Archival Report

Biological Psychiatry

Data-Driven Phenotypic Categorization for Neurobiological Analyses: Beyond DSM-5 Labels

Nicholas T. Van Dam, David O'Connor, Enitan T. Marcelle, Erica J. Ho, R. Cameron Craddock, Russell H. Tobe, Vilma Gabbay, James J. Hudziak, F. Xavier Castellanos, Bennett L. Leventhal, and Michael P. Milham

ABSTRACT

BACKGROUND: Data-driven approaches can capture behavioral and biological variation currently unaccounted for by contemporary diagnostic categories, thereby enhancing the ability of neurobiological studies to characterize brain-behavior relationships.

METHODS: A community-ascertained sample of individuals (N = 347, 18–59 years of age) completed a battery of behavioral measures, psychiatric assessment, and resting-state functional magnetic resonance imaging in a cross-sectional design. Bootstrap-based exploratory factor analysis was applied to 49 phenotypic subscales from 10 measures. Hybrid hierarchical clustering was applied to resultant factor scores to identify nested groups. Adjacent groups were compared via independent samples t tests and chi-square tests of factor scores, syndrome scores, and psychiatric prevalence. Multivariate distance matrix regression examined functional connectome differences between adjacent groups.

RESULTS: Reduction yielded six factors, which explained 77.8% and 65.4% of the variance in exploratory and constrained exploratory models, respectively. Hybrid hierarchical clustering of these six factors identified two, four, and eight nested groups (i.e., phenotypic communities). At the highest clustering level, the algorithm differentiated functionally adaptive and maladaptive groups. At the middle clustering level, groups were separated by problem type (maladaptive groups; internalizing vs. externalizing problems) and behavioral type (adaptive groups; sensation-seeking vs. extraverted/emotionally stable). Unique phenotypic profiles were also evident at the lowest clustering level. Group comparisons exhibited significant differences in intrinsic functional connectivity at the highest clustering level in somatomotor, thalamic, basal ganglia, and limbic networks.

CONCLUSIONS: Data-driven approaches for identifying homogenous subgroups, spanning typical function to dysfunction, not only yielded clinically meaningful groups, but also captured behavioral and neurobiological variation among healthy individuals.

Keywords: Hierarchical clustering, Multivariate distance matrix regression, Phenotypes, Psychopathology, RDoC, Resting state fMRI

http://dx.doi.org/10.1016/j.biopsych.2016.06.027

The limitations of categorical definitions of psychiatric illness for clinical practice (1) and psychiatric research (2) are increasingly apparent. Although diagnostic labels defined in nosological systems such as the DSM-5 (3) are needed for clinical practice, these systems impede the search for pathophysiological markers using epidemiologic, genetic, and neuroimaging approaches (4). Given growing recognition of these limitations, the Research Domain Criteria (RDoC) project has called for the development of a new nosology (5). In response, empirical data are being used to identify target phenotypic domains and constructs to characterize psychopathology and guide psychological and neurobiological investigations.

Not surprisingly, how to best delineate phenotypic domains or constructs to guide a nonsyndromal research framework is uncertain (6). Inherent to this pursuit is the varying utility of categorical and dimensional frameworks. Although dimensional models of psychopathology are widely supported (7–9), fully dimensional perspectives have limitations with respect to clinical decision making (10,11). Further, it is unclear whether it is more expedient to derive phenotypic targets from existing models (based on existing data and theory), data-driven analytic approaches (12,13), or some combination. Psychiatric classification systems have variable derivations spanning clinical and research observations (e.g., DSM) as well as empirical assessment (e.g., Achenbach System of Empirically Based Assessment), with other entities exhibiting a combination (e.g., RDoC). Consensus-driven methods [e.g., DSM; also arguably RDoC (14)] can certainly provide valuable

SEE COMMENTARY ON PAGE e41

ISSN: 0006-3223

insights; however, data-driven approaches may be crucial for identifying more behaviorally refined biological phenotypes (15) to address the profound heterogeneity evident in health and illness (16,17). Fair *et al.* (18) and Karalunas *et al.* (19) recently demonstrated the potential value of delineating groups by similarity and dissimilarity of individual phenotypic profiles (e.g., neuropsychological profiles, temperament profiles). Adopting community detection methodologies from graph theory, they successfully identified 1) six distinct neuropsychological profiles that capture normal variation and are modified by attention-deficit/hyperactivity disorder (18) and 2) three temperamental phenotypes that showed intriguing biological differences as well as differential clinical outcomes (19).

Here, we used the Nathan Kline Institute-Rockland Sample (NKI-RS) (20), a deeply phenotyped, community-ascertained multimodal imaging sample. Using data from adult participants (18–59 years of age) in the NKI-RS, we aimed to identify data-driven phenotypes, based on core behavioral features representing several domains of function (including personality or temperament, symptom features, interpersonal functioning, and behavioral tendencies). Our first aim was to identify phenotypic dimensions that accurately represent meaningful variation across multiple domains of behavior. Accordingly, we conducted a bootstrap-based exploratory factor analysis (EFA) on 49 subscales derived from 10 measures obtained for 347 participants. The second aim was to identify a nested hierarchy of homogenous participant groups via hybrid hierarchical clustering (HHC) of participants, based on the factor profiles that we previously identified. To provide a phenotypic characterization of the participant groupings identified, we used DSM-IV labels and the Achenbach Adult Self-Report (ASR) (21), neither of which were included in the factor analysis. The third aim was to examine multivariate intrinsic brain functional connectivity differences among adjacent clusters and groups (derived from the first two aims).

METHODS AND MATERIALS

Participants

Participants were recruited as part of the NKI-RS (20), a community-based sample of approximately 1000 participants 6–85 years of age. To ascertain a cohort approximating a representative sample, exclusion criteria were minimal. Notably, comorbid medical conditions and medications (including psychotropics) were permitted. Written informed consent was obtained from all participants in accordance with local institutional review board oversight. The following inclusion criteria were applied: 1) 18–59 years of age; 2) absence of serious head injury or major neurological disorder; 3) negative history of bipolar disorder or psychosis; 4) negative drug test for commonly used illicit drugs with no therapeutic analogs or applications; and 5) at least 95% completion of each self-report measure examined.

Subject Phenotyping

All participants completed the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (22), and the Edinburgh Handedness Inventory (23), in addition to measures reflecting clinical symptom domains, personality or temperament, and broad behavioral characteristics (see Supplemental Table S1) (20). We selected subscales rather than individual items or full-scale scores to balance depth of assessment and amount of data per subject.

Magnetic Resonance Imaging Acquisition

Imaging data were collected on a 3T Siemens TIM Trio system (Siemens Healthcare GmbH, Erlangen, Germany) equipped with 32-channel head coil. Both structural and resting state functional magnetic resonance imaging (R-fMRI) data were acquired. The structural image was a T1-weighted magnetization prepared gradient echo sequence: repetition time = 1900 ms, echo time = 2.52 ms, flip angle = 9°, 176 slices, 1 mm^3 isotropic voxels. The R-fMRI data were acquired via multiband echo-planar imaging (24) with the following parameters: volumes = 900, repetition time = 645 ms, echo time = 30 ms, flip angle = 60°, 3 mm^3 isotropic voxels.

Phenotypic Analysis

Data Screening. All self-report data were checked for univariate and multivariate outliers. We also tested the assumption of missingness at random (25). Missing data were imputed using an expectation-maximization algorithm (26).

Dimension Reduction. For each participant, we created a multidimensional phenotypic profile using 49 subscale scores obtained from 10 questionnaires. Given modest intercorrelations among the questionnaires, we next performed an EFA to obtain a reduced set of dimensions. Analyses were done on age- and gender-regressed residuals of the 49 subscale scores to minimize the impact of these demographic variables on clustering (27). Parallel analysis (28) of 10,000 permutations of the raw data (29) was used to determine number of factors, comparing eigenvalues for each factor from the raw data with the 95th percentile of eigenvalues from the permutations. Maximum likelihood factor estimation with varimax rotation was used to estimate six factor loadings for each subscale score. Confidence intervals for factor loadings were estimated from 10,000 bootstrapped resamplings (30). To minimize factor intercorrelations, factor loadings overlapping zero (95% confidence interval) or exhibiting values less than |0.25| were set to zero in a restricted model. Factor scores (i.e., equivalent of latent value per factor) were computed using regression estimation (31).

Clustering Analysis. Categorical approaches (e.g., DSM) confer the ability to delineate unique combinations of above-threshold impairments in subsets of individuals. Building on this strength, a central goal of the present work is to identify phenotypically distinct groupings of participants among the larger sample (akin to categories), based on their dimensional six-factor phenotypic profiles. Specifically, we implemented HHC (32) using tree-structured vector quantization (33) to identify nested participant groups based on Euclidean distances between participant factor score profiles. HHC combines agglomerative and divisive clustering by 1) identifying mutual clusters (i.e., groups of data that are exceptionally close to one another and as a group, distant from all others)

via agglomerative clustering, 2) implementing constrained divisive clustering (retaining mutual clusters), and 3) applying additional divisive clustering, which explores the division of mutual clusters. In combination with visual examination of the dendrogram, we used the Calinski-Harabasz criterion (CHC) (34) at each cluster number to inform cut decisions (i.e., where to divide into subgroups).

Cluster Comparisons. Pairwise comparisons were made among adjacent clusters at each level using EFA, ASR (21), and SCID profiles. EFA scores were fully dimensional, whereas the ASR and SCID diagnoses were categorical. A categorical version of ASR scores was achieved by mimicking common clinical strategies for its use, which apply cut-scores to identify meaningful psychopathology. For each grouping of participants identified via cluster analysis, we calculated the percentage of individuals exhibiting standardized scores (*T* scores) greater than or equal to 60 within each ASR domain. The cutoff of 60 was chosen to increase sensitivity to subthreshold symptoms of potential relevance. EFA scores were compared using independent-sample *t* tests, whereas ASR proportions and psychiatric diagnoses were compared using chi-square tests. We also conducted *t* tests on continuous ASR scores, reported in the Supplemental Results.

MRI Data Processing

R-fMRI Preprocessing. Data were preprocessed using the Configurable Pipeline for Analysis of Connectomes (http://fcp-indi.github.io), which combines tools from AFNI (http://afni.nimh.nih.gov/afni), FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), and Advanced Normalization Tools (http://stnava.github.io/ANTs), using Nipype (35). Preprocessing included 1) motion correction, 2) mean-based intensity normalization, 3) nuisance signal regression, 4) temporal band-pass filtering (0.01–0.1 Hz), 5) coregistration of functional to structural images using boundary-based registration (36) using FSL's FLIRT (37), 6) normalization of functional image to Montreal Neurological Institute 152 template (38) by applying a nonlinear transform from Advanced Normalization Tools, and 7) smoothing with a full width at half maximum 6 mm Gaussian kernel. Nuisance regression removed linear and quadratic trends to account for scanner drift, 24 motion parameters, and five nuisance signals, identified via the component correction approach (CompCor) (39).

Multivariate Distance Matrix Regression. At each level of the hierarchy identified via HHC, we used multivariate distance matrix regression (MDMR) (40) to compare voxelwise functional connectivity profiles between adjacent phenotypic groups (e.g., cluster 1 [C1] vs. cluster 2 [C2], C1a vs. C1b, C2a vs. C2b).

MDMR was performed on a voxel-by-voxel basis. At each voxel, the following three steps were carried out. First, for each participant, Pearson's correlations were computed between the target voxel and all other voxels within a specified brain mask; this step generated, for each participant, a whole-brain functional connectivity map for the target voxel. Second, a between-participant distance matrix was computed, reflecting the distance between the connectivity maps obtained for the target voxel in two different participants. Distance was defined as $\sqrt{(2 * (1 - r))}$, where r is the spatial correlation of the connectivity maps obtained at the target voxel in two different individuals; the range of this distance metric is 0 (perfectly correlated) to 2 (perfectly negatively correlated), where 1 reflects no correlation. Importantly, these distances are calculated independent of any phenotypic relationships. Third, a pseudo-*F* statistic was computed to provide mathematical evaluation of the relationship between the variability in the distance matrix [computed in the second step (41)] and the variable of interest (i.e., group membership). Thus, the pseudo-*F* value at each voxel tells us whether the functional connectivity profiles for that voxel varied among individuals as a function of phenotypic group membership (e.g., C1 vs. C2, C1a vs. C1b, C2a vs. C2b).

MDMR was applied using the Connectir package in R (http://czarrar.github.io/connectir) on resampled, 4 mm^3 isotropic voxels. Computations were constrained to a study-specific group mask, including only voxels present across all participants and contained in a 25% probability gray matter Montreal Neurological Institute mask. The MDMR model (at each level of the dendrogram) specified cluster membership (categorical) and age, sex, hand laterality, and mean framewise displacement as covariates.

Consistent with prior work, voxelwise significance of the pseudo-*F* statistic was determined via estimation of the null distribution with random permutation ($n = 10{,}000$) for each cluster–community comparison (40). Recent work has raised concerns about potentially inflated type I error rates with random field theory (RFT) cluster thresholding approaches (42,43). These concerns are primarily applicable to parametric tests; nonparametric approaches are likely less prone to such inflations. However, we chose a more conservative thresholding approach than those used in prior studies implementing MDMR. Specifically, we corrected for multiple comparisons using cluster-based permutation ($n = 5000$) with a height threshold of Z greater than or equal to 2.33 ($p < .01$) and cluster extent probability of p less than .05.

Given that lower levels of the hierarchy had few participants per group, statistical power in these comparisons would be expected to be notably lower. For illustrative purposes, we repeated analyses using less stringent criteria (voxelwise $p < .05$; RFT-corrected $p < .05$) and included these in the Supplemental Results.

RESULTS

Demographic and diagnostic information is provided in Table 1. Details about data screening are provided in the Supplemental Results.

Dimension Reduction

Parallel analysis suggested seven factors (Supplemental Figure S1), though the seventh factor exhibited factor loadings whose bootstrap-based 95% confidence intervals overlapped zero. Thus a six-factor solution was estimated with maximum likelihood–based EFA, accounting for 77.8% of the variance. The constrained model (eliminating low-loading items) accounted for 65.4% of the total variance; standardized factor loadings are presented in Supplemental Table S3. Correlations between latent factor scores and regression estimates are

Table 1. Demographic and Diagnostic Characteristics

Characteristic	Full Sample (N = 347)
Age, Years	37.5 ± 13.6
Female	66.0 (229)
Right-Handed	91.9 (319)
Ethnicity	
Hispanic/Latino	13.8 (47)
Racial Background	
White/Caucasian	67.6 (230)
Black/African American	20.3 (69)
Asian	7.6 (26)
Native American/Pacific Islander	1.2 (4)
Other	3.2 (11)
Lifetime Psychiatric History[a]	
Any disorder	49.1 (167)
Depression	21.2 (72)
Anxiety[b]	9.4 (32)
Substance use disorder	26.2 (89)
ADHD	1.8 (6)
Current Psychiatric History[a]	
Any disorder	10.9 (37)
Depression	2.6 (9)
Anxiety[b]	4.7 (16)
Substance use disorder	4.1 (14)
ADHD	0.9 (3)
Number of Lifetime Psychiatry Diagnoses	
0	52.3 (178)
1	25.3 (86)
2	12.4 (42)
3 or more	10.0 (34)

Values are mean ± SD or % (n).
ADHD, attention-deficit/hyperactivity disorder.
[a]n = 340; diagnostic information missing for 7.
[b]Excluding obsessive-compulsive disorder and posttraumatic stress disorder.

displayed in Supplemental Table S4. Primary factor loadings and example items for each subscale are provided in Figure 1. The six retained factors were interpreted as follows (see Figure 1): 1) general distress and impairment, 2) conscientiousness, 3) sensation and risk seeking, 4) frustration intolerance, 5) contextual sensitivity, and 6) neuroticism and negative affect. Details about each of the factors are provided in the Supplemental Results.

Cluster Analysis

Visual inspection of the dendrogram suggested three clear cutpoints, yielding two, four, and eight groups, respectively. The largest CHC value was observed at k = 2 and stable subgroups (operationalized as values of k wherein CHC did not change appreciably from the prior solution [i.e., local minima]) at k = 4 and k = 8. Visual examination of participant-by-participant squared Euclidean distance matrices supported the face validity of these cutpoints (see Figure 2).

Phenotypic Cluster Differences

At all three levels of the dendrogram (i.e., two-, four-, and eight-cluster solutions), the participant groups differed from one another with respect to their phenotypic profiles. Due to space limitations, we limit reported findings beyond the first level (C1 vs. C2) to those along the C2 arm (which exhibited more psychopathology-like patterns). Phenotypic results along the C1 arm as well as for both groups together are provided in the Supplemental Results.

Level 1. Results at level 1 were robust, reflecting broad-reaching group differences that spanned nearly all domains. More specifically, C1 participants exhibited higher levels of adaptive functionality and C2 higher levels of maladaptive functionality; significant differences were noted in nearly all measures included in the three phenotypic profiles examined (EFA, ASR, psychiatric diagnosis [SCID]; see Figure 3, Supplemental Figure S2). To facilitate comparison, we also depicted phenotypic findings as heatmaps (and using a continuous score on the ASR) in Supplemental Figure S3.

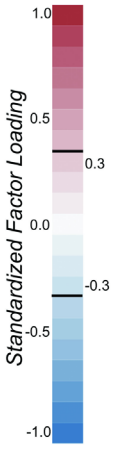
Level 2. The second level (k = 4) subdivided C1 (C1a and C1b) and C2 (C2a and C2b). C2 (functionally maladaptive group) was further divided into internalizing (C2a) and externalizing (C2b) problem characteristics. C2a had significantly higher ASR scores across all internalizing domains (Supplemental Figure S3). C2a also exhibited significantly higher rates of any lifetime psychiatric diagnosis and lifetime depression. C2b exhibited significantly higher levels of sensation and risk seeking on the EFA factor profile and significantly higher levels of ASR externalizing problems (see Figure 3).

Level 3. The third level (k = 8) divided the four clusters from level 2 into eight total subclusters (two clusters at level 3 for each cluster at level 2). Significant pairwise differences in ASR domains were largely a difference of magnitude (see Figure 3, Supplemental Figures S2, S3). Overall there were few significant pairwise differences between subclusters in DSM diagnoses, though notably C2a2 exhibited more current psychopathology than C2a1.

Multivariate Intrinsic Connectivity Differences Among Clusters

With permutation-based cluster correction, only MDMR findings from the first level (C1 vs. C2) survived multiple comparison correction (see Figure 4; see Supplemental Table S8 for functional peaks). This is not surprising given the larger number of participants in each group at the first level (C1-functionally adaptive: n = 165; C2-functionally maladaptive: n = 115), compared to the lower levels, which subdivided the sample into four and eight groups, respectively.

Three clusters were identified for the first-level comparison of C1 versus C2. The largest cluster (k = 133,248 mm^3, p = .0005) included the bilateral primary and secondary somatosensory cortices, as well as premotor, motor, and supplementary motor regions and was approximately centered on the midline near the supplementary motor area (X = 0, Y = –20, Z = 54); it also extended bilaterally to the lateral temporal



Factor	Subscale	Example Item
F1.General Distress & Impairment	TSC40 - Depression	[In the past month I have experienced]…sadness.
	TSC40 - Anxiety	[In the past month I have experienced]…feeling tense all the time.
	TSC40 - Dissociation	[In the past month I have experienced]… "spacing out".
	TSC40 - Sex Abuse	[In the past month I have experienced]… sexual problems.
	TSC40 - Sex Problems	[In the past month I have experienced]…low sex drive.
	TSC40 - Sleep Disturb	[In the past month I have experienced]…restless sleep.
	BDI - Somatic Sx's	I am too tired or fatigued to do a lot of the things I used to do.[a]
	BDI – Cognitive & Affective Sx's	I feel more worthless as compared to other people.[a]
	ATQ - Discomfort	I find loud noises to be very irritating.
F2.Conscientiousness	NEO - Conscientiousness	I am a productive person who always gets the job done.
	UPPS - Perseverance	Once I start a project, I almost always finish it. (R)
	ATQ - Activation Control	I hardly ever finish things on time. (R)
	CAARS - Inattention & Memory Problems	I have trouble getting started on a task. (R)
	UPPS - Premeditation	My thinking is usually careful and purposeful. (R)
	ATQ - Attentional Control	When I am trying to focus my attention, I am easily distracted. (R)
F3.Sensation and Risk Seeking	UPPS - Sensation Seeking	I sometimes like doing things that are a bit frightening.
	DOSP – Recreation Risk	[How likely are you to]… try bungee jumping?
	DOSP - Health & Safety Risk	How likely are you to drive home after you had three drinks in the last two hours?
	ATQ - High Intensity Pleasure	I would probably not enjoy a fast, wild carnival ride. (R)
	DOSP – Financial Risk	How likely are you to co-sign a new car loan for a friend?
	DOSP – Social Risk	How likely are you to speak your mind about an unpopular issue at a social occasion?
	DOSP – Ethical Risk	How likely are you to cheat on an exam?
F4.Frustration Intolerance	UPPS - Negative Urgency	When I feel rejected, I will often say things that I later regret.
	UPPS - Positive Urgency	I am always able to keep my feelings under control. (R)
	ATQ - Frustration	I seldom become irritated when someone is late. (R)
	ATQ - Inhibitory Control	It is easy for me to hold back my laughter in a situation where it is not appropriate. (R)
	CAARS – Impulsivity & Emotional Liability	My moods are unpredictable.
	IRI - Personal Distress	I tend to lose control during emergencies.
	ATQ - Fear	Loud noises sometimes scare me.
	CAARS – Hyperactivity & Restlessness	I'm bored easily.
F5. Contextual Sensitivity	NEO - Openness	I am intrigued by the patterns I find in art and nature.
	IRI - Empathic Concern	Other people's misfortunes do not usually disturb me a great deal. (R)
	ATQ - Affective Perceptual Sensitivity	I am often consciously aware of how the weather seems to affect my mood.
	IRI - Fantasy	After seeing a play or movie, I have felt as though I were one of the characters.
	IRI - Perspective Taking	I believe there are two sides to every question and try to look at them both.
	NEO - Agreeableness	Some people think of me as cold and calculating. (R)
	ATQ - Affective Sensitivity	When I am resting with my eyes closed, I sometimes see visual images.
	ICU - Uncaring	I am concerned about the feeling of others. (R)
	ATQ - Neutral Perceptual Sensitivity	I often notice visual details in the environment.
	ICU - Unemotional	I do not show my emotions to others. (R)
	ATQ - Sad	I rarely feel sad after saying good-bye to friends or relatives. (R)
	ICU - Callousness	I feel bad or guilty when I do something wrong. (R)
F6.Neuroticism & Negative Affect	STAI - Anxiety	Some unimportant thought runs through my mind and bothers me.
	NEO - Extraversion	I like to have a lot of people around me. (R)
	CAARS – Self-Concept Problems	My past failures make it hard for me to believe in myself.
	ATQ - Positive Affect	It doesn't take much to evoke a happy response in me. (R)
	NEO - Neuroticism	When I'm under a great deal of stress, sometimes I feel like I'm going to pieces.
	ATQ - Sociability	I usually enjoy being with people. (R)
	STAI - Depression	I am happy. (R)

Figure 1. Factors and their corresponding subscales (along with example items) identified by exploratory analysis with 10,000 bootstrap resamplings. Subscale names are provided in the middle column. Sample items for each subscale are provided in the far right column. Color bar on the left provides an index to the shading of each subscale relative to its standardized loading on the factor. ATQ, Adult Temperament Questionnaire; BDI, Beck Depression Inventory-Second Edition; CAARS, Conners' Adult ADHD Rating Scale; DOSP, Domain-Specific Risk-Taking Scale; ICU, Inventory of Callous and Unemotional Traits; IRI, Interpersonal Reactivity Index; NEO, NEO-Five Factor Inventory; (R), reverse-scored item in scale; STAI, Spielberger State-Trait Anxiety Inventory; TSC40, Trauma Symptom Checklist; UPPS, Impulsive Behavior Scale. [a]These sample items were selected from the response option corresponding to 2 on a 0–3 scale.

lobes. These regions have been implicated in bodily self-awareness (44) and interoception (45), and identified as comprising a major hub (46) and critical network (47) of the functional brain. Additionally, emerging evidence suggests an important role for the somatosensory–somatomotor hub in high prevalence psychopathology (48,49).

The second largest cluster (k = 27,323 mm^3, p = .0061) was approximately centered on the left thalamus (X = –16, Y = –20, Z = 14). In addition to a large thalamic contribution, the cluster included limbic regions (e.g., hippocampus, amygdala), decision-making regions (e.g., caudate, putamen), and various language (e.g., lingual gyrus) and vision regions (e.g., fusiform gyrus). It also extended from the left to right thalamus into the right caudate and putamen. The cluster comprised thalamic and basal ganglia regions commonly implicated in models of mental illness that emphasize thalamocortical and frontostriatal contributions (49,50).

Finally, the third cluster (k = 14,528 mm^3, p = .0167) was approximately centered on the right hippocampus (X = 32, Y = –24, Z = –18) and largely comprised right limbic regions (e.g., hippocampus and amygdala), as well as the caudate and putamen, fusiform gyrus, and middle and posterior insula. The regions implicated in this cluster, especially the right hippocampus and amygdala, are commonly associated with automated emotional processing (51), particularly of the kind related to high prevalence psychopathological alterations (52). Regions herein have also been identified as part of the medial temporal lobe subsystem of the default mode network (53).

See the Supplemental Results for MDMR findings obtained at the second and third levels of the dendrogram using a more liberal thresholding strategy.

DISCUSSION

Traditional psychiatric nosology comprises heterogeneous categories with few meaningful neurobiological correlates. The present findings illustrate the utility of data-driven approaches to 1) derive relevant phenotypic dimensions from diverse measures, 2) identify interpretable groups from

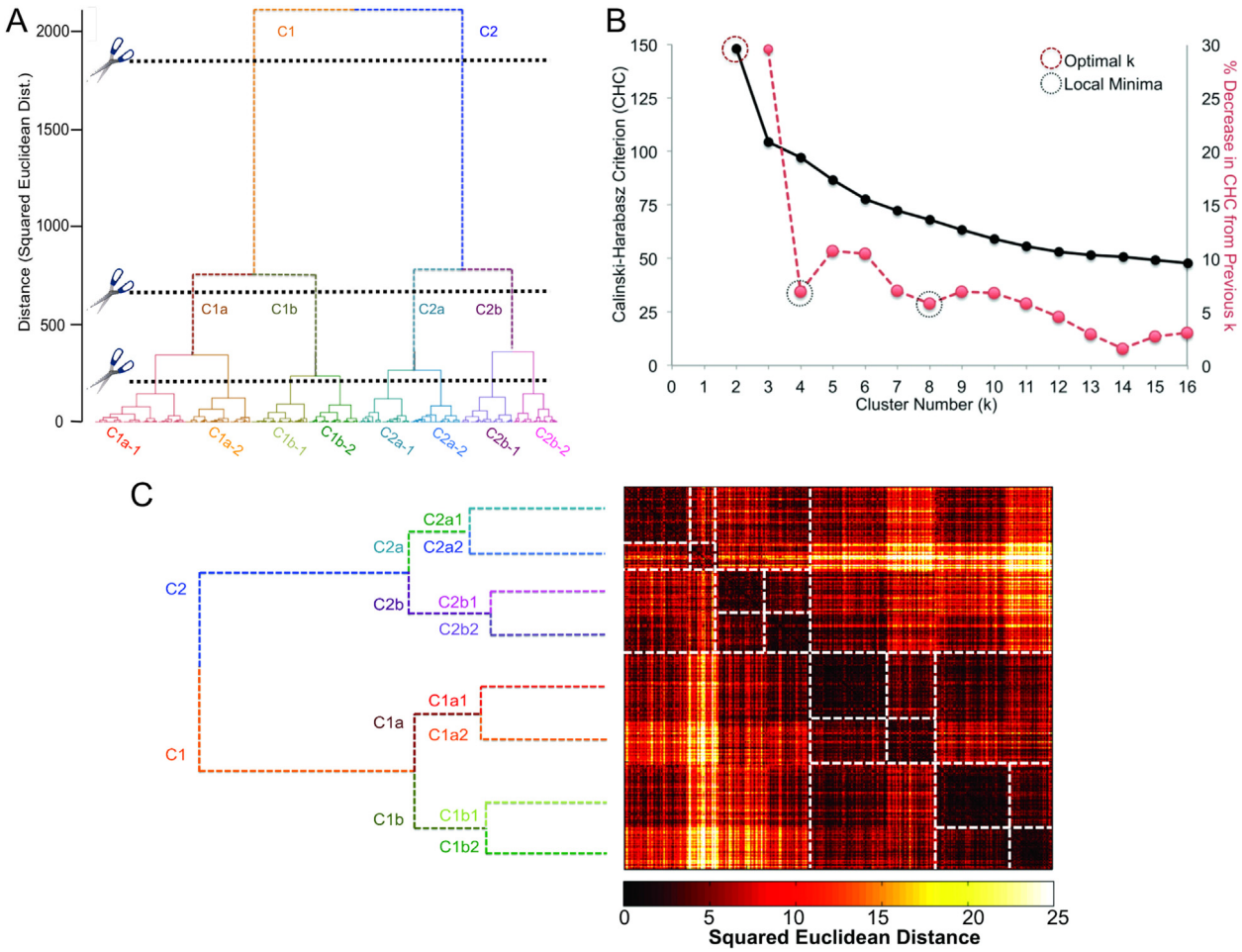


Figure 2. Results of hierarchical clustering. Panels in this figure depict the decision criteria used to ascertain clustering levels and resultant groups, as well as the similarity and dissimilarity of the groups as a function of correlation and squared Euclidean distance between factor scores (subject by subject). Panel **(A)** shows the dendrogram resulting from hybrid hierarchical clustering. It also shows the various levels at which the dendrogram was cut and the resultant groups. Panel **(B)** shows the Calinski-Harabasz criterion (CHC; black line), used as a decision aid for dendrogram cutting, as a function of cluster (C) number. The red line depicts percent change in the CHC value from one cluster to the next. Because the CHC did not exhibit a typical pattern (i.e., elevation at some cluster level), we defined stability (i.e., minimal change from one cluster number to the next) as our goal in deciding where to cut the dendrogram. Panel **(C)** again depicts the dendrogram, but relative to the squared Euclidean distance matrix. Groups and subgroups are outlined with dashed lines to help visualize group membership and increased similarity or decreased dissimilarity.

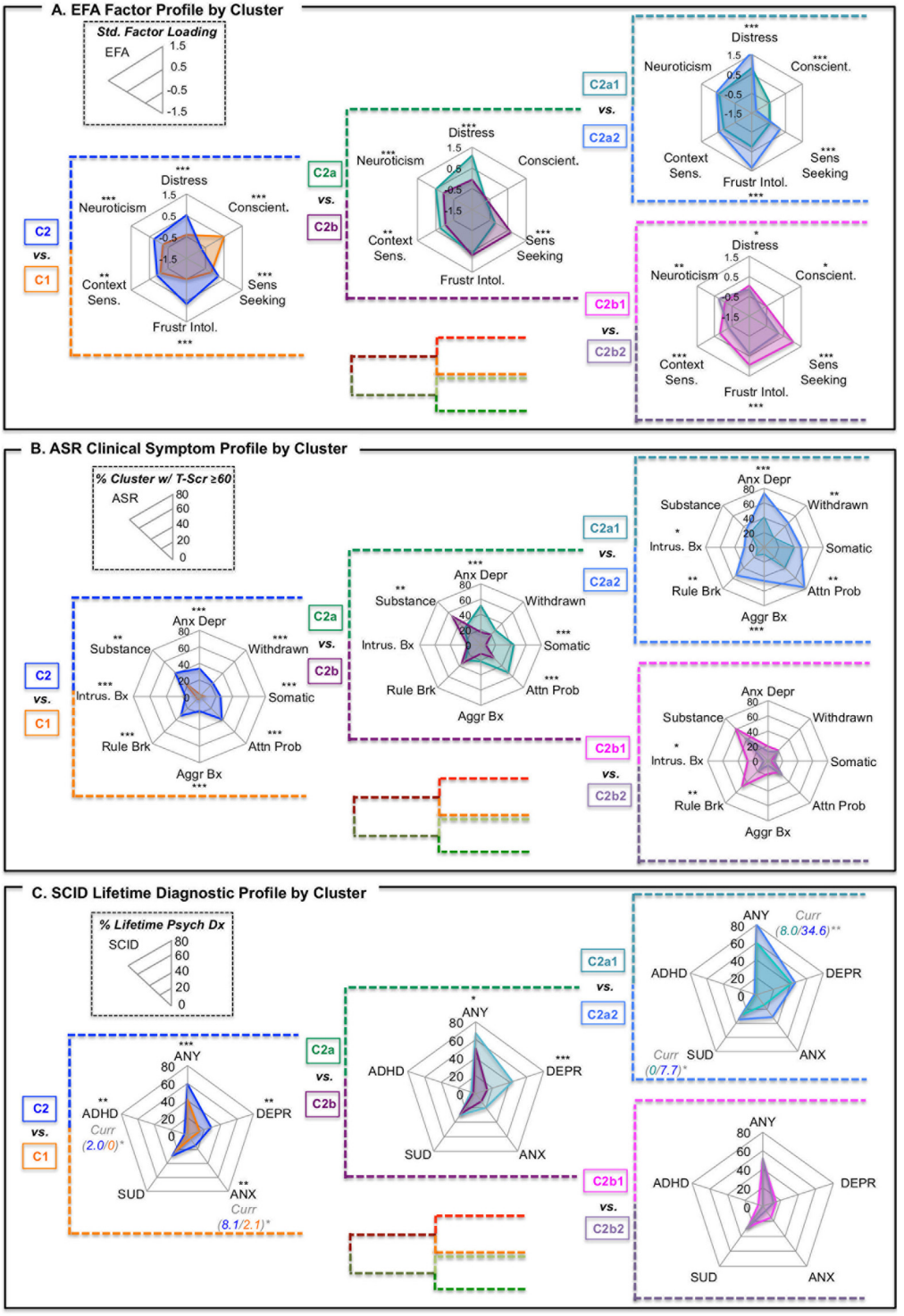
hierarchical clustering of dimensional variables, and 3) use nested groups to examine potential neurobiological differences.

Our findings suggest that inclusion of instruments with normal distributions (i.e., not truncated due to an assessment floor, such as an absence of symptoms [e.g., NEO Five Factor Inventory, Adult Temperament Questionnaire]) can be critical to defining relevant groups in population-based classification and among high-prevalence conditions (e.g., anxiety, depression, substance use). Although most individuals in the functionally adaptive group had truncated scores on syndrome-focused or problem-focused measures, our inclusion of bipolar scales allowed us to delineate further subgroups. The present work also highlights the value of including assessment tools that capture positive or protective factors rather than focusing solely on syndrome characteristics or problems. For example, conscientiousness, a personality characteristic associated with health and well-being (54), was a large contributor to differentiating groups at the highest and lowest levels of the nested hierarchical classification. Protective factors are rarely assessed in syndrome- or problem-focused assessments, though they can have important implications for presentation and prognosis among neuropsychiatric conditions (55).

Within the limitations of the current sample size, the present work demonstrated the value of pursuing nested subgroups within both the functionally maladaptive (C2) and functionally adaptive (C1) participant groupings. Of concern, it is possible that the groups merely recapitulated the theories on which some of the measures were predicated. It is important to note, however, that no single scale had its subscores distributed in a manner that could explain all components derived from the factor analysis (arguably, the NEO Five Factor Inventory came the closest). We observed novel combinations of subscales in the factors derived, reflecting a range of

psychiatric symptoms, personality features, and protective or risk factors.

The present data also show the value of using broad behavioral characteristics for group identification to examine potential neurobiological differences. Multivariate comparisons of intrinsic connectivity were implemented at three different clustering levels, though the findings only passed stringent multiple comparisons correction at the first level, where power was the largest. The connectome differences observed between the functionally maladaptive and adaptive groups (C1 vs. C2) at the first level were evident within the 1) somatomotor network, 2) thalamic and basal



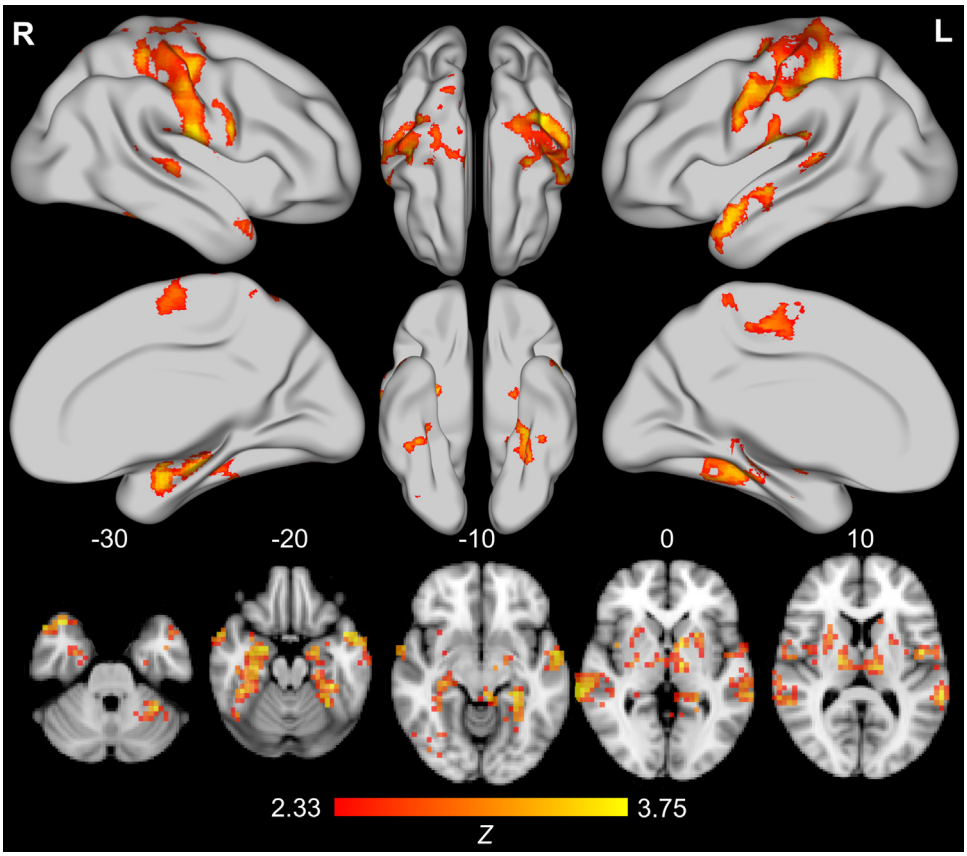


Figure 4. Results from multivariate distance matrix analysis of the functional connectome between cluster 1 (C1) and cluster 2 (C2). Adjacent groups at the highest level of hierarchical clustering (level 1: n_{C1} = 165, n_{C2} = 115) are displayed. Rendered brains and axial slices reflect multivariate distance matrix regression comparing intrinsic connectivity between groups; findings represent conversion of pseudo-*F* test results to *Z* values via permutation testing (10,000 resamplings of data) and permutation-based cluster correction (5000 resamplings of data) with cluster formation set at *p* less than .01 and extent threshold set at *p* less than .05. Note that images are presented in neurological convention (left [L] = right [R], R = L). The bottom row of axial slices represents significant findings at Montreal Neurological Institute axis *Z* values of –30, –20, –10, 0, and 10. From left to right, top to bottom, the top two thirds of the figure depict the lateral and medial surface of the right hemisphere, the dorsal and ventral surface of the right, then left, hemisphere, and finally, the dorsal and medial surface of the left hemisphere. Only level 1 results survived cluster permutation testing correction for multiple comparisons.

ganglia regions, and 3) amygdala and extended hippocampal complex.

The somatomotor network is a somewhat novel functional target in the context of identifying potential imaging biomarkers of the tendency toward psychiatric illness. Although known as a key hub (46,47) of the functional brain, the potential role in psychopathology of the somatomotor network is only beginning to be recognized (48,49). The somatosensory–motor network is intimately involved in bodily self-consciousness and interoception (44,45), processes that are increasingly implicated in predictive outcome models (56), especially for neuropsychiatric illness (57). At a coarse level of group differentiation (e.g., functionally maladaptive vs. functionally adaptive), these results underscore the potential importance of subjective valuation and bodily states in making interactional predictions that may be fundamentally altered as part of the pathophysiology of psychiatric illness.

Somewhat less novel, connectomic alterations in the thalamus and basal ganglia, as well as the amygdala and extended hippocampal complex, underscore the importance of basic cognitive functions (e.g., attention, working memory), as well as reward and emotion or saliency in delineating adaptive from maladaptive function. This latter finding lends support to the RDoC-style approach to examining domains rather than syndromes or diseases. On the whole, the intrinsic connectivity findings indicated potential new targets associated with adaptive and maladaptive function, while affirming existing targets, which suggests neurobiological validity of our data-driven group assignments.

Figure 3. Factor, clinical symptom, and lifetime psychiatric profiles visualized as radar plots by cluster (C) or group at three levels of hierarchical clustering, showing expansion within C2. All panels in this figure represent different measures pertaining to the same clusters [i.e., C2 is the same group of individuals, showing **(A)** variation in factor profiles, **(B)** Achenbach Adult Self-Report (ASR) clinical symptom profiles, and **(C)** lifetime psychiatric diagnosis]. Panel **(A)** represents mean values by cluster for each of the six factors from the exploratory factor analysis (EFA). Plots represent a standard loading of –1.5 at the origin and 1.5 at the maximum for each of the six factors. Panel **(B)** represents percent of individuals within a cluster exhibiting *T* scores ≥60 (1 SD above the mean; approaching clinical importance) for eight domains from the ASR. Plots represent zero at the center and 25% at the periphery (unless otherwise denoted) for each of the eight domains. Panel **(C)** represents percent of individuals within a cluster exhibiting a lifetime (i.e., past or current) psychiatric diagnosis (any diagnosis [ANY]; depressive disorder [DEPR]; anxiety disorder (ANX), excluding obsessive-compulsive disorder and posttraumatic stress disorder; substance use disorder [SUD]; attention-deficit/hyperactivity disorder [ADHD]). Where significant differences in current psychiatric diagnosis were observed, group percentages and significance are represented next to the diagnosis in italics. Plots represent zero at the center and 60% at the periphery (unless otherwise denoted) for each of the five diagnostic categories. Note that diagnoses are not mutually exclusive. Significant group differences are represented by asterisks: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Aggr Bx, aggressive behavior; Anx Depr, anxiety and depression; Atten Prob, attention problems; Consient, Conscientiousness; Context Sens., contextual sensitivity; Dx, diagnosis; Frust Intol, frustration intolerance; Intrus Bx, intrusive behavior; Rule Brk, rule breaking behavior; SCID, Structured Clinical Interview for DSM-IV Axis I Disorders; Sens Seeking, sensation seeking.

Limitations

A number of steps were taken to protect against over- or underfitting the number of factors and provide robust estimates of which subscales loaded on which dimension. Factor number was determined via parallel analysis, factor structure was based on bootstrapping the raw data, and bootstrap-based confidence intervals were computed for estimates of standardized factor loadings. Despite all this, a separate confirmatory sample will be necessary to validate the replicability of the factor structure and the groups identified.

The different response formats of the questionnaires could potentially have led to clustering of items by comparable response format and hence, recapitulation of the original scale factor structure. There was little evidence of such a result, though. To overcome clustering as a function of similar response formats, all questions would have to be administered with the same response format. However, such an approach would require extremely large samples and known psychometric properties of the scales would be forfeited.

Significant connectome-wide differences decreased dramatically beyond the first level of hierarchical classification. Results at lower levels were only significant with less stringent, RFT multiple comparisons correction (see the Supplement). There are at least three plausible reasons for this decrease in significant findings: sample size, group homogeneity, and demographic differences. Average sample size was halved for each incremental level of the hierarchy. Thus, one contributor may merely be decreased power. Another potential factor is increasing subgroup homogeneity: group factor profiles became progressively more similar to one another at lower levels of the hierarchy. Demographic differences may also contribute to smaller connectome-wide differences at more refined levels of subgroup detection; groups may become more demographically similar as they become more behaviorally homogenous. However, cluster analyses were conducted on age and gender regression-residuals, mitigating some concerns regarding the influence of these demographic variables.

Although biomarker differences may diminish as subgroups become more similar, it is exactly these kinds of comparisons that will ultimately permit differential classification based on differences in neurobiology (58). Larger samples within specific diagnostic categories or problem domains will likely provide more power to detect differences among more similar groups (18,19,59) and can also remedy some of the problems that arise in exploratory analyses requiring complex multiple comparisons correction. By combining population-based, data-driven categorization and diagnostically focused pattern assessments, we can begin to compare the value of different classification methods. Currently, diagnostic heterogeneity is so marked that only extremely large samples (59) or time-consuming separation into individual diagnostic criteria permit head-to-head comparison of both methods (60).

Finally, although the present work focused on phenotypic information alone for group classification or detection, alternative approaches might permit classification based entirely on neurobiology or on a combination of neurobiology and behavioral features. We opted to avoid classification purely on neurobiology as such approaches are complex (61), and interpretation of behavioral characteristics can be tricky when groups are clustered entirely on neurobiology. A potentially valuable alternative is analytic approaches that simultaneously consider phenotypic and neurobiological information (62).

Conclusions

Psychiatry is limited by a lack of validated biobehavioral tests and extensive heterogeneity within diagnostic categories. In addition to consensus-based approaches to reframing the current nosology (e.g., RDoC), data-driven approaches to delineating homogenous subgroups, spanning adaptive to maladaptive function can yield clinically meaningful groups with potentially important neurobiological differences. In doing so, it will be important to consider not only symptoms of psychiatric illness (i.e., deficits or problems), but also features of psychiatric health (i.e., strengths or protective factors). Examination of a broad array of phenotypic characteristics, in combination with neurobiological differences, may improve our understanding of the pathophysiology of mental illness and provide new preventative and treatment strategies. Although these analyses will require large samples and advanced analytic approaches, the present study is evidence that such efforts can yield new hypotheses, as well as support existing theories, helping to focus biological psychiatry on those areas that may yield the highest return on investment.

ACKNOWLEDGMENTS AND DISCLOSURES

This study was supported by National Institute of Mental Health Grant Nos. R01MH094639 (to MPM), R01MH081218 (to FXC), R01MH083246 (to FXC), and R21MH084126, with additional funding for personnel and administrative support provided by a grant from the Child Mind Institute (1FDN2012-1 to MPM) and funds from the New York State Office of Mental Health, the Research Foundation for Mental Hygiene, the Brain Research Foundation, and the Stavros Niarchos Foundation. Additional funds were provided by gifts from Phyllis Green, Randolph Cowen, and Joseph P. Healey.

All data are publicly available via the International Neuroimaging Data-Sharing Initiative. Data can be accessed at http://fcon_1000.projects.nitrc.org. A data usage agreement is required for access to detailed phenotypic information.

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

ARTICLE INFORMATION

From the Center for the Developing Brain (NTVD, DO, ETM, EJH, RCC, MPM), Child Mind Institute; Departments of Psychiatry and Neuroscience (VG), Icahn School of Medicine at Mount Sinai; The Child Study Center at NYU Langone Medical Center (FXC), New York; Center for Biomedical Imaging and Neuromodulation (NTVD, DO, RCC, RHT, FXC, BLL, MPM), Nathan Kline Institute, Orangeburg, New York; Department of Psychiatry (BLL), University of California, San Francisco, San Francisco, California; and Department of Psychiatry (JJH), University of Vermont, Burlington, Vermont.

NTVD is currently affiliated with the Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York. ETM is currently affiliated with the Department of Psychology, University of California–Berkeley, Berkeley, California.

Address correspondence to Michael P. Milham, M.D., Ph.D., Center for the Developing Brain, Child Mind Institute, 445 Park Avenue, New York, NY 10022; E-mail: michael.milham@childmind.org.

Received Jan 5, 2016; revised Jun 29, 2016; accepted Jun 30, 2016.

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.biopsych.2016.06.027.

REFERENCES

1. Wakefield JC (2010): Misdiagnosing normality: Psychiatry's failure to address the problem of false positive diagnoses of mental disorder in a changing professional environment. J Ment Health 19:337–351.
2. Cuthbert B, Insel T (2013): Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Med 11:126.
3. American Psychiatric Association. (2013): Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC: American Psychiatric Press.
4. Hyman SE (2010): The diagnosis of mental disorders: The problem of reification. Annu Rev Clin Psychol 6:155–179.
5. Insel TR (2014): The NIMH Research Domain Criteria (RDoC) Project: Precision medicine for psychiatry. Am J Psychiatry 171:395–397.
6. Casey BJ, Craddock N, Cuthbert BN, Hyman SE, Lee FS, Ressler KJ (2013): DSM-5 and RDoC: Progress in psychiatry research? Nat Rev Neurosci 14:810–814.
7. Kraemer HC, Noda A, O'Hara R (2004): Categorical versus dimensional approaches to diagnosis: methodological challenges. J Psychiatr Res 38:17–25.
8. Widiger TA, Samuel DB (2005): Diagnostic categories or dimensions? A question for the Diagnostic And Statistical Manual Of Mental Disorders–fifth edition. J Abnorm Psychol 114:494–504.
9. Yee CM, Javitt DC, Miller GA (2015): Replacing dsm categorical analyses with dimensional analyses in psychiatry research: The research domain criteria initiative. JAMA Psychiatry 72:1159–1160.
10. First MB (2005): Clinical utility: A prerequisite for the adoption of a dimensional approach in DSM. J Abnorm Psychol 114:560–564.
11. Weinberger DR, Glick ID, Klein DF (2015): Whither Research Domain Criteria (RDoC)?: The good, the bad, and the ugly. JAMA Psychiatry 72:1161–1162.
12. Biswal BB, Mennes M, Zuo X-N, Gohel S, Kelly C, Smith SM, *et al*. (2010): Toward discovery science of human brain function. Proc Natl Acad Sci U S A 107:4734–4739.
13. Akil H, Martone ME, Van Essen DC (2011): Challenges and Opportunities in Mining Neuroscience Data. Science 331:708–712.
14. Kraemer H (2015): REsearch Domain Criteria (RDoC) and the dsm–Two methodological approaches to mental health diagnosis. JAMA Psychiatry 72:1163–1164.
15. Insel TR, Cuthbert BN (2015): Brain disorders? Precisely. Science 348: 499–500.
16. McGorry P, Keshavan M, Goldstone S, Amminger P, Allott K, Berk M, *et al*. (2014): Biomarkers and clinical staging in psychiatry. World Psychiatry 13:211–223.
17. Wardenaar KJ, de Jonge P (2013): Diagnostic heterogeneity in psychiatry: Toward an empirical solution. BMC Med 11:201.
18. Fair DA, Bathula D, Nikolas MA, Nigg JT (2012): Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. Proc Natl Acad Sci U S A 109:6769–6774.
19. Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, Nigg JT (2014): Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: Toward biologically based nosologic criteria. JAMA Psychiatry 71:1015–1024.
20. Nooner KB, Colcombe S, Tobe R, Mennes M, Benedict M, Moreno A, *et al*. (2012): The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. Front Neurosci 6:152.
21. Achenbach TM, Rescorla LA (2003): Manual for the ASEBA Adult Forms & Profiles. Burlington, VT: Research Center for Children, Youth, & Familes, University of Vermont.
22. First MB, Spitzer RL, Gibbon M, Williams JBW (1996): Structured Clinical Interview for DSM-IV Axis I Disorders. New York, NY: Biometrics Research, New York State Psychiatric Institute.
23. Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.
24. Xu J, Moeller S, Auerbach EJ, Strupp J, Smith SM, Feinberg DA, *et al*. (2013): Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage 83:991–1001.
25. Potthoff RF, Tudor GE, Pieper KS, Hasselblad V (2006): Can one assess whether missing data are missing at random in medical studies? Stat Methods Med Res 15:213–234.
26. Gold MS, Bentler PM (2000): Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. Struct Equat Model 7:319–355.
27. Hudziak JJ, Achenbach TM, Althoff RR, Pine DS (2007): A dimensional approach to developmental psychopathology. Int J Methods Psychiatr Res 16(suppl 1):S16–S23.
28. Horn JL (1965): A rationale and test for the number of factors in factor analysis. Psychometrika 30:179–185.
29. Castellan NJ (1992): Shuffling arrays: Appearances may be deceiving. Behav Res Methods Instrum Comput 24:72–77.
30. Zientek L, Thompson B (2007): Applying the bootstrap to the multivariate case: Bootstrap component/factor analysis. Behav Res Methods 39:318–325.
31. Grice JW (2001): Computing and evaluating factor scores. Psychol Methods 6:430–450.
32. Chipman H, Tibshirani R (2006): Hybrid hierarchical clustering with applications to microarray data. Biostatistics 7:286–301.
33. Gersho A, Gray RM (1992): In: Vector Quantization and Signal Compression. Dordrecht, the Netherlands: Kluwer Academic.
34. Caliński T, Harabasz J (1974): A dendrite method for cluster analysis. Commun. Stat 3:1–27.
35. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, *et al*. (2011): Nipype: A flexible, lightweight and extensible neuroimaging data processing framework. Front Neuroinform 5:13.
36. Greve DN, Fischl B (2009): Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48:63–72.
37. Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.
38. Avants BB, Epstein CL, Grossman M, Gee JC (2008): Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12:26–41.
39. Behzadi Y, Restom K, Liau J, Liu TT (2007): A Component Based Noise Correction Method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37:90–101.
40. Shehzad A, Kelly C, Reiss PT, Craddock RC, Emerson JW, McMahon K, *et al*. (2014): A multivariate distance-based analytic framework for connectome-wide association studies. Neuroimage 93:74–94.
41. Reiss PT, Stevens MHH, Shehzad Z, Petkova E, Milham MP (2010): On distance-based permutation tests for between-group comparisons. Biometrics 66:636–643.
42. Woo C-W, Krishnan A, Wager TD (2014): Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. Neuroimage 91:412–419.
43. Eklund A, Nichols T, Knutsson H (2016): Can parametric statistical methods be trusted for fMRI based group studies? Proc Natl Acad Sci U S A 113:7900–7905.
44. Blanke O (2012): Multisensory brain mechanisms of bodily self-consciousness. Nat Rev Neurosci 13:556–571.
45. Craig AD (2002): How do you feel? Interoception: The sense of the physiological condition of the body. Nat Rev Neurosci 3:655–666.
46. Tomasi D, Shokri-Kojori E, Volkow ND (2016): High-resolution functional connectivity density: Hub locations, sensitivity, specificity, reproducibility, and reliability. Cereb Cortex 26:3249–3259.
47. Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, *et al*. (2009): Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A 106:13040–13045.
48. Gabbay V, Ely BA, Li Q, Bangaru SD, Panzer AM, Alonso CM, *et al*. (2013): Striatum-based circuitry of adolescent depression and anhedonia. J Am Acad Child Adolesc Psychiatry 52(628–641):e613.
49. Cui H, Zhang J, Liu Y, Li Q, Li H, Zhang L, *et al*. (2016): Differential alterations of resting-state functional connectivity in generalized anxiety disorder and panic disorder. Hum Brain Mapp 37: 1459–1473.

50. Buckholtz Joshua W, Meyer-Lindenberg A (2012): Psychopathology and the human connectome: Toward a transdiagnostic model of risk for mental illness. Neuron 74:990–1004.
51. Dyck M, Loughead J, Kellermann T, Boers F, Gur RC, Mathiak K (2011): Cognitive versus automatic mechanisms of mood induction differentially activate left and right amygdala. Neuroimage 54: 2503–2513.
52. Fairhall SL, Sharma S, Magnusson J, Murphy B (2010): Memory related dysregulation of hippocampal function in major depressive disorder. Biol Psychol 85:499–503.
53. Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL (2010): Functional-anatomic fractionation of the brain's default network. Neuron 65:550–562.
54. Bogg T, Roberts BW (2004): Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. Psychol Bull 130:887–919.
55. Belcher AM, Volkow ND, Moeller FG, Ferré S (2014): Personality traits and vulnerability or resilience to substance use disorders. Trends Cogn Sci 18:211–217.
56. Barrett LF, Simmons WK (2015): Interoceptive predictions in the brain. Nat Rev Neurosci 16:419–429.
57. Seth AK (2013): Interoceptive inference, emotion, and the embodied self. Trends Cogn Sci 17:565–573.
58. Kapur S, Phillips AG, Insel TR (2012): Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 17:1174–1179.
59. 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.
60. Borsboom D, Cramer AOJ (2013): Network analysis: An integrative approach to the structure of psychopathology. Annu Rev Clin Psychol 9:91–121.
61. Yang Z, Xu Y, Xu T, Hoy CW, Handwerker DA, Chen G, *et al*. (2014): Brain network informed subject community detection in early-onset schizophrenia. Sci Rep 4:5549.
62. Chen G, Sullivan PF, Kosorok MR (2013): Biclustering with heterogeneous variance. Proc Natl Acad Sci U S A 110:12253–12258.