

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

Paris Alexandros Laloucis^{*,**1,2**}, Stephen J. Wood^{**1,3,4**}, Lianne Schmaal^{**3,4**}, Katharine Chisholm^{**1,5**}, Sian Lowri Griffiths^{**1,2,**6****}, Renate L.E.P. Reniers^{**1,2,6**}, Alessandro Bertolino^{**7**}, Stefan Borgwardt^{**8,9**}, Paolo Brambilla^{**10,11**}, Joseph Kambeitz^{**12,**6****}, Rebekka Lencer^{**13**}, Christos Pantelis^{**14,**6****}, Stephan Ruhrmann^{**15**}, Raimo K. R. Salokangas^{**16**}, Frauke Schultze-Lutter^{**17,18,19**}, Carolina Bonivento^{**20**}, Dominic Dwyer^{**12**}, Adele Ferro^{**11**}, Theresa Haidl^{**15**}, Marlene Rosen^{**15**}, Andre Schmidt^{**8,**6****}, Eva Meisenzahl^{**17**}, Nikolaos Koutsouleris^{**12,**†****}, Rachel Upthegrove^{**1,2,21,**†****}, and the PRONIA Consortium^{*****}

¹Institute for Mental Health, University of Birmingham, Birmingham, UK; ²Centre for Human Brain Health, University of Birmingham, Birmingham, UK; ³Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, Australia; ⁴Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia; ⁵Department of Psychology, Aston University, Birmingham, UK; ⁶Institute of Clinical Sciences, University of Birmingham, Birmingham, UK; ⁷Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy; ⁸Department of Psychiatry, University of Basel, Basel, Switzerland; ⁹Department of Psychiatry and Psychotherapy, Center of Brain, Behavior and Metabolism, University of Lübeck, Germany; ¹⁰Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy; ¹¹Department of Neurosciences and Mental Health, IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; ¹²Department of Psychiatry and Psychotherapy, Ludwig Maximilians University, Munich, Germany; ¹³Department of Psychiatry, University of Münster, Münster, Germany; ¹⁴Melbourne Neuropsychiatry Centre, University of Melbourne, Melbourne, Australia; ¹⁵Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany; ¹⁶Department of Psychiatry, University of Turku, Turku, Finland; ¹⁷Department of Psychiatry and Psychotherapy, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany; ¹⁸Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Surabaya, Indonesia; ¹⁹University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland; ²⁰IRCCS "E. Medea" Scientific Institute, San Vito al Tagliamento (Pn), Italy; ²¹Early Intervention Service, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

*To whom correspondence should be addressed; 52 Pritchatts Road, B15 2SA, Birmingham, UK; e-mail: pal532@student.bham.ac.uk

[†]Joint Senior Authors.

[‡]See PRONIA consortium author list in [supplementary material](#).

Diagnostic heterogeneity within and across psychotic and affective disorders challenges accurate treatment selection, particularly in the early stages. Delineation of shared and distinct illness features at the phenotypic and brain levels may inform the development of more precise differential diagnostic tools. We aimed to identify prototypes of depression and psychosis to investigate their heterogeneity, with common, comorbid transdiagnostic symptoms. Analyzing clinical/neurocognitive and grey matter volume (GMV) data from the PRONIA database, we generated prototypic models of recent-onset depression (ROD) vs. recent-onset psychosis (ROP) by training support-vector machines to separate patients with ROD from patients with ROP, who were selected for absent comorbid features (pure groups). Then, models were applied to patients with comorbidity, ie, ROP with depressive symptoms (ROP+D) and ROD participants with sub-threshold psychosis-like features (ROD+P), to measure their positions within the affective-psychotic continuum. All models were independently validated in a replication sample.

Comorbid patients were positioned between pure groups, with ROP+D patients being more frequently classified as ROD compared to pure ROP patients (clinical/neurocognitive model: $\chi^2 = 14.874$; $P < .001$; GMV model: $\chi^2 = 4.933$; $P = .026$). ROD+P patient classification did not differ from ROD (clinical/neurocognitive model: $\chi^2 = 1.956$; $P = 0.162$; GMV model: $\chi^2 = 0.005$; $P = .943$). Clinical/neurocognitive and neuroanatomical models demonstrated separability of prototypic depression from psychosis. The shift of comorbid patients toward the depression prototype, observed at the clinical and biological levels, suggests that psychosis with affective comorbidity aligns more strongly to depressive rather than psychotic disease processes. Future studies should assess how these quantitative measures of comorbidity predict outcomes and individual responses to stratified therapeutic interventions.

Key words: psychosis/depression/transdiagnostic/machine learning/MRI/gray matter volume/comorbidity

Introduction

Treatments for mental illness are currently based on categorical structures built on patterns of syndromes and their course, rather than etiology.¹ The biological and clinical overlaps between these syndromes, and significant heterogeneity in outcomes, has become more apparent in recent years.²⁻⁴ Advancement in both pharmacological and psychotherapeutic interventions has stalled, potentially as a result of continued focus on invalid disease categories.^{5,6} The need for better treatments is particularly acute in psychosis and depression, which constitute major mental health challenges to the world's population.⁷⁻¹³ The legacy of a Jaspers based hierarchical approach to symptom structures suggests that positive psychotic symptoms are of primary importance within psychotic spectrum disorders.¹⁴ Yet the categorical and hierarchical division of psychotic disorders into affective and non-affective has been contested for decades,¹⁵ with a clear demonstration of the presence of affective symptoms in psychosis and psychotic symptoms in affective disorders.^{8,11,16}

Heterogeneity is particularly noticeable in early and developing stages of illness, with a high prevalence of affective symptoms across disorders.¹⁷ The comorbidity of depression in early psychosis has been largely regarded as secondary to the primary disorder (psychosis), reinforcing a categorical, hierarchical approach.^{18,19} There are increasing calls to use empirical evidence to develop alternative aetiological informed structures.²⁰ The use of multidimensional item response modeling to predict psychosis biotypes has been shown to transcend traditional diagnostic boundaries; with the suggestion of an underlying transdiagnostic dimension across psychotic diagnoses.²¹⁻²³ However, there are valid reasons why a categorical approach to mental illness has persisted; a significant number of individuals' presentation will "fit" within distinct categories of mental illness and the course and outcome of their treatment can be predicted from such diagnostic structures.²⁴ When disorders are fully formed, course, as well as symptom structure, enables clearer distinction between categories. Imaging studies also show shared areas of interest, including the hippocampus and cerebellum,²⁵ the prefrontal cortex and insula²⁶ and both depression and psychosis have been associated with heightened brain activation in regions central to emotional processing²⁶⁻²⁸ with similarities particularly prominent in the early stages of illness.²⁶⁻²⁹

The distinction between depression and schizophrenia is possible by structural brain data, but also, more challenging in the early stages of illness when symptoms and course are more heterogeneous.²⁹ Transdiagnostic processes of mental health disorders are descriptively transdiagnostic (ie, being present in multiple disorders, without regard to how or why) or mechanistically transdiagnostic (ie, reflecting neurobiological, physiological, or functional mechanisms).^{30,31} Both depression and psychosis are associated with transdiagnostic features of

working memory, executive functioning, and verbal fluency deficits.^{32,33} The importance of certain mechanistically transdiagnostic symptoms is potentially hidden in categorical structures, and they remain under-investigated.³⁴

Complex psychopathology and heterogeneity in developing mental health disorders present the opportunity of fuller exploration of the significance of potential transdiagnostic symptoms, to provide further insight into aetiopathogenetic pathways of symptoms and through this to advance diagnostic structures.^{35,36} However, novel approaches and powerful statistical tools such as machine learning techniques could help provide this deeper understanding by detecting complex patterns of data across diagnostic structures, and the delineation of shared and distinct features of these illnesses at the phenotypic and brain levels. This may inform the development of more precise differential diagnostic tools and improve the development of new treatments.³⁵

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

Materials and Methods

Study Design

Data were taken from the discovery and replication samples of the PRONIA study, an EU-FP7 funded seven-center study aiming to optimize candidate biomarkers for the prediction and staging of mental health disorders. Details of the PRONIA study sites, recruitment protocol and quality control procedures are described in a previous publication³⁷ and in **supplementary methods** (1.1, 1.2, 1.3, **tables S2 and S3**).

Inclusion and Exclusion Criteria

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

(6) any medical indication against MRI. ROP and ROD inclusion criteria can be found in [supplementary section 1.1](#).

Group Identification

Pure ROP: any ROP patient who had a Beck Depression Inventory-II (BDI-II) score of 13 or lower, which is indicative of absent or minimal depressive symptoms.^{38,39} Pure ROD: any ROD patient who had a Positive and Negative Symptom Scale (PANSS)⁴⁰ positive subscale score of no more than 7 and no Structured Interview of Psychosis-risk Syndromes positive (SIPS-P) severity score of 3 or more on any item.

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

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

MRI Imaging Data Acquisition, Quality Control, and Preprocessing

Participants underwent a multi-modal MRI protocol. A minimal harmonization protocol, which the MR sequences across the different scanners had to comply with is described in the [supplementary methods \(1.3\)](#). In the current study, T1-weighted structural MRI (sMRI) images of the participants were analyzed. The sMRI images of six healthy traveling volunteers who were scanned at all sites with same parameters were also analyzed as part of a calibration study. The images were processed using the open-source CAT12 toolbox (version r1155; <http://dbm.neuro.uni.jena.de/cat12/>) (see [supplementary methods 1.4](#)). Employing generalization theory,^{41,42} a between-site voxel reliability map (G coefficient map) was computed from the analysis of the GMV maps that were derived from the calibration study. During our neuroimaging-based machine learning analyses the G coefficient maps were used for reliability-based voxel masking.

Classification Models

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

A second model using GMV whole-brain voxel-wise data as features was developed in the pure ROD and ROP groups and then applied to the ROP+D and ROD+P groups. The developed clinical/cognitive and GMV models were combined by using decision values from the pure clinical and pure GMV models (to build a model that learns from the meta-data) in a stacking-based data fusion framework.^{43,44} Finally, all the models were applied to an independent replication sample.

Support Vector Machine Learning Analysis

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

Imbalanced learning was corrected for, by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes. A linear kernel was used with 11 C values (0.0156, 0.0312, 0.0625, 0.1250, 0.2500, 0.5000, 1, 2, 4, 8, and 16) to optimize the choice of C value and create an ensemble of predictive models to be applied to the CV₂ data to produce a single average robust prediction.

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

Finally, a stacking-based data fusion framework^{43–46} was used to examine whether the combination of the clinical/neurocognitive and the neuroimaging-based models would provide a superior classification accuracy (see [supplementary Methods 1.7](#)).

Independent Validation

All our models were validated in our independent replication sample ($N = 262$) (see [supplementary methods table S4](#)) which was collected at a different timescale. The same group identification criteria were applied.

Supplementary Analyses

A number of supplementary exploratory analyses, including correlation analyses between decision scores from our models, association and comparison analyses between correctly and misclassified patients, GMV comparison between groups, decision score group comparisons, and regression analyses with 9-month functional outcomes were conducted and can be found in [supplementary material](#) (Section 2.6).

Results

Demographic Information

Data from 154 participants with ROP and 146 patients with ROD were included in the analysis as our Discovery sample.

Thirty-eight ROP patients were included in the pure ROP group and 90 ROD patients in the pure ROD group. The mean age of the pure ROP group was 26.5 [SD 6.8] and the mean age of the pure ROD group was 26.5 [SD 6.6]. There were 25 male and 13 female patients in the pure ROP group and 45 male and 45 female patients in the pure ROD group.

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

The independent validation sample consisted of 161 patients with ROP and 131 patients with ROD. Fifty ROP patients were included in the pure ROP group and 53 ROD patients in the pure ROD group. The ROP+D and ROD+P groups consisted of 90 and 25 patients, respectively. A full description is provided in [supplementary table S4](#).

Machine Learning Analyses

Internal Validation of the Pure Group Differential Classifiers

Clinical and Neurocognitive Data. A repeated nested pooled cross-validation model with classifiers of clinical and cognitive variables predicted pure diagnostic groups with a balanced accuracy (BAC) of 79.3%; 95% CI [77.2, 82.3] and an area under the curve (AUC) of 0.86 ([table 2](#) and [figure 1A](#)). Assignment to the ROP category by the clinical classifier was driven by reduced scores in the RSA and elevated scores in the WSS, RSA, and SPIA. ROD group classification was informed by increased scores in the SPIA, WSS, together with reduced scores in the DSST. The contribution of the features was calculated by feature weights and by cross-validation ratio ([figure 2](#)).

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

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

Separability of Comorbid Groups. Clinical/Neurocognitive Data The trained pure classification system comprising the collection of 11 clinical/neurocognitive models generated by the repeated nested cross-validation scheme on pure

groups was then applied to the comorbid groups (ROP+D and ROD+P) to produce decision scores measuring ROP vs. ROD likeness. This model had a BAC of 62.5% and an AUC of 0.66. Misclassifications showed a directionality toward the ROD group, with 63% of ROP+D patients being classified as ROD; $Z = 1.276$, $P = 0.0385$) (see [table 2](#) and [figure 1A](#)). ROP+D patients were more frequently classified as ROD compared to pure ROP patients ($\chi^2 = 14.874$; $P < 0.001$). In contrast, the assignment precision of ROD+P and ROD patients did not differ ($\chi^2 = 1.956$; $P = 0.162$).

GMV Data. The trained pure classification system (comprising of 11 GMV models generated by the repeated nested cross-validation scheme on pure groups) was then applied to the comorbid groups (ROP+D and ROD+P) to produce decision scores measuring ROP vs. ROD likeness. This produced a BAC of 47.8% and AUC of 0.43. Misclassifications showed a directionality toward the ROD group, with 80.4% of ROP+D patients being classified as ROD; $Z = 0.713$, $P = 0.344$) (see [table 2](#) and [figure 1B](#)). Similarly to the clinical/neurocognitive model ROP+D patients were more frequently classified as ROD compared to pure ROP patients ($\chi^2 = 4.933$; $P = .026$). In contrast, the assignment precision of ROD+P and ROD patients did not differ ($\chi^2 = 0.005$; $P = .943$).

Stacking. When applied to the comorbid groups, the combined model predicted a diagnostic group with a BAC of 58.5% and an area under the curve (AUC) of 0.66 (see [table 1](#) and [figure 1C](#)).

Independent Validation. Application of our models to the independent validation sample replicated findings very well (pure clinical and neurocognitive model BAC 76.2; pure imaging model BAC 49.9; pure stacking model BAC 78.2). Full results from the independent validation analysis can be found in [supplementary table S5](#).

Supplementary Analysis. See [supplementary file S2](#) for additional exploratory analysis results.

Discussion

Using repeated nested cross-validation techniques we built classification models based on transdiagnostic clinical and neurocognitive features and GMV data, together with a combined model integrating all data modalities, to classify prototype diagnostic groups of ROD and ROP participants without comorbidity. Eighty-seven percent of patients with pure ROP and ROD were accurately ascribed to their diagnostic group. Applying this model to groups with comorbidity, 88% of patients with ROD and psychotic features were ascribed to their primary diagnostic group (depression) whereas only 37% of patients with ROP and depressive features were ascribed to their primary diagnostic group (psychosis). The shift of

Table 1. Sample Sociodemographics

	ROP Group	ROD Group	t/z ²	P	Pure ROP Group	Pure ROD Group	t/z/χ ²	P	ROP+D Group	ROD+P Group	t/z/χ ²	P
Sample sizes, No.	154	146			38	90			92	25		
Participants per site, No. (%)												
Basel	23 (7.7)	17 (5.7)			3 (2.3)	8 (6.9)			14 (12)	6 (5.1)		
Birmingham	10 (4.7)	10 (3.3)			4 (3.1)	0 (0)			10 (8.5)	2 (1.7)		
Cologne	27 (9)	27 (9)			4 (3.1)	19 (14.8)			22 (18.8)	3 (2.6)		
Milan	13 (4.3)	7 (2.3)	χ ² = 8.9	.257	8 (6.3)	5 (3.9)	χ ² = 21.0	.002	1 (0.9)	0 (0)	χ ² = 6.1	.517
Munich	46 (15.3)	47 (15.7)			10 (7.8)	35 (27.3)			31 (26.5)	7 (6)		
Turku	22 (7.3)	13 (4.3)			11 (8.6)	7 (5.5)			7 (6)	3 (2.6)		
Udine	12 (4)	21 (7)			2 (1.6)	12 (9.4)			3 (2.6)	3 (2.6)		
Age, mean (SD)	24.7 (5.4)	25.5 (6.1)			22.9	26.5 (6.6)			9.59	24.6 (4.8)	23.8 (3.9)	.449
Sex (Male/Female)	94/60	67/79	χ ² = 6.9	.009	25/13	45/45	χ ² = 2.6	.101	57/35	12/13	χ ² = 0.81	.368
Education, mean (SD)	13.9 (2.4)	15.1 (7.5)	t = 1.85	.064	13.9 (2.2)	15.8 (9.3)	t = 1.22	.224	14.0 (2.6)	13.5 (2.3)	t = -0.838	.404
Educational years repeated, mean (SD)	1.8 (2.5)	2.3 (2.7)	t = 1.47	.141	.83 (.87)	1.1 (1.8)	t = 1.00	.319	3.1 (4.8)	5.0 (5.4)	t = 1.70	.091
Having partnership most of the time in the year before study inclusion, No. (%)	72 (24.7)	85 (29.2)	χ ² = 4.1	.036	17 (13.4)	51 (40.2)	χ ² = 1.6	.194	48 (41.4)	18 (15.5)	χ ² = 2.9	.085
Population density in living area, mean (SD), inhabitants/km ²	3717.8 (2532.3)	3529.8 (2377.1)	t = -0.544	.587	4498.3 (2724.6)	3022.1 (2274.3)	t = -2.640	.010	3447.9 (2338.5)	5152.1 (2361.8)	t = 2.547	.013

Sample sizes, participants per study site, age, sex, education, partnership status, population density.

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

	True Positive, No.	True Negative, No.	False Positive, No.	False Negative, No.	Balanced Accuracy, %	AUC	Model P
Clinical/Neurocognitive Pure Model ROP-ROD	29	74	16	9	79.3	0.86	<.001
Applied Clinical/Neurocognitive Model ROP+D-ROD+P	36	19	6	56	57.6	0.71	NA
Applied Clinical and Neurocognitive Model Validation Pure ROP-ROD	34	39	14	16	70.8	0.78	NA
Applied Clinical and Neurocognitive Model Validation ROP+D-ROD+P	27	20	5	63	55	0.56	NA
GMV Pure Model ROP-ROD	15	77	13	23	62.5	0.7	<.001
Applied GMV Model ROP+D-ROD+P	19	20	5	73	50.3	0.47	NA
Applied GMV Model Validation Pure ROP-ROD	17	38	15	33	52.8	0.59	NA
Applied GMV Model Validation ROP+D-ROD+P	31	17	8	59	51.2	0.6	NA
Combined Model ROP-ROD	30	72	18	8	79.5	0.87	NA
Applied Combined Model ROP+D-ROD+P	49	19	6	43	64.6	0.71	NA
Applied Combined Model Validation Pure ROP-ROD	36	38	15	14	71.8	0.78	NA
Applied Combined Model Validation ROP+D-ROD+P	31	18	7	59	53.2	0.56	NA

comorbid psychosis patients towards the depression prototype was observed both at the clinical and biological levels. This suggests that when comorbid with affective symptoms, psychoses align more strongly to the disease processes of depressive than psychotic disorders. Using GMV measures only for classification, comorbid groups ROP patients with depressive symptoms largely resembled pure ROD. Results were generalizable to our independent validation sample.

Our findings suggest that clinical and neurocognitive transdiagnostic symptoms may have differential weight within psychosis and depression presentation with and without comorbidity. In pure groups, these symptoms could be seen as under the hierarchical umbrella of psychosis or depression, and may accurately reflect underlying pathology in these groups. However, in the face of comorbidity, transdiagnostic symptoms lean more to the depression domain. Given the prevalence and importance of depressive comorbidity in early psychosis this may support a model where depression could be more intrinsically important than is currently considered in early phases of illness.^{4,47} When participants with ROP exhibited even mild depressive symptoms, their GMV classification was more likely to lie within the depression group than psychosis suggesting a potential depressive biological phenotype that exists in the psychosis spectrum. The implications of these findings suggest that understanding heterogeneity of brain structures may need to include a specific focus on symptoms that may often be masked by more acute (eg, positive) symptoms and simple solutions that some symptoms may be transdiagnostic, potentially belie the complexity of individual etiology and psychopathology. Furthermore, these findings indicate a need to rethink current diagnostic classification to better reflect the biological reality and eventually develop better treatment options.

A classical diagnostic hierarchy in the structure of personal illness, reflected in current nosological classification systems, posits that mental disorders of the primary diagnosis have more weight and primacy over symptoms from lower classes, which are seen as secondary or comorbid.^{18,19} We found that comorbid symptoms affect diagnostic structures in different ways. Sub-clinical psychotic-like symptoms appeared to not alter the signature of ROD patients whilst depressive symptoms had a profound effect on ROP patients' classification accuracy.

Our findings add evidence to the debate around the validity of the system on which the DSM is built upon,⁴⁸ suggesting that a descriptivist position in the diagnosis of mental disorder is not sufficient and that a novel multivariate approach of mental disorder is more appropriate. Comorbidity in psychiatric disorders presents a clinical as well as nosological challenge. There may be a significant interplay between sets or clusters of symptoms over the development of disorder from prodrome via onset to

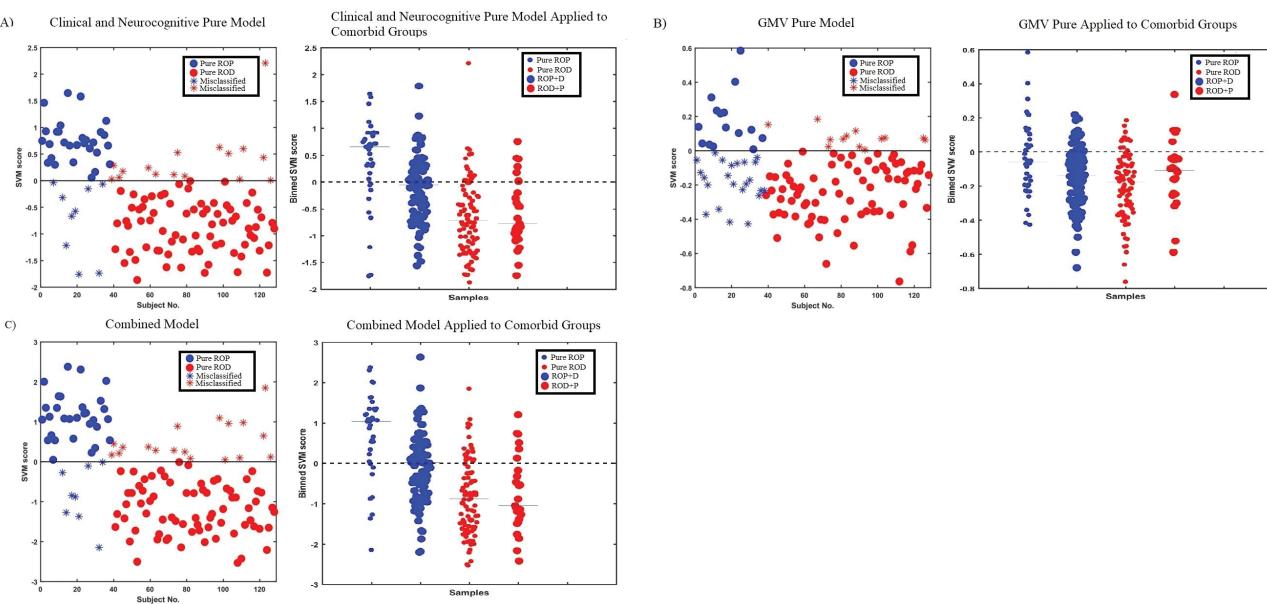


Fig. 1. Classification performance of the pure and applied clinical and neurocognitive, GMV, and combined models. (A) Pure clinical and neurocognitive classification balanced accuracy 79.3%, sensitivity 76.3%, specificity 82.2%, AUC 0.86. Applied Clinical classification balanced accuracy 57.6%, sensitivity 39.1%, specificity 76%, AUC 0.71. (B) Classification performance of the pure GMV model and the applied GMV model. Pure GMV classification balanced accuracy 62.5%, sensitivity 39.5%, specificity 85.6%, AUC 0.70. Applied GMV classification balanced accuracy 50.3%, sensitivity 20.7%, specificity 80%, AUC 0.47. (C) Classification performance of the combined model and the applied combined model. Stacked classification balanced accuracy 79.5%, sensitivity 78.9%, specificity 80%, AUC 0.87. Applied stacked classification balanced accuracy 64.6%, sensitivity 53.3%, specificity 76%, AUC 0.71.

potential chronicity. The frequency of depression within ROP may be a primary driver, rather than being a secondary symptom. If identified correctly, novel symptoms may be new targets that if treated effectively, ameliorate the other, eg, positive symptoms.

Recent onset disorders may constitute groups of phenotypically highly individual symptoms with underlying etiopathology, and it may be that personalized treatments could be tailored accordingly. Our SVM classification model found that different types of anhedonia (social and physical) were important in the classification of both psychosis and depression. Anhedonia has been suggested as a possible biomarker for depression⁴⁹ and has been found to be associated with decreased activation in ventral basal ganglia areas, the dorsal anterior cingulate, middle frontal gyrus, and medial frontal gyrus both in schizophrenia and depression.⁵⁰ Our GMV model revealed that orbitofrontal areas were higher weighted in the classification to the ROD group.

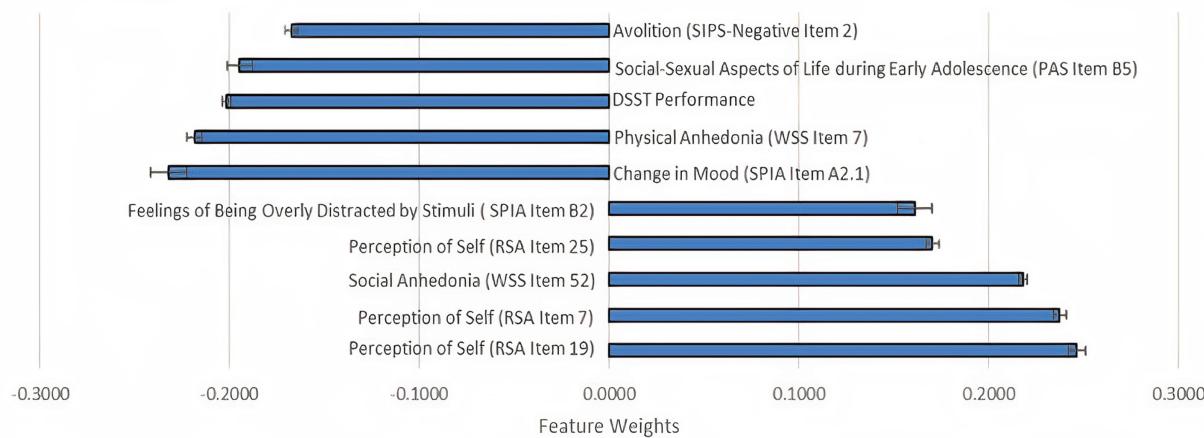
The relationship between psychotic and affective symptoms has been central to the dilemma of psychiatric classification. Substantial clinical and genomic evidence shows that schizophrenia and affective disorders may be distributed across a dimensional spectrum.⁵¹ However, the dimensional spectra model does not allow for either the clinical reality of a complex and changing symptom profile, nor the investigation of clinical features commonly seen across all disorders; some of which may be of primary importance. Concerning the neurobiology of

schizophrenia and depression, the majority of previous studies are based in subjects with depression or psychosis, but only rarely in both^{29,52} and not previously in highly mixed recent-onset comorbid disorders. Our results suggest that while GMV showed some distinction in prototype (pure) groups, when presented with complex comorbid groups, which may be the majority in clinical practice, there was a significant lack of any point of rarity between disorders. This builds on previous work suggesting distinction is more challenging, but also that neuroimaging based data do not support categorical classification in recent-onset disorders.²⁹ In recent onset disorders, early cognitive processes related to depression can both drive other more severe symptoms and/or be seen in isolation: for example, anhedonia could be an early indication of negative symptom clusters or a core feature of co-morbid depression.⁵³

Strengths and Limitations

The strengths of the present analysis include sufficiently large data, robust collection of clinical and imaging data from both depression and psychosis groups, independent validation analysis together with a novel approach to a challenging and essential clinically relevant research question, which speaks to the validity of diagnoses as the cornerstone of psychiatric practice. Our results however should be interpreted with certain caution due to the limitations of the study. Regarding our definition of the

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



B. Cross-Validation Ratio. Sum of the median weights across all CV1 folds divided by the standard error.

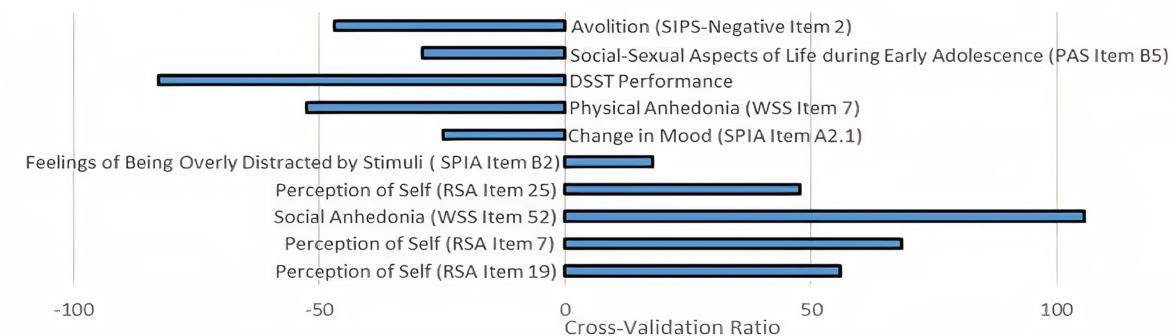


Fig. 2. Feature weights and cross-validations ratios of the most significant features. (A) Feature weights. Derived from 1000 random permutations of the outcome labels and features. (B) Cross-validation ratio. Sum of the median weights across all CV1 folds divided by the standard error.

ROD+P group, we used a SIPS-P item score of 2 or more which is not a marker of formal psychotic symptoms, and thus would only measure low levels of psychotic-like symptoms. However, to supplement this, we used an SPI-A COGDIS item score of 3 or more. We did not include core symptomatology measures such as the PANSS and the BDI in our features because primary groups are defined with these measures, and therefore including them would risk a circular analysis. Finally, there were more subjects identified in the pure ROD and comorbid ROP groups; ideally, groups of equal size would have been used. Nevertheless, we addressed this imbalance in our analysis, by increasing the C value in the minority class by multiplying it by the inverse ratio of the training class sizes.

Conclusions

Findings from this large, multi-modal, replicated machine learning classification study in recent-onset disorders suggest that whilst there may be a small subset

of prototypically pure individuals with clear categorical disorder, the majority of patients share a number of transdiagnostic features, primarily from the depression domain. The brain structure of psychosis patients with co-morbid depressive symptoms largely resembles that of depression. The increasing interest in heterogeneity of early disorders and transdiagnostic symptoms as novel treatment targets needs to be fully informed of potential depression related co-morbidity. Our analysis in recent-onset groups also highlights that both categorical and transdiagnostic approaches may ultimately fail at an individual patient level, as neither recognize the possibility of pluripotent pathways that are both stage and context-dependent. The implications, particularly for early intervention and prevention in mental health disorders is that ultimately a personalized medicine approach, encompassing the full potential of comorbidity, may be necessary to improve outcomes. Future studies should investigate the utility of targeting such transdiagnostic depression features

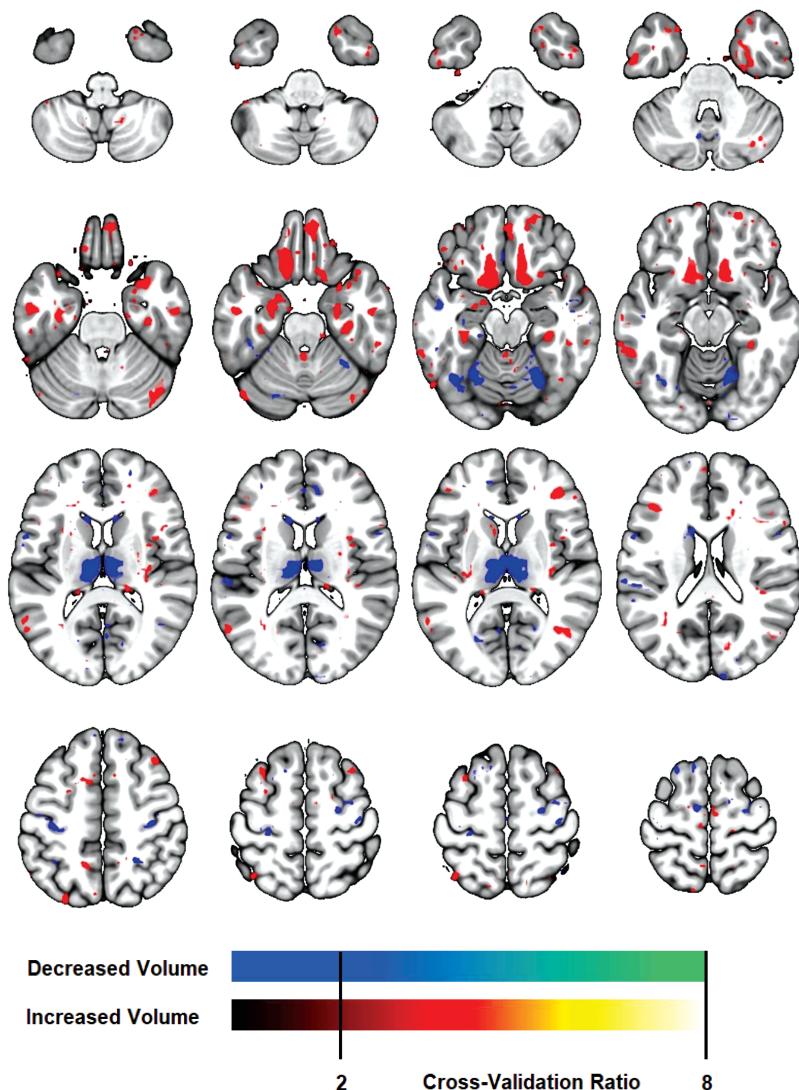


Fig. 3. Significant regions in the imaging classification model. ROP GMV reductions in the thalamus and the cerebellum, ROD GMV reductions in orbitofrontal, limbic, and paralimbic regions.

to elucidate their prognostic value, and develop new treatments.

Supplementary Material

Supplementary material is available at <https://academic.oup.com/schizophreniabulletin/>.

Funding

PRONIA is a Collaboration Project funded by the European Union under the 7th Framework Programme under grant agreement number 602152.

Acknowledgments

R.U. reports grants from Medical Research Council, grants from National Institute for Health Research:

Health Technology Assessment, grants from European Commission - Research: The Seventh Framework Programme, and personal speaker fees from Sunovion, outside the submitted work. C.P. has participated on Advisory Boards for Janssen-Cilag, Astra-Zeneca, Lundbeck, and Servier. He has received honoraria for talks presented at educational meetings organised by Astra-Zeneca, Janssen-Cilag, Eli-Lilly, Pfizer, Lundbeck and Shire. All other authors report no conflicts of interest.

References

- Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3(3):223–230.
- Kambeitz J, Kambeitz-Ilankovic L, Leucht S, et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology*. 2015;40(7):1742–1751.

3. Kambeitz J, Cabral C, Sacchet MD, et al. Detecting neuroimaging biomarkers for depression: a meta-analysis of multivariate pattern recognition studies. *Biol Psychiatry*. 2017;82(5):330–338.
4. Upthegrove R, Lalousis P, Mallikarjun P, et al. The psychopathology and neuroanatomical markers of depression in early psychosis. *Schizophr Bull*. 2021;47(1):249–258. doi:10.1093/schbul/sbaa094
5. Wong EH, Yocco F, Smith MA, Lee CM. Challenges and opportunities for drug discovery in psychiatric disorders: the drug hunters' perspective. *Int J Neuropsychopharmacol*. 2010;13(9):1269–1284.
6. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cognit Ther Res*. 2012;36(5):427–440.
7. Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of Schizophrenia: findings from the global burden of disease study 2016. *Schizophr Bull*. 2018;44(6):1195–1203. doi:10.1093/schbul/sby058
8. WHO | Depression and Other Common Mental Disorders. WHO. 2017. http://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/. Accessed May 20, 2019.
9. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat*. 2016;12:357–373.
10. Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jönsson B; CDBE2010 study group; European Brain Council. The economic cost of brain disorders in Europe. *Eur J Neurol*. 2012;19(1):155–162.
11. Conley RR, Ascher-Svanum H, Zhu B, Faries DE, Kinon BJ. The burden of depressive symptoms in the long-term treatment of patients with schizophrenia. *Schizophr Res*. 2007;90(1-3):186–197.
12. Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of depression in Europe. *J Ment Health Policy Econ*. 2006;9(2):87–98.
13. Dieleman JL, Baral R, Birger M, et al. US spending on personal health care and public health, 1996–2013. *JAMA*. 2016;316(24):2627–2646. doi:10.1001/jama.2016.16885
14. Bürgy M. The concept of psychosis: historical and phenomenological aspects. *Schizophr Bull*. 2008;34(6):1200–1210. doi:10.1093/schbul/sbm136
15. Quattrone D, Di Forti M, Gayer-Anderson C, et al.; EU-GEI WP2 Group. Transdiagnostic dimensions of psychopathology at first episode psychosis: findings from the multinational EU-GEI study. *Psychol Med*. 2019;49(8):1378–1391.
16. Ohayon MM, Schatzberg AF. Prevalence of depressive episodes with psychotic features in the general population. *Am J Psychiatry*. 2002;159(11):1855–1861.
17. Rickwood D, Van Dyke N, Telford N. Innovation in youth mental health services in Australia: common characteristics across the first headspace centres. *Early Interv Psychiatry*. 2015;9(1):29–37.
18. Birchwood M, Iqbal Z, Upthegrove R. Psychological pathways to depression in schizophrenia: studies in acute psychosis, post psychotic depression and auditory hallucinations. *Eur Arch Psychiatry Clin Neurosci*. 2005;255(3):202–212.
19. Foulds GA, Bedford A. Hierarchy of classes of personal illness. *Psychol Med*. 1975;5(2):181–192.
20. Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, Reed GM. Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychol Sci Public Interest*. 2017;18(2):72–145.
21. Maj M. The need for a conceptual framework in psychiatry acknowledging complexity while avoiding defeatism. *World Psychiatry Off J World Psychiatr Assoc WPA*. 2016;15(1):1–2. doi:10.1002/wps.20291
22. Tamminga CA, Ivleva EI, Keshavan MS, et al. Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry*. 2013;170(11):1263–1274.
23. Dwyer DB, Kalman JL, Budde M, et al. An investigation of psychosis subgroups with prognostic validation and exploration of genetic underpinnings: the PsyCourse study. *JAMA Psychiatry*. 2020;77(5):523–533.
24. Potuzak M, Ravichandran C, Lewandowski KE, Ongür D, Cohen BM. Categorical vs dimensional classifications of psychotic disorders. *Compr Psychiatry*. 2012;53(8):1118–1129.
25. Meisenzahl EM, Seifert D, Bottlender R, et al. Differences in hippocampal volume between major depression and schizophrenia: a comparative neuroimaging study. *Eur Arch Psychiatry Clin Neurosci*. 2010;260(2):127–137. doi:10.1007/s00406-009-0023-3
26. Busatto GF. Structural and functional neuroimaging studies in major depressive disorder with psychotic features: a critical review. *Schizophr Bull*. 2013;39(4):776–786.
27. Kumari V, Peters E, Guinn A, et al. Mapping depression in Schizophrenia: a functional magnetic resonance imaging study. *Schizophr Bull*. 2016;42(3):802–813.
28. Broome MR, He Z, Iftikhar M, Eyden J, Marwaha S. Neurobiological and behavioural studies of affective instability in clinical populations: a systematic review. *Neurosci Biobehav Rev*. 2015;51:243–254.
29. Koutsouleris N, Meisenzahl EM, Borgwardt S, et al. Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*. 2015;138(Pt 7):2059–2073.
30. Harvey AG, Murray G, Chandler RA, Soehner A. Sleep disturbance as transdiagnostic: consideration of neurobiological mechanisms. *Clin Psychol Rev*. 2011;31(2):225–235.
31. Harvey AG, Watkins E, Mansell W. *Cognitive Behavioural Processes Across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment*. New York: Oxford University Press; 2004.
32. Moritz S, Birkner C, Kloss M, et al. Executive functioning in obsessive-compulsive disorder, unipolar depression, and schizophrenia. *Arch Clin Neuropsychol*. Published online 2002;17:7.
33. Carrà G, Crocamo C, Bartoli F, et al. The mediating role of depression in pathways linking positive and negative symptoms in schizophrenia. A longitudinal analysis using latent variable structural equation modelling. *Psychol Med*. Published online March 8, 2019;50:1–9. doi:10.1017/S0033291719000321
34. Krueger RF, Eaton NR. Transdiagnostic factors of mental disorders. *World Psychiatry*. 2015;14(1):27–29.
35. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. 2018;14:91–118.
36. Klöppel S, Abdulkadir A, Jack CR Jr, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. *Neuroimage*. 2012;61(2):457–463.

37. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al.; PRONIA Consortium. 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):1156–1172.
38. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck depression inventories -IA and -II in psychiatric outpatients. *J Pers Assess*. 1996;67(3):588–597.
39. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II (BDI-II). Published online January 1, 1996. <https://www.scienceopen.com/document?vid=9feb932d-1f91-4ff9-9d27-da3bda716129>. Accessed February 24, 2020.
40. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276.
41. Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. *Behav Res Methods*. 2006;38(3):542–547.
42. Robert L. Brennan. Generalizability Theory and Classical Test Theory: Applied Measurement in Education: Vol 24, No 1. Published 2011. <https://www.tandfonline.com/doi/abs/10.1080/08957347.2011.532417>. Accessed May 20, 2019.
43. Wolpert DH. Stacked generalization. *Neural Netw*. 1992;5(2):241–259. doi:[10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
44. R. Polikar. Ensemble based systems in decision making. *IEEE Journals & Magazine*. Published 2006. <https://ieeexplore.ieee.org/document/1688199>. Accessed May 20, 2019.
45. Sabbagh D, Ablin P, Varoquaux G, Gramfort A, Engemann DA. Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states. *Neuroimage*. 2020;222:116893.
46. Karrer TM, Bassett DS, Derntl B, et al. Brain-based ranking of cognitive domains to predict schizophrenia. *Hum Brain Mapp*. 2019;40(15):4487–4507.
47. Upthegrove R, Marwaha S, Birchwood M. Depression and Schizophrenia: cause, consequence, or trans-diagnostic issue? *Schizophr Bull*. 2017;43(2):240–244.
48. Stein DJ, Phillips KA, Bolton D, Fulford KW, Sadler JZ, Kendler KS. What is a mental/psychiatric disorder? From DSM-IV to DSM-V. *Psychol Med*. 2010;40(11):1759–1765.
49. Der-Avakian A, Markou A. The neurobiology of anhedonia and other reward-related deficits. *Trends Neurosci*. 2012;35(1):68–77.
50. Zhang B, Lin P, Shi H, et al. Mapping anhedonia-specific dysfunction in a transdiagnostic approach: an ALE meta-analysis. *Brain Imaging Behav*. 2016;10(3):920–939.
51. Smeland OB, Bahrami S, Frei O, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. *Mol Psychiatry*. Published online January 4, 2019;25. doi:[10.1038/s41380-018-0332-x](https://doi.org/10.1038/s41380-018-0332-x)
52. Koutsouleris N, Gaser C, Jäger M, et al. Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. *Neuroimage*. 2008;39(4):1600–1612.
53. Barkus E, Badcock JC. A transdiagnostic perspective on social Anhedonia. *Front Psychiatry*. 2019;10:216.
54. Walsh-Messinger J, Jiang H, Lee H, Rothman K, Ahn H, Malaspina D. Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning. *Psychiatry Res*. 2019;278:27–34.
55. Jauhar S, Krishnadas R, Nour MM, Cunningham-Owens D, Johnstone EC, Lawrie SM. Is there a symptomatic distinction between the affective psychoses and schizophrenia? A machine learning approach. *Schizophr Res*. 2018;202:241–247.