RC, Ritchie MD, Satterthwaite TD, Murray RM, Di Forti M, Ciufolini S, Zanetti MV, Wolf DH, Pantelis C, Crespo-Facorro B, Busatto GF, Davatzikos C, Koutsouleris N, Dazzan P. Psychosis brain subtypes validated in first-episode cohorts and related to illness remission: results from the PHENOM consortium. Mol Psychiatry. 2023 May;28(5):2008-2017. doi: 10.1038/s41380-023-02069-0.

31. Erin C McGlade 1 2, Doug Hyun Han 3, Sun Mi Kim 3, Xianfeng Shi 1, Kirsten Cline 1 2, Deborah Yurgelun-Todd 1 2, Perry F Renshaw. Proton magnetic resonance spectroscopy (MRS) in individuals with internet gaming. Front Psychiatry. 2022 Dec 22;13:1031947.
32. Ezeoke A, Mellor A, Buckley P, Miller B. A systematic, quantitative review of blood autoantibodies in schizophrenia. Schizophr Res. 2013;150:245–251

33. Fossati, A.; Di Ceglie, A.; Acquarini, E.; Barratt, E.S. Psychometric properties of an Italian version of the Barratt Impulsiveness Scale-11 (BIS-11) in nonclinical subjects. J. Clin. Psychol. 2001, 57, 815–828
34. Gaebel W, Zielasek J. Focus on psychosis. Dialogues Clin Neurosci. 2015 Mar; 17 (1): 9-18.
35. Gillinder L, Warren N, Hartel G, Dionisio S, O'Gorman C. EEG findings in NMDA encephalitis—a systematic review. Seizure 2018; 65: 20–24
36. Goldberg, T. E. et al. Cognitive Improvement Afer Treatment With Second-Generation Antipsychot-ic Medications in First-Episode Schizo.pdf. Arch. Gen. Psychiatry 64, 1115–1122 (2007). A review of altered biochemistry in the anterior cingulate cortex of first-episode psychosis.
37. Goodwin RD, Fergusson DM, Horwood LJ (2004): Early anxious/withdrawn behaviours predict later internalising disorders. J Child Psychol Psychiatry Allied Discip. https://doi.org/10.1111/j.1469-7610.2004.00279.x

38. Graus F, Titulaer MJ, Balu R, Benseler S, Bien CG, Cellucci T, Cortese I, Dale RC, Gelfand JM, Geschwind M, Glaser CA, Honnorat J, Höftberger R, Iizuka T, Irani SR, Lancaster E, Leypoldt F, Prüss H, Rae-Grant A, Reindl M, Rosenfeld MR, Rostásy K, Saiz A, Venkatesan A, Vincent A, Wandinger KP, Waters P, Dalmau J. A clinical approach to diagnosis of autoimmune encephalitis. Lancet Neurol. 2016 Apr;15(4):391-404.
39. Haijma SV, Van Haren N, Cahn W, Koolschijn PCM, Hulshoff Pol HE, Kahn RS. Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. Schizophr Bull. 2013;39:1129–38
40. Hamideh Bakhshinezhad, Mohammad Darharaj, Yasha Fayaz Feyzi, Shirzad Babaei, Reza Ahadi, Behnammaddin Jamei, Mahsa Pourhamzeh, Abdolhadi Daneshi, Reza Arezoomandan. The Relationship Between Brain Metabolites Alterations and Neuropsychological Deficits in Patients with Methamphetamine Use Disorder: A Proton Magnetic Resonance Spectroscopy Study. Arch Clin Neuropsychol. 2022 Jan 17;37(1):160-172.

41. Harsanyi S, Kupcova I, Danisovic L, Klein M. Selected Biomarkers of Depression: What Are the Effects of Cytokines and Inflammation? Int J Mol Sci. 2022 Dec 29;24(1):578.
42. Hellmann-Regen J, Clemens V, Grözinger M, Kornhuber J, Reif A, Prvulovic D, Goya-Maldonado R, Wiltfang J, Gruber O, Schüle C, Padberg F, Ising M, Uhr M, Friede T, Huber C, Manook A, Baghai TC, Rupprecht R, Heuser I. Effect of Minocycline on Depressive Symptoms in Patients With Treatment-Resistant Depression: A Randomized Clinical Trial. JAMA Netw Open. 2022 Sep 1;5(9):e2230367.
43. Heine J, Prüss H, Bartsch T, Ploner CJ, Paul F, Finke C. Imaging of autoimmune encephalitis-relevance for clinical practice and hippocampal function. Neuroscience 2015; 309: 68–83
44. Herken J, Prüss H. Red flags: clinical signs for identifying autoimmune encephalitis in psychiatric patients. Front Psychiatry 2017; 8: 25
45. Hofstra MB, Van Der Ende J, Verhulst FC (2002): Child and Adolescent Problems Predict DSM-IV Disorders in Adulthood: A 14-Year Follow-up of a Dutch Epidemiological Sample. J Am Acad Child

Adolesc Psychiatry. https://doi.org/10.1097/00004583-200202000-00012

46. Holroyd, CB, and Yeung, N. Motivation of extended behaviors by anterior cingulate cortex. Trends Cognit Sci (2012) 16(2):122–28.
47. Honnorat N, Dong A, Meisenzahl-Lechner E, Koutsouleris N, Davatzikos C. Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. Schizophr Res. 2019;214:43–50.
48. Lally, O. Ajnakina, B. Stubbs, et al. Remission and recovery from first-episode psychosis in adults: systematic review and meta- analysis of long-term outcome studies. Br J Psychiatry, 211 (2017), pp. 350-358
49. Jian-Neng Li 1, Xi-Long Liu 2, Long Li. Prefrontal GABA and glutamate levels correlate with impul-sivity and cognitive function of prescription opioid addicts: A 1 H-magnetic resonance spectroscopy study. Psychiatry Clin Neurosci. 2020 Jan;74(1):77-83.

50. Joyce DW, Kormilitzin A, Smith KA, Cipriani A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. npj Digit Med 2023; 6: 1–7.
51. Kaczkurkin AN, Sotiras A, Baller EB, Barzilay R, Calkins ME, Chand GB, et al. Neurostructural Heterogeneity in Youths With Internalizing Symptoms. Biol Psychiatry. 2020;87:473–82.
52. Kay SR, Fiszbeln A, Opler LA (1987) The positive and negative syndrome scale (panss) manual. Schizophrenia Bull 13(2):261
53. Kavakbasi E, Sampson E, Mills NT, Hori H, Schwarte K, Hohoff C, Schubert KO, Clark SR, Fourrier C, Baune BT. Inflammation-stratified augmentation of vortioxetine with celecoxib: Results from a double-blind, randomized, placebo-controlled trial in major depressive disorder. J Neurochem. 2023 Aug 27. doi: 10.1111/jnc.15946. Online ahead of print.
54. Kelley BP, Patel SC, Marin HL, Corrigan JJ, Mitsias PD, Griffith B. Autoimmune encephalitis: pathophysiology and imaging review of an

overlooked diagnosis. AJNR Am J Neuroradiol 2017; 38: 1070–78.

55. Keshavan MS. Characterizing transdiagnostic premorbid biotypes can help progress in selective prevention in psychiatry. World Psychiatry. 2021;20(2):231-232.
56. Lee R, Leighton SP, Thomas L, Gkoutos GV, Wood SJ, Fenton SH, Deligianni F, Cavanagh J, Mallikarjun PK. Prediction models in first-episode psychosis: systematic review and critical appraisal. Br J Psychiatry. 2022 Jan 24;220(Spec Iss 4 Themed Iss Precision Medicine and Personalised Healthcare in Psychiatry):1-13. doi: 10.1192/bjp.2021.219
57. Lord E, Willems M, Lapointe FJ, Makarenkov V (2017) Using the stability of objects to determine the number of clusters in datasets. Inf Sci 393:29–46
58. Lubeiro A, Rueda C, Hernández JA, Sanz J, Sarramea F, Molina V. Identification of two clusters within schizophrenia with different structural, functional and clinical characteristics. Prog Neuropsychopharmacol Biol Psychiatry. 2016;64:79–86.

59. M. Laruelle, L. S. Kegeles, and A. Abi-Dargham, "Glutamate, dopamine, and schizophrenia," Ann NY Acad Sci, vol. 1003, pp. 138-158, 2003.
60. Matcheri S Keshavan 1, Rachel M Dick, Vaibhav A Diwadkar, Debra M Montrose, Konasale M Pra-sad, Jeffrey A Stanley. Striatal metabolic alterations in non-psychotic adolescent offspring at risk for schizophrenia: a (1)H spectroscopy study. Schizophr Res 2009 Nov;115(1):88-93
61. McCutcheon, R.A., Marques, T.R., Howes, O.D., 2020. Schizophrenia—an overview. JAMA psychiatry 77, 201–210
62. Moreno-Küstner, B., Martín, C. & Pastor, L. Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. PLoS One 13, e0195687 (2018).
63. Musazzi, L., Treccani, G., Mallei, A. & Popoli, M. Te Action of Antidepressants on the Glutamate System: Regulation of Glutamate Release and Glutamate Receptors. Biol. Psychiatry 73, 1180–1188 (2013).
64. Oeztuerk OF, Pigoni A, Wenzel J, Haas SS, Popovic D, Ruef A, Dwyer DB, Kambeitz-Ilankovic L, Ruhrmann S, Chisholm K, Lalousis P, Griffiths SL,

Lichtenstein T, Rosen M, Kambeitz J, Schultze-Lutter F, Liddle P, Upthegrove R, Salokangas RKR, Pantelis C, Meisenzahl E, Wood SJ, Brambilla P, Borgwardt S, Falkai P, Antonucci LA, Koutsouleris N; PRONIA Consortium. The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study. Eur Arch Psychiatry Clin Neurosci. 2022 Apr;272(3):403-413.

65. Oeztuerk OF, Pigoni A, Antonucci LA, Koutsouleris N. Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half-century studies. Eur Arch Psychiatry Clin Neurosci. 2022 Apr;272(3):381-393. doi: 10.1007/s00406-021-01295-3.
66. Patton, J.H.; Stanford, M.S.; Barratt, E.S. Factor structure of the Barratt impulsiveness scale. J. Clin. Psychol. 1995, 51, 768–774.
67. Pigoni A, Delvecchio G, Squarcina L, Bonivento C, Girardi P, Finos L, Crisanti C, Balestrieri M, D'A-gostini S, Stanley JA, Brambilla P. Sex differences in brain metabolites in anxiety and mood disor-ders.

Psychiatry Res Neuroimaging. 2020 Nov 30;305:111196

68. Pigoni A, Delvecchio G, Dusi N, Schiena G, Andreella A, Finos L, Cecchetto F, Perlini C, Gloria Rossetti M, Ferro A, Bellani M, Lasalvia A, Ruggeri M, Brambilla P; PICOS and First projects. Insula volumes in first-episode and chronic psychosis: A longitudinal MRI study. Schizophr Res. 2022 Mar;241:14-23. doi: 10.1016/j.schres.2021.12.048.
69. Pigoni A, Dwyer D, Squarcina L, Borgwardt S, Crespo-Facorro B, Dazzan P, Smesny S, Spaniel F, Spalletta G, Sanfelici R, Antonucci LA, Reuf A, Oeztuerk OF, Schmidt A, Ciufolini S, Schönborn-Harrisberger F, Langbein K, Gussew A, Reichenbach JR, Zaytseva Y, Piras F, Delvecchio G, Bellani M, Ruggeri M, Lasalvia A, Tordesillas-Gutiérrez D, Ortiz V, Murray RM, Reis-Marques T, Di Forti M, Koutsouleris N, Brambilla P; ClassiFEP group. Classification of first-episode psychosis using cortical thickness: A large multicenter MRI study. Eur Neuropsychopharmacol. 2021 Jun;47:34-47
70. Pollak et al., Antibodies in the Diagnosis, Prognosis, and Prediction of Psychotic Disorders

71. Pollak TA, Beck K, Irani SR, Howes OD, David AS, McGuire PK. Autoantibodies to central nervous system neuronal surface antigens: psychiatric symptoms and psychopharmacological implications. Psychopharmacology (Berl) 2016; 233: 1605–21.
72. Pollak TA, Lennox BR, Müller S, Benros ME, Prüss H, Tebartz van Elst L, Klein H, Steiner J, Frodl T, Bogerts B, Tian L, Groc L, Hasan A, Baune BT, Endres D, Haroon E, Yolken R, Benedetti F, Halaris A, Meyer JH, Stassen H, Leboyer M, Fuchs D, Otto M, Brown DA, Vincent A, Najjar S, Bechter K. Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. Lancet Psychiatry. 2020 Jan;7(1):93-108.
73. Pollak TA, McCormack R, Peakman M, Nicholson TR, David AS. Prevalence of anti-N-methyl-D-aspartate (NMDA) receptor [corrected] antibodies in patients with schizophrenia and related
74. Pollak TA, Rogers JP, Nagele RG, Peakman M, Stone JM, David AS, McGuire P. Antibodies in the

Diagnosis, Prognosis, and Prediction of Psychotic Disorders. Schizophr Bull. 2019 Jan 1;45(1):233-246.

75. Porcelli S, Marsano A, Caletti E, Sala M, Abbiati V, Bellani M, Perlini C, Rossetti MG, Mandolini GM, Pigoni A, Paoli RA, Piccin S, Lazzaretti M, Fabbro D, Damante G, Bonivento C, Ferrari C, Rossi R, Pedrini L, Serretti A, Brambilla P; GECOBIP Group. Temperament and Character Inventory in Bi-polar Disorder versus Healthy Controls and Modulatory Effects of 3 Key Functional Gene Variants. Neuropsychobiology. 2017;76(4):209-221.
76. Punitha Subramaniam 1, Andrew Prescot 2, Erin McGlade 3, Perry Renshaw 3, Deborah Yurgelun-Todd. Examination of gamma-aminobutyric acid and glutamate-glutamine levels in association with impulsive behavior in adolescent marijuana users. Drug Alcohol Depend. 2022 Apr 1;233:109326.
77. Ramain J, Conus P, Golay P. Subtyping based on premorbid profile: A strategy to personalize treatment in first-episode affective psychosis. Early Interv Psychiatry. 2022 Jan;16(1):51-60.
78. Romer AL, Elliott ML, Knodt AR, et al.. Pervasively thinner neocortex as a transdiagnostic feature of

general psychopathology. Am J Psychiatry. 2021;178(2):174-182.

79. S Marques, F Godinho, AL Melo, D. Barrocas First-episode psychosis: What does it mean? European Psychiatry, 33 (2016), p. s258, 10.1016/j.eurpsy.2016.01.656
80. Schmitt SE, Pargeon K, Frechette ES, Hirsch LJ, Dalmau J, Friedman D. Extreme delta brush: a unique EEG pattern in adults with anti-NMDA receptor encephalitis. Neurology 2012; 79: 1094–100.
81. Scott JG, Gillis D, Ryan AE, Hargovan H, Gundarpi N, McKeon G, Hatherill S, Newman MP, Parry P, Prain K, Patterson S, Wong RCW, Wilson RJ, Blum S. The prevalence and treatment outcomes of antineuronal antibody-positive patients admitted with first episode of psychosis. BJPsych Open. 2018 Mar 16;4(2):69-74.
82. Segal A, Parkes L, Aquino K, Kia SM, Wolfers T, Franke B, Hoogman M, Beckmann CF, Westlye LT, Andreassen OA, Zalesky A, Harrison BJ, Davey CG, Soriano-Mas C, Cardoner N, Tiego J, Yücel M, Braganza L, Suo C, Berk M, Cotton S, Bellgrove MA, Marquand AF, Fornito A. Regional, circuit and

network heterogeneity of brain abnormalities in psychiatric disorders. Nat Neurosci. 2023 Sep;26(9):1613-1629

83. Shang P, Ho AM, Tufvesson-Alm M, Lindberg DR, Grant CW, Orhan F, Eren F, Bhat M, Engberg G, Schwieler L, Fatouros-Bergman H, Imbeault S, Iverson RM, Dasari S, Piehl F, Cervenka S, Sellgren CM, Erhardt S, Choi DS. Identification of cerebrospinal fluid and serum metabolomic biomarkers in first episode psychosis patients. Transl Psychiatry. 2022 Jun 3;12(1):229.
84. Squarcina L, Stanley JA, Bellani M, Altamura CA, Brambilla P. Epidemiol Psychiatr Sci. 2017 Apr;26(2):122-128.
85. Stork C, Renshaw P. Mitochondrial dysfunction in bipolar disorder: evidence from magnetic resonance spectroscopy research. Mol Psychiatry. (2005) 10:900–19.
86. Sugihara G, Oishi N, Son S, Kubota M, Takahashi H, Murai T. Distinct Patterns of Cerebral Cortical Thinning in Schizophrenia: A Neuroimaging Data-Driven Approach. Schizophr Bull. 2017;43:900–6

87. Takeuchi H, Siu C, Remington G, Fervaha G, Zipursky RB, Foussias G, Agid O. Does relapse contrib-ute to treatment resistance? Antipsychotic response in first- vs. second-episode schizophrenia. Neuropsychopharmacology. 2019 May;44(6):1036-1042. doi: 10.1038/s41386-018-0278-3. Epub 2018 Nov 22. PMID: 30514883; PMCID: PMC6462044.
88. The RAISE early treatment program for first-episode psychosis: background, rationale, and study design. J.M. Kane, N.R. Schooler, P. Marcy, et al. J Clin Psychiatry, 76 (2015), pp. 240-246
89. Tibshirani R, Walther G (2005) Cluster validation by prediction strength. J Comput Graph Stat 14(3):511–528
90. Titulaer MJ, McCracken L, Gabilondo I, Armangué T, Glaser C, Iizuka T, Honig LS, Benseler SM, Kawachi I, Martinez-Hernandez E, Aguilar E, Gresa-Arribas N, Ryan-Florance N, Torrents A, Saiz A, Rosenfeld MR, Balice-Gordon R, Graus F, Dalmau J. Treatment and prognostic factors for long-term outcome in patients with anti-NMDA receptor encephalitis: an observational cohort study. Lancet Neurol. 2013 Feb;12(2):157-65.

91. Varley J, Taylor J, Irani SR. Autoantibody-mediated diseases of the CNS: structure, dysfunction and therapy. Neuropharmacology 2018; 132: 71–82.
92. Varol E, Sotiras A, Davatzikos C. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage. 2016;145:346–64.
93. Varol E, Sotiras A, Davatzikos C. MIDAS: Regionally linear multivariate discriminative statistical mapping. Neuroimage. 2018;174:111–26
94. Veronika Husarova 1, Michal Bittsansky 2, Igor Ondrejka 3, Dusan Dobrota. Prefrontal grey and white matter neurometabolite changes after atomoxetine and methylphenidate in children with attention deficit/hyperactivity disorder: a (1)H magnetic resonance spectroscopy study. Psychiatry Res. 2014 Apr 30;222(1-2):75-83.
95. W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, "Statistical Parametric Mapping: The Analysis of Functional Brain Images," Stat. Parametr. Mapp. Anal. Funct. Brain Images, 2007, doi: 10.1016/B978-0-12-372560-8.X5000-1.

96. Whitehurst TS, Osugo M, Townsend L, Shatalina E, Vava R, Onwordi EC, Howes O. Proton Magnetic Resonance Spectroscopy of N-acetyl Aspartate in Chronic Schizophrenia, First Episode of Psychosis and High-Risk of Psychosis: A Systematic Review and Meta-Analysis. Neurosci Biobehav Rev. 2020 Dec;119:255-267
97. Xiao Y, Liao W, Long Z, Tao B, Zhao Q, Luo C, et al. Subtyping Schizophrenia Patients Based on Patterns of Structural Brain Alterations. Schizophr Bull. 2022;48:241–50.
98. Yoshimura B, Takaki M. Anti-NMDA receptor antibody positivity and presentations without seizure, involuntary movement, hypoventilation, or tumor: a systematic review of the literature. J Neuropsychiatry Clin Neurosci 2017; 29: 267–74.
99. Z.-X. Xi, D. A. Baker, H. Shen, D. S. Carson, and P. W. Kalivas, "Group II metabotropic glutamate receptors modulate extracellular glutamate in the nucleus accumbens," Journal of Pharmacology and Experimental Therapeutics, vol. 300, no. 1, pp. 162-171, 2002.

100. Zhao Q, Cao H, Zhang W, Li S, Xiao Y, Tamminga CA, et al. A subtype of institutionalized patients with schizophrenia characterized by pronounced subcortical and cognitive deficits. Neuropsychopharmacology. 2022;47:2024–32.