Aus der
Klinik für Psychiatrie und Psychotherapie
Klinikum der Ludwig-Maximilians-Universität München

Phenotypical exploration and predictive utility of formal thought disorder in individuals with recent-onset of psychosis

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von:
Ömer Faruk Öztürk

aus:

Gaziantep / Türkei

Jahr
2024

Mit Genehmigung der Medizinischen Fakultät der

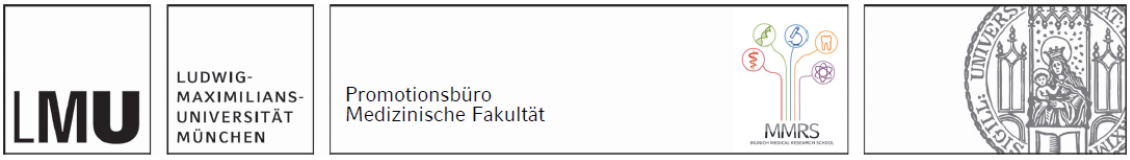
Ludwig-Maximilians-Universität zu München

Erstes Gutachten:	Prof. Dr. Nikolaos Koutsouleris
Zweites Gutachten:	Prof. Dr. Alkomiet Hasan
Drittes Gutachten:	Priv. Doz. Dr. Wolfgang Strube
Viertes Gutachten:	Prof. Dr. Michael Landgrebe

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 04.03.2024

Affidavit



Affidavit

Öztürk, Ömer Faruk

Surname, first name

Street

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Phenotypical exploration and predictive utility of formal thought disorder in individuals with recent onset of psychosis

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Bickenbach, 10.07.2023 — Ömer Faruk Öztürk

place, date — Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Öztürk, Ömer Faruk

Surname, first name

Street

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Phenotypical exploration and predictive utility of formal thought disorder in individuals with recent onset of psychosis

is congruent with the printed version both in content and format.

Bickenbach, 10.07.2023

place, date

Ömer Faruk Öztürk

Signature doctoral candidate

Table of content

List of abbreviations

ARMS: at-risk mental state

CHARMS: Clinical High At Risk Mental State

CHR: The Clinical High-Risk

DALY: the total disability adjusted life years

DSM - V -TR: Diagnostic and Statistical Manual of Mental Disorders Fifth Edition Text Revision

DTI: Diffusion tensor imaging

fALFF: fractional amplitude of low-frequency fluctuations

FThD: Formal thought disorder

GABA: Gamma-Aminobutyric Acid

GF: Global Functioning

GVM: gray matter volume

HiTOP: The Hierarchical Taxonomy of Psychopathology

ICD - 10: International Classification of Diseases 10th Revision

NMDA: N-methyl-D-aspartate

PANSS: the Positive and Negative Syndrome Scale

p_{fdr}: p-value corrected for the positive false discovery rate

PRONIA: Personalised Prognostic Tool for Early Psychosis Management

ROP: recent-onset psychosis

SANS: the Scale for the Assessment of Negative Symptoms

sMRI: structural MRI

SVM: support vector machine

WAIS: The Wechsler Adult Intelligence Scale

WMV: white matter volume

YLD: the years lived with disability

List of publications

1. Antonucci, L. A., Penzel, N., Sanfelici, R., Pigoni, A., Kambeitz-Ilankovic, L., Dwyer, D., Ruef, A., sen Dong, M., **Öztürk, Ö. F**., Chisholm, K., Haidl, T., Rosen, M., Ferro, A., Pergola, G., Andriola, I., Blasi, G., Ruhrmann, S., Schultze-Lutter, F., Falkai, P., … Koutsouleris, N. (2022). Using combined environmental-clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. *British Journal of Psychiatry*, *220*(4), 229–245.
2. Dwyer, D. B., Buciuman, M. O., Ruef, A., Kambeitz, J., sen Dong, M., Stinson, C., Kambeitz-Ilankovic, L., Degenhardt, F., Sanfelici, R., Antonucci, L. A., Lalousis, P. A., Wenzel, J., Urquijo-Castro, M. F., Popovic, D., **Oeztuerk, O. F.**, Haas, S. S., Weiske, J., Hauke, D., Neufang, S., … Koutsouleris, N. (2022). Clinical, Brain, and Multilevel Clustering in Early Psychosis and Affective Stages. *JAMA Psychiatry*.
3. Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., **Oeztuerk, O.**, Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., … Meisenzahl, E. (2021). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients with Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*, *78*(2), 195–209.
4. Koutsouleris, N., Pantelis, C., Velakoulis, D., McGuire, P., Dwyer, D. B., Urquijo-Castro, M.-F., Paul, R., Dong, S., Popovic, D., **Oeztuerk, O.**, Kambeitz, J., Salokangas, R. K. R., Hietala, J., Bertolino, A., Brambilla, P., Upthegrove, R., Wood, S. J., Lencer, R., Borgwardt, S., … Johannes, L. (2022). Exploring Links Between Psychosis and Frontotemporal Dementia Using Multimodal Machine Learning. *JAMA Psychiatry*.
5. **Oeztuerk, O. F.**, Pigoni, A., Antonucci, L. A., & Koutsouleris, N. (2022). **Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half-century studies.** In *European Archives of Psychiatry and Clinical Neuroscience* (Vol. 272, Issue 3, pp. 381–393).
6. **Oeztuerk, O. F.**, Pigoni, A., Wenzel, J., … Koutsouleris, N. (2022). **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*.
7. Pigoni, A., Dwyer, D., Squarcina, L., Borgwardt, S., Crespo-Facorro, B., Dazzan, P., Smesny, S., Spaniel, F., Spalletta, G., Sanfelici, R., Antonucci, L. A., Reuf, A., **Oeztuerk, O. F.**, Schmidt, A., Ciufolini, S., Schönborn-Harrisberger, F., Langbein, K., Gussew, A., Reichenbach, J. R., … Brambilla, P. (2021). Classification of first-episode psychosis using cortical thickness: A large multicenter MRI study. *European Neuropsychopharmacology*, *47*, 34–47.
8. Popovic, D., Ruef, A., **Oztuerk, O. F.**, Dong, M. S., & Paul, R. (2020). Traces of trauma: a multivariate pattern analysis of childhood trauma, brain structure, and clinical phenotypes. *Biological Psychiatry*, *88*(11), 829–842.
9. Strube, W., Cimpianu, C. L., Ulbrich, M., **Öztürk, Ö. F.**, Schneider-Axmann, T., Falkai, P., Marshall, L., Bestmann, S., & Hasan, A. (2022). Unstable Belief Formation and Slowed Decision-making: Evidence That the Jumping-to-Conclusions Bias in Schizophrenia Is Not Linked to Impulsive Decision-making. *Schizophrenia Bulletin*, *48*(2), 347–358.
10. Wenzel, J., Haas, S. S., Dwyer, D. B., Ruef, A., **Oeztuerk, O. F.**, Antonucci, L. A., von Saldern, S., Bonivento, C., Garzitto, M., Ferro, A., Paolini, M., Blautzik, J., Borgwardt, S., Brambilla, P., Meisenzahl, E., Salokangas, R. K. R., Upthegrove, R., Wood, S. J., Kambeitz, J., … Rohner, H. (2021). Cognitive subtypes in recent onset psychosis: distinct neurobiological fingerprints? *Neuropsychopharmacology*, *46*(8), 1475–1483.
11. Buciuman, M.-O., **Oeztuerk, O. F**., Popovic, D., …, Koutsouleris, N. (2023). Structural and functional brain patterns predict formal thought disorder's severity and its persistence in recent-onset psychosis: Results from the PRONIA Study. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. https://doi.org/10.1016/j.bpsc.2023.06.001

1. Introductory summary

1.1. Formal thought disorder (FThD)

Formal thought disorder (FThD) is a transdiagnostic psychopathological alteration emerging in schizophrenia, unipolar depression, and bipolar disorder. (T. Kircher et al., 2018; Roche, Creed, et al., 2015) FThD as a multifaceted psychopathology comprises disturbances in the structure of thought, such as derailment, thought blocking, and distractible speech, that impairs speech, and therefore social interaction. (T. Kircher et al., 2014; Roche, Creed, et al., 2015) Formal thought disorder (FThD) is conceptualized as one of the required items i.e. "Disorganized speech" in criteria A of DSM-V for schizophrenia diagnosis. According to the ICD-10, characteristic symptoms of schizophrenia include thought disorders, perceptual disturbances, blunted and inappropriate affect as well as cognitive impairments such as executive dysfunction and deficits in working memory. (McCutcheon et al., 2020) Schizophrenia is a chronic disease with 0,30 - 0,66 % lifetime prevalence that affects 10.2 - 20.2 new individuals in 100000 people worldwide each year. (McGrath et al., 2008; van Os & Kapur, 2009) Schizophrenia causes 1,1 % of the total disability-adjusted life years (DALY) and 2,8 % of the years lived with disability (YLD) (Jablensky, 2000; Picchioni & Murray, 2007) and has been ranked as the 12th most disabling disorder globally in 2016 according to a global survey (Vos et al., 2017).

Previous literature revealed FThD as a fundamental characteristic of psychosis that impacts adversely on social outcomes and functioning levels of patients. (Andreasen & Grove, 1986; Hart & Lewine, 2017; Jerónimo et al., 2018; T. Kircher et al., 2014, 2018) 55% of patients experiencing first-episode psychosis have been affected by various FThD dimensions. (Roche et al., 2016a) FThD intensity has been found positively correlated with the number of acute psychotic episodes, and negatively correlated with quality of life and therapeutic compliance. (Cavelti et al., 2016; Roche et al., 2016a; Tan et al., 2014) Moreover, further findings showed that FThD severity positively associated with patients' unemployment risk (Marengo & Harrow, 1987) and increased rehospitalization rate, as well as negatively associated with perceived quality of life (Sigaudo et al., 2014) and overall life adjustment. (Marengo & Harrow, 1987) Chronic FThD has been found inversely correlated with functioning in work life, but positively correlated with the number of relapses and readmissions into hospitals. (Harrow & Marengo, 1986). However, we noted that most findings regarding the prognostic impact of FThD severity and course were restricted to chronic or heterogeneous populations of patients covering a wide range of age. (Roche et al., 2016a; Roche, Segurado, et al., 2015) We noted that studies investigating the clinical impact and prognostic importance of FThD at early or prodromal stages of psychosis are rare. Therefore, we reviewed the literature systematically (Oeztuerk et al., 2022) to explore the associations between FThD severity and clinical outcomes such as functioning levels, and neurocognition in various domains at cross-sectional and the longitudinal course. As discussed in our review paper (Oeztuerk et al., 2022), scores on "illogical thinking", "poverty of content of speech", and "referential cohesion" (Bearden et al., 2011) as well as scores on the disorganization dimension in the Comprehensive Assessment of At-Risk Mental States (CAARMS) (Demjaha et al., 2012) and "thought blocking" and "tangentiality" (Thompson et al., 2013) have predicted transition to psychosis in clinical high-risk populations. Consistently, "disorganized communication" at baseline as well as persistent "disorganized communication" have been associated with a higher risk for developing a psychotic episode. (DeVylder et al., 2014) Cognitive basic symptoms such as "subjective thought blockage", "interference", and "pressure" as well as "disturbances of abstract thinking"

and “expressive and receptive speech” have been found as subthreshold representations of observable FThD predicting transition to psychosis. (Caplan et al., 1989; Klosterkötter et al., 2001, 2008; Schultze-Lutter et al., 2012) Furthermore, attentional deficits at baseline have been associated with FThD severity. (Ilonen et al., 2010; Nuechterlein et al., 1986) Executive dysfunctions have been found predictive for persistent FThD. (Xu et al., 2014) (Caplan et al., 1989; Klosterkötter et al., 2001, 2008; Schultze-Lutter et al., 2012) Moreover, “thought and communication disturbances” at risk population have been associated with lower functioning levels in social life and at work. (E. Y. H. Chen et al., 1996)

This evidence and outlined research findings suggest the “forgotten” clinical relevance of FThD that calls for a comprehensive investigation of heterogenous FThD manifestations at prodromal or early stages of psychotic experiences and exploration of the interplay between FThD severity and short- and long-term clinical outcomes. However, the accumulating findings and research awareness in early detection of individuals at risk as well as in predictive prognostics and possible preventive interventions in psychiatric diseases, such clinical applications targeting subgroups of patients experiencing predominantly a subset of psychotic psychopathological symptoms e.g., FThD are still missing. Therefore, this presented dissertation entitled “Phenotypical exploration and predictive utility of formal thought disorder in individuals with recent-onset of psychosis” explored the FThD psychopathology in a dimensional approach in early phases of an evolving psychosis and searched for FThD-related neuroimaging patterns as potential predictive biomarkers. As a cumulative work in the frame of the presented dissertation, (Oeztuerk et al., 2022) we first explored the clinical relevance of heterogenous FThD severity by thoroughly investigating whether and how FThD severity is associated with relevant impacts of psychosis such as functioning and neurocognitive impairments already present in early stages by applying unsupervised machine learning algorithms to the clinical data in Paper I (Oeztuerk et al., 2021). Thereafter, we explored the neurobiological correlates of cross-sectional and longitudinal FThD patterns in structural and functional neuroimaging data as potential early diagnostic and prognostic biomarkers by applying supervised machine learning algorithms in Paper II (Buciuman et al., 2023).

Aims of the PhD thesis:

My cumulative dissertation aimed to study first the existing literature on early diagnostic and predictive potential of FThD severity and chronicity by conducting a systematic review. (Oeztuerk et al., 2022) We goaled in Paper I (Oeztuerk et al., 2021) to develop a stratification methodology differentiating patients’ clusters with divergent FThD representations and to explore clinical impacts of such clustering approach by statistical examination of neurocognitive performances and functioning levels of patients in various clusters. We purposed in Paper II (Buciuman et al., 2023) to investigate whether resting-state functional and structural MRI data at baseline could cross-sectionally differentiate FThD clusters and whether FThD cluster membership from baseline to 1-year follow-up could be longitudinally predicted based on these data modalities.

1.2. Clustering FThD subgroups in early psychosis

Research in the last few decades focused on predictive clinical applications based on machine learning algorithms to better understand psychopathological and neurobiological underpinnings of psychiatric disorders and to recognize chronic adverse effects. (Dwyer et al., 2018; Iniesta et al., 2016; Insel & Cuthbert, 2015; Koutsouleris et al., 2018) Recent studies applying machine

learning and pattern recognition techniques revealed the potential utility of these techniques for addressing the issue of clinical heterogeneity in psychiatric disorders. (Bzdok & Meyer-Lindenberg, 2018; Dwyer et al., 2018; Marquand et al., 2016) Using these methods allowed us to study the early prognostic importance of FThD severity in the heterogeneous clinical representation of psychosis with the aim to identify subgroups of patients suffering from distinct psychotic symptomatology. (Iniesta et al., 2016) Therefore, in the first paper (Oeztuerk et al., 2021) of my cumulative PhD project, we investigated associations between FThD severity and neurocognitive performance and functioning levels in various domains by analyzing the clinical data from 279 patients who experienced recent-onset psychosis (ROP). We assessed observed positive FThD through: i) "conceptual disorganization" (PANSS [*the Positive and Negative Syndrome Scale*] (Kay, S R Fiszbein A, 1987) P2) item, (T. Kircher et al., 2014) and ii) "poverty of content speech" (SANS [the Scale for the Assessment of Negative Symptoms] (Andreasen, 1989) 10) item, whereas we assessed observed negative FThD through iii) "difficulty in abstract thinking" (PANSS N5) item, iv) "increased latency of response" (SANS 12) item, and v) "poverty of speech" (SANS 9) item.

We conducted a multi-step clustering analysis to stratify patients experiencing ROP into subgroups demonstrating differential FThD severity. This analysis comprised four analytical steps, namely (i) the *ClValid* (Burock et al., 2008); validity assessment method, (ii) *NbClust* (Charrad et al., 2014); controlling for the ideal quantity of clusters, (iii) *ClusterStability* (Lord et al., 2017); conducting a stability test, and (iv) *predict.strength* (Tibshirani & Walther, 2005); conducting a generalizability test. In brief, subsequent to the process of data scaling, we first evaluated *k-means*, *hierarchical clustering*, and *partitioning around medoids* (PAM) for their validity through utilization of the *ClValid* (Burock et al., 2008) software package in R. The *ClValid* algorithm employs a methodology of assessing solutions generated from multiple clustering algorithms across a spectrum 2 to 10 clusters. We determined the optimal algorithms and cluster quantities through a consensus approach among calculated internal validity and stability indices. (*supplementary section 3* of Paper I (Oeztuerk et al., 2021)) Then, we reevaluated the algorithms selected as winners through the *NbClust* (Charrad et al., 2014) software package in R, which facilitated the identification of the most valid quantity of clusters (*supplementary section 3* of Paper I (Oeztuerk et al., 2021)). The application of these two validity test yielded k-means portioning and hierarchical clustering as optimal solutions. Thereafter, we assessed the hierarchical clustering solution for its clinical validity and robustness through inspection of both silhouette width figure and the *cophenetic distances* (*supplementary section 4* of Paper I (Oeztuerk et al., 2021)). Subsequently, we assessed the k-means partitioning results for their stability and generalizability as winning algorithm chosen through preceding procedures. (*supplementary sections 5 and 6* of Paper I (Oeztuerk et al., 2021)). Specifically, we employed the *ClusterStability* (Lord et al., 2017) package in R and tested the k-means partitioning solution again for its robustness to eliminate any potential effects that could have arisen due to particular packages or functions. Ultimately, we explored the generalizability of the winning partitioning algorithm for the predicted clustering strength through employing the *predict.strength* (Tibshirani & Walther, 2005) package in R that incorporates the essentials of cross-validation into k-means clustering algorithm.

The multi-step clustering analysis undertaken has indicated a two-cluster solution based on the k-means algorithm as the winning stratification that demonstrated the best validity, stability, and generalizability. This operation stratified two FThD clusters of patients, "FThD-High" (n = 75) and "FThD-Low" (n = 204) (*Figure 1* of Paper I (Oeztuerk et al., 2021)). This clustering solution was explored for further statistical associations between the FThD subgroups regarding overall as well

as syndrome specific illness intensity measures, functioning measures and neurocognition at baseline. The statistical analyses conducted on the clustering solution revealed that there were notable differences in the levels of psychopathological symptom severity between the two FThD subgroups. Specifically, these differences were observed in positive FThD symptoms as measured by “conceptual disorganization” (PANSS P2) (p_{fdr} < 0.001, median = 4 vs. 1, r = 0.488) and “poverty of content of speech” (SANS 10) (p_{fdr} < 0.001, median = 3 vs. 0, r = 0.712), and in negative FThD symptoms as measured by “difficulty in abstract thinking” (PANSS N5) (p_{fdr} < 0.001, median = 3 vs. 1, r = 0.503), “increased latency of response” (SANS 12) (p_{fdr} < 0.001, median = 3 vs. 0, r = 0.653) and “poverty of speech” (SANS 9) (p_{fdr} < 0.001, median = 2 vs. 0, r = 0.611) (*Figure 1* of Paper I (Oeztuerk et al., 2021)). Additionally, a series of supplementary analyses were conducted (i) to eliminate possible associations between overall and syndrome specific disease severity, (ii) to explore potential interplays between the five FThD-related symptoms, that were input features for our multi-step clustering analysis, and FThD subgroups (“FThD-High” and “FThD-Low”), and (iii) to strengthen the specificity of the cluster solution related to the five FThD specific symptoms by excluding the possible confounding effect of positive or negative symptom patterns. The findings of these supplementary analyses were presented in *supplementary sections 7 and 8* of Paper I (Oeztuerk et al., 2021) in detail demonstrating the clinical validity of the clustering solution and that it is specifically driven by FThD severity. Subsequent investigations of further items extracted from PANSS and SANS, including their subscales, excluding FThD-related items yielded clustering solutions that were notably less stable with a lower level of generalizability than the FThD-driven clustering solution.

The statistical analyses on functioning measures indicated significant differences between the two FThD subgroups in the *GF-Social* and *GF-Role* assessments (*Table 2 and Figure 3* of Paper I (Oeztuerk et al., 2021)). The findings indicated that the “FThD-High” subgroup demonstrated significantly reduced social functioning levels measured by *GF-Social* scores when compared to the “FThD-Low” group in the highest lifetime (p_{fdr} < 0.001, r = − 0.216), past year (p_{fdr} < 0.001, r = − 0.219) and baseline variables (p_{fdr} < 0.001, r = − 0.269), respectively. (Oeztuerk et al., 2021) A consistent trend was observed in role functioning levels measured by the *GF-Role* scores that the “FThD-High” group demonstrated significantly reduced role functioning levels when compared to those of the “FThD-Low” group across various time points, including in the highest lifetime (p_{fdr} < 0.001, r = − 0.229), past year (p_{fdr} = 0.001, r = − 0.242) and baseline variables (p_{fdr} < 0.001, r = − 0.259), respectively. (Oeztuerk et al., 2021) The findings of the statistical analyses on neurocognitive measurements indicated that the “FThD-High” group in multiple domains, such as *verbal and semantic fluency*, *verbal short-term memory*, and *abstract reasoning* (*Table 2* of Paper I (Oeztuerk et al., 2021)). The “FThD-High” group exhibited lower scores in both, the *WAIS Vocabulary* (p_{fdr} = 0.002, r = − 0.200) and *WAIS-Matrices* (p_{fdr} = 0.010, r = − 0.166). Similar results were observed in the *phonological verbal fluency* (p_{fdr} < 0.001, r = − 0.235) and *semantic fluency* (p_{fdr} < 0.001, r = − 0.326) scores. (Oeztuerk et al., 2021) Comparable findings were also observed in *the forward* (p_{fdr} = 0.002, r = − 0.204) and *backward* (p_{fdr} = 0.015, r = − 0.151) *digit span* scores

so that individuals in the “FThD-High” group performed worse than individuals in the “FThD-Low” group in these neurocognitive domains. (Oeztuerk et al., 2021)

1.3. Multivariate structural and functional neuroimaging patterns predictive of FThD

Up until now, a limited number of cross-sectional studies have indicated correlations between structural gray matter volume changes within language networks involving frontal and temporal regions, and FThD. (Cavelti, Kircher, et al., 2018; Sans-Sansa et al., 2013) Specifically, total FThD scores in individuals with schizophrenia have been shown associated with white matter volume reductions within tracts related to language network (Cavelti, Winkelbeiner, et al., 2018). Furthermore, FThD scores in specific subdomains have been shown associated with fractional anisotropy in the right posterior cingulum bundle, inferior longitudinal fascicle and anterior thalamic radiation (Stein et al., 2022). Moreover, scores in different FThD sub-dimensions exhibited distinct associations with brain volume reductions in prefrontal lobe, orbitofrontal cortex, superior temporal gyrus – STG, amygdala-hippocampus, cerebellum vermis, nucleus accumbens. (Sumner et al., 2018b) Vita et al. also yielded similar findings showing an inverse correlation between distractible speech and illogicality and left prefrontal lobe volume, as well as between incoherence and tangentiality and left and right lobe volume among younger population diagnosed with schizophrenia. (Vita et al., 1995) FThD-associated structural brain surrogates have predominantly been documented in individuals who suffer from chronic schizophrenia. Notwithstanding, current research endeavors have yet to investigate the FThD-associated structural alterations that are observed in the initial stages of psychotic and longitudinal predictive approaches that examine the progression of FThD contingent upon structural brain changes.

Likewise, few task-based and resting-state fMRI studies using univariate region-of-interest-based approaches have found associations between functional abnormalities within the language and executive networks and total FThD (Cavelti, Kircher, et al., 2018; Horn et al., 2012; T. T. J. Kircher et al., 2001; Sumner et al., 2018a; Wensing et al., 2017), as well as associations for different FThD subdomains in schizophrenia (J. Chen et al., 2021; Fuentes-Claramonte et al., 2021; T. T. J. Kircher et al., 2001; Winkelbeiner et al., 2018). Additionally, recent studies explored FThD-associated functional resting-state changes by applying a multivariate framework to demonstrate the whole-brain patterns. For instance, a study combining a seed-based whole-brain resting-state functional connectivity with machine learning algorithms identified robust clusters associated with FThD symptom dimensions. (J. Chen et al., 2021) Although the seed-based functional connectivity metrics have received considerable attention, one study has indicated an association between other voxel-level measures such as degree centrality and FThD (Dey et al., 2021). The fractional amplitude of low-frequency fluctuations (fALFF) (Zang et al., 2007) is another whole brain voxel-level connectivity metric that captures the voxel-level intensity of spontaneous fluctuation during rest reflecting local functional integrity and inter-relating to higher-order functional connectivity based on spatial and temporal dimensions of rs-fMRI data (Fu et al., 2018; Tomasi

et al., 2016). Researchers have reliably found fALFF alterations in functional connectivity data from patients with schizophrenia. (Hoptman et al., 2010; Yu et al., 2014) Resting-state functional changes related to FThD or other psychopathological dimensions of psychosis have not been investigated through fALFF alterations, yet.

Distinct structural and functional changes in the brain have been associated with the clinical heterogeneous and transdiagnostic presentation of FThD in psychotic disorders, peculiarly in chronic stages, however, early diagnostic and prognostic relevance of such multidimensional brain surrogates of FThD have not been investigated thoroughly, yet. (Jerónimo et al., 2018; T. Kircher et al., 2014, 2018; Roche, Creed, et al., 2015) Hence, the second paper of my cumulative PhD project (Paper II (Buciuman et al., 2023)) filled this gap by exploring multivariate structural and functional correlates of FThD at the cross-sectional and longitudinal level based on the resting-state fALFF, gray matter and white matter volume data at baseline and the 1-year follow-up. For the analyses in Paper II (Buciuman et al., 2023), we analyzed structural and resting-state functional MRI data from 233 individuals (47,6% female) with recent-onset psychosis. we had to exclude 46 participants from the study population explored in Paper I (Oeztuerk et al., 2021) due to missing MRI or follow-up data as well as due to insufficient neuroimaging data quality for few cases. We published all details of data acquisition and processing for both structural and resting state functional MRI data in the main manuscript and supplementary in Paper II (Buciuman et al., 2023) of this cumulative thesis. The open-source toolbox *NeuroMiner* (version 1.1; http://www.pronia.eu/neurominer) has been used to apply all supervised machine learning models. We presented all these models in detail in the supplementary of Paper II (Buciuman et al., 2023). We applied an *outer-leave-site-out inner-pooled nested cross-validation* structure, including scaling, regression of age and sex, modality-specific site correction, PCA-based dimensionality reduction and standardization of the obtained components to preprocess both the structural and functional MRI data (detailed in the supplementary of Paper II (Buciuman et al., 2023)). Then, we trained the models with *binary L2-regularized support vector machine* (SVM) algorithms with a *Maximum Relevance Minimum Redundancy* filter to select differentiating features among each of the three fALFF data sub-bands separately (slow-5, slow-4, slow-3), the gray matter volume maps, and 3) the white matter volume maps to classify FThD-High vs FThD-Low subgroups at baseline as well as persistent vs non persistent FThD at 1-year-follow-up. All the details on these machine learning techniques were published in the supplementary of Paper II (Buciuman et al., 2023) and previous studies from our group (Koutsouleris et al., 2021)).

We presented all model performances for cross-sectional and longitudinal classifying high- and low-FThD patients based on resting-state fALFF, gray matter and white matter volume data in *Table 2, Table S4 and Table S5* of Paper II(Buciuman et al., 2023). Our cross-sectional analyses at baseline showed that the supported vector machine (SVM) algorithm based on gray matter volume could classify the high and low FThD subgroups above chance level (Balanced accuracy (BAC) = 60.8%, Sensitivity = 63.8%, Specificity = 57.7%, p_{fdr} = .002), while the other data modalities did not perform significantly different from chance level. (Buciuman et al., 2023) Volume

increases within cingulate cortex regions including the prefrontal control network and dorsal attentional network and volume decreases in the visual network were the main predictive patterns differentiating the FThD-High subgroup from the FThD-Low subgroup (*Figure 1 and Figure 2* of Paper II (Buciuman et al., 2023)). In our longitudinal analyses regarding high FThD persistence, we revealed that all imaging data modalities performed significantly above chance level in differentiating patients with persistently high FThD from those with other symptom courses (slow-5 fALFF: BAC = 73.2%, Sensitivity = 83.3%, Specificity = 63.1%, $p_{fdr} < .001$; slow-4 fALFF: BAC = 72.3%, Sensitivity = 83.3%, Specificity = 62.4%, $p_{fdr} < .001$; slow-3 fALFF: BAC = 68.0%, Sensitivity = 75.0%, Specificity = 61.0%, $p_{fdr} < .001$; GMV: BAC = 62.7%, Sensitivity = 75.0%, Specificity = 50.4%, $p_{fdr} = .048$; WMV: BAC = 73.1%, Sensitivity = 91.7%, Specificity = 54.6%, $p_{fdr} < .001$). (Buciuman et al., 2023) Stacking all data modalities reached a performance of BAC = 77.0% (Sensitivity = 100%, Specificity = 53.9%, $p_{fdr} = .048$) (*Table S4* of Paper II (Buciuman et al., 2023)). Specifically, widely distributed deactivations and activations within brain networks, such as the *default-mode network*, *dorsal attention network*, and *salience network*, as well as the cerebellum were the main predictive patterns of high FThD persistence with specificity for the different frequency sub-bands fALFF (*Figure 2, Figure 3A, B, C* of Paper II (Buciuman et al., 2023)). In addition, predictive GMV patterns of high FThD persistence consisted of increments within the dorsal attention and salience networks and decrements within regions of the ventral attention and visual networks (*Figure 2, Figure 3D* of Paper II (Buciuman et al., 2023)) were significantly overlapping with the predictive GMV patterns of high FThD at cross-sectional level (*Figure 2* of Paper II (Buciuman et al., 2023)). Lastly, predictive WMV patterns included decreases within frontal tracts and more restricted increases within subcortical tracts (*Figure 3E* of Paper II (Buciuman et al., 2023)).

In our post-hoc analyses, we found positive correlations between the decision scores of the cross-sectional GMV-based model and all fALFF models predicting high FThD persistence (slow-5: $r(151) = .29$, $p_{fdr} < .001$; slow-4: $r(151) = .39$, $p_{fdr} < .001$; slow-3: $r(151) = .58$, $p_{fdr} < .001$). (Buciuman et al., 2023) We did not observe such positive correlations between the decision scores of the cross-sectional GMV-based model and the decision scores coming from the GMV or WMV persistence models (*Table S5* of Paper II (Buciuman et al., 2023)). Moreover, there were no significant correlations between antipsychotic medication (cumulative clozapine equivalent dosage) and the decision scores of any of the significant models at cross-sectional level(GMV: $r(231) = -0.09$, $p_{fdr} = .26$) or any of the persistence classifiers at follow up (Slow-5: $r(151) = -.001$, $p_{fdr} = .98$; Slow-4: $r(151) = -.19$, $p_{fdr} = .06$; Slow-3: $r(151) = -.11$, $p_{fdr} = .26$; GMV: $r(151) = -.01$, $p_{fdr} = .98$; WMV: $r(151) = -.15$, $p_{fdr} = .18$). (Buciuman et al., 2023)

1.4. Summary and outlook of the presented dissertation

In our first study (Oeztuerk et al., 2021), we dissected differing FThD severity as a crucial psychopathological alteration in young individuals experiencing ROP by employing clustering techniques. Specifically, we explored the associative impact of FThD severity on neurocognitive performance and functioning levels in various domains. The findings in Paper I (Oeztuerk et al., 2021)

revealed clusters of individuals experiencing differential FThD who might benefit purposeful early interventions targeting the FThD-related neurocognitive and functioning adverse effects.

In our second study (Buciuman et al., 2023), we demonstrated that multivariate structural and functional brain alterations spanning large-scale brain networks can differentiate FThD subgroups differentiated in Paper I (Oeztuerk et al., 2021) cross-sectionally at baseline. We found that part of the structural and functional changes can predict longitudinally the FThD persistence. Specifically, our findings showed at the cross-sectional level that multivariate patterns of GMV including alterations within the salience, dorsal attention, visual and ventral attention networks classified high vs. low FThD ROP subgroups above chance-level. Our longitudinal analyses revealed that distributed increased and decreased resting-state brain activity within all fALFF sub-bands, as well as GMV patterns at 1-year follow-up overlapping with the cross-sectional alterations and frontal WMV decrease at 1-year follow-up predicted the persistence of high FThD severity. Therefore, Paper II (Buciuman et al., 2023) proved neurobiological correlates of the differential clinical presentations of FThD severity found in Paper I (Oeztuerk et al., 2021) and accentuate the clinical impact of early differential exploration dimensional psychopathology combined with multivariate neuroimaging diagnostics.

The static nature of the categorical diagnostic system challenges the current clinical practice of psychiatry and the dynamic nature of disease courses needs to be accounted for in early diagnostics, prognostics and intervention. The research efforts calling for a paradigm shift motivate further transdiagnostic studies exploring the heterogeneity of clinical manifestations within a categorical diagnosis i.e. depression or schizophrenia, as well as common comorbidity of psychopathological alterations between diagnoses, especially in earlier phases of psychiatric disorders. The newer conceptualizations of psychiatric disorders from one disease to a continuum of transdiagnostic spectra i.e. The Hierarchical Taxonomy of Psychopathology (HiTOP) and/or of dynamic disease courses with a clinical staging approach i.e. Clinical High At Risk Mental State (CHARMS) paradigm underline the critical role of dimensional psychopathology approach to increase the clinical efficiency of early diagnostic and preventive efforts.

Overall, our findings provide evidence supporting the framework of symptom-based dimensional understanding of psychiatric disorders as opposed to traditional diagnoses for representing psychopathological heterogeneity (Bzdok & Meyer-Lindenberg, 2018; Dwyer et al., 2018; Insel & Cuthbert, 2015; Marquand et al., 2016) by showing that heterogenous formal thought disorder symptom severity can be biologically depicted and improve early recognition of populations in risk for poor clinical impacts beginning already at early stages of psychosis. To conclude, the presented PhD dissertation explored FThD in psychotic disorders considering the increasing evidence on the pitfalls of traditional psychiatric diagnoses, a possible upcoming paradigm change and the revolutionary developments of computational machine learning techniques promising early spectrum-wise diagnostics and preventive as well as long-term individual and clinical severity-oriented therapeutics.

2. Contribution to the PhD publications and Appendix

2.1. Contribution to Paper I

I served as first author for the phenomenological clustering manuscript (Paper I) on formal thought disorder in patients with recent onset psychosis. I worked as a study physician in PRONIA team and at the early diagnostic outpatient clinic for two years. During this clinical research period of my residency-PhD Track, I conducted clinical examinations of the patients, wrote clinical reports, provided individual visiting our outpatient or inpatient clinic and their families with psychoeducation on clinical high risk for psychosis, psychotic disorders, and depression. I recruited subjects in the PRONIA study, organized baseline and follow-up examination, conducted clinical interview, collected clinical data from blood samples to neuroimaging data as well as to clinical assessments through baseline and follow-up visits. Under the supervision by Prof. Koutsouleris, I developed the conceptual framework of the paper. I wrote the clustering algorithms and performed all clustering analyses on myself. I wrote the first draft of the paper, including the manuscript, tables, figures, and supplementary material with an immense help of Dr. Antonucci. Finally, I organized the submission process und implemented further analyses and improvements during consortium-intern and peer review process.

2.2. Contribution to paper II

Since the second manuscript (Paper II) exploring multivariate formal thought disorder patterns in structural and functional MRI, included PRONIA sample, I have served as a shared first author and study physician with my contributions by my clinical work for PRONIA study mentioned for the first paper. As part of my PhD project, I have drawn the framework of psychopathological and multivariate neuroimaging pattern analysis under the supervision of Prof. Koutsouleris. I served as shared first author with Ms. Buciuman, with whom I wrote the first draft of the paper together. I wrote the introduction and the discussion, while Ms. Buciuman conducted the analyses, wrote the methods and the results. Ms. Buciuman and I managed the whole consortium-intern and peer review process until the submission.

2.3. Contribution to review paper (Appendix)

I conceptualized the review paper and underlined the clinical relevance of formal thought disorder in psychosis a possible early diagnostic, suggest this core psychopathology in evolving psychosis as a potential target for preventive interventions. I design the systematic review according to PRISMA paradigm and performed a literature search in three different databases composing a half century time period. I teamed up with Mr. Pigoni and Ms Dr. Antonucci and organized all meetings related to literature search, writing and submission. I served as first and corresponding author and wrote the first draft on myself. I implemented revisions from Mr. Pigoni and Ms Dr. Antonnuci and managed further improvements of the review paper to submission.

Paper I

European Archives of Psychiatry and Clinical Neuroscience

The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study

Oemer Faruk Oeztuerk[1,2,3] · Alessandro Pigoni[4] · Julian Wenzel[5] · Shalaila S. Haas[6] · David Popovic[1,2] · Anne Ruef[1] · Dominic B. Dwyer[1] · Lana Kambeitz-Ilankovic[5] · Stephan Ruhrmann[5] · Katharine Chisholm[7] · Paris Lalousis[8] · Sian Lowri Griffiths[8] · Theresa Lichtenstein[5] · Marlene Rosen[5] · Joseph Kambeitz[5] · Frauke Schultze-Lutter[9] · Peter Liddle[10] · Rachel Upthegrove[7,8] · Raimo K. R. Salokangas[11] · Christos Pantelis[12,13] · Eva Meisenzahl[9] · Stephen J. Wood[14,15,16] · Paolo Brambilla[17] · Stefan Borgwardt[18] · Peter Falkai[1] · Linda A. Antonucci[1,19,20] · Nikolaos Koutsouleris[1,3,21] · the PRONIA Consortium

Linda A. Antonucci and Nikolaos Koutsouleris have contributed equally to the work.

1 Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Nussbaumstr. 7, 80336 Munich, Germany
2 International Max Planck Research School for Translational Psychiatry, Munich, Germany
3 Max Planck Institute for Psychiatry, Munich, Germany
4 MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy
5 Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Cologne, Germany
6 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA
7 School of Psychology, Aston University, Birmingham, UK
8 Institute for Mental Health, University of Birmingham, Birmingham, UK
9 Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany
10 Division of Psychiatry and Applied Psychology, Institute of Mental Health, University of Nottingham, Nottingham, UK
11 Department of Psychiatry, University of Turku, Turku, Finland
12 Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia
13 Melbourne Health, Melbourne, Australia
14 School of Psychology, University of Birmingham, Birmingham, UK
15 Orygen, The National Centre of Excellence for Youth Mental Health, Melbourne, Australia
16 Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia
17 Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy
18 Department of Psychiatry, University Psychiatric Clinic, Psychiatric University Hospital, University of Basel, Basel, Switzerland

Keywords: Formal thought disorder · Early psychosis · Clustering · Functioning · Neurocognition

Received: 1 May 2021 / Accepted: 27 August 2021

Published Online: 17 September 2021

Citation: Oeztuerk, O.F., Pigoni, A., Wenzel, J. et al. The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study. Eur Arch Psychiatry Clin Neurosci 272, 403–413 (2022). https://doi.org/10.1007/s00406-021-01327-y

European Archives of Psychiatry and Clinical Neuroscience
https://doi.org/10.1007/s00406-021-01327-y

ORIGINAL PAPER

The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study

Oemer Faruk Oeztuerk[1,2,3] · Alessandro Pigoni[4] · Julian Wenzel[5] · Shalaila S. Haas[6] · David Popovic[1,2] · Anne Ruef[1] · Dominic B. Dwyer[1] · Lana Kambeitz-Ilankovic[5] · Stephan Ruhrmann[5] · Katharine Chisholm[7] · Paris Lalousis[8] · Sian Lowri Griffiths[8] · Theresa Lichtenstein[5] · Marlene Rosen[5] · Joseph Kambeitz[5] · Frauke Schultze-Lutter[9] · Peter Liddle[10] · Rachel Upthegrove[7,8] · Raimo K. R. Salokangas[11] · Christos Pantelis[12,13] · Eva Meisenzahl[9] · Stephen J. Wood[14,15,16] · Paolo Brambilla[17] · Stefan Borgwardt[18] · Peter Falkai[1] · Linda A. Antonucci[1,19,20] · Nikolaos Koutsouleris[1,3,21] · the PRONIA Consortium

Received: 1 May 2021 / Accepted: 27 August 2021

Abstract

Background Formal thought disorder (FTD) has been associated with more severe illness courses and functional deficits in patients with psychotic disorders. However, it remains unclear whether the presence of FTD characterises a specific subgroup of patients showing more prominent illness severity, neurocognitive and functional impairments. This study aimed to identify stable and generalizable FTD-subgroups of patients with recent-onset psychosis (ROP) by applying a comprehensive data-driven clustering approach and to test the validity of these subgroups by assessing associations between this FTD-related stratification, social and occupational functioning, and neurocognition.

Methods 279 patients with ROP were recruited as part of the multi-site European PRONIA study (Personalised Prognostic Tools for Early Psychosis Management; www.pronia.eu). Five FTD-related symptoms (conceptual disorganization, poverty of content of speech, difficulty in abstract thinking, increased latency of response and poverty of speech) were assessed with Positive and Negative Symptom Scale (PANSS) and the Scale for the Assessment of Negative Symptoms (SANS).

Results The results with two patient subgroups showing different levels of FTD were the most stable and generalizable clustering solution (predicted clustering strength value = 0.86). FTD-High subgroup had lower scores in social ($p_{fdr} < 0.001$) and role ($p_{fdr} < 0.001$) functioning, as well as worse neurocognitive performance in semantic ($p_{fdr} < 0.001$) and phonological verbal fluency ($p_{fdr} < 0.001$), short-term verbal memory ($p_{fdr} = 0.002$) and abstract thinking ($p_{fdr} = 0.010$), in comparison to FTD-Low group.

Conclusions Clustering techniques allowed us to identify patients with more pronounced FTD showing more severe deficits in functioning and neurocognition, thus suggesting that FTD may be a relevant marker of illness severity in the early psychosis pathway.

Keywords Formal thought disorder · Early psychosis · Clustering · Functioning · Neurocognition

Introduction

Psychotic disorders are closely linked with neurocognitive and functional impairments [17, 49] that frequently precede disease onset and persist after remission of the acute illness [22, 37]. Formal thought disorder (FTD) is a multifaceted construct of disturbances in thought, communication and language, such as loosening of associations, blocking, semantic and phonemic paraphasia [28, 42]. Previous literature revealed that FTD is not only a core feature of psychosis but that it is also associated with adverse social and

Linda A. Antonucci and Nikolaos Koutsouleris have contributed equally to the work.

✉ Oemer Faruk Oeztuerk
oemer.oeztuerk@med.uni-muenchen.de

Extended author information available on the last page of the article

Published online: 17 September 2021

functional outcomes in psychotic patients [3, 21, 25, 27, 28]. More specifically, FTD is associated with an increased (re-) hospitalization rate [44], unemployment risk [34], reduced quality of life [48] and adjustment abilities indexed by occupational functioning and self-support [34]. Harrow et al. [20] highlighted that patients with schizophrenia experiencing enduring FTD after the acute phase of psychosis showed lower occupational functioning levels and higher relapse/rehospitalisation rates. Moreover, thought and communication disturbances in youth at clinical high risk for psychosis were associated with reduced social and occupational functioning outcomes and a higher risk for transition to the established disease [46]. Furthermore, cognitive basic symptoms including subjective thought blockage, interference and pressure as well as disturbances of abstract thinking and expressive and receptive speech that might be regarded as subclinical presentations of observable FTD have been demonstrated to predict subsequent psychosis [45]. In summary, these findings may point towards FTD playing an important role in explaining the behavioural, psychopathological and functional heterogeneity of disease manifestations in both early and prodromal phases of psychosis. So far, this aspect has remained under-investigated, [42, 43] as research on clinical markers of psychosis has focused rather on positive and negative symptoms as well as on cognitive phenotypes of the disorders [44]. However, this traditional approach may not fully capture the clinical heterogeneity of psychosis in terms of both disease course and severity [21, 24, 27, 42]. FTD may represent a clinical fingerprint of disease severity [42] as it encompasses observable speech-related, and cognitive impairments of psychosis. As findings from a recent systematic review indicated, FTD—especially disorganization—might have early diagnostic and prognostic relevance in the early stages of psychosis [39]. The main findings of this systematic review showed that FTD severity predicted poor social functioning, unemployment, relapses, rehospitalisations, as well as correlations between attentional deficits, executive functions and FTD severity, and highlighted the predictive potential of executive dysfunctions for sustained FTD. Therefore, FTD stratification could help clinician to detect a subgroup of patients at risk of developing poor disease outcomes, who may need early preventive interventions targeting specifically FTD-related deficits.

Recently, there has been great interest in machine learning and pattern recognition techniques, which have shown to be promising tools for addressing clinical heterogeneity in psychiatric disorders [7, 15, 35]. Among these, unsupervised clustering algorithms allow us to explore the subgroup structure of psychopathological phenomena in a quantitative and potentially unbiased way [23]. Using unsupervised machine learning techniques to investigate the role of FTD role in the heterogeneity of psychosis phenotypes would allow (i) identifying more homogeneous clinical subgroups experiencing differential illness manifestations, and (ii) exploring the interdependence of possible FTD clusters with other phenotypic expressions of psychosis.

Thus, this study aims for the first time (1) to evaluate whether it is possible to identify robust subtypes of patients with recent-onset psychosis (ROP) that are characterized by distinct FTD patterns, (2) to investigate whether this FTD-related stratification is associated with clinical (i.e., the Global Social (GF-Social) and Role (GF-Role) Functioning), and neurocognitive (i.e., Wechsler Adult Intelligence Scale (WAIS-III; premorbid verbal intelligence), Phonological and Semantic Verbal Fluency (VF—P & S), and Auditory Forward and Backward Digit Span (ADS—Forward and Backward) phenotypes at an early stage of the disease, and (3) to explore the potential associations between FTD-related symptoms, functioning and neurocognition.

Methods

The PRONIA (Personalised Prognostic Tool for Early Psychosis Management) study recruited patients into the recent-onset psychosis (ROP) study group if, i) they fulfilled the DSM-IV-TR criteria for affective and non-affective psychotic episode lifetime, ii) the psychotic episode was present within the past 3 months, and iii) the onset of psychosis occurred within the past 24 months. Exclusion criteria were treatments with antipsychotic medication for longer than 90 days (cumulative number of days) at or above minimum dosage of the 1st episode psychosis range of (DGPPN) S3 guideline. [16] General inclusion and exclusion criteria are detailed in the Supplementary Sect. 1. Based on these criteria, we were able to analyse clinical and neurocognitive data from 279 individuals experiencing a ROP between 15 and 40 years of age (Table 1). Patients were recruited by the PRONIA Consortium between January 2014 and December 2017 at ten sites across five countries (Table 1). All adult participants gave their written informed consent prior to study inclusion. Participants younger than 18 years provided written, informed assent, and their caregivers written, informed consent before being enrolled in the study.

1) Psychopathological assessment of the severity of formal thought disorder

The FTD severity has been assessed with different scales [1, 8, 14] since Kraepelin and Bleuer postulated the importance of earlier manifestation of this clinical phenomenon in an evolving psychosis [21]. The Thought and Language disorder (TALD) scale from Kircher et al. [28] is a validated instrument for assessing FTD. Indeed, it allows clinicians to examine the multifaceted nature of FTD and distinguish

Table 1 Study-associated sociodemographic

	Formal thought disorder related symptom severity			
	Low	High	χ^2	*p* value
Age, median	24	23		0.011
Female, No. (%)	91 (44)	28 (37)	0.7053	0.401
Education year, median	14	12		< 0.001
Participants per site, No				
The Ludwig-Maximilian-University Munich	76	20	16.452	0.058
The University of Cologne	30	15		
The University of Münster	6	4		
The University of Düsseldorf	2	3		
The University of Basel	18	6		
The University of Turku	33	7		
The University of Milan	15	10		
The University of Udine	8	3		
The University of Bari	1	3		
The University of Birmingham	16	3		

positive and negative thought disorder with subjective and objective components. We operationalized only observed positive and negative FTD with items from the Positive and Negative Symptom Scale (PANSS) [26] and the Scale for the Assessment of Negative Symptoms (SANS) [2] with a psychopathological orientation on the Thought and Language disorder (TALD) scale from Kircher et al. [28]

More in detail, observed positive FTD was assessed through:

- conceptual disorganization (PANSS P2) item, reflecting the following psychopathological alterations listed in the TALD scale; tangentiality, circumstantiality, derailment, dissociation of thinking, cross talk, and logorrhoea [26, 28].
- poverty of content speech (SANS 10) item is included in the TALD scale [2, 28]

On the other hands, observed negative FTD was assessed through:

- the difficulty in abstract thinking (PANSS N5) item is called as concretism in the TALD scale, that reflects the same psychopathological alteration [26, 28].
- increased latency of response (SANS 12) item is called as slowed thinking in the TALD scale that also reflects the same psychopathological alteration [2, 28].
- poverty of speech (SANS 9) item is included in the TALD scale [2, 28].

These five FTD-related symptoms were used as features to cluster patients with ROP into FTD subgroups. Notably, PANSS N6 item (i.e., "Lack of spontaneity and flow of conversation") and SANS 11 item (i.e., "Blocking") individual scores reflecting following TALD scale subjective negative items: "Poverty of thought", "Dysfunction of thought initiative", "Intentionality and expressive speech dysfunction" and "Inhibited thinking", were not included as a feature on purpose in the present study to avoid redundancy between objective and subjective FTD assessments [28, 33].

Trained clinicians assessed psychopathology of each participant and interrater reliability tests were performed regularly to calibrate PANSS [Intraclass Correlation (ICC) = 0.79] across study sites. Interrater reliability test for SANS is not available.

2) Clustering FTD subgroups based on psychopathological patterns

Our first aim was to identify a stable and generalizable ROP patient stratification based on distinct patterns of FTD identified by means of data-driven unsupervised machine learning techniques. After scaling the data (Supplementary Sect. 2), we applied the following steps to investigate and compare alternative FTD subgroup solutions.

2a) Validity and stability of clustering methods

First, we used the *ClValid* [6] package to assess three clustering algorithms: (i) k-means, (ii) hierarchical clustering, and (iii) partitioning around medoids (PAM) with the number of clusters ranging from 2 to 10. *ClValid* [6] compares solutions from several clustering algorithms while the numbers of clusters vary. This allows researchers to choose the optimal algorithm and number of clusters with a majority rule based on a battery of internal validity and stability measures (for a full description, see [6]). We tested average and ward linkage for the hierarchical clustering due to previous simulation study from Walesiak and Dudek [54] reporting them as the best linkages for ordinal data. The average linkage considers the distance between two clusters as the average distance between each point in one cluster to every point in the other cluster, whereas ward linkage is a method that minimizes the error sum of squares between the clusters over all the variables. We selected the optimal algorithms and numbers of clusters by applying a majority rule among computed internal validity and stability measures (Supplementary Sect. 3).

Second, we retested the winner algorithms determined with majority rule in the first step with the *NbClust* [11] package for the optimal number of clusters (Supplementary Sect. 3). Many indices have been reported previously to decide for a valid clustering solution in a given dataset. *NbClust* provides an automatized scan through 26 validity indices such Calinski–Harabasz (CH) Index [9], Davies–Bouldin (DB) Index [13], Silhouette Index [46] and

reports the optimal number of clusters for a given clustering algorithm with majority rule. Following the first step, we tested the average and ward linkage for the hierarchical clustering algorithm, which was one of the winner algorithms (Supplementary Sect. 3).

2b) Stability and generalizability of the optimal cluster solutions

Given that the first and second steps resulted in more than one optimal solution, namely k-means and hierarchical algorithms (Supplementary Sect. 3), we inspected the solutions from the hierarchical clustering for their clinical validity and robustness with the figure for silhouette width as well as the cophenetic correlation in a third step (Supplementary Sect. 4).

Thereafter, we examined the stability and generalizability of k-means algorithm solution selected based on the previous steps (Supplementary Sect. 5 and 6). To exclude possible package or function-specific effects, we retested the robustness of the k-means-based clustering solution with the R package *ClusterStability* [31]. This package allows screening popular validity measures and provides researchers with a global stability (ST) index and an individual ST-index ranging from 0 to 1, where 1 indicates very strong stability (Supplementary Sect. 5).

Lastly, we investigated the predicted clustering strength of the partitioning algorithms using the *predict.strength* package of Tibshirani et al. [52], which applies the principles of cross-validation, well established in supervised machine learning, to the unsupervised case. *Predict.strength* performs n-fold random resampling, partitions the resampled population into training and test data folds, and clusters these with varying numbers of clusters through m 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. In the present study, subjects were randomly resampled over 500 iterations and partitioned each time into test and training datasets using two-fold cross-validation. The highest prediction strength over cut off value 0.80 was chosen as the optimal *predict.strength* value, following published procedures [52] (Supplementary Sect. 6).

The k-means clustering solution survived these validity, stability and generalizability tests and was chosen for further association tests with global and syndromal measures of disease severity, such as the PANSS and SANS total scores, as well as each PANSS and SANS subscales between the identified FTD subgroups. Respective results were reported in Supplementary Sect. 7. To test the specificity of the k-means clustering solution, we run several sanity analyses using items from PANSS and SANS subscales that are not related to FTD and reported the results in Supplementary Sect. 8.

We compared age, educational years, clinical outcomes; Global Social (GF-Social) and Role (GF-Role) Functioning scores, and neuropsychological performances; abstract reasoning, verbal fluency, processing speed, verbal short-term and working memory between FTD subgroups using the Mann–Whitney-*U* test [32] after checking for normal distribution with the Shapiro–Wilk test [48]. Distributions of sex, site and FTD subgroups, respectively, were compared using the χ^2 tests [36] (Tables 1 and 2). All analyses and univariate statistical comparisons were conducted with R version 3.5.2. We used the False Discovery Rate (FDR) [5] to correct all *P* values for the multiple comparisons. *P* values of the Sociodemographic: Age, sex, education, and the number of participants per site, *P* values of the clinical outcomes and *P* values of the neurocognition were considered dependent and corrected using FDR. We provided effect sizes calculated with the *wilcoxonR* function in the R package *rcompanion* for each nonparametric statistical comparison (Fig. 1). We tested the FDR-corrected significance of correlations and provided the correlogram that is a graphical representation of the correlation matrix of all included variables. (Fig. 2).

3a) Association of the clustering solution with clinical outcomes; social and role functioning

We investigated the association of the selected clustering solution with clinical outcomes by performing between-groups statistical comparisons of baseline the GF-Social and GF-Role functioning scores including the retrospective assessments of highest functioning levels lifetime as well as past year. [10, 12] Trained clinicians assessed the functioning level of each participant and interrater reliability tests were performed regularly to calibrate the Global Functioning scales; GF-Social [Intraclass Correlation (ICC) = 0.945 and GF-Role (ICC = 0.924)] across study sites [29]. We compared the differences in social and role functioning scores between two clusters with the Mann–Whitney-*U* test [32] or with the Welch's two-sample *t*-test [54] after checking for normal distribution with the Shapiro–Wilk test [48].

3b) Association of FTD-defined subgroups and neurocognitive performance

We analysed the following neurocognitive domains; the WAIS-III; premorbid verbal intelligence assessing visual processing and abstract reasoning, the VF—P & S assessing verbal fluency and processing speed and the ADS—F & B assessing verbal short-term memory and verbal working memory from the neurocognitive battery of the PRONIA study (Supplementary Table 1), because previous literature has shown a significant association between these neurocognitive domains and FTD as measured by Thought, Language,

Table 2 Clinical, functioning and neurocognition differences in individuals with recent-onset psychosis

Characteristics	Formal thought disorder related symptom severity			
	Low	High	*p* value	p_{fdr} value
Global functioning Social scale rated at baseline				
Highest lifetime score, median	8	8	< 0.001	< 0.001
Highest score in past year, median	7	6	< 0.001	< 0.001
Lowest score in past year, median	5	5	0.005	0.005
Current score, median	6	5	< 0.001	< 0.001
Global functioning Role scale rated at baseline				
Highest lifetime score, median	8	8	< 0.001	< 0.001
Highest score in past year, median	7	6	< 0.001	< 0.001
Lowest score in past year, median	5	4	0.001	0.002
Current score, median	6	5	< 0.001	< 0.001
Neurocognition at baseline				
WAIS—premorbid verbal intelligence, median	10	9	0.001	0.002
WAIS—Matrices, median	10	9	0.008	0.010
Phonological Verbal Fluency, median	13	11	< 0.001	< 0.001
Semantic Verbal Fluency, median	21	16	< 0.001	< 0.001
Forward Digit Span, median	9	8	< 0.001	0.002
Backward Digit Span, median	6	6	0.015	0.015

and Communication (TLC) [1] and TALD [28] scale [38, 41, 42, 51, 53]. Welch's two-sample *t*-test [55] or Mann–Whitney-*U* test [32] (based on the results of the Shapiro–Wilk test [48]) were used to assess differences between FTD subgroups in their neurocognitive performances.

We ran several sanity analyses with the following aims: (1) to test whether our clustering solution was associated with global and syndromal disease severity, (2) to investigate the interaction between FTD-related symptoms, (3) to corroborate the specificity of our clustering solution with FTD-related symptoms, (4) to exclude that the clustering algorithms provide us with a solution more prone to detect negative symptom pattern and (5) to test whether our clustering protocol recognizes a positive symptom pattern, and reported results in the Supplementary Sects. 7 and 8 in details. These sanity analyses showed us that our FTD-driven clustering solution was clinically valid and specific to FTD-related symptom severity. Analyses with non-FTD items from PANSS and SANS as well as their subscale showed less stable and less generalizable clustering solutions than the FTD-driven clustering solution.

Results

Our multi-step clustering analyses identified a k-means algorithm-based two-cluster solution as the most stable and generalizable stratification approach. This approach delineated two FTD subgroups, FTD-High and FTD-Low ($n = 75$ vs. 204) in our ROP patient cohort (Fig. 1). The clustering solution was significantly informed by (1) observed positive FTD as measured by conceptual disorganization (PANSS P2) (median = 4 vs. 1, $p_{fdr} < 0.001$, $r = 0.488$) and poverty of content of speech (SANS 10) (median = 3 vs. 0, $p_{fdr} < 0.001$, r = 0.712), and (2) observed negative FTD as measured by difficulty in abstract thinking (PANSS N5) (median = 3 vs. 1, $p_{fdr} < 0.001$, $r = 0.503$), increased latency of response (SANS 12) (median = 3 vs. 0, $p_{fdr} < 0.001$, r = 0.653) and poverty of speech (SANS 9) (median = 2 vs. 0, $p_{fdr} < 0.001$, $r = 0.611$) (Fig. 1).

As reported in Table 1, we observed no significant interaction of FTD-informed subgroups with sex ($\chi^2 = 0.7053$, $p_{fdr} = 0.401$, phi = 0.050) and site ($\chi^2 = 16.452$, $p_{fdr} = 0.073$, phi = 0.243). Furthermore, the distribution of missing data did not show significant differences between FTD subgroups (Supplementary Table 5). At baseline, the FTD-High group was younger than the FTD-Low group (median = 23 vs. 24, $p_{fdr} = 0.022$, $r = 0.153$) in group level comparison (Supplementary Fig. 5). The two groups also differed in their education level: with fewer years in education in FTD-High group than in FTD-Low group (median = 12 vs. 14, $p_{fdr} = 0.002$, $r = 0.209$). 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 ($p_{fdr} = 0.779$, $r = 0.017$) but was associated with more pronounced negative symptoms as measured by SANS total

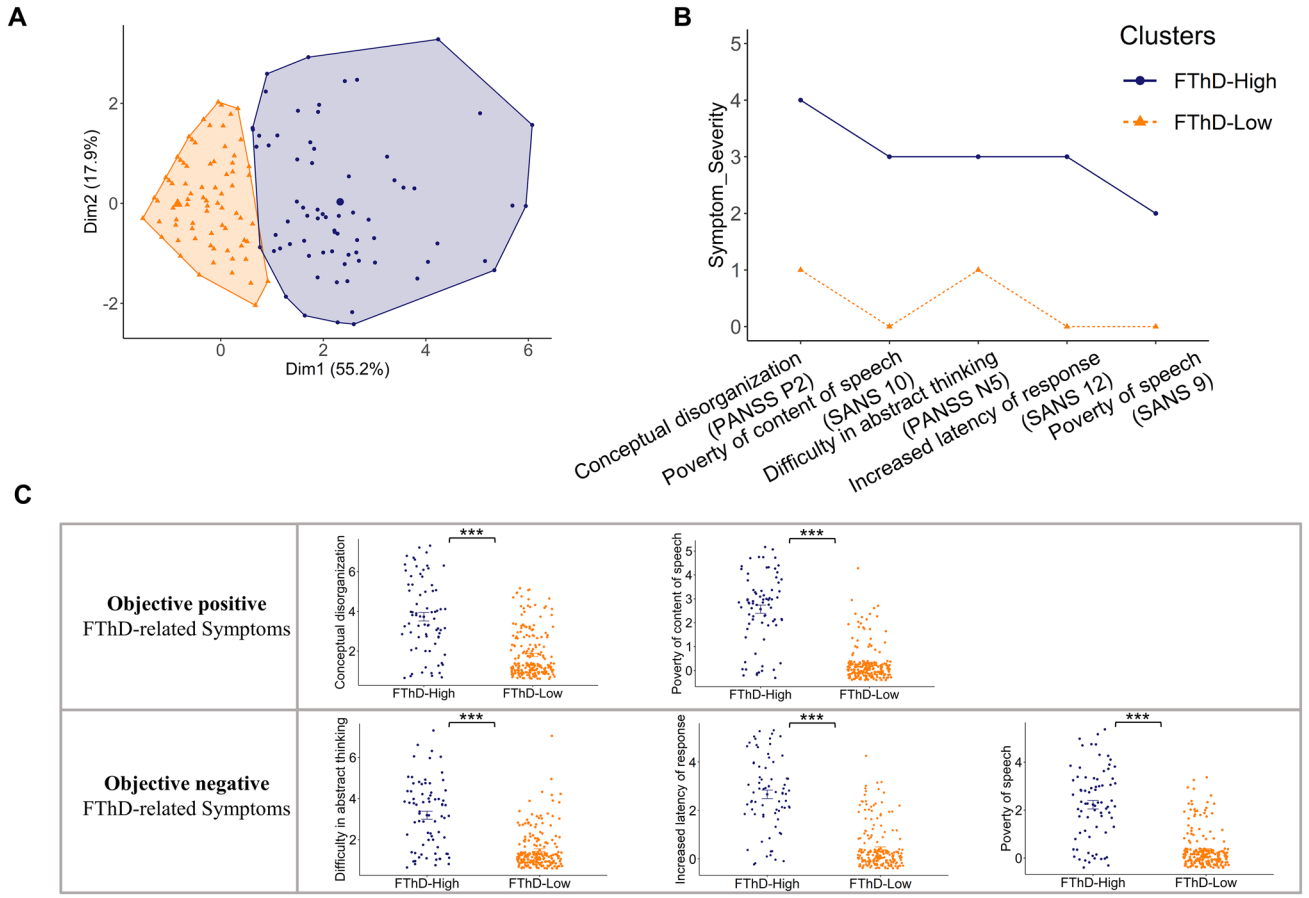


Fig. 1 The Psychopathological Comparison of FTD subgroups. **A** Represents the results of the principal component analysis in two-dimensional space, **B** the difference between medians of FTD-related symptom severity, **C** the distributions of each FTD-related symptom and their statistical comparisons with Wilcoxon rank-sum test. Statistical significances are shown $***p_{fdr} < 0.001$

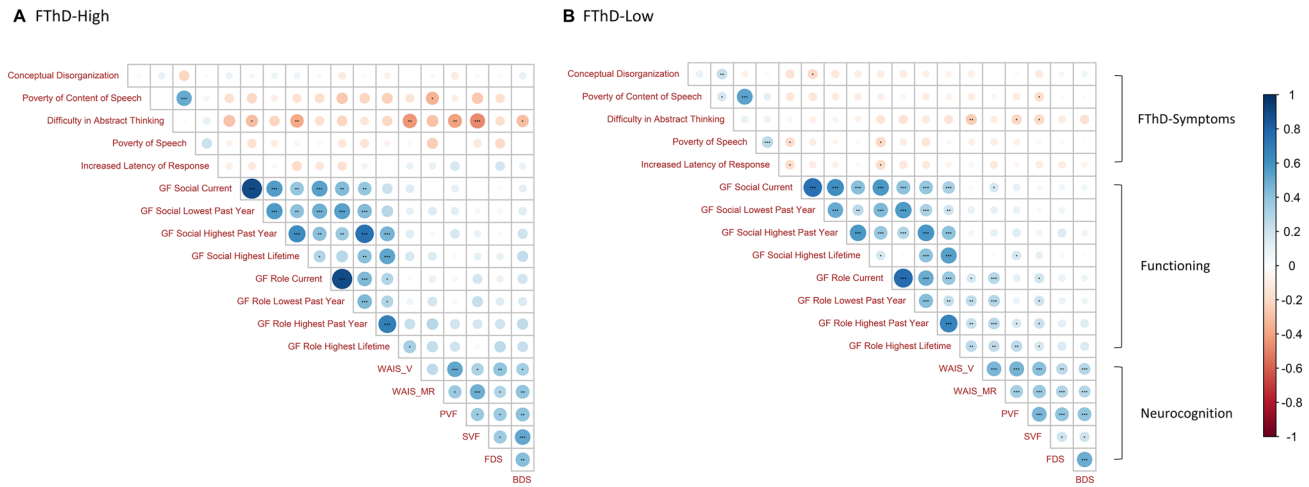


Fig. 2 Correlogram among FTD-related symptoms, functioning and neurocognition; **A** FTD-High **B** FTD-Low. Different colours: red = negative or blue = positive represent the direction of correlations, different size of the circles represents the strength of the correlations. Statistical Significances are shown; $*p_{fdr} < 0.05$, $**p_{fdr} < 0.01$, $***p_{fdr} < 0.001$

score at baseline ($p_{fdr} < 0.001$, $r = -0.551$) in the high FTD subgroup (Supplementary Table 6).

Comparison of functioning between FTD subgroups

FTD subgroups differed significantly in the GF-Social and GF-Role instruments (Table 2 and Fig. 3). The GF-Social scores were lower in the FTD-High compared to the FTD-Low group in the highest lifetime ($p_{fdr} < 0.001$, $r = -0.216$), past year ($p_{fdr} < 0.001$, $r = -0.219$) and baseline variables ($p_{fdr} < 0.001$, $r = -0.269$). Similarly, GF-Role scores were lower in the FTD-High vs. FTD-Low group in the highest lifetime ($p_{fdr} < 0.001$, $r = -0.229$), past year ($p_{fdr} = 0.001$, $r = -0.242$) and baseline variables ($p_{fdr} < 0.001$, $r = -0.259$).

Comparison of neurocognitive performance between FTD subgroups

Comparisons of neurocognitive measures between FTD subgroups showed significant differences in verbal and semantic fluency, verbal short-term memory and abstract reasoning (Table 2). The WAIS-Vocabulary ($p_{fdr} = 0.002$, $r = -0.200$) and WAIS-Matrices ($p_{fdr} = 0.010$, $r = -0.166$) scores were lower in the FTD-High group than in the FTD-Low group. We found a similar pattern of results in the phonological verbal fluency ($p_{fdr} < 0.001$, $r = -0.235$) and semantic fluency ($p_{fdr} < 0.001$, $r = -0.326$) scores, as well as in the forward ($p_{fdr} = 0.002$, $r = -0.204$) and backward ($p_{fdr} = 0.015$, $r = -0.151$) digit span scores, i.e., FTD-High individuals always performed worse than FTD-Low group in these neurocognitive domains.

Correlation analyses of FTD-related symptoms, functioning and neurocognition

The correlation analyses in the FTD-High subgroup revealed that "difficulty in abstract thinking" correlated negatively with functioning GF-Social lowest in past year: ($p_{fdr} = 0.038$, $r = -0.30$); GF-Social highest in lifetime: ($p_{fdr} = 0.006$, $r = -0.37$)), and neurocognitive domains (WAIS-Vocabulary: ($p_{fdr} = 0.002$, $r = -0.43$); backward digit span: ($p_{fdr} = 0.042$, $r = -0.30$); phonological verbal fluency: ($p_{fdr} = 0.003$, $r = -0.40$); and semantic fluency: ($p_{fdr} < 0.001$, $r = -0.48$)). Poverty of content of speech item was significantly associated with WAIS-Matrices ($p_{fdr} = 0.016$, $r = -0.35$) in the FTD-High subgroup. The correlations between other FTD-related symptoms and other variables in functioning or neurocognition were either non-significant or significant with a small effect size ($r < 0.30$) in both subgroups (Fig. 2).

Discussion

In this multisite naturalistic study, we addressed for the first time the psychopathological heterogeneity of patients with recent-onset psychosis from the perspective of the core syndrome of FTD with unsupervised machine learning algorithms. Furthermore, we investigated the broader quantitative associations between FTD, neurocognitive performance and level of functioning in the early stage of psychosis with group-level statistical comparisons. Our multi-step clustering analysis yielded a solution with two FTD subgroups as the most optimal stratification scheme after running the following four checkpoints; (i) validity, (ii) re-evaluation of validity results and unbiased determination of the winning algorithm, (iii) stability test and (iv) generalizability test for the best clustering solution. In summary, the winning clustering solution revealed two stable subgroups of patients with high and low severity of FTD, which were independent of global disease severity in this early stage of the disease. The FTD-High subgroup showed significant impairment of all functional domains and significantly lower neurocognitive performance in verbal and semantic fluency, short-term verbal memory and abstract thinking. Moreover, the median age at baseline was one year less in both sexes in the FTD-High subgroup and an earlier differentiating peak of male distribution was observed in the FTD-High subgroup in the late adolescence and early adulthood. (Supplementary Fig. 5) This is in keeping with previous clinical observations of worse prognostic long-term outcomes in males. [19, 30]

Our results are also in line with the previous literature showing that thought disorders are negatively associated

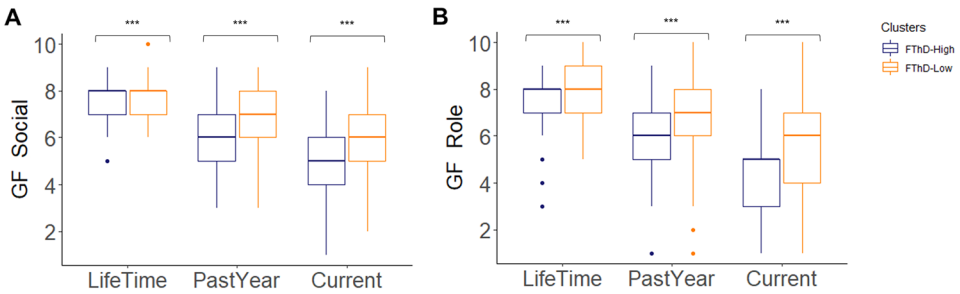


Fig. 3 The Comparison of functioning levels in social and role functioning domains. Statistical comparisons are conducted with the Welch two-sample or the Mann–Whitney-*U* tests based on the distribution of the data. Statistical Significances are shown; ***$p_{fdr} < 0.001$

with role functioning [20], and thought and communication disturbances are related to poorer social and role functioning levels [4]. Retrospective clinical assessments at study inclusion indicated that FTD subgroups started to deviate in symptom and disability measures as well as in social and role functioning already in the year prior to the study. The more specific assessment of social and role functioning differences between FTD subgroups extended this observation to the lifetime scale that may point to FTD subtypes being as a sensitive prognostic marker for a later manifestation of psychosis. [4, 18, 56] These findings may represent a starting point for further investigation of these alterations that may help in identify an early diagnostic and interventional window in which interventions focusing on FTD-related impairments may be beneficial in improving functioning during the early stages of psychosis.

Furthermore, we found that FTD stratification was associated with reduced verbal and semantic fluency, and impairments in short-term verbal memory and abstract thinking. This association between FTD and neurocognitive performance in semantic processing, executive functioning, abstract thinking adds to the previous literature. [38, 41, 42, 50, 52] We may speculate that the lower neurocognitive performance of FTD-High group could be seen from a causal consequential perspective and may drive the impaired social and occupational functioning in this subgroup requiring regular and frequent follow-up examinations. [40, 57] However, further studies addressing this interrelation between FTD and neurocognitive performance are warranted to validate this speculation. Moreover, future studies should be conducted to understand whether the relationship between cluster assignment and neurocognition is moderated or mediated by the severity of negative symptoms.

Limitations

A validated specific scale to assess FTD and SAPS (Scale for the Assessment of Positive Symptoms) are missing in the presented study. We assessed a restricted part of FTD spectrum with the overlapping symptoms from PANSS and SANS that mapped on to the TALD. This reduces the complexity of the possible interpretations of our results. Therefore, our findings should be considered as preliminary, and further studies employing a more thorough investigation of FTD manifestations through much more assessments (i.e., TALD, SAPS) are warranted to understand the degree of replicability of our findings. Moreover, the presented study did not differ affective and non-affective psychosis. Another limitation is the cross-sectional nature of the analysed data from the baseline examinations, which significantly limits any causal speculation. Lastly, a missing external validation sample restricts the generalizability of these clustering solutions. Further studies applying presented stratification scheme at the single patient level in an external sample are neededQuery.

Conclusions

The presented multi-step clustering study demonstrated subgroups of patients with distinct clinical presentations of FTD who may have divergent preventive and therapeutic needs related to differential FTD severity. Our findings elucidate how unsupervised machine learning techniques may provide novel insight about the associations between psychopathology, neurocognition and functioning.

In summary, our findings suggest that FTD may be a relevant marker of illness severity in the early psychosis pathway. The reciprocal associations between FTD, functional, and neuropsychological phenotypes of psychosis emphasize the importance of specific treatment pathways for people with more severe FTD. Furthermore, they highlight how FTD may potentially represent a target variable for individualized psycho-, socio-, logotherapeutic interventions aimed at improving neurocognition abilities and personal functioning. Prospective studies should further test this promising perspective.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00406-021-01327-y.

Acknowledgements We thank all subjects for their participation in the study. Oemer Faruk Oeztuerk and David Popovic are supported by International Max Planck Research School for Translational Psychiatry in the frame of Clinician Scientist Program funded by the Else-Kröner-Fresenius-Stiftung. Linda A. Antonucci's salary is funded by the Structural European Funding of the Italian Minister of Education (Attraction and International Mobility—AIM—action, grant agreement No 1859959).

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Andreasen NC (1986) Scale for the assessment of thought, language, and communication (TLC). Schizophr Bull 12(3):473–482

2. Andreasen NC (1989) The scale for the assessment of negative symptoms (SANS). Br J Psychiatry 155:49–52
3. Andreasen NC, Grove WM (1986) Thought, language, and communication in schizophrenia: diagnosis and prognosis. Schizophr Bull 12(3):348–359
4. Bearden CE, Wu KN, Caplan R, Cannon TD (2011) Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. J Am Acad Child Adolesc Psychiatry 50(7):669–680
5. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B (Methodol) 57(1):289–300
6. 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
7. Bzdok D, Meyer-Lindenberg A (2018) Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 3(3):223–230
8. Caissie SL, Liddle PF, Ngan ETC, Quested DJ, Bates AT, White R, Weg R, Anderson CM (2002) Thought and Language Index: an instrument for assessing thought and language in schizophrenia. Br J Psychiatry 181(04):326–330
9. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theor Methods 3(1):1–27
10. Carrión RE, McLaughlin D, Goldberg TE, Auther AM, Olsen RH, Olvet DM, Correll CU, Cornblatt BA (2013) Prediction of functional outcome in individuals at clinical high risk for psychosis. JAMA Psychiat 70(11):1133–1142
11. 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.v061.i06
12. Cornblatt BA, Auther AM, Niendam T, Smith CW, Zinberg J, Bearden CE, Cannon TD (2007) Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. Schizophr Bull 33(3):688–702
13. Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 2:224–227
14. Docherty NM, DeRosa M, Andreasen NC (1996) Communication disturbance in schizophrenia and mania. Arch Gen Psychiatry 53:358
15. Dwyer DB, Falkai P, Koutsouleris N (2018) Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 14(1):91
16. Falkai P, Gaebel W, Weinmann S, Wobrock T (2006) DGPPN (Hrsg.) S3–Praxisleitlinien in Psychiatrie und Psychotherapie Band 1: Behandlungsleitlinie Schizophrenie
17. Gaebel W, Zielasek J (2015) Focus on psychosis. Dialogues Clin Neurosci 17(1):9–18. https://doi.org/10.31887/DCNS.2015.17.1/wgaebel
18. Gooding DC, Coleman MJ, Roberts SA, Shenton ME, Levy DL, Erlenmeyer-Kimling L (2012) Thought disorder in offspring of schizophrenic parents: findings from the New York High-Risk Project. Schizophr Bull 38(2):263–271
19. Hafner H, Maurer K, Loffler W, Riecher-Rossler A (1993) The influence of age and sex on the onset of early course of schizophrenia. Br J Psychiatry 162:80–86. https://doi.org/10.1192/bjp.162.1.80
20. Harrow M, Marengo JT (1986) Schizophrenic thought disorder at followup : its persistence and prognostic significance. Schizophr Bull 12(3):373
21. Hart M, Lewine RRJ (2017) Rethinking thought disorder. Schizophr Bull 43(3):514–522
22. Hou C-L, Xiang Y-T, Wang Z-L, Everall I, Tang Y, Yang C, Xu M-Z, Correll CU, Jia F-J (2016) Cognitive functioning in individuals at ultra-high risk for psychosis, first-degree relatives of patients with psychosis and patients with first-episode schizophrenia. Schizophr Res 174(1–3):71–76
23. Iniesta R, Stahl D, McGuffin P (2016) Machine learning, statistical learning and the future of biological research in psychiatry. Psychol Med 46(12):2455–2465
24. Insel TR, Cuthbert BN (2015) Brain disorders? Precisely. Science 348(6234):499–500
25. Jerónimo J, Queirós T, Cheniaux E, Telles-Correia D (2018) Formal thought disorders—historical roots. Front Psychiatry 9(November):1–5
26. Kay SR, Fiszbeln A, Opler LA (1987) The positive and negative syndrome scale (panss) manual. Schizophrenia Bull 13(2):261
27. Kircher T, Bröhl H, Meier F, Engelen J (2018) Formal thought disorders: from phenomenology to neurobiology. Lancet Psychiatry 5(6):515–526
28. Kircher T, Krug A, Stratmann M, Ghazi S, Schales C, Frauenheim M, Turner L, Fährmann P, Hornig T, Katzev M, Grosvald M, Müller-Isberner R, Nagels A (2014) A rating scale for the assessment of objective and subjective formal thought and language disorder (TALD). Schizophr Res 160(1–3):216–221
29. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, Paolini M, Chisholm K, Kambeitz J, Haidl T (2018) Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiat 75(11):1156–1172
30. Labad X, Cobo J, Usall J, Ochoa S, Kulkarni J (2012) Gender differences in schizophrenia and first-episode psychosis: a comprehensive literature review. Schizophr Res Treat 2012:1–9
31. 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
32. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 18(1):50–60
33. Marder SR, Galderisi S (2017) The current conceptualization of negative symptoms in schizophrenia. World Psychiatry 16:14–24
34. Marengo JT, Harrow M (1987) Schizophrenic thought disorder at follow-up. Arch Gen Psychiatry 44:651
35. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF (2016) Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. Biol Psychiatry Cogn Neurosci Neuroimaging 1(5):433–447
36. McHugh ML (2013) The chi-square test of independence. Biochemia Medica 23(2):143–149
37. Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F (2018) Course of clinical high-risk states for psychosis beyond conversion. Eur Arch Psychiatry Clin Neurosci 268(1):39–48
38. Nagels A, Fährmann P, Stratmann M, Ghazi S, Schales C, Frauenheim M, Turner L, Hornig T, Katzev M, Müller-Isberner R, Grosvald M, Krug A, Kircher T (2016) Distinct neuropsychological correlates in positive and negative formal thought disorder syndromes: the thought and language disorder scale in endogenous psychoses. Neuropsychobiology 73(3):139–147
39. Oeztuerk OF, Pigoni A, Antonucci LA, Koutsouleris N (2021) 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 1–13. https://doi.org/10.1007/s00406-021-01295-3
40. Racenstein JM, Penn D, Harrow M, Schleser R (1999) Thought disorder and psychosocial functioning in schizophrenia: the concurrent and predictive relationships. J Nerv Ment Dis 187(5):281–289

41. Remberk B, Namysłowska I, Rybakowski F (2012) Clinical and cognitive correlates of formal thought disorder in early onset schizophrenia. Activitas Nervosa Superior Rediviva 54(2):68–76
42. Remberk B, Namysłowska I, Rybakowski F (2012) Cognitive impairment and formal thought disorders in parents of early-onset schizophrenia patients. Neuropsychobiology 65(4):206–215
43. Roche E, Creed L, Macmahon D, Brennan D, Clarke M (2015) The epidemiology and associated phenomenology of formal thought disorder: a systematic review. Schizophr Bull 41(4):951–962
44. Roche E, Lyne J, O'Donoghue B, Segurado R, Behan C, Renwick L, Fanning F, Madigan K, Clarke M (2016) The prognostic value of formal thought disorder following first episode psychosis. Schizophr Res 178(1–3):29–34
45. Rodrigues-Amorim D, Rivera-Baltanás T, López M, Spuch C, Olivares JM, Agís-Balboa RC (2017) Schizophrenia: a review of potential biomarkers. J Psychiatr Res 93:37–49
46. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65
47. Schultze-Lutter F, Michel C, Schmidt SJ, Schimmelmann BG, Maric NP, Salokangas RKR, Riecher-Rössler A, van der Gaag M, Nordentoft M, Raballo A (2015) EPA guidance on the early detection of clinical high risk states of psychoses. Eur Psychiatry 30(3):405–416
48. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611
49. Sigaudo M, Crivelli B, Castagna F, Giugiario M, Mingrone C, Montemagni C, Rocca G, Rocca P (2014) Quality of life in stable schizophrenia: The relative contributions of disorganization and cognitive dysfunction. Schizophr Res 153(1–3):196–203
50. Stouten LH, Veling W, Laan W, van der Helm M, van der Gaag M (2017) Psychosocial functioning in first-episode psychosis and associations with neurocognition, social cognition, psychotic and affective symptoms. Early Interv Psychiatry 11(1):23–36
51. Tan EJ, Rossell SL (2014) Building a neurocognitive profile of thought disorder in schizophrenia using a standardized test battery. Schizophr Res 152(1):242–245
52. Tibshirani R, Walther G (2005) Cluster validation by prediction strength. J Comput Graph Stat 14(3):511–528
53. Ventura J, Thames AD, Wood RC, Guzik LH, Hellemann GS (2010) Disorganization and reality distorsion in schizophrenia: a meta-analysis of the relationship between positive symptoms and neurocognitive deficits. Schizophr Res 121:1–14
54. Walesiak M, Dudek A (2010) Finding groups in ordinal data: an examination of some clustering procedures. Classification as a Tool for Research. Springer, pp 185–192
55. Welch BL (1947) The generalization of students' problem when several different population variances are involved. Biometrika 34(1/2):28–35
56. Wilcox J, Winokur G, Tsuang M (2012) Predictive value of thought disorder in new-onset psychosis. Compr Psychiatry 53(6):674–678
57. Yalınçetin B, Ulaş H, Var L, Binbay T, Akdede BB, Alptekin K (2016) Relation of formal thought disorder to symptomatic remission and social functioning in schizophrenia. Compr Psychiatry 70:98–104

Authors and Affiliations

Oemer Faruk Oeztuerk[1,2,3] · Alessandro Pigoni[4] · Julian Wenzel[5] · Shalaila S. Haas[6] · David Popovic[1,2] · Anne Ruef[1] · Dominic B. Dwyer[1] · Lana Kambeitz-Ilankovic[5] · Stephan Ruhrmann[5] · Katharine Chisholm[7] · Paris Lalousis[8] · Sian Lowri Griffiths[8] · Theresa Lichtenstein[5] · Marlene Rosen[5] · Joseph Kambeitz[5] · Frauke Schultze-Lutter[9] · Peter Liddle[10] · Rachel Upthegrove[7,8] · Raimo K. R. Salokangas[11] · Christos Pantelis[12,13] · Eva Meisenzahl[9] · Stephen J. Wood[14,15,16] · Paolo Brambilla[17] · Stefan Borgwardt[18] · Peter Falkai[1] · Linda A. Antonucci[1,19,20] · Nikolaos Koutsouleris[1,3,21] · the PRONIA Consortium

1 Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Nussbaumstr. 7, 80336 Munich, Germany

2 International Max Planck Research School for Translational Psychiatry, Munich, Germany

3 Max Planck Institute for Psychiatry, Munich, Germany

4 MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy

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

6 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, USA

7 School of Psychology, Aston University, Birmingham, UK

8 Institute for Mental Health, University of Birmingham, Birmingham, UK

9 Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany

10 Division of Psychiatry and Applied Psychology, Institute of Mental Health, University of Nottingham, Nottingham, UK

11 Department of Psychiatry, University of Turku, Turku, Finland

12 Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia

13 Melbourne Health, Melbourne, Australia

14 School of Psychology, University of Birmingham, Birmingham, UK

15 Orygen, The National Centre of Excellence for Youth Mental Health, Melbourne, Australia

16 Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

17 Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy

18 Department of Psychiatry, University Psychiatric Clinic, Psychiatric University Hospital, University of Basel, Basel, Switzerland

[19] Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari "Aldo Moro", Bari, Italy

[20] Department of Education, Psychology and Communication Science, University of Bari "Aldo Moro", Bari, Italy

[21] Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

The clinical relevance of formal thought disorder in the early stage of psychosis:

results from PRONIA - Cohort

Supplementary Material

List of Contents:

1. Inclusion and Exclusion Criteria and general methodological information for the PRONIA study

The Personalized Prognostic Tool for Early Psychosis Management (PRONIA) study was registered at the German Clinical Trials Register (DRKS00005042) and the local research ethics committees at each collaboration site approved the study. (Koutsouleris et al., 2018) The PRONIA study included participants with age between 15 and 40 years, with sufficient language skills for participation and providing informed consent/assent. MRI compatibility is another general inclusion criterion due to the main neuroimaging aspect of the study. Individuals are excluded if they have IQ below 70. Current or past head trauma with loss of consciousness longer than 5 minutes, current or past neurological or somatic disorders potentially affecting the structure or functions of the brain, current or past alcohol dependence as well as polysubstance dependence within the past six months are other general exclusion criteria.

Individuals are included in the recent onset of psychosis (ROP) study group, if they fulfil the DSM-IV-TR criteria for affective and non-affective psychotic episode lifetime and if the psychotic episode is present within the past three months. Another ROP study group-specific inclusion criterion is that the onset of psychosis occurred within the past 24 months. ROP study group-specific exclusion criterion is antipsychotic medication for longer than 90 days (cumulative number of days) at or above minimum dosage of the 1st episode psychosis range of DGPPN (German Association for Psychiatry, Psychotherapy and Psychosomatics) S3 guideline.

The clinical evaluation of psychopathological symptom severity, as well as the clinical assessments of clinical outcomes, social and role functioning, have been conducted by trained physicians or clinical psychologist and interrater reliability tests were performed to minimize any site or rater effect. The neuropsychological assessments with 12 paper-pencil and tablet-based tests adapted for the PRONIA consortium (Supplementary Table 1) were administered by trained personnel. All neuropsychological tests were instructed by trained clinical psychologists and

lasted for about 2 hours on average. All participants underwent the same tasks in the same order and were guided by standardized instructions. In language-dependent tasks, English speaking participants or participants who were insufficient in German were evaluated based on test material in English.

2. Scaling Data

Since the clustering results depend on the (dis)similarity between two observations, scaling the data is necessary to have consistent distance measures and to exclude the effect of different ranges in data. We scaled the data to a 0 - 1 range to avoid inconsistency by detecting (dis)similarities among observations due to different range of PANSS (rated on a scale from 1: not present to 7: extremely severe) and SANS (rated on a scale from 0: not at all to 6: severe). After subtracting 1 from PANSS items (PANSS score - 1) to get a range of PANSS scale from 0: not present to 6: extremely severe, we scaled data to a 0 - 1 range from both PANSS score and SANS score with the following function in RStudio: scaletorange = function(x){(x - min(x))/diff(range(x))} . Therefore, the minimum; not present in the SANS and the PANSS scale were treated mathematically as equal in the algorithms. This scaling provided the clustering algorithms with better detection of similar patterns as well as a better differentiation of dissimilarity between the pairs of observations in the given clinical dataset.

3. The decision for k-means clustering and re-evaluation of the clustering stability

Even though clustering algorithms are useful computational methods to recognize multivariate patterns in a given data, these are not straightforward approaches. There are important steps to get robust, replicable and generalizable clustering solutions with clinical utility. In general, we explored the clinical heterogeneity in psychosis in a data-driven manner. First, we applied *ClValid* (Burock et al., 2008) package for three algorithms; k-means, hierarchical and partitioning around medoids. Hierarchical clustering algorithms were tested for two different linkages: average and

ward. As the results (Supplementary Table 2 and 3) show, k-means and hierarchical clustering using average linkage with 2 clusters were the most valid solutions.

Second, we retested these results with *NbClust* (Charrad et al., 2014) package if the optimal number, two is a robust solution. The *NbClust* package provided us with two possible solutions: hierarchical clustering using average linkage with four clusters and k-means with two clusters. We considered also the second optimal solutions; hierarchical clustering using average linkage with two clusters and hierarchical clustering using ward linkage with three clusters. These results are displayed in Supplementary Figure 1. We report the reason for the exclusion of these hierarchical algorithm solutions in the following section. Then we examined the stability and generalizability of the k-means solution with two clusters that we report in Supplementary Section 5 & 6. Thus, we tested many common indices for each step and majority rule showed two as the optimal number of clusters and k-means algorithm provided the most valid, stable and generalizable clustering solution.

4. Exclusion the hierarchical clustering solutions

The *ClValid* package comparing hierarchical clustering with average linkage and Ward linkage, k-means and partitioning around medoids resulted in two as the optimal numbers of clusters for hierarchical clustering using average linkage. We retested the possible optimal number of clusters with the *NbClust* package that provided us with different optimal numbers for average and ward linkage. These solutions for the hierarchical clustering were re-evaluated with the figure for silhouette width as well as the cophenetic correlation and this clustering method was not selected for further analysis due to low cophenetic correlation values, unbalanced cluster size and amount of wrongly clustered subjects showing negative silhouette width that reduces the clinical relevance of those results. (Supplementary Figure 2)

5. The re-evaluation of the cluster stability

The *ClusterStability* (Lord et al., 2017) package was used to retest the robustness of the clustering solution with the k-means algorithm. This R package considers the common validity measures; the Calinski–Harabasz (Caliński & Harabasz, 1974), Silhouette (Caliński & Harabasz, 1974), Dunn (Dunn†, 1974) and Davies–Bouldin (Davies & Bouldin, 1979) measures and provides researchers with a global stability index (global ST-index) and an index for the stability of individual objects (individual ST-index) ranging from 0 to 1, where 1 indicates very strong stability. In the presented study, we applied *ClusterStability* with 500 replications for the k-means algorithms with cluster numbers of two. All global ST-indexes, as well as all individual ST-indexes, were 1. To illustrate the decreasing stability with varying optimal numbers of clusters, we displayed the outputs of global ST-indexes in Supplementary Figure 3.

6. The generalizability of the cluster solution

The *predict.strength* (Tibshirani & Walther, 2005) package is an implementation of cross-validation, a widely used generalization method in supervised machine learning algorithms, for a given unsupervised machine learning question. The principals of the *predict.strength* algorithm are n-fold random resampling and partitioning of observations in training and test sets through m iterations in a given data where the proportion of pair – observations falling into the same clusters is computed. Thibshirani et. Whalter compared 2-fold cross-validation with 5-fold, that did not increase the predictive strength performance, and they continued with 2-fold cross-validation for the further analyses in the original publication. (Tibshirani & Walther, 2005) They provided predicting strength values between 0.8 – 0.9 as optimal. Therefore, subjects were randomly resampled 500 times with a 2-fold cross-validation manner and the number of clusters with highest prediction strength over the cut off value 0.80 was chosen as an optimal number of clusters in the presented study (Supplementary Figure 4).

7. Comparison of psychopathological subdomains of clusters

In order to test whether our clustering solution was associated with global and syndromal disease severity, we did group level statistical comparison using the PANSS and SANS total scores, as well as each PANSS and SANS subscales between the identified FTD subgroups and reported the results showing the psychopathological surrogates of the clustering solution in the Supplementary Table 6. Notably, as reported in the Supplementary Table 6 the statistical comparison of PANSS total or subscale scores between two FTD subgroups did not differ from each other that indicates the specificity of these cluster solution for FTD. In contrast, the SANS scores showed a strong association with our FTD-informed patient subgroups that is in keeping with previous findings linking FTD, negative and cognitive symptom domains to poorer functioning in psychotic disorders. (Cacciotti-Saija et al., 2018; Gerritsen et al., 2019) The NIMH-MATRICS consensus statement previously highlighted five negative symptom domains; blunted affect, alogia, asociality, anhedonia and avolition for a more comprehensive examination of negative symptoms. (Kirkpatrick et al., 2006) The SANS covers a larger part of these defined negative symptoms, whereas the PANSS negative subscale does not incorporate negative symptoms such avolition and anhedonia. (Daniel, 2013; Garcia-Portilla et al., 2015) The different operationalization of the negative symptoms construct in the PANSS and SANS might explain the discrepancy in our results between PANSS negative subscale and SANS.

8. The cluster solution specificity and sanity analyses

To investigate the interaction between FTD-related symptoms; Conceptual Disorganization, Poverty of Content of Speech, Difficulty in Abstract Thinking, Increased Latency of Response, Poverty of Speech and FTD subgroups; FTD-High and FTD-Low, we ran factorial ANOVA (Jaccard & Jaccard, 1998) and reported the results in the Supplementary Table 7 and Supplementary Figure 6. The FTD-related symptoms and the FTD subgroups were entered in the factorial ANOVA to test the main effects of these symptoms and of these subgroups as well as their interaction with each other. Our findings showed a significant main effect of FTD subgroups

($p < 0.001$), a significant main effect of FTD-symptoms ($p < 0.001$) as well as a significant interaction between FTD-symptoms and FTD subgroups ($p = 0.025$).

Moreover, to corroborate the specificity of our clustering solution with FTD-related symptoms, we run the multi-step clustering protocol with PANSS and SANS items that are not related to FTD (Supplementary Figure 7). To exclude that the clustering algorithms provide us with a solution more prone to detect negative symptom pattern, we applied our clustering protocol using (i) 6 items from PANSS negative subscale without the item Difficulty in Abstract Thinking which has been used for the FTD-driven clustering solution (Supplementary Figure 8) and (ii) using all other 17 items from SANS excluding FTD-related symptoms; Poverty of Content of Speech, Increased Latency of Response and Poverty of Speech (Supplementary Figure 9). To test whether our clustering protocol recognizes a positive symptom pattern using 6 items from PANSS positive subscale without Conceptual Disorganization, we applied our clustering protocol to these variables. As reported in the Supplementary Figure 10, the clustering analysis using PANSS positive items without FTD-related Conceptual Disorganization as input has also provided a stable (global ST(stability)-index = 0.99) and highly generalizable (predict.strength value=0.88) two-cluster solution with k-means algorithm. We compared the original FTD-driven solution (n = 279; $n_{FTD\text{-}High\ =\ 75,}\ n_{FTD\text{-}Low\ =\ 204}$) and the cluster solution driven by positive items without Conceptual Disorganization (n=279; $n_{cluster1=\ 128,}\ n_{cluster2\ =\ 151}$) in their size. We also compared the clustering assignment of each participant to observe the proportion of participants who have been assigned into Cluster 1 or Cluster 2. These comparisons showed that (i) the cluster solutions are different in their size and (ii) 149 / 279 participants have been assigned to other clusters. These results together with the results of the factorial ANOVA also strengthen that the clustering solutions are driven by the contribution of different psychopathological domains and that the symptoms entered to clustering algorithm as inputs matter.

To sum up, these sanity analyses showed us that our FTD-driven clustering solution was clinically valid and specific to FTD-related symptom severity. Analyses with non-FTD items from PANSS and SANS as well as their subscale showed less stable and less generalizable clustering solutions than the FTD-driven clustering solution.

8. Comparison of neurocognitive performances corrected for years of education

Comparisons of neurocognitive measures corrected for years of education between FTD subgroups showed significant differences in verbal and semantic fluency, verbal short-term memory and abstract reasoning (Supplementary Table 8). The WAIS-Vocabulary (p_{fdr} = 0.042, r = 0.129) and WAIS-Matrices (p_{fdr} = 0.048, r = 0.130) scores were lower in the FTD-High group than in the FTD-Low group. We found a similar pattern of results in the phonological verbal fluency (p_{fdr} = 0.009, r = 0.196) and semantic fluency (p_{fdr} = 0.012, r = 0.288) scores, as well as in the forward (p_{fdr} = 0.016, r = 0.159) digit span scores, i.e., FTD-High individuals always performed worse than FTD-Low group in these neurocognitive domains. The statistical significance was not observed in the backward (p_{fdr} = 0.108, r = 0.099) digit span scores.

9. Supplementary Tables 1 - 7

Supplementary Table 1. The neurocognitive battery used in the PRONIA study in order of administration. [a] test used for analysis, [b] revised version of the Hopkins' Verbal Learning Test for the University of Turku.

The name of the neurocognitive assessment	The cognitive domains	Administration
Rey-Osterrieth Complex Figure (ROCF) (Gagnon et al., 2003; A M Hubley, 1996; Anita M Hubley & Tremblay, 2002; Osterrieth, 1944)	visuo-spatial construction; visuo-spatial memory (short- and long-term)	paper-pencil format with tablet support
Diagnostic Analysis of Non-Verbal Accuracy (DANVA-2-AF) (Nowicki Jr & Carton, 1993; S Nowicki, 2000; Stephen Nowicki & Duke, 1994)	social cognition	tablet-based
Auditory Digit Span Forward & Backward trials (ADS-F&B) [a,] (Orsini et al., 1987; Wechsler, 2008)	verbal short-term memory verbal working memory	auditory presentation of numbers by recorded (male) voice
Verbal Fluency, Phonological & Semantic trials (VF-P&S) [a,] (Borkowski et al., 1967; Harrison et al., 2000)	verbal fluency; in a phonemic ('S'-words) and in a semantic ('Animals') condition	named words were recorded and written down by an examiner
Rey Auditory Verbal Learning Test (RAVLT)[b,] (McMinn et al., 1988; Rey, 1964)	short-term verbal memory long-term verbal memory	auditory presentation of word list by recorded (male) voice
Trail Making Task, A & B trials (TMT-A&B) (Horton Jr. & Hartlage, 1994)	processing speed sequencing graphomotor capacity visual attention search ability and flexibility	paper-pencil format
Continuous Performance Test Identical Pairs version (CPT-IP) (Bellani & Brambilla, 2008; Cornblatt et al., 1988, 1989)	selective visual attention sustained visual attention	tablet-based
Self-Ordered Pointing Test (SOPT) (Gillett, 2007; Milner et al., 1985; Petrides & Milner, 1982)	short-term visuospatial memory short-term working memory	tablet-based
Digit Symbol Substitution Test (DSST) (Keefe et al., 2004; Wechsler, 2008)	sustained attention working memory processing speed	paper-pencil format
Salience Attribution Task (SAT-SV) (J P Roiser et al., 2009; Jonathan P Roiser et al., 2010, 2012)	explicit and implicit salience adaptive and aberrant salience	tablet-based
Wechsler Adult Intelligence Scale (WAIS-III) [a,] (Wechsler, 2008)		
Vocabulary	premorbid verbal intelligence	paper-pencil format
Matrices	visual processing and abstract reasoning	paper-pencil format

Supplementary Table 2. The results from the *ClValid* package comparing algorithms; hierarchical using ward linkage, k-means and partitioning around medoids (pam) for internal validity and stability measures. Results in bold show the optimal clustering method represented with [a] and the optimal number of clusters represented with [b] for each measure.

Clustering Methods	No. of clusters	Stability measures				Internal validity measures		
		APN	AD	ADM	FOM	Connectivity	Dunn	Silhouette
Hierarchical	**2** [b]	0.2859	0.5916	0.2264	0.2169	**42.4440** [a]	0.0894	0.3354
	3	0.4275	0.5469	0.2560	0.2121	55.1198	0.1072	0.3492
	4	0.3615	0.4897	0.2217	0.2082	65.0385	0.1072	0.2116
	5	0.3970	0.4676	0.2181	0.2061	86.2671	0.1072	0.2290
	6	0.4187	0.4514	0.2271	0.2049	103.6698	0.1072	0.2471
	7	0.4398	0.4384	0.2368	0.2011	108.6286	0.1162	0.2568
	8	0.4059	0.4219	0.2338	0.1976	110.9214	0.1162	0.3027
	9	0.3860	0.4061	0.2263	0.1956	125.0655	0.1162	0.3114
	10 [b]	0.3469	0.3875	0.2101	**0.1903** [a]	129.9125	0.1162	0.3364
k-means	**2** [b]	**0.1081**[a]	0.5312	**0.0813** [a]	0.2071	46.7008	**0.1516** [a]	**0.4586** [a]
	3	0.1832	0.4873	0.1224	0,2030	66.0560	0.1072	0.3813
	4	0.2687	0.4668	0.1883	0.2037	89.5635	0.1072	0.3469
	5	0.3620	0.4611	0.2128	0.2044	99.2310	0.1187	0.3623
	6	0.3977	0.4435	0.2192	0.1999	111.0052	0.1187	0.3652
	7	0.4216	0.4299	0.2288	0.1989	119.0147	0.1280	0.3715
	8	0.3651	0.4095	0.2136	0.1965	124.5056	0.1280	0.3237
	9	0.3716	0.3974	0.2116	0.1950	122.3010	0.1280	0.3356
	10	0.3191	0.3697	0.1845	0.1903	133.2766	0.1348	0.3613
pam	2	0.1495	0.5314	0.1011	0.2128	66.1817	0.0894	0.4125
	3	0.1802	0.4816	0.1123	0.2036	75.1012	0.0971	0.3354
	4	0.3385	0.4734	0.1868	0.2021	97.5560	0.1072	0.3558
	5	0.3097	0.4480	0.1755	0.1995	102.8889	0.1072	0.2725
	6	0.2850	0.4230	0.1735	0.1984	110.4143	0.1072	0.2899
	7	0.3268	0.4150	0.1908	0.1954	135.4651	0.1091	0.2938
	8	0.3783	0.4099	0.2110	0.1939	120.7849	0.1222	0.3104
	9	0.3217	0.3921	0.2054	0.1931	134.3079	0.1091	0.3345
	10 [b]	0.2825	**0.3669** [a]	0.1711	0.1907	133.4821	0.1022	0.3515

Supplementary Table 3. The results from the *ClValid* package comparing algorithms; hierarchical using average linkage, k-means and partitioning around medoids (pam) for internal validity and stability measures. Results in bold show the optimal clustering method represented with [a] and the optimal number of clusters represented with [b] for each measure.

Clustering Methods	No. of clusters	Stability measures				Internal validity measures		
		APN	AD	ADM	FOM	Connectivity	Dunn	Silhouette
Hierarchical	**2** [b]	**0.0181** [a]	0.6283	**0.0337** [a]	0.2287	**7.3155** [a]	**0.3078** [a]	**0.5420** [a]
	3	0.1290	0.6124	0.1135	0.2229	29.2960	0.1451	0.3660
	4	0.1341	0.5644	0.1626	0.2180	57.2714	0.1591	0.4047
	5	0.1674	0.5383	0.1530	0.2086	61.6738	0.1715	0.3763
	6	0.1732	0.5323	0.1507	0.2068	62.3905	0.1715	0.3694
	7	0.2273	0.5248	0.1633	0.2046	65.2083	0.1715	0.3613
	8	0.3137	0.5247	0.1952	0.2003	67.8194	0.1715	0.3609
	9	0.2664	0.4857	0.1733	0.1982	83.8452	0.1980	0.3495
	10	0.2770	0.4820	0.1738	0.1973	83.9563	0.1980	0.3440
k-means	2	0.0941	0.5314	0.0736	0.2072	46.7008	0.1516	0.4586
	3	0.1664	0.4973	0.1393	0,2052	67.6091	0.1072	0.3911
	4	0.2332	0.4776	0.1647	0.2069	85.4873	0.1187	0.4029
	5	0.2584	0.4594	0.1722	0.2007	122.3821	0.1313	0.3877
	6	0.3147	0.4537	0.1990	0.2001	122.8413	0.1313	0.3826
	7	0.3047	0.4345	0.1901	0.1982	131.0464	0.1313	0.3688
	8	0.3139	0.4205	0.1984	0.1971	139.2278	0.1400	0.3647
	9	0.3634	0.4273	0.2149	0.1965	134.9710	0.1601	0.3570
	10	0.3621	0.4066	0.2076	0.1914	153.7004	0.1690	0.3653
pam	2	0.1495	0.5314	0.1011	0.2128	66.1817	0.0894	0.4125
	3	0.1802	0.4816	0.1123	0.2036	75.1012	0.0971	0.3354
	4	0.3385	0.4734	0.1868	0.2021	97.5560	0.1072	0.3558
	5	0.3097	0.4480	0.1755	0.1995	102.8889	0.1072	0.2725
	6	0.2850	0.4230	0.1735	0.1984	110.4143	0.1072	0.2899
	7	0.3268	0.4150	0.1908	0.1954	135.4651	0.1091	0.2938
	8	0.3783	0.4099	0.2110	0.1939	120.7849	0.1222	0.3104
	9	0.3217	0.3921	0.2054	0.1931	134.3079	0.1091	0.3345
	10 [b]	0.2825	**0.3669** [a]	0.1711	**0.1907** [a]	133.4821	0.1022	0.3515

Supplementary Table 4: Represents the results of the Shapiro – Wilk normality test for each statistical comparison, *P* values are shown only with 4 decimals.

	FTD - High		FTD - Low	
	w	***P* value**	**w**	***P* value**
Age	0.9401	0.002	0.9519	< 0.001
Education year	0.9385	0.001	0.9766	0.002
GF Social highest lifetime	0.8425	< 0.001	0.8783	< 0.001
GF Social highest in past year	0.9360	0.001	0.9253	< 0.001
GF Social lowest in past year	0.9515	0.008	0.9505	< 0.001
GF Social current	0.9441	0.003	0.9465	< 0.001
GF Role highest lifetime	0.8144	< 0.001	0.8932	< 0.001
GF Role highest in past year	0.9475	0.004	0.8695	< 0.001
GF Role lowest in past year	0.9414	0.002	0.9531	< 0.001
GF Role current	0.9514	0.007	0.9480	< 0.001
WAIS – premorbid verbal intelligence	0.9758	0.193	0.9821	0.018
WAIS - Matrices	0.9767	0.233	0.9669	< 0.001
Phonological Verbal Fluency	0.9534	0.011	0.9907	0.253
Semantic Verbal Fluency	0.9813	0.390	0.9930	0.498
Forward Digit Span	0.9620	0.031	0.9719	0.001
Backward Digit Span	0.9535	0.010	0.9643	< 0.001
PANSS Total	0.9757	0.166	0.9803	0.006
PANSS Positive	0.9534	0.008	0.9771	0.002
PANSS Negative	0.9141	< 0.001	0.9171	< 0.001
PANSS General	0.9594	0.018	0.9742	0.001
SANS Total	0.9648	0.037	0.9467	< 0.001
SANS Blunting	0.9233	< 0.001	0.8112	< 0.001
SANS Alogia	0.9007	< 0.001	0.6786	< 0.001
SANS Avolition	0.8930	< 0.001	0.8926	< 0.001
SANS Anhedonia	0.8931	< 0.001	0.9002	< 0.001
SANS Attention	0.8885	< 0.001	0.7884	< 0.001

Supplementary Table 5: Represents the number of subjects with missing values per cluster.

	FTD - High	FTD - Low	*P*-value
Education	0	1	1
WAIS premorbid verbal Intelligence	4	18	0.502
WAIS Matrices	6	21	0.762
Phonological verbal fluency	4	12	1
Semantic verbal fluency	5	12	1
Forward digit span	3	11	0.895
Backward digit span	3	11	0.895

Supplementary Table 6: Comparison of two FTD subgroups in PANSS and SANS.

	Formal Thought Disorder-related symptom severity			
Characteristics	**High**	**Low**	**p_{fdr} value**	***p* value**
PANSS at baseline				
Total, median	65.5	67	0.780	0.7780
Positive, median	16	18	0.392	0.275
Negative, median	16	14	0.438	0.351
General, median	32	34	0.506	0.456
SANS at baseline				
Total, median	46.5	16	< 0.001	< 0.001
Blunting, median	15	3	< 0.001	< 0.001
Alogia, median	9	0	< 0.001	< 0.001
Avolition, median	7	4	< 0.001	< 0.001
Anhedonia, median	11	6	< 0.001	< 0.001
Attention, median	4	0	< 0.001	< 0.001

Supplementary Table 7: Representation of results of factorial ANOVA using scaled FTD-related Symptom Severity data.

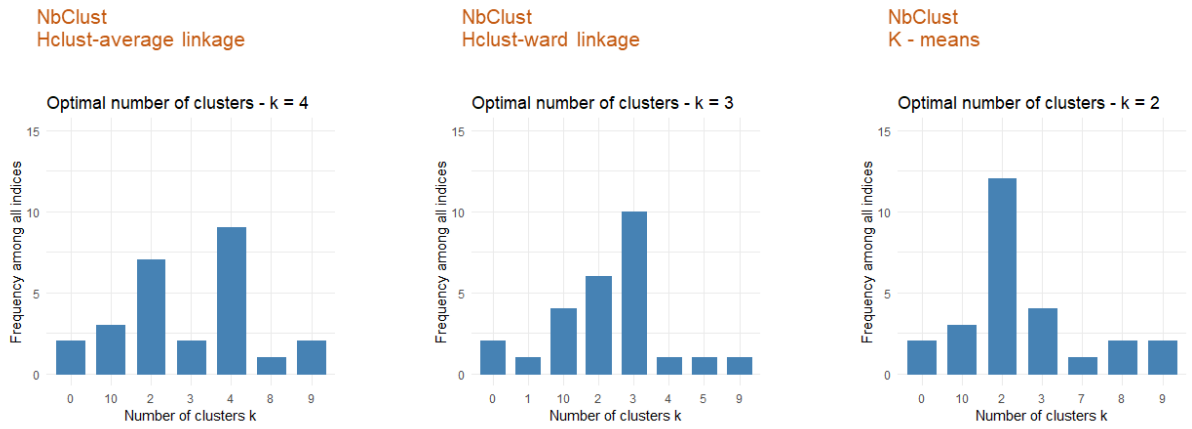
	Df	Sum Sq	Mean Sq	F	Pr(>F)
Clusters	1	30.52	30.519	865.356	< 0.001 ***
Features	4	1.13	0.283	8.013	< 0.001 ***
Clusters: Features	4	0.39	0.098	2.786	0.025 *
Residuals	1385	48.85	0.035		

Supplementary Table 8: Neurocognition Differences corrected for years of education in Individuals with Recent-Onset Psychosis

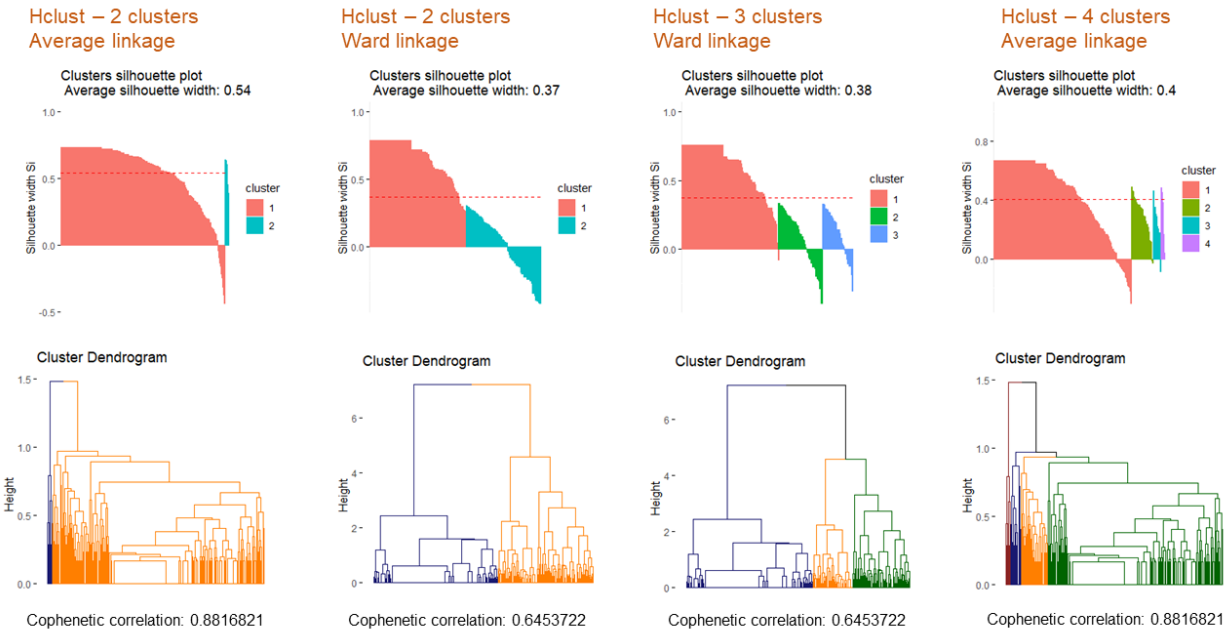
	Formal Thought Disorder related symptom severity			
Characteristics	**Low**	**High**	***p* value**	**p_{fdr} value**
Neurocognition at baseline				
WAIS - premorbid verbal intelligence, median	10	9	0.078	0.042
WAIS - Matrices, median	10	9	0.040	0.048
Phonological Verbal Fluency, median	13	11	0.001	0.009
Semantic Verbal Fluency, median	21	16	0.004	0.012
Forward Digit Span, median	9	8	0.008	0.016
Backward Digit Span, median	6	6	0.108	0.108

10. Supplementary Figures 1 - 10

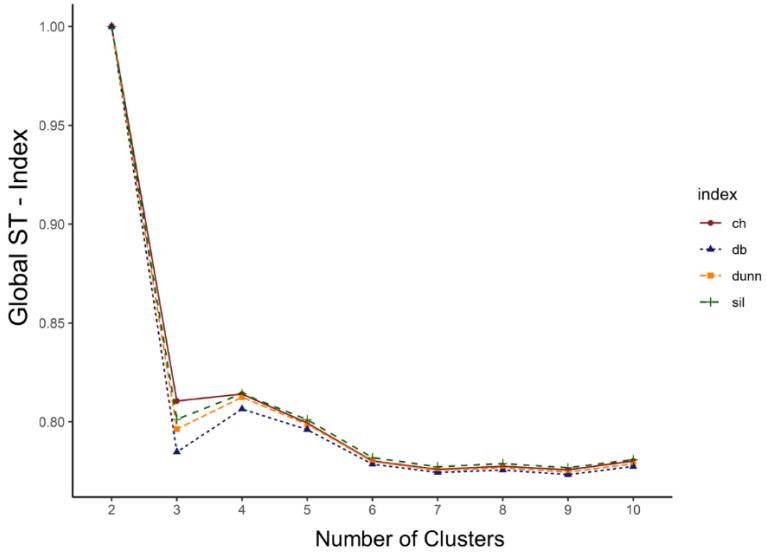
Supplementary Figure 1: The results from *NbClust* package showing the different optimal number of clusters.



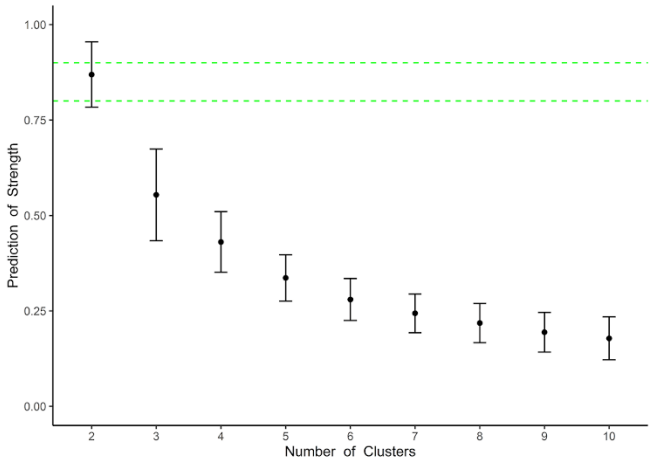
Supplementary Figure 2: The quality check for the solutions with hierarchical clustering algorithms**.**



Supplementary Figure 3: Representation of the decreasing stability with an increasing number of clusters based on four stability indices; ch:Calinski–Harabasz, db: Davies–Bouldin, sil: Silhouette.

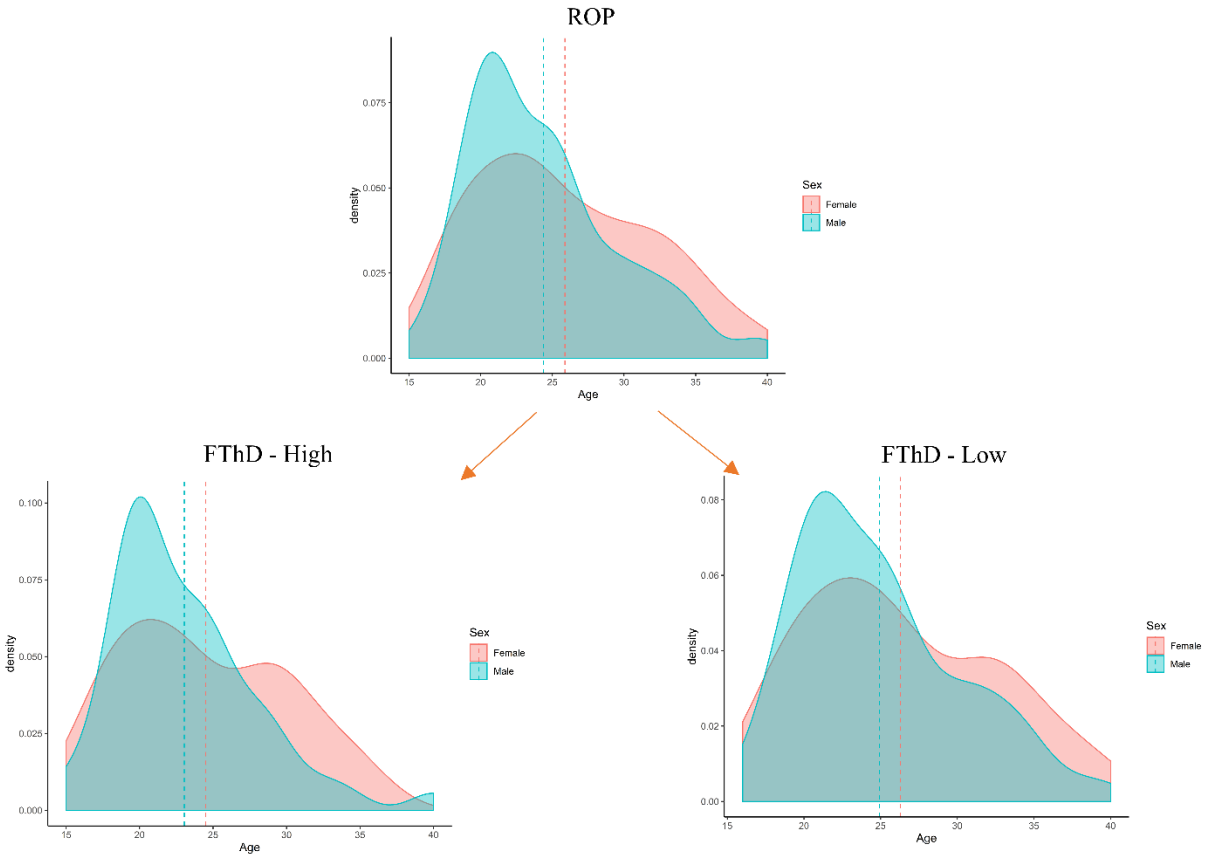


Supplementary Figure 4: Representation of the settings and results of the generalizability test with *predict.strength* package

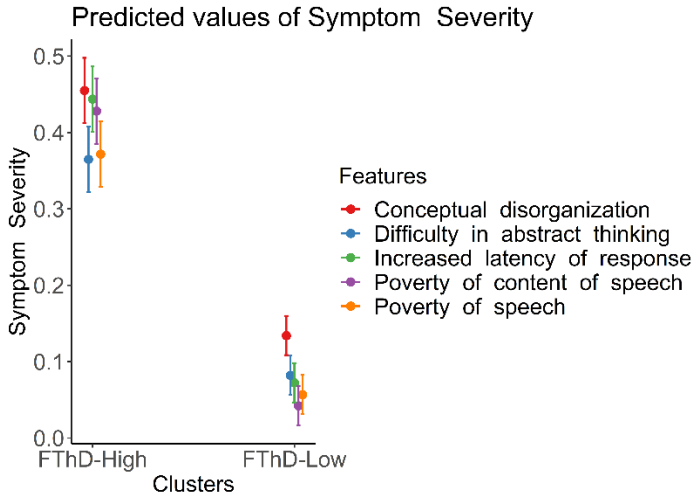



```
Prediction strength
Clustering method: kmeans
Maximum number of clusters: 10
Resampled data sets: 500
Mean pred.str. for numbers of clusters: 1 0.8662567 0.5610063 0.4301843 0.3382322 0.2800358 0.2447623 0.2208781 0.1966664 0.1743017
Cutoff value: 0.8
Largest number of clusters better than cutoff: 2
```

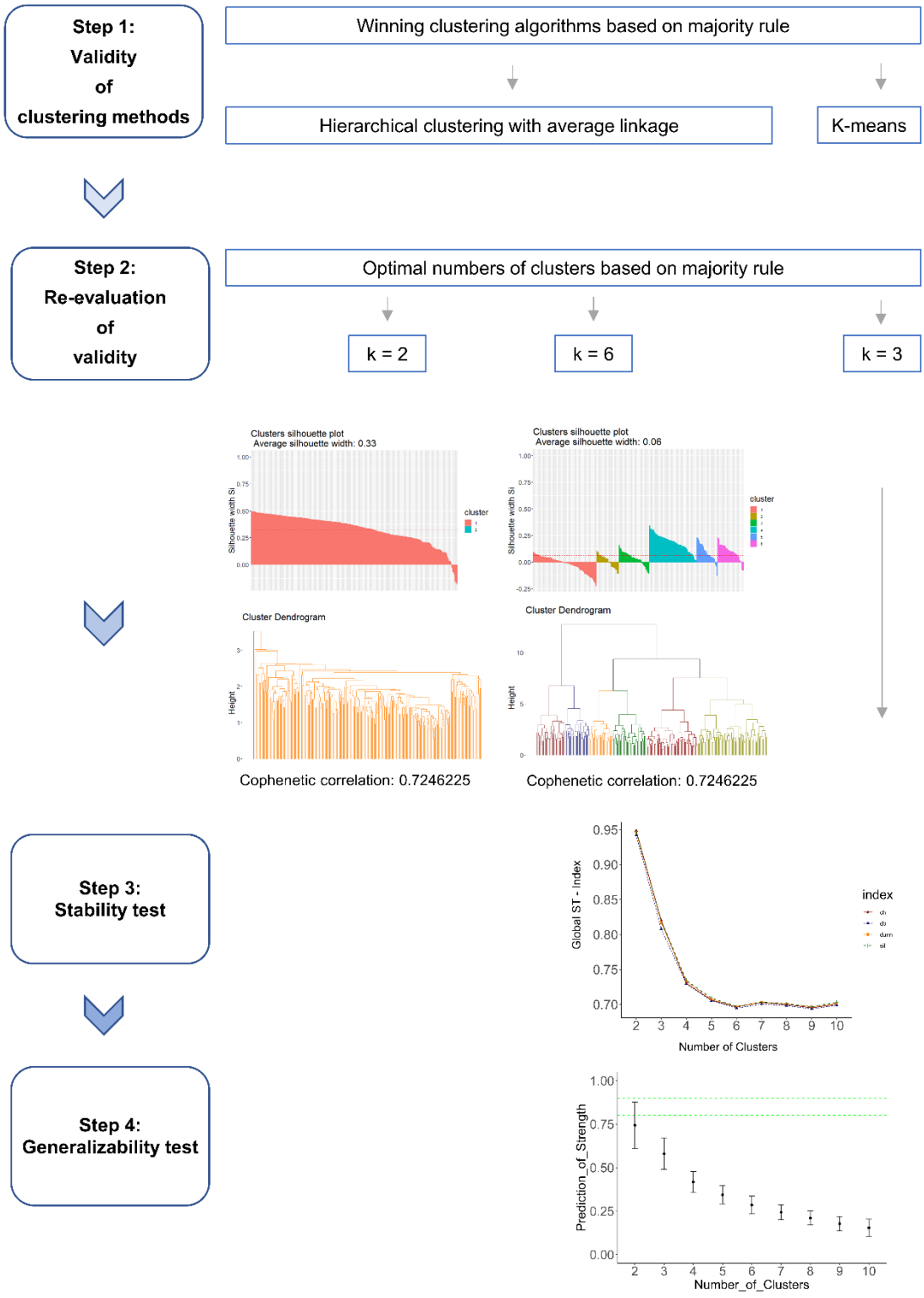
Supplementary Figure 5: Representation of the age and sex distribution of study participants with a recent-onset psychosis from the PRONIA cohort.



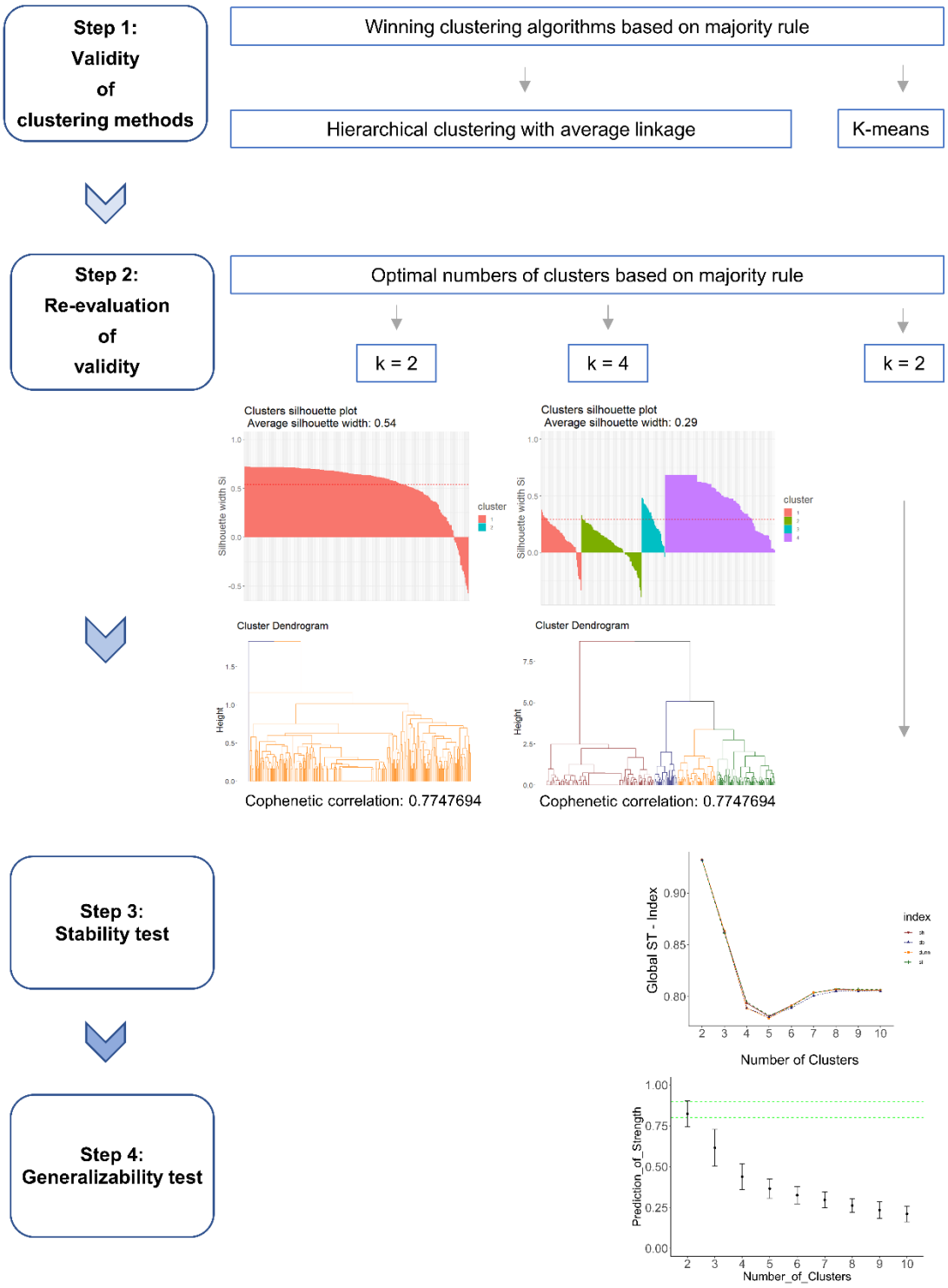
Supplementary Figure 6: Representation of the predicted values of symptom severity in FTD-clusters; results of factorial ANOVA using scaled FTD-related Symptom Severity data.



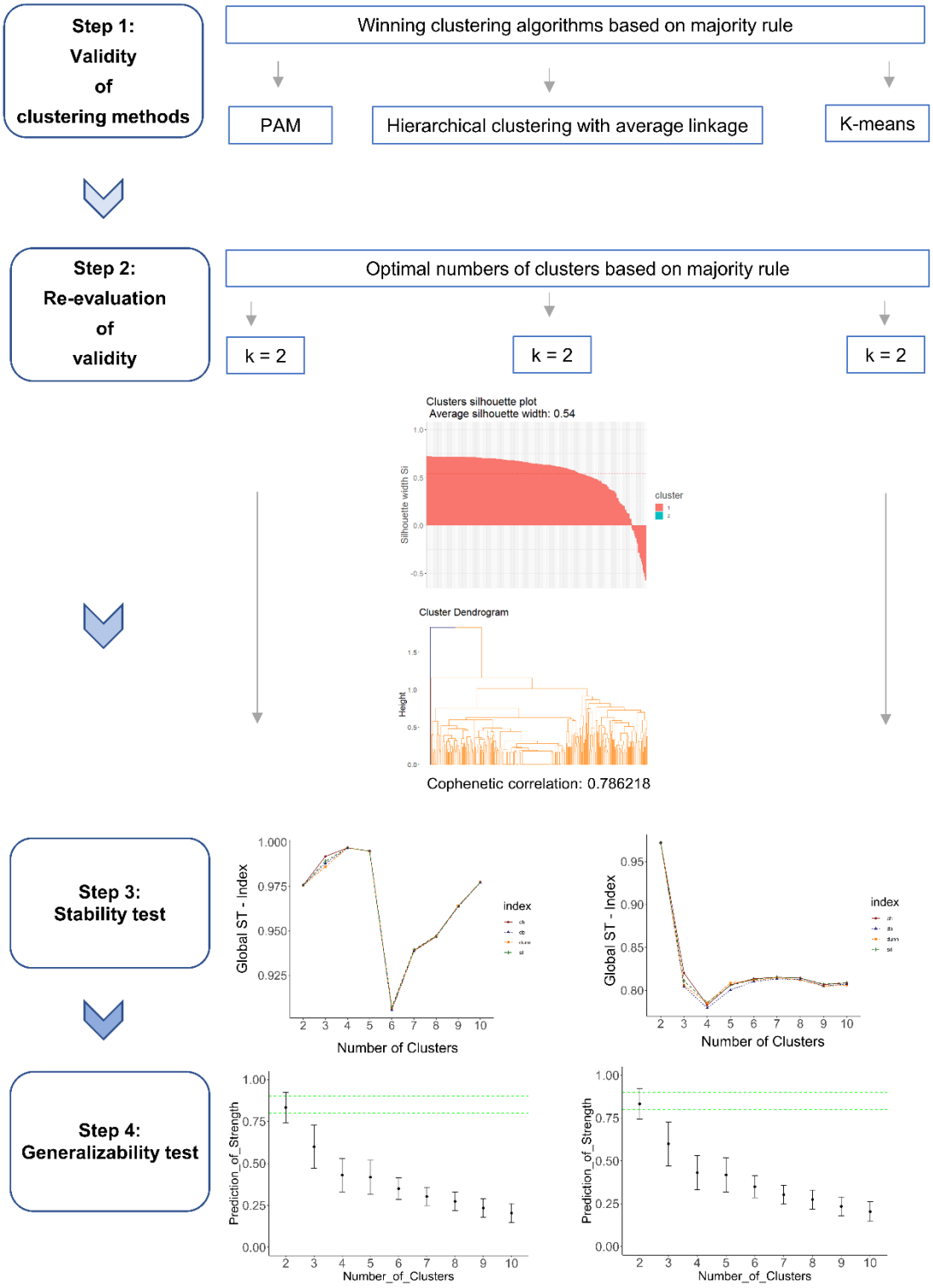
Supplementary Figure 7: Representation of the multi-step clustering solutions using PANSS and SANS items that are not related to FTD as input variables.



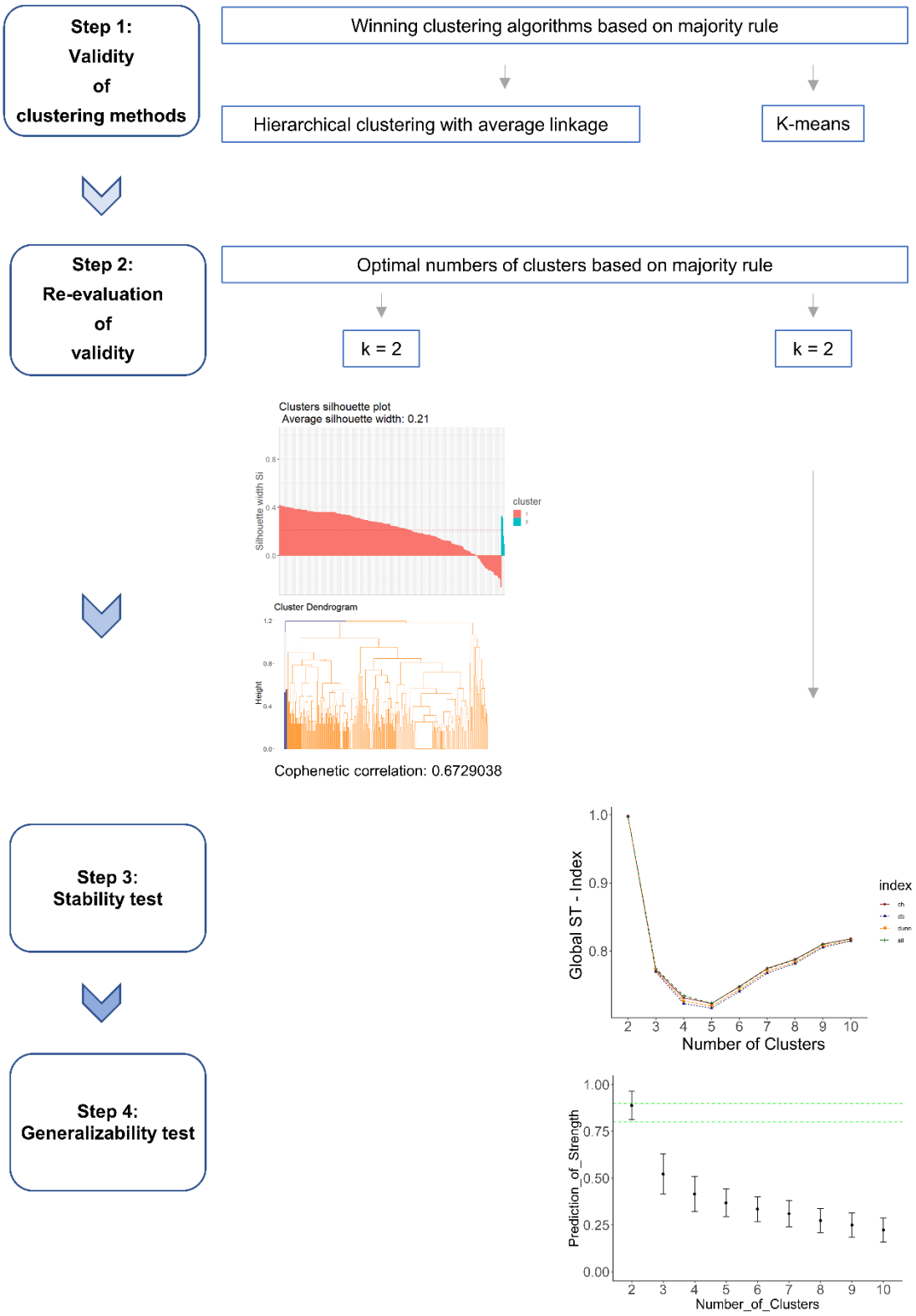
Supplementary Figure 8: Representation of the multi-step clustering solutions using PANSS negative subscale items that are not related to FTD as input variables.



Supplementary Figure 9: Representation of the multi-step clustering solutions using SANS items that are not related to FTD as input variables.



Supplementary Figure 10: Representation of the multi-step clustering solutions using PANSS positive subscale items that are not related to FTD as input variables.



References

Bellani, M., & Brambilla, P. (2008). The use and meaning of the continuous performance test in schizophrenia. *Epidemiology and Psychiatric Sciences*, *17*(3), 188–191.

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140.

Burock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statictical Software*, *25*(4).

Cacciotti-Saija, C., Langdon, R., Ward, P. B., Hickie, I. B., & Guastella, A. J. (2018). Clinical symptoms predict concurrent social and global functioning in an early psychosis sample. *Early Intervention in Psychiatry*, *12*(2), 177–184.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, *3*(1), 1–27.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*.

Cornblatt, B. A., Lenzenweger, M. F., & Erlenmeyer-Kimling, L. (1989). The continuous performance test, identical pairs version: II. Contrasting attentional profiles in schizophrenic and depressed patients. *Psychiatry Research*, *29*(1), 65–85.

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

Daniel, D. G. (2013). Issues in Selection of Instruments to Measure Negative Symptoms. *Schizophrenia Research*, *150*(2–3), 343–345.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *2*, 224–227.

Dunn†, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, *4*(1), 95–104.

Gagnon, M., Awad, N., Mertens, V. B., & Messier, C. (2003). Comparing the Rey and Taylor Complex Figures: A Test-Retest Study in Young and Older Adults. *Journal of Clinical and Experimental Neuropsychology*, *25*(6), 878–890. https://doi.org/10.1076/jcen.25.6.878.16480

Garcia-Portilla, M. P., Garcia-Alvarez, L., Saiz, P. A., Al-Halabi, S., Bobes-Bascaran, M. T., Bascaran, M. T., Muñiz, J., & Bobes, J. (2015). Psychometric evaluation of the negative syndrome of schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *265*(7), 559–566.

Gerritsen, C., Maheandiran, M., Lepock, J., Ahmed, S., Kiang, M., Bagby, R. M., & Mizrahi, R. (2019). Negative symptoms in the clinical high-risk state for psychosis: Connection with cognition and primacy in impacting functioning. *Early Intervention in Psychiatry*, *June*, eip.12843.

Gillett, R. (2007). Assessment of working memory performance in self-ordered selection tests. *Cortex*, *43*(8), 1047–1056.

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *British Journal of Clinical Psychology*, *39*(2), 181–191.

Horton Jr., A. M., & Hartlage, L. C. (1994). The Halstead-Reitan neuropsychology test battery: Theory and clinical interpretation second edition. *Archives of Clinical Neuropsychology*,

9(3), 289–290.

Hubley, A M. (1996). Modification of the Taylor Complex Figure: a comparable figure to the Rey-Osterrieth Figure? Edgeworth Series in Quantitative Behavioural. *Sci., Pap. No. ESQBS*, 96–97.

Hubley, Anita M, & Tremblay, D. (2002). Comparability of Total Score Performance on the Rey–Osterrieth Complex Figure and a Modified Taylor Complex Figure. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 370–382. https://doi.org/10.1076/jcen.24.3.370.984

Jaccard, J., & Jaccard, J. (1998). *Interaction effects in factorial analysis of variance* (Issue 118). Sage.

Keefe, R. S. E., Goldberg, T. E., Harvey, P. D., Gold, J. M., Poe, M. P., & Coughenour, L. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research*, *68*(2–3), 283–297.

Kirkpatrick, B., Fenton, W. S., Carpenter, W. T., & Marder, S. R. (2006). The NIMH-MATRICS consensus statement on negative symptoms. *Schizophrenia Bulletin*, *32*(2), 214–219.

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

Lord, E., Willems, M., Lapointe, F. J., & Makarenkov, V. (2017). Using the stability of objects to

determine the number of clusters in datasets. *Information Sciences*, *393*, 29–46.

McMinn, M. R., Wiens, A. N., & Crossen, J. R. (1988). Rey Auditory-Verbal Learning Test: Development of norms for healthy young adults. *The Clinical Neuropsychologist*, *2*(1), 67–87.

Milner, B., Petrides, M., & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology*, *4*(3), 137–142.

Nowicki Jr, S., & Carton, J. (1993). The measurement of emotional intensity from facial expressions. *The Journal of Social Psychology*, *133*(5), 749–750.

Nowicki, S. (2000). Manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2. *Atlanta, GA: Department of Psychology, Emory University*.

Nowicki, Stephen, & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*(1), 9–35.

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*, *8*(6), 537–548.

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. [Test of copying a complex figure; contribution to the study of perception and memory.]. *Archives de Psychologie*, *30*, 206–356.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal-and temporal-lobe lesions in man. *Neuropsychologia*, *20*(3), 249–262.

Rey, A. (1964). L'Examen Clinique en Psychologie L'Examen Clinique en Psychologie. *Paris: Presses Universitaires de France*.

Roiser, J P, Stephan, K. E., Den Ouden, H. E. M., Barnes, T. R. E., Friston, K. J., & Joyce, E. M. (2009). Do patients with schizophrenia exhibit aberrant salience? *Psychological Medicine*, *39*(2), 199–209.

Roiser, Jonathan P, Howes, O. D., Chaddock, C. A., Joyce, E. M., & McGuire, P. (2012). Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. *Schizophrenia Bulletin*, *39*(6), 1328–1336.

Roiser, Jonathan P, Stephan, K. E., den Ouden, H. E. M., Friston, K. J., & Joyce, E. M. (2010). Adaptive and aberrant reward prediction signals in the human brain. *Neuroimage*, *50*(2), 657–664.

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*(3), 511–528.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*, 498.

Paper II

Biological Psychiatry: Cognitive Neuroscience and Neuroimaging

Structural and functional brain patterns predict formal thought disorder's severity and its persistence in recent-onset psychosis: Results from the PRONIA Study

Madalina-Octavia Buciuman[1,2,3‡]; Oemer Faruk Oeztuerk[1,2,3‡]; David Popovic[1,2,3]; Paolo Enrico[1,20]; Anne Ruef[1]; Nadia Bieler[1]; Elif Sarisik[1]; Johanna Weiske[1]; Mark Sen Dong[1]; Dominic B. Dwyer[1,18]; Lana Kambeitz-Ilankovic[4]; Shalaila S. Haas[5]; Alexandra Stainton[6]; Stephan Ruhrmann[4]; Katharine Chisholm[7]; Joseph Kambeitz[4]; Anita Riecher-Rössler[8]; Rachel Upthegrove[9]; Frauke Schultze-Lutter[10,11,12]; Raimo K. R. Salo-kangas[13]; Jarmo Hietala[13]; Christos Pantelis[14,15]; Rebekka Lencer[16]; Eva Meisenzahl[10]; Stephen J. Wood[6,17,18]; Paolo Brambilla[19,20]; Stefan Borgwardt[21]; Peter Falkai[1,3]; Linda A. Antonucci[1,22,23]; Alessandro Bertolino[22]; Peter Liddle[24]; Nikolaos Koutsouleris[1,3,25, CA] for the PRONIA Consortium[#]

‡ These authors contributed equally to the work

PRONIA consortium is described in the Acknowledgments

[1] Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany
[2] International Max Planck Research School for Translational Psychiatry, Munich, Germany
[3] Max Planck Institute for Psychiatry, Munich, Germany
[4] Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Germany
[5] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[6] Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia
[7] School of Psychology, Aston University, United Kingdom
[8] Faculty of Medicine, University of Basel, Basel, Switzerland
[9] Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom
[10] Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany
[11] Department of Psychology, Faculty of Psychology, Airlangga University, Surabaya, Indonesia
[12] University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland
[13] Department of Psychiatry, University of Turku, Turku, Finland
[14] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne, Carlton South, Vic, Australia
[15] NorthWestern Mental Health, Royal Melbourne Hospital, Parkville, VicMelbourne, Australia
[16] Department of Psychiatry and Psychotherapy, University of Münster, Germany
[17] School of Psychology, University of Birmingham, United Kingdom
[18] Orygen, the National Centre of Excellence for Youth Mental Health, Melbourne, Australia
[19] Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy
[20] Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy
[21] Department of Psychiatry and Psychotherapy, University of Lübeck, Lüebeck, Germany
[22] Department of Basic Medical Sciences, Neuroscience and Sense Organs – University of Bari "Aldo Moro", Bari, Italy
[23] Department of Translational Biomedicine and Neuroscience (DiBraiN) – University of Bari "Aldo Moro", Bari, Italy
[24] Division of Mental Health and Clinical Neuroscience, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom
[25] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

Running title: Neuroimaging predictors of FThD in early psychosis

Keywords: Formal thought disorder, recent-onset psychosis, neuroimaging, predictive modeling, subtyping, early recognition

Received: 22 February 2023 / Accepted: 12 June 2023

Published online: 19 June 2023

Citation: Buciuman, M. O., Oeztuerk, O. F., Popovic, D., Enrico, P., Ruef, A., Bieler, N., ... & PRONIA Consortium. (2023). Structural and functional brain patterns predict formal thought disorder's severity and its persistence in recent-onset psychosis: Results from the PRONIA Study. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 8(12), 1207-1217. https://doi.org/10.1016/j.bpsc.2023.06.001

Structural and functional brain patterns predict formal thought disorder's severity and its persistence in recent-onset psychosis: Results from the PRONIA Study

Madalina-Octavia Buciuman[1,2,3‡]; Oemer Faruk Oeztuerk[1,2,3‡]; David Popovic[1,2,3]; Paolo Enrico[1,20]; Anne Ruef[1]; Nadia Bieler[1]; Elif Sarisik[1]; Johanna Weiske[1]; Mark Sen Dong[1]; Dominic B. Dwyer[1,18]; Lana Kambeitz-Ilankovic[4]; Shalaila S. Haas[5]; Alexandra Stainton[6]; Stephan Ruhrmann[4]; Katharine Chisholm[7]; Joseph Kambeitz[4]; Anita Riecher-Rössler[8]; Rachel Upthegrove[9]; Frauke Schultze-Lutter[10,11,12]; Raimo K. R. Salokangas[13]; Jarmo Hietala[13]; Christos Pantelis[14,15]; Rebekka Lencer[16]; Eva Meisenzahl[10]; Stephen J. Wood[6,17,18]; Paolo Brambilla[19,20]; Stefan Borgwardt[21]; Peter Falkai[1,3]; Linda A. Antonucci[1,22,23]; Alessandro Bertolino[22]; Peter Liddle[24]; Nikolaos Koutsouleris[1,3,25, CA] for the PRONIA Consortium[#]

‡ These authors contributed equally to the work

PRONIA consortium is described in the Acknowledgments

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

[2] International Max Planck Research School for Translational Psychiatry, Munich, Germany

[3] Max Planck Institute for Psychiatry, Munich, Germany

[4] Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Germany

[5] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6] Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

[7] School of Psychology, Aston University, United Kingdom

[8] Faculty of Medicine, University of Basel, Basel, Switzerland

[9] Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom

[10] Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany

[11] Department of Psychology, Faculty of Psychology, Airlangga University, Surabaya, Indonesia

[12] University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

[13] Department of Psychiatry, University of Turku, Turku, Finland

[14] Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne, Carlton South, Vic, Australia

[15] NorthWestern Mental Health, Royal Melbourne Hospital, Parkville, VicMelbourne, Australia

[16] Department of Psychiatry and Psychotherapy, University of Münster, Germany

[17] School of Psychology, University of Birmingham, United Kingdom

[18] Orygen, the National Centre of Excellence for Youth Mental Health, Melbourne, Australia

[19] Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy

[20] Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy

[21] Department of Psychiatry and Psychotherapy, University of Lübeck, Lüebeck, Germany

[22] Department of Basic Medical Sciences, Neuroscience and Sense Organs – University of Bari "Aldo Moro", Bari, Italy

[23] Department of Translational Biomedicine and Neuroscience (DiBraiN) – University of Bari "Aldo Moro", Bari, Italy

[24] Division of Mental Health and Clinical Neuroscience, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[25] Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

[CA] Corresponding author:
Nikolaos Koutsouleris, Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstr. 7, D-80336 Munich, Germany, Tel.: +49 (0) 89 4400, Fax: 0049 (0) 89 4400 55776, E-mail: Nikolaos.koutsouleris@med.uni-muenchen.de.

Running title: Neuroimaging predictors of FThD in early psychosis

Keywords: Formal thought disorder, recent-onset psychosis, neuroimaging, predictive modeling, subtyping, early recognition

Abstract

Background: Formal thought disorder (FThD) is a core feature of psychosis, and its severity and long-term persistence relates to poor clinical outcomes. However, advances in developing early recognition and management tools for FThD are hindered by a lack of insight into the brain-level predictors of FThD states and progression at the individual level.

Methods: 233 individuals with recent-onset psychosis were drawn from the multi-site European Prognostic Tools for Early Psychosis Management study. Support vector machine classifiers were trained within a cross-validation framework to separate two FThD symptom-based subgroups (high vs. low FThD severity), using cross-sectional whole-brain multi-band fractional amplitude of low frequency fluctuations (fALFF), gray-matter volume (GMV) and white-matter volume (WMV) data. Moreover, we trained machine learning models on these neuroimaging readouts to predict the persistence of high FThD subgroup membership from baseline to 1-year follow-up.

Results: Cross-sectionally, multivariate patterns of GMV within the salience, dorsal attention, visual and ventral attention networks separated the FThD severity subgroups (BAC=60.8%). Longitudinally, distributed activations/deactivations within all fALFF sub-bands (BAC_{slow-5}=73.2%, BAC_{slow-4}=72.9%, BAC_{slow-3}=68.0), GMV patterns overlapping with the cross-sectional ones (BAC=62.7%) and smaller frontal WMV (BAC=73.1%) predicted the persistence of high FThD severity from baseline to follow-up, with a combined multi-modal balanced accuracy of BAC=77%.

Conclusions: We report first evidence of brain structural and functional patterns predictive of FThD severity and persistence in early psychosis. These findings open the avenue for the development of neuroimaging-based diagnostic, prognostic and treatment options for the early recognition and management of FThD and associated poor outcomes.

1. Introduction

Early psychosis is characterized by marked symptom and outcome heterogeneity, which poses challenges for the development and implementation of tailored prevention and treatment options. In this context, modern frameworks such as the RDoC (Research Domain Criteria) (1) or HiTOP (The Hierarchical Taxonomy Of Psychopathology) (2) highlight the relevance of symptom-based stratifications for the identification of more homogeneous neuroetiological pathways which could guide personalized early recognition, prevention and treatment.

As a multidimensional psychopathological construct, formal thought disorder (FThD) is a core and dynamically evolving feature of psychotic disorders (3,4). FThD encompasses symptoms of conceptual disorganization such as derailment, incoherence, tangentiality, neologisms, and thought blocking. Importantly, FThD has been associated with long-term adverse outcomes such as increased hospitalization, reduced quality of life, social, and occupational functioning, both in chronic schizophrenia patients (3,5–9), and in patients with recent-onset psychosis (ROP) (10–12). Employing a data-driven methodology for FThD stratification, a recent multi-site study from our group identified two ROP subgroups that differed in their FThD severity and provided support for the association between FThD intensity and functioning impairments, along with neurocognitive deficits (13). Longitudinally, FThD has been shown to exhibit a persistent symptom course in almost 40% of schizophrenia patients (7) and FThD persistence has been transdiagnostically associated with particularly poor clinical outcomes (14). Collectively, these results highlight the prognostic value of FThD symptoms and their persistence for relevant clinical outcomes. However, the currently limited insight into the brain-level alterations that underly FThD at the individual level hinders the development of FThD-tailored prevention and intervention tools.

Regarding brain anatomy, correlational evidence suggests structural gray matter volume (GMV) alterations within frontotemporal language networks associated with FThD in schizophrenia at the cross-sectional level (9,15). Furthermore, FThD sub-dimensions have been differentially related to lower volume in specific brain regions. An extensive review (16) reported that lower volume in six gray matter regions (superior temporal gyrus, orbitofrontal cortex, prefrontal lobe, amygdala-hippocampus, cerebellum vermis, nucleus accumbens) distinctly related to positive, negative, and global FThD dimensions. Additionally, lower white matter volume (WMV) within language-related tracts have been associated with global scores of FThD in schizophrenia (17), and fractional anisotropy disruptions in the right posterior cingulum bundle, inferior longitudinal fasciculus and anterior thalamic radiation were differentially associated with FThD subdomains, such as disorganization and incoherence (18). Despite such FThD-associated structural brain abnormalities reported in patients with chronic schizophrenia, the presence of similar changes underlying or predicting future FThD at first

episode has rarely been studied. Vita et al. (1995) found that distractible speech and illogicality were inversely correlated with left prefrontal lobe volume, whereas incoherence and tangentiality were inversely correlated with left and right lobe volume in younger patients with schizophrenia (19). However, longitudinal approaches aiming to predict the course of FThD based on structural brain abnormalities have not been implemented so far, despite their potential for guiding preventive interventions.

Similarly, functional abnormalities within the language and executive networks have been related to FThD symptoms in schizophrenia using task-based and resting-state fMRI measures (9,20–23), with specificity for different FThD subdomains (24–27). Recent studies have increasingly transitioned from a univariate region-of-interest-based approach to a multivariate framework for analyzing functional MRI, to model the whole-brain functional correlates of FThD dimensions. For instance, Chen et al. (2021) used a whole-brain resting-state functional connectivity approach combined with machine learning algorithms to identify robust clusters associated with FThD dimensions (24). Another promising resting-state fMRI measure is the fractional amplitude of low frequency fluctuations (fALFF), a frequency-domain metric capturing the intensity of spontaneous fluctuations during rest at the voxel-level (28). fALFF is thought to reflect local functional integrity and is closely related to higher-order functional connectivity metrics (29,30). fALFF is commonly measured within multiple frequency sub-bands which seem to originate from distinct functional brain systems, encompassing both spatial and temporal dimensions of fMRI data (31–33). fALFF alterations have been consistently found in schizophrenia patients (33,34), and could therefore provide a promising measure for capturing complex patterns underlying FThD symptoms.

In summary, literature suggests that structure-function interdependencies underly FThD symptoms and calls for a better understanding of the predictive value of such changes for the development and persistence of FThD. Therefore, understanding potential neurobiological predictors of FThD severity and persistence at the individual level represents an important, yet uncharted research area, which could lead to a better characterization of the pathways underlying symptom heterogeneity in early psychosis. Our current study aimed to address this gap by investigating (I) whether the two ROP FThD subgroups previously identified by our group can be cross-sectionally differentiated based on whole-brain voxel-level gray matter volume, white matter volume, multi-band functional resting-state fALFF, and combined brain patterns, and (II) whether the persistence of high-FThD severity after 1-year of follow-up can be predicted by the same baseline neuroimaging modalities.

2. Methods

2.1. Study sample

Participants with ROP were recruited within a multisite, longitudinal study (PRONIA – Prognostic Tools for Early Psychosis Management, https://www.pronia.eu/), which included ten sites across five European countries (study protocol presented in Text S1 and Figure S1). General PRONIA inclusion and exclusion criteria are presented in Table S1. ROP patients fulfilled the DSM-IV-TR criteria for a lifetime affective and non-affective psychotic episode, having the psychotic episode within the past 3 months, with the onset of the psychotic episode occurring within the past 24 months. Being treated with antipsychotic medication for longer than 90 days at or above the minimum dosage of the first-episode psychosis range of the German Association for Psychiatry and Psychotherapy (DGPPN) S3 guideline (35) was a specific exclusion criterion. For the current analyses, we excluded 46 out of the 279 available ROP patients included in our previous clustering study (13) due to missing or low-quality structural or functional MRI scans, leading to the inclusion of 233 ROP patients (47,6% female, average age = 24.6 (SD = 5.6)) in the current study (exclusion steps are presented in Figure S1A). Additionally, for the persistence analyses, we additionally excluded 80 participants due to missing clinical follow-up data required for FThD subgroup assignment, leaving 153 ROP patients in the persistence analyses (43% female, average age = 24 (SD = 5.5)) (Figure S1A). The distribution of the ROP patients across diagnosis categories based on the *ICD-10 classification of mental and behavioural disorders* by *World Health Organization* (WHO) is provided in Figure S2.

Adult participants gave written informed consent, while patients younger than 18 years-of-age and their guardians provided written informed assent and consent, respectively. The study was registered at the German Clinical Trials Register (DRKS00005042) and approved by the local research ethics committees of all sites.

2.2. Formal thought disorder subgroup definition and assignment

The ROP patients were assigned to either a high or a low FThD subgroup based on a clustering analysis of FThD symptoms described in our previous work (13). In summary, the authors used 5 items from the Positive and Negative Symptom Scale (*PANSS)*(36) conceptual disorganization (P2); difficulty in abstract thinking (N5)) and the Scale for the Assessment of Negative Symptoms (*SANS)*(37); poverty of speech (9), poverty of content speech (10), increased latency of response (12)) in order to operationalize FThD as proposed elsewhere (3). Based on this, the authors used an R-based clustering validation package (38) to derive a two-cluster k-means-based solution corresponding to high and low FThD subgroups (13).

In the current study, we used a subsample of the original ROP sample used by Oeztuerk et

al., 2021 (13), according to MRI data availability (Text S1; Figure S1B). Cross-sectionally, we classified the high vs low FThD subgroups based on their structural and functional imaging data, directly using the subgroup assignment of the patients from our previous study (13). The same clustering model was used to assign individuals to the FThD subgroups at their 1-year follow-up study visit, using the squared Euclidean distance from each cluster center in the k-means based clustering model. This allowed us to identify and classify individuals with persistent high FThD severity based on the original clustering solution of the previous study (13), without building any new clustering models.

The two baseline FThD subgroups, as well as the persistent vs non-persistent high FThD symptom (at follow-up) groups were compared in terms of clinical and socio-demographic characteristics.

2.3. MRI data acquisition and processing

2.3.1. Structural MRI data

The MRI data was minimally harmonized between the different PRONIA sites to preserve the heterogeneity of MRI data acquisition present in the clinical context (39)(Text S2). The acquisition parameters used at each PRONIA site are presented in Table S2.

Images were pre-processed using a standardized pipeline implemented in the Statistical Parametric Mapping software (SPM12, http://www.fil.ion.ucl.ac.uk/spm)-based CAT12 toolbox (version r1155; http://dbm.neuro.uni-jena.de/cat12/). Briefly, the scans underwent segmentation (into white matter, gray matter, and cerebrospinal fluid) and registration to the *Montreal Neurological Institute* (MNI-152) space based on the DARTEL algorithm. More details on the CAT12 preprocessing procedures are provided in the Text S2 and in previous studies from our group (39,40). Following high-dimensional registration, the gray matter and white matter maps were modulated using the Jacobian determinants obtained during the registration to produce GMV and WMV maps, respectively. The volumetric images were resliced to a resolution of 3 mm^3 and corrected for total intracranial volume using global scaling. Images with a quality score lower or equal to C (satisfactory) based on the CAT12 Quality Assurance framework were excluded from further analyses (detailed in Text S2).

2.3.2. Resting-state fMRI data

The rs-fMRI data was collected using Echo Planar Imaging (EPI) sequences, a repetition time of 3 seconds and a total of 200 volumes across all sites (acquisition parameters per site are detailed in Table S3).

Data preprocessing was conducted using a pipeline previously developed in the PRONIA consortium (41), mainly based on SPM12 and the Resting State fMRI data analysis Toolkit (REST, version 1.848; http://www.restfmri.net/) (Text S2). In brief, the first 8 volumes were discarded, images were slice-time corrected and realigned to the first volume. The functional maps were co-registered to the T1 images, resliced and normalized to the common MNI space. Then, nuisance covariates were regressed out, including white matter, cerebrospinal fluid and Friston 24 motion parameters. Lastly, the images were smoothed, motion-corrected using time-series despiking and detrended (detailed in Text S2). Images with more than 38.5% of volumes with a mean framewise displacement over 0.50 mm were discarded (42).

The fALFF was computed as the ratio of the integrated power spectrum of the frequency range of interest to that of the entire frequency range, after time series were transformed to the frequency domain using a fast Fourier transform. Based on previous literature showing different properties and generators of the activity in different low-frequency sub-bands (31,43), we chose three different sub-bands: slow-5 (0.01 – 0.027 Hz), slow-4 (0.027 – 0.073 Hz) and slow-3 (0.073 – 0.198 Hz). The fALFF maps were z-score standardized voxel-wise within the gray matter mask for each participant.

2.4. Machine learning procedure

All machine learning models were built using the MATLAB-based open-source toolbox NeuroMiner (version 1.1; http://www.pronia.eu/neurominer), within a nested repeated cross-validation framework as detailed in Text S3. Based on the baseline structural and functional MRI data, supervised classifiers were trained to separate patients with high and low FThD cross-sectionally, as well as patients with persistent high-FThD (high-FThD group membership both at baseline and the 1-year follow-up) from those with any other FThD symptom courses.

Both the structural and functional MRI data were preprocessed within an outer-leave-site-out/ inner-pooled nested cross-validation structure for the cross-sectional models, and a pooled nested cross-validation structure for the persistence models, given the insufficient number of persistent cases available across different sites. Preprocessing steps included scaling to a range of 0 to 1, regression of covariates such as age, sex and symptom severity (*PANSS* total score), modality-specific site correction (based on generalizability theory-based voxel-level reliability maps (44) (Figure S3) and site correction of the voxel-level data based on the site means as detailed in Text S3), dimensionality reduction using principal component analysis, and standardization of the obtained components. Following preprocessing, we trained binary L2-regularized support vector machine classifiers with a Maximum Relevance Minimum Redundancy filter for feature selection to discriminate between the high and low FThD groups at baseline and between persistent vs. non-persistent high-FThD based on: 1) each of the

three fALFF sub-bands separately (slow-5, slow-4, slow-3), 2) the GMV maps, and 3) the WMV maps. Lastly, we trained stacked multimodal SVM models on the classifier scores derived from the models trained on the individual data modalities that performed above chance level, to assess the combined discriminative performance of structural and functional brain images. The models' significance was assessed using a label permutation approach and the obtained *P*-values were corrected using the false discovery rate to account for the multiple classifiers trained. The predictive features were visualized using a combination of feature importance (sign-based consistency) and feature stability (cross-validation ratio) metrics, as detailed in Text S3. Additionally, the performance differences between single-modality classifiers and the stacked multimodal models were tested for statistical significance by means of Quade tests followed by post-hoc pairwise classifier comparisons (45).

Post-hoc, we investigated the associations between the decision scores of significant classifiers using Pearson's correlation coefficients. Moreover, we explored the potential influence of medication on the models by computing Pearson's correlation coefficients between the significant classifiers' decision scores and antipsychotic medication converted into chlorpromazine equivalents, as well as cumulative dosages of selective serotonin reuptake inhibitors (SSRI).

Supplementary analyses were conducted to evaluate possible methodological moderators of classification performance: 1) all main models were similarly trained on the non-standardized fALFF data; 2) all main models were trained without correcting for total symptomatology (*PANSS* total score); 3) FThD persistence prediction models were trained within an outer leave-site-out cross-validation framework despite the small sample size to further evaluate the impact of site effects; 4) potential differential relationships between positive and negative FThD subscores and the models' decision scores were evaluated using slope interaction analyses (Figure S4).

3. Results

3.1. Sociodemographic and clinical characteristics of the FThD subgroups

Sociodemographic and clinical comparisons of the two baseline FThD-based ROP subgroups, as well as between the patients with persistent and non-persistent high FThD are presented in Table 1. Briefly, high FThD patients at baseline were younger than those with low FThD (t(df)=2.46(231), p=.02), were more likely to have a stable partnership ($\chi 2$(df)=8.49(6), p=.02), but were similar in terms of other sociodemographic characteristics (Table 1). High FThD patients had higher levels of positive *PANSS* score (t(df)=-4.76(230), p<.001), negative *PANSS* score (t(df)=-10.93(230), p<.001), and *PANSS* general symptoms score (t(df)=-

4.09(226), p<.001). Lastly, high FThD patients had higher scores on all *SANS* subscales (Blunting: t(df)=-8.82(231), p<.001; Alogia: t(df)=-18.52(231), p<.001; Apathy: t(df)=-4.67(231), p<.001; Anhedonia: t(df)=-4.34(222), p<.001; Attention: t(df)=-6.85 205), p<.001)(Table 1).

Furthermore, patients with persistently high FThD were younger (t(df)=-2.05(151), p=.04), had fewer education years (t(df)=-3.87(151), p<.001), and were less likely to have worked in the past year ($\chi 2$(df)=-3.87(151), p=.003). Also, persistently high FThD patients had higher levels of general (t(df)=2.80(151), p=.006) and negative *PANSS* symptoms (t(df)=5.10(151), p<.001) and higher levels of affective blunting (t(df)=4.07(151), p<.001), alogia (t(df)=5.05(151), p<.001), apathy (t(df)=2.14(151), p=.03) and attention deficits (t(df)=4.45(134), p<.001) according to the *SANS* (Table 1).

3.2. Cross-sectional analyses

At baseline, high and low FThD subgroups could only be separated above chance level using the GMV data (Balanced Accuracy (BAC)=60.8%, Sensitivity=63.8%, Specificity=57.7%, p_{FDR}=.002, Table 2). The GMV pattern predicted high FThD group membership based on higher GMV in cingulate cortex regions pertaining to the prefrontal control network and dorsal attentional network as well as lower GMV in the visual network (Figures 1 and 2).

3.3. FThD persistence analyses

All imaging data modalities performed significantly above chance level in separating patients with persistently high FThD from those with other FThD symptom courses (slow-5 fALFF: BAC=73.2%, Sensitivity=83.3%, Specificity=63.1%, p_{FDR}<.001; slow-4 fALFF: BAC=72.3%, Sensitivity=83.3%, Specificity=62.4%, p_{FDR}<.001; slow-3 fALFF: BAC=68.0%, Sensitivity=75.0%, Specificity=61.0%, p_{FDR}<.001; GMV: BAC=62.7%, Sensitivity=75.0%, Specificity=50.4%, p_{FDR}=.048; WMV: BAC=73.1%, Sensitivity=91.7%, Specificity=54.6%, p_{FDR}<.001). Stacking all data modalities generated a performance of BAC=77.0% (Sensitivity=100%, Specificity=53.9%, p_{FDR}=.048), which was significantly higher than that of the GMV-based model, but not that of the fALFF- or WMV-based models, as evidenced by a Quade test followed by post-hoc comparisons (Table S4). Moreover, supplementary leave-site-out analyses showed that the slow-5-, slow-3- and WMV-based models remained statistically significant with a maximal multimodal BAC of the stacked model of 69% (Table S5).

fALFF predictive patterns of high FThD persistence consisted of distributed deactivations and activations within large-scale brain networks, such as the default-mode network, dorsal attention network, and salience network, as well as the cerebellum, with specificity for frequency sub-bands (Figure 2, Figure 3A,B,C). Moreover, high FThD persistence was

predicted by higher GMV within the dorsal attention and salience networks and lower GMV within regions of the ventral attention and visual networks (Figures 2 and 3D). Lastly, lower WMV within frontal tracts and higher WMV within subcortical tracts predicted the persistence of high FThD severity (Figure 3E).

3.4. Post-hoc analyses

Correlation analyses showed that the decision scores of the cross-sectional GMV-based model were positively correlated with those of the slow-4-based model predicting high FThD persistence ($r(151)$=.20, p_{FDR}<.05), as well as with those of the GMV- ($r(151)$=.60, p_{FDR}<.001) and WMV-based ($r(151)$=.30, p_{FDR}<.001) FThD persistence prediction models (Table S6). Moreover, we found significant positive correlations between the decisions scores of the persistence models trained on the different modalities (Table S6).

Furthermore, there were no significant correlations between antipsychotic medication and the decision scores of any of the significant cross-sectional models (GMV: $r(231)$=-0.09, p=.26) or any of the persistence classifiers (Slow-5: $r(151)$=-.001, p=.98; Slow-4: $r(151)$=-.19, p=.06; Slow-3: $r(151)$=-.11, p=.26; GMV: $r(151)$=-.01, p=.98; WMV: $r(151)$=-.15, p=.18). Similarly, no significant correlations were found between cumulative dosage of SSRI medication and the decision scores of the cross-sectional GMV model (r(231)=-0.12, p=.071). None of the patients included in the persistence analyses had been medicated with SSRIs.

Using the non-standardized fALFF did not have an influence on which models performed above chance-level but led to performance decreases of the FThD persistence prediction models (Table S7). Furthermore, removing the *PANSS* total score correction from the pipeline additionally led to the slow-5 fALFF sub-band being cross-sectionally predictive of high vs low FThD at baseline, without influencing the other significant results of the *PANSS*-corrected models (Table S8). Lastly, the slope interaction analyses showed similar relationships between the decision scores of all significant models and the positive and negative FThD subdimension scores (Figure S4).

4. Discussion

The current study provides preliminary evidence that FThD severity in recent-onset psychosis (ROP) can be cross-sectionally delineated and longitudinally predicted by multivariate structural and functional brain patterns spanning large-scale brain networks. Specifically, we show that GMV differences cross-sectionally separate between high vs. low FThD ROP subgroups with a relatively low BAC of 60.8%. Furthermore, the baseline activity in all fALFF sub-bands, as well as GMV and WMV patterns predicted the persistence of high FThD severity over one year with BACs of up to 77%.

The cross-sectional results describing the separability of FThD severity subgroups based on GMV data are in line with univariate correlational studies showing associations between FThD symptoms and GMV alterations in schizophrenia (9). The higher GMV within fronto-cingulate regions and lower GMV localized within occipital regions, predictive of high FThD in our study, are consistent with previously findings in patients with schizophrenia (46,47). However, in contrast to other studies (9,15), we did not find lower GMV within the language networks to be particularly predictive of high-FThD in our sample. In addition to the multivariate methodology employed here, these inconsistencies could highlight the specificity of the predictive patterns reported here for delineating FThD heterogeneity in the early stages of psychosis, where the influence of general illness chronicity and medication effects are more limited. Our cross-sectional findings provide preliminary evidence for an FThD-discriminative pattern of limited clinical utility. Future studies showing cross-sectional classifications with higher specificity and sensitivity are needed to reach individual level clinical implications.

In contrast to the limited cross-sectional findings, we found that all structural and functional neuroimaging modalities investigated were predictive of 1-year high-FThD persistence. Predictive volumetric brain patterns included lower/higher GMV that partially overlapped with those found to be predictive of high-FThD at the cross-sectional level, as well as lower WMV most pronounced within frontal tracts. Furthermore, distributed functional alterations within the fALFF sub-bands across large-scale brain networks were predictive of high-FThD persistence. These functional patterns overlapped with those identified in previous studies, highlighting alterations of language, executive and default-mode networks as a possible substrate for positive (i.e., conceptual disorganization) and negative FThD symptoms (i.e., alogia, difficulty in abstract thinking)(20,24,25,27). Together with the WMV findings, our results suggest that connectivity patterns within and between long-range brain networks may represent relevant early biomarkers of FThD symptom progression, in line with the *disconnection hypothesis* of schizophrenia (48). Future studies may benefit from exploring the prognostic value of additional measures such as dynamic functional connectivity and diffusion tensor imaging, which provide more fine-grained information on the function and structure of these neural systems.

The current study addresses an important gap in the FThD stratification literature, providing a novel exploration of the predictive value of multi-modal neuroimaging data for disentangling the heterogeneity of FThD severity and persistence in ROP using machine learning methods operating in a multi-site context. Collectively, our findings support the utility of symptom-based hierarchical approaches in psychiatric nosology (2) for identifying specific brain systems associated with psychopathological dimensions such as FThD, with noteworthy implications for personalized diagnosis and intervention. Firstly, the predictive patterns reported provide

exploratory pathophysiological evidence linking specific brain networks to FThD severity and persistence. Following validation in future clinical studies, such imaging-based biomarkers could translate into targets for novel treatments (e.g. brain stimulation-enhanced neurocognitive training), tailored to patients suffering from severe FThD. Secondly, we provide preliminary evidence for the prognostic utility of neuroimaging measures for the stratification of FThD persistence, which has been associated with poor clinical outcomes (7). Although the small sample size available for this analysis calls for future validation studies, our results encourage the future development of neuroimaging-based prognostic tools for the early recognition of patients at-risk for progressing to chronic FThD syndromes, as a first step towards the indicated prevention of such debilitating disease trajectories.

In addition to the small sample size and lack of an external replication sample, which limit claims regarding the out-of-sample generalizability of present findings, several limitations of our study are noteworthy. Firstly, we did not use a construct-specific FThD scale, but instead used FThD subdomains corresponding to the *Thought and Language Disorder (TALD*)(3) scale (conceptual disorganization, poverty of content of speech, difficulty in abstract thinking, poverty of speech and increased latency of response), as measured by *PANSS* and *SANS* items. Although having an FThD-proxy within widely used clinical scales can facilitate clinical translation, further studies are needed to investigate FThD more broadly through validated scales (i.e. *TALD* or speech sample analyses(49)), to further evaluate the specificity of our findings for the FThD construct. Secondly, regarding FThD progression, we restricted our analyses to persistently high FThD symptom courses. The chronicity of FThD symptoms in comparison to single-time measurements has been associated with particularly poor clinical outcomes (7) and can be regarded as a more reliable prognostic marker of FThD severity for preventive clinical applications. This is further supported by our models, which performed with higher sensitivity in predicting high FThD persistence compared to single timepoint severity. Nonetheless, future studies may benefit from a manifold investigation of FThD longitudinal dynamics which may unravel symptom course patterns delineated by distinct structural and functional brain alterations.

6. Conclusion

Our findings align with modern symptom-based psychiatric frameworks by providing first evidence for multivariate structural and functional brain alterations within large-scale brain networks predictive of FThD severity and its persistence in recent-onset psychosis. These preliminary results open the avenue for the development of refined early spectrum-wise diagnostics with preventive and long-term clinical outcome-oriented therapeutics, guided by and involving integrated neuroimaging data domains within larger multi-site samples.

Acknowledgements

Disclosures of Conflicts of Interest

Dr Bertolino reports speaker fees from Otsuka, Lundbeck, Angelini and Rovi outside of the submitted work. Dr Haas has received support by the NIH National Institute of Mental Health, grant T32MH122394, unrelated to the submitted work. Dr Hietala reports personal fees from Orion ltd, personal fees from Lundbeck, personal fees from Otsuka and other from Takeda during the conduct of the study. Dr Koutsouleris, Dr Ruhrmann, Dr Riecher-Rossler report grants from European Union over the duration of the study. Dr Meisenzahl and Dr Koutsouleris hold patent US20160192889A1 ('Adaptive pattern recognition for psychosis risk modelling'). Dr Koutsouleris reports speaker fees from Otsuka, Roche and Angelini outside of the submitted work. Dr Pantelis reports grants from Australian NHMRC during the study, and personal fees from Lundbeck, Australia Pty Ltd outside the submitted work. Dr Upthegrove reports speaker fees from Sunovion, Otsuka and Vitaris outside the submitted work as well as unpaid officership with the British Association for Pharmacology - Honorary General Secretary 2021-2024. She serves as Deputy Editor for The British Journal of Psychiatry. Dr Falkai reports he has received research support/honoraria for lectures or advisory activities from Boehringer-Ingelheim, Janssen, Lundbeck, Otsuka, Recordati and Richter outside the submitted work. Dr Pantelis was supported by an Australian National Health and Medical Research Council (NHMRC) L3 Investigator Grant (1196508) outside the submitted work. Dr Lana Kambeitz-Ilankovic reports receiving a NARSAD Young Investigator Award of the Brain & Behavior Research Foundation No◦ 28474 (PI: LK-I) outside the submitted work.

All other co-authors did not report any other financial disclosures or other conflicts of interests within or outside of the scope of the submitted work.

Funding Sources

PRONIA is a Collaboration Project funded by the European Union under the 7th Framework Programme under grant agreement n° 602152. Oemer Faruk Oeztuerk and David Popovic were supported by the Else-Kröner-Fresenius-Foundation through the Clinician Scientist Program 'EKFS-Translational Psychiatry'.

The PRONIA consortium

The following authors participated in the screening, recruitment, rating, examination, and follow-up of the study participants used in the current analyses. The authors are listed in alphabetical order according to the institution of affiliation.

Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany: Shalaila Haas, Alkomiet Hasan, Claudius Hoff, Ifrah Khanyaree, Camilla Krämer, Aylin Melo, Susanna Muckenhuber-Sternbauer, Yanis Köhler, Oemer Faruk Oeztuerk, Nora Penzel, David Popovic, Adrian Rangnick, Sebastian von Saldern, Rachele Sanfelici, Moritz Spangemacher, Ana Tupac, Maria Fernanda Urquijo, Johanna Weiske, Antonia Wosgien.

Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany: Karsten Blume, Dennis Hedderich, Dominika Julkowski, Nathalie Kaiser, Thorsten Lichtenstein, Ruth Milz, Alexandra Nikolaides, Tanja Pilgram, Mauro Seves, Martina Wassen.

Department of Psychiatry (Psychiatric University Hospital, UPK), University of Basel, Switzerland: Christina Andreou, Laura Egloff, Fabienne Harrisberger, Ulrike Heitz, Claudia Lenz, Letizia Leanza, Amatya Mackintosh, Renata Smieskova, Erich Studerus, Anna Walter, Sonja Widmayer.

Institute of Mental Health & School of Psychology, University of Birmingham, United Kingdom: Chris Day, Sian Lowri Griffiths, Mariam Iqbal, Mirabel Pelton, Pavan Mallikarjun, Alexandra Stainton, Ashleigh Lin.

Department of Psychiatry, University of Turku, Finland: Alexander Denissoff, Anu Ellilä, Tiina From, Markus Heinimaa, Tuula Ilonen, Päivi Jalo, Heikki Laurikainen, Antti Luutonen, Akseli Mäkela, Janina Paju, Henri Pesonen, Reetta-Liina Säilä, Anna Toivonen, Otto Turtonen.

Department of Psychiatry (Psychiatric University Hospital LVR/HHU Düsseldorf), University of Düsseldorf, Germany: Sonja Botterweck, Norman Kluthausen, Gerald Antoch, Julian Caspers, Hans-Jörg Wittsack.

Department of Basic Medical Science, Neuroscience and Sense Organs - University of Bari Aldo Moro: Giuseppe Blasi, Giulio Pergola, Grazia Caforio, Leonardo Fazio, Tiziana Quarto, Barbara Gelao, Raffaella Romano, Ileana Andriola, Andrea Falsetti, Marina Barone, Roberta Passiatore, Marina Sangiuliano.
Department of Psychiatry and Psychotherapy of the University of Münster, Germany: Marian Surmann, Olga Bienek, Udo Dannlowski.
General Electric Global Research Inc., USA.
Ana Beatriz Solana, Manuela Abraham, Timo Schirmer.
Workgroup of Paolo Brambilla, University of Milan, Italy:
Department of Neuroscience and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy: Carlo Altamura, Marika Belleri, Francesca Bottinelli, Adele Ferro, Marta Re
Programma2000, Niguarda Hospital, Milan: Emiliano Monzani, Maurizio Sberna
San Paolo Hospital, Milan: Armando D'Agostino, Lorenzo Del Fabro
Villa San Benedetto Menni, Albese con Cassano (CO): Giampaolo Perna, Maria Nobile, Alessandra Alciati
Workgroup of Paolo Brambilla at the University of Udine, Italy.
Department of Medical Area, University of Udine, Udine, Italy: Matteo Balestrieri, Carolina Bonivento, Giuseppe Cabras, Franco Fabbro.
IRCCS Scientific Institute "E. Medea", Polo FVG, Udine: Marco Garzitto, Sara Piccin.

References

1. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Vol. 167, American Journal of psychiatry. Am Psychiatric Assoc; 2010. p. 748–51.
2. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. J Abnorm Psychol. 2017;126(4):454.
3. Kircher T, Krug A, Stratmann M, Ghazi S, Schales C, Frauenheim M, et al. A rating scale for the assessment of objective and subjective formal thought and language disorder (TALD). Schizophr Res. 2014;160(1–3):216–21.
4. Roche E, Creed L, Macmahon D, Brennan D, Clarke M. The Epidemiology and Associated Phenomenology of Formal Thought Disorder: A Systematic Review. Schizophr Bull. 2015;41(4):951–62.
5. Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. Schizophr Bull. 1986;12(3):348–59.
6. Hart M, Lewine RRJ. Rethinking thought disorder. Schizophr Bull. 2017;43(3):514–22.
7. Marengo JT, Harrow M. Schizophrenic Thought Disorder at Follow-up. Arch Gen Psychiatry. 1987;44.
8. Sigaudo M, Crivelli B, Castagna F, Giugiario M, Mingrone C, Montemagni C, et al. Quality of life in stable schizophrenia: The relative contributions of disorganization and cognitive dysfunction. Schizophr Res. 2014;153(1–3):196–203.
9. Cavelti M, Kircher T, Nagels A, Strik W, Homan P. Is formal thought disorder in schizophrenia related to structural and functional aberrations in the language network? A systematic review of neuroimaging findings. Schizophr Res. 2018;199:2–16.
10. 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. Vol. 272, European Archives of Psychiatry and Clinical Neuroscience. Springer Science and Business Media Deutschland GmbH; 2022. p. 381–93.
11. Remberk B, Namysłowska I, Rybakowski F. Clinical and cognitive correlates of formal thought disorder in early onset schizophrenia. Activitas Nervosa Superior Rediviva. 2012;54(2):68–76.
12. Roche E, Lyne J, O'Donoghue B, Segurado R, Behan C, Renwick L, et al. The prognostic value of formal thought disorder following first episode psychosis. Schizophr Res. 2016;178(1–3):29–34.
13. Oeztuerk OF, Pigoni A, Wenzel J, Haas SS, Popovic D, Ruef A, et al. The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study. Eur Arch Psychiatry Clin Neurosci. 2021 Apr 1;
14. Harrow M, Marengo JT. Schizophrenic Thought Disorder at Followup: Its Persistence and Prognostic Significance. Schizophr Bull. 1986 Jan 1;12(3):373–93.
15. Sans-Sansa B, McKenna PJ, Canales-Rodríguez EJ, Ortiz-Gil J, López-Araquistain L, Sarró S, et al. Association of formal thought disorder in schizophrenia with structural brain abnormalities in language-related cortical regions. Schizophr Res. 2013;146(1–3):308–13.
16. Sumner PJ, Bell IH, Rossell SL. A systematic review of the structural neuroimaging correlates of thought disorder. Neurosci Biobehav Rev. 2018;84:299–315.

17. Cavelti M, Winkelbeiner S, Federspiel A, Walther S, Stegmayer K, Giezendanner S, et al. Formal thought disorder is related to aberrations in language-related white matter tracts in patients with schizophrenia. Psychiatry Res Neuroimaging. 2018;279(November 2017):40–50.
18. Stein F, Buckenmayer E, Brosch K, Meller T, Schmitt S, Ringwald KG, et al. Dimensions of Formal Thought Disorder and Their Relation to Gray- and White Matter Brain Structure in Affective and Psychotic Disorders. Schizophr Bull. 2022;1–10.
19. Vita A, Dieci M, Giobbio GM, Caputo A, Ghiringhelli L, Comazzi M, et al. Language and thought disorder in schizophrenia: brain morphological correlates. Schizophr Res. 1995;15(3):243–51.
20. Horn H, Jann K, Federspiel A, Walther S, Wiest R, Müller T, et al. Semantic network disconnection in formal thought disorder. Neuropsychobiology. 2012 Jul;66(1):14–23.
21. Kircher TTJ, Bulimore ET, Brammer MJ, Williams SCR, Broome MR, Murray RM, et al. Differential activation of temporal cortex during sentence completion in schizophrenic patients with and without formal thought disorder [Internet]. 2001. Available from: www.elsevier.com/locate/schres
22. Sumner PJ, Bell IH, Rossell SL. A systematic review of task-based functional neuroimaging studies investigating language, semantic and executive processes in thought disorder. Neurosci Biobehav Rev. 2018;94:59–75.
23. Wensing T, Cieslik EC, Müller VI, Hoffstaedter F, Eickhoff SB, Nickl-Jockschat T. Neural correlates of formal thought disorder: An activation likelihood estimation meta-analysis. Hum Brain Mapp. 2017;38(10):4946–65.
24. Chen J, Wensing T, Hoffstaedter F, Cieslik EC, Müller VI, Patil KR, et al. Neurobiological substrates of the positive formal thought disorder in schizophrenia revealed by seed connectome-based predictive modeling. Neuroimage Clin. 2021 Jan 1;30.
25. Fuentes-Claramonte P, López-Araquistain L, Sarró S, Sans-Sansa B, Ortiz-Gil J, Maristany T, et al. Brain functional correlates of formal thought disorder in schizophrenia: Examining the frontal/dysexecutive hypothesis. Psychol Med. 2021;51(14):2446–53.
26. Kircher TTJ, Liddle ; P F, Brammer MJ, Williams SCR, Murray RM, Mcguire PK. Neural Correlates of Formal Thought Disorder in Schizophrenia Preliminary Findings From a Functional Magnetic Resonance Imaging Study. 2001.
27. Winkelbeiner S, Cavelti M, Federspiel A, Kunzelmann K, Dierks T, Strik W, et al. Decreased blood flow in the right insula and middle temporal gyrus predicts negative formal thought disorder in schizophrenia. Schizophr Res. 2018;9–11.
28. Zang YF, Yong H, Chao-Zhe Z, Qing-Jiu C, Man-Qiu S, Meng L, et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 2007 Mar;29(2):83–91.
29. Fu Z, Tu Y, Di X, Du Y, Pearlson GD, Turner JA, et al. Characterizing dynamic amplitude of low-frequency fluctuation and its relationship with dynamic functional connectivity: An application to schizophrenia. Vol. 180, NeuroImage. Academic Press Inc.; 2018. p. 619–31.
30. Tomasi D, Shokri-Kojori E, Volkow ND. Temporal changes in local functional connectivity density reflect the temporal variability of the amplitude of low frequency fluctuations in gray matter. PLoS One. 2016 Apr 1;11(4).
31. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. 1995.

32. Liang X, Wang J, Yan C, Shu N, Xu K, Gong G, et al. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: A resting-state functional MRI study. PLoS One. 2012 Mar 6;7(3).
33. Yu R, Chien YL, Wang HLS, Liu CM, Liu CC, Hwang TJ, et al. Frequency-specific alternations in the amplitude of low-frequency fluctuations in schizophrenia. Hum Brain Mapp. 2014 Feb;35(2):627–37.
34. Hoptman MJ, Zuo XN, Butler PD, Javitt DC, D'Angelo D, Mauro CJ, et al. Amplitude of low-frequency oscillations in schizophrenia: A resting state fMRI study. Schizophr Res. 2010 Mar;117(1):13–20.
35. Falkai P, Gaebel W, Weinmann S, Wobrock T. DGPPN (Hrsg.) S3–Praxisleitlinien in Psychiatrie und Psychotherapie Band 1: Behandlungsleitlinie Schizophrenie. 2006.
36. Kay, S R Fiszbein A L a O. Positive and Negative Syndrome Scale (Panss) Rating Criteria. Schizophr Bull [Internet]. 1987;13(2):261–76. Available from: www.ncbi.nlm.nih.gov
37. Andreasen NC. The scale for the assessment of negative symptoms (SANS). British Journal of Psychiatry. 1989;155:49–52.
38. Burock G, Pihur V, Datta S, Datta S. clValid: An R Package for Cluster Validation. Journal of Statictical Software. 2008;25(4).
39. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry. 2018;75(11):1156–72.
40. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients with Clinical High-Risk Syndromes and Recent-Onset Depression. JAMA Psychiatry. 2021 Feb 1;78(2):195–209.
41. Haas SS, Antonucci LA, Wenzel J, Ruef A, Biagianti B, Paolini M, et al. A multivariate neuromonitoring approach to neuroplasticity-based computerized cognitive training in recent onset psychosis. Neuropsychopharmacology. 2021 Mar 1;46(4):828–35.
42. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage. 2014;84:320–41.
43. Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, et al. The oscillating brain: complex and reliable. Neuroimage. 2010;49(2):1432–45.
44. Cronbach LJ, Rajaratnam N, Gleser GC. Theory of generalizability: A liberalization of reliability theory. British Journal of Statistical Psychology. 1963;16(2):137–63.
45. Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. J Am Stat Assoc. 1979;74(367):680–3.
46. Palaniyappan L, Mahmood J, Balain V, Mougin O, Gowland PA, Liddle PF. Structural correlates of formal thought disorder in schizophrenia: An ultra-high field multivariate morphometry study. Schizophr Res. 2015 Oct 1;168(1–2):305–12.
47. Stein F, Meller T, Brosch K, Schmitt S, Ringwald K, Pfarr JK, et al. Psychopathological Syndromes across Affective and Psychotic Disorders Correlate with Gray Matter Volumes. Schizophr Bull. 2021;47(6):1740–50.
48. Friston KJ. The disconnection hypothesis. Schizophr Res. 1998;30(2):115–25.
49. Çokal D, Sevilla G, Jones WS, Zimmerer V, Deamer F, Douglas M, et al. The language profile of formal thought disorder. NPJ Schizophr. 2018 Dec 1;4(1).
50. Beck AT, Steer RA, Brown G. Beck depression inventory–II. Psychol Assess. 1996;

51. Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;

Table 1. Sociodemographic and clinical comparison of the high vs low FThD subgroups at baseline and of the persistence-based subgroups.

	Baseline subgroup membership				Baseline to follow-up high FThD persistence vs. other FThD courses			
	High FThD Subgroup	Low FThD Subgroup	t/χ2 value (df)	*P*-value	Persistent High FThD	Non-persistent High FThD	t/χ2 value (df)	*P*-value
Socio-demographic characteristics								
Participants, N	58	175			12	141		
Participants per site, N (%)								
Munich	17 (29.3)	66 (37.7)	1.34 (4)	.25	4 (33.3)	47 (33.3)	0.00 (4)	1.0
Turku	7 (12.1)	31 (17.7)	1.02 (4)	.31	2 (16.7)	23 (16.3)	0.00 (4)	.98
Milan	10 (17.2)	14 (8.0)	4.03 (4)	**.05**	4 (33.3)	13 (9.2)	6.51 (4)	**.01**
Birmingham	2 (3.4)	8 (4.6)	0.13 (4)	.71	0 (0.0)	8 (5.7)	0.72 (4)	.40
Basel	4 (6.9)	18 (10.3)	0.59 (4)	.44	0 (0.0)	17 (12.1)	1.63 (4)	.20
Udine	2 (3.4)	7 (4.0)	0.04 (4)	.85	0 (0.0)	6 (4.3)	0.53 (4)	.47
Cologne	13 (22.4)	25 (14.3)	2.11 (4)	.15	0 (0.0)	23 (16.3)	2.30 (4)	.13
Münster	3 (5.2)	6 (3.4)	0.36 (4)	.55	2 (16.7)	4 (2.8)	5.61 (4)	**.02**
Age, yrs., mean (SD)	23.9 (5.5)	26.0 (5.7)	2.46 (231)	**.02**	22.3 (5.2)	25.8 (5.7)	-2.05 (151)	**.04**
Male sex, N (%)	36 (62.1)	96 (54.9)	0.92 (4)	.34	6 (50.0)	81 (57.4)	0.25 (4)	.61
Education yrs., mean (SD)	12.9 (3.1)	14.9 (9.0)	1.65 (231,0)	.10	10.6 (2.3)	14.2 (3.2)	-3.87 (151)	**<.001**
Steady partnership/ married, N (%)	53 (91.4)	129 (73.7)	8.49 (6)	**.01**	11 (91.7)	109 (77.3)	1.39 (6)	.49
Work in the last year, N (%)	17 (31.5)	70 (41.4)	3.37 (6)	.19	3 (25.0)	58 (41.7)	11.51 (6)	**.003**
First-degree relatives with psychosis, N (%)	1.3 (1.8)	1.1 (1.6)	-1.08 (231)	.28	2 (16.7)	15 (10.6)	1.0 (16)	.99
Birth Complications, N (%)	12 (20.7)	29 (16.7)	0.48 (4)	.49	2 (16.7)	21 (15.0)	0.0 (4)	.88
Psychopharma-cological medication, N (%)	53 (91.4)	148 (85.1)	1.50 (4)	.22	11 (91.7)	120 (85.7)	0.33 (4)	.56
Chlorpromazine-equivalent antipsychotic dosages (mg), mean(SD)	650.8 (4082.5)	1225.5 (3932.4)	-0.94 (231,0)	.35	264.1 (207.6)	807.4 (4586.1)	-0.41 (151)	.68
Clinical characteristics								
Beck Depression Inventory, mean (SD)	18.9 (10.8)	20.4 (12.2)	0.75 (199)	.46	16.2 (13.4)	19.7 (11.5)	-0.92 (135)	.36
PANSS General, mean (SD)	42.4 (10.4)	32.2 (9.0)	-7.21 (230)	**<.001**	42.6 (8.1)	34.0 (10.3)	2.80 (151)	**.006**
PANSS Negative, mean (SD)	23.5 (8.0)	13.3 (5.5)	-10.93 (230)	**<.001**	24.4 (4.9)	14.4 (6.6)	5.10 (151)	**<.001**
PANSS Positive, mean (SD)	21.7 (6.1)	17.4 (5.8)	-4.76 (230)	**<.001**	19.6 (5.2)	18.6 (6.5)	0.49 (151)	.62
SANS Blunting, mean (SD)	17.4 (9.7)	6.3 (7.8)	-8.82 (231)	**<.001**	19.0 (7.5)	7.8 (9.3)	4.07(151)	**<.001**
SANS Alogia, mean (SD)	12.4 (5.7)	1.9 (2.8)	-18.52 (231)	**<.001**	10.8 (5.6)	3.4 (4.8)	5.05 (151)	**<.001**
SANS Apathy, mean (SD)	9.6 (4.8)	6.2 (4.8)	-4.67 (231)	**<.001**	9.4 (5.1)	6.2 (5.0)	2.14 (151)	**.03**

SANS Anhedonia, mean (SD)	13.6 (7.5)	8.6 (7.3)	-4.34 (222)	**<.001**	11.2 (6.9)	8.8 (7.6)	1.03 (143)	.31
SANS Attention, mean (SD)	6.5 (4.1)	2.5 (3.3)	-6.85 (205)	**<.001**	7.5 (2.8)	2.6 (3.5)	4.45 (134)	**<.001**

Note. BDI – Beck Depression Inventory (50); PANSS – Positive and Negative Schizophrenia Symptoms; SANS – Scale for the Assessment of Negative Symptoms.

*Statistically significant *P*-values (at an $\alpha = .05$) are presented in bold.

Table 2. Performance metrics for all cross-sectional and persistence models predicting high vs low FThD subgroup membership at baseline and differentiating between persistent vs non-persistent high FThD from baseline to follow-up based on multiband fALFF data, gray matter and white matter volume.

Cross-sectional high vs. low FThD subgroup classification								
	Leave-site-out nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**P_{FDR} value**
Rs-fMRI slow-5	63.8	48.6	56.2	0.63	29.1	80.2	9.3	.053
Rs-fMRI slow-4	53.4	*55.4*	54.4	0.57	28.4	78.2	6.7	.056
Rs-fMRI slow-3	67.2	45.7	56.5	0.63	29.1	80.8	9.9	.053
GMV	63.8	57.7	60.8	0.66	33.3	82.8	16.1	**.01**
WMV	56.9	51.4	54.2	0.63	28.0	78.3	6.2	.19
Prediction of high FThD persistence (from baseline to 1-year follow-up)								
	Pooled nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**P_{FDR} value**
Rs-fMRI slow-5	83.3	63.1	73.2	0.78	16.1	97.8	13.9	**<.001**
Rs-fMRI slow-4	83.3	62.4	72.9	0.83	15.9	97.8	13.7	**<.001**
Rs-fMRI slow-3	75.0	61.0	68.0	0.76	14.1	96.6	10.7	**<.001**
GMV	75.0	50.4	62.7	0.76	11.4	95.9	7.3	**.048**
WMV	91.7	54.6	73.1	0.75	14.7	98.7	13.4	**<.001**
Stacked model – all modalities	100.0	53.9	77.0	0.87	15.6	100.0	15.6	**.048**

Note. P-values were adjusted using the false-discovery rate (FDR) correction for multiple comparisons and the values exceeding the threshold for statistical significance based on the permutation testing procedure employed (α = .05) are presented in bold.

AUC – area under the curve, BAC – Balanced Accuracy, PPV – positive predictive value, NPV – negative predictive value, PSI – prognostic summary index

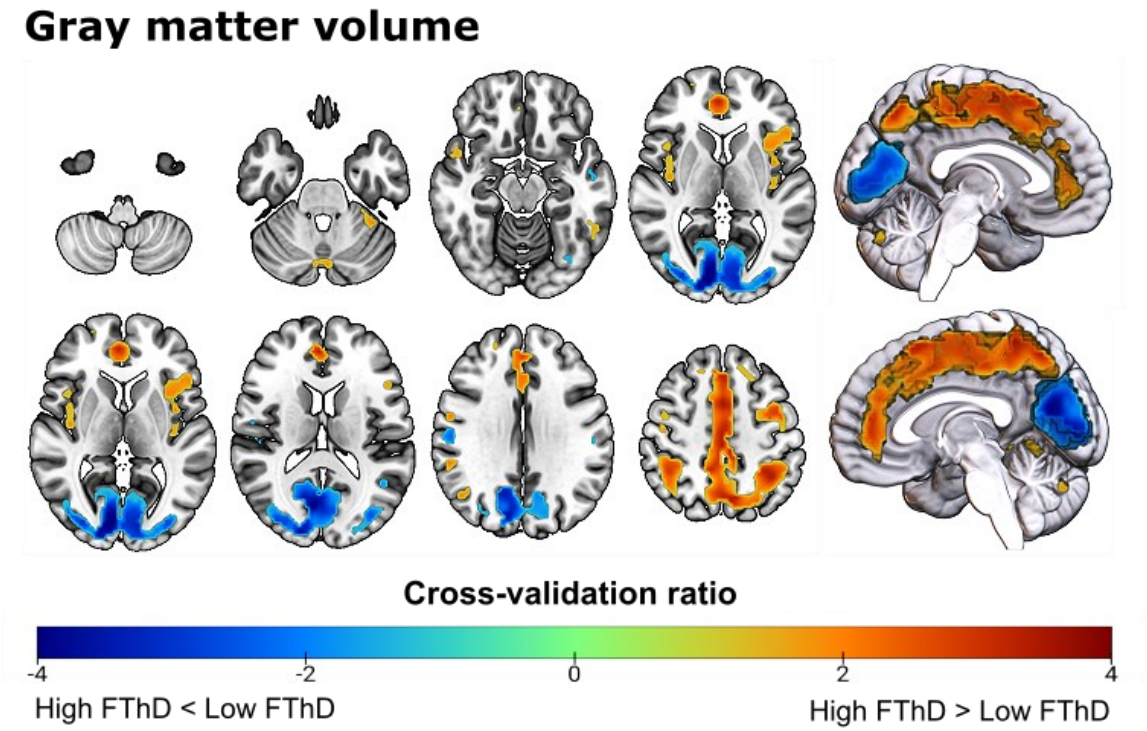


Figure 1. Reliable features for predicting membership to the high vs low formal thought disorder clusters at baseline based on GMV data aquired at the same study visit. The patterns of the other data modalities explored (fALFF sub-bands and WMV) are not presented here, as the models performed at chance level. The reliability of the features is displayed using a grand mean cross-validation ratio, thresholded based on FDR-corrected sign-based consistency maps at α=.05 (detailed in the Supplement). Warm colors represent voxels with higher gray-matter volume for individuals in the high formal thought disorder subgroup, while cold colors indicate lower volume for this subgroup.

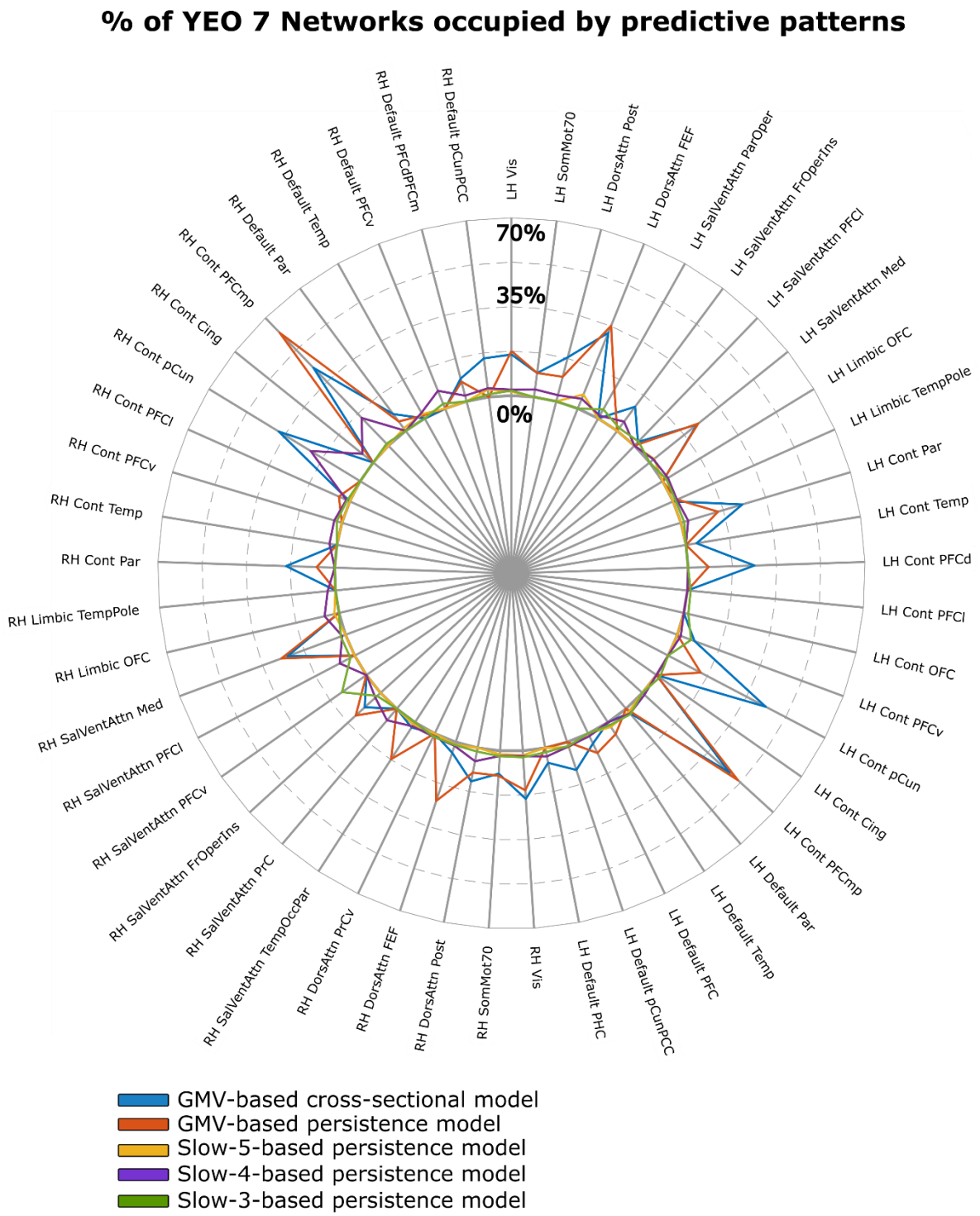


Figure 2. Percentage of the YEO 7 resting-state networks' parcellations (51) **occupied by the predictive patterns of the significant cross-sectional and longitudinal fALFF- and GMV-based models.** Voxels were considered to be predictive based on the derived sign-based consistency metric of feature importance (detailed in the Supplementary methods) thresholded at 1.3, corresponding to $\alpha = .05$. Only YEO resting-state network parcellations with at least 1% occupation by any of the models' predictive patterns were included in the figure.

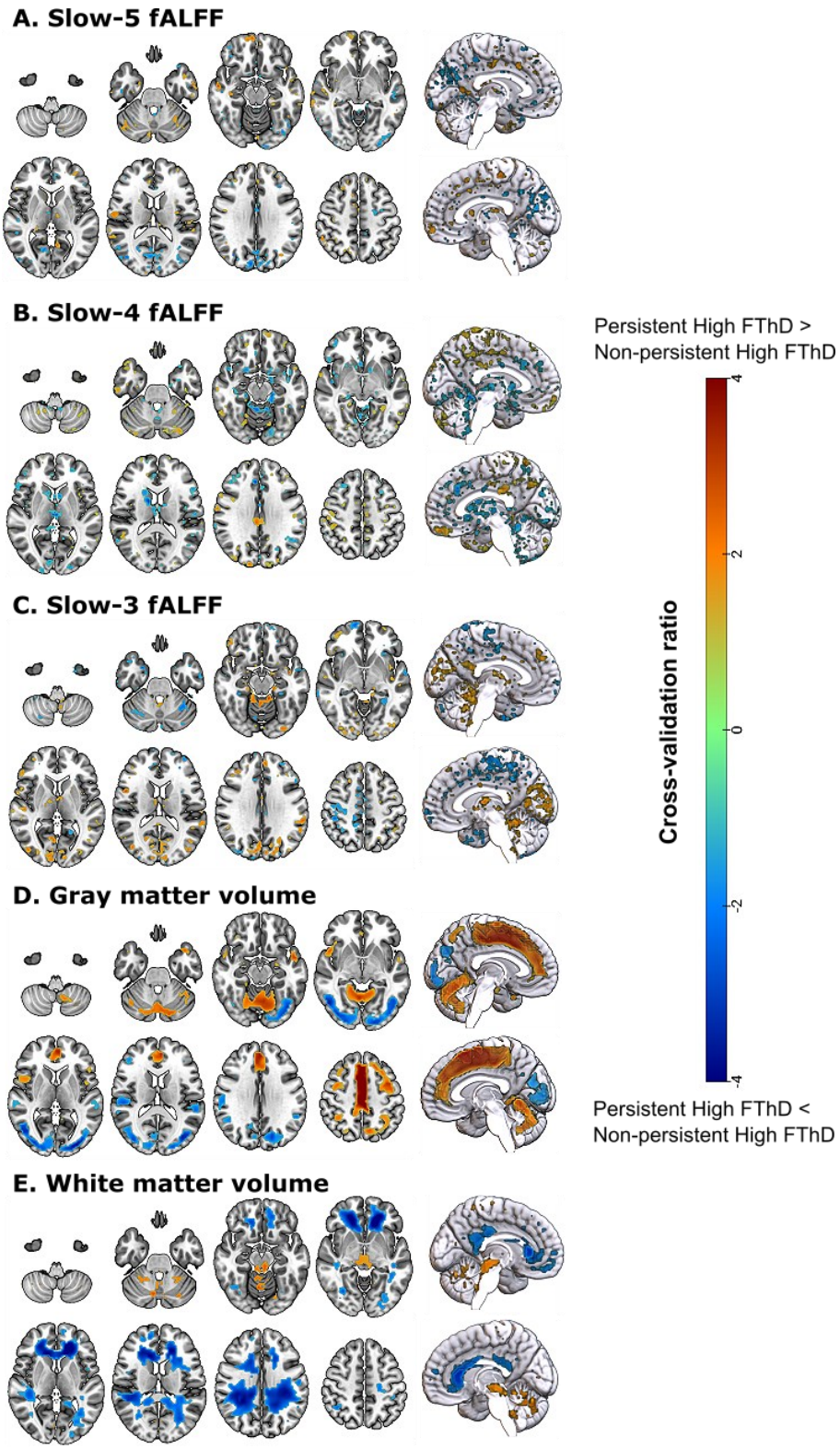


Figure 3. Reliable features for predicting the persistence of high FThD symptomatology from baseline to follow-up relative to other symptom courses based on **A.** slow-5 fALFF data, **B.** slow-4 fALFF data, **C.** slow-3 fALFF data, **D.** gray matter volume data, and **E.** white matter volume data. The reliability of the features is displayed using a grand mean cross-validation ratio, thresholded based on *FDR*-corrected sign-based consistency maps at α=.05 (detailed in Text S3). Warm colors represent voxels with higher activity/volume for individuals with persistently high FThD from baseline to follow-up, while cold colors indicate lower activity/volume for this subgroup.

SUPPLEMENTARY INFORMATION

Structural and Functional Brain Patterns Predict Formal Thought Disorder Severity and Its Persistence in Recent-Onset Psychosis: Results From the PRONIA Study

Buciuman *et al.*

Table of Contents

Abbreviations

Abbreviation	Description
CV	Cross-validation
fALFF	Fractional amplitude of low frequency fluctuations
FThD	Formal thought disorder
G-theory	Generalizability theory
GMV	Gray matter volume
PANSS	Positive and Negative Symptom Scale (1)
PRONIA	Prognostic Tools for Early Psychosis Management Study
ROP	Recent-onset psychosis
SANS	Scale for the Assessment of Negative Symptoms (2)
SVM	Support vector machine
WMV	White matter volume

Supplementary Texts

Text S1. PRONIA consortium sample and study protocol

The follow-up scheme and data modalities collected at each timepoint of the longitudinal multisite, multi-modal PRONIA consortium study (Prognostic Tools for Early Psychosis Management, https://www.pronia.eu/) have been presented in previous studies of our group (3,4) and are graphically depicted in Figure S1A.

Regarding the PRONIA sample, a number of 5547 individuals covering a catchment population of over 5 million Europeans were screened for study eligibility between 2/15/2014 and 5/1/2017 initially across 7 European sites from Germany (München), Italy (Udine), Finland (Turku), Switzerland (Basel), and the United Kingdom (Birmingham). The total PRONIA study sample was split into a discovery/model training sample of individuals which have been followed up for at least 18 months and a replication/model validation sample which also included people coming from 3 subsequently added PRONIA sites (Münster, Bari, Düsseldorf). For the current analysis we only included participants coming from one of the newly added PRONIA replication sites (Münster, Germany), due to the high amount of missing social functioning follow-up data for the other two sites. Additionally, we combined the original discovery and replication PRONIA samples in order to maximize our sample size, given the reduced number of subjects with high FThD severity and persistence.
Regarding exclusion criteria, from the 409 ROP patients initially included based on general PRONIA study inclusion criteria (Table S1), 279 ROP patients had the clinical and demographic data required complete and were included in our previous clustering paper (5). Additionally, for the current analyses we additionally excluded 8 participants coming from the Bari and Düsseldorf sites due to insufficient healthy controls for site calibration coming from these sites, 14 participants due to missing structural or functional MRI scans, and 24 participants due to insufficient rs-fMRI data quality (Text S2). Additionally, a sample of 96 healthy controls matched for age, sex and education years were used for computing the means for the calibration of site effects in the rs-fMRI data (Text S2) and a larger sample of 465 available healthy controls was used for computing beta coefficients for correcting the effects of age, sex and mean framewise displacement the patient data, while preserving potential clinically relevant interactions between these variables and our outcome measure.

Text S2. MRI data acquisition and processing

Neuroimaging data acquisition within the PRONIA study followed a minimal harmonization procedure in order to retain the heterogeneity inherent in clinical settings. T1-weighted structural sequence parameters that were required for all sites included the acquisition of data

at a resolution of 1 mm^3, a field of view offering a full 3D coverage of the brain and maximizing the contrast between the cortical ribbon and the white matter, as well as enhancing the signal-to-noise ratio by means of the relaxation time, echo time and other relevant parameters. Aquisition parameters used at each of the PRONIA sites are presented in Table S2 for the structural MRI data and Table S3 for the resting-state functional MRI data.

After visual inspection and automatic defacing and anonymization using the Freesurfer toolbox, the T1-weighted structural MRI images were pre-processed using the default pipeline implemented in the SPM12-based open-source CAT12 toolbox (version r1207; http://dbm.neuro.unijena.de/cat12/). All the preprocessing steps included in the pipeline are explained in detail in the CAT12 manual (http://www.neuro.unijena.de/cat12/CAT12-Manual.pdf). In brief, these included: 1) Spatially Adaptive Non-Local Means (SANLM) filtering (6) as a first denoising step, 2) data segmentation using the Adaptive Maximum A Posteriori (AMAP) segmentation procedure, which models local variations of intensity distributions as slowly varying spatial functions and leads to a homogeneous segmentation across cortical and subcortical regions, 3) application of a Markov Random Field technique incorporating prior spatial information of adjacent voxels into the segmentation estimation produced by AMAP as a second denoising step, 4) Local Adaptive Segmentation (LAS) in order to adjust the scans for white matter (WM) inhomogeneities and varying gray matter (GM) intensities due to differing iron content in cortical and subcortical areas, 5) the tissue segments produced by AMAP were additionally processed using a partial volume segmentation algorithm modeling tissues with intensities between GM and WM, as well as between GM and cerebrospinal fluid (CSF), and 6) scans were registered to the Montreal Neurological Institute (MNI) template based on 555 healthy controls from the IXI database (http://www.braindevelopment.org) using a high-dimensional DARTEL.

Registered GM and WM segments were then modulated by multiplying them with the Jacobian determinants obtained during registration to produce GM volume (GMV) and, respectively, WM volume (WMV) maps. Then, the GMV and WMV maps were smoothed with an 8 mm full-width-at-half maximum (FWHM) Gaussian kernel. Lastly, images with a weighted quality score equal to C (satisfactory) or lower as determined by the CAT12 Quality Assurance framework were eliminated. The weighted overall image quality (IQR) combines measurements of noise contrast ratio and spatial resolution of the images before pre-processing and is automatically computed by the CAT12 pipeline using the tissue segmentation, as explained in more detail in the CAT12 manual (http://www.neuro.unijena.de/cat12/CAT12-Manual.pdf).

The resting-state fMRI data was preprocessed according to a pipeline previously developed within the PRONIA consortium and also detailed in previous studies from the consortium (7). The pipeline was mainly based on the Statistical Parametric Mapping software (SPM, version 12-6685; http://www.fil.ion.ucl.ac.uk/spm) and the REsting State fMRI data analysis Toolkit (REST, version 1.8; (8); http://www.restfmri.net/). First preprocessing steps included discarding the first 8 volumes, slice-time correction of the remaining 192 volumes and realignment of the functional images to the first volume. Then, the translations in each direction and the rotations in angular motion about each axis was computed for each scan in order to get the time course of head motion. The framewise displacement parameters (FD) were calculated for each subject by setting the FD for the first volume of a run to zero and participants with more than 38.5% of volumes with a mean FD of more than 0.50 mm were excluded from further analyses based on previous guidelines (9). Subsequently, the images were co-registered to the T1 scans in individual native space, resliced using 4th-degree B-Spline interpolation and normalized to the MNI space using the deformation fields generated by CAT12 for the T1-weighted scans. Gray matter (>0.2), white matter (>0.2) and cerebrospinal fluid (>0.05) masks were then created and Friston 24 motion parameters (six motion parameters, six temporal derivatives, six quadratic terms and six quadratic expressions of the derivatives of motion estimates) (10) were computed. Then, individual signal estimates with WM and CSF regressed were created and the images were gray-matter masked and smoothed using a 6-mm full width at half-maximum Gaussian kernel. Motion correction was additionally conducted based on a time-series despiking procedure (Wavelet Despiking; BrainWavelet Toolbox, http://www.brainwavelet.org/) and the nuisance variables previously computed were regressed out (Friston 24 motion parameters, WM and CSF residuals). Linear trend removal was then performed as a last preprocessing step.

Subsequently, the fractional amplitude of low frequency fluctuations (fALFF) was calculated by means of the algorithm implemented in the REST toolbox, such that a fast Fourier transformation (FFT) was applied on the time series and the fALFF was measured as the ratio of the integrated power spectrum of the interest low-frequency range to that of the entire frequency range. Based on previous literature highlighting distinct fALFF sub-bands (11,12), we used three different frequencies of interest: 1) slow-5: 0.01 – 0.027 Hz, 2) slow-4: 0.027 – 0.078 Hz, and 3) slow-3: 0.078 – 0.198 Hz. Finally, the fALFF maps were Z-standardized at voxel level (13).

Text S3. Machine learning pipeline

Machine learning models were build using a MATLAB-based open-source toolbox developed by our group (NeuroMiner, version 1.1; http://www.pronia.eu/neurominer). For the cross-

sectional models, we used a nested cross-validated approach with a leave-site-out approach on the outer cycle (8 folds corresponding to the different sites and 10 permutations) in order to improve generalizability across different sites and a pooled approach on the inner cycle (5 folds, 10 permutations) for hyperparameter optimization. Given the reduced sample size and unavailability of persistent cases at multiple sites, we employed a pooled nested cross-validation structure (5 folds, 10 permutations both on the inner and outer cycle) for the FThD persistence prediction models.

Neuroimaging data preprocessing was embedded within the CV structure and included regressing out the effects of nuisance covariates (age, sex, and general illness effects as measured by the total *PANSS* score for both structural and functional MRI, and additionally mean framewise displacement for the functional data), correction of site effects, PCA-based dimensionality reduction by retaining 40%-60%-80% of the total signal energy (optimized within the CV structure), median-based standardization of the obtained components and winsorization of outliers (± 4 standard deviations). The correction of age and sex has been conducted using beta coefficients computed within a larger sample of healthy controls (N = 465), in order to preserve potentially relevant clinical variance associated with these variables within the ROP sample and only eliminate purely nuisance effects. Furthermore, the GM and WM volume data were divided by the total intracranial volume and subsequently scaled (shifting the range to [0,1] without modifying the shape of the distribution at voxel-level) prior to the other preprocessing steps.The site correction methods employed differed for the structural and functional data, based on previous analyses of the efficiency of different correction methods within the PRONIA consortium. Specifically, as the fALFF data was found to be particularly prone to site effects, we first used a global mean adjustment technique, by subtracting the voxel-wise mean value of each site from the overall mean and then subtracting this mean difference from each voxel of participants coming from the other sites. The correction of global site effects for the fALFF data was computed based on a sample of healthy controls matched for age, sex and education year across sites (N = 96). Following this, we used a reliability map derived based on Generalizability theory (G-theory (14)) from the data of 6 traveler subjects available in the PRONIA sample (healthy controls scanned at 6 from the 7 initial PRONIA sites) in order to disentangle and quantify the variance associated with site effects. The reliability maps generated for each data modality are presented in Figure S3 of this supplement. Based on this reliability map, only the most reliable 25% - 50% - 75% voxels (optimized within the CV structure) across sites were retained for model training. For the GMV and WMV data, we only employed the G-theory-based reliability map correction, shown to be adequate for accounting for site effects for structural MRI data (3,4).

Following preprocessing, separate L2-regularized support vector machine classifiers were trained to discriminate high from low FThD individuals at baseline, as well as individuals with persistently high FThD from baseline to the 1-year follow-up from those with different FThD symptom courses, based on each of the three fALFF sub-bands of interest (slow-5, slow-4, slow-3) and based on the GMV and WMV data. The L2-regularized SVM algorithm used was based on the *LIBLINEAR* library (https://www.csie.ntu.edu.tw/~cjlin/liblinear/) with a tolerance of 0.01 and we optimized the C parameter in the [0.015625, 0.0625, 0.25, 1, 4, 16] range, using hyperplane weighting in order to account for unbalanced classes. Moreover, a Maximum Relevance Minimum Redundancy filter based on the algorithm scores (15) was embedded within model training as a feature selection method.

A permutation-based approach was employed in order to determine the statistical significance of the models such that the class labels were permuted 1000 times and the balanced accuracy for each of these models was calculated. Then, the number of times that the balanced accuracy of these permuted models was equal or higher to the observed accuracy of our models was devided by the total number of permutations and this value was considered as the probability of the model's performance being a chance level occurence within a normal distribution. The obtained *P*-values for each unimodal model were further FDR-corrected within each of the classification problems in order to account for multiple comparisons.

Following this, models that performed statistically above chance level were further evaluated in terms of their predictive features by means of a combined metric of feature stability and feature importance. Specifically, the cross-validation ratio is considered to be an index of feature stability or consistency across cross-validation folds and is computed as the sum of the selected CV1 median weights across all the CV2 folds divided by the standard error of the selected CV1 weights for each voxel, as detailed elsewhere (4). We thresholded the cross-validation ratio maps based on feature importance as measured using a sign-based consistency metric, in order to only retain voxels significantly contributing to the predictive model (at $\alpha = .05$). The sign-based consistency measure assigns feature importance based on the times that a specific feature has the same sign (positive/negative) across the ensemble, adapted from the method of Gómez-Verdejo et al. (2019) (16) and detailed in Koutsouleris et al., 2021 (4). Briefly, this procedure produces a consistency between 0 (equally positive and negative weights across the ensemble) and 1 (perfect sign consistency), based on which z-scores and associated normal cumulative distribution-based standard *P* values for each feature are derived. Subsequently, these *P*-values were corrected using the false discovery rate and the $-\log_{10} P_{FDR}$-values were multiplied by the sign of the mean weight for that feature

across the ensemble, in order to also capture the direction of the predictive feature effect across the cross-validation structure for visualization and interpretability purposes.

Lastly, the decision scores coming from the single-modality classifiers that performed significantly above chance level were stacked in order to evaluate their complementary value for each classification problem (cross-sectional and persistence analyses). For the stacked model, the unimodal decision scores were median-standardized and outliers (± 4 standard deviations) were winsorized within the same CV structure as the unimodal classifiers. Following this, L2-regularized SVM classifiers were trained on the unimodal decision scores with the same parameters as for the single-modality ones and model significance was tested based on the same permutation approach. Moreover, we ran a global Quade test (17) followed by post-hoc pairwise *t*-test comparisons in order to identify whether the performance of the stacked model combining all predictive modalities was significantly different from the performances of the single-model classifiers that performed above chance level.

Supplementary Tables

Table S1. General and study group specific inclusion and exclusion criteria used in the Personalized Prognostic Tools for Early Psychosis Management (PRONIA) study, based on Koutsouleris et al., 2021 (4) and Dwyer et al., 2022 (18).

Study group	Inclusion Criteria	Exclusion Criteria
General	1. Age between 15 and 40 years 2. Sufficient language abilities for participation 3. Ability to provide consent / assent	1. IQ below 70 2. Hearing is not sufficient for neuro-cognitive testing 3. Current or past head trauma with loss of consciousness (> 5 min) 4. Current or past known neurological disorder of the brain 5. Current or past known somatic disorder potentially affecting the structure or functioning of the brain 6. Current or past alcohol dependence 7. Current poly-substance dependence or within the past six months 8. Any contra-indication for MRI
Recent-onset psychosis	1. The presence of a psychotic episode as determined by the DSM-IV-TR affective or non-affective psychotic episode category 2. The psychotic episode was present in the past 3 months. 3. The onset of psychosis has been within the past 24 months.	Intake of antipsychotic medication for longer than 90 cumulative days with a daily dose rate at or above the minimum dosage threshold defined by the DGPPN S3 Guidelines for the treatment of first-episode psychosis (19)
Healthy Controls		1. Any current or past DSM-IV axis disorder 2. A positive familial history (1st degree relatives) of affective or non-affective psychoses or major affective disorders; 3. Intake of psychotropic medication or drugs more than 5 times/year or in the month before study inclusion.

Table S2. Neuroimaging acquisition parameters for the structural MRI data across the different PRONIA sites. Adapted with minimal modifications from Koutsouleris et al. (2018) (3).

Site	Scanner Model	Field Strength	Flip Angle	Coil Channels	Voxel Size (mm)	TR (ms)	TE (ms)	FOV (mm)	Number of slices
Munich	Philips Ingenia	3T	8	32	0.97 x 0.97 x 1	9.5	5.5	250 x 250	190
Milan	Philips Achieva Intera	1.5T	12	8	0.93 x 0.93 x 1	Short - est (8.1)	Short - est (3.7)	240 x 240	170
Cologne	Philips Achieva	3T	8	8	0.97 x 0.97 x 1	9.5	5.5	250 x 250	190
Basel	SIEMENS Verio	3T	8	12	1 x 1 x 1	2000	3.4	256 x 256	176
Birmingham	Philips Achieva	3T	8	32	1 x 1 x 1	8.4	3.8	288 x 287	175
Turku	Philips Ingenuity	3T	7	32	1 x 1 x 1	8.1	3.7	256 x 256	176
Udine	Philips Achieva	3T	12	8	0.93 x 0.93 x 1	Short - est (8.1)	Short - est (3.7)	240 x 240	170
Münster	SIEMENS Prisma fit	3T	8	12	1 x 1 x 1	2130	2.3	256 x 256	192

Note. TR – repetition time; TE – echo time; FOV – field of view.

Table S3. Neuroimaging acquisition parameters for the resting-state functional MRI data across the different PRONIA sites.

Site	Scanner Model	Field Strength	Flip Angle	Coil Channels	Voxel Size (mm)	TR (ms)	TE (ms)	FOV (mm)	Number of slices	Number of volumes	Slice order
Munich	Philips Ingenia	3T	90	32	2.88 x 2.88 x 3	3000	30	230 x 230	53	200	Ascending
Milan	Philips Achieva Intera	1.5T	90	8	3 x 3 x 3	3000	32	240 x 240	45	200	Interleaved
Cologe	Philips Achieva	3T	90	8	2.88 x 2.88 x 3	3000	30	230 x 230	53	200	Ascending
Basel	SIEMENS Verio	3T	82	12	2.98 x 2.98 x 3	3000	28	256 x 256	34	200	Interleaved
Birmingham	Philips Achieva	3T	85	32	3 x 3 x 3	3000	34.5	240 x 240	52	200	Interleaved
Turku	Philips Ingenuity	3T	90	32	3 x 3 x 3	3000	30	240 x 240	53	200	Interleaved
Udine	Philips Achieva	3T	90	8	3 x 3 x 3	3000	32	240 x 240	45	200	Interleaved
Münster	SIEMENS Prisma fit	3T	90	12	3 x 3 x 3	3000	26	256 x 256	51	200	Interleaved

Note. TR – repetition time; TE – echo time; FOV – field of view.

Table S4. Comparison between the performance of the persistence prediction models trained on individual data modalities and that of the stacked model combining all data modalities.

We first ran a global Quade test (17) in order to test whether there were any significant differences between any of the models' performances within the respective research question. Provided the Quade test was significant, post-hoc pairwise *t*-test comparisons were run in order to identify which unimodal model performances significantly differed from the stacked model combining all modalities. The performance metrics for all the unimodal models are presented in Table 2 in the main document.

Models compared	Global Quade Test		Post-hoc tests	
	W(df)	p_{FDR}	*t*(df)	p_{FDR}
Persistence models				
	3.57 (5, 245)	**.003**		
Stacked model – Slow-5			1.03(98)	.21
Stacked model – Slow-4			0.21(98)	.43
Stacked model – Slow-3			1.87(98)	.08
Stacked model – GMV			3.44(98)	**.03**
Stacked model – WMV			1.21(98)	.17

Note. P values are adjusted using the false-discovery rate correction for multiple comparisons and statistically significant values (at α = .05) are presented in bold.

Table S5. Performance metrics of models trained using an outer leave-site-out cross-validation framework for the high FThD persistence models. All models were built using the same preprocessing and training parameters as those described for the main analyses in section 2.4. of the main manuscript and Text S3 of this supplement.

Prediction of high FThD persistence (from baseline to 1-year follow-up)								
	Leave-site-out nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	P_{FDR} **value**
Rs-fMRI slow-5	66.7	59.6	63.1	.58	12.3	95.5	7.8	**.03**
Rs-fMRI slow-4	50.0	63.8	56.9	.62	10.5	93.8	4.3	>.05
Rs-fMRI slow-3	66.7	58.2	62.4	.65	11.9	95.3	7.3	**.05**
GMV	50.0	56.0	53.0	.64	8.8	92.9	1.8	>.05
WMV	83.3	48.9	66.1	.67	12.2	97.2	9.4	**.003**
Stacked model	83.3	54.6	69.0	.76	13.5	97.5	11.0	**.003**

Table S6. Associations between the decision scores of the unimodal cross-sectional and persistence classifiers. Pearson's correlation coefficients were calculated between all the significant cross-sectional and persistence models trained on the individual data modalities.

Model	1.	2.	3.	4.	5.
1. GMV – cross-sectional					
2. Slow-5 fALFF – persistence	-.10 [-.26, .06]				
3. Slow-4 fALFF – persistence	**.21 [.05, .35]**	**.42 [.28, .55]**			
4. Slow-3 fALFF – persistence	.15 [-.01, .30]	**.62 [.52, .71]**	**.76 [.68, .82]**		
5. GMV – persistence	**.60 [.50, .70]**	.11 [-.05, .26]	**.31 [.16, .44]**	**.28 [.13, .42]**	
6. WMV – persistence	**.30 [.15, .44]**	**.20 [.05, .40]**	**.31 [.16, .45]**	**.31 [.16, .45]**	**.50 [.36, .60]**

Note. The values in square brackets represent the 95% confidence interval for the Pearson's *r* coefficients. *P* values were adjusted using the false-discovery rate correction for multiple comparisons and statistically significant values (at α = .05) are presented in bold.

Table S7. Performance metrics of models trained using non-standardized fALFF data instead of the z-scored images. All models were built using the same preprocessing and training parameters, as well as the same cross-validation structures as those described for the main analyses in section 2.4. of the main manuscript and Text S3 of this supplement.

Cross-sectional high vs. low FThD subgroup classification								
	Leave-site-out nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**p_{FDR} value**
Rs-fMRI slow-5	56.9	48.9	52.9	.55	27.0	77.3	4.3	>.05
Rs-fMRI slow-4	59.6	47.1	53.4	.57	27.0	78.1	5.1	>.05
Rs-fMRI slow-3	56.1	52.6	54.4	.58	27.8	78.6	6.5	>.05
Prediction of high FThD persistence (from baseline to 1-year follow-up)								
	Pooled nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**p_{FDR} value**
Rs-fMRI slow-5	75.0	58.2	66.6	.76	13.2	96.5	9.7	**<.001**
Rs-fMRI slow-4	75.0	56.0	65.5	.74	12.7	96.3	9.0	**<.001**
Rs-fMRI slow-3	66.7	58.2	62.4	.74	11.9	95.3	7.3	**.002**

Table S8. Performance metrics of models trained without correcting for general illness effects as measured using the PANSS total score. Apart from removing the regression of the total *PANSS* score from the data preprocessing steps, all models were built using the same preprocessing and training parameters, as well as the same cross-validation structures as those described for the main analyses in section 2.4. of the main manuscript and Text S3 of this supplement.

Cross-sectional high vs. low FThD subgroup classification								
	Leave-site-out nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**p_{FDR} value**
Rs-fMRI slow-5	69.0	49.7	59.3	.63	31.2	82.9	14.1	**.004**
Rs-fMRI slow-4	50.0	55.4	52.7	.57	27.1	77.0	4.1	>.05
Rs-fMRI slow-3	55.2	56.0	55.6	.61	29.4	79.0	8.4	>.05
GMV	58.6	57.1	57.9	.61	31.2	80.6	11.8	**.041**
WMV	60.3	50.9	55.6	.58	28.9	79.5	8.4	>.05
Prediction of high FThD persistence (from baseline to 1-year follow-up)								
	Pooled nested cross-validation							
	Sensitivity, %	**Specificity, %**	**BAC, %**	**AUC**	**PPV, %**	**NPV, %**	**PSI**	**p_{FDR} value**
Rs-fMRI slow-5	83.3	62.4	72.9	.78	15.9	97.8	13.7	**<.001**
Rs-fMRI slow-4	83.3	62.4	72.9	.82	15.9	97.8	13.7	**<.001**
Rs-fMRI slow-3	75.0	58.2	66.6	.72	13.2	96.5	9.7	**<.001**
GMV	75.0	51.1	63.0	.71	11.5	96.0	7.5	**.002**
WMV	83.3	52.5	67.9	.72	13.0	97.4	10.4	**<.001**

Supplementary Figures

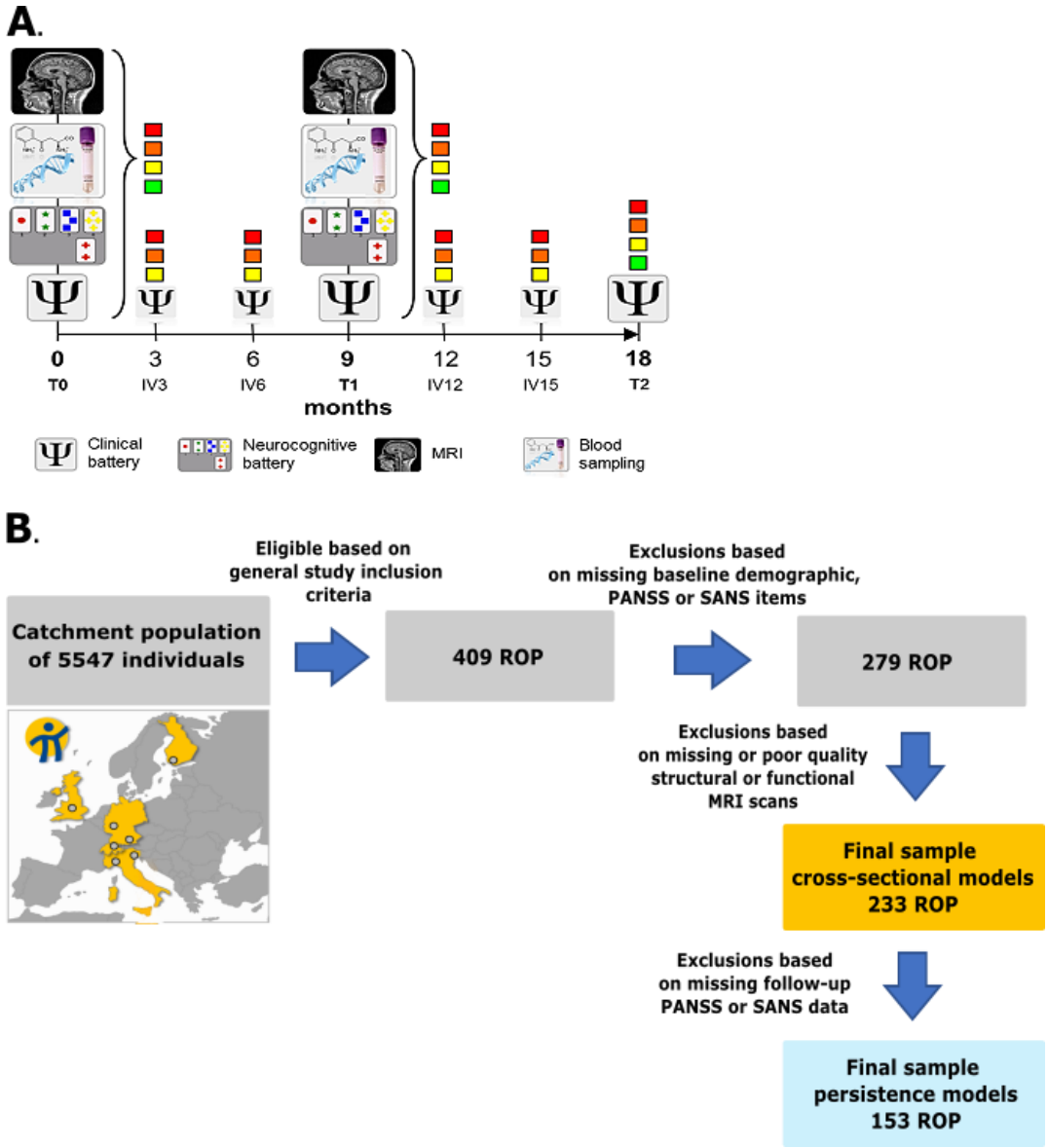


Figure S1. PRONIA study follow-up protocol and sampling procedure for the current study.

A. The PRONIA study employed a multi-modal longitudinal design, following healthy individuals and patients over the course of 18 months. The colored squares represent the clinical study group assessed at each specific timepoint as follows: green – healthy controls, yellow – patients with recent-onset depression, orange – persons with a clinical high-risk for psychosis, red – patients with recent-onset psychosis. **B.** Participants from 8 European sites were included in the current analyses. Apart from the general inclusion criteria of the PRONIA study (Table S1), we further excluded participants due to missing baseline or follow-up PANSS or SANSS data, missing or insufficient quality of the structural or functional MRI scans data (detailed in the Supplementary methods).

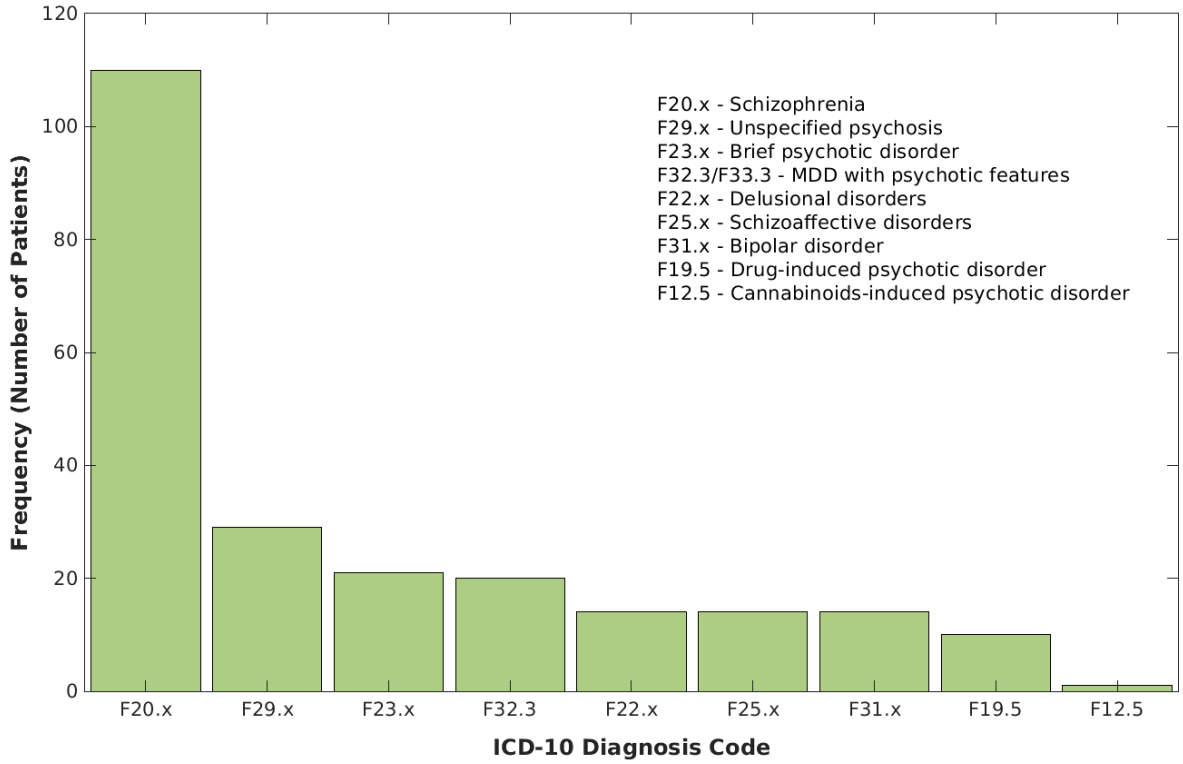


Figure S2. Distribution of ROP patients across diagnosis categories based on the *ICD-10 classification of mental and behavioral disorders* by *World Health Organization* (WHO). The number of patients per category within the full cross-sectional sample used for the current study (N = 233) is presented on the y-axis. The ICD-10 diagnosis codes corresponding to recent-onset psychosis associated disorders are presented on the x-axis.

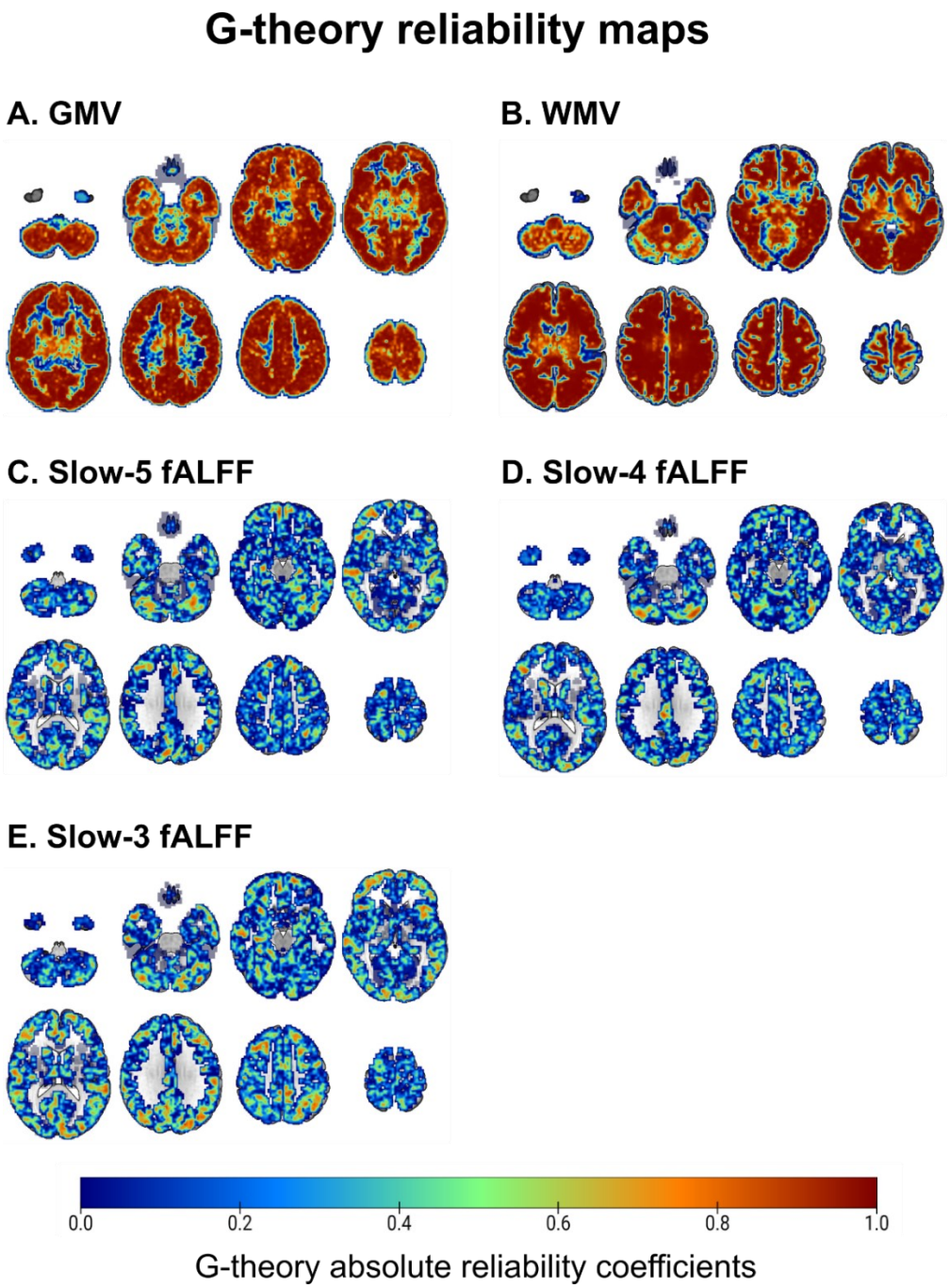


Figure S3. Distribution of the absolute reliability coefficients derived from the G-theory analysis based on the PRONIA traveler subjects for each neuroimaging data modality (A-E). The absolute reliability coefficients reflect the degree of consistency in the rank ordering of voxels across sites and the consistency in the elevations of the raw scores across sites (also known as the Phi coefficient or index of dependability). An intensity of 1 (warm colors) indicates perfect reliability across sites (100%) for the respective voxel. The most reliable 25% - 50% - 75% voxels within the maps (optimized within the CV structure) across sites were retained for model training.

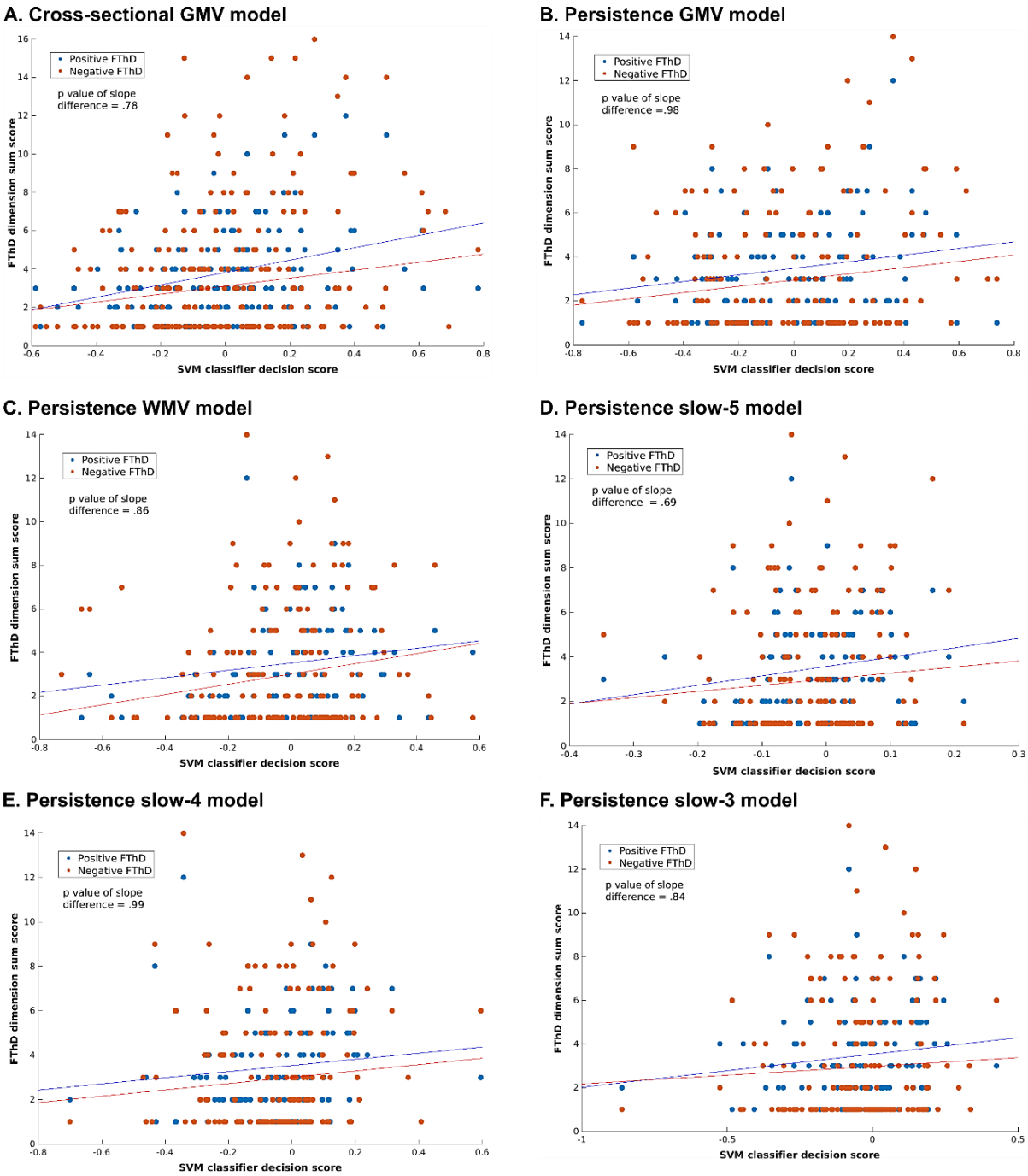


Figure S4. Slope differences for the relationships between the decision scores of the predictive models and positive and negative FThD subscores. Positive and negative FThD subscores were defined by taking the sum score of the corresponding *PANSS/SANS* items used for clustering by Oeztuerk et al., 2022 (5) as follows: objective positive FThD-related symptoms: conceptual disorganization (*PANSS* P2), poverty of content speech (*SANS* 10); objective negative FThD-related symptoms: difficulty in abstract thinking (*PANSS* N5), increased latency of response (*SANS* 12) and poverty of speech (*SANS* 9). The data distribution and least squares regression slopes for the relationships between the decision scores of each significant model (A-F) and each of the positive and negative sum scores were plotted using MATLAB. The significance of the slope difference for the two FThD-associated symptom dimensions at an α = .05 was evaluated.

Supplementary References

1. Kay, S R Fiszbein A L a O. Positive and Negative Syndrome Scale (Panss) Rating Criteria. Schizophr Bull [Internet]. 1987;13(2):261–76. Available from: www.ncbi.nlm.nih.gov
2. Andreasen NC. The scale for the assessment of negative symptoms (SANS). British Journal of Psychiatry. 1989;155:49–52.
3. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry. 2018;75(11):1156–72.
4. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients with Clinical High-Risk Syndromes and Recent-Onset Depression. JAMA Psychiatry. 2021 Feb 1;78(2):195–209.
5. Oeztuerk OF, Pigoni A, Wenzel J, Haas SS, Popovic D, Ruef A, et al. The clinical relevance of formal thought disorder in the early stages of psychosis: results from the PRONIA study. Eur Arch Psychiatry Clin Neurosci. 2021 Apr 1;
6. Manjón J V, Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. Journal of Magnetic Resonance Imaging. 2010;31(1):192–203.
7. Haas SS, Antonucci LA, Wenzel J, Ruef A, Biagianti B, Paolini M, et al. A multivariate neuromonitoring approach to neuroplasticity-based computerized cognitive training in recent onset psychosis. Neuropsychopharmacology. 2021 Mar 1;46(4):828–35.
8. Song XW, Dong ZY, Long XY, Li SF, Zuo XN, Zhu CZ, et al. REST: a toolkit for resting-state functional magnetic resonance imaging data processing. PLoS One. 2011;6(9):e25031.
9. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage. 2014;84:320–41.
10. Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage. 2013;64:240–56.
11. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. 1995.
12. Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, et al. The oscillating brain: complex and reliable. Neuroimage. 2010;49(2):1432–45.
13. Zang YF, Yong H, Chao-Zhe Z, Qing-Jiu C, Man-Qiu S, Meng L, et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 2007 Mar;29(2):83–91.
14. Cronbach LJ, Rajaratnam N, Gleser GC. Theory of generalizability: A liberalization of reliability theory. British Journal of Statistical Psychology. 1963;16(2):137–63.
15. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005;27(8):1226–38.

16. Gómez-Verdejo V, Parrado-Hernández E, Tohka J, Initiative ADN. Sign-consistency based variable importance for machine learning in brain imaging. Neuroinformatics. 2019;17:593–609.
17. Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. J Am Stat Assoc. 1979;74(367):680–3.
18. Dwyer DB, Buciuman MO, Ruef A, Kambeitz J, Sen Dong M, Stinson C, et al. Clinical, Brain, and Multilevel Clustering in Early Psychosis and Affective Stages. JAMA Psychiatry. 2022 Jul 1;
19. Falkai P, Gaebel W, Weinmann S, Wobrock T. DGPPN (Hrsg.) S3–Praxisleitlinien in Psychiatrie und Psychotherapie Band 1: Behandlungsleitlinie Schizophrenie. 2006.

Appendix A: Review Paper

European Archives of Psychiatry and Clinical Neuroscience

Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half century studies

Oemer Faruk Oeztuerk[1,2] · Alessandro Pigoni[3] · Linda A. Antonucci[1,4,5] · Nikolaos Koutsouleris[1,6,7]

[1] Department of Psychiatry and Psychotherapy, Ludwig Maximilian-University Munich, Nussbaumstr. 7, 80336 Munich, Germany
[2] International Max Planck Research School for Translational Psychiatry, Munich, Germany
[3] MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy
[4] Department of Basic Medical Sciences, Neuroscience and Sense Organs–University of Bari "Aldo Moro", Bari, Italy
[5] Department of Education, Psychology and Communication Science–University of Bari "Aldo Moro", Bari, Italy
[6] Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[7] Max-Planck Institute of Psychiatry, Munich, Germany

Keywords: Formal thought disorder · Clinical high risk · Psychosis · Functioning · Neurocognition

Received: 14 March 2021 / Accepted: 4 July 2021

Published Online: 14 July 2021

Citation: Oeztuerk, O.F., Pigoni, A., Antonucci, L.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.
Eur Arch Psychiatry Clin Neurosci 272, 381–393 (2022).
https://doi.org/10.1007/s00406-021-01295-3

European Archives of Psychiatry and Clinical Neuroscience (2022) 272:381–393
https://doi.org/10.1007/s00406-021-01295-3

INVITED REVIEW

Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half-century studies

Oemer Faruk Oeztuerk[1,2] · Alessandro Pigoni[3] · Linda A. Antonucci[1,4,5] · Nikolaos Koutsouleris[1,6,7]

Received: 14 March 2021 / Accepted: 4 July 2021 / Published online: 14 July 2021

Abstract

Recent review articles provided an extensive collection of studies covering many aspects of format thought disorders (FTD) among their epidemiology and phenomenology, their neurobiological underpinnings, genetics as well as their transdiagnostic prevalence. However, less attention has been paid to the association of FTD with neurocognitive and functioning deficits in the early stages of evolving psychosis. Therefore, this systematic review aims to investigate the state of the art regarding the association between FTD, neurocognition and functioning in the early stages of evolving psychotic disorders in adolescents and young adults, by following the PRISMA flowchart. A total of 106 studies were screened. We included 8 studies due to their reports of associations between FTD measures and functioning outcomes measured with different scales and 7 studies due to their reports of associations between FTD measures and neurocognition. In summary, the main findings of the included studies for functioning outcomes showed that FTD severity predicted poor social functioning, unemployment, relapses, re-hospitalisations, whereas the main findings of the included studies for neurocognition showed correlations between attentional deficits, executive functions and FTD, and highlighted the predictive potential of executive dysfunctions for sustained FTD. Further studies in upcoming years taking advantage of the acceleration in computational psychiatry would allow researchers to re-investigate the clinical importance of FTD and their role in the transition from at-risk to full-blown psychosis conditions. Employing automated computer-assisted diagnostic tools in the early stages of psychosis might open new avenues to develop targeted neuropsychotherapeutics specific to FTD.

Keywords Formal thought disorder · Clinical high risk · Psychosis · Functioning · Neurocognition

✉ Oemer Faruk Oeztuerk
oemer.oeztuerk@med.uni-muenchen.de

1 Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Nussbaumstr. 7, 80336 Munich, Germany

2 International Max Planck Research School for Translational Psychiatry, Munich, Germany

3 MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy

4 Department of Basic Medical Sciences, Neuroscience and Sense Organs–University of Bari "Aldo Moro", Bari, Italy

5 Department of Education, Psychology and Communication Science–University of Bari "Aldo Moro", Bari, Italy

6 Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

7 Max-Planck Institute of Psychiatry, Munich, Germany

Introduction

Formal thought disorders (FTD) are psychopathological alterations that can emerge in different psychiatric disorders such as schizophrenia, major depressive disorder, and mania. [1, 2] FTD are a multidimensional construct involving thought, language, and communication disturbances, such as loosening of associations, blocking, semantic and phonemic paraphasia. [1, 3] It has taken one and half centuries to formulate FTD as such a construct, starting from Esquirol pointing to a primary pathology in coordinating ideas [4] (1838) and from Griesinger's differentiation [5] between "formal deviations" and "false contents" (1867). Further clinical developments over the twentieth century led to further conceptualizations of FTD, such as "dementia praecox" [6] from Kraepelin (1919), "derailment" [7] from Schneider (1930), "concrete thinking" [8] from Goldstein (1944), "loosening of associations" [9] from Bleuler's

(1950), and Andreasen's (1985) revision of FTD as a multidimensional psychopathological construct of disturbances in thought, language and communication [10]. (Jerónimo et al. 2018 for more details [11]).

FTD severity has been assessed with different scales since Kraepelin and Bleuler postulated the importance of earlier manifestation of this clinical phenomenon in an evolving psychosis [12]. However, before the development of specific scales such as—in chronological order—Andreasen's scale for the assessment of thought, language and communication (TLC) [10], the Thought and Language Index (TLI) [13], the Thought Disorder Index (TDI) [14], and the scale for the assessment of Thought, Language and Disorder (TALD) [3], FTD severity has been assessed through items part of non-FTD-specific psychopathological scales, such as the Positive and Negative Syndrome Scale (PANSS) [15], the Scale for the Assessment of Positive Symptoms (SAPS) [16], the Scale for the Assessment of Negative Symptoms (SANS) [17] and the Brief Psychiatric Rating Scale (BPRS) [18]. These non-specific scales usually address one or only a few psychopathological aspects of FTD. [2] Moreover, the heterogeneity [2] of the specific scales makes quantitative comparison of findings in literature difficult, as they capture various psychopathological aspects of FTD.

Recent review articles provided an extensive collection of studies covering many aspects of FTD among their epidemiology and phenomenology, their neurobiological underpinnings, genetics, their neurological correlates, as well as their transdiagnostic prevalence. [1, 2, 11, 12] However, literature regarding the investigation of FTD as early clinical signs of an evolving psychosis, as well as their prognostic importance, is rather limited. Indeed, the majority of findings addressed FTD relevance for disease course either in chronic patients or in patients with a highly heterogeneous age range (18–65). [19, 20] The limited literature in early stages of psychosis revealed FTD not only as a core feature of psychosis but also in association with several psychosis-related adverse outcomes, such as functional impairments and cognitive deficits. [2, 3, 11, 12, 21] These findings are particularly relevant, as functional and neurocognitive impairments [22, 23] frequently precede disease onset and persist after remission of the acute illness in psychotic disorders. [24, 25]. Consistently, given perspective changes and contributions in the last decades regarding early diagnostic tools aimed at recognizing such persisting impairments with an impact on real-world prognostic outcomes, research focus has been turned on risk groups and younger patients experiencing first-episode psychosis. [26–29] With specific respect to FTD, only very few studies have highlighted that their severity has been associated with increased (re-)hospitalization rate [19], unemployment risk [30] and reduced quality of life [31]. Moreover, patients with schizophrenia experiencing enduring FTD after the acute phase of psychosis showed lower occupational functioning levels and higher relapse rates. [32] On the other hand, FTD have been associated with attentional executive functions deficits at the early stages of psychosis, [33, 34], even if this association has not been fully replicated yet. Furthermore, executive dysfunctions at the early stages of psychosis seem to predict FTD severity at follow up [35]. However, despite these findings and the increasing interest in early diagnostics and prevention in psychotic disorders, literature on prodromal state and first-episode psychosis so far did not provide any target for preventive interventions based on the core psychopathological changes such as FTD predicting adversity in clinical outcomes, yet. A potential way to understand this is to thoroughly investigate whether and how FTD are related to relevant outcomes of psychosis, i.e., deficits already present in their early stages, such as functioning and neurocognitive impairments.

Therefore, we conducted a systematic review regarding the state of the art of the association between FTD, functioning outcomes and neurocognitive impairments in the early and prodromal stages of psychotic disorders. We summarized the main findings of the included studies providing evidence related to the early diagnostic and prognostic potential of FTD in association with functioning and neurocognition. Furthermore, we discussed possible reasons for the paucity of studies investigating the clinical relevance of FTD in early psychosis and presented new perspectives for potential future investigations on FTD with the help of modern computational analytical techniques that might serve as computer-assisted early diagnostic tools. We concluded with an outlook for future research targeting the potential preventive and predictive role of FTD in psychosis trajectories.

Methods

We conducted two separate systematic literature searches; (i) association between FTD and functioning outcomes and (ii) association between FTD and neurocognition, both in the early and prodromal stages of psychosis, by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [36] statement. Two of the authors (Ö.F.Ö and A.P.) independently conducted the systematic literature searches in PubMed, PsychINFO and Web of Science covering the last half-century until 2019.

For the literature search regarding the association between FTD and functioning outcomes, we used the following search terms combination: ("formal thought disorder" OR "thought disorder") AND (child OR adolescent OR young adult) AND (psychosis OR "early-onset schizophrenia" OR "first episode psychosis" OR "recent onset" OR "high risk" OR prodrom) AND ("GAF" OR "GF" OR "functioning" OR "disability" OR occupational OR social. For the literature

search regarding the association between FTD and neurocognition, we used the following search terms combination: ("formal thought disorder" OR "thought disorder") AND (child OR adolescent OR young adult) AND (psychosis OR "early onset schizophrenia" OR "first episode psychosis" OR "recent onset" OR "high risk" OR prodrom) AND (neurocognition OR neurocognitive or cognition or cognitive or neuropsycho OR memory OR executi OR process).

Articles had to meet the following inclusion criteria: (1) reported statistically significant associations between FTD and either functioning outcomes, or neurocognitive measures; (2) did include high-risk groups [37] such as ultra-high risk (UHR), attenuated psychotic symptoms (APS), the brief limited intermittent psychotic episode (BLIPS) and genetic risk and deterioration syndrome (GRD) or prodromal phases or early stage of psychosis such as first-episode psychosis (FEP), early psychosis (EP) (Onset of Psychosis < 2 years); (3) did include children, adolescents or young adults (< 35 years of age); (4) did include data of functioning outcomes or neurocognition. Articles were excluded for the following reasons: (a) not in English, (b) sample composed only of participants with chronic psychosis or with an onset of disease > 2 years (c) sample composed only of adult patients (18–65 age range) without any specific subgroups analyses based on age ranges (d) did include individuals with drug-induced psychosis (e) did include individuals with psychosis due to medical conditions. Case series, book chapters, literature reviews, conference papers, meeting abstracts or meta-analyses were also excluded.

Based on these criteria, as a first step, two authors have screened the titles and abstracts separately in three different databases (PsycINFO, Medline and Web of Science), and the results of included and excluded articles were discussed among co-authors. Based on the abstracts screening, records for which an inclusion or exclusion decision could not be taken were listed. Thus, in a second round, the same two authors have screened the full texts of these articles to check their eligibility for inclusion. Records for which both authors agreed on their inclusion were then proved for their eligibility again through further full-text investigation.

The procedure used for the selection of studies for each literature search is reported in Figs. 1 and 2. The included studies are summarized in Tables 1 and 2.

Results

Functioning outcomes and FTD

We screened 106 studies in total and included 8 of them. These studies were included due to the presence of significant associations between FTD in adolescents or young adults and their reports, and functioning outcomes measured with different scales. Overall, the included studied showed that higher severity of FTD predicted poor social functioning, unemployment, relapses, re-hospitalisations. Three studies, namely Harrow. [38], Marengo [30] and Racenstein et al. [39], reported results from the Chicago Follow-up Study, assessing the importance of FTD through a comprehensive measure of bizarre-idiosyncratic thinking in the acute phase of inpatients at baseline, and at follow-ups taking place 1 ½, 2, 4 and 8 years after hospital discharge. They reported that approximately 30% of patients diagnosed with schizophrenia exhibited persisting thought disorders during follow-up periods and that such patients with enduring thought disorder signs were also a poor-outcome group. Specifically, they highlighted a poorer prognosis with lower occupational functioning levels (subsequent unemployment 82% and higher relapse/re-hospitalisation rates: psychosis 68% and rehospitalization 59%) in schizophrenia patients experiencing enduring FTD after the acute phase. Moreover, they showed that FTD and functioning levels are correlated in the first 8 year of schizophrenia, with a stronger correlation between FTD and occupational functioning compared with social functioning.

Kotov et al. [40] interviewed 628 inpatients with first-episode psychosis six times over two decades in an epidemiologic cohort. They showed that a four-factor model (reality distortion, disorganization, inexpressivity, and apathy/asociality) could significantly predict functional outcomes. Interestingly, they observed that apathy/asociality predicted impairments in global functioning, social functioning, role functioning, and life satisfaction, whereas inexpressivity predicted lower residential independence. Minor et al. [41] assessed the positive FTD via the Communication Disturbances Index (CDI) [42] and speech production in early psychosis (EP) individuals and controls, and explored their association to real-world outcomes. They reported large differences in both positive FTD and speech production between EP individuals and controls, and showed that positive FTD and affective reactivity were associated with poor social and role functioning only in EP. Moreover, they found that positive FTD and affective reactivity were consistently accounting for poor social functioning (up to 56% of social functioning's variance) and associated with poor role functioning (accounting for up to 46% of the variance) in EP.

Roche et al. [19] evaluated the relationship between FTD features (namely, disorganization, verbosity and poverty of speech) and social and occupational functioning outcomes at baseline and 1 year later in a first-episode psychosis cohort. This study found that only disorganization was associated with functional outcome, specifically social functioning and that the longitudinal course of disorganization remained significantly associated with social functioning on multivariate analysis. Moreover, they reported that higher baseline severity of disorganization predicted

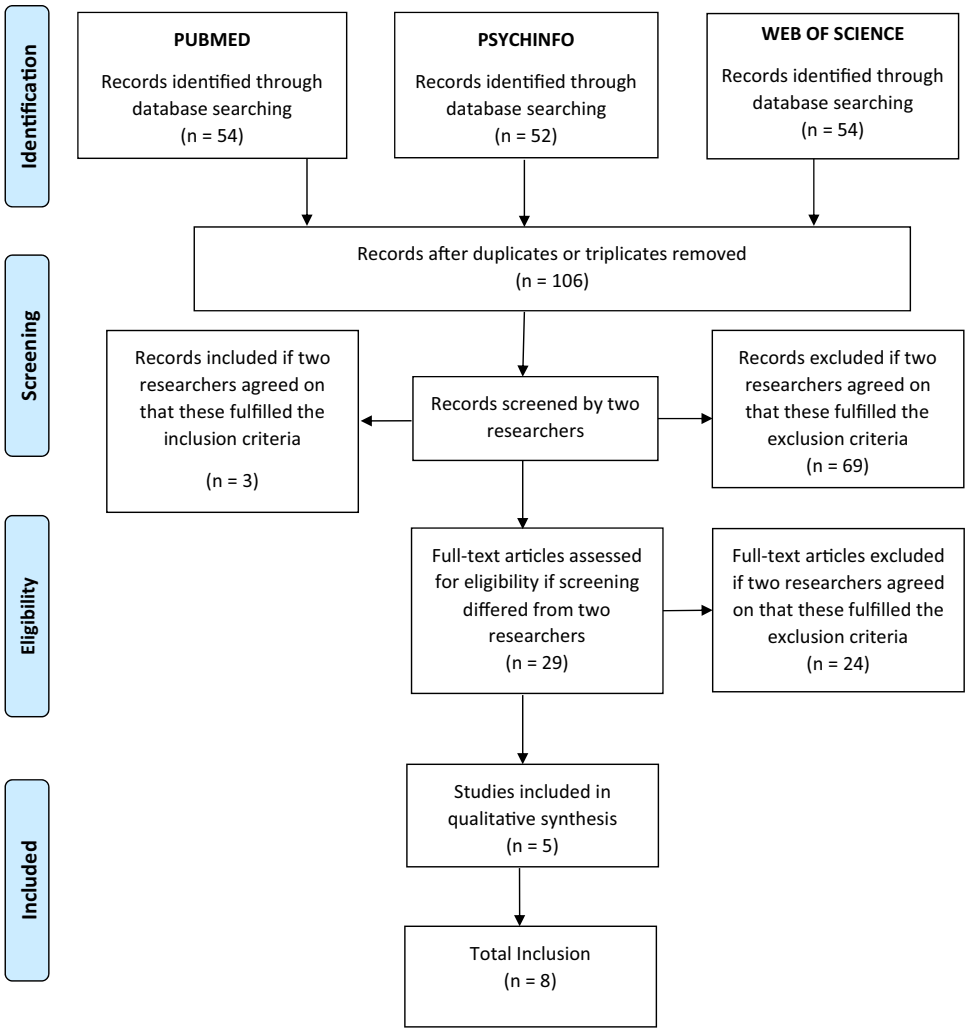


Fig. 1 Prisma Graphs representing the inclusion of the studies related to functioning and FTD association

a greater number of hospitalisations and prolonged hospitalisation during the first year of illness. Burton et al. [43] tested the efficacy of early intervention for youth at risk of developing psychosis in a multisite national trial dataset, where they followed participants prospectively at 6, 12, and 24 months, and examined the relationships between baseline symptoms and longitudinal global social and role functioning. They found that higher baseline negative symptoms and deteriorated thought process predicted worse social and role functioning for up to 2 years among adolescents and young adults at risk for psychosis, and that the changing effect of negative symptoms on social functioning over time was moderated by positive symptoms. Bearden et al. [44] examined the association of baseline FTD assessed by the Kiddie Formal Thought Disorder Rating Scale (K-FTDS) [45] and linguistic cohesion with conversion to psychosis and social and role outcome at follow-up (approximately 1 year later) by analyzing transcribed speech samples in individuals with a clinical high-risk for psychosis. They reported that baseline poverty of content (POC) and referential cohesion predicted significantly social and role functioning at follow-up.

Neurocognition and FTD

We screened 106 studies in total and included 7 of them. These studies were included due to the existence of significant associations between FTD in adolescents or young adults, and neurocognition. The study from Minor et al. [41] was included both in this section and on the one regarding functioning, as this study investigated the association between FTD and both domains. Overall, the included studies for the associations between FTD and neurocognition showed heterogeneous findings. Still, some consistent findings were observed for what concerns associations between attentional deficits, executive functions and FTD.

Minor et al. [41] explored associations between positive FTD assessed by the Communication Disturbances Index (CDI) [42] and speech production assessed by 2 min records about negative (affective condition) and neutral (baseline,

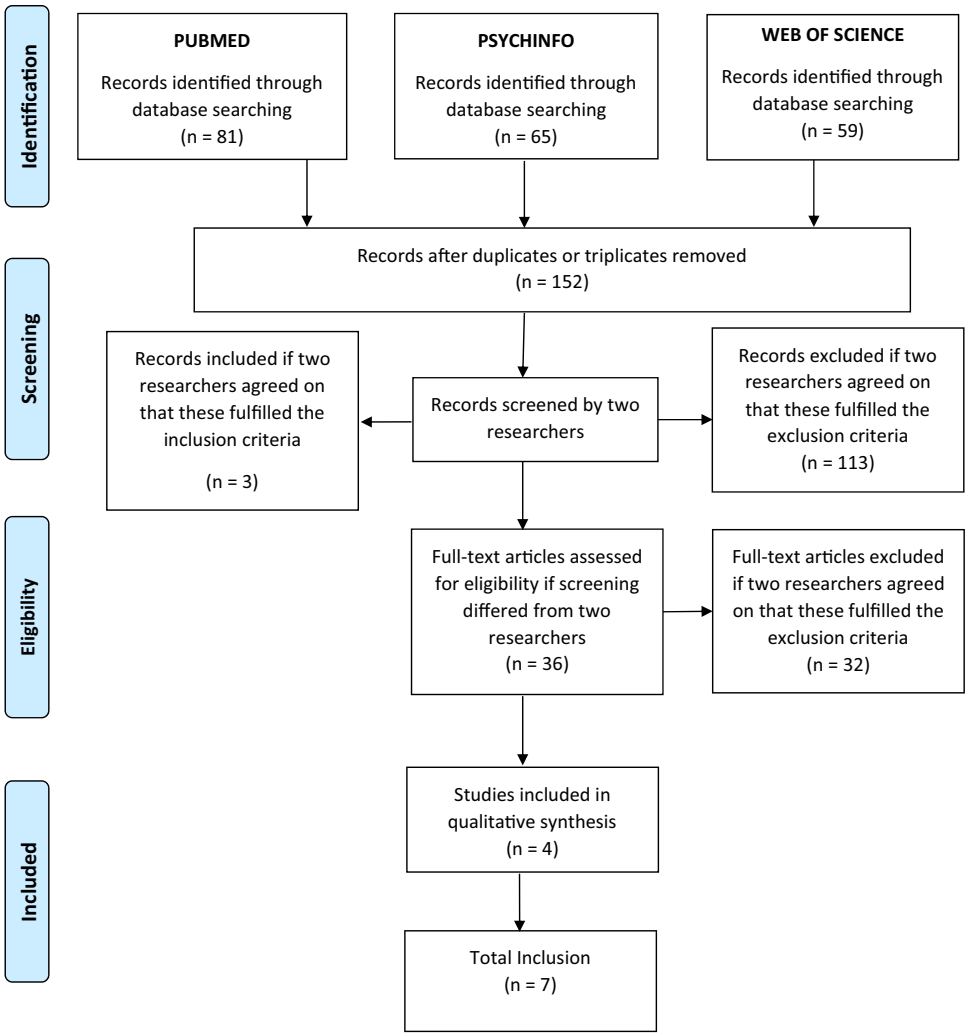


Fig. 2 Prisma Graphs representing the inclusion of the studies related to neurocognition and FTD association

cognitive conditions) memories in early psychosis (EP) individuals and controls. They employed a single task for baseline condition, where participants completed only one-back visual working memory test, whereas a dual task for cognitive condition, where participants completed the one-back visual working memory test while simultaneously generating speech. They showed that cognitive reactivity appeared to have a less defined role than affective reactivity in positive FTD. Based on these findings, they suggested that the cognitive load effect may not translate to positive FTD or speech production in early psychosis. In contrast, affective systems played a prominent role in positive FTD that EP individuals exhibited a steeper increase in positive FTD from the baseline to the affective condition compared to controls.

Nuechterlein et al. [33] investigated the association between three cognitive tests in the attention domain (namely, two versions of the continuous performance test (CPT) and forced-choice span of apprehension task) and FTD assessed through Brief Psychiatric Rating Scale (BPRS) Conceptual Disorganization subscores, Rorschach Thought Disorder Index factors, and negative symptoms assessed by BPRS Anergia factor scores in inpatients experiencing an early phase of schizophrenic disorders. They retested 32 patients after clinical stabilization to address the extent to which continued attentional deficits were associated with specific symptomatology during the hospitalized period. Their results indicated that signal discrimination deficits were consistently related to the presence of negative symptoms and that the correlation of the attentional deficits to FTD was significant, even though to a less extent. In clinically stabilized outpatients, they reported significant associations between the level of signal discrimination measured by the CPT, specifically on the degraded-stimulus CPT, and TDI factor scores for Fluid Thinking and Associative Disorganization during the inpatient period. They also observed that the outpatient signal discrimination deficits were significantly correlated with inpatient schizophrenic modes of thinking measured by the Rorschach Thought Disorder Index and with formal thought disorder measured by the BPRS Conceptual Disorganization rating.

Table 1 A list of included studies for associations between FTD and functioning outcomes

First author	Publication year	Published journal	Study groups	Sample size	Mean age	FTD measures	Outcome measures	Main findings and conclusion
Harrow, M	1986	Schizophrenia Bulletin	SCZ, other psychotic patients, and non-psychotic patients	191	23	Bizarre-idiosyncratic thinking	Occupational functioning, subsequent unemployment, relapse/re-hospitalisation rates (LKPS and SCS)	Subsequent unemployment and higher relapse/re-hospitalisation rates in schizophrenia patients experiencing enduring FTD
Marengo, JT	1987	Arch Gen Psychiatry	SCZ, other psychotic, and non-psychotic patients	191	23	Bizarre-idiosyncratic thinking	Occupational functioning, subsequent unemployment, relapse/re-hospitalisation rates (LKPS and SCS)	Patients with enduring thought disorder signs proved to be a poor-outcome group
Racenstein, JM	1999	J Nerv Ment Dis	SCZ, other psychotic, and non-psychotic patients	191	23	Bizarre-idiosyncratic thinking	Occupational functioning, subsequent unemployment, relapse/re-hospitalisation rates (LKPS and SCS)	FTD and functioning correlated in the first eight-year of schizophrenia. A stronger correlation between FTD and occupational than social functioning
Kotov, R	2016	Journal of Abnormal Psychology	First-admission inpatients with psychosis	628	30	SAPS and SANS	GAF, QLS (Social and role functioning), residential independency	A four-factors model (reality distortion, disorganization, inexpressivity, and apathy/asociality) having a stable and replicable validity in predicting outcomes
Minor, KS	2016	J Abnorm Psychol	Early stages of psychosis	38	24.89	CDI	GFS and GFR	Positive FTD and affective reactivity were consistently accounting for poor social functioning and associated with poor role functioning in some cases in EP
Roche, E	2016	Schizophrenia Research	FEP	680	33.42	Disorganization, verbosity, poverty of speech	Social and occupational functioning (MIRECC GAF subscales), number of hospitalisations	Higher baseline severity of disorganization predicted a greater number of hospitalisations and prolonged hospitalisation during the first year of illness

Table 1 (continued)

First author	Publication year	Published journal	Study groups	Sample size	Mean age	FTD measures	Outcome measures	Main findings and conclusion
Burton, CZ	2019	Schizophrenia Research	CLR, CHR, EFEP	327	16.69	SIPS	GFS and GFR	Baseline negative symptoms and thought disorder appeared to predict the functional outcome for up to 2 years among adolescents and young adults at risk for psychosis
Bearden, CE	2011	Journal of the American Academy of Child and Adolescent Psychiatry	CHR	105	16.66	Illogical thinking, poverty of content (POC), and referential cohesion	GFS and GFR	Transited to psychosis, predicted significantly social and role functioning at follow-up

FTD formal thought disorder, *CDI* communication disturbances index, *FEP* first episode psychosis, *EP* early psychosis, *CLR* clinical low risk, *CHR* clinical high risk, *UHR* ultra-high risk, *HR* high risk, *LR* low risk, *ARMS* at-risk mental state, *APS* attenuated psychosis syndrome, *APSS* attenuated positive symptom syndrome, *LKPS* levenstein, klein, and pollack scale, *SCS* strauss and carpenter scale, *OPCRIT* operational criteria for psychotic illness tool, *SAPS* scale for the assessment of positive symptoms, *SANS* scale for the assessment of negative symptoms, *CAARMS* comprehensive assessment of at-risk mental states, *SIPS* structured interview for prodromal syndromes, *SCZ* schizophrenia, *SOPS* scale of prodromal symptoms, *WERCAP* Washington early recognition center affectivity and psychosis screen, *GAF* global assessment of functioning, *GFS* global functioning scale-social *GFR* global functioning scale-role, *QLS* quality of life scale

Xu et al. [35] reported in a 1 year prospective study of language disorganization in patients with first-episode schizophrenia-spectrum disorders and investigated executive functions as a predictor of persistent FTD. They investigated the FTD using the Clinical Language Disorder Rating Scale (CLANG) [46] subdividing language abnormalities into syntactic, semantic, and production levels. They found that poorer performances in sustained attention and attention allocation/planning at illness onset were associated with an increased risk of having residual levels of semantic disorganization of language after 1 year, whereas poorer sustained attention was associated with increased risk of residual production problems. Pawelcyk et al. [47] evaluated pragmatic language functions in patients with first-episode psychosis, parents of the patients and healthy controls. Their results showed that the assessed groups varied in their ability to comprehend implicit information and to understand emotional prosody, as well as in processing language information regarding general knowledge, and in the effectiveness of interpersonal communication. More specifically, patients with first-episode psychosis performed significantly worse than healthy controls in all these neurocognitive domains. On the other hand, they found that the assessed groups did not differ regarding humour comprehension, understanding of linguistic prosody, understanding of both written and picture metaphors, or in their ability to process language information in the context of oral messages.

Caplan et al. [48] examined children with psychosis and investigated their use of discourse devices as well as the relation between FTD assessed with the K-FTDS and discourse deficits. They reported that patients not showing loose of associations used fewer words and conjunctions than controls. Furthermore, they observed that in the patients with loose of associations lexical cohesion correlated negatively with the total Intelligence Quotient (IQ) score, and with performance IQ, but not with verbal IQ. Remberk et al. [49] assessed the association between FTD, assessed through the TLC scale, and several neurocognitive domains in inpatients with early-onset schizophrenia-spectrum disorder and matched healthy controls. They assessed associations between psychopathological symptoms, cognitive functions, and FTD. The study showed that in patients FTD severity positively correlated with the number of non-perseverative errors on the Wisconsin Card Sorting Test (WCST), and with disturbances in semantic verbal fluency.

Ilonen et al. [34] explored a sample composed of clinical high-risk individuals (CHR), patients with psychosis and non-psychotic/non-CHR individuals. All individuals were administered a neuropsychological battery investigating verbal comprehension, perceptual organisation, working memory and processing speed, as well as measures of executive function and perceptual and thinking accuracy. They tested whether patients with CHR can be distinguished from

Table 2 A list of included studies for associations between FTD and neurocognition

First author	Publication year	Published journal	Study groups	Sample size	Mean age	FTD measures	Outcome measures	Main findings and conclusion
Minor, KS	2016	J Abnorm Psychol	EP	38	24.89	CDI	Single- and dual-task one-back visual working, memory tests	Affective, but not cognitive, systems play a critical role in positive FTD
Nuechterlein, KH	1986	Schizophrenia Bulletin	Early phase of SSD	32	22.3	The Rorschach TDI, the BPRS conceptual disorganization rating	CPT and forced-choice span of apprehension	The only significant correlations of the outpatient signal discrimination indices with inpatient positive symptoms were with conceptual disorganization
Xu, Jia-Qi	2014	Schizophrenia research	First-episode, SSD	60	25.28	CLANG	HSCT, MCT, modified SET, LNST, modified WCST	Poorer performances in sustained attention and attention allocation/planning at illness onset were associated with an increased risk of having residual semantic levels of language disorganization after one year
Pawelczyk, A	2018	Psychiatry research	FEP, HC, parents of FEP	34, 34, 32	20.85, 20.21, 49.44	The right, hemisphere language battery	TMT part A, TMT part B, DST–B, DST-F	Pragmatic dysfunctions may act as vulnerability markers of schizophrenia
Caplan, R	1992	Journal of the American academy of child and adolescent psychiatry	SCZ	31	10.2	The kiddie formal thought disorder rating scale	The Wechsler intelligence test for children revised	Lexical cohesion correlated negatively with full-scale IQ and performance IQ scores but not with the verbal IQ scores in the patients with loose of associations
Remberk, B	2012	Progress in Neuro-Psycho, pharmacology and Biological Psychiatry	SSD, HC	32, 32	16.7, 16.7	TLC, KRT	WCST, SVFT, PVFT, DST-B, DST-F	Thought disorder was correlated with executive dysfunction and disturbance in semantic verbal fluency

Table 2 (continued)

First author	Publication year	Published journal	Study groups	Sample size	Mean age	FTD measures	Outcome measures	Main findings and conclusion
Ilonen, T	2010	Psychiatry research	CHR, psychotics, non-psychotic/non-CHR individuals	22, 67, 187	15.6, 15.7, 15.5	The Rorschach, PTI	Verbal comprehension, perceptual organisation, working memory, processing speed, executive function, perceptual and thinking accuracy	The deficits were comparable in severity to those observed in adolescents with psychotic diagnoses and that patients at CHR for psychosis displayed mild-to-moderate executive impairment, without any impairment in intellectual functioning

FTD formal thought disorder, *CDI* communication disturbances index, *EP* early-stage psychosis, *FEP* first-episode psychosis, *CHR* clinical high risk, *HC* healthy control, *CPT* continuous performance test, *SCZ* schizophrenia, *SSD* schizophrenic spectrum disorders, *TDI* thought disorder index, *TLC* thought, language and communication scale, BPRS brief psychiatric rating scale, *CLANG* clinical language disorder rating scale, PTI perceptual thinking index, *KRT* Kent–Rosanoff test, *HSCT* Hayling sentence completion test, *MCT* monotone counting test, *SET* six element test, *LNST* letter number sequence test, WCST Wisconsin card sorting test, *TMT* trail making test, *DST-B* digit span test backward, *DST-F* digit span test forward, *SVFT* semantic verbal fluency test, *PVFT* phonological verbal fluency test

psychotic and non-psychotic/non-CHR individuals using neuropsychological tests. They found that adolescents with CHR displayed poorer visual form perception and thinking disorder compared to non-psychotic/non-CHR individuals. They reported that the deficits observed in CHR were comparable in severity to those observed in adolescents with psychotic diagnoses and that patients with CHR displayed mild-to-moderate executive impairment, but no impairment in intellectual functioning.

Discussion

Our literature survey showed that there was an increased interest for FTD in the early stages of psychosis from the late 70 s to late 80 s. Harrow et al. investigated in a few studies the course of psychosis after the first episode with a focus on FTD and formalized the following questions already in 1986 [32]: "Is thought disorder a frequent characteristic of acute schizophrenia? Does it persist in some or many patients with schizophrenia after the acute phase? And is it linked to other aspects of psychopathology? Does the presence of thought disorder after the acute phase predict subsequent clinical course and outcome?" These important questions reflect the core motivation for this systematic review covering the second half of the last century. Still, not all these questions have been answered. Even though the clinical importance of FTD has been observed from famous psychiatrists such as Kraepelin, Schneider, Bleuler, in the first half of the twentieth century, not so many studies in high-risk or stages of psychosis have been published due to methodological and phenomenological difficulties in the assessment of FTD with its multidimensional construct. [1, 2, 11, 12] Of note, early intellectual efforts on prevention and early diagnostics in psychiatry were started to be prominent only in the last 3 decades. Furthermore, the various measures of FTD among the included studies make a comparative interpretation of the associations between FTD and functioning and neurocognition difficult. Nevertheless, our review showed that FTD aspects linked to disorganization seem to be the most prominent in early psychosis, because of their consistent and replicated associations with both functioning and neurocognition aspects. Therefore, on the one hand, findings from this systematic review indicated that FTD, especially disorganization, might potentially have an early diagnostic and prognostic relevance in psychotic disorders, as (i) significant associations with functioning and cognition have been reported in both first-episode and clinical high-risk cohorts, and (ii) significant associations have been found in several subdomains of the functioning and neurocognition construct, spanning from general, to social, to role functioning, and hospitalization rate, for what concerns functioning, and from attention, executive function, and verbal IQ,

for what concerns neurocognition. On the other hand, the predictive and generalizability potential of such symptoms in terms of disease and risk trajectories has not been fully explored yet in the early stages of psychotic disorders, given the paucity of studies included. In light of these few, but consistent associations across studies, we think that findings from this review suggest the urgency of spending more efforts in understanding the role of FTD into the risk pathways of psychosis. This is further testified by recent studies showing that FTD dimensions affect 55% of those presenting with first-episode psychosis and are associated with acute clinical presentation, poor quality of life and worse therapeutic relationships. [19, 50, 51] Furthermore, FTD were associated positively with patients' unemployment risk [30], and negatively with their perceived quality of life [31] and overall life adjustment, based on indexes like work functioning, life disruptions and self-support. [30].

Moreover, given the significant associations between FTD, neurocognition and functioning impairments here reported, and given that these impairments very often precede the onset of full-blown psychosis and persist after the acute phase is resolved [24, 25], we speculate that FTD should be considered not only as a core psychosis characteristic but also as a feature of key importance to be targeted in early identification and intervention programs [19, 32]. This is further testified by the fact that six of the excluded studies (listed in Supplementary Table 1), although out of the review scope and thus not included in the "Results", highlighted significant associations between FTD and transition to psychosis in high-risk groups. Indeed, Bearden et al. [44] reported that illogical thinking, POC, and referential cohesion, distinguished putatively CHR individuals who transited to psychosis from those who did not transit. Similarly, Demjaha et al. [52] found that scores on the negative and on the disorganization/cognitive dimensions in the Comprehensive Assessment of At-Risk Mental States (CAARMS) were associated with a transition to psychosis during the 24-month-follow-up in an at-risk mental state (ARMS)-cohort. Thompson et al. [53] found that the presence of FTD, especially thought blocking and tangentiality, together with elevated mood, predicted transition to psychosis. Consistently, Devlyder et al. [54] found that disorganized communication (i.e., subthreshold thought disorder) was associated with an increased hazard for psychosis onset, both at baseline and as a trajectory of high persistent disorganized communication. Katsure et al. [55] found that converter CHR showed more severe symptom scores for unusual thought content, disorganized speech, and emotional disturbances items, compared to non-converter CHR. Mamah et al. [56] found a significant association between disorganized communication and associated with psychosis conversion. Finally, Brucato et al. [57] showed that SIPS/SOPS Unusual Thought Content and Disorganized Communication subscales, measures of attenuated odd delusions and thought disorder, were very good predictors of psychosis, whereas attenuated suspiciousness and perceptual abnormalities were not associated with conversion.

Consistently with findings from our review, disorganization seems to be the most notable FTD symptom predicting transition to psychosis in high-risk groups. The POC was another FTD symptom that commonly associated with social and role functioning as well as with transition to psychosis. Not only these findings are consistent with the associations between FTD, functioning and neurocognition we have discussed, but they also further speak in favor of the clinical relevance of FTD in the pathophysiology of psychosis and encourage future research to longitudinally investigate their role in the clinical trajectories of individuals at-risk for psychosis. A potential way to methodologically accomplish this future research direction could be represented by further efforts in designing computational methods that might help clinicians in evaluating risk categories of patients at their first presentation also through the presence of FTD, especially in the disorganization domain. For example, some machine learning and computational pattern recognition studies provided promising results in automated computational speech or text analyses to improve the clinical utility and objective quantifiability of FTD. The earlier effort has appeared already in 2009 in the literature: Strous et al. [58] showed that patients with schizophrenia could be classified from non-affected individuals with 83.3% accuracy based on written text characteristics. Consistently, Bedi et al. [59] reported that speech features derived from a latent semantic analysis could successfully predict subsequent transition to psychosis (followed up quarterly for up to 2.5 years) in CHR. Other findings from the same group showed that an automated machine learning speech classifier could discriminate the speech of recent-onset psychosis patients from that of healthy individuals with a cross-validated predictive accuracy of 79%. [60] Furthermore, Mota et al. [61] showed that speech disorganization measured by graph connectedness could correctly predict schizophrenia diagnosis at 6-month follow-up with 91.67% accuracy in patients undergoing first clinical contact for recent-onset psychosis and 21 well-matched healthy subjects. Despite promising, these findings need to be externally replicated in greater samples to fulfil the challenging state-of-art criterion of machine learning applications in computational psychiatry [62].

Nevertheless, overall, these finding and those from our systematic review would support the relevance of FTD for several psychosis-associated outcomes, from neurocognition, to functioning, to transition to psychosis, and that applying cutting-edge methodological techniques to further FTD investigations may be a promising way to deliver clinicians with time-efficient and objective clinical tools

supporting diagnostic and prognostic procedures. Many more future studies are warranted.

Limitations

This review has some limitation. Mainly, studies have employed different assessment strategies to evaluate FTD, neurocognition and functioning. As a matter of fact, this high variability did not allow us to conduct any direct comparative analysis. Therefore, the results in this review can be only partially interpreted. We also acknowledge that our results are limited only to the published studies and that there was no study excluded due to non-significant associations. Therefore, a possible bias through studies showing non-significant associations between FTD, neurocognition and functioning could not be excluded.

Conclusion

The reviewed studies showed that FTD severity is significantly associated with poor social functioning, unemployment, relapses of psychosis and re-hospitalisations, as well as transition to psychosis in high-risk groups. The results also showed significant associations between FTD and attention performance, executive functions, and verbal abilities. Machine learning algorithms show good potential for understanding the prognostic value of FTD in the risk trajectories of psychosis and encourage the development of computer-assisted early diagnostic tools targeting FTD, especially disorganization. Further studies taking advantage of the acceleration in computational psychiatry using methods such as unsupervised, supervised machine learning algorithms, deep learning techniques, natural language processing, sound, and rhythm analyses in records of patients' clinical evaluations, would hopefully allow researchers to develop novel clinical markers for FTD. Such automated and time friendly computer-assisted diagnostic tools could give researchers and clinicians the chance to re-investigate the clinical importance of FTD starting from high risk and early stage of psychosis, and to fill the gap in the literature that might open new avenues to develop targeted neuropsychotherapeutics specific to FTD.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00406-021-01295-3.

Funding Open Access funding enabled and organized by Projekt DEAL. Oemer Faruk Oeztuerk is supported by International Max Planck Research School for Translational Psychiatry in the frame of the Clinician Scientist Program funded by the Else-Kröner-Fresenius-Stiftung. Linda A. Antonucci's salary is funded by the Structural European Funding of the Italian Minister of Education (Attraction and International Mobility—AIM— action, Grant No. 1859959).

Declarations

Conflict of interest The authors state no conflict of interest.

References

1. Roche E, Creed L, Macmahon D, Brennan D, Clarke M (2015) The epidemiology and associated phenomenology of formal thought disorder: a systematic review. Schizophr Bull 41(4):951–962
2. Kircher T, Bröhl H, Meier F, Engelen J (2018) Formal thought disorders: from phenomenology to neurobiology. Lancet Psychiatry 5(6):515–526
3. Kircher T, Krug A, Stratmann M et al (2014) A rating scale for the assessment of objective and subjective formal thought and language disorder (TALD). Schizophr Res 160(1–3):216–221
4. Esquirol É (1838) Des Maladies Mentales Considérées Sous Les Rapports Médical, Hygiénique et Médico-Légal, Vol 1. chez JB Baillière
5. Griesinger W (1882) Mental pathology and therapeutics, vol 69. W. Wood & Company, London
6. Kraepelin E (1919) Dementia praecox and paraphrenia. Livingstone, London
7. Schneider C (1930) Die Psychologie der Schizophrenen. The psychology of schizophrenics. Thieme, Leipzig
8. Goldstein K (1939) Methodological approach to the study of schizophrenic thought disorder. Lang thought Schizophr 1944:17–40
9. Bleuler E (1950) Dementia Praecox or the Group of Schizophrenias, translated by J. Zinkin. International Universities Press, New York
10. Andreasen NC (1986) Scale for the assessment of thought, language, and communication (TLC). Schizophr Bull 12(3):473–482
11. Jerónimo J, Queirós T, Cheniaux E, Telles-Correia D (2018) Formal thought disorders – historical roots. Front Psychiatry. 9(572(November)):1–5
12. Hart M, Lewine RRJ (2017) Rethinking thought disorder. Schizophr Bull 43(3):514–522
13. Caissie SL, Liddle PF, Ngan ETC et al (2002) Thought and language index: an instrument for assessing thought and language in schizophrenia. Br J Psychiatry 181(04):326–330
14. Johnston MH, Holzman PS (1979) Assessing schizophrenic thinking: A clinical and research instrument for measuring thought disorder. Jossey-Bass
15. Kay SR, Fiszbein AL (1987) Positive and negative syndrome scale (Panss) rating criteria. Schizophr Bull 13(2):261–276. https://doi.org/10.1016/0006-3223(91)91039-t

16. Andreasen NC (1984) Scale for the assessment of positive symptoms (SAPS). University of Iowa, Iowa City
17. Andreasen NC (1989) The scale for the assessment of negative symptoms (SANS). Br J Psychiatry 155:49–52
18. Overall JE, Gorham DR (1962) The brief psychiatric rating scale. Psychol Rep 10(3):799–812
19. Roche E, Lyne J, O'Donoghue B et al (2016) The prognostic value of formal thought disorder following first episode psychosis. Schizophr Res 178(1–3):29–34
20. Roche E, Segurado R, Renwick L et al (2015) Language disturbance and functioning in first episode psychosis. Psychiatry Res 235:29–37
21. Andreasen NC, Grove WM (1986) Thought, language, and communication in schizophrenia: diagnosis and prognosis. Schizophr Bull 12(3):348–359
22. Gaebel W, Zielasek J (2015) Focus on psychosis. Dialogues Clin Neurosci 17(1):9
23. Stouten LH, Veling W, Laan W, van der Helm M, van der Gaag M (2017) Psychosocial functioning in first-episode psychosis and associations with neurocognition, social cognition, psychotic and affective symptoms. Early Interv Psychiatry 11(1):23–36
24. Hou C-L, Xiang Y-T, Wang Z-L et al (2016) Cognitive functioning in individuals at ultra-high risk for psychosis, first-degree relatives of patients with psychosis and patients with first-episode schizophrenia. Schizophr Res 174(1–3):71–76
25. Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F (2018) Course of clinical high-risk states for psychosis beyond conversion. Eur Arch Psychiatry Clin Neurosci 268(1):39–48
26. Insel TR, Cuthbert BN (2015) Brain disorders? precisely. Science (80-) 348(6234):499–500
27. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S et al (2018) Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiat 75(11):1156–1172
28. Iniesta R, Stahl D, McGuffin P (2016) Machine learning, statistical learning and the future of biological research in psychiatry. Psychol Med 46(12):2455–2465
29. Dwyer DB, Falkai P, Koutsouleris N (2018) Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 14(1):91–118
30. Marengo JT, Harrow M (1987) Schizophrenic thought disorder at follow-up. Arch Gen Psychiatry 44:651
31. Sigaudo M, Crivelli B, Castagna F et al (2014) Quality of life in stable schizophrenia: the relative contributions of disorganization and cognitive dysfunction. Schizophr Res 153(1–3):196–203
32. Harrow M, Marengo JT (1986) Schizophrenic thought disorder at followup: its persistence and prognostic significance. Schizophr Bull 12(3):373–393
33. Nuechterlein KH, Edell WS, Norris M, Dawson ME (1986) Attentional vulnerability indicators, thought disorder, and negative symptoms. Schizophr Bull 12(3):408–426
34. Ilonen T, Heinimaa M, Korkeila J, Svirskis T, Salokangas RKR (2010) Differentiating adolescents at clinical high risk for psychosis from psychotic and non-psychotic patients with the Rorschach. Psychiatry Res 179(2):151–156
35. Xu JQ, Hui CLM, Longenecker J et al (2014) Executive function as predictors of persistent thought disorder in first-episode schizophrenia: a one-year follow-up study. Schizophr Res 159(2–3):465–470
36. Moher D, Liberati A, Tetzlaff J, Altma DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097
37. Fusar-Poli P, Borgwardt S, Bechdolf A et al (2013) The psychosis high-risk state: a comprehensive state-of-the-art review. JAMA Psychiat 70(1):107–120
38. Harrow M, Grossman LS, Silverstein ML, Meltzer HY, Kettering RL (1986) A longitudinal study of thought disorder in manic patients. Arch Gen Psychiatry 43(8):781–785
39. Racenstein JM, Penn D, Harrow M, Schleser R (1999) Thought disorder and psychosocial functioning in schizophrenia: the concurrent and predictive relationships. J Nerv Ment Dis 187(5):281–289
40. Kotov R, Foti D, Li K, Bromet EJ, Hajcak G, Ruggero CJ (2016) Validating dimensions of psychosis symptomatology: neural correlates and 20-year outcomes. J Abnorm Psychol 125(8):1103–1119
41. Minor KS, Marggraf MP, Davis BJ, Mehdiyoun NF, Breier A (2016) Affective systems induce formal thought disorder in early-stage psychosis. J Abnorm Psychol 125(4):537–542
42. Docherty NM, DeRosa M, Andreasen NC (1996) Communication disturbances in schizophrenia and mania. Arch Gen Psychiatry 53(4):358–364. https://doi.org/10.1001/archpsyc.1996.01830040094014
43. Burton CZ, Tso IF, Carrion RE et al (2019) Baseline psychopathology and relationship to longitudinal functional outcome in attenuated and early first episode psychosis. Schizophr Res 212:157–162
44. Bearden CE, Wu KN, Caplan R, Cannon TD (2011) Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. J Am Acad Child Adolesc Psychiatry 50(7):669–680
45. Caplan R, Guthrie D, Fish B, Tanguay PE, David-Lando G (1989) The kiddie formal thought disorder rating scale: clinical assessment, reliability, and validity. J Am Acad Child Adolesc Psychiatry 28(3):408–416
46. Chen EYH, Lam LCW, Kan CS et al (1996) Language disorganisation in schizophrenia: validation and assessment with a new clinical rating instrument. Hong Kong J Psychiatry 6(1):4
47. Pawełczyk A, Łojek E, Żurner N, Gawłowska-Sawosz M, Pawełczyk T (2018) Higher-order language dysfunctions as a possible neurolinguistic endophenotype for schizophrenia: evidence from patients and their unaffected first degree relatives. Psychiatry Res 267(May):63–72
48. Caplan R (1992) Communication deficits and formal thought disorder in schizophrenic children. J Am Acad Child Adolesc Psychiatry 31(1):151–159
49. Remberk B, Namysłowska I, Rybakowski F (2012) Clinical and cognitive correlates of formal thought disorder in early onset schizophrenia. Act Nerv Super Rediviva 54(2):68–76
50. Cavelti M, Homan P, Vauth R (2016) The impact of thought disorder on therapeutic alliance and personal recovery in schizophrenia and schizoaffective disorder: an exploratory study. Psychiatry Res 239:92–98
51. Tan EJ, Thomas N, Rossell SL (2014) Speech disturbances and quality of life in schizophrenia: differential impacts on functioning and life satisfaction. Compr Psychiatry 55(3):693–698
52. Demjaha A, Valmaggia L, Stahl D, Byrne M, McGuire P (2012) Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. Schizophr Bull 38(2):351–359
53. Thompson A, Nelson B, Bruxner A et al (2013) Does specific psychopathology predict development of psychosis in ultra high-risk (UHR) patients? Aust N Z J Psychiatry 47(4):380–390
54. DeVylder JE, Muchomba FM, Gill KE et al (2014) Symptom trajectories and psychosis onset in a clinical high-risk cohort: the relevance of subthreshold thought disorder. Schizophr Res 159(2–3):278–283
55. Katsura M, Ohmuro N, Obara C et al (2014) A naturalistic longitudinal study of at-risk mental state with a 2.4 year follow-up at a specialized clinic setting in Japan. Schizophr Res 158(1–3):32–38
56. Mamah D, Musau A, Mutiso VN et al (2016) Characterizing psychosis risk traits in Africa: a longitudinal study of Kenyan adolescents. Schizophr Res 176(2–3):340–348

57. Brucato G, Masucci MD, Arndt LY et al (2017) Baseline demographics, clinical features and predictors of conversion among 200 individuals in a longitudinal prospective psychosis-risk cohort. Psychol Med 47(11):1923–1935
58. Strous RD, Koppel M, Fine J, Nachliel S, Shaked G, Zivotofsky AZ (2009) Automated characterization and identification of schizophrenia in writing. J Nerv Ment Dis 197(8):585–588
59. Bedi G, Carrillo F, Cecchi GA et al (2015) Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophr. https://doi.org/10.1038/npjschz.2015.30
60. Corcoran CM, Carrillo F, Fernández-Slezak D et al (2018) Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17(1):67–75
61. Mota NB, Copelli M, Ribeiro S (2017) Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. npj Schizophr 3(1):1–10
62. Sanfelici R, Dwyer DB, Antonucci LA, Koutsouleris N (2020) Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state of the art. Biol Psychiatry 88(4):349–360

References

Andreasen, N. C. (1989). The scale for the assessment of negative symptoms (SANS). *British Journal of Psychiatry*, *155*, 49–52.

Andreasen, N. C., & Grove, W. M. (1986). Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophrenia Bulletin*, *12*(3), 348–359.

Bearden, C. E., Wu, K. N., Caplan, R., & Cannon, T. D. (2011). Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(7), 669–680.

Buciuman, M.-O., Oeztuerk, O. F., Popovic, D., Enrico, P., Ruef, A., Bieler, N., Sarisik, E., Weiske, J., Dong, M. Sen, Dwyer, D. B., Kambeitz-Ilankovic, L., Haas, S. S., Stainton, A., Ruhrmann, S., Chisholm, K., Kambeitz, J., Riecher-Rössler, A., Upthegrove, R., Schultze-Lutter, F., … Koutsouleris, N. (2023). Structural and functional brain patterns predict formal thought disorder's severity and its persistence in recent-onset psychosis: Results from the PRONIA Study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/https://doi.org/10.1016/j.bpsc.2023.06.001

Burock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statictical Software*, *25*(4).

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(3), 223–230.

Caplan, R., Guthrie, D., Fish, B., Tanguay, P. E., & David-Lando, G. (1989). The kiddie formal thought disorder rating scale: clinical assessment, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *28*(3), 408–416.

Cavelti, M., Homan, P., & Vauth, R. (2016). The impact of thought disorder on therapeutic alliance and personal recovery in schizophrenia and schizoaffective disorder: An exploratory study. *Psychiatry Research*, *239*, 92–98.

Cavelti, M., Kircher, T., Nagels, A., Strik, W., & Homan, P. (2018). Is formal thought disorder in schizophrenia related to structural and functional aberrations in the language network? A systematic review of neuroimaging findings. *Schizophrenia Research*, *199*, 2–16.

Cavelti, M., Winkelbeiner, S., Federspiel, A., Walther, S., Stegmayer, K., Giezendanner, S., Laimböck, K., Dierks, T., Strik, W., Horn, H., & Homan, P. (2018). Formal thought disorder is related to aberrations in language-related white matter tracts in patients with schizophrenia. *Psychiatry Research - Neuroimaging*, *279*(November 2017), 40–50.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*.

Chen, E. Y. H., Lam, L. C. W., Kan, C. S., Chan, C. K. Y., Kwok, C. L., GH, N. D., & Chen, R. Y. L. (1996). Language disorganisation in schizophrenia: validation and assessment with a new clinical rating instrument. *Hong Kong Journal of Psychiatry*, *6*(1), 4.

Chen, J., Wensing, T., Hoffstaedter, F., Cieslik, E. C., Müller, V. I., Patil, K. R., Aleman, A., Derntl, B., Gruber, O., Jardri, R., Kogler, L., Sommer, I. E., Eickhoff, S. B., & Nickl-Jockschat, T. (2021).

Neurobiological substrates of the positive formal thought disorder in schizophrenia revealed by seed connectome-based predictive modeling. *NeuroImage: Clinical*, *30*.

Demjaha, A., Valmaggia, L., Stahl, D., Byrne, M., & McGuire, P. (2012). Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. *Schizophrenia Bulletin*, *38*(2), 351–359.

DeVylder, J. E., Muchomba, F. M., Gill, K. E., Ben-David, S., Walder, D. J., Malaspina, D., & Corcoran, C. M. (2014). Symptom trajectories and psychosis onset in a clinical high-risk cohort: The relevance of subthreshold thought disorder. *Schizophrenia Research*, *159*(2–3), 278–283.

Dey, A., Dempster, K., MacKinley, M., Jeon, P., Das, T., Khan, A., Gati, J., & Palaniyappan, L. (2021). Conceptual disorganization and redistribution of resting-state cortical hubs in untreated first-episode psychosis: A 7T study. *Npj Schizophrenia*, *7*(1).

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1).

Fu, Z., Tu, Y., Di, X., Du, Y., Pearlson, G. D., Turner, J. A., Biswal, B. B., Zhang, Z., & Calhoun, V. D. (2018). Characterizing dynamic amplitude of low-frequency fluctuation and its relationship with dynamic functional connectivity: An application to schizophrenia. In *NeuroImage* (Vol. 180, pp. 619–631). Academic Press Inc.

Fuentes-Claramonte, P., López-Araquistain, L., Sarró, S., Sans-Sansa, B., Ortiz-Gil, J., Maristany, T., Salvador, R., McKenna, P. J., & Pomarol-Clotet, E. (2021). Brain functional correlates of formal thought disorder in schizophrenia: Examining the frontal/dysexecutive hypothesis. *Psychological Medicine*, *51*(14), 2446–2453.

Harrow, M., & Marengo, J. T. (1986). Schizophrenic Thought Disorder at Followup: Its Persistence and Prognostic Significance. *Schizophrenia Bulletin*, *12*(3), 373–393.

Hart, M., & Lewine, R. R. J. (2017). Rethinking thought disorder. *Schizophrenia Bulletin*, *43*(3), 514–522.

Hoptman, M. J., Zuo, X. N., Butler, P. D., Javitt, D. C., D'Angelo, D., Mauro, C. J., & Milham, M. P. (2010). Amplitude of low-frequency oscillations in schizophrenia: A resting state fMRI study. *Schizophrenia Research*, *117*(1), 13–20.

Horn, H., Jann, K., Federspiel, A., Walther, S., Wiest, R., Müller, T., & Strik, W. (2012). Semantic network disconnection in formal thought disorder. *Neuropsychobiology*, *66*(1), 14–23.

Ilonen, T., Heinimaa, M., Korkeila, J., Svirskis, T., & Salokangas, R. K. R. (2010). Differentiating adolescents at clinical high risk for psychosis from psychotic and non-psychotic patients with the Rorschach. *Psychiatry Research*, *179*(2), 151–156.

Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine*, *46*(12), 2455–2465.

Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. *Science*, *348*(6234), 499–500.

Jablensky, A. (2000). Epidemiology of schizophrenia: the global burden of disease and disability. In *Eur Arch Psychiatry Clin Neurosci* (Vol. 250).

Jerónimo, J., Queirós, T., Cheniaux, E., & Telles-Correia, D. (2018). Formal Thought Disorders – Historical Roots. *Frontiers in Psychiatry*, *9:572*(November), 1–5.

Kay, S R Fiszbein A, L. a O. (1987). Positive and Negative Syndrome Scale (Panss) Rating Criteria. *Schizophrenia Bulletin*, *13*(2), 261–276. www.ncbi.nlm.nih.gov

Kircher, T., Bröhl, H., Meier, F., & Engelen, J. (2018). Formal thought disorders: from phenomenology to neurobiology. *The Lancet Psychiatry*, *5*(6), 515–526.

Kircher, T., Krug, A., Stratmann, M., Ghazi, S., Schales, C., Frauenheim, M., Turner, L., Fährmann, P., Hornig, T., Katzev, M., Grosvald, M., Müller-Isberner, R., & Nagels, A. (2014). A rating scale for the assessment of objective and subjective formal thought and language disorder (TALD). *Schizophrenia Research*, *160*(1–3), 216–221.

Kircher, T. T. J., Liddle, ; P F, Brammer, M. J., Williams, S. C. R., Murray, R. M., & Mcguire, P. K. (2001). *Neural Correlates of Formal Thought Disorder in Schizophrenia Preliminary Findings From a Functional Magnetic Resonance Imaging Study*.

Klosterkötter, J., Hellmich, M., Steinmeyer, E. M., & Schultze-Lutter, F. (2001). *Diagnosing Schizophrenia in the Initial Prodromal Phase*.

Klosterkötter, J., Schultze-Lutter, F., & Ruhrmann, S. (2008). Kraepelin and psychotic prodromal conditions. *European Archives of Psychiatry and Clinical Neuroscience*, *258*(SUPPL. 2), 74–84. https://doi.org/10.1007/s00406-008-2010-5

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., … Meisenzahl, E. (2021). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients with Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*, *78*(2), 195–209.

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., & Haidl, T. (2018). Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry*, *75*(11), 1156–1172.

Lord, E., Willems, M., Lapointe, F. J., & Makarenkov, V. (2017). Using the stability of objects to determine the number of clusters in datasets. *Information Sciences*, *393*, 29–46.

Marder, S. R., & Galderisi, S. (2017). *The current conceptualization of negative symptoms in schizophrenia*. *February*, 14–24.

Marengo, J. T., & Harrow, M. (1987). Schizophrenic Thought Disorder at Follow-up. *Arch. Gen. Psychiatry*, *44*.

Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J., & Beckmann, C. F. (2016). Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(5), 433–447.

McCutcheon, R. A., Reis Marques, T., & Howes, O. D. (2020). Schizophrenia - An Overview. In *JAMA Psychiatry* (Vol. 77, Issue 2, pp. 201–210). American Medical Association.

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. In *Epidemiologic Reviews* (Vol. 30, Issue 1, pp. 67–76).

Nuechterlein, K. H., Edell, W. S., Norris, M., & Dawson, M. E. (1986). Attentional vulnerability indicators, thought disorder, and negative symptoms. *Schizophrenia Bulletin*, *12*(3), 408–426.

Oeztuerk, O. F., Pigoni, A., Antonucci, L. A., & Koutsouleris, N. (2022). Association between formal thought disorders, neurocognition and functioning in the early stages of psychosis: a systematic review of the last half-century studies. In *European Archives of Psychiatry and Clinical Neuroscience* (Vol. 272, Issue 3, pp. 381–393). Springer Science and Business Media Deutschland GmbH.

Oeztuerk, O. F., Pigoni, A., Wenzel, J., Haas, S. S., Popovic, D., Ruef, A., Dwyer, D. B., Kambeitz-Ilankovic, L., Ruhrmann, S., Chisholm, K., Lalousis, P., Griffiths, S. L., Lichtenstein, T., Rosen, M., Kambeitz, J., Schultze-Lutter, F., Liddle, P., Upthegrove, R., Salokangas, R. K. R., … Koutsouleris, N. (2021). 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*.

Picchioni, M. M., & Murray, R. M. (2007). Schizophrenia. In *British Medical Journal* (Vol. 335, Issue 7610, pp. 91–95).

Remberk, B., Namysłowska, I., & Rybakowski, F. (2012). Cognitive impairment and formal thought disorders in parents of early-onset schizophrenia patients. *Neuropsychobiology*, *65*(4), 206–215.

Roche, E., Creed, L., Macmahon, D., Brennan, D., & Clarke, M. (2015). The Epidemiology and Associated Phenomenology of Formal Thought Disorder: A Systematic Review. *Schizophrenia Bulletin*, *41*(4), 951–962.

Roche, E., Lyne, J., O'Donoghue, B., Segurado, R., Behan, C., Renwick, L., Fanning, F., Madigan, K., & Clarke, M. (2016a). The prognostic value of formal thought disorder following first episode psychosis. *Schizophrenia Research*, *178*(1–3), 29–34.

Roche, E., Lyne, J., O'Donoghue, B., Segurado, R., Behan, C., Renwick, L., Fanning, F., Madigan, K., & Clarke, M. (2016b). The prognostic value of formal thought disorder following first episode psychosis. *Schizophrenia Research*, *178*(1–3), 29–34.

Roche, E., Segurado, R., Renwick, L., McClenaghan, A., Sexton, S., Frawley, T., Chan, C. K., Bonar, M., & Clarke, M. (2015). Language disturbance and functioning in first episode psychosis. *Psychiatry Research*, *235*, 29–37.

Sans-Sansa, B., McKenna, P. J., Canales-Rodríguez, E. J., Ortiz-Gil, J., López-Araquistain, L., Sarró, S., Dueñas, R. M., Blanch, J., Salvador, R., & Pomarol-Clotet, E. (2013). Association of formal thought disorder in schizophrenia with structural brain abnormalities in language-related cortical regions. *Schizophrenia Research*, *146*(1–3), 308–313.

Schultze-Lutter, F., Ruhrmann, S., Fusar-Poli, P., Bechdolf, A., Schimmelmann, B. G., & Klosterkötter, J. (2012). Basic Symptoms and the Prediction of First-Episode Psychosis. In *Current Pharmaceutical Design* (Vol. 18).

Sigaudo, M., Crivelli, B., Castagna, F., Giugiario, M., Mingrone, C., Montemagni, C., Rocca, G., & Rocca, P. (2014). Quality of life in stable schizophrenia: The relative contributions of disorganization and cognitive dysfunction. *Schizophrenia Research*, *153*(1–3), 196–203.

Stein, F., Buckenmayer, E., Brosch, K., Meller, T., Schmitt, S., Ringwald, K. G., Pfarr, J. K., Steinsträter, O., Enneking, V., Grotegerd, D., Heindel, W., Meinert, S., Leehr, E. J., Lemke, H., Thiel, K., Waltemate, L., Winter, A., Hahn, T., Dannlowski, U., … Kircher, T. (2022). Dimensions of Formal Thought Disorder and Their Relation to Gray- and White Matter Brain Structure in Affective and Psychotic Disorders. *Schizophrenia Bulletin*, 1–10.

Sumner, P. J., Bell, I. H., & Rossell, S. L. (2018a). A systematic review of task-based functional neuroimaging studies investigating language, semantic and executive processes in thought disorder. *Neuroscience and Biobehavioral Reviews*, *94*, 59–75.

Sumner, P. J., Bell, I. H., & Rossell, S. L. (2018b). A systematic review of the structural neuroimaging correlates of thought disorder. *Neuroscience and Biobehavioral Reviews*, *84*, 299–315.

Tan, E. J., Thomas, N., & Rossell, S. L. (2014). Speech disturbances and quality of life in schizophrenia: Differential impacts on functioning and life satisfaction. *Comprehensive Psychiatry*, *55*(3), 693–698.

Thompson, A., Nelson, B., Bruxner, A., O'Connor, K., Mossaheb, N., Simmons, M. B., & Yung, A. (2013). Does specific psychopathology predict development of psychosis in ultra high-risk (UHR) patients? *Australian and New Zealand Journal of Psychiatry*, *47*(4), 380–390.

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*(3), 511–528.

Tomasi, D., Shokri-Kojori, E., & Volkow, N. D. (2016). Temporal changes in local functional connectivity density reflect the temporal variability of the amplitude of low frequency fluctuations in gray matter. *PLoS ONE*, *11*(4).

van Os, J., & Kapur, S. (2009). Schizophrenia. In *The Lancet* (Vol. 374, Issue 9690, pp. 635–645). Elsevier B.V.

Vita, A., Dieci, M., Giobbio, G. M., Caputo, A., Ghiringhelli, L., Comazzi, M., Garbarini, M., Mendini, A. P., Morganti, C., Tenconi, F., Cesana, B., & Invernizzi, G. (1995). Language and thought disorder in schizophrenia: brain morphological correlates. *Schizophrenia Research*, *15*(3), 243–251.

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., Abdulkader, R. S., Abdulle, A. M., Abebo, T. A., Abera, S. F., Aboyans, V., Abu-Raddad, L. J., Ackerman, I. N., Adamu, A. A., Adetokunboh, O., Afarideh, M., Afshin, A., Agarwal, S. K., Aggarwal, R., … Murray, C. J. L. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1211–1259.

Wensing, T., Cieslik, E. C., Müller, V. I., Hoffstaedter, F., Eickhoff, S. B., & Nickl-Jockschat, T. (2017). Neural correlates of formal thought disorder: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *38*(10), 4946–4965.

Winkelbeiner, S., Cavelti, M., Federspiel, A., Kunzelmann, K., Dierks, T., Strik, W., Horn, H., & Homan, P. (2018). Decreased blood flow in the right insula and middle temporal gyrus predicts negative formal thought disorder in schizophrenia. *Schizophrenia Research*, 9–11.

Xu, J. Q., Hui, C. L. M., Longenecker, J., Lee, E. H. M., Chang, W. C., Chan, S. K. W., & Chen, E. Y. H. (2014). Executive function as predictors of persistent thought disorder in first-episode schizophrenia: A one-year follow-up study. *Schizophrenia Research*, *159*(2–3), 465–470.

Yu, R., Chien, Y. L., Wang, H. L. S., Liu, C. M., Liu, C. C., Hwang, T. J., Hsieh, M. H., Hwu, H. G., & Tseng, W. Y. I. (2014). Frequency-specific alternations in the amplitude of low-frequency fluctuations in schizophrenia. *Human Brain Mapping*, *35*(2), 627–637.

Zang, Y. F., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., Li-Xia, T., Tian-Zi, J., & Yu-Feng, W. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development*, *29*(2), 83–91.

Acknowledgements

The great scientific environment supporting translational research for my PhD in combination to my residency education in psychiatry in frame of a clinician-scientist program IMPRS-TP made this thesis possible. Hereby, I would like to express my special thanks to Prof. Dr. Andrea Schmitt and Prof. Dr. Elisabeth Binder.

I would like to thank my clinical supervisor Prof. Dr. Nikolaos Koutsouleris for his great support and choosing me for his research group as a PhD Student. I am very grateful for have being part of his team, his worldwide scientific network, and his pioneering publications that abled me to put together my clinical work focusing an early diagnostic and prevention with cutting edge computational research in precision psychiatry. He motivated me to explore formal thought disorder and helped me anytime, when I struggled on my way to this dissertation. His extraordinary success in science and medicine gave me a strong motivation for conveying my career as a clinician-scientist in modern psychiatry.

I would like to thank the PRONIA Munich Team and the PRONIA Consortium for this unique scientific sprit, for the excellent experience of working in a great multisided project and for being a very productive international research team. This precious experience to work as a study physician at Koutsouleris Group and for the PRONIA Consortium let me gain priceless experiences in clinical research in precision psychiatry and scientific collaboration.

I would like to thank Ms. Buciuman, for her great cooperation for our collaborative paper, for her patience and perseverance as well as scientific skills. Without her sustainable support, it would not be possible to make our second paper published.

I would like to thank my postdoc supervisor Assoc. Prof. Dr. Linda Antonucci for her patience and unlimited enduring support in very long writing and review processes of our publications. Her mentorship helped me to keep on going on those difficult PhD days.

I would like to thank Prof. Dr. Alkomiet Hasan for his review of my cumulative PhD Thesis and for being a inspiring role model. His extraordinary success in science and medicine has strongly motivated me to continue my career in modern psychiatry with an emphasis on evidence-based practice.

I would like to thank Prof. Dr. Peter Falkai for his boundness vision for research in psychiatry and ageless spirit to improve the psychiatric care. The unique scientific environment for translational psychiatry at the Department of Psychiatry and Psychotherapy of Ludwig-Maximilian University Hospital allowed me to complete an excellent residency education in psychiatry and psychotherapy and a very competitive PhD in medical sciences focusing on translational research in computational psychiatry.

Finally, this work would never have happened without the inexpressible and constant support and patience of my wife Léonie and my family, who never stopped being on my side on all those days.