**A framework for brain-derived dimensions of psychopathology**

Tristram A. Lett, Nilakshi Vaidya, Tianye Jia, Elli Polemiti, Tobias Banaschewski, Arun L.W. Bokde, Herta Flor, Antoine Grigis, Hugh Garavan, Penny Gowland, Andreas Heinz, Rüdiger Brühl, Jean-Luc Martinot, Marie-Laure Paillère Martinot, Eric Artiges, Frauke Nees, Dimitri Papadopoulos Orfanos, Herve Lemaitre, Tomáš Paus, Luise Poustka, Argyris Stringaris, Lea Waller, Zuo Zhang, Jeanne Winterer, Yuning Zhang, Michael N. Smolka, Robert Whelan, Ulrike Schmidt, Julia Sinclair, Henrik Walter, Jianfeng Feng, Trevor W. Robbins, Sylvane Desrivières, Andre Marquand, Gunter Schumann, IMAGEN Consortium, environMENTAL Consortium

**Word count: 2977**

**Centre for Population Neuroscience and Stratified Medicine (PONS), Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Germany** (Tristram Lett, PhD; Nilakshi Vaidya, MSc; Elli Polemiti, PhD; Prof Gunter Schumann, MD);

**Centre for Population Neuroscience and Precision Medicine (PONS), Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, China** (Tianye Jia, PhD; Prof Jianfeng Feng, PhD; Prof Trevor Robbins FRS, Prof Gunter Schumann, MD);

**Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany** (Prof Tobias Banaschewski, MD; Prof Frauke Nees, PhD);

**Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland (**Prof Arun Bokde, PhD**);**

**Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany** (Prof Herta Flor, PhD; Prof Frauke Nees, PhD);

**Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany** (Prof Herta Flor, PhD);

**NeuroSpin, CEA, Université Paris-Saclay, Gif-sur-Yvette, France** (Antoine Grigis, PhD; Dimitri Orfanos, PhD);

**Departments of Psychiatry and Psychology, University of Vermont, Burlington, Vermont, USA** (Prof Hugh Garavan, PhD);

**Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom** (Prof Penny Gowland, PhD);

**Department of Psychiatry and Psychotherapy CCM, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany** (Prof Andreas Heinz, MD; Lea Waller, BSc; Jeanne Winterer, BSc; Prof Henrik Walter, MD);

**Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany** (Rüdiger Brühl, PhD);

**Institut National de la Santé et de la Recherche Médicale, INSERM U1299 ‘Trajectoires développementales en psychiatrie’, Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS UMR9010, Centre Borelli, Gif-sur-Yvette, France** (Prof Jean-Luc Martinot, MD; Marie-Laure Paillère Martinot, MD);

**AP-HP, Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France** (Marie-Laure Paillère Martinot, MD);

**Institut National de la Santé et de la Recherche Médicale, INSERM U A10 "Trajectoires développementales en psychiatrie"; Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette; and Psychiatry Department, EPS Barthélémy Durand, Etampes, France** (Eric Artiges, MD);

**Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig-Holstein, Kiel University, Kiel Germany** (Prof Frauke Nees, PhD);

**Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, Bordeaux, France** (Herve Lemaitre, PhD);

**Department of Psychiatry, Faculty of Medicine and Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada** (Prof Tomáš Paus, MD);

**Departments of Psychiatry and Psychology, University of Toronto, Toronto, Ontario, Canada** (Prof Tomáš Paus, MD);

**Department of Child and Adolescent Psychiatry, Center for Psychosocial Medicine, University Hospital Heidelberg, Heidelberg, Germany** (Prof Luise Poustka, MD);

**Division of Psychiatry and Department of Clinical, Educational & Health Psychology, University College London, UK** (Prof Argyris Stringaris, MD);

**Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, United Kingdom** (Zuo Zhang, PhD; Prof Sylvane Desrivières, PhD);

**Department of Psychological Medicine, Centre for Research in Eating and Weight Disorders, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK** (Prof Ulrike Schmidt, MD);

**South London and Maudsley NHS Foundation Trust, London, UK** (Prof Ulrike Schmidt, MD);

**Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.**(Jeanne Winterer, BSc);

**Psychology Department, University of Southampton, Southampton. UK**

(Yuning Zhang PhD);

**Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany** (Prof Michael Smolka, MD);

**School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland** (Prof Robert Whelan, PhD);

**Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK** (Prof Julia Sinclair, FRCPsych);

**Department of Psychology and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK** (Prof Trevor Robbins, FRS);

**Donders Institute/Radboud UMC, Nijmegen, the Netherlands** (Prof Andre Marquand, PhD);

**Department of Psychiatry, University of Cambridge, Cambridge, UK** (Prof Gunter Schumann, MD);

Correspondence to:

Dr. Tristram Lett, Centre for Population Neuroscience and Stratified Medicine (PONS), Department of Psychiatry and Neuroscience, Charité Universitätsmedizin Berlin, 10117, Germany

**tristram.lett@charite.de**

Prof. Gunter Schumann, Centre for Population Neuroscience and Precision Medicine (PONS), Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, 200437, China and Department of Psychiatry and Neuroscience, Charité Universitätsmedizin Berlin, 10117, Germany

**gunter\_schumann@fudan.edu.cn;** [**gunter.schumann@charite.de**](mailto:gunter.schumann@charite.de)

**Key Points**

**Question:** Can existing psychiatric assessments be enhanced by multimodal brain neuroimaging to create neurobiological dimensions of psychopathology?

**Findings:** In this longitudinal population-based study, we identified six psychopathology scores derived from ICD-10/DSM-5 clinical symptoms that are defined by shared brain mechanisms characterized by brain structure, function, and connectivity (*r* range, .19 to .40; p<.0083).

**Meaning:** By identifying symptom groups that are specifically linked to quantifiable neurobiological measures, our approach enables the development of precise interventions that target biological mechanisms of psychiatric disorders and allows for quantitative assessment of comorbidity.

**Abstract**

**Importance:** Psychiatric diagnoses are not defined by neurobiological measures hindering the development of therapies targeting mechanisms underlying mental illness. Research confined to diagnostic boundaries yields heterogeneous biological results, whereas transdiagnostic studies often investigate individual symptoms in isolation.

**Objective:** We propose a framework that groups clinical symptoms compatible with ICD-10/DSM-5 according to their covariation and shared brain mechanisms.

**Design:** The study was conducted in two samples, the population-based, the IMAGEN Cohort (longitudinal assessments at 14, 19, and 23 years; study duration from 2010 to the present) and the cross-diagnostic STRATIFY/ESTRA samples (study duration from 2016 to 2023).

**Setting:** The samples are from multicenter studies from eight clinical research hospitals in: Germany, UK, France, and Ireland.

**Participants:** For the population-based IMAGEN study, 794 of 1253 23-year-old participants had complete assessments including complete clinical assessments and neuroimaging data across all timepoints. For the cross-diagnostic STRATIFY/ESTRA samples, 209 of 485 had complete clinical and neuroimaging data. The sample included healthy controls, and patients with alcohol-use disorder, major depressive disorder, anorexia nervosa, and bulimia nervosa.

**Exposure:** We employ sparse generalized canonical correlation analysis to integrate diverse data from clinical symptoms and seven brain imaging modalities.

**Main Measures:** The prediction of symptom features was our main outcome. The model was developed in the training from IMAGEN Study at age 23 (70%), then applied in the remaining holdout test sample (30%), the independent STRATIFY/ESTRA patient sample, and longitudinally in IMAGEN.

**Results:** In total, 1003 participants were included (425 male; 578 female) with the age of STRATIFY/ESTRA subjects ranging from 18-26 years. The reassembly of existing ICD-10/DSM-5 symptoms reveals six cross-diagnostic psychopathology scores related to excitability and impulsivity, depressive mood and distress, emotional and behavioral dysregulation, stress pathology, eating pathology, and social fear and avoidance symptoms. They were consistently associated with multimodal neuroimaging components in the training sample (r=0.26 to 0.40, p<0.0083), the independent test sample (r=0.14 to 0.26, p<0.05), and in psychiatric patients (r=0.12 to 0.19, p<0.05).

**Conclusion and Relevance:**

The identification of symptom groups of mental illness robustly defined by precisely characterized brain mechanisms enables the characterization of dimensions of psychopathology based upon quantifiable neurobiological measures.

**Introduction**

There has been a growing imperative within psychiatric neuroscience to uncover the biological mechanisms underlying mental health and disease to develop more effective treatments1. A major challenge lies in the classification of psychiatric disorders since their categorization does not follow biological mechanisms. Biological links distinguishing diagnostic criteria, including brain structure2, function3, and connectivity4, are limited, pointing to shared neurobiological substrates across mental illnesses. Dysfunctions within one mechanism affect the clinical presentation of more than one diagnosis, giving rise to comorbidity5. For a more nuanced understanding of psychiatric symptoms to be achieved, objective means of patient stratification and identification of robust psychiatric biomarkers are needed.

This need is perhaps most apparent in the efforts of biology-driven initiatives, like the National Institute of Mental Health's Research Domain Criteria (RDoC) framework1. RDoC aims to provide data about biological and behavioral processes related to mental-health and mental-illness. It is not designed to categorize psychiatric disorders. The Hierarchical-Taxonomy-Of-Psychopathology (HiTOP) maintains a clinical characterization applying clinical spectra and hierarchy6. HiTOP constructs are not driven by the biology underlying psychiatric liability. A unifying framework that considers the complex biological variation and the clinical variation concurrently to characterize nosology is warranted.

A potential solution to this challenge is to use existing clinical measures to optimize the link between symptoms and biology – which may lead to the discovery of novel biomarkers and targets for treatment development. We employed a data-driven strategy to integrate information from multiple domains, including clinical symptoms, brain structure (white-matter fractional anisotropy, cortical thickness, and surface area), as well as intrinsic (resting-state fMRI) and extrinsic (task fMRI) brain function. Our model integrates distinct, multimodal neuroimaging features, revealing their linear relationships with shared psychiatric symptoms across different disorders. We characterized in a single statistical model a wide variety of psychiatric symptoms and their covariance with a comprehensive multimodal characterization of the brain and established the reproducibility of our model by validating it in two samples with similar clinical and neuroimaging assessments: the longitudinal population-based IMAGEN study and its clinical follow-up study, STRATIFY/ESTRA. Our analysis aims at a novel framework that combines clinical usefulness with biological validity by harnessing current clinical assessments and quantifiable neurobiological measures, such as comprehensive functional and structural neuroimaging data.

**Methods**

**Study Design and Participants**

The analysis was carried out on 794 adolescents (366 male; 428 female) from the population-based IMAGEN-cohort, with neuroimaging assessments at 14, 19, and 23 years with an additional psychological assessment at 16 years7. 209 (59 male; 150 female) participants were included from the cross-disorder STRATIFY/ESTRA clinical sample of patients with major depressive disorder, alcohol-use disorder, anorexia nervosa, bulimia nervosa, and healthy controls, whose assessments were similar to IMAGEN at age 23. All studies received ethical approval and obtained informed consent (for detailed information, see Supplementary Methodology). Our training model was established in IMAGEN participants at 23 years, and applied to the test sample, earlier IMAGEN neuroimaging assessments, and the STRATIFY/ESTRA sample. All participants were self-reported as Western Europeans. Details of the cohorts are available in the supplementary material. Sample sizes are lower than initial recruitment since we only analyzed participants with complete clinical assessments of the Development-and-Well-Being-Assessment (DAWBA)8, Strengths-and-Difficulties-Questionnaire (SDQ)9, Alcohol-Use-Disorders-Identification-Test (AUDIT)10, and neuroimaging data including T1-weighted structural MRI, diffusion weighted images, resting state and task-based functional MRI (fMRI).

**Neuroimaging procedures**

Magnetic resonance imaging acquisition and processing were performed according to IMAGEN guidelines, and details are available in supplementary materials.

**Statistical analysis**

Details of the sparse generalized canonical correlation analysis (SGCCA) model and its derivation have been described elsewhere11,12. Sparse canonical correlation analysis is common among neuroimaging analyzes13–15. Prior to inclusion into the SGCCA model, each data view was corrected for age at time of MRI scan, sex, and site. Details of model parameters, optimization, and application are available in supplementary materials. SGCCA uses cross-covariance matrices of two or more sets of vectors (or data-views) to find the linear combinations (or components) of these data-views (clinical or neuroimaging data) that have maximum correlation with each other using gradient descent. We include all data-views into a unified model; thus, describing multimodal functional, structural, and diffusion MRI relationships in the context of cross-disorder symptom-scores (Figure 1).

**Results**

We established an optimized SGCCA model (Figure 1), reducing the number of collinear variables in our data-views while maximizing the variance explained. The optimal L1 sparsity for all data-views was lambda = 0.3 after 1000 permutations at each of the 10 steps (z-statistic = 12.6, eFigure 3A). We selected ten components as the point at which the cumulative average variance explained (AVE) of the full model levels off at 40.4% (eFigure 3B). The stability selection was performed by randomly selecting 50% of the training data without replacement 10000 times and retaining the clinical items, brain regions, and resting-state brain mode connectivity variables that appeared in 90% of the subsampled SGCCA models (eFigure 3CD). The final model (selected variables, ten components, and λ1=1.0) explained 52.7% of the variance among all data-views (eFigure 3E).

In the training data, ten canonical components we investigated were significant (Z=4.4 to 31.9, AVEinner=0.025 to 0.037, ppermuted<1.0x10-4, eFigure 3F). In the test data, the first six models remained significant (Z=1.8 to 10.3, AVEinner=0.008 to 0.017, ppermuted=0.048 to <1.0x10-4, eFigure 3F). Since the six components were significant for the model’s inner canonical correlation, we consider these to be components of interest. For an overview of the contribution of the composition of individual clinical items to psychopathology scores for the six components of interest, we calculated the mean DAWBA clinical subdomains, the AUDIT, and SDQ subscales for the structural coefficients (correlation between each psychopathology score and clinical items). The training, test, and STRATIFY/ESTRA samples were similar in terms of psychopathology (Figure 2). Based on these values, we categorized the psychopathology scores as: excitability and impulsivity, depressive mood and distress, emotional and behavioral dysregulation, stress pathology, eating pathology, and social fear and avoidance symptoms of components one to six, respectively (Figure 3).

Using SGCCA-regression to evaluate which neuroimaging scores were contributing to psychopathology scores, we found that each of the six symptom component scores predicted their corresponding neuroimaging components scores in the training, test and STRATIFY/ESTRA samples (all pbootstrapped<0.0083; Figure 4A) except for the stress pathology which was only nominally associated in the STRATIFY/ESTRA sample (pbootstrapped=0.026). The canonical correlations were moderate to low for: excitability and impulsivity (rtraining=0.26[0.18-0.33], rtest=0.22[0.10-0.35], rSTRATIFY/ESTRA=0.19[0.07-0.31]), depressive mood and distress (rtraining=0.30[0.20-0.38], rtest=0.22[0.09-0.35], rSTRATIFY/ESTRA=0.19[0.04-0.33]), emotional and behavioral dysregulation (rtraining=0.40[0.31-0.48], rtest=0.17[0.14-0.36], rSTRATIFY/ESTRA=0.19[0.06-0.30]), stress pathology (rtraining=0.32 [0.19-0.43], rtest=0.14[0.05-0.23], rSTRATIFY/ESTRA=0.12[0.01-0.22]), eating pathology (rtraining== 0.34[0.25-0.42], rtest=0.26[0.15-0.37], rSTRATIFY/ESTRA=0.15[0.12-0.34]), and social fear and avoidance symptoms (rtraining=0.31[0.25-0.42], rtest=0.18[0.15-0.35], rSTRATIFY/ESTRA=0.12[0.12-0.33]). This relationship was generally consistent in the IMAGEN sample at age 14 and age 19 (eFigure 4). Replicating the association in the test and STRATIFY/ESTRA data validates the prediction of clinical variates by the neuroimaging variates, since this data constitutes only independent, transformed scores from the SGCCA model.

Next, we asked which variables are driving the association between psychopathology scores and neuroimaging modality scores. We identified which coefficients (neuroimaging modality scores) were significantly associated with the psychopathology scores (pbootstrapped<0.0083, Figure 4B). Our SGCCA model is not limited to positive covariance since that would assume *a priori* that the direction of our neuroimaging values is better or worse clinically. Therefore, careful interpretation is needed since the coefficient’s direction can be negative or positive. The significant neuroimaging modality scores are correlated back to their corresponding data. This step is important for identifying which clinical items are most linked to brain regions; information that could be used to develop a parsimonious model to be applied in a clinical setting.

The psychopathological variables contributing to the excitability and impulsivity symptoms were primarily negatively associated with DAWBA items from the bipolar disorder section (PFDR<0.05; Figure 3). The stop-signal task (SST) negatively correlated in areas involved in frontoparietal executive function which mirrored the surface area correlation in the dorsolateral-prefrontal cortex, anterior-cingulate and inferior-parietal cortex (PFDR<0.05; Figure 5).

The depressive mood and distress score was correlated to DAWBA items in opposing directions for the bipolar and major depressive disorder sections. The score was negatively correlated with bipolar items related to “full of energy”, “more active”, “elevated mood”, and positively correlated with depressive items such as “miserable daily”, “impact of depression”, “tired or low energy”, “feelings of worthless guilt” (PFDR<0.05; Figure 3). Since the excitability and impulsive symptoms score is orthogonal to this score, we consider depressive features to be the defining feature of this component. The monetary incentive delay (MID) and SST scores both correlated with anterior and posterior-cingulate cortex activation, but in opposing directions (PFDR < 0.05; Figure 5).

The emotional and behavioral dysregulation score is positively correlated with questions related to poor concentration, impact on learning, and distress from the general anxiety disorder DAWBA section with additional correlations from social fears, depression, and panic attack items (PFDR<0.05; Figure 3). The insula and superior temporal gyrus were contributing to the emotional face task (EFT), MID, and SST with the latter correlating with medial-prefrontal cortex activation. The task loadings would suggest a common involvement of salience and ventral attention networks (PFDR<0.05; Figure 5).

The stress pathology score was primarily negatively correlated with items from the post-trauma stress disorder DAWBA questions (PFDR<0.05; Figure 3). The SST score correlated with activation in the dorsal anterior-cingulate cortex, insula, as well as the pre/post- central gyrus suggesting an involvement with default mode and salience networks (PFDR<0.05; Figure 5). The cortical thickness score correlated with thickness in the anterior and posterior cingulate, orbitofrontal, dorsolateral prefrontal, and insular cortices (PFDR<0.05; Figure 5).

The eating pathology score correlated with DAWBA items related to bulimia nervosa rather than anorexia nervosa; although, both were present (PFDR<0.05; Figure 3). The MID score correlated with activation in the striatum, medial-prefrontal cortex, and medial-temporal cortex suggesting an involvement of limbic and anterior salience networks (PFDR<0.05; Figure 5). The resting state brain modes score correlated with negative relationship between the high visual and language networks (medial temporal), and a positive relationship between the language network and the ventral default mode network (PFDR<0.05; Figure 5).

The clinical contribution for the social fear and avoidance score were split among social anxiety and panic attack items, indicating a specificity for social anxiety that is differentiated from physical panic symptoms. The resting state brain modes score correlated with negative relationship among the dorsal default mode network and both the visual and sensorimotor networks in the resting state brain modes (PFDR<0.05; Figure 5).

**Discussion**

We have developed a framework to characterize dimensions of psychopathology based on neurobiological measures. By constructing six symptom groups according to covariance and shared structural and functional neuroimaging features across seven modalities, we have provided mechanistic characterization and identified possible targets for therapeutic intervention. The neuroimaging correlates identified are specific to their symptom group, thus providing precise biomarkers and intervention targets. As our clinical characterization contains ICD-10/DSM-5 symptoms that were reassembled in a manner informed by their shared underlying biology, we preserved the clinical experience accumulated in existing psychopathological characterizations while optimizing them for neurobiological prediction.

The ability to link major psychiatric symptom groups to distinct neuroimaging modalities helps with biological understanding by providing quantifiable measures that are specific to each symptom group. Each component has biological characteristics that are independent of the other components. The model demonstrates predictive stability by replicating in both test and cross-disorder STRATIFY/ESTRA samples. While the model was developed among the IMAGEN participants at the age of 23, the clinical associations largely remained consistent at ages 14 and 19 suggesting that the neuroimaging variables may serve as early markers of severe symptoms.

The neuroimaging characterization of psychopathology provides new insight into multimodal brain relationships and their symptom-specific liability regions. Previous neuroimaging-CCA studies have identified between one and three significant components5,13–16. Our study describes six components that capture a wider range of the clinical continuum. The model weights neuroimaging features to provide a relative importance of the anatomical areas to the psychopathology scores. The neuroimaging features are derived from seven different neuroimaging modalities which is unique among psychiatric neuroimaging studies that typically focus on a single neuroimaging modality. Therefore, the putative biomarkers we identify are more comprehensive in describing the neurobiology of psychopathology.

The excitability and impulsive symptom score (component 1) was associated with a novel structure-function relationship, involving activation during the SST and surface area in overlapping regions including the inferior/medial-frontal gyrus, insula, inferior-parietal cortex, caudate, and putamen. These regions play an important role in cognitive control, attention, and response inhibition and are altered in bipolar patients and their relatives17,18. By demonstrating the contingency of functional activation of these brain areas implicated in behavioral inhibition19 on the regional surface area, our finding provides a more refined understanding of the biomarkers that could contribute to impaired inhibitory control20.

For the depressive mood and distress score (component 2), we identified a fronto-limbic brain network involved in top-down control and emotion integration21 that consists of overlapping activations during SST and MID tasks in the dorsolateral/medial-prefrontal cortex, posterior-cingulate cortex, precuneus, and limbic regions including the hippocampus and amygdala. This network is specific to depressive symptoms and distinct from component 1.

The emotional and behavioral dysregulation score (component 3) was the only component broadly associated with multiple diagnostic categories with items primarily related to anxiety. It was associated with all functional neuroimaging modalities, but not the structural modalities, in the amygdala, thalamus, and insula with involvement of the anterior salience network. Anxiety symptoms are frequently present in psychiatric disorders, particularly in internalizing and thought disorders22. The MID correlation between the MID component score and the MID activation is the inverse of the depressive mood and distress score. The differences in loading represent the variance in MID that separates depressive mood from anxiety rather than their co-morbidity. This specificity enables the potential targeting of biological features for possible mechanistic intervention.

The stress pathology score (component 4) was associated with the SST with the strongest positive loading in the anterior cingulate cortex consistent with hyperactivation in this region associated with emotional reactivity and vigilance. The clinical items contributing to the scores were predominately related to post-traumatic stress disorder, suggesting a link to prior trauma. This relationship is bolstered by the pivotal role of the anterior cingulate cortex in emotional reactivity and post-traumatic stress disorder23.

The eating pathology score (component 5) was strongly associated with the resting state brain modes and particularly connectivity between the ventral default mode, basal ganglia, and temporal networks. Both functional and structural associations with the temporal lobe have been reported in bulimia nervosa, where these networks are thought to be linked to social behavior and emotional stimuli24.

Social fear and avoidance score (component 6) was only associated with resting state brain modes, particularly with respect to connectivity in the dorsal default mode and left executive control networks suggesting neural mechanisms underlying deficits in cognitive control during experiences of fear25.

For each psychopathology component individually, the variance explained by the neuroimaging variates was moderate which limits the utility of these neuroimaging variates in a clinical setting. Our objective was to identify behavioral symptom groups informed by their underlying biology. The psychopathology features exhibit the highest covariance across clinical items and multimodal features simultaneously. These clinical and MRI features are orthogonal to each other, meaning that each subsequent component explains the residual variance not accounted for by the previous component. This approach arguably parses both clinical and biological heterogeneity. Consequently, our model offers a more precise understanding by directly identifying biomarkers associated with specific psychopathology components, free from the confounding effects of comorbidity.

There are limitations in translating these findings to clinical application. Our model was developed using a naturalistic sample, minimizing potential confounds from psychiatric treatment, such as medications, but likely missing out on psychiatric disorders with a lower prevalence (e.g., schizophrenia) or older age-at-onset (e.g., dementia). Further, all participants were of Western European origin, potentially limiting the generalizability of our findings to other ethnicities. Additionally, we have not mapped individual differences in the neuroimaging modalities to average brain functioning such as those used in normative modeling26. Furthermore, norms need to be established in these models. SGCCA and similar data-driven approaches offer flexibility in determining statistical models, including sparsity methodology, feature selection, and covariance optimization functions. While beneficial, a major drawback is the lack of standardization—no two CCA models are alike. A consensus is necessary on which biomarkers to include. Valid arguments exist for incorporating other psychiatric biomarkers beyond neuroimaging, including circulating markers, genomic/epigenetic profiles, electrophysiology, and neurochemical markers. This consensus is also needed for the type of clinical data included because these models are highly sensitive to biases in clinical input. The number of items contributing to a symptom can disproportionately affect its weight in the overall model. We coded the DAWBA skip rules using zeros assuming that omissions due to skip rules were rare in our samples8. This approach potentially imposes a covariance structure on the clinical input; however, bias in such models is currently inevitable. If we had only used entry items, information about symptom severity and frequency would have been lost. Symptom questionnaires without skip rules are potentially biased towards a specific diagnosis. Therefore, while we do not claim that our choice of statistical parameters or input data is optimal, we provide proof-of-principle that a reliable model can be produced, emphasizing the need for methodological consensus before these models are ready for clinical application. Therefore, our methodology in enriched datasets and patient populations may have more concrete implications for mental illness treatment.

In conclusion, jointly linking psychiatric symptoms to multimodal brain features lays the groundwork for a dimensional approach to psychiatry optimized for brain biomarkers. We present ‘proof-of-principle’ for a framework that points to quantifiable neurobiological measures enabling precise targeting of biological features for mechanistic intervention. Our results demonstrate the feasibility of SGCCA methodology to produce stable brain-linked psychopathology features but also highlight the need to have consensus among clinical and biological parameters for clinical application. Our framework for neurobiology-enhanced dimensions of psychopathology enables quantitative assessment of comorbidity necessary for precision medicine and demonstrates their potential to bridge the gap between psychiatric neuroscience and clinical treatment of mental disorders.

**Data sharing statement**

IMAGEN data are available from a dedicated database at <https://imagen-project.org>. STRATIFY/ESTRA data are available from the IMAGEN database at <https://stratify-project.org>. The SGCCA statistical model, which includes aggregated data, will be shared upon reasonable request to the corresponding authors at tristram.lett@charite.de or gunter.schumann@charite.de. The code that supports the findings of this study is available on GitHub at <https://github.com/trislett/sgcca-psychiatry-nosology.git>.

**Acknowledgments**

This work received support from the following sources: the European Union-funded Horizon Europe project ‘environMENTAL’ (101057429) and co-funding by UK Research and Innovation under the UK Government’s Horizon Europe funding guarantee (10041392 and 10038599); the Horizon 2020-funded European Research Council Advanced Grant ‘STRATIFY’ (695313); the German Research Foundation (COPE; 675346); the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT-2007-037286), the Medical Research Council and Medical Research Foundation (grants MR/R00465X/1 and MRF-058-0004-RG-DESRI: ‘ESTRA: Neurobiological underpinning of eating disorders: integrative biopsychosocial longitudinal analyses in adolescents’; MR/S020306/1 and MRF-058-0009-RG-DESR-C0759: ‘Establishing causal relationships between biopsychosocial predictors and correlates of eating disorders and their mediation by neural pathways’) the Medical Research Council (grant MR/W002418/1: ‘Eating Disorders: Delineating illness and recovery trajectories to inform personalized prevention and early intervention in young people (EDIFY)’ and the National Institute for Health and Research (NIHR) Biomedical Research Centre (BRC) and Maudsley NHS Foundation Trust (SLaM). The Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the National Key R&D Program of Ministry of Science and Technology of China (MOST 2023YFE0199700). The Medical Research Council Grant 'c-VEDA’ (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institute of Health (NIH) (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers), the Bundesministerium für Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Medical Research Foundation (MRF-058-0014-F-ZHAN-C0866) and the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01), NSFC grant (82150710554).

Further support was provided by grants from: - the ANR (ANR-12-SAMA-0004, AAPG2019 - GeBra), the Eranet Neuron (AF12-NEUR0008-01 - WM2NA; and ANR-18-NEUR00002-01 - ADORe), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l’Avenir (grant AP-RM-17-013 ), the Fédération pour la Recherche sur le Cerveauthe Ile-de-France Region (Action 16700103 -grant to QIM– VEAVE, 23002745–413 23002747); the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1) and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence. This paper represents independent research partly funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care or the European Union, the European Health and Digital Executive Agency (HADEA) or UKRI. European Research Council under a consolidator grant (MENTALPRECISION, 101001118).

Dr. Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire. He received conference support or speaker’s fee from Lilly, Medice, Novartis and Shire. He has been involved in clinical trials conducted by Shire & Viforpharma. He received royalties from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. The present work is unrelated to the above grants and relationships. Dr. Poustka served in an advisory or consultancy role for Roche and Viforpharm and received a speaker's fee from Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. Dr. Robbins provides consultancy for Cambridge Cognition and Supernus, and has a research grant with Shionogi. He has also received editorial honoraria from Elsevier and Springer-Nature. The present work is unrelated to the above grants and relationships. The other authors report no biomedical financial interests or potential conflicts of interest.

None of the funding organizations or sponsors had a role in: design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

We also thank the additional raters for the DAWBA-SDQ: Patricia Conrod, Maren Struve, Jani Penttilä, Viola Kappel, Yvonne Grimmer, Tahmine Fadai, Corinna Insensee, Andreas Becker, Andreas Becker-Isensee.

T.A.L. and G.S. wrote the manuscript, and feedback and editing were provided by all authors. T.J., A.M., N.V. ,E.P., contributed to statistical analysis. N.V., T.B., A.L.W.B., H.F., A.G., H.G., P.G., A.H., R.B., J.-L.M, M.-L.P.M., E.A. F.N., D.P.O., H.V., T.P., L.P., M.B., M.J.B., B.M.N., A.S., Z.Z., L.R., J.W., Y.Z., S. K., C. B., M.N.S, R.W, U.S., J.S., H.W. S.D., acquired the data. A.L.W.B., L.W., N.V., H.L., R.L., E.A., contributed to neuroimaging data processing and analysis. T.A.L., G.S., A.H., J.F., T.W.R., S.D., A.M. conceptualized the study. T.A.L and G.S. had full access to all data in the present study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**References**

1. Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. *Science*. 2015;348(6234):499-500.

2. Opel N, Goltermann J, Hermesdorf M, Berger K, Baune BT, Dannlowski U. Cross-Disorder Analysis of Brain Structural Abnormalities in Six Major Psychiatric Disorders: A Secondary Analysis of Mega- and Meta-analytical Findings From the ENIGMA Consortium. *Biol Psychiatry*. 2020;88(9):678-686.

3. Sprooten E, Rasgon A, Goodman M, et al. Addressing reverse inference in psychiatric neuroimaging: Meta-analyses of task-related brain activation in common mental disorders. *Hum Brain Mapp*. 2017;38(4):1846-1864.

4. Chen CP, Keown CL, Jahedi A, et al. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *Neuroimage Clin*. 2015;8:238-245.

5. Jia T, Ing A, Quinlan EB, et al. Neurobehavioural characterisation and stratification of reinforcement-related behaviour. *Nat Hum Behav*. 2020;4(5):544-558.

6. Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol*. 2017;126(4):454-477.

7. Schumann G, Loth E, Banaschewski T, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry*. 2010;15(12):1128-1139.

8. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41(5):645-655.

9. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry*. 1997;38(5):581-586.

10. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. *Addiction*. 1993;88(6):791-804.

11. Tenenhaus M, Tenenhaus A, Groenen PJF. Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods. *Psychometrika*. Published online May 23, 2017. doi:10.1007/s11336-017-9573-x

12. Tenenhaus A, Philippe C, Guillemot V, Le Cao KA, Grill J, Frouin V. Variable selection for generalized canonical correlation analysis. *Biostatistics*. 2014;15(3):569-583.

13. Xia CH, Ma Z, Ciric R, et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun*. 2018;9(1):3003.

14. Ing A, Sämann PG, Chu C, et al. Identification of neurobehavioural symptom groups based on shared brain mechanisms. *Nat Hum Behav*. 2019;3(12):1306-1318.

15. Moser DA, Doucet GE, Lee WH, et al. Multivariate Associations Among Behavioral, Clinical, and Multimodal Imaging Phenotypes in Patients With Psychosis. *JAMA Psychiatry*. 2018;75(4):386-395.

16. Modabbernia A, Reichenberg A, Ing A, et al. Linked patterns of biological and environmental covariation with brain structure in adolescence: a population-based longitudinal study. *Mol Psychiatry*. 2021;26(9):4905-4918.

17. Bora E, Yucel M, Pantelis C. Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *J Affect Disord*. 2009;113(1-2):1-20.

18. Cattarinussi G, Di Giorgio A, Wolf RC, Balestrieri M, Sambataro F. Neural signatures of the risk for bipolar disorder: A meta-analysis of structural and functional neuroimaging studies. *Bipolar Disord*. 2019;21(3):215-227.

19. Hampshire A, Chamberlain SR, Monti MM, Duncan J, Owen AM. The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage*. 2010;50(3):1313-1319.

20. Cotovio G, Oliveira-Maia AJ. Functional neuroanatomy of mania. *Transl Psychiatry*. 2022;12(1):29.

21. Admon R, Pizzagalli DA. Dysfunctional Reward Processing in Depression. *Curr Opin Psychol*. 2015;4:114-118.

22. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med*. 2007;146(5):317-325.

23. Herringa RJ. Trauma, PTSD, and the developing brain. *Curr Psychiatry Rep*. 2017;19(10). doi:10.1007/s11920-017-0825-3

24. Donnelly B, Touyz S, Hay P, Burton A, Russell J, Caterson I. Neuroimaging in bulimia nervosa and binge eating disorder: a systematic review. *J Eat Disord*. 2018;6(1). doi:10.1186/s40337-018-0187-1

25. Kim YK, Yoon HK. Common and distinct brain networks underlying panic and social anxiety disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2018;80:115-122.

26. Rutherford S, Barkema P, Tso IF, et al. Evidence for embracing normative modeling. *Elife*. 2023;12. doi:10.7554/eLife.85082

A screenshot of a computer screen

Description automatically generated

**Figure 1. Development of sparse generalized canonical correlation analysis (SGCCA) model in the IMAGEN study.** The SGCCA model incorporates eight distinct datasets (called data-views), consisting of both clinical assessments and neuroimaging modalities, from the IMAGEN study. This model is built using 70% of the participants as the training dataset, while the remaining 30% form the test group. Canonical correlation analysis is the method employed, which utilizes cross-covariance matrices of two or more sets of data-views to identify linear combinations (or components) that have maximal correlation. The training data serves several crucial purposes: firstly, for optimizing the model's parameters, including shrinkage parameters (sparsity); secondly, for determining the suitable number of components; and lastly, for performing stability selection (for details, refer to Supplementary Methodology). After establishing the optimal model parameters, the training data are refitted accordingly. Furthermore, ten thousand randomized models are generated by permuting subjects among each training data view. This allows us to evaluate the significance of the model within both the training and test datasets for each component. In the training data, the inner average variance explained (AVE) of the actual model is ranked and compared to the inner AVE of the randomized models. Similarly, the test data are fitted to both the actual and randomized models, and their inner AVEs are compared. Last, regression of the data view components is conducted, with clinical component scores as dependent variables and neuroimaging scores as independent variables. This entire process is repeated in both the training and test samples for each of the significant components. CCA, canonical correlation analysis; CT, cortical thickness; EFT, emotional face task; FA, fractional anisotropy; MID, monetary incentive delay task; SA, surface area; SST, stop-signal task.

A diagram of different emotions

Description automatically generated with medium confidence

**Figure 2. The aggregated loadings for the DAWBA sections and AUDIT questionnaire.**

The aggregate loading (radial axis) provides an overview of the relationship between the symptom scores and the clinical items for the IMAGEN training and test samples and the STRATIFY/ESTRA sample. The aggregated represents the mean of the AUDIT and DAWBA sections for the correlation between the psychopathology scores on the original clinical items. AUD, alcohol-use disorder; POS, positivity, PSY, psychosis; ED, eating-disorders; ADHD, attention deficit hyperactivity disorder; BPD, bipolar disorder; MDD, major depressive disorder; GAD, generalized anxiety disorder; OCD, obsessive-compulsive disorder; PTSD, post-traumatic stress disorder; Soc. Fears, social fears; Spc. Fears, specific fears.

**A chart of a variety of emotions

Description automatically generated with medium confidence**

**Figure 3. The clinical contribution to each psychopathology component.**

The top 25 items clinical structural coefficients are plotted according to their absolute value for each psychopathology component. The question from the clinical battery is on the y-axis with its corresponding section in parentheses, and the structural coefficients are on the x-axis. All items are significant (pFDR<0.05) after 10,000 bootstraps, and the error bars contain the 95% confidence interval. ADHD, attention deficit hyperactivity disorder; AUDIT, alcohol-use disorders identification test; BPD, bipolar disorder; ED, eating disorder; GAD, general anxiety disorder; MDD, major depressive disorder. OCD, obsessive compulsive disorder; PanAttack, panic attack; PTSD, post-traumatic stress disorder; SDQ, strength and difficulties questionnaire; SocFear, social fears.

**A screenshot of a graph

Description automatically generated**

**Figure 4. Neuroimaging features contribution to the psychopathology symptoms for each significant component.**

SGCCA-regression of the variates for the training, test IMAGEN samples and the STRATIFY/ESTRA sample with the psychopathology scores as the response variable and the neuroimaging predictor scores for: the emotional face task, the monetary incentive delay task, the stop-signal task, cortical thickness, surface area, the resting-state brain modes connectivity, and white matter fractional anisotropy. (A), regression model fit and 95% confidence intervals with variate psychopathology scores as the response variable and the neuroimaging predictor scores for: the emotional face task, the monetary incentive delay task, the stop-signal task, cortical thickness, surface area, the resting-state brain modes connectivity, and white matter fractional anisotropy. (B), bar charts of the model coefficients. Each regression model underwent 10000 bootstraps to determine the confidence interval and significance. The horizontal line represents the 95% confidence interval (2.5% to 97.5%). \* pbootstrap <0.05 \*\* pbootstrap <0.0083.

**A collage of different colored brain models

Description automatically generated**

**Figure 5. Neuroimaging loadings for each psychopathology score.**

Significant loadings (structural coefficients; rs) for each score are shown using 10000 bootstraps and after accounting for false discovery rate (PFDR < 0.05). Colors ranging from red to dark blue denote significant positive and negative rs values, respectively. The psychopathology components of interest were excitability and impulsivity (stop-signal task and surface area), depressive mood and distress (monetary incentive delay task and stop-signal task), emotional and behavioral dysregulation (emotional face task, stop-signal task, and monetary incentive delay task) , stress pathology (monetary incentive delay task), eating pathology (monetary incentive delay task and resting state fMRI brain modes), and social fear and avoidance (resting state fMRI brain modes). DMN, Default Mode Network; ECN, Executive Control Network.