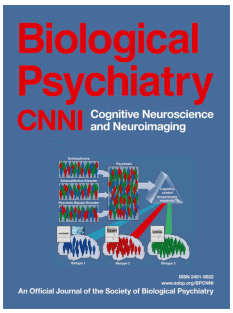


Voxel-wise multivariate analysis of brain-psychosocial associations in adolescents reveals six latent dimensions of cognition and psychopathology

Rick A. Adams, Cemre Zor, Agoston Mihalik, Konstantinos Tsirlis, Mikael Brudfors, James Chapman, John Ashburner, Martin P. Paulus, Janaina Mourão-Miranda

PII: S2451-9022(24)00085-5

DOI: https://doi.org/10.1016/j.bpsc.2024.03.006

Reference: BPSC 1204

To appear in: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*

Received Date: 3 August 2023

Revised Date: 15 March 2024

Accepted Date: 28 March 2024

Please cite this article as: Adams R.A, Zor C., Mihalik A., Tsirlis K., Brudfors M., Chapman J., Ashburner J., Paulus M.P & Mourão-Miranda J., Voxel-wise multivariate analysis of brain-psychosocial associations in adolescents reveals six latent dimensions of cognition and psychopathology, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2024), doi: https://doi.org/10.1016/j.bpsc.2024.03.006.

Voxel-wise multivariate analysis of brain-psychosocial associations in adolescents reveals six latent dimensions of cognition and psychopathology

Short title: Latent mental health and cognition modes in adolescents

Rick A Adams[1,2]*†, Cemre Zor[1]*, Agoston Mihalik[1,2,3], Konstantinos Tsirlis[1,2], Mikael Brudfors[4,5], James Chapman[1,2], John Ashburner[4], Martin P Paulus[6], Janaina Mourão-Miranda[1,2]

*equal contribution

[1]Centre for Medical Image Computing, Department of Computer Science, University College London, WC1V 6LJ, UK

[2]Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, WC1B 5EH, UK

[3]Department of Psychiatry, University of Cambridge, CB2 0SZ, UK

[4]Wellcome Centre for Human Neuroimaging, University College London, WC1N 3BG, UK

[5]School of Biomedical Engineering and Imaging Sciences, King's College London, SE1 7EH, UK

[6]Laureate Institute for Brain Research, Tulsa, Oklahoma, USA

†corresponding author: rick.adams@ucl.ac.uk

Centre for Medical Image Computing, Department of Computer Science, University College London, 90 High Holborn, London, WC1V 6LJ, UK

Keywords: brain-behaviour associations; partial least squares; neurodevelopment; psychopathology; machine learning; structural MRI

Abstract

Background

Adolescence heralds the onset of much psychopathology, which may be conceptualized as an emergence of altered covariation between symptoms and brain measures. Multivariate methods can detect such modes of covariation or latent dimensions, but none specifically relating to psychopathology have yet been found using population-level structural brain data. Using voxel-wise (instead of parcellated) brain data may strengthen latent dimensions' brain-psychosocial relationships, but this creates computational challenges.

Methods

We obtained voxel-wise grey matter density and psychosocial variables from the baseline (aged 9-10 years) Adolescent Brain and Cognitive Development cohort (n=11288), and employed a state-of-the-art segmentation method, sparse partial least squares, and a rigorous machine learning framework to prevent overfitting.

Results

We found six latent dimensions, four pertaining specifically to mental health. The mental health dimensions related to overeating, anorexia/internalizing, oppositional symptoms (all $p<0.002$) and ADHD symptoms ($p=0.03$). ADHD related to increased and internalizing related to decreased grey matter density in dopaminergic and serotonergic midbrain areas, whereas oppositional symptoms related to increased grey matter in a noradrenergic nucleus. Internalizing related to increased and oppositional symptoms to reduced grey matter density in insula, cingulate and auditory cortices. Striatal regions featured strongly, with reduced caudate nucleus grey matter in ADHD, and reduced putamen grey matter in oppositional/conduct problems. Voxel-wise grey matter density generated stronger brain-psychosocial correlations than brain parcellations.

Conclusions

Voxel-wise brain data strengthen latent dimensions of brain-psychosocial covariation and sparse multivariate methods increase their psychopathological specificity. Internalizing and externalizing are associated with opposite grey matter changes in similar cortical and subcortical areas.

Introduction

Large scale datasets comprising both 'biopsychosocial' and neuroimaging measures – such as UK Biobank (1) and the Adolescent Brain Cognitive Development (ABCD) study (2), each with n>10,000 participants – hold great promise for the discovery of complex associations between the brain, its genetic heritage, and its sociocultural environment. In such datasets one can go beyond standard univariate analyses based on *a priori* symptom clusters (e.g., diagnoses) or brain regions, and examine latent dimensions (or modes) of covariation across many brain and psychosocial variables (3), using multivariate statistical methods such as canonical correlation analysis (CCA) (4) and partial least squares (PLS) (5). The first attempt to do so applied CCA to psychosocial and functional connectivity measures in n=461 participants from the Human Connectome Project (6), and found a single mode of covariation resembling a 'general intelligence' factor that also included social outcomes (e.g. income, life satisfaction versus drug and alcohol use), relating to connectivity in predominantly 'default mode' brain areas (7). Similar modes to this have subsequently been found in other datasets (8,9) and using other methods (10).

A potential disadvantage of the multivariate approach, however, is that the resulting modes of brain-psychosocial covariation contain so many psychosocial variables that their complexity hinders interpretation: indeed, it has been difficult to robustly identify specifically mental health-related modes within large scale datasets that relate brain structure to psychopathology (7–9) (unless the psychosocial data are confined to mental health items only (11–13)). Furthermore, effect sizes are low (14).

One potential way to improve brain-psychosocial correlations may be to use voxel-wise instead of parcellation-based brain data, although this is rarely attempted in large-scale population datasets (with some exceptions (15)). This potentially greater accuracy comes at a substantial computational cost, however, and the risk that latent dimensions will be even harder to interpret and subject to much more 'overfitting' (16). Interpretability can be much improved, however, using sparse methods – such as sparse PLS – that find modes containing fewer variables (11,17) (and also parcellations – and frameworks such as NeuroSynth – can always be applied to voxel-wise data *post hoc*). Notably, if the underlying brain-psychosocial association is also sparse (e.g., the hippocampus-centred atrophy in dementia), sparse PLS can outperform CCA methods at recovering the 'true' associations in simulated data, with comparable out-of-sample correlations (10). Conversely, if the underlying brain-psychosocial

association is highly distributed, a sparse method will generally produce modes with weaker out-of-sample correlations than non-sparse methods (10).

'Overfitting' means discovering spurious associations that will not generalise to unseen datasets. When the dimensionality of the data exceeds the number of subjects – as voxel-wise imaging data does, by at least an order of magnitude – this problem is especially acute. It can be mitigated by using regularization, both to enable solutions when the dimensionality exceeds the sample size and to reduce overfitting, and a rigorous machine learning-based approach, in which associations detected in a training sample are then assessed for statistical significance in an unseen hold-out sample, using permutation-based statistics. Regularization can be performed during the optimization of CCA/PLS itself, as in L2-norm (ridge) or L1-norm (sparse) regularizations, which force weights to be small but non-zero, or sparse, respectively.

These analyses are based on the premise that robust estimations of the covariation patterns between imaging and behavioural measures can uncover novel patterns of brain structure or function underlying mental health disorders (18–23). Changes in structural brain measures in adolescence may relate to neurodevelopmental processes such as synaptic pruning and myelination, which are thought to be altered in some mental disorders (e.g., schizophrenia (24) and compulsive and impulsive disorders (25)). In particular, although novel dimensional and/or hierarchical symptom-based classifications have been proposed (26,27), it is likely that including neurobiological information – e.g., imaging – will help to delineate which disorders or symptoms are related or distinct, and describe their key modes of variation (28). For example, a study using CCA to link clinical and functional brain measures in children found some 'transdiagnostic' modes (comprising irritability, anxiety and ADHD, and also irritability and disruptive behaviour) and one unique mode (anxiety alone) (23). A similar study combining structural and functional imaging data with clinical data from 19 year olds found a similar transdiagnostic (irritability, anxiety/depression and inattention) mode and a more unique (ADHD-related) mode (11). Scores on such modes could potentially then be used for risk stratification (21) or outcome prediction (29). Doing so in adolescence is especially important: it is when many mental health disorders manifest (30), and thus may be the optimal time to try to mitigate or prevent their onset. However, mental health-related effects may be more weakly expressed in the brain in adolescence compared to adulthood, and adolescent brains differ from standard adult templates, e.g., those used to distinguish grey and white matter.

In this study, we address the above problems with high dimensional multivariate analyses by using a sparse method – sparse partial least squares (SPLS) (31) – to increase the

likelihood of identifying robust modes that are specific (e.g., to mental health), a rigorous machine learning framework (17) to prevent overfitting, and a state-of-the-art segmentation method (32) that generalises well to non-standard (e.g., adolescent) images. We chose not to constrain the psychosocial variables to only mental health-related items (except for a sensitivity analysis in the Supplement), because many other factors may affect mental health – and neurodevelopment more broadly – in childhood, e.g., poverty, local deprivation, access to recreational activities and racism. We analysed the ABCD dataset: the largest adolescent psychosocial and neuroimaging dataset (n=11880). It is a population (i.e., non-clinical) sample, but it is enriched for children with behavioural problems (33). We used only the structural magnetic resonance imaging (sMRI) data, as functional connectivity can have lower reliability (34) and replicability (23) and be more challenging to interpret, and whilst functional measures sometimes have a slightly stronger relationship to psychopathology than structural ones, the difference does not tend to be large (and sometimes the reverse is true) (3,11,12,14,14,29,34,35).

Our key questions were i) whether the computational cost of using voxel-wise grey matter variables (instead of parcellation-based measures, which are more standard at this scale) would be compensated by increasing the sensitivity (i.e., the brain-psychosocial correlation) of the results, and ii) whether employing a sparse approach (SPLS) could find specific, interpretable latent dimensions within these very high dimensional data – especially relating to psychopathology.

Methods

The ABCD study is a single-cohort prospective longitudinal study of n=11880 children aged 9-10 years (and their parents/guardians) recruited across 22 sites, to be followed up for at least 10 years (https://abcdstudy.org). We downloaded structural magnetic resonance imaging (sMRI) data, and behavioural, clinical, cognitive and sociodemographic (termed 'psychosocial' below) data from the (baseline) ABCD Study Curated Annual Release 2.0.1. The analysis is detailed in the Supplement and is summarised below and in Figure 1. The full acquisition protocol, including the small differences between scanners, is fully described elsewhere (2) (Section 2 and Appendix).

In brief, voxel-wise grey matter images were obtained from the sMRI data using the Multi-Brain toolbox (32), which produces spatially normalised tissue segmentations for each scan. The individual grey matter maps were then smoothed in SPM12 (www.fil.ion.ucl.ac.uk/spm) using a 10 mm FWHM Gaussian kernel, and then averaged to create an average grey matter map. A common mask was created, selecting voxels with a >10% probability of containing grey matter in the average grey matter map, and applied to each participant, resulting in a total of 124398 voxels, each scored with the probability of its containing grey matter in that participant (brain data). Psychosocial questionnaire variables potentially relevant to mental health were selected (see Figure 1 and Supplement). The data were then deconfounded (for sex, age, total brain volume, body mass index, and scanner type) and standardized (i.e. the standardized variables had mean zero and standard deviation one).

Sparse PLS was then employed to find latent dimensions (or modes) of maximal covariation between the voxel-wise grey matter density and psychosocial variables, defined by $\boldsymbol{Xu}$ and $\boldsymbol{Yv}$ (where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the data matrices containing brain and psychosocial data, respectively, and $\boldsymbol{u}$ and $\boldsymbol{v}$ are the weight vectors which indicate the contribution of each variable to the mode of covariation). The level of sparsity of the weight vectors is controlled by hyper-parameters c_u and c_v, which were optimised using five-fold cross-validation within the training set (80% of participants), using joint 'generalizability' and 'stability' optimization criteria. These select hyperparameters that maximise the mean out-of-sample correlation across all folds, and the average overlap of the weights, respectively. Once the hyper-parameters were optimised, the SPLS model was fitted to the whole training set, and assessed using the correlation of the latent dimensions in the hold-out set (20% participants). Permutation testing was used to assess the significance of the hold-out correlation. In addition to visualizing the

sparse weights, we also plotted the non-sparse brain loadings (the correlations of each variable with the mode score: see the Supplement) to show how each mode relates to the entire cortex.

To assess the benefit of using voxel-wise brain data, as opposed to parcellation-based brain data of substantially lower dimensionality, we repeated the machine learning analysis using mean grey matter volumes of 166 anatomical parcels of cortex, cerebellum and thalamus (Anatomical Labelling atlas 3, or AAL3 (36)), or of a network-based parcellation of 234 areas (the 'Schaefer-Choi-Buckner' atlas (37–40)), both detailed in the Supplement.

We performed a principal component analysis (PCA) of the psychosocial variables alone, and assessed the extent to which the brain-psychosocial SPLS modes captured the largest sources of variance in the psychosocial data (7) (see the Supplement for details).

Finally, we compared the extent to which psychosocial scores on the SPLS modes (from ABCD data collected at baseline, year 0) and scores from the psychosocial-only principal components correlated with important outcomes in years 1-3, such as school grades, suicidality and substance use: see the Supplement for details.

Results

Six modes relate voxel-wise brain grey matter to psychosocial measures

Six latent dimensions (or modes) of brain-psychosocial covariance achieved statistical significance (assessed using the Pearson correlation) in the hold-out set (Figures 2-4). We illustrate the loadings of both brain and psychosocial modes in Figures 2-4 (psychosocial variables with loadings of $|r|>0.15$ – up to a maximum of 40 – are shown in the Figures, but the full list of weights on each mode is in the Supplementary Data). The brain and psychosocial weights, parameter optimization details and correlations between brain and psychosocial scores in training and hold-out sets are plotted in the Supplement (Figures S1-6), along with the anatomical details of both loadings and weights on cortical, cerebellar and subcortical regions (Tables S5-S16 and S17-S22 respectively). We describe each of them in detail below.

Mode 1 (r=0.30, p<0.002; Figure 2A, Tables S5-6) relates higher intelligence and socioeconomic status (e.g., income, lack of neighbourhood violence/crime, extracurricular activities, with poor sleep and a single parent at the opposite end) to lower grey matter in frontoparietal, medial temporal and cerebellar regions and increased grey matter in primary sensorimotor and thalamic regions.

Mode 2 (r=0.24, p<0.002; Figure 2B, Tables S7-8) describes overeating (being overweight, overeating, binging, weight control) or issues related to sleep apnoea (snoring, overtired). High loadings (i.e. obesity) are associated with increased occipitotemporal and orbitofrontal (i.e. visual and limbic) grey matter.

Mode 3 (r=0.25, p<0.002; Figure 3A, Tables S9-11) ranges from high scorers who have a Hispanic background and family from outside the US, speak Spanish (or another language) besides English, and are Catholic, to low scorers who are not Hispanic, and whose parents speak English well (but no other language) and never married. High scorers had more grey matter in temporal (auditory, operculum) and parietal – i.e. language-related – areas and precuneus, and less grey matter in frontal and motor cortex (especially left-sided) and temporal pole.

Mode 4 (r=0.042, p=0.030; Figure 3B, Tables S11-12) comprises symptoms of ADHD, including the three key domains (hyperactivity, impulsivity and inattention). These symptoms related to reduced grey matter in subcortical and cortical limbic (and attentional) areas –

caudate (accounting for >60% of the weights), most of the thalamus, nucleus accumbens, hippocampus, amygdala, insula and orbitofrontal cortex – and increased grey matter in substantia nigra, ventral tegmental area, raphe nucleus, and areas of cortex, cerebellum and thalamus relating to sensorimotor function. Given the weakness of this correlation and the potential for confounding by medication or movement, we performed additional analyses to ensure the results were robust (see Supplement and Figures S4 and S7A).

Mode 5's (r=0.14, p<0.002; Figure 4A, Tables S13-14) psychosocial loadings can best be summarised as relating anorexia and internalizing symptoms such as anxiety (worries, fearful, nervous, self-conscious), depression (feeling worthless, unhappy, guilty) and obsessionality (can't ignore certain thoughts, picks nose/skin, fears s/he might do something bad) and lack of activities to increased grey matter in frontal (especially medial and orbital) and insular/auditory cortex and lower grey matter in raphe nucleus, substantia nigra and ventral tegmental area, and thalamus.

Mode 6 (r=0.10, p<0.002; Figure 4B, Tables S15-16) accounts for children with parental conflict and Oppositional Defiant Disorder (ODD) symptoms (arguing with authority, temper tantrums), and – loading more weakly – some Conduct Disorder symptoms (disobeying rules, cruelty, lying, threatening others) which were associated with increased grey matter in caudate, locus coeruleus, amygdala and right rostral prefrontal cortex, and reduced grey matter in insular, cingulate and auditory cortex, and nucleus accumbens.

The within-modality variances or cross-modality covariance in the hold-out set explained by each mode are shown in Figure 5A. In terms of brain-psychosocial covariance explained, mode 1 (15.7%) dominates the others (1.8-5.1%): in total, 13.3% brain variance, 10.1% psychosocial variance, and 34.2% brain-psychosocial covariance was explained by the six voxel-wise SPLS modes. Subgroup analyses, in which the brain-psychosocial correlations in the same six modes are assessed separately within distinct groups (by sex and socioeconomic status) are detailed in Table S23. All correlations were present in all subgroups, implying generalizability of the results.

Comparative analyses

Alternative analyses were conducted in which the voxel-wise brain grey matter probabilities were averaged within 166 anatomical parcels (AAL3 atlas) or 234 network-based parcels (Schaefer-Choi-Buckner atlas). The analyses were otherwise identical to the voxel-wise SPLS

analysis. The results are detailed in the Supplement (Figures S8-S15). The anatomical and network-based analyses found five and three (respectively) modes, all of which (except one) were very similar to the corresponding modes of the voxel-wise analysis: confusion matrices demonstrating the correlation between averaged brain and psychosocial scores on these modes are shown in Figure 5C (see Figure S7B for separate confusion matrices for brain and psychosocial modalities).

The main difference between the parcellation and voxel-wise analyses was that the correlations in most modes (except mode 4) were smaller when using the parcellation methods (Figure 5D). (The voxel-wise analysis also found more significant modes, but this was not a consistent attribute: e.g., if BMI was not included as a confound, the parcellation-based analyses found more modes – not shown.)

The sensitivity analysis using only mental health-related psychosocial variables found three mental health-related modes (Figures S16-18). One of these was similar to Mode 2, and the others related to mixtures of conduct problems and IQ, and internalizing and externalizing (Figures 5B and S19A; discussed in the Supplement).

See the Supplement for comparison of the brain-psychosocial modes from the main analysis with the psychosocial principal components (text and Figures S19B and S20); the latter had slightly stronger correlations with future outcomes than the former.

Discussion

The goal of this study was to examine whether latent dimensions of brain structure and psychosocial covariation – including those specifically relating to psychopathology – could be found in adolescents. Using a sparse partial least squares approach, we identified six latent dimensions (or modes of covariation) in n=11288 nine- and ten-year-olds, which explained a third of the brain-psychosocial covariance. Three modes showed moderate correlations (r=0.24-0.30) with psychosocial variables relating to intelligence and income, overeating, and being bilingual (with a Hispanic background). Four modes related to mental health – binge eating, ADHD symptoms, anorexia and internalizing, and Oppositional Defiant symptoms (respectively) – but the last three showed weaker correlations to grey matter density (r<0.15). Nevertheless, the fifth mode adds to recent evidence that dopamine function may relate to internalizing and depression symptoms (as well as ADHD), and the sixth mode links noradrenaline to conflict and defiance (discussed below). The voxel-wise SPLS analyses showed stronger brain-psychosocial correlations than parcellation-based analyses. Another advantage of the SPLS analysis is that the sparsity can improve specificity and interpretability – especially of the psychosocial weights, which for modes 2, 3 and 6 contain only 2, 8 and 7 psychosocial features (Figures S2-S4).

We now evaluate each mode in turn. Mode 1 related higher IQ and socioeconomic status to lower grey matter in attentional and executive (frontoparietal) regions, but higher grey matter in sensorimotor regions. Indeed, a previous multivariate analysis of ABCD data (9) found a relation between IQ, income and reduced anterior but increased posterior cortical thickness (mode 5). Note that thinner cortex is associated with higher intelligence in children aged ~10 years; in adults, intelligence relates to thicker cortex (41,42). Other studies have seen relationships of IQ to increased surface area (and reduced thickness) in similar prefrontal and parietal areas and the opposite relationships in occipital and motor areas (43,44) – just as in mode 1. IQ is most likely an effect rather than a cause of these changes, however. Regarding likely causes, socioeconomic status is strongly related to IQ (8) and deprived backgrounds may in part drive these cortical grey matter effects via a lack of educational stimulation (45). Indeed, mode 1's brain loadings are very similar to the associations of cortical thickness with socioeconomic status (46) (and of cortical surface area with family income (47)), which may reflect accelerated adolescent cortical thinning induced by deprivation. Indeed, lower cerebellar grey matter (which dominates the positive weights for this mode: Table S17) may relate to environmental aspects of socioeconomic status (48). Unfortunately, most

neuroscience studies ignore socioeconomic status despite its clear effects on neurobiology (49) and physical and mental health (50). Longitudinal analyses are required to disambiguate the respective contributions of IQ and socioeconomic status.

Mode 2 related overeating to increased occipitotemporal and orbitofrontal grey matter. This replicates a finding relating increased body mass index (BMI) to occipital thickness in >3000 children of similar age (51), and increased occipital and orbitofrontal thickness relates to BMI in adolescents (52). In adults, however, the opposite relationship exists – of BMI to thinner lateral occipital and ventromedial prefrontal cortices (53,54). Higher BMI's effect on slowing adolescent global cortical thinning (52) doesn't explain this localised pattern, however, as BMI was regressed out. Furthermore, although cortical gyrification reduces with extremes of weight loss and gain, and normalises with weight restoration in anorexia (55), this effect spares occipital cortex (51). Interestingly, polygenic risk for obesity relates to both lower lateral occipital area and lateral orbitofrontal thickness, with the latter showing a mediating effect on obesity itself (56). Thus this mode may comprise areas with functional relevance to cognitive models of obesity (52): i.e., processing of visual food cues and their reward value in goal-directed decision-making.

Mode 3 found covariance between being bilingual (with a Hispanic background), and increased (left) temporoparietal but decreased frontal grey matter. Although left parietal volume is associated with bilingualism (57), this mode may seem at odds with findings that bilingual (versus monolingual) children have more grey matter (i.e. less developmental loss) in both parietal (including precuneus) but also frontal regions (58), as do adults (59). Note, however, that bilingual proficiency (in bilingual samples) is related to *thinner* left frontal areas in both children (60) and adults (61). Similarly, interpreters learning a novel language showed increased grey matter in both left frontal and left superior temporal areas, associated with worse and better performance, respectively (62).

Mode 4 related ADHD symptoms to reduced caudate and increased substantia nigra grey matter. Previous meta-analyses in ADHD have noted reduced grey matter in striatum (63), insula, and amygdala (64) – these effects being more pronounced in untreated individuals (65). Other smaller studies have also found reduced grey matter in the thalamus in ADHD (66,67). To our knowledge, increased grey matter in substantia nigra/ventral tegmental area in ADHD is a novel finding, and does not appear to be an effect of stimulant medication. Similarly, motion effects are highly unlikely to drive specific relationships with the caudate and substantia nigra. Intriguingly, these results resemble the cortical area and subcortical volume reductions seen in children with ADHD diagnoses (n=2246, n=1934 controls) in the ENIGMA dataset

(68) but do not resemble the (minimal) changes in those same measures found in ADHD diagnoses (n=949, n=9787 controls) in the ABCD dataset itself (69). Contributory factors could include the quality of subcortical segmentation, reliability of diagnosis (versus symptom questionnaires), and balance of cases/controls.

Mode 5 related anorexia and internalizing symptoms (and lack of activities) to increased grey matter in medial/orbitofrontal and insular/auditory cortex. These brain areas are strongly associated with the influence of emotion on decision-making and regulation of emotion itself (ventromedial and orbitofrontal cortex, also frontal pole (70)) and inhibitory control (cingulate cortex), mood and interoception (insula), and auditory and language processing (Heschl's gyrus, Rolandic operculum). Unlike mode 4, this mode associated psychopathology with *reduced* grey matter in subcortical serotonergic and dopaminergic areas (and thalamus).

The cortical weights resemble the cortical thickness weights of mode 5 in another analysis of the same dataset (9), which likewise related to internalizing problems (but also attentional and externalizing problems, lower income and IQ). One of the strongest contributors to the mode was ventromedial prefrontal cortex: one of two cortical areas in which increased grey matter related to transdiagnostic pathology in a meta-analysis of 132 studies in children and adolescents (71). This region was also one of the most predictive areas for transdiagnostic mental health problems using functional connectivity data from the ABCD cohort (22,72).

This may indicate that thicker cortex, i.e. slower neurodevelopment, in these mainly cortical emotion-processing areas increases the risk of psychopathology. In older subjects, however, the opposite relation exists: the 'p' factor is associated with thinner cortex (especially in frontotemporal areas) in adults (73) and a large (n=15892) meta-analysis of voxel-based morphometry studies in adults showed that grey matter reduction in anterior cingulate and bilateral insula was common to schizophrenia, bipolar disorder, depression, obsessive-compulsive disorder, addiction, and anxiety (74). Similarly, the ENIGMA consortium found thinner frontotemporal cortex and increased thalamic volume (the reverse of mode 5) was common to the first four of these disorders (75). Overall psychopathology relates to globally decreased grey matter in adolescence (8-23 years) – perhaps indicating accelerated neurodevelopment (46) (or brain 'aging' (76)) – although internalizing symptoms were associated with global increases in the same dataset (20,77).

It is interesting to note that activities such as football loaded on the opposite end of this mode. Given this correlational analysis, whether activities are likely to reduce internalizing symptoms (and/or whether children without internalizing symptoms are more motivated to

engage in activities) is unclear, but other studies have found that exercise is an effective treatment for depression (78), including in adolescence (79).

Mode 6 related Oppositional Defiant Disorder (ODD) symptoms to lower grey matter in insula and cingulate cortex, but increased grey matter in caudate, locus coeruleus and amygdala. This mode was interesting as it differentiated ODD from ADHD symptoms in mode 4 (although in the repeated analyses, these modes were combined into a ADHD/ODD mode: see the Supplement). Previous studies have struggled to differentiate brain structural changes in ADHD and ODD because they are highly co-morbid (80,81). Work including very large studies and meta-analyses have found ODD relates to reduced grey matter in both insula (82–84) and cingulate (84,85), however – as here – and a smaller study found increased grey matter in right rostral prefrontal cortex (80), as is also seen in this mode. The novel contribution of this study is in demonstrating subcortical changes relating to ODD (82,85), namely, a larger locus coeruleus and smaller nucleus accumbens. These are interesting given the hypoarousal (86) and reduced representation of expected value found in ODD (87), and that successful treatments boost noradrenergic and dopaminergic function (88).

Comparing the voxel-wise and parcellation-based methods for the SPLS analysis, the most important (and reassuring) point is that most of the modes in the voxel-wise and two (network and anatomical, respectively) parcellation-based methods are very similar, with almost identical variables on the psychosocial side and strong correlations between the brain and psychosocial scores across the different methods (Figure 5C). The major difference is the voxel-wise approach generally yields higher out-of-sample correlations (Figure 5D). It would be interesting to perform a more systematic analysis, to assess correlation strength as a function of resolution, and gauge whether finer parcellations (of up to 1000 areas (37)) have similar performance to voxel-wise methods. Based on this preliminary analysis, however, the only advantage of using parcellated brain data in the SPLS analysis is the lower computational burden: otherwise, voxel-wise data here yielded stronger brain-psychosocial correlations.

Some limitations can hopefully be addressed using novel data and/or methods. It will be important to replicate these latent dimensions in a sample of similar age, and to assess their evolution in the ABCD sample during development. Longitudinal data are crucial to understanding how neurodevelopment relates to brain function: for example, both increased *and* decreased grey matter volumes have been associated with the same psychopathologies in the same studies (of internalizing (20) and autism/ADHD (18)). These could reflect both abnormally delayed and accelerated developmental processes (respectively), impeding the detection of effects from cross-sectional data. Longitudinal data will also potentially permit

more mechanistic insights than these cross-sectional data allow. Although the ABCD sample is very large, and enriched for psychopathology, the proportion with current psychiatric diagnoses is still low (the commonest is ADHD at 9%, with unspecified eating disorder at 8%, and oppositional defiant disorder and suicidal ideation both at 6%), and clinical datasets of similar size (e.g., the Healthy Brain Network (89)) may yield greater insight into (and stronger correlations in) brain-psychopathology relationships.

Despite their relatively weak correlations, however, modes 4, 5 and 6 suggest two further conclusions. One is that when latent variable discovery methods such as factor analysis are applied to ADHD and other psychopathology scales (e.g., depression, psychotic disorders, obsessive compulsive disorder (OCD)), the results differ depending on whether or not brain data are included. Using symptoms only, ADHD symptoms tend to be absorbed into an 'Externalizing' factor along with conduct and oppositional symptoms (26,27,90) or associated with OCD symptoms (91). When incorporating brain structure, ADHD tends to be separated from these other disorders (although not always (92)), as in mode 4 here, and likewise in the ENIGMA meta-analysis of n≈20,000 (75), and, interestingly, when independent component analysis is applied to symptom data instead of factor analysis (12).

Secondly, dopaminergic midbrain areas (bilateral substantia nigra pars compacta and left ventral tegmental area) feature among the highest loadings/voxel not just in the ADHD-related mode, but also in the internalizing mode. Interestingly, psychopathology related to increased grey matter in these dopaminergic areas in mode 4, but decreased grey matter in mode 5. A role for dopamine in depression has long been hypothesised (93), but is controversial (94,95). Nevertheless, recent work has shown that dopamine 2 receptor effects in prefrontal areas (also found in mode 5) are related to antidepressant actions and effortful stress responses (96), and boosting dopamine activity makes depressed subjects' brain responses to a reward-related task (in fMRI) more like controls' (97). Given the range of psychopathology in mode 5, dopamine may have relevance for not just depression, but other maladaptive stress responses. These points illustrate the advantage of including the brain in latent variable analyses of psychopathology.

Lastly, it is striking that prodromal psychosis symptoms did not contribute more strongly to the latent dimensions: we discuss this in the Supplement.

In conclusion, we used sparse partial least squares and a rigorous machine learning framework to uncover six modes of covariation between voxel-wise grey matter density and psychosocial variables from adolescents aged 9-10 years in the ABCD dataset. There were interesting relationships of psychopathology to changes in grey matter density: for overeating

and occipitotemporal and orbitofrontal increases, for ADHD symptoms and decreases in striatal (especially caudate) and cortical attentional areas but increases in substantia nigra and raphe, for internalizing and increases in prefrontal areas and decreases in raphe and substantia nigra, and for ODD symptoms and increases in caudate, locus coeruleus and amygdala. Notably, internalizing related to increased and oppositional problems to reduced grey matter density in insula, cingulate and auditory cortices.

The next goal will be to delineate the mechanistic processes underlying these associations, and whether they can be harnessed to generate or repurpose treatments (21,98), e.g., dopaminergic agents for depression (97,99), or transcranial magnetic stimulation. For example, correlating the modes with brain gene expression maps could establish further links to underlying neurobiology (100) and potential drug targets (98).

Funding

RAA is a Future Leaders Fellow (MR/W011751/1) and was previously an MRC Skills Development Fellow (MR/S007806/1) and has been supported by the NIHR UCLH Biomedical Research Centre and an EPSRC Platform Grant to the Centre for Medical Image Computing (EP/M020533/1). JMM, CZ, AM and KT were supported by the Wellcome Trust under Grant No. WT102845/Z/13/Z. MPP has been supported in part by The William K. Warren Foundation, the National Institute on Drug Abuse (U01 DA041089), and the National Institute of General Medical Sciences Center Grant Award Number (1P20GM121312). The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome (203147/Z/16/Z).

Author contributions

RAA contributed to the design and interpretation and wrote the paper. CZ designed the analysis, wrote analysis code, analysed the data and wrote the paper. AM wrote analysis code and contributed to the analysis. KT performed additional analyses and contributed to the figures and paper. MB and JA provided analysis code and assisted with analysis. JC and MPP assisted with analysis and interpretation. JMM supervised the project. All authors contributed to the paper.

Competing interests

The authors report no biomedical financial interests or potential conflicts of interest.

References

1. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, *et al.* (2016): Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 19: 1523–1536.

2. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, *et al.* (2018): The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci* 32: 43–54.

3. Genon S, Eickhoff SB, Kharabian S (2022): Linking interindividual variability in brain structure to behaviour. *Nat Rev Neurosci* 23: 307–318.

4. Hotelling H (1936): Relations Between Two Sets of Variates. *Biometrika* 28: 321–377.

5. Wold H (1985): Partial Least Squares. *Encyclopedia of Statistical Sciences*. Wiley, New York, pp 581–91.

6. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K, WU-Minn HCP Consortium (2013): The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80: 62–79.

7. Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, *et al.* (2015): A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci* 18: 1565–1567.

8. Alnæs D, Kaufmann T, Marquand AF, Smith SM, Westlye LT (2020): Patterns of sociocognitive stratification and perinatal risk in the child brain. *Proc Natl Acad Sci U S A* 117: 12419–12427.

9. Modabbernia A, Janiri D, Doucet GE, Reichenberg A, Frangou S (2021): Multivariate Patterns of Brain-Behavior-Environment Associations in the Adolescent Brain and Cognitive Development Study. *Biol Psychiatry* 89: 510–520.

10. Mihalik A, Chapman J, Adams RA, Winter NR, Ferreira FS, Shawe-Taylor J, *et al.* (2022): Canonical Correlation Analysis and Partial Least Squares for Identifying Brain-Behavior Associations: A Tutorial and a Comparative Study. *Biol Psychiatry Cogn Neurosci Neuroimaging* 7: 1055–1067.

11. Ing A, Sämann PG, Chu C, Tay N, Biondo F, Robert G, *et al.* (2019): Identification of neurobehavioural symptom groups based on shared brain mechanisms. *Nat Hum Behav* 3: 1306–1318.

12. Modabbernia A, Michelini G, Reichenberg A, Kotov R, Barch D, Frangou S (2022): Neural signatures of data-driven psychopathology dimensions at the transition to adolescence. *Eur Psychiatry* 65: e12.

13. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, *et al.* (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23: 28–38.

14. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, *et al.* (2022): Reproducible brain-wide association studies require thousands of individuals. *Nature* 603: 654–660.

15. Popovic D, Ruef A, Dwyer DB, Antonucci LA, Eder J, Sanfelici R, *et al.* (2020): Traces of Trauma: A Multivariate Pattern Analysis of Childhood Trauma, Brain Structure, and Clinical Phenotypes. *Biol Psychiatry* 88: 829–842.

16. Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, *et al.* (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *Neuroimage Clin* 22: 101796.

17. Mihalik A, Ferreira FS, Moutoussis M, Ziegler G, Adams RA, Rosa MJ, *et al.* (2020): Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships. *Biological Psychiatry* 87: 368–376.

18. Jacobs GR, Voineskos AN, Hawco C, Stefanik L, Forde NJ, Dickie EW, *et al.* (2021): Integration of brain and behavior measures for identification of data-driven groups cutting across children with ASD, ADHD, or OCD. *Neuropsychopharmacology* 46: 643–653.

19. Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, *et al.* (2018): Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* 9: 3003.

20. Kaczkurkin AN, Sotiras A, Baller EB, Barzilay R, Calkins ME, Chand GB, *et al.* (2020): Neurostructural Heterogeneity in Youths With Internalizing Symptoms. *Biol Psychiatry* 87: 473–482.

21. Goldstein-Piekarski AN, Ball TM, Samara Z, Staveland BR, Keller AS, Fleming SL, *et al.* (2022): Mapping Neural Circuit Biotypes to Symptoms and Behavioral Dimensions of Depression and Anxiety. *Biol Psychiatry* 91: 561–571.

22. Chen J, Tam A, Kebets V, Orban C, Ooi LQR, Asplund CL, *et al.* (2022): Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nat Commun* 13: 2217.

23. Linke JO, Abend R, Kircanski K, Clayton M, Stavish C, Benson BE, *et al.* (2021): Shared and Anxiety-Specific Pediatric Psychopathology Dimensions Manifest Distributed Neural Correlates. *Biol Psychiatry* 89: 579–587.

24. Paus T, Keshavan M, Giedd JN (2008): Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci* 9: 947–957.

25. Ziegler G, Hauser TU, Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, *et al.* (2019): Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. *Nat Neurosci* 22: 992–999.

26. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, *et al.* (2017): The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* 126: 454–477.

27. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH (2017): A hierarchical causal taxonomy of psychopathology across the life span. *Psychol Bull* 143: 142–186.

28. McTeague LM, Rosenberg BM, Lopez JW, Carreon DM, Huemer J, Jiang Y, *et al.* (2020): Identification of Common Neural Circuit Disruptions in Emotional Processing Across Psychiatric Disorders. *Am J Psychiatry* 177: 411–421.

29. Bertocci MA, Bebko G, Versace A, Iyengar S, Bonar L, Forbes EE, *et al.* (2017): Reward-related neural activity and structure predict future substance use in dysregulated youth. *Psychol Med* 47: 1357–1369.

30. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, *et al.* (2022): Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry* 27: 281–295.

31. Witten DM, Tibshirani R, Hastie T (2009): A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10: 515–534.

32. Brudfors M, Balbastre Y, Flandin G, Nachev P, Ashburner J (2020): Flexible Bayesian Modelling for Nonlinear Image Registration. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, *et al.*, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Cham: Springer International Publishing, pp 253–263.

33. Garavan H, Bartsch H, Conway K, Decastro A, Goldstein RZ, Heeringa S, *et al.* (2018): Recruiting the ABCD sample: Design considerations and procedures. *Dev Cogn Neurosci* 32: 16–22.

34. Dutt RK, Hannon K, Easley TO, Griffis JC, Zhang W, Bijsterbosch JD (2022): Mental health in the UK Biobank: A roadmap to self-report measures and neuroimaging correlates. *Hum Brain Mapp* 43: 816–832.

35. Ooi LQR, Chen J, Zhang S, Kong R, Tam A, Li J, *et al.* (2022): Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *Neuroimage* 263: 119636.

36. Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M (2020): Automated anatomical labelling atlas 3. *Neuroimage* 206: 116189.

37. Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, *et al.* (2018): Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* 28: 3095–3114.

38. Choi EY, Yeo BTT, Buckner RL (2012): The organization of the human striatum estimated by intrinsic functional connectivity. *J Neurophysiol* 108: 2242–2263.

39. Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BTT (2011): The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* 106: 2322–2345.

40. Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, *et al.* (2011): The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106: 1125–1165.

41. Schnack HG, van Haren NEM, Brouwer RM, Evans A, Durston S, Boomsma DI, *et al.* (2015): Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cereb Cortex* 25: 1608–1617.

42. Shaw P, Greenstein D, Lerch J, Clasen L, Lenroot R, Gogtay N, *et al.* (2006): Intellectual ability and cortical development in children and adolescents. *Nature* 440: 676–679.

43. Tadayon E, Pascual-Leone A, Santarnecchi E (2020): Differential Contribution of Cortical Thickness, Surface Area, and Gyrification to Fluid and Crystallized Intelligence. *Cereb Cortex* 30: 215–225.

44. Vuoksimaa E, Panizzon MS, Chen C-H, Fiecas M, Eyler LT, Fennema-Notestine C, *et al.* (2016): Is bigger always better? The importance of cortical configuration with respect to cognitive ability. *Neuroimage* 129: 356–366.

45. Tomasi D, Volkow ND (2021): Associations of family income with cognition and brain structure in USA children: prevention implications. *Mol Psychiatry* 26: 6619–6629.

46. Tooley UA, Bassett DS, Mackey AP (2021): Environmental influences on the pace of brain development. *Nat Rev Neurosci* 22: 372–384.

47. Noble KG, Houston SM, Brito NH, Bartsch H, Kan E, Kuperman JM, *et al.* (2015): Family income, parental education and brain structure in children and adolescents. *Nat Neurosci* 18: 773–778.

48. Kweon H, Aydogan G, Dagher A, Bzdok D, Ruff CC, Nave G, *et al.* (2022): Human brain anatomy reflects separable genetic and environmental components of socioeconomic status. *Sci Adv* 8: eabm2923.

49. Dotson VM, Duarte A (2020): The importance of diversity in cognitive neuroscience. *Ann N Y Acad Sci* 1464: 181–191.

50. Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, *et al.* (2020): Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *Lancet Public Health* 5: e140–e149.

51. Steegers C, Blok E, Lamballais S, Jaddoe V, Bernardoni F, Vernooij M, *et al.* (2021): The association between body mass index and brain morphology in children: a population-based study. *Brain Struct Funct* 226: 787–800.

52. Westwater ML, Vilar-López R, Ziauddeen H, Verdejo-García A, Fletcher PC (2019): Combined effects of age and BMI are related to altered cortical thickness in adolescence and adulthood. *Dev Cogn Neurosci* 40: 100728.

53. Medic N, Ziauddeen H, Ersche KD, Farooqi IS, Bullmore ET, Nathan PJ, *et al.* (2016): Increased body mass index is associated with specific regional alterations in brain structure. *Int J Obes (Lond)* 40: 1177–1182.

54. Veit R, Kullmann S, Heni M, Machann J, Häring H-U, Fritsche A, Preissl H (2014): Reduced cortical thickness associated with visceral fat and BMI. *Neuroimage Clin* 6: 307–311.

55. Bernardoni F, King JA, Geisler D, Birkenstock J, Tam FI, Weidner K, *et al.* (2018): Nutritional Status Affects Cortical Folding: Lessons Learned From Anorexia Nervosa. *Biol Psychiatry* 84: 692–701.

56. Opel N, Thalamuthu A, Milaneschi Y, Grotegerd D, Flint C, Leenings R, *et al.* (2021): Brain structural abnormalities in obesity: relation to age, genetic risk, and common

psychiatric disorders : Evidence through univariate and multivariate mega-analysis including 6420 participants from the ENIGMA MDD working group. *Mol Psychiatry* 26: 4839–4852.

57. Mechelli A, Crinion JT, Noppeney U, O'Doherty J, Ashburner J, Frackowiak RS, Price CJ (2004): Neurolinguistics: structural plasticity in the bilingual brain. *Nature* 431: 757.

58. Pliatsikas C, Meteyard L, Veríssimo J, DeLuca V, Shattuck K, Ullman MT (2020): The effect of bilingualism on brain development from early childhood to young adulthood. *Brain Struct Funct* 225: 2131–2152.

59. Olulade OA, Jamal NI, Koo DS, Perfetti CA, LaSasso C, Eden GF (2016): Neuroanatomical Evidence in Support of the Bilingual Advantage Theory. *Cereb Cortex* 26: 3196–3204.

60. Archila-Suerte P, Woods EA, Chiarello C, Hernandez AE (2018): Neuroanatomical profiles of bilingual children. *Dev Sci* 21: e12654.

61. Claussenius-Kalman HL, Vaughn KA, Archila-Suerte P, Hernandez AE (2021): Highly proficient, balanced bilingualism is related to thinner cortex in two cognitive control regions. *Ann N Y Acad Sci* 1485: 83–94.

62. Mårtensson J, Eriksson J, Bodammer NC, Lindgren M, Johansson M, Nyberg L, Lövdén M (2012): Growth of language-related brain areas after foreign language learning. *Neuroimage* 63: 240–244.

63. Norman LJ, Carlisi C, Lukito S, Hart H, Mataix-Cols D, Radua J, Rubia K (2016): Structural and Functional Brain Abnormalities in Attention-Deficit/Hyperactivity Disorder and Obsessive-Compulsive Disorder: A Comparative Meta-analysis. *JAMA Psychiatry* 73: 815–825.

64. Nakao T, Radua J, Rubia K, Mataix-Cols D (2011): Gray matter volume abnormalities in ADHD: voxel-based meta-analysis exploring the effects of age and stimulant medication. *Am J Psychiatry* 168: 1154–1163.

65. Frodl T, Skokauskas N (2012): Meta-analysis of structural MRI studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr Scand* 125: 114–126.

66. Batty MJ, Palaniyappan L, Scerif G, Groom MJ, Liddle EB, Liddle PF, Hollis C (2015): Morphological abnormalities in prefrontal surface area and thalamic volume in attention deficit/hyperactivity disorder. *Psychiatry Res* 233: 225–232.

67. Ivanov I, Bansal R, Hao X, Zhu H, Kellendonk C, Miller L, *et al.* (2010): Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder. *Am J Psychiatry* 167: 397–408.

68. Hoogman M, Muetzel R, Guimaraes JP, Shumskaya E, Mennes M, Zwiers MP, *et al.* (2019): Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. *Am J Psychiatry* 176: 531–542.

69. Bernanke J, Luna A, Chang L, Bruno E, Dworkin J, Posner J (2022): Structural brain measures among children with and without ADHD in the Adolescent Brain and Cognitive Development Study cohort: a cross-sectional US population-based study. *Lancet Psychiatry* 9: 222–231.

70. Bramson B, Folloni D, Verhagen L, Hartogsveld B, Mars RB, Toni I, Roelofs K (2020): Human Lateral Frontal Pole Contributes to Control over Emotional Approach-Avoidance Actions. *J Neurosci* 40: 2925–2934.

71. Li T, Wang L, Camilleri JA, Chen X, Li S, Stewart JL, *et al.* (2020): Mapping common grey matter volume deviation across child and adolescent psychiatric disorders. *Neurosci Biobehav Rev* 115: 273–284.

72. Wainberg M, Jacobs GR, Voineskos AN, Tripathy SJ (2022): Neurobiological, familial and genetic risk factors for dimensional psychopathology in the Adolescent Brain Cognitive Development study. *Mol Psychiatry* 27: 2731–2741.

73. Romer AL, Elliott ML, Knodt AR, Sison ML, Ireland D, Houts R, *et al.* (2021): Pervasively Thinner Neocortex as a Transdiagnostic Feature of General Psychopathology. *Am J Psychiatry* 178: 174–182.

74. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, *et al.* (2015): Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* 72: 305–315.

75. Opel N, Goltermann J, Hermesdorf M, Berger K, Baune BT, Dannlowski U (2020): 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* 88: 678–686.

76. Han LKM, Dinga R, Hahn T, Ching CRK, Eyler LT, Aftanas L, *et al.* (2021): Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol Psychiatry* 26: 5124–5139.

77. Kaczkurkin AN, Park SS, Sotiras A, Moore TM, Calkins ME, Cieslak M, *et al.* (2019): Evidence for Dissociable Linkage of Dimensions of Psychopathology to Brain Structure in Youths. *Am J Psychiatry* 176: 1000–1009.

78. Kvam S, Kleppe CL, Nordhus IH, Hovland A (2016): Exercise as a treatment for depression: A meta-analysis. *J Affect Disord* 202: 67–86.

79. Wang X, Cai Z-D, Jiang W-T, Fang Y-Y, Sun W-X, Wang X (2022): Systematic review and meta-analysis of the effects of exercise on depression in adolescents. *Child Adolesc Psychiatry Ment Health* 16: 16.

80. Vetter NC, Backhausen LL, Buse J, Roessner V, Smolka MN (2020): Altered brain morphology in boys with attention deficit hyperactivity disorder with and without comorbid conduct disorder/oppositional defiant disorder. *Hum Brain Mapp* 41: 973–983.

81. Noordermeer SDS, Luman M, Greven CU, Veroude K, Faraone SV, Hartman CA, *et al.* (2017): Structural Brain Abnormalities of Attention-Deficit/Hyperactivity Disorder With Oppositional Defiant Disorder. *Biol Psychiatry* 82: 642–650.

82. Rogers JC, De Brito SA (2016): Cortical and Subcortical Gray Matter Volume in Youths With Conduct Problems: A Meta-analysis. *JAMA Psychiatry* 73: 64–72.

83. Noordermeer SDS, Luman M, Oosterlaan J (2016): A Systematic Review and Meta-analysis of Neuroimaging in Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) Taking Attention-Deficit Hyperactivity Disorder (ADHD) Into Account. *Neuropsychol Rev* 26: 44–72.

84. Fahim C, He Y, Yoon U, Chen J, Evans A, Pérusse D (2011): Neuroanatomy of childhood disruptive behavior disorders. *Aggress Behav* 37: 326–337.

85. Bayard F, Nymberg Thunell C, Abé C, Almeida R, Banaschewski T, Barker G, *et al.* (2020): Distinct brain structure and behavior related to ADHD and conduct disorder traits. *Mol Psychiatry* 25: 3020–3033.

86. Hawes DJ, Gardner F, Dadds MR, Frick PJ, Kimonis ER, Burke JD, Fairchild G (2023): Oppositional defiant disorder. *Nat Rev Dis Primers* 9: 31.

87. White SF, Pope K, Sinclair S, Fowler KA, Brislin SJ, Williams WC, *et al.* (2013): Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. *Am J Psychiatry* 170: 315–323.

88. Pringsheim T, Hirsch L, Gardner D, Gorman DA (2015): The pharmacological management of oppositional behaviour, conduct problems, and aggression in children and adolescents with attention-deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder: a systematic review and meta-analysis. Part 1: psychostimulants, alpha-2 agonists, and atomoxetine. *Can J Psychiatry* 60: 42–51.

89. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, *et al.* (2017): An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data* 4: 170181.

90. Forbush KT, Watson D (2013): The structure of common and uncommon mental disorders. *Psychol Med* 43: 97–108.

91. Michelini G, Barch DM, Tian Y, Watson D, Klein DN, Kotov R (2019): Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Transl Psychiatry* 9: 261.

92. Cropley VL, Tian Y, Fernando K, Mansour L S, Pantelis C, Cocchi L, Zalesky A (2021): Brain-Predicted Age Associates With Psychopathology Dimensions in Youths. *Biol Psychiatry Cogn Neurosci Neuroimaging* 6: 410–419.

93. Kapur S, Mann JJ (1992): Role of the dopaminergic system in depression. *Biol Psychiatry* 32: 1–17.

94. Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, *et al.* (2017): Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry* 74: 790–797.

95. Schneier FR, Slifstein M, Whitton AE, Pizzagalli DA, Reinen J, McGrath PJ, *et al.* (2018): Dopamine Release in Antidepressant-Naive Major Depressive Disorder: A Multimodal [11C]-(+)-PHNO Positron Emission Tomography and Functional Magnetic Resonance Imaging Study. *Biol Psychiatry* 84: 563–573.

96. Wilke SA, Lavi K, Byeon S, Donohue KC, Sohal VS (2022): Convergence of Clinically Relevant Manipulations on Dopamine-Regulated Prefrontal Activity Underlying Stress Coping Responses. *Biol Psychiatry* 91: 810–820.

97. Liu Y, Admon R, Mellem MS, Belleau EL, Kaiser RH, Clegg R, *et al.* (2020): Machine Learning Identifies Large-Scale Reward-Related Activity Modulated by Dopaminergic Enhancement in Major Depression. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5: 163–172.

98. Ji JL, Helmer M, Fonteneau C, Burt JB, Tamayo Z, Demšar J, *et al.* (2021): Mapping brain-behavior space relationships along the psychosis spectrum. *Elife* 10: e66968.

99. Halahakoon DC, Kaltenboeck A, Martens M, Geddes JG, Harmer CJ, Cowen P, Browning M (2023): Pramipexole Enhances Reward Learning by Preserving Value Estimates. *Biol Psychiatry* S0006-3223(23)01356–2.

100. Whitaker KJ, Vértes PE, Romero-Garcia R, Váša F, Moutoussis M, Prabhu G, *et al.* (2016): Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc Natl Acad Sci U S A* 113: 9105–9110.

Figure Legends

Figure 1: Preprocessing and analysis pipeline

The left side of the figure summarises the preprocessing steps for the brain and psychosocial variables. See the Supplement for a full description of the sMRI and psychosocial preprocessing steps. Many psychosocial variables were excluded because the ratio of their most common answer to the remainder was >99:1, or because >10% of values were missing. Different SPLS analyses were conducted using either voxel-wise or parcellation-based grey matter data; the latter comprised 166 parcels in the AAL3 atlas, or 234 parcels in the network-based atlas. Within the SPLS and machine learning framework, the green box represents the 80% participants randomly selected to be in the training set, and the red box represents the 20% participants in the hold-out or 'test' set. Within the training set, 5 fold cross-validation is performed in order to optimise the SPLS hyper-parameters c_u and c_v (according to generalizability and stability criteria: see the Supplement): the folds are shown demarcated with grey lines and the validation sets in blue. Once the parameters have been optimised, the model is trained on the full training set and evaluated on the hold-out set, by permuting the labels 500 times (Abbreviations: BMI, body mass index; TBV, total brain volume).

Figure 2: Modes 1 and 2 – Psychosocial and brain loadings

A – The title of the figure gives this mode's main theme, its brain-psychosocial Pearson correlation and the permutation-based *p* value and the number of brain and psychosocial variables selected (i.e., with non-zero weights). The weights themselves are shown in the Supplement. Here the loadings – proportional to the correlation of each variable with the brain/psychosocial scores on that mode (also see Equation 3) – are displayed. The upper part of the figure shows the psychosocial loadings, coloured according to the questionnaires in the legend. Only the loadings of $|r|>0.15$ are shown, up to a maximum of 40 variables. On the bottom row, loadings on the brain are plotted, normalised to within -1 and 1: note that the cerebellum is not shown and subcortical loadings are projected on to the medial surface (see the Supplement for the loadings on cerebellum and specific subcortical regions). On the bottom right is a scatter plot showing the brain and psychosocial scores for each participant in the training (blue) and hold-out (brown) sets. Abbreviations: Matrix reasoning and Rey Auditory Verbal Learning Test (PearsonScores), Parent Demographics Survey (PDEM), Sports and Activities Involvement Questionnaire (SAIQ), Family History Assessment (FHX1), NIH Toolbox Tasks (TBX), Neighbourhood Safety/Crime Survey (NSC2), Parent Diagnostic Interview for DSM-5: Background Items (KSAD3).

B – This figure shows the results for mode 2 in the same format as Figure 2A. Note that the psychosocial weights are so sparse (2 variables) that the scores form a trimodal distribution (bottom right). Abbreviations: Child Behaviour Checklist (CBCL), Sleep Disturbance Scale (SDS), Parent Diagnostic Interview for DSM-5: Mental health diagnosis (KSAD4).

Figure 3: Modes 3 and 4 – Psychosocial and brain loadings

A – This figure shows the results for mode 3 in the same format as Figure 2A. Abbreviations: Parent Demographics Survey (PDEM), Acculturation Survey: Child (ACC1), Acculturation Survey: Parent (ACC2).

B – This figure shows the results for mode 4 in the same format as Figure 2A. Some considerable outliers in brain score are also visible on the left of the bottom right plot: these (training) subjects do not drive the test correlation, however. Abbreviations: Child Behaviour Checklist (CBCL), Parent General Behaviour Inventory – Mania (PGBI), Parent Diagnostic Interview for DSM-5: Background Items (KSAD3), Parent Diagnostic Interview for DSM-5: Mental health diagnosis (KSAD4).

Figure 4: Modes 5 and 6 – Psychosocial and brain loadings

A – This figure shows the results for mode 5 in the same format as Figure 2A. Abbreviations: Behavioural Inhibition/Behavioural Approach System Scale (BIS), Child Behaviour Checklist (CBCL), Parent Demographics Survey (PDEM), Sports and Activities Involvement Questionnaire (SAIQ), Family History Assessment (FHX1), Parent Diagnostic Interview for DSM-5: Background Items (KSAD3), Parent Diagnostic Interview for DSM-5: Mental health diagnosis (KSAD4).

B – This figure shows the results for mode 6 in the same format as Figure 2A. Abbreviations: Child Behaviour Checklist (CBCL), Parent General Behaviour Inventory – Mania (PGBI), Prosocial Behaviour Survey (PST2), Family Environment Scale – Family Conflict (FES2), Parent Diagnostic Interview for DSM-5: Background Items (KSAD3), Parent Diagnostic Interview for DSM-5: Mental health diagnosis (KSAD4), Parent Diagnostic Interview for DSM-5: Conduct Disorder (KSAD5).

Figure 5: Comparisons of analysis methods

A – This bar plot shows the percentage of variances explained in the hold-out set by each mode, both within each modality (brain or psychosocial) and across both modalities (brain-psychosocial covariance), for each of the six voxel-wise SPLS modes.

B – This confusion matrix plots the Pearson correlations between the averaged brain and psychosocial scores across individuals for each mode in the (voxel-wise) analysis using all psychosocial variables and the analysis using only mental health-related variables.

C – These confusion matrices plot the Pearson correlations between the averaged brain and psychosocial scores across individuals for each mode in the voxel-wise analysis and the anatomical (AAL3) parcellation analysis (left) or the network-based parcellation analysis (right). It is clear that the first three modes of each are very similar to the voxel-wise analysis, as is mode 4 in the anatomical parcellation analysis. See the Supplement for further details, and the similarity of the separate brain and psychosocial scores.

D – This bar plot shows the Pearson correlation strengths for all SPLS modes found in the voxel-wise analysis, and the corresponding (i.e. containing similar variables) modes in the anatomical (AAL3) parcellation and network parcellation analyses. Only three modes were found in the network parcellation-based analysis; five modes were found in the anatomical parcellation-based analysis, but the last one had no equivalent in the voxel-wise analysis so it is not plotted here. The correlations are stronger in the voxel-wise analysis for the modes that were common to all analyses.

ABCD Release 2.0.1 (n=11880)
• Behavioural, clinical, cognitive and demographic data (n=11850)
• Minimally processed sMRI data (n=11409)

sMRI data preprocessing
• Quality control (n=20 discarded)
• Light cleaning and segmentation
• Smoothing
• Masking for voxel-wise grey matter scores ± averaging over parcellation area

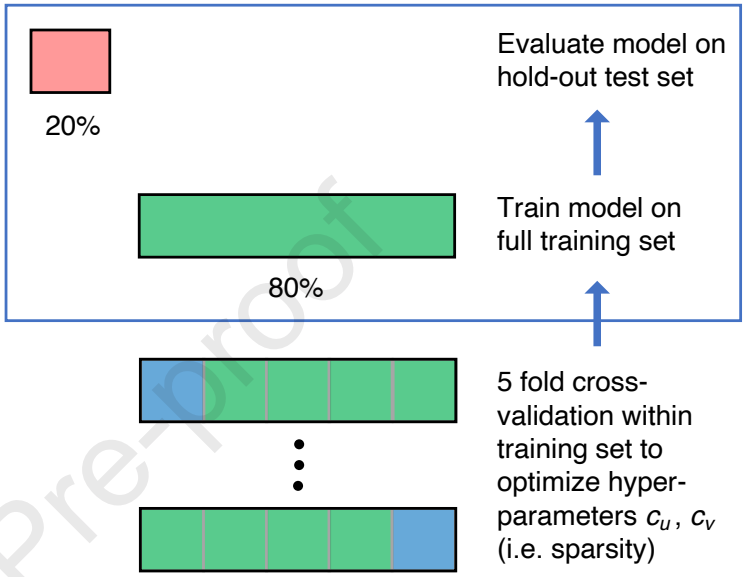
Psychosocial data preprocessing
• Variable selection (3036 variables)
• Recoding (2701 variables)
• One-hot encoding (2857 variables)
• Cognitive variables recoded (109 variables excluded)
• >99:1 ratio? (1745 variables excluded)
• >10% missing values (variables)? (553 variables excluded)
• >10% missing values (participants)? (n=77 discarded)

Common data preprocessing
• n=11288 participants with both brain (124398 voxels or 166/234 parcels) & psychosocial data (557 variables, excluding age and sex)
• Imputation
• Deconfounding (sex, age, TBV, BMI, scanner type)
• Standardization

SPLS and machine learning framework

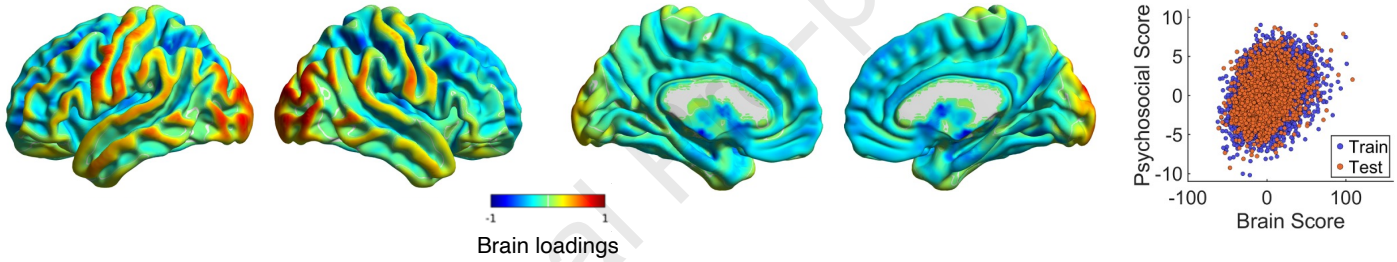
For each mode (latent dimension):

500 permutations for statistical testing



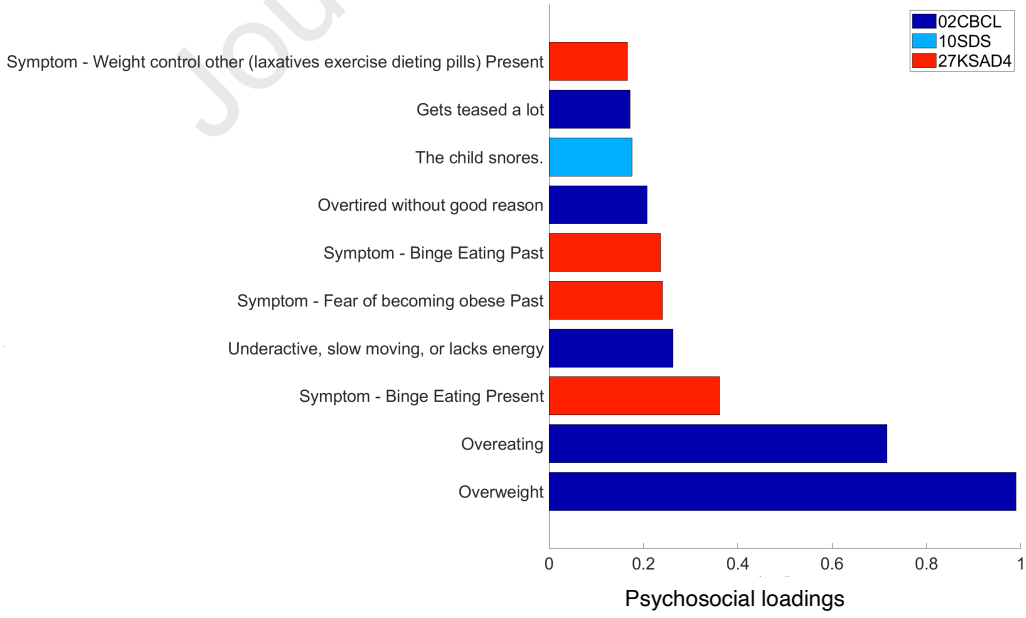
A

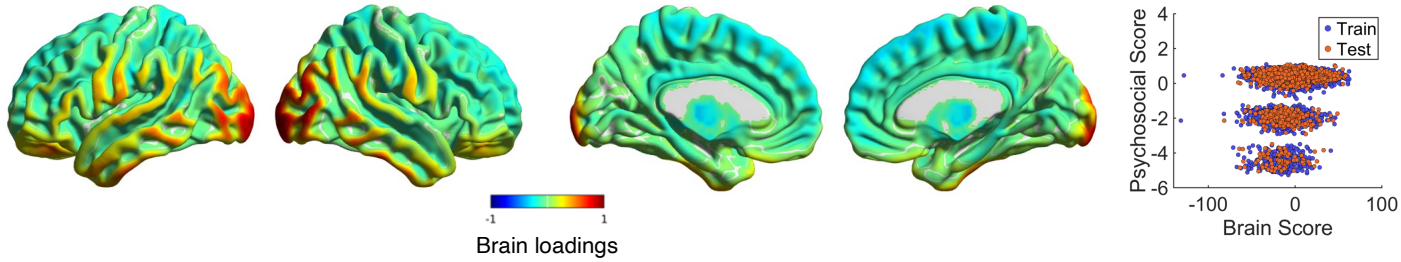
Mode 1: IQ & income, r=0.30, p<0.002, 5429 brain features, 48 psychosocial features



B

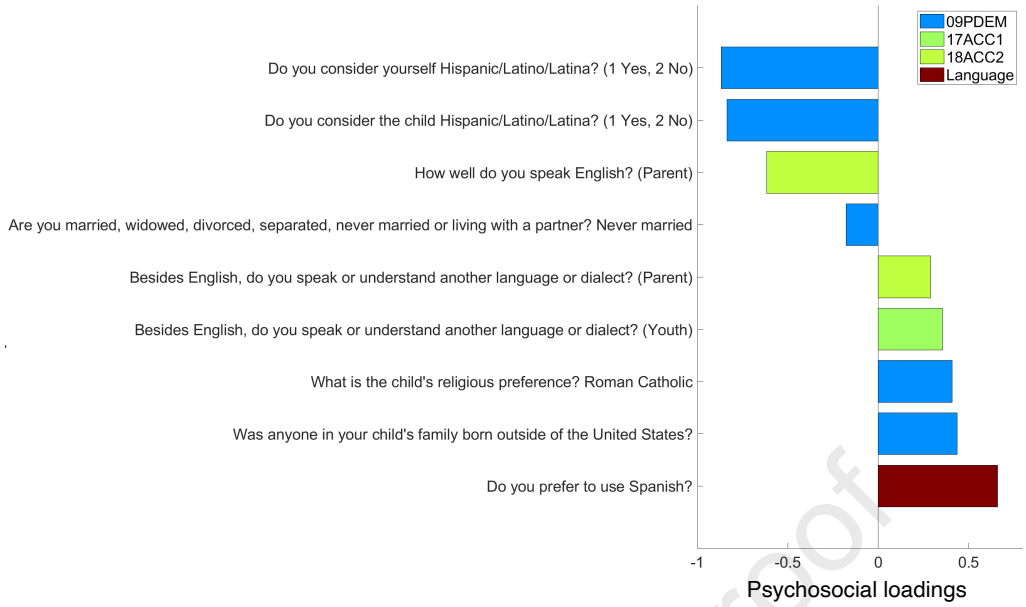
Mode 2: Overeating, r=0.24, p<0.002, 3055 brain features, 2 psychosocial features

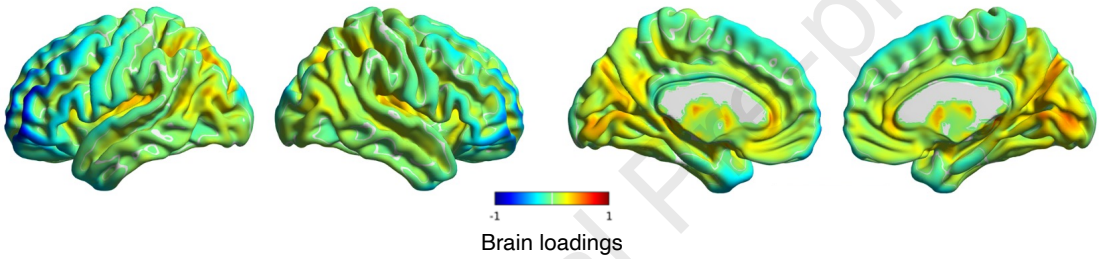


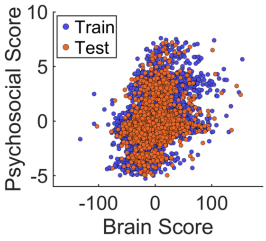


A

Mode 3: Bilingual (with a Hispanic background), r=0.24, p<0.002, 18947 brain features, 8 psychosocial features

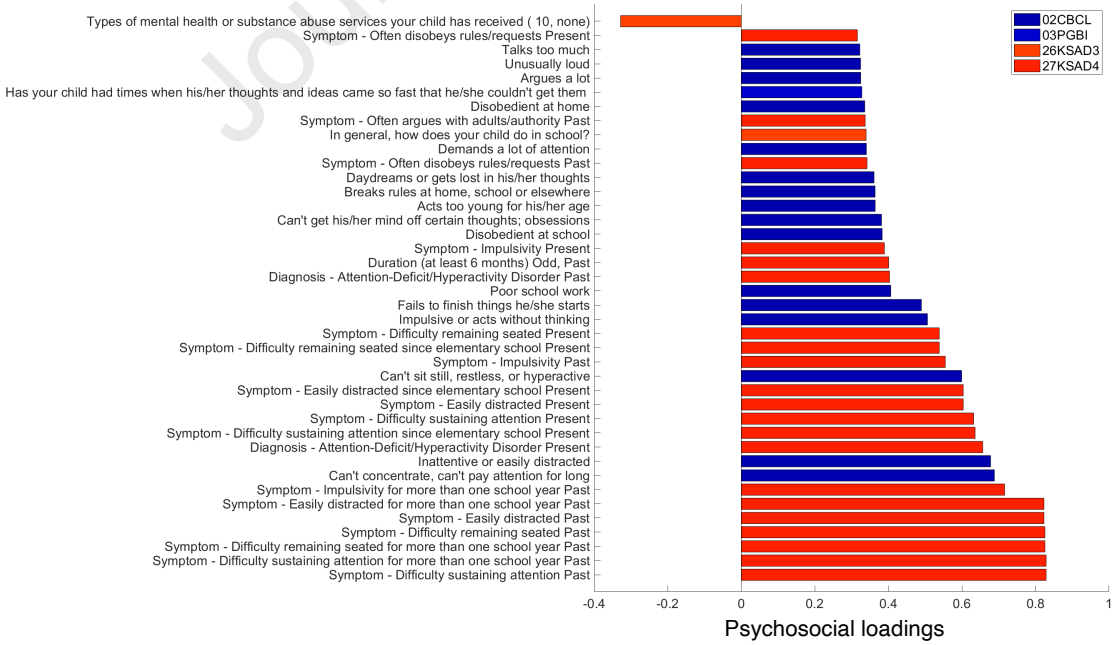


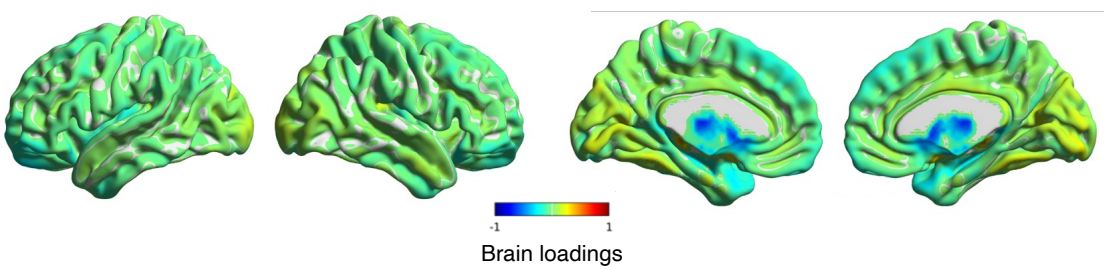


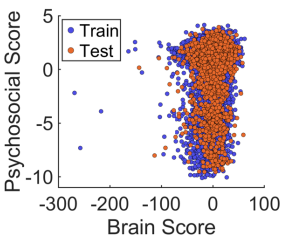


B

Mode 4: ADHD symptoms, r=0.042, p=0.030, 10335 brain features, 21 psychosocial features

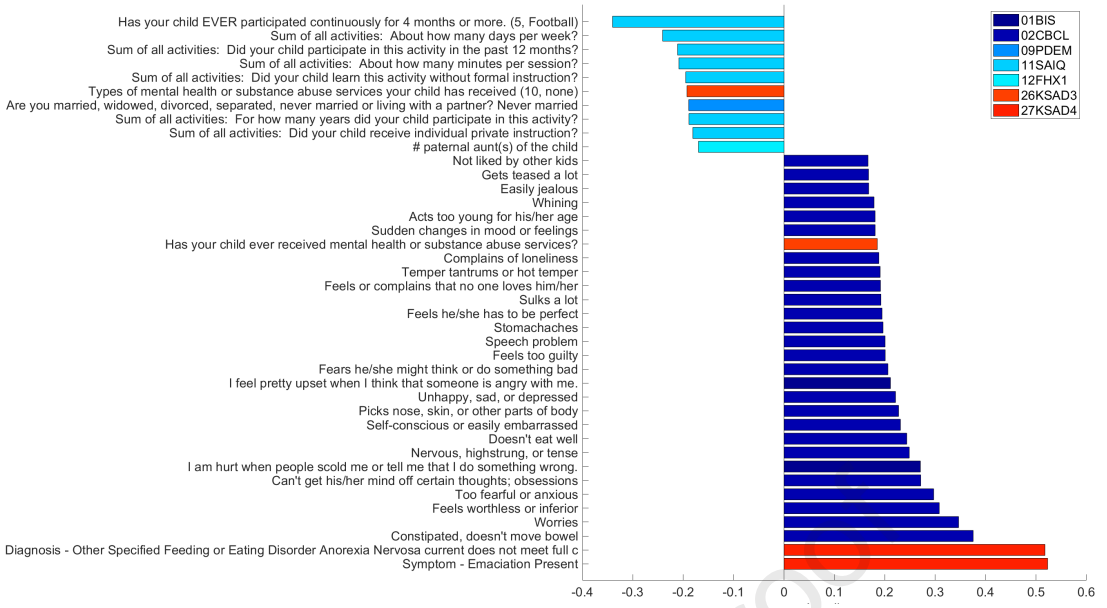


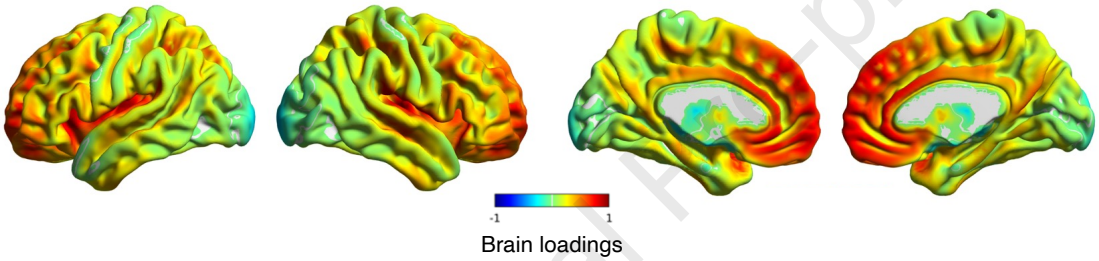


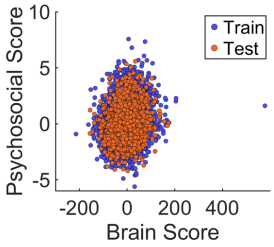


A

Mode 5: Anorexia & internalizing, r=0.14, p<0.002, 16973 brain features, 72 psychosocial features

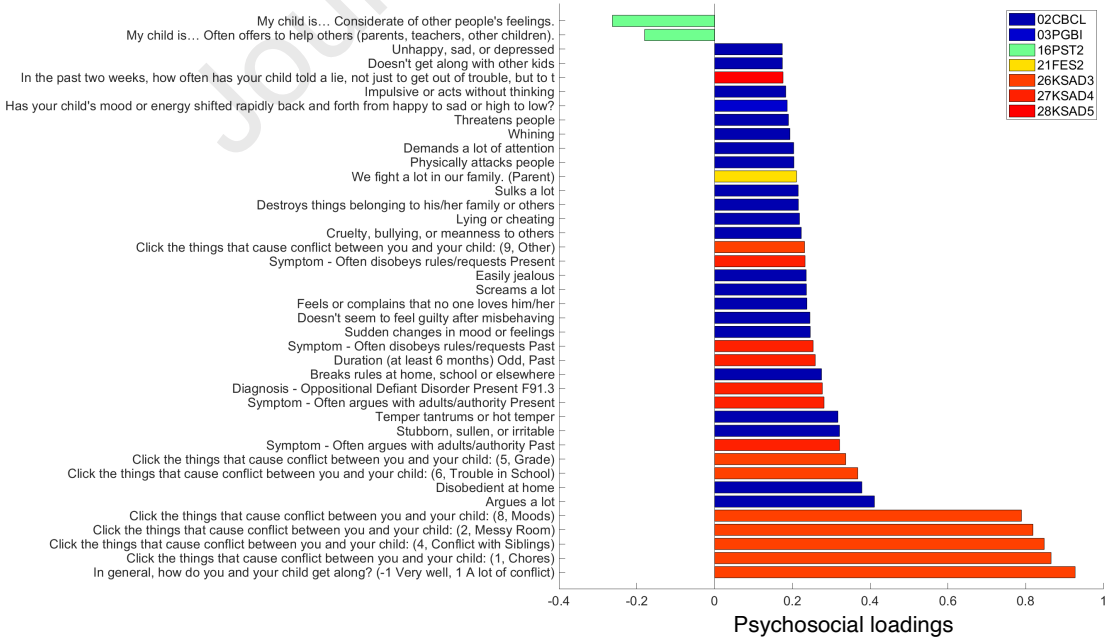


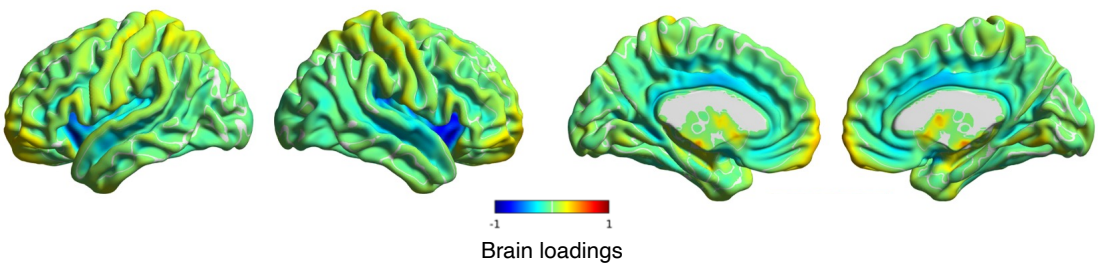


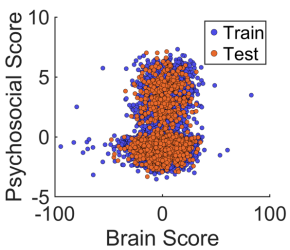


B

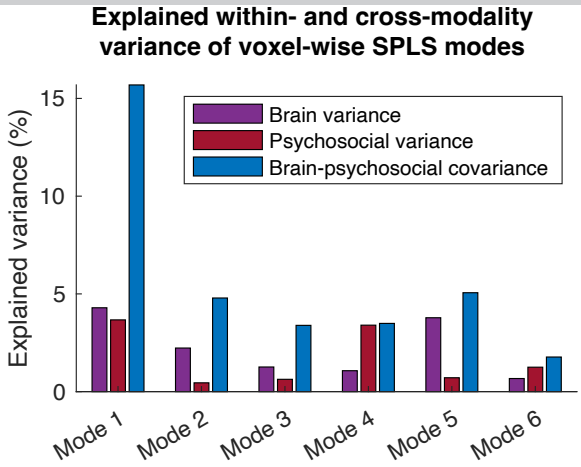
Mode 6: Oppositional Defiant symptoms, r=0.10, p<0.002, 10250 brain features, 7 psychosocial features



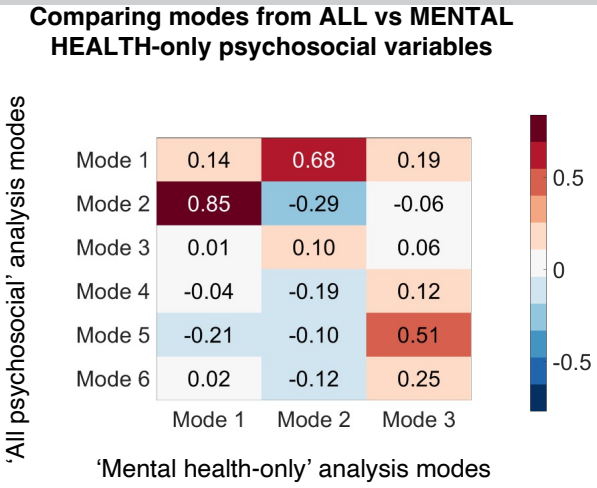




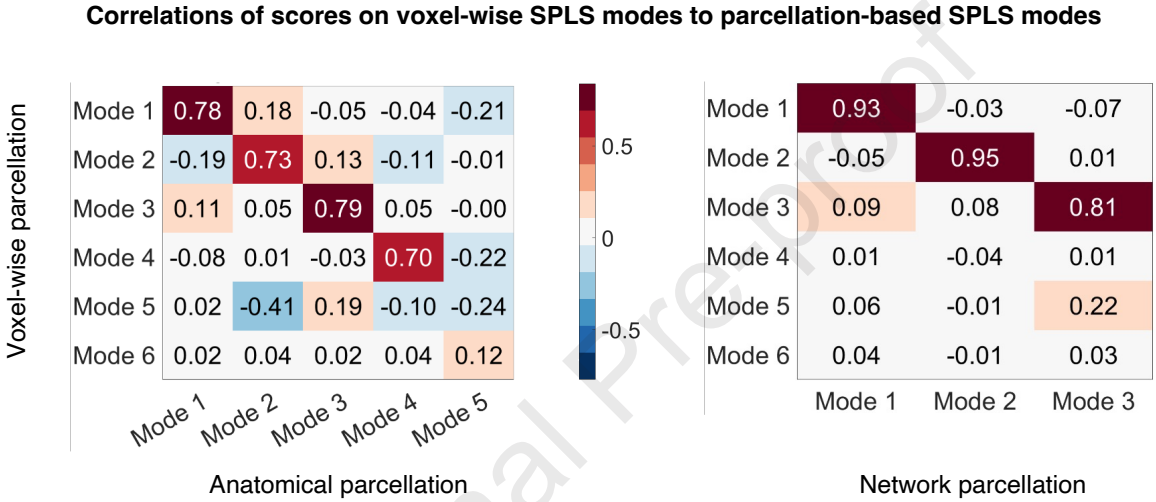
A



B



C



D