ARTICLE OPEN

Breaking the norm: population-scale deviations of brain structure in depression and anxiety

Julius Wiegert[1,2], Sebastián Marty-Lombardi[1,2], Jailan Oweda[1,2], Esra Lenz[1,2], Peter Ahnert[3], Klaus Berger[4], Hermann Brenner[5,6], Josef Frank[7], Hans J. Grabe[8], Karin Halina Greiser[9], Johanna Klinger-König[8], André Karch[4], Michael Leitzmann[10], Claudia Meinke-Franze[11], Rafael Mikolajczyk[12,13], Frauke Nees[14,15,16], Thoralf Niendorf[17], Oliver Sander[18], Carsten Oliver Schmidt[11], Steffi G. Riedel-Heller[19], Kerstin Ritter[20,21,22], Annette Peters[23,24,25], Tobias Pischon[26,27,28], Stephanie Witt[7,29], Johannes Nitsche[1,2], Joonas Naamanka[1,2,30], Sebastian Volkmer[1,2], Antonia Mai[1,2], Amrou Abas[1,2], Xiuzhi Li[1,2], Andreas Meyer-Lindenberg[2], Tobias Gradinger[1,2], Fabian Streit[1,2,29], Urs Braun[1,2] and Emanuel Schwarz[1,2,29]✉

© The Author(s) 2026

Structural brain alterations associated with depression and anxiety are subtle, heterogeneous, and difficult to characterize. We applied autoencoder-based normative modeling to contrastively learned structural MRI representations from two large population-based cohorts (German National Cohort, $N \approx 29{,}000$; UK Biobank, $N \approx 25{,}000$) to quantify individual deviations from normative brain structure across symptom dimensions of depression, anxiety, and, for contextualization, alcohol use.

Deviation magnitude increased with symptom severity for depressive and anxiety symptoms and was most pronounced in individuals with high alcohol use. Directional analyses revealed shared deviation patterns for depression and anxiety that were largely distinct from alcohol-related deviations, and these patterns generalized across cohorts. These affective-symptom-related patterns implicated distributed regional brain-structural variation. Individual deviation profiles improved classification of symptomatic status beyond demographic covariates, with gains concentrated at higher symptom severity.

Together, these findings indicate that affective symptoms are associated with reproducible, dimensional patterns of regional brain-structural deviation that extend beyond normative population variability, supporting transdiagnostic models of internalizing psychopathology.

Molecular Psychiatry (2026) 31:6034–6045; https://doi.org/10.1038/s41380-026-03691-4

INTRODUCTION

Major depressive disorder (MDD) and generalized anxiety disorder (GAD) are among the most prevalent psychiatric conditions worldwide and contribute substantially to the global burden of disease [1–3]. Both, MDD and GAD are characterized by disturbances across mood, motivation, cognition, and sleep, and are associated with substantial functional impairment and increased mortality risk [4–6]. Despite distinct diagnostic criteria,

[1]Hector Institute for Artificial Intelligence in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [3]Universität Leipzig, Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany. [4]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [5]German Cancer Research Center (DKFZ), Heidelberg, Germany. [6]Network Aging Research, Heidelberg University, Heidelberg, Germany. [7]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [8]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [9]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [10]Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany. [11]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [12]Institute for Medical Epidemiology, Biometrics, and Informatics, Medical Faculty of the Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [13]German Center for Mental Health (DZPG), partner site Halle-Jena-Magdeburg, Halle (Saale), Germany. [14]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig-Holstein, Kiel University, Kiel, Germany. [15]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [16]Institute of Medical Psychology, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. [17]Berlin Ultrahigh Field Facility (B.U.F.F.), Max Delbrueck Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany. [18]Clinic for Rheumatology and Hiller Research Center, University Hospital, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany. [19]Institute of Social Medicine, Occupational Health and Public Health (ISAP), Leipzig University, Leipzig, Germany. [20]Department of Machine Learning, Hertie Institute for AI in Brain Health, University of Tübingen, Tübingen, Germany. [21]Department of Psychiatry and Neurosciences, Charité - Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Berlin, Germany. [22]Tübingen AI Center, Tübingen, Germany. [23]Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Neuherberg, Germany. [24]Chair of Epidemiology, Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-Universität München, Munich, Germany. [25]German Center for Mental Health (DZPG), partner site Munich, Munich, Germany. [26]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Molecular Epidemiology Research Group, Berlin, Germany. [27]Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. [28]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Biobank Technology Platform, Berlin, Germany. [29]German Center for Mental Health (DZPG), Partner Site Mannheim - Heidelberg - Ulm, Heidelberg, Germany. [30]SleepWell Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
✉email: emanuel.schwarz@zi-mannheim.de

Received: 11 February 2026 Revised: 21 May 2026 Accepted: 3 June 2026
Published online: 18 July 2026

both disorders affect overlapping psychological domains and co-occur frequently.

Although MDD and GAD are defined as distinct diagnostic entities, they show substantial overlap in symptom domains, treatment response, and biological risk factors. Comorbidity rates reach up to 26% [7], and genetic correlations are high [8]. Evidence from neuroimaging further suggests shared neural substrates [9], complicating attempts to disentangle disorder-specific mechanisms. Both disorders exhibit pronounced clinical heterogeneity, including variation in symptom profiles, age of onset, and disease course [10, 11].

This clinical heterogeneity is mirrored on the biological level, particularly in attempts to predict MDD using brain imaging. A recent ENIGMA benchmark demonstrated that predicting MDD from structural MRI (sMRI) remains highly challenging [12]. Conventional supervised learning approaches implicitly assume biological similarity within diagnostic categories and separability between diagnoses, while also failing to exploit the large availability of healthy imaging data in population-based cohorts.

Normative modeling provides a complementary framework for studying psychiatric and neurodegenerative disorders by characterizing individual variation relative to a reference model learned from healthy populations [13, 14]. This approach enables the quantification of subject-specific deviations and has been applied across a wide range of conditions, including autism, dementia, and mood disorders [15–18]. However, depending on the underlying feature representation, normative models may emphasize localized or summary-level deviations (e.g., voxel- or region-of-interest-based), which can limit their ability to capture highly distributed patterns of structural variation.

In this work, we pursued three aims: first, to characterize distributed brain-structural variation associated with depressive and anxiety symptoms; second, to quantify individual deviation magnitude relative to normative variability; and third, to assess the directional organization of these deviations using alcohol-related symptoms as a comparator, given their well-documented structural brain alterations [19, 20]. Together, these aims target both the magnitude and organization of brain-structural deviations while addressing transdiagnostic and individual heterogeneity.

To address these aims, we adopted a two-stage modeling strategy. First, we learned data-driven structural MRI representations from minimally processed sMRI data using contrastive representation learning. Recent contrastive approaches enable end-to-end feature learning without reliance on predefined regional summaries [21], and by comparing brain images at the population level, they may encourage representations that are sensitive to inter-individual differences expressed across distributed brain regions. Second, we applied autoencoder-based deep normative modeling to these contrastively learned representations, modeling the distribution of variation observed in healthy individuals and quantifying subject-specific deviations in latent space [22, 23]. This separation of representation learning and normative modeling allows the normative model to operate on a compact, data-driven embedding of distributed brain-structural differences.

Depressive and anxiety symptoms are present in the general population [24]. Thus we leveraged population-scale cohorts from the German National Cohort (NAKO) and the UK Biobank (UKB) [25, 26]. sMRI is available at scale in both cohorts, exhibits high test--retest reliability [27], and provides a stable readout of cumulative neurobiological variation well suited for large-scale normative modeling. Moreover, large meta-analyses demonstrate that gray matter volume (GMV) alterations in depression and anxiety disorders are spatially distributed rather than regionally focal [28, 29]. The NAKO cohort served as the discovery dataset, with external validation performed in the UK Biobank.

MATERIALS AND METHODS

This study was approved by the Ethics Committee II of the University of Heidelberg Medical Faculty Mannheim (protocol no. 2025-827). NAKO and UKB data collection were approved by the relevant local ethics committees in Germany and the United Kingdom and were conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent. Unless stated otherwise, all model training was performed in the NAKO cohort; UK Biobank data were used exclusively for evaluation.

Group definitions

Current depressive symptoms were assessed using the depression scale of the Patient Health Questionnaire-9 (PHQ-9) [30], anxiety symptoms using the anxiety scale of the Generalized Anxiety Disorder-7 (GAD-7) [31], and alcohol use using the consumption scale of the Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) [32]. These instruments were selected because they provide ordinal severity measures available in both cohorts.

Participants were stratified into symptom severity groups based on established cutoff scores to define interpretable severity strata: PHQ-9 none/minimal (0–4), mild (5–9), moderate (10–14), moderately severe (15–19), and severe (20–27) [30]; GAD-7 minimal (0–4), mild (5–9), moderate (10–14), and severe (15–21) [31]; and AUDIT-C low (0–4), increasing (5–7), higher risk (8–10), and possible dependence (11–12) [32]. To ensure sufficient statistical power while retaining resolution at higher symptom levels, individuals with AUDIT-C scores above 7 were further split into groups with scores of 8, 9, and ≥10.

Healthy controls (HCs) were defined using harmonized thresholds across instruments (PHQ-9 < 5, GAD-7 < 5, AUDIT-C < 8). To minimize psychiatric contamination, individuals with lifetime or self-reported diagnoses of mood, anxiety, or alcohol use disorders were excluded using cohort-specific diagnostic information.

Cohorts - German national cohort (NAKO)

The NAKO is a population-based study designed to investigate the causes and early stages of common chronic diseases [25]. The study enrolled 205,000 participants (aged 20–69 years) from 18 centers across Germany, of whom 30,861 underwent MRI at five centers with dedicated facilities [33]. The baseline assessment, conducted between 2014 and 2019, comprised a comprehensive list of examinations, standardized face-to-face medical interviews and touchscreen-based self-report questionnaires, as well as biomaterial collection.

We performed an initial data filtering step by excluding samples with missing questionnaire data (PHQ-9, GAD-7, or AUDIT-C), diagnostic data (MINI or self-reported diagnosis), or confounder data (age, sex, or scan site). Participants with a history of stroke ($N = 301$) or Parkinson's disease ($N = 23$) were also excluded to reduce potential bias from neurodegenerative conditions. In total, 1570 samples were excluded due to missing questionnaire or diagnostic data. The final group distributions and demographics used in subsequent analyses are summarized in Table 1.

Cohorts - UK biobank (UKB)

The UKB is a population-based cohort of approximately 500,000 participants aged 40–69 years at recruitment, with sMRI data available for 49,279 individuals [26]. Individuals with neurodegenerative diseases ($N = 2589$; ICD-10 codes) were excluded. After further excluding participants with missing questionnaire data (PHQ-9, GAD-7, or AUDIT-C) or missing confounder information (age, sex, or total intracranial volume (TIV)), 24,838 individuals remained (Table 1).

Because PHQ-9 and GAD-7 scores were available in UKB, the same HC criteria were applied as in NAKO. However, questionnaire assessments in UKB were completed independently of the MRI scan date [34], resulting in substantially larger temporal offsets between imaging and symptom measures (median absolute deviation: 742 days in UKB vs. 15 days in NAKO). This temporal mismatch represents an inherent limitation. In addition to questionnaire-based definitions, ICD-10 diagnoses and self-reported diagnoses for MDD, GAD, ANX, and AUD were included in UKB analyses.

Magnetic resonance imaging

Structural T1-weighted MRI data were analyzed. In NAKO, images were acquired using 3D MPRAGE (TR = 2300 ms, TE = 2.98 ms, TI = 900 ms) on a 3T Siemens Skyra scanner with 1 mm isotropic resolution. Images were

processed using CAT12 (v12.9), including bias correction, segmentation, normalization to MNI152NLin2009cAsym space, modulation, and resampling. For computational tractability while preserving macro-anatomical structure relevant for whole-brain representation learning, GMV images were downsampled to a standardized $80 \times 80 \times 80$ voxel grid using spline interpolation, a resolution successfully used in 3D CNN-based brain MRI studies [35, 36]. Voxel intensities were subsequently standardized.

In UKB, T1-weighted images were acquired using 3D MPRAGE (TR = 2000 ms, TE = 2.01–2.03 ms, TI = 880 ms) on a 3T Siemens scanner. Images were preprocessed using FSL (v5.0.10) FAST and registered to the same MNI template using FLIRT. Identical downsampling and standardization procedures were applied to ensure comparability across cohorts.

Site- and scanner-related variation was addressed using linear residualization and additionally, where appropriate, by including the corresponding variables as covariates in downstream analyses.

Overview of the analytical pipeline

In brief, the analytical pipeline consisted of three steps. First, self-supervised contrastive learning was used to transform each sMRI scan into a compact representation of distributed inter-individual brain-structural variation. Second, an autoencoder trained exclusively on HCs was used to model the normative range of these representations; deviations were defined as the difference between observed and reconstructed embeddings. Third, these deviation profiles were analyzed with respect to their magnitude, direction, regional correlates, and ability to classify symptom-defined groups. Thus, the analyses quantify departures from a learned multivariate structural reference space rather than focal voxelwise abnormalities or diagnostic markers. Details are provided below and in the Supplementary Material.

Contrastive feature extraction using momentum contrast

A schematic overview of the representation learning and normative modeling pipeline is provided in Supplementary Figure 1.

To derive compact, multivariate representations of inter-individual variation in brain structure, we applied momentum contrast (MoCo) for self-supervised feature extraction on GMV images [37], using a three-dimensional convolutional neural network (3D CNN) encoder. MoCo was trained on the full NAKO MRI dataset, including HCs and symptomatic participants. As training was fully unsupervised and did not use diagnostic information, this procedure does not introduce label leakage and follows standard practice in representation learning. MoCo was trained to produce 256-dimensional embeddings, a commonly used embedding size in prior contrastive learning work that balances representational capacity with downstream model complexity [38–40].

All deep learning models were implemented in PyTorch (version 2.2.0) [41].

Normative modeling using autoencoder

MoCo embeddings were linearly residualized with respect to covariates age, age squared, sex, scanning site, and TIV to reduce variance attributable to non-pathological sources, and subsequently min--max scaled to ensure numerical stability.

The residualized embeddings were input to a fully connected autoencoder trained exclusively on HCs from the NAKO cohort. HC data were split into training, validation, and test sets (70/15/15%), ensuring model selection and generalization assessment without exposure to symptomatic subjects. The autoencoder mapped the 256-dimensional input to a 50-dimensional bottleneck and back to the original space, and reconstruction error-derived deviations were interpreted relative to the learned normative manifold rather than as probabilistic scores. Architectural details, bottleneck selection, optimization settings, and training/validation diagnostics are provided in Supplementary Sections 2.1–2.3 and Supplementary Tables 8–9.

Normative deviations were quantified using reconstruction errors from the HC-trained autoencoder. Because the autoencoder partially reintroduced confounding structure, deviation vectors were subjected to a second linear residualization using the same covariates. Prior to residualization, deviations showed widespread associations with age, age squared, sex, and age × sex; after residualization, only a single weak association with sex remained.

Shift analysis of mahalanobis distances

To quantify how symptomatic groups deviate from normative variability at the group level, we performed a shift analysis comparing the full distribution of deviations rather than relying on summary statistics such as central tendencies.

For each cohort (NAKO and UKB), Mahalanobis distances in the 256-dimensional deviation space were computed relative to a cohort-specific HC reference distribution. For an individual deviation vector x, the Mahalanobis distance was defined as

$$D_M(x) = \sqrt{(x - \mu_{\mathrm{HC}})^{\mathsf{T}} \Sigma_{\mathrm{HC}}^{-1} (x - \mu_{\mathrm{HC}})},$$

where μ_{HC} and $_{\mathrm{HC}}$ denote the mean vector and covariance matrix estimated from the HC reference sample, respectively [42]. This yielded a covariance-adjusted summary of how strongly each individual's deviation profile departed from the HC reference distribution. The mean vector and covariance matrix were estimated from a random 90% subset of HCs, and distances were subsequently computed for symptomatic participants and the remaining 10% HC holdout set.

Cohort-specific HC reference distributions were used because Mahalanobis distances depend on the mean and covariance structure of the reference population. Accordingly, absolute deviation magnitudes are not directly comparable across cohorts. Stability of reference parameters was assessed by bootstrap resampling (Supplementary Section 2.5).

Distributional shift was quantified as the exponentially weighted area in which the Mahalanobis distance distribution of a symptomatic group exceeded that of HCs. This weighting emphasizes larger deviations, yielding a scalar metric between 0 and 1 that summarizes the extent to which symptomatic variability exceeds normative heterogeneity (details in Supplementary Section 2.4). As this metric is designed to capture distribution-level differences, it is intended for group-level comparison.

Uncertainty in shift estimates was quantified using nonparametric bootstrap resampling (1000 iterations) of subjects within each group, from which 95% confidence intervals were derived.

Significance testing of mahalanobis distances

To assess covariate-adjusted group differences in deviation magnitude, we performed individual-level inference using linear regression models with heteroskedasticity-consistent (HC3) standard errors. For each symptomatic group, log-transformed Mahalanobis distances were regressed on binary group membership (symptomatic vs. HC).

Regression-based inference was used to allow covariate adjustment. Log-transformation and robust standard errors were used to mitigate non-normality and the influence of extreme values.

Statistical significance of group membership was assessed using type III ANOVA on the fitted models, performed separately for each symptomatic group. Resulting p-values reflect covariate-adjusted differences in individual deviation magnitude between symptomatic participants and HCs and complement the distributional shift analysis. To control for multiple testing across diagnostic groups, p-values were corrected using the Benjamini--Hochberg false discovery rate (FDR) procedure [43]. Unless stated otherwise, significance was assessed at an FDR threshold of $\alpha = 0.05$.

Directional analysis

In addition to the scalar shift analysis, we examined the geometry of normative deviations after projection into a low-dimensional space for visualization. Deviations were defined as the difference between each subject's original and reconstructed feature representations in the deconfounded 256-dimensional deviation space.

For visualization, subject-level deviation vectors were standardized and projected into two dimensions using principal component analysis (PCA). The first two components explained 21% of the variance. Subjects were then grouped according to symptom domain (PHQ-9, GAD-7, AUDIT-C), severity level, and comorbidity status.

Robust linear models correcting for covariates were fitted in symptomatic subjects only, relating each principal deviation score (PC1, PC2) to symptom severity (PHQ-9 or GAD-7) in separate models, with HC3 standard errors. Multiple testing was controlled using Benjamini-Hochberg FDR correction. Effect sizes were summarized using standardized regression coefficients (β_{std}).

Group-level deviation vectors were estimated as geometric medians. Stability was assessed using 1000 bootstrap resamples, from which 1-SD uncertainty ellipses were derived.

All group-level deviation vectors and bootstrap ellipses were centered relative to the HC centroid. This analysis is intended as a qualitative assessment of relative deviation directions and overlap between groups in

Table 1. Demographics in the NAKO and UKB cohort.

Group	Criterion	Count	Age $\mu(\pm\sigma)$	Females (%)
NAKO cohort				
HC	–	15941	48 (±12)	39
AUDIT-C (8)	AUDIT-C sum =8	633	49 (±13)	15
AUDIT-C (9)	AUDIT-C sum =9	249	49 (±13)	10
AUDIT-C ≥ 10	AUDIT-C sum ≥10	111	51 (±13)	11
MDD (mild)	PHQ-9 sum ∈ [5, 9]	6829	51 (±11)	53
MDD (moderate)	PHQ-9 sum ∈ [10, 14]	1464	46 (±12)	56
MDD (moderately severe)	PHQ-9 sum ∈ [15, 19]	423	45 (±12)	55
MDD (severe)	PHQ-9 sum ≥20	181	47 (±12)	55
MDD (MINI)	MINI Diagnosis MDD	4752	59 (±11)	54
MDD (Diagnosis)	Doctor's Diagnosis	3636	50 (±11)	60
GAD (mild)	GAD-7 sum ∈ [5, 9]	5566	47 (±12)	54
GAD (moderate)	GAD-7 sum ∈ [10, 14]	1058	46 (±12)	56
GAD (severe)	GAD-7 sum ≥ 15	320	46 (±12)	59
ANX/PD (Diagnosis)	Doctor's Diagnosis	1716	49 (±12)	59
UKB cohort				
HC	–	13582	65 (±7)	53
AUDIT-C (8)	AUDIT-C sum =8	845	63 (±7)	32
AUDIT-C (9)	AUDIT-C sum =9	611	62 (±7)	25
AUDIT-C ≥10	AUDIT-C sum ≥10	631	63 (±7)	22
AUD	ICD-10	192	64 (±8)	25
MDD (mild)	PHQ-9 sum ∈ [5, 9]	3357	61 (±7)	63
MDD (moderate)	PHQ-9 sum ∈ [10, 14]	770	61 (±8)	62
MDD (moderately severe)	PHQ-9 sum ∈ [15, 19]	267	60 (±7)	66
MDD (severe)	PHQ-9 sum ≥ 20	132	58 (±6)	70
MDD (ICD-10)	ICD-10	754	62 (±8)	65
MDD (Diagnosis)	Doctor's Diagnosis	3 918	63 (±7)	66
GAD (mild)	GAD-7 sum ∈ [5, 9]	2927	61 (±8)	66
GAD (moderate)	GAD-7 sum ∈ [10, 14]	559	61 (±8)	65
GAD (severe)	GAD-7 sum ∈ [15, 21]	276	59 (±7)	67
GAD (ICD-10)	ICD-10	19	63 (±7)	66
ANX (ICD-10)	ICD-10	654	63 (±7)	65
ANX (Diagnosis)	Doctor's Diagnosis	2646	63 (±7)	63

HC, healthy controls, *AUDIT-C* Alcohol Use Disorders Identification Test-Consumption; *PHQ-9*, Patient Health Questionnaire-9, *MDD*, major depressive disorder, *MINI*, MINI International Neuropsychiatric Interview, *GAD-7*, Generalized Anxiety Disorder-7, *GAD*, generalized anxiety disorder, *ANX/PD*, anxiety or panic disorder, *ICD-10*, International Classification of Diseases, 10th Revision.

the projected space, rather than a formal inferential test in the full high-dimensional representation.

Explaining normative deviations

To relate the learned latent deviation dimensions to interpretable neuroanatomical features, we performed a post hoc, heuristic analysis linking deviations to regional gray matter measures. Direct gradient-based attribution methods (e.g., saliency maps) are not applicable in this setting because multiple deconfounding and residualization steps break the differentiable mapping between voxel-level inputs and final deviation estimates. Moreover, the normative model is spatially unconstrained and operates on distributed whole-brain representations rather than predefined regions. Accordingly, the following analysis is exploratory in nature and intended for hypothesis generation rather than definitive anatomical localization.

The analysis comprised three steps. First, deviation dimensions were screened for associations with symptom severity. Second, relationships between regional GMV measures and latent embedding dimensions were quantified. Third, symptom-associated deviation dimensions were aggregated to derive regional relevance maps.

In the first step, questionnaire sum scores (PHQ-9, GAD-7, AUDIT-C) were correlated with each of the 256 deviation dimensions using Spearman rank correlations. To account for multiple testing (3×256 tests), we applied the Storey--Tibshirani q-value procedure [44]. Statistical significance was defined as $q < 0.05$, corresponding to control of the positive FDR (pFDR).

In the second step, elastic net regression [45] was used to predict each covariate-adjusted MoCo embedding dimension from 99 covariate-adjusted FreeSurfer GMV features. Elastic net regularization was chosen for its robustness to multicollinearity and its ability to yield sparse and stable region-embedding associations.

In the third step, regional coefficients from the elastic net models were weighted by the corresponding significant dimension-symptom correlations and summed across all selected dimensions, yielding a single heuristic relevance score per brain region and symptom domain.

Classification analysis

To quantify the predictive value of the deviations beyond demographic effects, we performed classification analyses using center, age, age squared, sex, and the sex × age interaction as baseline covariates.

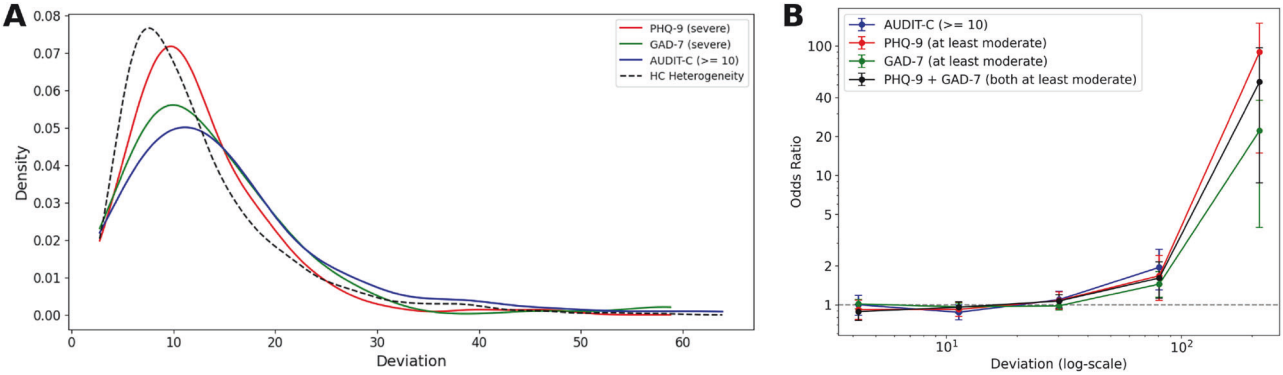


Fig. 1 Distribution of normative deviations and their association with symptom status. A Density estimates of Mahalanobis distance (deviation) from the norm for three exemplary symptomatic groups. The dashed black line represents the natural heterogeneity among independent healthy controls (HCs). **B** Relationship between deviation from the normative distribution and the odds ratio (OR) of having Patient Health Questionnaire-9 (PHQ-9) ≥10, Generalized Anxiety Disorder-7 (GAD-7) ≥10, and an Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) ≥10. Deviation scores were stratified into 5 bins with edges defined on a base-10 logarithmic scale, spanning the minimum to maximum observed distances. This ensured relatively finer resolution in the distribution tail where case enrichment was expected. A strong increase in OR with increasing deviation illustrates that individuals further from the norm are substantially more likely to be symptomatic. Error bars denote 95% confidence intervals obtained via 1000 bootstrap resamples. In the highest deviation bin, there was 1 HC, 0 AUDIT-C samples, 6 PHQ-9 samples, 7 GAD-7 samples, and 6 PHQ-9 + GAD-7 samples.

All classification analyses were performed exclusively in the UKB cohort, ensuring strict separation between representation learning and supervised evaluation.

We evaluated two model configurations: (i) confounders only, and (ii) normative deviation features plus confounders, where the latter included the 256-dimensional deviation vectors derived from the autoencoder.

Binary classifications (symptomatic group vs. HC) were performed using logistic regression with elastic net regularization. The elastic net mixing parameter (ℓ_1 ratio) was tuned within each training fold using an inner stratified 80/20 split. To prevent data leakage, all features were standardized within each training fold using parameters estimated from the corresponding training data only.

Model performances were evaluated using stratified 10-fold cross-validation and summarized reporting the mean and standard deviation of the area under the ROC curve (AUC) and balanced accuracy (BACC), with a fixed decision threshold of 0.5.

RESULTS

Shift analysis

Figure 1A illustrates how deviations in symptomatic PHQ-9, GAD-7, and AUDIT-C groups diverge from normative variation defined by an independent HC holdout set ($N = 1594$). Figure 1B shows that the magnitude of deviation is associated with the likelihood of belonging to at least moderately symptomatic groups. We quantified the proportion of each group whose deviations exceeded the range observed among HCs (Fig. 2A, henceforth referred to as "shift").

This analysis therefore asks whether symptomatic groups contain a larger fraction of individuals with deviation profiles outside the range typically observed among HCs.

At the symptomatic group level, participants with higher symptom severity levels for MDD and GAD (encoded on a scale from 1 [mild] to 4 [severe]) exhibited significantly larger shifts. Spearman rank correlations on the group level between symptom severity and shift revealed strong positive associations when jointly analyzing the PHQ-9 and GAD-7 groups ($p = 0.92$, $p = 0.006$), i.e., levels 1 to 4 for PHQ-9 and 2 to 4 for GAD-7 and their corresponding shift values.

The group with the highest shift was alcohol-related, with a shift of over 20% in the top AUDIT-C (≥10) group (Fig. 2A). Recent, severe depressive and anxiety symptoms exhibited larger deviations than lifetime diagnoses assessed via clinical interviews or self-report (Supplementary Figure 3). P-values for all groups are provided in Supplementary Tables 10 and 11. Sex-stratified analyses showed comparable trends (Supplementary Figure 4).

As shown in Fig. 2B, deviation shifts in UKB followed a pattern that was qualitatively consistent with NAKO. Because Mahalanobis distances and shift estimates were defined relative to cohort-specific HC reference distributions, absolute deviation magnitudes are not directly comparable across cohorts. Cross-cohort comparisons are therefore interpreted descriptively, focusing on the relative ordering of symptom strata and the recurrence of severity-related patterns rather than on absolute differences in magnitude. Under this interpretation, the relative ordering of symptom strata by severity was broadly preserved across cohorts.

Directional analysis

To examine how psychiatric symptom profiles manifest in the space of normative brain deviations, we analyzed the geometry of group-level deviation vectors relative to the healthy reference. Participants were stratified by symptom domain (PHQ-9, GAD-7, AUDIT-C), severity level, and comorbidity status, allowing us to assess both the extent of heterogeneity within symptom groups and the presence of higher-level structural patterns across domains.

Group-representative deviation vectors were estimated as geometric medians and projected into two principal components for visualization (Fig. 3A). Shaded ellipses reflect bootstrap uncertainty, highlighting substantial within-group heterogeneity.

Despite substantial within-group heterogeneity, consistent meta-patterns emerged. Alcohol-related groups deviated along a direction largely distinct from the mood--anxiety axis traced by PHQ-9 and GAD-7 groups, while increasing AUDIT-C thresholds showed progressively larger deviations. Depressive and anxiety symptom groups overlapped extensively, with higher symptom levels shifting further along a shared mood--anxiety-related direction. Individuals with concurrent PHQ-9/GAD-7 elevation occupied intermediate positions rather than forming a separable cluster, supporting a continuous spectrum of overlapping deviation patterns. Participants with elevated AUDIT-C and concurrent depressive or anxiety symptoms clustered between alcohol-related and mood--anxiety-associated directions, consistent with mixed or additive deviation profiles. Comparable patterns were observed for interview-based and self-reported diagnoses (Supplementary Figure 5).

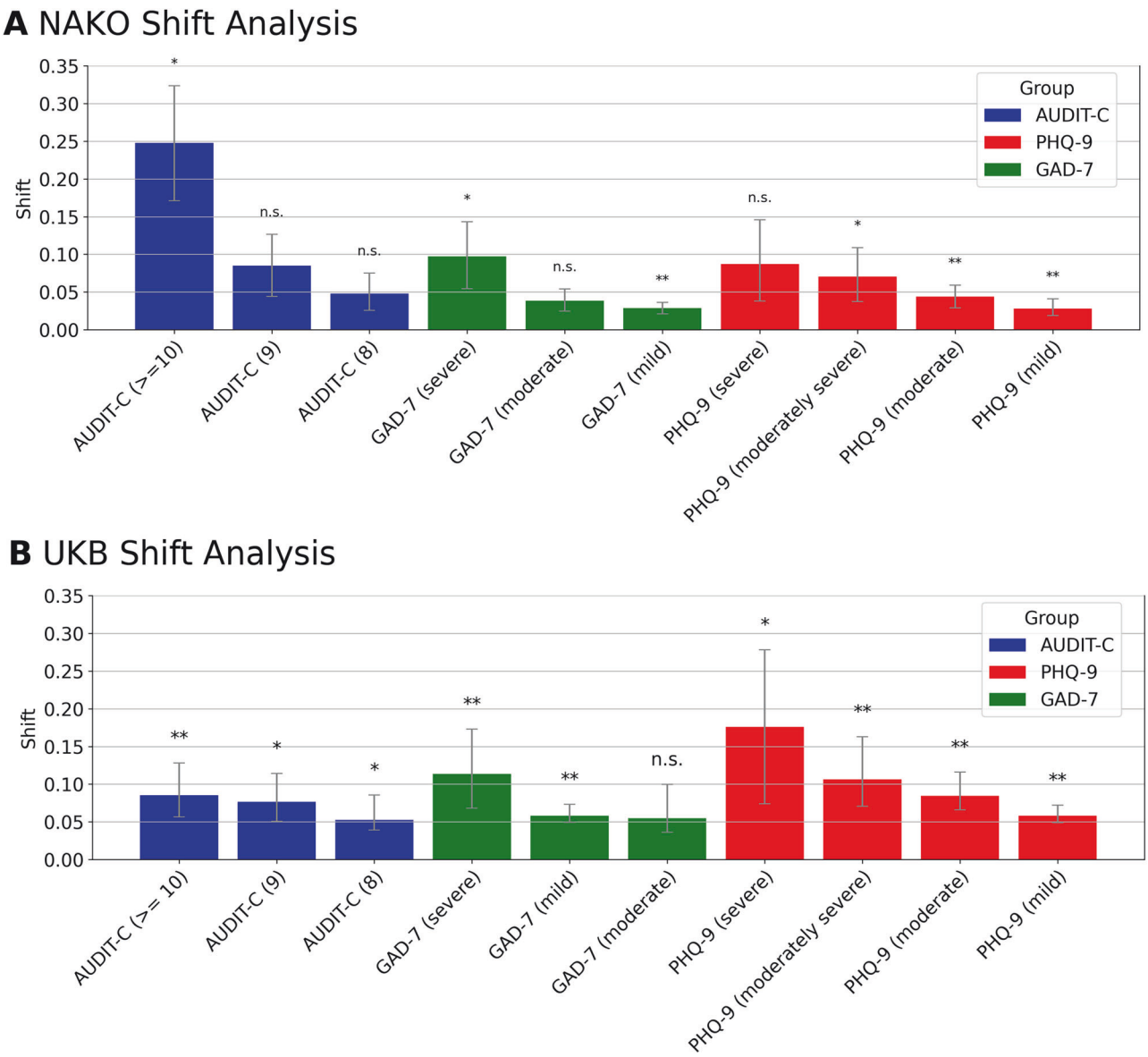


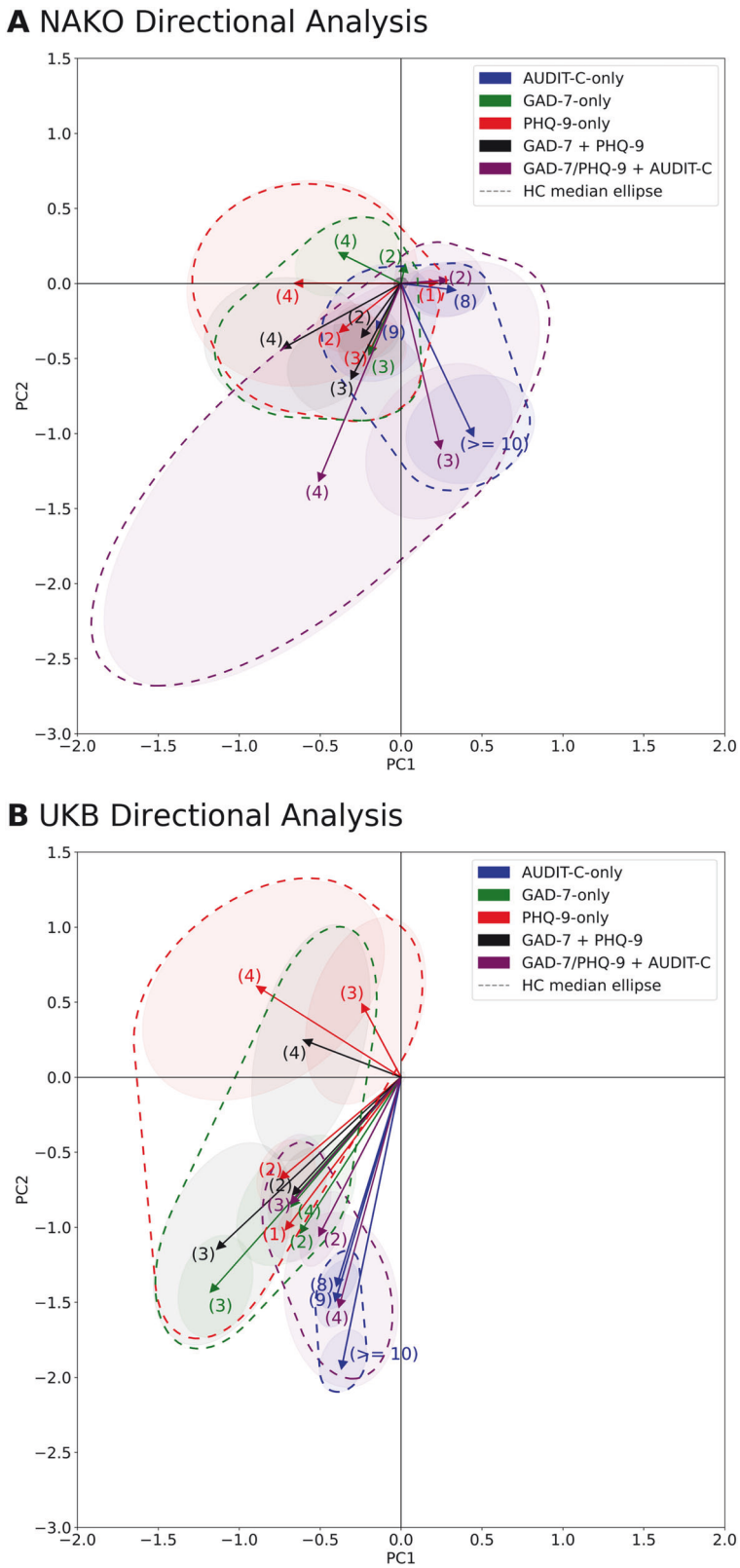
Fig. 2 Cohort-specific shifts in normative deviations across symptomatic groups and cohorts. Fraction of symptomatic group deviations unexplained by healthy control variability (*shift*) for all symptomatic groups for NAKO cohort (German National Cohort) (**A**) and the UKB cohort (UK Biobank) (**B**). Because shifts were computed relative to cohort-specific HC reference distributions, values are not directly comparable as absolute deviation magnitudes across cohorts and should be interpreted within cohort. To assess statistical significance, underlying Mahalanobis distances were modeled as the dependent variable in a linear regression with group membership as the main predictor, adjusting for sex, age, age-squared, and sex-age interactions. Asterisks indicate the level of statistical significance after multiple testing correction using Benjamini-Hochberg correction: * $p_{FDR} < 0.05$, ** $p_{FDR} < 0.01$. Error bars indicate 95% confidence intervals obtained from 1000 bootstrap samples. Symptomatic groups are defined by AUDIT-C (Alcohol Use Disorders Identification Test-Consumption), PHQ-9 (Patient Health Questionnaire-9), and GAD-7 (Generalized Anxiety Disorder-7) thresholds.

A sensitivity analysis examining the stability of deviation directions under sex stratification is provided in Supplementary Figure 6.

In symptomatic subjects ($n = 13{,}003$), GAD-7 scores were associated with PC1 deviations ($q = 0.003$), whereas PHQ-9 scores were not ($q = 0.077$); conversely, both PHQ-9 ($q = 0.004$) and GAD-7 ($q = 0.022$) were associated with PC2. Effect sizes were uniformly small (standardized $\beta \approx 0.02$–0.03). In the UKB external validation sample (symptomatic subjects-only; $n = 9010$), PC1 was associated with both PHQ-9 ($\beta_{std} = 0.030$, $q = 5.5 \times 10^{-3}$) and GAD-7 ($\beta_{std} = -0.033$, $q = 4.1 \times 10^{-3}$), whereas PC2 was associated only with GAD-7 ($\beta_{std} = 0.040$, $q = 5.1 \times 10^{-4}$) and not with PHQ-9 ($q = 0.054$).

To assess whether the geometric relationships observed in NAKO generalize to an independent cohort, UKB deviation vectors were projected into the directional PCA deviation space (Fig. 3B). Despite substantial within-group heterogeneity, several higher-level patterns are consistent across cohorts. Alcohol-related symptom groups again deviate along directions that are largely distinct from the mood-anxiety axis traced by PHQ-9 and GAD-7 groups. Mood and anxiety symptom strata occupy overlapping regions of the deviation space, and participants with concurrent mood-anxiety and alcohol-related symptom elevation fall in intermediate positions between the corresponding directions, which appears to be the case in our analysis.

At the same time, cohort-specific differences were visually evident. In UKB, symptom groups appeared more displaced from the HC centroid than in NAKO, including lower symptom-burden and alcohol-related strata. Because deviation vectors were



centered and interpreted relative to cohort-specific HC references, this apparent displacement should be interpreted descriptively rather than as evidence for larger absolute deviation magnitudes in UKB. It may reflect differences in cohort composition, assessment context, population health, or residual cohort-specific effects, rather than a change in the relative geometry of deviation patterns.

Consistency of the directional patterns was further supported by analyses based on ICD-10 and self-reported diagnoses in the UKB (Supplementary Figure 5).

Fig. 3 Directional organization of normative deviations across symptom groups and cohorts. Directional analyses based on principal component analysis (two dimensions, 21% explained variance) in the NAKO cohort (German National Cohort) (**A**) and the UKB cohort (UK Biobank) (**B**). Shaded ellipses denote the ±1 SD contour of the bootstrap (1000 resamples) distribution of the group-level geometric median vectors. The dark ellipse around the origin indicates the median deviation among healthy controls (HCs). Numbers indicate severity levels. ``MDD/GAD + ALC" denotes individuals with either acute MDD or GAD symptoms and elevated alcohol use (AUDIT-C ≥ 10). For PHQ-9, levels (1) to (4) correspond to mild, moderate, moderately severe, and severe symptoms, respectively. For GAD-7, levels (2) to (4) correspond to mild, moderate, and severe symptoms. The scattered area includes the ellipses of GAD, PHQ, GAD+PHQ groups (red), AUDIT-C groups (blue) and comorbid groups (purple). Because deviation vectors were centered and interpreted relative to cohort-specific HC references, apparent distances from the HC centroid should not be interpreted as directly comparable absolute deviation magnitudes across cohorts.

A sensitivity analysis stratifying UKB participants by MRI--questionnaire interval showed that group-level deviation directions were highly similar to the unfiltered estimates for both the shortest and longest interval quintiles when group sizes were sufficiently large, whereas estimates became unstable only under restrictive minimum-size thresholds with few retained groups (Supplementary Figure 7).

Importance of brain regions

We first examined which latent deviation dimensions were associated with symptom severity. Of the 256 normative deviation dimensions, 14 were significantly associated with PHQ-9 scores (5.4%), 4 with GAD-7 scores (1.6%), and 35 with AUDIT-C scores (13.7%) after multiple-testing correction (q-value <0.05).

This mapping should be interpreted as a heuristic summary of distributed associations between latent deviation dimensions and regional GMV features, rather than as evidence for focal anatomical effects. Elastic net models explained a moderate fraction of variance in the MoCo embedding dimensions (median $R^2 = 0.22$, range =0.10–0.40; Supplementary Table 14), further supporting cautious interpretation of the regional relevance maps.

Aggregating regional GMV associations via our heuristic mapping procedure yielded distributed regional relevance patterns rather than focal effects. Figure 4 shows the regions contributing most consistently to symptom-associated deviation dimensions. Across all three symptom domains, cerebellar regions emerged as the most prominent contributors. For AUDIT-C, these associations were predominantly negative, indicating that higher alcohol-related symptom severity was linked to deviation dimensions associated with lower cerebellar GMV. Additional negative contributions were observed in posterior cortical regions, including the cuneus (Supplementary Figures 11 and 12).

Direct correlations between deconfounded FreeSurfer GMVs and symptom scores yielded markedly different spatial patterns (Supplementary Figures 11 and 13), indicating that the normative deviation space captures diffuse structure not apparent in univariate region-symptom analyses.

Classification performance

Across symptom domains, models incorporating normative deviations consistently outperformed confounder-only baselines, with performance generally increasing with symptom severity (Table 2). Mean AUC and BACC generally increased with symptom severity, although variability also increased as group sizes decreased. Improvements were also observed for anxiety-related groups, although gains were smaller and less stable than for depression.

In alcohol-related groups, models incorporating normative deviations again outperformed the confounder-only baseline, with performance increasing alongside symptom severity. It reached peak classification performance in the highest alcohol consumption group (AUC 72+/−3, BACC 66 +/−3). Using normative deviations yielded modest but consistent improvements over deconfounded FreeSurfer features across diagnostic groups (ΔBACC: median =2%, mean =2 ± 2% SD; ΔAUC: median =2%, mean =1 ± 2% SD; Supplementary Table 15). When restricting analyses to ICD-10-defined diagnostic groups, balanced accuracy increased further (ΔBACC≈+4–5%), with corresponding AUC gains for depressive and anxiety disorders (ΔAUC≈+3–4%), while alcohol consumption showed a small decrease in AUC (ΔAUC≈−1%). These results indicate that normative deviations carry information beyond demographic covariates, particularly at higher symptom severity.

DISCUSSION

In this study, we examined how depressive, anxiety-related, and alcohol-related symptoms relate to deviations in distributed gray matter structure in large population-based cohorts. We found that higher symptom severity was associated with greater deviation from normative structural variation. Depression and anxiety showed strongly overlapping deviation patterns, whereas alcohol-related symptoms deviated along a largely distinct direction, consistent with dimensional rather than diagnosis-specific organization.

Several choices in the normative modeling pipeline warrant consideration. MoCo-based contrastive representation learning was used to obtain compact features sensitive to distributed inter-individual structural differences rather than to optimize voxelwise image reconstruction. This is relevant for subtle psychiatric effects, where reconstruction objectives may prioritize dominant anatomical variance unrelated to symptoms. The 256-dimensional embedding size pragmatically balances representational capacity and downstream tractability, following prior contrastive learning work [38–40]. The normative model should therefore be interpreted as operating on a compact, data-driven structural representation rather than directly on voxelwise anatomy.

The directional overlap of MDD and GAD and specificity against AUD-related symptoms is consistent with the well-established clinical and biological overlap of MDD and GAD [46], including their high genetic correlation [47, 48] and the considerably lower overlap of either disorder with AUD. Importantly, this geometric organization should be interpreted in light of the contrastive representation learning stage used to derive the feature space. Contrastive objectives emphasize relative differences between individuals at the population level, rather than absolute regional abnormalities or deviations from an anatomical template. As a consequence, the resulting normative deviations reflect population-relative configurations of distributed brain structure rather than focal deficits. This representation bias may naturally favor dimensional and transdiagnostic organization, as variation is encoded along continuous axes of difference rather than partitioned into disorder-specific patterns. As some individuals with AUD show pronounced affective symptoms [49], we anticipated that such individuals would show an intermediate deviation profile. This was supported by our data, reinforcing the relevance of directional deviation analyses for capturing transdiagnostic neurobiological structure at the population level. The observed separation between internalizing and externalizing symptom profiles also aligns well with the dimensional structure

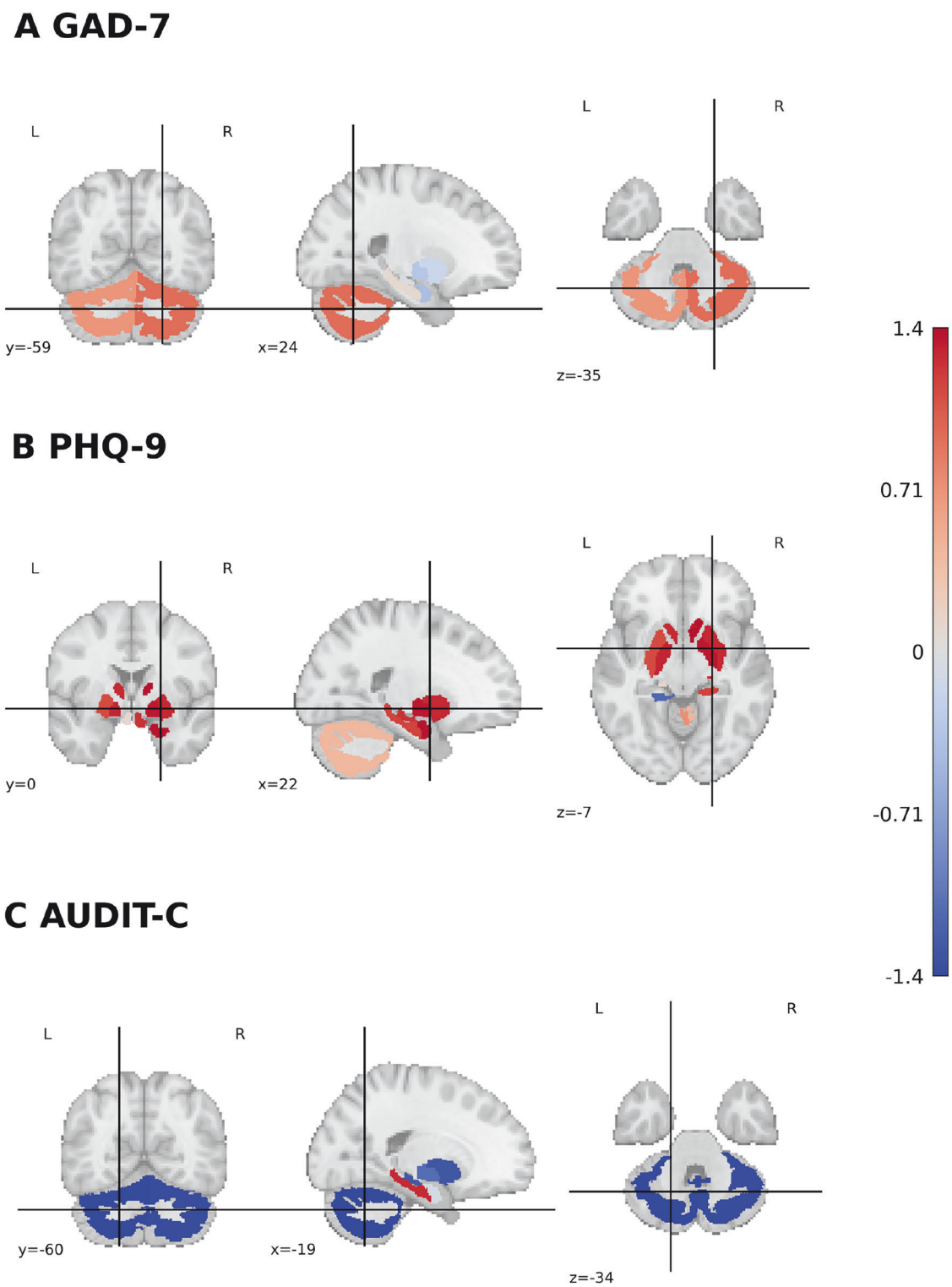


Fig. 4 Regional associations between normative brain deviations and symptom scores. Z-scored associations between brain normative deviations and symptom scores for (**A**) GAD-7 (Generalized Anxiety Disorder-7), (**B**) PHQ-9 (Patient Health Questionnaire-9), and (**C**) AUDIT-C (Alcohol Use Disorders Identification Test-Consumption). Higher values reflect stronger regional associations with individual symptom expression. Associations are z-scored across all symptom scores and brain regions to emphasize relative patterns. These maps provide a heuristic summary of distributed associations between symptom-related deviation dimensions and regional GMV features, and should not be interpreted as maps of focal anatomical effects.

proposed by the Hierarchical Taxonomy of Psychopathology (HiTOP) [50], which groups MDD and GAD within the internalizing spectrum and AUD within the externalizing domain, although it is important to note that some AUD-related alterations in brain structure may arise from the direct neurotoxic effects of alcohol rather than from an externalizing liability.

More broadly, and as a longer-term perspective, this approach may become useful for contextualizing psychiatric diagnoses as

population cohorts increasingly provide imaging and phenotypic data across wider disorder spectra. Future analyses, beyond the scope of the present study, could extend this framework to additional psychiatric phenotypes, including psychotic and bipolar-spectrum conditions, and examine whether normative brain-structural deviation profiles correspond more closely to diagnostic categories, transdiagnostic spectra, or shared symptom dimensions. Analogous to genetic correlations revealing shared etiological architecture across disorders, including the close relationship between schizophrenia and bipolar disorder [51], population-scale normative imaging may help assess convergent or distinct deviation profiles and contribute to evaluating frameworks such as HiTOP, while avoiding the assumption that shared deviation geometry necessarily reflects shared etiology.

The observed improvements in individual-level prediction were concentrated at higher levels of symptom severity, particularly for depression. This pattern is consistent with the limited sensitivity of sMRI to subtle, heterogeneous alterations characteristic of common psychiatric conditions, and suggests that normative brain-structural deviations primarily enhance detection of more pronounced departures from the healthy population rather than enabling robust discrimination across the full symptom spectrum. When evaluated relative to conventional ROI-based GMV features, our normative deviations provided clearer benefits at the level of broader ICD-10-defined diagnostic groups. This likely reflects the ability of contrastive, normative representations to integrate distributed patterns of variation that are not well captured by region-wise measures. For ICD-10-defined AUD, divergence between BACC and AUC may be attributable to the smaller number of available cases ($N = 192$), highlighting the sensitivity of rank-based performance estimates to sample size in this context.

A post hoc mapping of deviation dimensions to regional GMV suggested that cerebellar regions contributed most consistently to symptom-associated deviation dimensions across domains. The strongest effects were observed for AUDIT-C and were negative in bilateral cerebellar cortex, indicating that higher alcohol-related symptom burden was linked to deviation dimensions associated with lower cerebellar GMV, consistent with prior evidence of cerebellar vulnerability in chronic alcohol use [52, 53]. Smaller, less pronounced contributions were also observed in subcortical regions, including the thalamus and caudate. In contrast, cortical contributions were weak and spatially distributed, with no dominant cortical region emerging as a focal contributor for depressive or anxiety-related symptoms. This lack of strong cortical localization is consistent with prior observations of distributed neural correlates in MDD [54] and psychiatric disorders in general [55], and suggests that affective-symptom-related deviations are better interpreted as multivariate, spatially distributed patterns than as isolated cortical abnormalities.

Several limitations should be considered. Temporal offsets between MRI acquisition and symptom assessment in UKB may attenuate brain--symptom associations, particularly for state-dependent symptoms; however, interval-stratified sensitivity analyses showed broadly stable group-level deviation directions when sufficient group sizes were retained. More generally, deviations were modest, heterogeneous, and primarily apparent at the group level, with limited individual-level classification performance. The analyses were restricted to sMRI and therefore do not capture functional, neurochemical, or molecular processes that may be more proximally related to symptoms. Moreover, representations were learned without direct symptom supervision, meaning that some symptom-relevant variation may not be encoded in the learned deviation space. Because Mahalanobis distances, shift estimates, and directional deviation vectors were defined relative to cohort-specific HC references, cross-cohort differences in absolute deviation magnitude or distance from the HC centroid are not directly interpretable. Region-level maps should also be interpreted cautiously, as the spatially unconstrained normative model was not optimized for anatomical localization and post hoc mappings provide indirect summaries; regional GMV features explained only a moderate fraction of variance in MoCo dimensions. Finally, although HCs were screened for low depressive, anxiety, and alcohol-related symptoms and available mood, anxiety, and alcohol-use diagnoses, the harmonized analysis did not allow comprehensive exclusion of other psychiatric conditions. This reflected the focus of available NAKO mental-health phenotypes as well as the unavailability of the required UKB diagnostic data for re-analysis during the revision period. Inclusion of individuals with unmeasured psychiatric conditions would likely increase normative heterogeneity and attenuate deviation estimates, particularly for MDD and GAD if these conditions share affective symptoms or structural correlates. As with all observational analyses, causal inferences cannot be drawn.

In summary, our analyses demonstrate that distributed normative deviations associated with depressive, anxiety, and alcohol-related symptoms show severity-dependent increases in magnitude and structured directional organization at the population level. Operating in a high-dimensional, image-derived representation space enabled the detection of subtle, spatially distributed deviation patterns. The overlap between depression- and anxiety-related deviations and their distinction from alcohol-related profiles is consistent with dimensional, transdiagnostic models of psychopathology. Although predictive gains are modest, the observed deviation axes may provide a way to stratify individuals according to continuous patterns of brain-structural variation that cut across conventional diagnostic categories. In large cohorts, such deviation-based representations may therefore be useful for identifying population-level neurobiological gradients of symptom burden, comorbidity, and severity rather than for making categorical individual-level diagnoses.

Table 2. Balanced accuracy (BACC) and area under the receiver operating characteristic curve (AUC) for all classifications.

Target	Confounders		+ Deviations	
	BACC (%)	AUC (%)	BACC (%)	AUC (%)
GAD-7 (2)	**58 ± 2**	**61 ± 2**	58 ± 2	61 ± 2
GAD-7 (3)	58 ± 6	62 ± 5	**61 ± 2**	**64 ± 3**
GAD-7 (4)	59 ± 9	62 ± 12	**65 ± 8**	**65 ± 6**
ANX ICD-10	55 ± 3	60 ± 4	**62 ± 2**	**63 ± 4**
ANX (diag.)	56 ± 2	58 ± 2	**57 ± 2**	**60 ± 2**
PHQ-9 (1)	58 ± 1	**61 ± 1**	**59 ± 2**	61 ± 1
PHQ-9 (2)	61 ± 2	65 ± 2	**63 ± 2**	**67 ± 2**
PHQ-9 (3)	59 ± 6	65 ± 7	**61 ± 6**	**69 ± 7**
PHQ-9 (4)	63 ± 11	65 ± 26	**70 ± 7**	**74 ± 7**
MDD ICD-10	57 ± 3	61 ± 3	**62 ± 3**	**65 ± 3**
MDD (diag.)	58 ± 1	61 ± 1	**59 ± 1**	**62 ± 1**
AUDIT-C (8)	**60 ± 3**	**65 ± 3**	60 ± 3	65 ± 4
AUDIT-C (9)	63 ± 3	68 ± 4	**64 ± 4**	**69 ± 3**
AUDIT-C ≥10	63 ± 3	68 ± 3	**66 ± 3**	**72 ± 3**
AUD ICD-10	55 ± 9	56 ± 10	**62 ± 4**	**63 ± 4**

Values are mean ± standard deviation over 10-fold stratified cross-validation, sorted by target. Targets are defined by symptom measures: GAD-7 (Generalized Anxiety Disorder-7), PHQ-9 (Patient Health Questionnaire-9), AUDIT-C (Alcohol Use Disorders Identification Test-Consumption), and ICD-10 diagnosis from hospital records or self-reported physician-reported diagnoses (diag.). Bold values indicate the best classifier for each target; in case of ties, the classifier with the smaller standard deviation is preferred, and if ties remain, the one requiring fewer features is chosen (Occam's razor).

DATA AVAILABILITY

Access to and use of NAKO data and biosamples can be obtained via the electronic application portal (https://transfer.nako.de). Access to UK Biobank data requires application through the registration and application portal (http://ukbiobank.ac.uk/register-apply).

CODE AVAILABILITY

All analysis were conducted in Python (version 3.12). All analysis scripts and the utilized packages are available via GitHub at https://github.com/wiegertj/DeepNormativeModeling.

REFERENCES

1. GBD 2019 Mental Disorders Collaborators, G. M. D. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. The Lancet Psychiatry. 2022;9:137–50. https://doi.org/10.1016/S2215-0366(21)00395-3.
2. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry. 2015;72:334–41. https://doi.org/10.1001/jamapsychiatry.2014.2502.
3. Wu Y, Wang L, Tao M, Cao H, Yuan H, Ye M, et al. Changing trends in the global burden of mental disorders from 1990 to 2019 and predicted levels in 25 years. Epidemiol Psychiatr Sci. 2023;32:e63 https://doi.org/10.1017/S2045796023000756.
4. Cui L, Li S, Wang S, Wu X, Liu Y, Yu W, et al. Major depressive disorder: hypothesis, mechanism, prevention and treatment. Signal Transduct Target Ther. 2024;9:30 https://doi.org/10.1038/s41392-024-01738-y.
5. Wittchen HU, Hoyer J. Generalized anxiety disorder: nature and course. J Clin Psychiatry. 2001;62:15–9. Suppl. 11 discussion 20-1
6. Isometsä ET. Suicides in mood disorders in psychiatric settings in nordic national register-based studies. Front Psychiatry. 2020;11:721 https://doi.org/10.3389/fpsyt.2020.00721.
7. Brown TA, Campbell LA, Lehman CL, Grisham JR, Mancill RB. Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. J Abnorm Psychol. 2001;110:585–99. https://doi.org/10.1037/0021-843X.110.4.585.
8. Kendler KS, Gardner CO, Gatz M, Pedersen NL. The sources of co-morbidity between major depression and generalized anxiety disorder in a swedish national twin sample. Psychol Med. 2006;37:453–62. https://doi.org/10.1017/S0033291706009135.
9. Chen Z, Tang Y, Liu X, Li W, Hu Y, Hu B, et al. Edge-centric connectome-genetic markers of bridging factor to comorbidity between depression and anxiety. Nat Commun. 2024;15:10560 https://doi.org/10.1038/s41467-024-55008-0.
10. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR D study. J Affect Disord. 2015;172:96–102. https://doi.org/10.1016/j.jad.2014.10.010.
11. Nandi A, Beard JR, Galea S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. BMC Psychiatry. 2009;9:31 https://doi.org/10.1186/1471-244X-9-31.
12. Belov V, Erwin-Grabner T, Aghajani M, Aleman A, Amod AR, Basgoze Z, et al. Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. Sci Rep. 2024;14:1084 https://doi.org/10.1038/s41598-023-47934-8.
13. Rutherford S, Kia SM, Wolfers T, Fraza C, Zabihi M, Dinga R, et al. The normative modeling framework for computational psychiatry. Nat Protoc. 2022;17:1711–34. https://doi.org/10.1038/s41596-022-00696-5.
14. Bozek J, Griffanti L, Lau S, Jenkinson M. Normative models for neuroimaging markers: Impact of model selection, sample size and evaluation criteria. Neuroimage. 2023;268:119864 https://doi.org/10.1016/j.neuroimage.2023.119864.
15. Floris DL, Wolfers T, Zabihi M, Holz NE, Zwiers MP, Charman T, et al. Atypical brain asymmetry in autismâ€"a candidate for clinically meaningful stratification. Biol Psychiatry: Cogn Neurosci Neuroimaging. 2021;6:802–12. https://doi.org/10.1016/j.bpsc.2020.08.008.
16. Wolfers T, Rokicki J, Alnaes D, Berthet P, Agartz I, Kia SM, et al. Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. Hum Brain Mapp. 2021;42:2546–55. https://doi.org/10.1002/hbm.25386.
17. Verdi S, Marquand AF, Schott JM, Cole JH. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. Brain. 2021;144:2946–53. https://doi.org/10.1093/brain/awab165.
18. Sun X, Sun J, Lu X, Dong Q, Zhang L, Wang W, et al. Mapping neurophysiological subtypes of major depressive disorder using normative models of the functional connectome. Biol Psychiatry. 2023;94:936–47. https://doi.org/10.1016/j.biopsych.2023.05.021. Mood Disorder: A Lifespan Perspective on Neurobiology
19. Heikkinen N, Niskanen E, Könönen M, Tolmunen T, Kekkonen V, Kivimäki P, et al. Alcohol consumption during adolescence is associated with reduced grey matter volumes. Addiction. 2017;112:604–13. https://doi.org/10.1111/add.13697.
20. Li L, Yu H, Liu Y, Meng YJ, Li XJ, Zhang C, et al. Lower regional grey matter in alcohol use disorders: evidence from a voxel-based meta-analysis. BMC Psychiatry. 2021;21:247 https://doi.org/10.1186/s12888-021-03244-9.
21. Eitel F, Schulz M-A, Seiler M, Walter H, Ritter K. Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. Exp Neurol. 2021;339:113608 https://doi.org/10.1016/j.expneurol.2021.113608.
22. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. Hum Brain Mapp. 2018;40:944–54.
23. Pinaya WHL, Scarpazza C, Garcia-Dias R, Vieira S, Baecker L, F da Costa P, et al. Using normative modelling to detect disease progression in mild cognitive impairment and alzheimer's disease in a cross-sectional multi-cohort study. Sci Rep. 2021;11:15746.
24. Walther, L, Vogelgesang F, Rosario AS, Kersjes C, Thom J, Peitz D et al. Depressive und angstsymptomatik bei erwachsenen in deutschland: Ergebnisse aus dem panel gesundheit in deutschland 2024. *J Health Monit* 18 (2025). https://doi.org/10.25646/13554.
25. Schipf S, Schöne G, Schmidt B, Günther K, Stübs G, Greiser KH, et al. Die Basiserhebung der NAKO Gesundheitsstudie: Teilnahme an den Untersuchungsmodulen, QualitÃ¤tssicherung und Nutzung von Sekundärdaten. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2020;63:254–66. https://doi.org/10.1007/s00103-020-03093-z.
26. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9. https://doi.org/10.1038/s41586-018-0579-z.
27. Hedges EP, Dimitrov M, Zahid U, Brito Vega B, Si S, Dickson H, et al. Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. Neuroimage. 2021;246:118751.
28. Liang J, Yu Q, Liu Y, Qiu Y, Tang R, Yan L, et al. Gray matter abnormalities in patients with major depressive disorder and social anxiety disorder: a voxel-based meta-analysis. Brain Imaging Behav. 2023;17:749–63. https://doi.org/10.1016/j.jad.2011.03.049.
29. Bora E, Fornito A, Pantelis C, Yücel M. Gray matter abnormalities in major depressive disorder: a meta-analysis of voxel based morphometry studies. J Affect Disord. 2011;138:9–18.
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
31. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166:1092–7. https://doi.org/10.1001/archinte.166.10.1092.
32. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Arch Intern Med. 1998;158:1789–95. https://doi.org/10.1001/archinte.158.16.1789.
33. Bamberg F, Schlett CL, Caspers S, Ringhof S, Günther M, Hirsch JG, et al. Baseline MRI Examination in the NAKO Health Study. Dtsch Arztebl International. 2024;121:587–93. https://doi.org/10.3238/arztebl.m2024.0151.
34. Dutt RK, Hannon K, Easley TO, Griffis JC, Zhang W, Bijsterbosch JD. Mental health in the UK biobank: A roadmap to self-report measures and neuroimaging correlates. Hum Brain Mapp. 2021;43:816–32. https://doi.org/10.1002/hbm.25690.
35. Shabanian, M, Eckstein, E, Chen, H & DeVincenzo, J Classification of neurodevelopmental age in normal infants using 3d-cnn based on brain mri 2373-8 (2019) https://doi.org/10.1109/BIBM47256.2019.8983399.
36. Frey, M & Nau, M Memory efficient brain tumor segmentation using an autoencoder-regularized u-net In Crimi, A & Bakas, S (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 388-96 (Springer International Publishing, 2020) https://doi.org/10.1007/978-3-030-46640-4_37.
37. He, K, Fan, H, Wu, Y, Xie, S & Girshick, R Momentum contrast for unsupervised visual representation learning In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729-38 (2020) https://doi.org/10.1109/CVPR42600.2020.00975.
38. Chen, X, Xie, S & He, K An empirical study of training self-supervised vision transformers In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9620-9 (2021) https://doi.org/10.1109/ICCV48922.2021.00950.
39. Li, J, Selvaraju RR, Gotmare AD, Joty S, Xiong C, Hoi SCH. Align before fuse: vision and language representation learning with momentum distillation In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21 (Curran Associates Inc., Red Hook, NY, USA, 2021).
40. Qiu, J, Zhu J, Liu S, Han W, Zhanf J, Duan C, et al. Automated cardiovascular record retrieval by multimodal learning between electrocardiogram and clinical report In Hegselmann, S et al. (eds.) *Proceedings of the 3rd Machine Learning for Health Symposium, vol. 225 of Proceedings of Machine Learning Research*, 480-97 (PMLR, 2023). URL https://proceedings.mlr.press/v225/qiu23a.html.
41. A Paszke, S Gross, S Chintala, G Chanan, E Yang, Z DeVito et al. Automatic differentiation in PyTorch (2017).

42. Reprint of: Mahalanobis, p.c. (1936). on the generalised distance in statistics. Sankhya A. 2018;80:1–7. https://doi.org/10.1007/s13171-019-00164-5. URL
43. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Society: Series B (Methodological). 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
44. Storey JD. A direct approach to false discovery rates. J R Stat Soc Series B: Stat Methodol. 2002;64:479–98. https://doi.org/10.1111/1467-9868.00346.
45. Zou H, Hastie T. Regularization and variable selection via the Elastic Net. J R Stat Soc Series B: Stat Methodol. 2005;67:301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.
46. Sindermann L, Redlich R, Opel N, Böhnlein J, Dannlowski U, Leehr EJ. Systematic transdiagnostic review of magnetic-resonance imaging results: Depression, anxiety disorders and their co-occurrence. J Psychiatr Res. 2021;142:226–39. https://doi.org/10.1016/j.jpsychires.2021.07.022.
47. Mei L, Gao Y, Chen M, Zhang X, Yue W, Zhang D, et al. Overlapping common genetic architecture between major depressive disorders and anxiety and stress-related disorders. Prog Neuro-Psychopharmacol Biol Psychiatry. 2022;113:110450 https://doi.org/10.1016/j.pnpbp.2021.110450.
48. Hettema JM. What is the genetic relationship between anxiety and depression? Am J Med Genet Part C: Semin Med Genet. 2008;148C:140–6. https://doi.org/10.1002/ajmg.c.30171.
49. Castillo-Carniglia A, Keyes KM, Hasin DS, Cerdá M. Psychiatric comorbidities in alcohol use disorder. Lancet Psychiatry. 2019;6:1068–80. https://doi.org/10.1016/S2215-0366(19)30222-6.
50. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. J Abnorm Psychol. 2017;126:454–77. https://doi.org/10.1037/abn0000258.
51. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984–94.
52. Sullivan EV, Zahr NM, Saranathan M, Pohl KM, Pfefferbaum A. Convergence of three parcellation approaches demonstrating cerebellar lobule volume deficits in alcohol use disorder. NeuroImage: Clinical. 2019;24:101974 https://doi.org/10.1016/j.nicl.2019.101974.
53. Torvik A, Torp S. The prevalence of alcoholic cerebellar atrophy: A morphometric and histological study of an autopsy material. J Neurol Sci. 1986;75:43–51. https://doi.org/10.1016/0022-510X(86)90049-3.
54. Winter NR, Leenings R, Ernsting J, Sarink K, Fisch L, Emden D, et al. Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. JAMA Psychiatry. 2022;79:879–88. https://doi.org/10.1001/jamapsychiatry.2022.1780.
55. Segal A, Parkes L, Aquino K, Kia SM, Wolfers T, Franke B, et al. Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. Nat Neurosci. 2023;26:1613–29. https://doi.org/10.1038/s41593-023-01404-6.

ACKNOWLEDGEMENTS

The project was conducted with data (NAKO-711) from the German National Cohort (NAKO) (http://www.nako.de/). The NAKO is funded by the Federal Ministry of Education and Research (BMBF) [project funding reference numbers: 01ER1301A/B/C, 01ER1511D, and 01ER1801A/B/C/D and 01ER2301A/B/C], federal states of Germany and the Helmholtz Association, the participating universities and the institutes of the Leibniz Association. We thank all participants who took part in the NAKO study and the staff of this research initiative. We also thank the participants and scientists involved in making the UK Biobank resource available (http://www.ukbiobank.ac.uk/). This study was conducted under UK Biobank application number 162313. The project was supported by the DZPG (German Centre for Mental Health Research) and by the BMBF (German Ministry of Education and Research) grant 01EE2303E. Fabian Streit is supported by a 2023 NARSAD Young Investigator Grant (#31537) from the Brain & Behavior Research Foundation with support from the Families for Borderline Personality Disorder Research. This work was supported by the Hector foundation II and was endorsed by German Center for Mental Health (DZPG).

AUTHOR CONTRIBUTIONS

Julius Wiegert: Conceptualization, Methodology, Formal analysis, Data curation, Writing -- original draft. Sebastián Marty-Lombardi: Validation (code analysis), Writing -- review & editing. Jailan Oweda: Data curation (imaging preprocessing). Esra Lenz: Data curation (data extraction). Antonia Mai, Amrou Abas, Xiuzhi Li: Data curation (imaging preprocessing support and genomic data processing). Johannes Nitsche: Methodology (statistical analysis design), Writing -- review & editing. Joonas Naamanka, Sebastian Volkmer, Andreas Meyer-Lindenberg: Writing -- review & editing. Tobias Gradinger, Fabian Streit, Urs Braun, Emanuel Schwarz: Conceptualization, Study design, Supervision. Peter Ahnert, Klaus Berger, Hermann Brenner, Hans J. Grabe, Karin Halina Greiser, Johanna Klinger-König, André Karch, Michael Leitzmann, Claudia Meinke-Franz, Rafael Mikolajczyk, Thoralf Niendorf, Oliver Sander, Carsten Oliver Schmidt, Steffi G. Riedel-Heller, Annette Peters, Tobias Pischon: Resources (NAKO dataset), Writing -- review & editing. Frauke Nees, Stephanie Witt, Kerstin Ritter, Josef Frank: Writing -- review & editing. All authors have reviewed and approved the manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

HJG has received travel grants and speakers honoraria from Neuraxpharm, Servier, Indorsia and Janssen Cilag. ES received speaker fees from bfd buchholz-fachinformationsdienst GmbH, Lundbeckfonden, and Janssen-Cilag GmbH, as well as editorial fees from Lundbeckfonden and the Wellcome Trust. AML has received consultancy honoraria from AbbVie, Janssen-Cilag GmbH, Boehringer-Ingelheim, Daimler und Benz Stiftung, Helmut Horten Stiftung, Neurotorium/Lundbeckfonden, Hector Stiftung, Endosane Pharmaceuticals, Elsevier, von Behring-Röntgen-Stiftung, The LOOP Zürich, ECNP, Teva, Medical Research Council/UKRI, Heinrich-Lanz-Stiftung, Johnson & Johnson, Lundbeckfonden, and the Wellcome Trust. He has received lecture honoraria from pro Mente Akademie GmbH, Schön Klinik, Janssen-Cilag, Evangelische Hochschule Ludwigsburg, Landesärztekammer Baden-Württemberg, Klinikum Ingolstadt, PSY (Psychiatrie und Psychotherapie Update Refresher, FOMF), Consorcio Mexicano de Neuropsicofarmacologa (MCNP), Universität Klagenfurt, and Universität Norwalk/USA. He has received editorial honoraria (as editor, etc.) from ECNP/Neuroscience Applied and JSPS. He has received authorship honoraria from Beltz Verlag, Thieme Verlag, and Kohlhammer Verlag. He has received project funding from BMBF, DFG, Hector Stiftung, Klaus Tschira Stiftung, and MWK.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-026-03691-4.

Correspondence and requests for materials should be addressed to Emanuel Schwarz.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026