



Published in final edited form as:

Biol Psychiatry. 2020 March 01; 87(5): 473–482. doi:10.1016/j.biopsych.2019.09.005.

Neurostructural Heterogeneity in Youth with Internalizing Symptoms

Antonia N. Kaczkurkin, Ph.D.^a, Aristeidis Sotiras, Ph.D.^{b,c,d}, Erica B. Baller, M.D., M.S.^a, Ran Barzilay, M.D., Ph.D.^a, Monica E. Calkins, Ph.D.^a, Ganesh B. Chand, Ph.D.^{b,d}, Zaixu Cui, Ph.D.^a, Guray Erus, Ph.D.^{b,d}, Yong Fan, Ph.D.^{b,d}, Raquel E. Gur, M.D., Ph.D.^{a,b}, Ruben C. Gur, Ph.D.^{a,b,e}, Tyler M. Moore, Ph.D.^a, David R. Roalf, Ph.D.^a, Adon F.G. Rosen, B.S.^a, Kosha Ruparel, MSE^a, Russell T. Shinohara, Ph.D.^{d,f}, Erdem Varol, Ph.D.^{b,d,g}, Daniel H. Wolf, M.D., Ph.D.^{a,d}, Christos Davatzikos, Ph.D.^{b,d}, Theodore D. Satterthwaite, M.D.^{a,d}

^aDepartment of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

^bDepartment of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

^cDepartment of Radiology, Washington University, St. Louis, MO, 63110, USA

^dCenter for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA

^ePhiladelphia Veterans Administration Medical Center, Philadelphia, PA 19104

^fDepartment of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA

^gDepartment of Statistics, Center for Theoretical Neuroscience, Grossman Center for the Statistics of Mind, Columbia University, New York, NY 10027

Abstract

Background: Internalizing disorders such as anxiety and depression are common psychiatric disorders, frequently begin in youth, and exhibit marked heterogeneity in treatment response and clinical course. Given that symptom-based classification approaches do not align with underlying neurobiology, an alternative approach is to identify neurobiologically-informed subtypes based on brain imaging data.

Corresponding author: Antonia N. Kaczkurkin, Ph.D., Richards Building, 5th Floor, Suite 5A, 3700 Hamilton Walk, Philadelphia, PA 19104-6085, antoniak@pennmedicine.upenn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

DISCLOSURES

Dr. Shinohara has received legal consulting and advisory board income from Genentech/Roche. Dr. Barzilay serves on the scientific board and reports stock ownership in 'Taliaz Health', with no conflict of interest relevant to this work. All other authors report no biomedical financial interests or potential conflicts of interest.

Methods: We used a recently developed semi-supervised machine learning method (HYDRA) to delineate patterns of neurobiological heterogeneity within youth with internalizing symptoms using structural data collected at 3T from a sample of 1,141 youth.

Results: Using volume and cortical thickness, cross-validation methods indicated two highly stable subtypes of internalizing youth (ARI=.66; permutation-based $p_{fidr} < .001$). Subtype 1, defined by smaller brain volumes and reduced cortical thickness, was marked by impaired cognitive performance and higher levels of psychopathology than both Subtype 2 and typically developing youth. Using resting-state fMRI and diffusion images not considered during clustering, we found that Subtype 1 also showed reduced amplitudes of low-frequency fluctuations in fronto-limbic regions at rest and reduced fractional anisotropy in several white matter tracts. In contrast, Subtype 2 showed intact cognitive performance, greater volume, cortical thickness, and amplitudes during rest compared to Subtype 1 and typically developing youth, despite still showing clinically significant levels of psychopathology.

Conclusions: We identified two subtypes of internalizing youth differentiated by abnormalities in brain structure, function, and white matter integrity, with one subtype showing poorer functioning across multiple domains. Identification of biologically-grounded internalizing subtypes may assist in targeting early interventions and assessing longitudinal prognosis.

Keywords

internalizing; heterogeneity; youth; structure; volume; cortical thickness

INTRODUCTION

Internalizing disorders, including depression and anxiety disorders, are the most common psychiatric conditions (1), and together result in an enormous worldwide burden of illness (2). In contrast to cardiovascular or neurodegenerative disorders, internalizing disorders often begin in youth, leading to a lifetime of morbidity (1, 3). At present, diagnosis of internalizing disorders remains driven by presenting symptoms, as codified in the *Diagnostic and Statistical Manual of Mental Health Disorders (DSM-5)*. However, such diagnoses often lack specificity, as evinced by the high degree of comorbidity with other psychiatric disorders (4, 5) and marked heterogeneity in both treatment response and longitudinal outcome (6).

Emerging evidence from epidemiology, genetics, and clinical neuroimaging often does not support the current diagnoses codified in the *DSM(7-9)*. One alternative is to evaluate dimensions of symptoms that cross diagnostic boundaries (10). However, dimensional models based on symptoms do not account for the neurobiological mechanisms underlying psychiatric symptoms. A classification approach that parses heterogeneous clinical syndromes based on neurobiological data would be a significant advancement for the field (11). Intensifying efforts are being made to identify neurobiologically-informed subtypes using machine learning techniques. Using such an approach, patients are clustered into disease sub-groups according to shared patterns in imaging or other data types in order to reveal the heterogeneous biological mechanisms that underlie comorbid disorders. Recent work has delineated neurobiological subtypes in Alzheimer's disease (12-14), depression

(15, 16), and psychosis (17-19) in adults as well as in attention-deficit/hyperactivity disorder in youth (20, 21).

To our knowledge, as of yet there have been no efforts to parse neurobiological heterogeneity in youth with internalizing symptoms. Accordingly, the aim of the current study was to delineate patterns of neurostructural heterogeneity in internalizing symptoms in relation to typically developing controls among 1,141 youth using data-driven machine learning techniques. Both volume and cortical thickness were included as genetic studies suggest that these two measures can provide complementary but distinct information (22). These subtypes were then evaluated using independent clinical, cognitive, and neuroimaging data that was not used in the clustering process.

METHODS AND MATERIALS

Participants

A total of 1,601 participants ages 8-23 years received multi-modal neuroimaging, clinical phenotyping, and cognitive assessment as part of the Philadelphia Neurodevelopmental Cohort (PNC), a large community-based sample of youth (23, 24). After standard exclusion criteria (including medical disorders and structural image quality; see Supplement), 715 participants met screening criteria for an anxiety and/or depressive disorder and 426 were typically developing youth with no psychiatric diagnoses ($n=1,141$ total). As expected, the internalizing group showed a greater percentage of females than typically developing youth. Furthermore, modality-specific quality assurance was performed and resulted in a sample of $n=840$ with resting-state functional MRI (rsfMRI) and $n=923$ with diffusion imaging data (Supplement). Finally, a subsample with data on gestational age at birth ($n=282$) was used to examine birth history (25). The institutional review boards of the University of Pennsylvania and the Children's Hospital of Philadelphia approved the study procedures. All participants provided written informed consent after receiving a complete description of the study.

Clinical assessment

As described in detail in our previous work (23, 24, 26) and in the Supplement, assessment of lifetime psychopathology was conducted using GOASSESS, a structured screening interview based on a modified version of the K-SADS (27). We included participants in the internalizing group if they met criteria for any anxiety and/or depressive disorder including agoraphobia, generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, separation anxiety disorder, social anxiety disorder, specific phobia, or major depressive disorder (Table 1).

Clinical and cognitive factor analyses

As prior (28, 29), to provide a dimensional summary of the diverse psychopathology data, we used a confirmatory bifactor analysis (30, 31) to model four orthogonal factors (anxious-misery, psychosis, behavioral, and fear) plus a general factor, overall psychopathology, which represents the symptoms common across all psychiatric disorders (Supplement). Cognition was assessed using the University of Pennsylvania Computerized Neurocognitive Battery (CNB), which has been described in detail elsewhere (32). Fourteen cognitive tests

evaluating aspects of cognition were summarized with exploratory factor analysis into three domains: 1) executive function and complex reasoning, 2) social cognition, and 3) episodic memory (Supplement). Reading skills were measured with the *Wide Range Achievement Test, 4th Edition* (WRAT-4) reading subscale (33).

Image acquisition, quality assurance, and image processing

Image acquisition, processing, and quality assurance procedures for volume, cortical thickness, rsfMRI, and DTI measures have been previously described (24, 34, 35) and are detailed in the Supplement. Structural images were processed using the top-performing tools included in ANTs (36). Functional connectivity among brain regions is primarily attributable to correlations between low-frequency fluctuations in regional activation patterns (37). Therefore, we computed the voxel-wise amplitude of low-frequency fluctuations (ALFF) as the sum over frequency bins in the low-frequency (0.01-0.08 Hertz) band of the power spectrum (37). While there are many different resting state functional connectivity measures, here we use ALFF based on prior work showing abnormal resting state fluctuations in those with psychopathology (37-43). ALFF also allows us to compare our structural measures to a resting state measure using the same atlas, allowing for correspondence of brain regions across modalities. Cortical thickness, volume, and ALFF were summarized in anatomic regions within gray matter defined on an individual basis using a top-performing multi-atlas labeling procedure with joint label fusion (44). Fractional anisotropy maps were calculated from DTI using FSL (45) and summarized in tracts defined by the JHU white-matter tractography atlas (46).

Parsing heterogeneity with semi-supervised machine learning

To identify neurostructural subtypes within youth with internalizing symptoms, we used a recently developed, semi-supervised machine learning tool, HYDRA (Heterogeneity through Discriminative Analysis) (12). In contrast to fully-supervised learning techniques (i.e., support vector machines (SVM) or random forests) which cannot distinguish between subtypes of cases (Figure 1A), HYDRA clusters cases based on their differences from controls by finding multiple linear hyperplanes, which together form a convex polytope (Figure 1B). In contrast to unsupervised clustering techniques (i.e., k-means or community detection), HYDRA does not cluster patients based on their similarity, which is a process that is vulnerable to confounding inter-individual variations that are irrelevant to disease (e.g., due to age or sex).

HYDRA defined neurostructural subtypes using the volume of 112 cortical and subcortical regions as well as the cortical thickness of 98 regions, adjusted for age and sex. Consistent with studies using this technique (12), we derived multiple clustering solutions requesting 2 to 10 clusters in order to obtain a range of possible solutions. The ARI was calculated using 10-fold cross validation to evaluate the stability of each solution; the solution with the highest ARI value was selected for subsequent analyses. If instead a one cluster solution exists, then the reproducibility of the solutions will be poor. Permutation testing was used to statistically evaluate the stability of observed ARI values in comparison to a null distribution (see Supplement).

Group-level statistical analyses

After parsing subtypes of internalizing youth based on structural data, we sought to 1) define how the subtypes differed on demographics, psychopathology, and cognition, 2) understand what structural features (thickness, volume) drove the subtypes discovered, and 3) investigate differences between the subtypes in two independent neuroimaging sequences not used in clustering (ALFF from rsfMRI and fractional anisotropy from DTI). Both linear and nonlinear age effects were modeled using penalized splines within a generalized additive model, which assesses a penalty on nonlinearity using restricted maximum likelihood (REML) in order to avoid over-fitting (47, 48). Age, sex, and image quality (see Supplement for details) were modeled as follows:

$$\text{Region} = \text{spline}(\text{age}) + \text{sex} + \text{image quality rating} + \text{group}$$

Omnibus ANOVAs and pairwise post-hoc tests were corrected for multiple comparisons by controlling the False Discovery Rate (FDR, $Q < 0.05$). Interactions between group and age as well as group and sex were also evaluated. Finally, sensitivity analyses were conducted that excluded those on psychiatric medications and included race as an additional covariate.

Data and code availability

See <https://github.com/PennBBL/KaczurkinHeterogenInternalizing> for all data analysis code used in this manuscript and a wiki detailing what each script does and the order the scripts were run. Data from the Philadelphia Neurodevelopmental Cohort can be accessed at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2. The HYDRA code can be found at <https://github.com/evanol/HYDRA>.

RESULTS

HYDRA identifies subtypes of internalizing youth with a high degree of stability

HYDRA identified k neurostructural subtypes from 210 regional brain features (volume and cortical thickness) after adjusting for age and sex. Evaluation of cluster stability using 10-fold cross-validation exhibited a well-defined peak at $k = 2$ (Figure 2), suggesting the existence of two highly reproducible subtypes (ARI = .66) within internalizing youth. Finding a reproducible solution for $k > 1$ suggests that there is structure in the data (in other words, the data is not homogeneous), since the reproducibility of the solution would be poor if the data were instead characterized by a 1 cluster solution. Permutation results further demonstrated a significantly higher ARI for the 2-subtype solution compared to a null distribution ($p_{fdr} < .001$).

Subtype demographics

As an initial step, we evaluated the demographics of our neurostructural subtypes (Table S1). Groups differed in age, with Subtype 1 being slightly older than typically developing youth; no other age effects were significant. All subsequent analyses controlled for age and sex. Subtype 1 also had lower maternal education than Subtype 2 or typically developing youth, while Subtype 2 did not differ from typically developing youth in this regard. In addition,

both Subtype 1 and Subtype 2 showed higher levels of trauma exposure than typically developing youth; moreover, Subtype 1 had higher levels of trauma exposure compared to Subtype 2. In a subsample of subjects with data on gestational age at birth (n=282), Subtype 1 had a lower gestational age on average than typically developing youth, but did not significantly differ from Subtype 2.

Elevated psychopathology is present in Subtype 1

Next, we evaluated whether the internalizing neurostructural subtypes differed in terms of psychopathology. Psychopathology symptoms were summarized as factors which reflect anxious-misery, psychosis, behavioral, fear, and overall psychopathology. As expected based on the inclusion criteria (patients all met criteria for an internalizing disorder), Subtype 1 and 2 both showed elevated psychopathology symptoms compared to typically developing youth in all domains (Table S1). Accordingly, we focused on differences between the subtypes. Subtype 1 showed greater overall psychopathology symptoms (Figure 3A), behavioral symptoms (Figure 3B), and fear symptoms (Figure 3C) compared to Subtype 2. No significant differences were found between the subtypes for anxious-misery or psychosis. These results demonstrate that Subtype 1 shows a higher burden of psychiatric symptoms than both Subtype 2 and typically developing youth across multiple domains.

Subtype 1 is marked by impaired cognition

We then evaluated whether the subtypes differed in cognitive performance and reading skills. Notably, this independent data was not used in clustering. Subtype 1 showed significantly reduced overall accuracy relative to Subtype 2 and typically developing youth, with no difference found between typically developing youth and Subtype 2 (Table S1). Analyses of the specific accuracy factors revealed that Subtype 1 showed reduced performance relative to Subtype 2 on all factors including executive function/complex reasoning (Figure 3D), social cognition (Figure 3E), and episodic memory (Figure 3F; Table S1). Compared to typically developing youth, Subtype 1 had lower performance only for executive function/complex reasoning and social cognition, but not for episodic memory. Subtype 2 did not significantly differ from typically developing youth on any of these measures. In terms of academic skills, Subtype 1 demonstrated lower reading skills than both Subtype 2 and typically developing youth, who did not differ from each other (Table S1).

Subtypes display markedly divergent patterns of brain structure

Having identified two subtypes of internalizing youth based on total gray matter volume and cortical thickness data, we examined the structural features that drove this clustering. The results demonstrated that Subtype 1 showed smaller regional volumes than either Subtype 2 or typically developing youth for all 112 regions (Figure 4A and S1A; Table S2). Likewise, Subtype 1 had thinner cortex in 96 out of 98 regions compared to Subtype 2 (Figure 4B; Table S2) and in 95 regions compared to typically developing youth (Figure S1B; Table S2). Compared to typically developing youth, Subtype 2 demonstrated greater volume in 111 regions (Figure S2A; Table S2) and greater cortical thickness in 81 regions (Figure S2B; Table S2). Interactions between group and age as well as group and sex were evaluated but found to be nonsignificant except for volume of occipital fusiform gyrus. The results were

similar when examining total brain volume (TBV) or intracranial volume (ICV; Table S2), as TBV and ICV are highly correlated in this developmental sample (Figure S3).

Subtype 1 shows abnormalities in resting-state and diffusion measures

Finally, to further understand differences in these neurostructural subtypes, we examined two independent imaging modalities which were not used in clustering: ALFF from rsfMRI and fractional anisotropy from DTI. ALFF was reduced in Subtype 1 compared to Subtype 2 in 40 frontal cortex and limbic regions, including bilateral middle/superior frontal gyrus, right amygdala, and right hippocampus (Figure 4C; Table S3). Subtype 1 also showed reduced amplitudes in 25 of these regions compared to typically developing youth (Figure S1C; Table S3). Conversely, Subtype 2 demonstrated greater ALFF in 13 regions compared to typically developing youth (Figure S2C; Table S3). These results suggest that Subtype 1 shows abnormalities in the resting-state power spectrum in regions associated with executive functioning and affective processing.

Differences between Subtype 1 and Subtype 2 were also apparent in fractional anisotropy. Subtype 1 showed reduced fractional anisotropy compared to Subtype 2 in 10 out of 18 white matter tracts including the inferior longitudinal fasciculus, uncinate fasciculus, anterior thalamic radiation, corticospinal tracts, parahippocampal cingulum bundle, superior longitudinal fasciculi, and forceps minor (Figure 4D; Table S4). Subtype 1 also showed reduced fractional anisotropy in 8 tracts compared to typically developing youth (Figure S1D; Table S4). Subtype 2 showed relatively similar fractional anisotropy to typically developing youth, with only the left and right anterior thalamic radiations demonstrating higher levels in Subtype 2 (Figure S2D; Table S4). These results emphasize that Subtype 1 has reduced white matter integrity in several key tracts that link frontal cortex and limbic brain regions. The interrelationships between structure, ALFF, fractional anisotropy, and cognition are shown in Figure S4.

Sensitivity analyses provide convergent results

Sensitivity analyses were conducted after excluding the minority of participants who were taking psychotropic medications at the time of imaging (included: $n=1,037$). In this subsample, the pattern of results for demographics, psychopathology, and cognition/academic skills remained highly similar (Table S5). For the structural results, sensitivity analyses yielded nearly identical results (Table S6). In addition, 27 out of 41 resting-state ALFF regions remained significant between the subtypes (Table S7). The fractional anisotropy results were quite similar, with 9 out of 10 tracts remaining significant between Subtype 1 and 2 (Table S8). We also conducted sensitivity analyses while including race as an additional covariate. The results remained similar for structure (Table S9), ALFF (Table S10), and fractional anisotropy (Table S11). Additionally, as a supplemental analysis to the bifactor model, we also examined subtype group differences in each categorical diagnosis separately (Table S12). The results were consistent with the bifactor model, with Subtype 1 and Subtype 2 showing greater symptoms than typically developing youth for all diagnostic categories. Furthermore, Subtype 1 showed greater symptoms than Subtype 2 on disorders related to fear (social anxiety disorder, agoraphobia, and PTSD) and behavioral problems (ODD and conduct disorder). Lastly, controlling for total brain volume (TBV) and average

cortical thickness before clustering with HYDRA produces clusters with a very low adjusted Rand index ($ARI < .20$), a measure of out-of-sample reproducibility.

DISCUSSION

Capitalizing on a large sample of youth and recent advances in semi-supervised machine learning, we identified two reliable neurostructural subtypes of internalizing disorders. Subtype 1 was marked by elevated levels of psychopathology, impaired cognition, and multiple deficits apparent on multi-modal imaging. These deficits included smaller gray matter volumes and thinner cortices, reduced ALFF in fronto-limbic cortex, and reduced integrity of white matter tracts. In contrast, Subtype 2 had preserved cognitive functioning and brain integrity despite clinically significant levels of psychopathology. These results provide a new account of the heterogeneity in brain structure and function present in youth with internalizing disorders.

Heterogeneous neurostructural abnormalities in internalizing disorders

The pattern of deficits revealed between the subtypes of internalizing youth illustrates the detrimental effects associated with abnormal structural development. Reduced volume and cortical thickness are associated with numerous detrimental effects including deficits in cognitive functioning (52, 53), impaired academic skills (53-55), and greater psychopathology (56-61). Our results show widespread effects for both volume and cortical thickness, with Subtype 1 showing deficits across the entire brain compared to Subtype 2 and typically developing youth. Group differences were apparent when examining total gray matter volume, TBV, or ICV, consistent with the high interrelationships between these variables in developmental samples (49). When TBV is controlled for along with age and sex before clustering, no reliable clustering solution emerges. There are two possible explanations for these results. One possibility is that the distributed pattern found may mediate the separation of the groups. A second hypothesis is that this distributed pattern directly occurs as a result of smaller ICV, as smaller heads have smaller volumes, and are also associated with lower IQ and greater psychopathology in studies of prematurity (50, 51). However, the very high correlation between ICV and TBV in this sample does not allow us to disentangle these competing hypotheses.

These structural abnormalities are likely the result of a combination of genetic and environmental effects (62). Environmental factors, such as low SES and childhood adversity, are associated with chronic exposure to stress hormones (63, 64), which have been shown to impact the development of structures related to psychopathology (65, 66) and cognition (67, 68). In line with this prior research, our subtypes show an association between neurostructural deficits and lower SES, greater trauma exposure, greater levels of psychopathology, and impaired cognition. This is consistent with prior work from our group and others showing a robust relationship between psychopathology and structural brain deficits (56-61). Additionally, it is possible that these differences were established in utero. While we found that Subtype 1 had a lower gestational age than typically developing youth in a subset of the data, Subtype 1 did not differ from Subtype 2, suggesting that preterm birth may not account for the differences between the subtypes. However, we only had

gestational age data for a relatively small number of subjects, thus, our analyses may have been underpowered to detect a significant effect. Thus, we cannot rule out the possibility that our results are related to birth complications.

We expand on previous work by showing related deficits in two independent modalities, with neurostructural deficits associated with reduced resting-state ALFF and fractional anisotropy in white matter tracts. Reduced resting-state ALFF in fronto-limbic regions may reflect dysregulated regional spontaneous neural activity (37) in executive functioning regions, consistent with poorer cognitive performance. Our results are also consistent with prior studies showing reduced resting-state ALFF in frontal regions in children with ADHD (37). Additionally, reduced white matter integrity in tracts such as the inferior longitudinal fasciculus, uncinate fasciculus, and forceps minor is consistent with previous studies implicating these tracts in depression (69, 70), ADHD (71, 72), and poorer cognitive functioning (73).

In contrast to the deficits seen in Subtype 1, Subtype 2 was characterized by preserved brain structure and cognitive functioning, but still showed high levels of psychopathology. This may suggest compensatory mechanisms, whereby individuals with greater brain reserve can compensate for deficits typically associated with psychopathology, allowing for preserved cognition (74). These results demonstrate the impact of abnormal structural development on cognitive and affective functioning, with greater neural resources potentially mitigating detrimental effects on cognition. However, this does not explain why preserved brain structure, function, and cognition does not also protect against psychiatric symptoms, suggesting disparate pathways to apparently similar manifestations of psychopathology (75).

Advances in parsing neurobiological heterogeneity in youth

The results of the current study provide both conceptual and methodological advances in the classification of internalizing disorders in youth. Prior studies have primarily used symptom-based diagnostic categories to explore associated neurobiological mechanisms in a case-control design, or examined associations between dimensional clinical phenotypes and imaging measures across diagnostic categories. More recent efforts have used clustering techniques to identify subtypes in Alzheimer's disease using structural data (12-14), in depression using resting-state connectivity data (15, 16), and in psychosis using multimodal data (17-19, 76-78). However, several of these studies defined subtypes using symptoms or cognitive performance rather than imaging measures. Within youth, functional connectivity data has been used to reveal neurobiological subtypes of attention-deficit/hyperactivity disorder using community detection (20, 21). While clustering methods have also been used previously to identify subtypes of internalizing adults (79), these methods have clustered on symptoms only and then related symptom subtypes to physiological measures. The current study extends this research by using structural data to delineate neuroanatomical subtypes of internalizing symptoms, which were then evaluated using independent cognitive and neuroimaging measures. Additionally, while the majority of prior studies on heterogeneity have considered adults, our results build upon this work by parsing neurostructural heterogeneity in a community-based sample of youth.

In addition, this study is distinctive from prior efforts in its methodological approach. Until relatively recently, the majority of approaches for understanding the neurobiological differences underlying psychiatric symptoms (case-control, linear SVMs) have assumed that a single discriminative pattern differentiates those with psychiatric symptoms from healthy controls. However, as noted above, there has been a movement towards using methods that parse neurobiological heterogeneity within patient groups (12-21). While unsupervised methods cluster patients based on how similar they are to each other, one advantage of the approach taken here is that the semi-supervised learning procedure implemented in HYDRA allows us to cluster patients by how different they are from typically developing youth, yielding subtype-specific neurobiological signatures that differ from controls. In contrast, traditional clustering techniques are susceptible to splitting the data on non-specific factors such as age or sex, producing clusters that may not be aligned well on psychopathology. In the current study, HYDRA allowed us to identify two subtypes of internalizing youth that differed from controls on multiple clinical, cognitive, and imaging measures of interest.

Several limitations should be noted. First, longitudinal designs are needed to determine the trajectory of these neurostructural differences over time in youth. A second limitation of the current study is the lack of data collected on household income as a measure of SES. However, we used maternal education as a proxy for SES as prior work shows that maternal level of education is more strongly associated with several developmental outcomes than other measures of SES (80-82). Third, the current study only had gestational age data from a subsample of the cohort (25), thus future work would benefit from replicating these results in a sample with data on gestational age as well as information on pregnancy and birth complications. Fourth, future work would benefit from performing similar clustering in an older sample with internalizing symptoms, as ICV and TBV are highly correlated in this developmental sample, making it difficult to disentangle the relative contributions of each to these clustering patterns. Fifth, the putative mechanisms (e.g., environmental vs. genetic) driving the differences between the subtypes could be tested by integrating genetic data into future work. Sixth and finally, while machine learning can find reliable and distinct subtypes, this does not rule out the possibility that dimensional approaches can also provide useful and important information.

Taken together, this study provides important data regarding neurostructural heterogeneity in youth with internalizing disorders. Given the early age of onset found in anxiety and depressive symptoms during development (3), biomarkers that dissect heterogeneous neural patterns in a developmental sample may aid in identifying youth at risk for these symptoms. A greater understanding of how abnormalities in the brain give rise to these symptoms in youth is critical for the development of earlier and more effective treatments that may reduce the negative long-term outcomes associated with internalizing symptoms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institute of Mental Health (RC2 grants MH089983 and MH089924 to REG, K99MH117274 to ANK, R01MH107703 and R01MH113550 to TDS, R01NS085211 to RTS, R01MH112847 to RTS and TDS, R01MH107235 to RCG, R01MH11365 to DHW, and R01MH112070 and R01EB022573 to CD), the Dowshen Program for Neuroscience, the Lifespan Brain Institute at the Children's Hospital of Philadelphia and Penn Medicine, the NARSAD Young Investigator Award (ANK), and the Center for Biomedical Computing and Image Analysis (CBICA) at Penn for developing statistical analyses (RTS & TDS) and multivariate pattern analysis software (AS & TDS). An earlier version of this manuscript is available as a preprint on bioRxiv.

REFERENCES

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005): Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 62: 593–602. [PubMed: 15939837]
2. McLean CP, Asnaani A, Litz BT, Hofmann SG (2011): Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *J Psychiatr Res*. 45: 1027–1035. [PubMed: 21439576]
3. Roza SJ, Hofstra MB, Van Der Ende J, Verhulst FC (2003): Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: A 14-year follow-up during childhood, adolescence, and young adulthood. *Am J Psychiatry*. 160: 2116–2121. [PubMed: 14638580]
4. Kessler RC, Chiu WT, Demler O, Walters EE (2005): Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 62: 617–627. [PubMed: 15939839]
5. Hettema JM (2008): What is the genetic relationship between anxiety and depression? *Am J Med Genet Part C Semin Med Genet*. 148: 140–146.
6. Martin JLR, Sainz-Pardo M, Furukawa TA, Martin-Sanchez E, Seoane T, Galan C (2007): Benzodiazepines in generalized anxiety disorder: heterogeneity of outcomes based on a systematic review and meta-analysis of clinical trials. *J Psychopharmacol*. 21: 774–782. [PubMed: 17881433]
7. Zhao Y, Chen L, Zhang W, Xiao Y, Shah C, Zhu H, et al. (2017): Gray Matter Abnormalities in Non-comorbid Medication-naïve Patients with Major Depressive Disorder or Social Anxiety Disorder. *EBioMedicine*. 21: 228–235. [PubMed: 28633986]
8. Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS (2006): A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry*. 163: 857–864. [PubMed: 16648327]
9. Etkin A, Wager TD (2007): Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am J Psychiatry*. 164: 1476–1488. [PubMed: 17898336]
10. Krueger RF (1999): The Structure of Common Mental Disorders. *Arch Gen Psychiatry*. 56: 921. [PubMed: 10530634]
11. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. (2010): Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 167: 748–751. [PubMed: 20595427]
12. Varol E, Sotiras A, Davatzikos C (2017): HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage*. 145: 346–364. [PubMed: 26923371]
13. Dong A, Toledo JB, Honnorat N, Doshi J, Varol E, Sotiras A, et al. (2017): Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*. 140: 735–747. [PubMed: 28003242]
14. Dong A, Honnorat N, Gaonkar B, Davatzikos C (2016): CHIMERA: Clustering of heterogeneous disease effects via distribution matching of imaging patterns. *IEEE Trans Med Imaging*. 35: 612–621. [PubMed: 26452275]

15. 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. [PubMed: 27918562]
16. Tokuda T, Yoshimoto J, Shimizu Y, Okada G, Takamura M, Okamoto Y, et al. (2018): Identification of depression subtypes and relevant brain regions using a data-driven approach. *Sci Rep.* 8: 1–13. [PubMed: 29311619]
17. Brodersen KH, Deserno L, Schlagenhaut F, Lin Z, Penny WD, Buhmann JM, Stephan KE (2014): Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage Clin.* 4: 98–111. [PubMed: 24363992]
18. Clementz BA, Sweeney JA, Hamm JP, Ivleva EI, Ethridge LE, Pearlson GD, et al. (2016): Identification of distinct psychosis biotypes using brain-based biomarkers. *Am J Psychiatry.* 173: 373–384. [PubMed: 26651391]
19. Sun H, Lui S, Yao L, Deng W, Xiao Y, Zhang W, et al. (2015): Two patterns of white matter abnormalities in medication-naive patients with first-episode schizophrenia revealed by diffusion tensor imaging and cluster analysis. *JAMA Psychiatry.* 72: 678–686. [PubMed: 25993492]
20. Gates KM, Molenaar PCM, Iyer SP, Nigg JT, Fair DA (2014): Organizing heterogeneous samples using community detection of GIMME-Derived resting state functional networks. *PLoS One.* 9: 1–11.
21. Costa Dias TG, Iyer SP, Carpenter SD, Cary RP, Wilson VB, Mitchel SH, et al. (2015): Characterizing heterogeneity in children with and without ADHD based on reward system connectivity. *Dev Cogn Neurosci.* 11: 155–174. [PubMed: 25660033]
22. Winkler AM, Kochunov P, Blangero J, Almasy L, Zilles K, Fox PT, et al. (2010): Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage.* 53: 1135–1146. [PubMed: 20006715]
23. Satterthwaite TD, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA, et al. (2016): The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage.* 124: 1115–1119. [PubMed: 25840117]
24. Satterthwaite TD, Elliott MA, Ruparel K, Loughhead J, Prabhakaran K, Calkins ME, et al. (2014): Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *Neuroimage.* 86: 544–553. [PubMed: 23921101]
25. Nassar R, Kaczurkin AN, Xia CH, Sotiras A, Pehlivanova M, Moore TM, et al. (2018): Gestational Age is Dimensionally Associated with Structural Brain Network Abnormalities Across Development. *Cereb Cortex.* 29: 2102–2114.
26. Calkins ME, Merikangas KR, Moore TM, Burstein M, Behr A, Satterthwaite TD, et al. (2015): The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry.* 56: 1356–1369. [PubMed: 25858255]
27. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. (1997): Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *J Am Acad Child Adolesc Psychiatry.* 36: 980–988. [PubMed: 9204677]
28. Kaczurkin AN, Moore TM, Calkins ME, Ciric R, Detre JA, Elliott MA, et al. (2018): Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses. *Mol Psychiatry.* 23: 1981–1989. [PubMed: 28924181]
29. Shanmugan S, Wolf DH, Calkins ME, Moore TM, Ruparel K, Hopson RD, et al. (2016): Common and dissociable mechanisms of executive system dysfunction across psychiatric disorders in youth. *Am J Psychiatry.* 173: 517–526. [PubMed: 26806874]
30. Reise SP, Moore TM, Haviland MG (2010): Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *J Pers Assess.* 92: 544–559. [PubMed: 20954056]
31. Gibbons RD, Hedeker DR (1992): Full-information item bi-factor analysis. *Psychometrika.* 57: 423–436.
32. Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC (2015): Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology.* 29: 235–46. [PubMed: 25180981]

33. Wilkinson G, Robertson G (2006): Wide Range Achievement Test, Fourth Edition. Lutz, FL: Psychological Assessment Resources, Inc.
34. Ciric R, Wolf DH, Power JD, Roalf DR, Baum GL, Ruparel K, et al. (2017): Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*. 154: 174–187. [PubMed: 28302591]
35. Roalf DR, Quarmley M, Elliott MA, Satterthwaite TD, Vandekar SN, Ruparel K, et al. (2016): The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort. *Neuroimage*. 125: 903–919. [PubMed: 26520775]
36. Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. (2014): Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage*. 99: 166–179. [PubMed: 24879923]
37. Yu-feng Z, Yong H, Chao-zhe Z, Qing-jiu C, Man-qiu S, Meng L, et al. (2007): Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev*. 29: 83–91. [PubMed: 16919409]
38. Liu X, Chen J, Shen B, Wang G, Li J, Hou H, et al. (2018): Altered Intrinsic Coupling between Functional Connectivity Density and Amplitude of Low-Frequency Fluctuation in Mild Cognitive Impairment with Depressive Symptoms. *Neural Plast*. 2018: 1–8.
39. Liu J, Ren L, Womer FY, Wang J, Fan G, Jiang W, et al. (2014): Alterations in amplitude of low frequency fluctuation in treatment-naïve major depressive disorder measured with resting-state fMRI. *Hum Brain Mapp*. 35: 4979–4988. [PubMed: 24740815]
40. Bing X, Qiu MG, Ye Z, Zhang JN, Min L, Han C, et al. (2013): Alterations in the cortical thickness and the amplitude of low-frequency fluctuation in patients with post-traumatic stress disorder. *Brain Res*. 1490: 225–232. [PubMed: 23122880]
41. Hoptman MJ, Zuo X-N, Butler PD, Javitt DC, D’Angelo D, Mauro CJ, Milham MP (2010): Amplitude of low-frequency oscillations in schizophrenia: a resting state fMRI study. *Schizophr Res*. 117: 13–20. [PubMed: 19854028]
42. Zhou J, Yao N, Fairchild G, Zhang Y, Wang X (2015): Altered hemodynamic activity in conduct disorder: A resting-state fMRI investigation. *PLoS One*. 10: 1–12.
43. Wang P, Yang J, Yin Z, Duan J, Zhang R, Sun J, et al. (2019): Amplitude of low-frequency fluctuation (ALFF) may be associated with cognitive impairment in schizophrenia: a correlation study. *BMC Psychiatry*. 19: 1–10. [PubMed: 30606141]
44. Wang H, Suh JW, Das SR, Pluta JB, Craige C, Yushkevich PA (2013): Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell*. 35: 611–623. [PubMed: 22732662]
45. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012): FSL. *Neuroimage*. 62: 782–790. [PubMed: 21979382]
46. Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS, et al. (2008): Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *Neuroimage*. 39: 336–347. [PubMed: 17931890]
47. Wood SN (2004): Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc*. 99: 673–686.
48. Wood SN (2001): mgcv: GAMs and generalized ridge regression for R. *R News*. 1: 20–25.
49. Royle NA, Booth T, Valdés Hernandez MC, Penke L, Murray C, Gow AJ, et al. (2013): Estimated maximal and current brain volume predict cognitive ability in old age. *Neurobiol Aging*. 34: 2726–2733. [PubMed: 23850342]
50. Jensen RB, Juul A, Larsen T, Mortensen EL, Greisen G (2015): Cognitive ability in adolescents born small for gestational age: Associations with fetal growth velocity, head circumference and postnatal growth. *Early Hum Dev*. 91: 755–760. [PubMed: 26411400]
51. Johnson S, Marlow N (2011): Preterm birth and childhood psychiatric disorders. *Pediatr Res*. 69: 22–28.
52. Karama S, Ad-dab Y, Haier RJ, Deary IJ, Lyttelton OC, Lepage C, Evans AC (2009): Intelligence Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence*. 37: 145–155. [PubMed: 20161325]
53. Nave G, Jung WH, Linnér RK, Kable JW, Koellinger PD (2019): Are Bigger Brains Smarter? Evidence From a Large-Scale Preregistered Study. *Psychol Sci*. 30: 43–54. [PubMed: 30499747]

54. Hair NL, Hanson JL, Wolfe BL, Pollak SD (2015): Association of Child Poverty, Brain Development, and Academic Achievement. *JAMA Pediatr.* 169: 822–829. [PubMed: 26192216]
55. Mackey AP, Finn AS, Leonard JA, Jacoby-Senghor DS, West MR, Gabrieli CF, Gabrieli JDE (2015): Neuroanatomical Correlates of the Income–Achievement Gap. *Psychol Sci.* 26: 925–933. [PubMed: 25896418]
56. 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. [PubMed: 21865529]
57. Valera EM, Faraone SV, Murray KE, Seidman LJ (2007): Meta-Analysis of Structural Imaging Findings in Attention-Deficit/Hyperactivity Disorder. *Biol Psychiatry.* 61: 1361–1369. [PubMed: 16950217]
58. Castellanos XF, Lee PP, Sharp W, Jeffries NO, Greenstein DK, Clasen LS, et al. (2002): Developmental Trajectories of Brain Volume Abnormalities in Children and Adolescents With Attention-Deficit/Hyperactivity Disorder. *JAMA.* 288: 1740–1748. [PubMed: 12365958]
59. Arnone D, McIntosh AM, Ebmeier KP, Munafò MR, Anderson IM (2012): Magnetic resonance imaging studies in unipolar depression: Systematic review and meta-regression analyses. *Eur Neuropsychopharmacol.* 22: 1–16. [PubMed: 21723712]
60. Kaczurkin 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.* in press.
61. Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, et al. (2007): Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc Natl Acad Sci.* 104: 19649–19654. [PubMed: 18024590]
62. Andersen SL (2003): Trajectories of brain development: Point of vulnerability or window of opportunity? *Neurosci Biobehav Rev.* 27: 3–18. [PubMed: 12732219]
63. Lupien S, McEwen B, Gunnar M, Heim C (2009): Effects of stress throughout the lifespan on the brain, behaviour, and cognition. *Nat Rev Neurosci.* 10: 434–445. [PubMed: 19401723]
64. McEwen BS, Gianaros PJ (2010): Central role of the brain in stress and adaptation: Links to socioeconomic status, health, and disease. *Ann N Y Acad Sci.* 1186: 190–222. [PubMed: 20201874]
65. Heim C, Newport DJ, Mletzko T, Miller AH, Nemeroff CB (2008): The link between childhood trauma and depression: Insights from HPA axis studies in humans. *Psychoneuroendocrinology.* 33: 693–710. [PubMed: 18602762]
66. De Bellis M, Keshavan M, Clark D, Casey B, Giedd J, Boring A, et al. (1999): Developmental Traumatology Part II: Brain Development. *Biol Psychiatry.* 45: 1271–1284. [PubMed: 10349033]
67. 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. [PubMed: 25821911]
68. Piccolo LR, Merz EC, He X, Sowell ER, Noble KG (2016): Age-related differences in cortical thickness vary by socioeconomic status. *PLoS One.* 11: 1–18.
69. Bhatia KD, Henderson LA, Hsu E, Yim M (2018): Reduced integrity of the uncinate fasciculus and cingulum in depression: A stem-by-stem analysis. *J Affect Disord.* 235: 220–228. [PubMed: 29656270]
70. Lai CH, Wu Y Te (2016): The white matter microintegrity alterations of neocortical and limbic association fibers in major depressive disorder and panic disorder: The comparison. *Med (United States).* 95: 1–6.
71. Qiu MG, Ye Z, Li QY, Liu GJ, Xie B, Wang J (2011): Changes of Brain structure and function in ADHD children. *Brain Topogr.* 24: 243–252. [PubMed: 21191807]
72. van Ewijk H, Heslenfeld DJ, Zwiers MP, Buitelaar JK, Oosterlaan J (2012): Diffusion tensor imaging in attention deficit/hyperactivity disorder: A systematic review and meta-analysis. *Neurosci Biobehav Rev.* 36: 1093–1106. [PubMed: 22305957]
73. Ursache A, Noble KG (2016): Socioeconomic status, white matter, and executive function in children. *Brain Behav.* 6: 1–13.
74. Stern Y (2009): Cognitive reserve. *Neuropsychologia.* 47: 2015–2028. [PubMed: 19467352]

75. Barzilay R, Calkins ME, Moore TM, Boyd RC, Jones JD, Benton TD, et al. (2019): Neurocognitive functioning in community youth with suicidal ideation: gender and pubertal effects. *Br J Psychiatry*. in press.
76. Bell MD, Corbera S, Johannesen JK, Fiszdon JM, Wexler BE (2013): Social cognitive impairments and negative symptoms in schizophrenia: Are there subtypes with distinct functional correlates? *Schizophr Bull.* 39: 186–196. [PubMed: 21976710]
77. Geisler D, Walton E, Naylor M, Roessner V, Lim KO, Charles Schulz S, et al. (2015): Brain structure and function correlates of cognitive subtypes in schizophrenia. *Psychiatry Res - Neuroimaging.* 234: 74–83.
78. Lewandowski KE, Sperry SH, Cohen BM, Öngür D (2014): Cognitive variability in psychotic disorders: A cross-diagnostic cluster analysis. *Psychol Med.* 44: 3239–3248. [PubMed: 25066202]
79. Grisanzio KA, Goldstein-Piekarski AN, Wang MY, Ahmed APR, Samara Z, Williams LM (2018): Transdiagnostic symptom clusters and associations with brain, behavior, and daily function in mood, anxiety, and trauma disorders. *JAMA Psychiatry.* 75: 201–209. [PubMed: 29197929]
80. Bradley R, Corwyn R (2013): Age and ethnic variations in family process mediators of SES In: Bornstein M, Bradley R, editors. *Socioecon Status, Parenting, Child Dev.* Mahwah, NJ: Lawrence Erlbaum Associates, pp 161–188.
81. Fluss J, Ziegler JC, Warszawski J, Ducot B, Richard G, Billard C (2009): Poor reading in French Elementary School: The interplay of cognitive, behavioral, and socioeconomic factors. *J Dev Behav Pediatr.* 30: 206–216. [PubMed: 19412126]
82. Raviv T, Kessenich M, Morrison FJ (2004): A mediational model of the association between socioeconomic status and three-year-old language abilities: The role of parenting factors. *Early Child Res Q.* 19: 528–547.

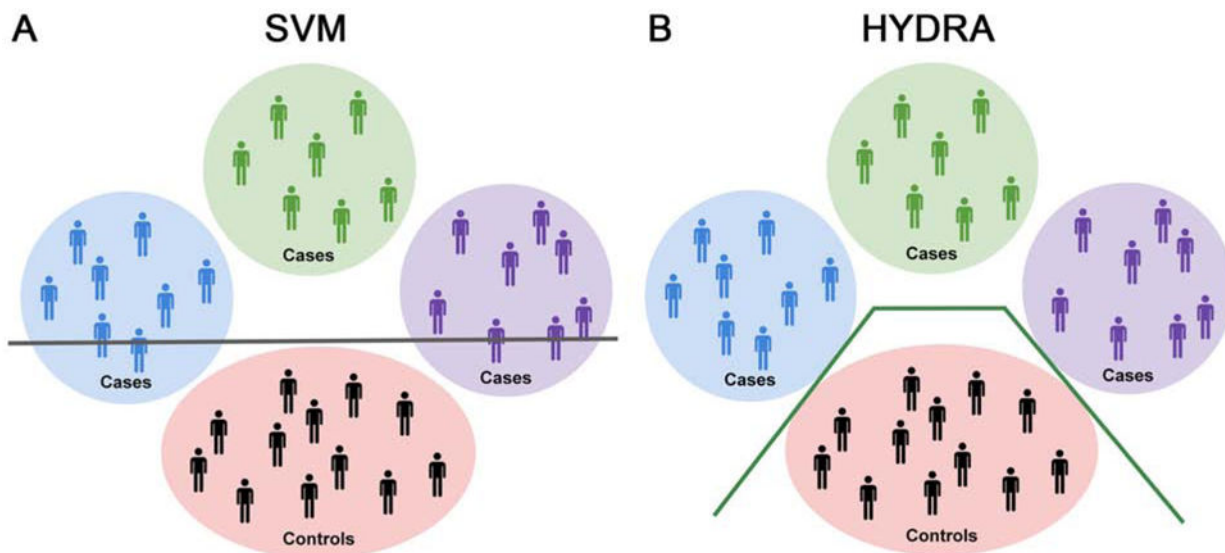


Figure 1. Schematic representing the utility of HYDRA over SVMs for parsing heterogeneity.
A) Schematic illustrating the use of a linear support vector machine (SVM) to separate cases from controls with a separating hyperplane, shown here as a gray line. Heterogeneity within the cases is represented by the blue, green, and purple circles. As can be seen in this schematic, linear SVMs do not capture the heterogeneity that exists in the cases. **B)** Conversely, HYDRA is able to classify each cluster of cases separately from the controls. This is accomplished by using multiple classifiers that form linear hyperplanes (green lines) whose segments separate the clusters of cases from the controls. The goal is to estimate k hyperplanes that distinguish the controls and cases with the largest margin, thus allowing HYDRA to identify heterogeneous groups within the cases.

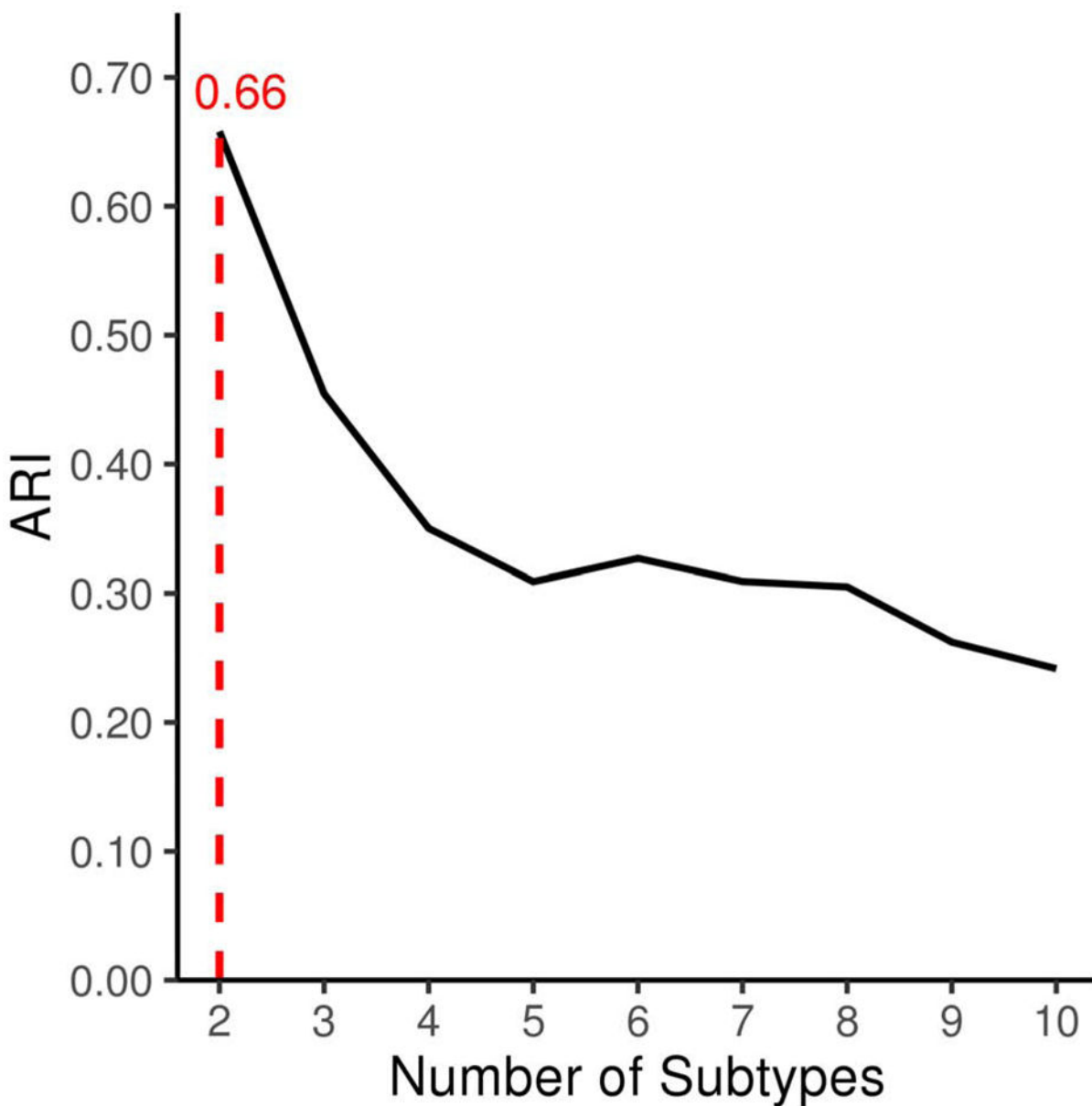


Figure 2. HYDRA identifies 2 subtypes of internalizing youth with a high level of stability. Cross-validated adjusted Rand index (ARI) for 2-10 cluster solutions obtained with HYDRA. The ARI was computed with 10-fold cross-validation to quantify the similarity between different clustering results while controlling for grouping by chance, resulting in a more conservative estimation of the overlap between clustering solutions. The figure shows a clear peak at the 2-cluster solution (shown with a dotted line), suggesting our data has 2 subtypes of internalizing youth which show a high degree of stability (ARI = .66).

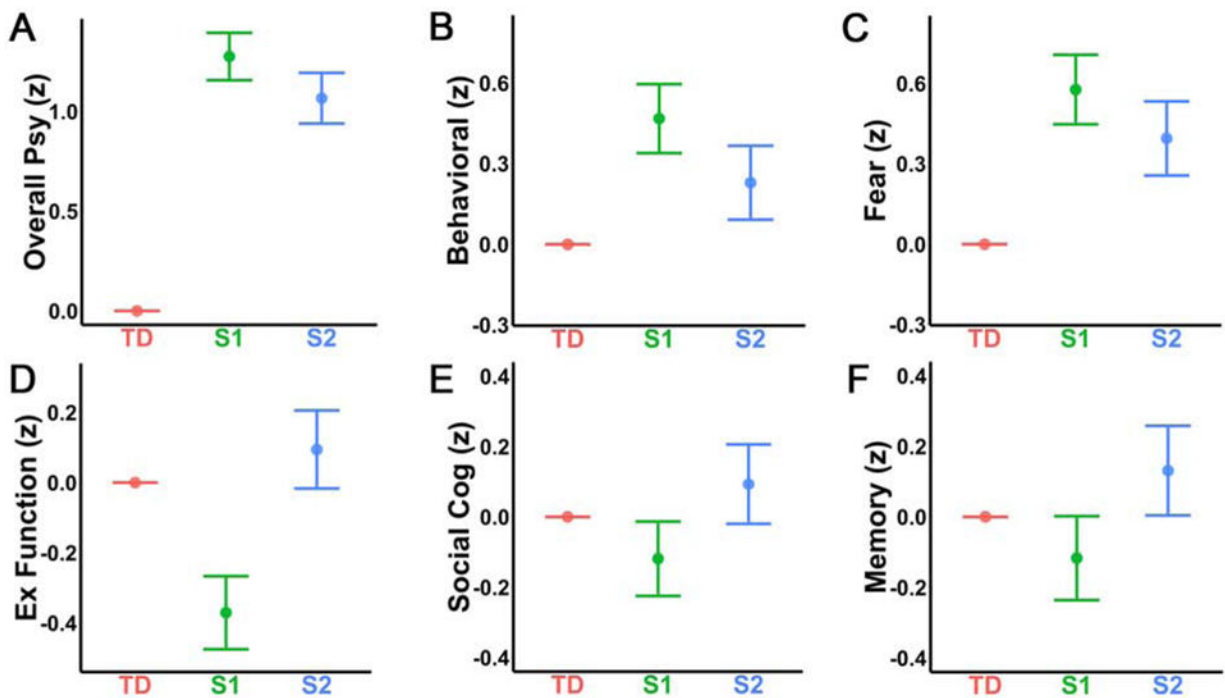


Figure 3. Subtype 1 shows greater psychopathology and poorer cognitive performance.

Estimates are shown on the Y-axis from the fitted model testing for group differences in each domain. Each vertical line represents the 95% confidence interval (CI), with the comparison group (typically developing youth: TD) represented by its mean line. The subtype is significantly different from TD if its corresponding CI does not contain 0 (the mean of TD). To examine differences in psychopathology, symptoms were summarized into anxious-misery, psychosis, behavioral, fear, and overall psychopathology factors (28, 29, 60). As expected, both Subtype 1 (S1) and Subtype 2 (S2) showed greater levels of psychopathology compared to TD across all psychopathology factors, thus, we focus on the differences between S1 and S2. S1 showed higher levels of **A**) overall psychopathology, **B**) behavioral symptoms, and **C**) fear symptoms than S2. There were no significant differences between S1 and S2 for the anxious-misery or psychosis factors. In terms of cognition, S1 showed significantly lower performance on **D**) executive functioning tasks than both TD and S2; S2 did not differ from TD. S1 also performed more poorly relative to the other two groups in terms of **E**) social cognition, while S2 again did not differ from TD. Additionally, S1 performed significantly below S2 on **F**) episodic memory; however, neither S1 nor S2 significantly differed from TD.

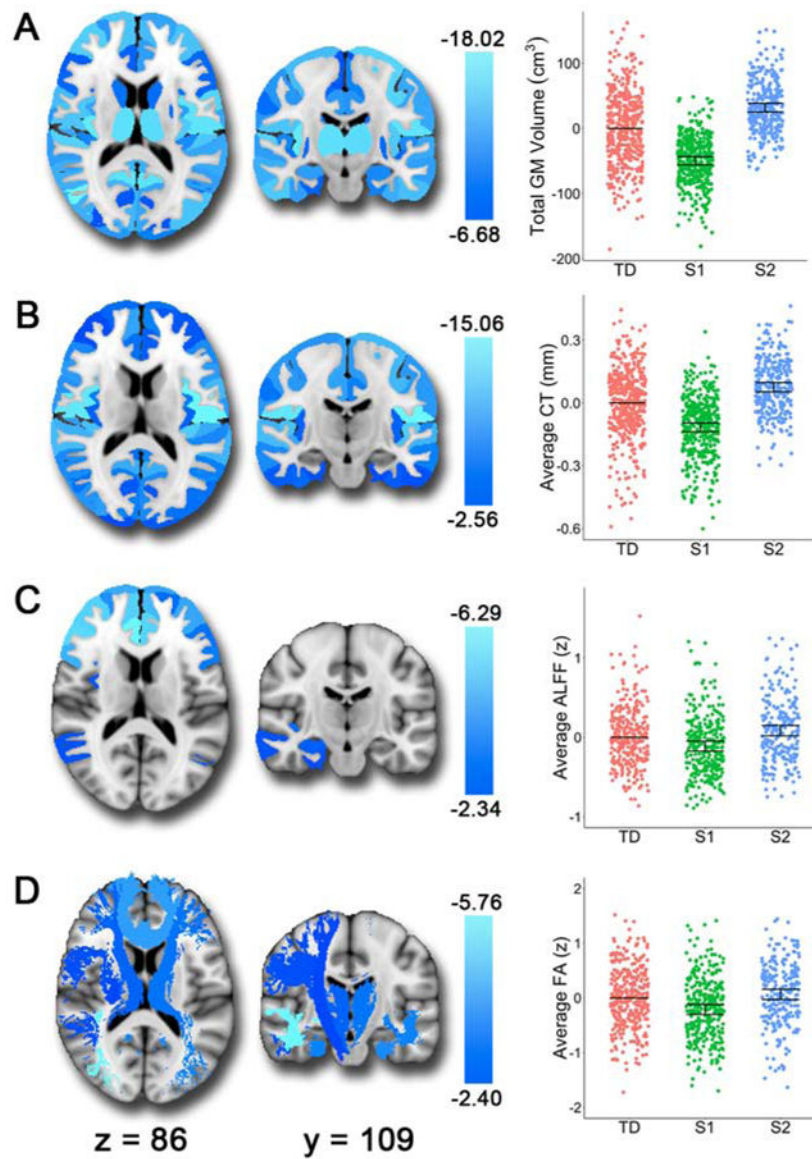


Figure 4. Subtype 1 (S1) shows smaller volume, thinner cortex, lower resting-state ALFF, and reduced white matter integrity relative to Subtype 2 (S2).

The brain images show the t-values for the S1>S2 contrast. In the scatterplots, we show the estimates from the fitted GAM model with all three groups for comparison. Each vertical line represents the 95% confidence interval (CI), with the comparison group (typically developing youth: TD) represented by its mean line. The subtype is significantly different from TD if its corresponding CI does not contain 0 (the mean of TD). **A)** S1 showed smaller volumes than S2 consistently across the brain. **B)** In terms of cortical thickness, S1 shows reduced cortical thickness compared to S2 in all regions except the left and right entorhinal cortices. **C)** S1 also demonstrated reduced resting-state ALFF (amplitude of low-frequency fluctuations) in frontal regions, the right amygdala, and the right hippocampus compared to S2. **D)** Finally, relative to S2, we found that S1 showed reduced fractional anisotropy in white matter tracts including the inferior longitudinal fasciculi, uncinate fasciculus, anterior

thalamic radiation, corticospinal tract, parahippocampal cingulum bundle, superior longitudinal fasciculus, and forceps minor.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Summary of demographic data

	TD (n=426)		S1 (n=403)		S2 (n=312)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age (years)	14.68	4.05	15.35	3.35	15.10	3.66
	N	Percent	N	Percent	N	Percent
Gender						
Female	208	49%	240	60%	179	57%
Male	218	51%	163	40%	133	43%
Trauma exposure						
No trauma	295	69%	146	36%	142	46%
1 trauma	86	20%	113	28%	92	29%
2+ traumas	45	11%	144	36%	78	25%
Internalizing Disorders (N=715)						
Agoraphobia	-	-	51	4%	27	2%
Generalized Anxiety Disorder	-	-	13	1%	14	1%
Major Depression	-	-	109	10%	82	7%
Obsessive-Compulsive Disorder	-	-	26	2%	17	1%
Panic	-	-	10	.9%	4	.4%
PTSD	-	-	112	10%	56	5%
Separation Anxiety	-	-	29	3%	34	3%
Social Anxiety	-	-	193	17%	125	11%
Specific Phobia	-	-	230	20%	184	16%
Co-morbid Disorders						
ADHD	-	-	77	7%	70	6%
Anorexia	-	-	10	.8%	5	.4%
Bulimia	-	-	2	.2%	3	.3%
Conduct Disorder	-	-	65	6%	19	2%
Mania	-	-	7	.6%	5	.4%
Oppositional Defiant Disorder	-	-	200	18%	106	9%
Psychosis-spectrum	-	-	169	15%	106	9%

Note. *Due to comorbidity, individual participants may be present in more than one category of lifetime prevalence.

KEY RESOURCES TABLE

Resource Type	Specific Reagent or Resource	Source or Reference	Identifiers	Additional Information
Add additional rows as needed for each resource type	Include species and sex when applicable.	Include name of manufacturer, company, repository, individual, or research lab. Include PMID or DOI for references; use “this paper” if new.	Include catalog numbers, stock numbers, database IDs or accession numbers, and/or RRIDs. RRIDs are highly encouraged; search for RRIDs at https://scicrunch.org/resources .	Include any additional information or notes if necessary.
Deposited Data; Public Database	Philadelphia Neurodevelopmental Cohort	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2	NA	
Software; Algorithm	HYDRA (Heterogeneity through Discriminative Analysis)	https://github.com/evarol/HYDRA	NA	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript