



HHS Public Access

Author manuscript

Biol Psychiatry. Author manuscript; available in PMC 2021 August 30.

Published in final edited form as:

Biol Psychiatry. 2020 July 01; 88(1): 9–17. doi:10.1016/j.biopsych.2020.02.015.

Methods and Challenges for Assessing Heterogeneity

Eric Feczko,

Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, Oregon.

Damien A. Fair

Advanced Imaging Research Center, Oregon Health & Science University, Portland, Oregon.

Abstract

The widely acknowledged homogeneity assumption limits progress in refining clinical diagnosis, understanding mechanisms, and developing new treatments for mental health disorders. This homogeneity assumption drives both a comorbidity and a heterogeneity problem, where two different approaches tackle the problems. One, a unifying approach, tackles the comorbidity problem by assuming that a single general psychopathology factor underlies multiple disorders. Another, a multifactorial approach, tackles the heterogeneity problem by assuming that disorders comprise multiple subtypes driven by multiple discrete factors. We show how each of these approaches can make useful contributions to mental health–related research and clinical practice. For example, the unifying approach can develop a rapid assessment tool that may be clinically valuable for triaging cases. The multifactorial approach can reveal subtypes that are differentially responsive to treatments and highlight distinct mechanisms leading to similar phenotypes. Because both approaches tackle different problems, both have different limitations. We describe the statistical frameworks that incorporate and adjudicate between both approaches (e.g., the bifactor model, normative modeling, and the functional random forest). Such frameworks can identify whether sets of disorders are more affected by heterogeneity or comorbidity. Therefore, future studies that incorporate such frameworks can provide further insight into the nature of psychopathology.

Keywords

Bifactor; modeling; Comorbidity; Functional; random; forest; Heterogeneity; Normative modeling; Psychopathology

Despite efforts over the past century (1–3), most psycho- pathological processes and biological links to most mental health disorders remain unknown. Convergent evidence demonstrates instability in mental health disorder categoriza- tion. Biological links such as single nucleotide polymorphisms (4,5), brain structure (6,7), or functional connectivity (8,9) to specific disorders are limited. Different disorders may occur sequentially. For example, diagnosis of generalized anxiety or major depression increases the risk of receiving

Address correspondence to Eric Feczko, M.B.I., Ph.D., Oregon Health & Science University, Department of Behavioral Neuroscience, 3181 SW Sam Jackson Park Road, Portland, OR 97239-3098, feczko@ohsu.edu.

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

the other diagnosis in the future (10–12). Profound comorbidity exists across diagnostic labels thought to separate internalizing and externalizing disorders; such studies reveal high phenotypic correlations between diagnosis pairs (13,14). Furthermore, a common source may drive such comorbidity; for example, the same genetic risk factors can underlie many diagnoses (4,15,16)—from autism spectrum disorder (ASD) to schizophrenia. These issues show that the homogeneity assumption is unfounded; disorders are not homogeneous discrete entities. However, most studies still treat typical and diagnosed cohorts as homogeneous discrete entities. From the homogeneity assumption, the comorbidity (17) and heterogeneity (18,19) problems emerge and limit psychiatric research (20–26).

THE COMORBIDITY AND HETEROGENEITY PROBLEMS EMERGE FROM THE HOMOGENEITY ASSUMPTION

The comorbidity problem assumes that the same mechanism may underlie two different diagnoses. Unifying approaches attempt to reveal such mechanisms (Figure 1A). Conceptually, one models these common mechanisms as a single factor, a general psychopathology (p) [see (27)]. For example, one might develop a model of p from genetic markers by aggregating genetic data across many mental health conditions and mapping the genetic data to a single factor (28).

The heterogeneity problem comprises two tenets (18). First, different underlying mechanisms may drive the same disorder. Second, factors important for mental health outcomes may be unrelated to such mechanisms, depending on the outcome of interest. The multifactorial approach (Figure 1B) tackles heterogeneity by assuming that psychopathology comprises discrete conditions driven by different factors (29). For example, one might measure mental health across dimensional constructs and model these constructs from neuroimaging data to test their validity (30,31).

These two approaches account for many of the current methods in use to answer questions revolving around the heterogeneity problem, such as bifactor models (Figure 2; discussed below). While it seems like the approaches tackle different problems, it is often unclear whether comorbidity, heterogeneity, or both problems may be present across a subset of mental health disorders. In the context of any mental health study, both approaches are critical and prompt consideration.

The Unifying Approach Addresses the Comorbidity Problem

As noted, several lines of evidence drive the assumption of a p factor (27,32) underlying multiple disorders. Consistent with the idea of a common mechanism, many individuals are diagnosed with different disorders across the life span, while profound comorbidity is present across multiple diagnoses. Taken together, such evidence might suggest that a p factor may drive multiple disorders or dimensional traits.

While prior studies posited a p factor, direct evidence for common mechanisms comes from recent large longitudinal cohorts in adults (National Epidemiologic Study of Alcohol and Related Conditions [NESARC]) (33) and children (Dunedin Multidisciplinary Health and Development Study) (32). Such cohorts enable test-retest reliability and help further

validate findings. From over 30,000 individuals in NESARC (33), investigators compared 2- and 3-factor nested models (Figure 2B) to an alternative bifactor model (Figure 2A), where a single nonspecific factor was modeled as affecting every disorder plus the 3 factors. In these bifactor models, no relationships were modeled between the specific factors to make the model comparable with the others. The investigators show that this bifactor model better captured diagnoses than the 2- or 3-factor uncorrelated model. However, as noted by the investigators (33), this was a preliminary study. The clinical assessments were messy and incomplete; the data were also limited to two time points yet captured an age range of 18 to 64 years; the inputs from the assessments were dichotomous diagnoses and not continuous symptom counts; and the true relationships between all sub-factors may account for the p factor.

Independent investigators addressed these questions with the Dunedin study (32). The Dunedin study examined 1037 children longitudinally as the children developed through adulthood. The study allowed investigators to model multiple ages in a more narrow range, from 18 to 38 years, and include symptom counts from in-depth structured interviews. Critically, the investigators also examined whether the p factor improved model fits when the subfactors internalizing, externalizing, and thought disorder were allowed to correlate with one another (Figure 2A, gray dotted line). The investigators found that inclusion of the p factor fit the data better than exclusion, suggesting a common mechanism.

General p Factor Is Associated With Biological and Behavioral Dimensions Important for Mental Health.—As the investigators (32,33) note, their studies cannot determine the p factor's biological underpinnings. Further studies measured the association between the p factor and brain and genetic variation. Studies of the Philadelphia Neurodevelopmental Cohort led the charge in identifying the p factor's brain associations (34–36). Using cross-sectional samples of over 1000 children ages 11 to 23 years, investigators found that increased p showed altered perfusion of the dorsal anterior cingulate cortex (34), reduced gray matter volume across the brain (35), and reduced brain activation in task control systems when engaged in a working memory task (36). While these exciting, landmark studies suggest a p factor's biological underpinnings, future studies that investigate brain-psychopathology associations will enable investigators to derive stronger inferences.

Similarly, few studies examined genetic associations with p (28,37–39). Early work using 1569 twin pairs from the Tennessee Twin Study assessed the associations among the p factor, neuroticism, and its heritability (39). Replicating the NESARC, bifactor model fits improved over correlated factor model fits. Additionally, neuroticism correlated more with the general bifactor than specific factors such as externalizing or internalizing. Furthermore, the twin analysis revealed that this association was partly heritable, suggesting the general bifactor's biological and structural basis. A subsequent study using almost 2000 children from the Generation R Study found further p factor biological and cognitive links (28). Unlike previously mentioned studies, the investigators incorporated commonly used, child developmental, parent and teacher rating scales and tested whether the observed general bifactor reflected cognitive and/or temperament measures. The general bifactor associated

with IQ and negative affectivity and was partly heritable. Taken together, both studies provide consistent evidence of a general bifactor's biological associations.

Clinical Rapid Assessment Tools May Emerge From Investigations Into the p Factor.—Relative to a p factor's many biological underpinnings, a formal quantitative assessment of a p factor shows clinical utility. In situations with limited resources for mental health evaluation, an inexpensive and effective system for triaging patients may reduce costs and improve outcomes. An analogous example would be the Mini-Mental State Exam, a 15-minute assessment tool. The Mini-Mental State Exam screens for elderly general impairment without measuring specific cognitive functions (40). Recognizing a general assessment tool's clinical utility, investigators calibrated and validated a p rapid screening assessment (41). Using computerized structured interview data, investigators developed a bifactor model for the interview using a 4500-child training dataset. Using the weights from the bifactor model, bifactor scores were simulated from 1000 new children, and investigators compared their clinical diagnoses and scores via a receiver operating characteristic curve. The area under the curve was 0.76, suggesting some generalizability for the instrument. Furthermore, the p factor showed a large association with trauma exposure ($r = .3$), but only subtle associations with brain structure, indicating reliability for the first-pass tool (41).

Analogically, blood pressure also highlights the strengths and limitations of p factors. Elevated blood pressure remains an excellent clinical sign in many cardiovascular diseases. As with p factors, there are many ways to use high blood pressure for clinical management of many distinct cardiovascular and other health events. Indeed, reducing overall blood pressure can reduce the risk of many future health problems. However, despite blood pressure's association with many unique diseases, by itself it does not provide information regarding the mechanisms of any given disorder. For example, while high blood pressure is correlated with poorer general health and related to diabetes (42), kidney disease (43), psychosocial stress (44), poor nutrition (45), lack of exercise (46), and genetic predispositions (47), it alone cannot provide information identifying specific underlying issues.

The Multifactorial Approach Addresses the Heterogeneity Problem

Alternatively, one can assume that diagnoses and/or phenotypes can be driven by multiple distinct factors independent of a p factor. By extension, one must also assume that factor misspecifications due to our current knowledge drive the heterogeneity problem. Such assumptions may enable the discovery of subtypes from dimensional or categorical outcomes not found through a unifying approach.

Subtypes Can Be Categorical or Dimensional.—Defining the borders of different clinical subtypes is a broad complex field and beyond this review's scope. One can reduce mental health subtype research into categorical or dimensional subtype studies. Such reductionism enables researcher consortiums to derive inferences that may be more reproducible across studies. Not all approaches produce categorical and dimensional subtypes. For example, the Research Domain Criteria consortium used expert opinion to

define a priori dimensional traits. They then validate the traits by testing whether collected data fit these traits and how the traits are linked to biological types (48). Consortia such as the Hierarchical Taxonomy of Psychopathology identify categorical hierarchically organized levels of diagnosis, which can represent distinct clinical subtypes (30). These subtypes are further validated using independent samples, and biological links drive future insights. Studies identifying categorical subtypes may examine transdiagnostic cohorts, because subtypes may cross diagnostic boundaries. Identifying dimensional subtypes transdiagnostically may help understand a given trait's biological underpinnings.

Mental Health Subtypes May Provide Insights Into Diagnostic Categories or Factors.—Using multifactorial approaches, investigators discovered previously unknown subtypes biologically and clinically linked to developmental outcomes. Unlike the p factor, the discovery of subtypes is extensive and beyond the scope of this review. Below, a few examples cover two categorical (attention-deficit/hyperactivity disorder [ADHD], ASD) and two dimensional outcomes (executive function, temperament). With 106 children, ADHD subtyping using resting-state functional connectivity magnetic resonance imaging and ADHD symptoms revealed insights into ADHD and typical subtypes (31). Specifically, the investigators found that similar subtypes existed within ADHD and typical samples and were associated with impulsive characteristics, independent of diagnosis. Such findings suggest that clinical subtypes are nested within broader typical heterogeneity, and such nesting may affect health outcomes independent of mechanism. Investigators discovered similar ASD subtype nesting using network-based functional connectivity (49). From 250 children within the Autism Brain Imaging Data Exchange Consortium, investigators found two subtypes associated with general within- and between- network connectivity in both ASD and control samples. Subtyping can improve the prediction of other outcomes. Through whole-brain measures of cortical shape (e.g., thickness and surface area) from 220 children within the Autism Brain Imaging Data Exchange Consortium, investigators discovered three subtypes. One subtype presented with cortical thickening and mild symptoms, another with cortical thinning and additional language impairments, and a third with altered shape and additional language impairments but not thinning. When these subtypes were incorporated into a predictive model of Autism Diagnostic Observation Schedule, Autism Diagnostic Observation Schedule scores were better predicted than if the subtypes were not included, suggesting that subtyping can improve clinical instruments.

Nested transdiagnostic categorical subtypes may require dimensional subtyping to further understand factors underlying mental health. Some studies discovered unknown temperament or executive function subtypes across ASD, ADHD, and typical cohorts. Three temperament subtypes, “mild,” “irritable,” and “surgent,” were identified from 437 children with or without ADHD followed longitudinally (50). Subtypes showed distinct amygdala functional connectivity patterns, and children composing the “irritable” subtype were more likely to receive a new comorbid diagnosis during longitudinal follow-up, indicating that these subtypes are clinically and biologically insightful. Furthermore, subtypes appeared independent of clinical diagnosis, suggesting that such subtypes are nested within typical (non-ADHD) variation. Another study (51) suggested the same using 1000 transdiagnostic children (ASD/ADHD/control) from a multisite study: a 320-case

discovery sample from the Children's National Hospital System, and a 692-case replication sample from the Kennedy Krieger Institute. The investigators identified three executive function profiles: one reflected problems with flexibility and emotional regulation; a second reflected problems with inhibition; and a third reflected issues with working memory, planning, and organization. Using brain imaging data to examine biological links to the three groups, the investigators found that executive function subtypes, but not clinical subtypes, observed differential brain activation in the right intraparietal gyrus during a dual-attention task. Taken together, these studies suggest that the same individual belongs in multiple subtypes, and some subtypes nest within typical variation.

Clinically Useful Subtypes May Emerge From the Multifactorial Approach.—A multifactorial approach may also prove useful for clinical treatments. A popular and recent use case is whether functional connectivity magnetic resonance imaging can be used to identify depression subtypes that may be more optimized for particular therapies (52–54). For example, a study of 122 treatment-resistant depression cases revealed two subtypes (53). One had negative subcallosal cingulate cortex to left insula connectivity and responded to medication but not cognitive behavioral therapy. The other had positive connectivity and responded to therapy but not medication. Another study was conducted on two independent datasets comprising 710 and 477 individuals, respectively (52). Investigators identified four subtypes using functional connectivity magnetic resonance imaging in over 1100 participants with treatment-resistant depression. The investigators further demonstrated that these four subtypes responded differentially to transcranial magnetic stimulation treatment, suggesting that these identified subtypes are clinically useful.

CONSIDERATION OF ONLY ONE PROBLEM LIMITS RESEARCH IMPACT

Because unifying and multifactorial approaches tackle different problems, each carries different limitations. As shown above, a p factor is useful for research and clinical applications. Clinically, a first-pass instrument that measures p factor could benefit resource-scarce environments. In research, modeling a p factor may account for other model misspecifications and provide insights regarding biological mechanisms common to multiple disorders. However, “teasing out” this latter benefit requires considering multifactorial approaches, for p factors might represent different common mechanisms for different sets of diagnoses. The unifying approach has a heterogeneity problem.

However, one must cautiously adopt multifactorial approaches to address the heterogeneity problem. Critically, most subtyping methods cannot discover unknown subtypes tied to a question (18). To validate subtypes, one must be very careful to use cross-validation, randomized holdouts, and/or independent testing to avoid overfitting (54). Such validation does not make the subtypes useful. The appropriate subtypes are likely going to differ based on whether one is interested in diagnosis, prognosis, treatment response, mechanism, etc. (18). Indeed, identified subtypes can share underlying mechanisms, whereas others remain distinct; the multifactorial approach has a comorbidity problem.

ADDRESSING BOTH COMORBIDITY AND HETEROGENEITY PROBLEMS CAN PRODUCE MORE INFORMATIVE FINDINGS

Despite evidence for the p factor across multiple disorders, many investigators caution against its existence across every disorder (32,33,41). Instead, most studies consider both comorbidity and heterogeneity problems within the single study. These studies adopt analytic frameworks that enable an investigator to assess whether comorbidity and/or heterogeneity are important for the question of interest. A careful and critical evaluation of these frameworks is needed to ensure that studies adopt best practices and standards.

The Bifactor Framework Incorporates Both Problems

A common framework is comparisons between bifactor (Figure 2A) and correlated factor models (Figure 2B) using confirmatory factor analysis (27,28,32–35,37–39). With correlated factor models, measures (e.g., symptom counts from structured diagnostic interviews) are mapped to distinct factors (beige and gray circles) reflecting psychopathology traits that may distinguish outcomes (blue and yellow histograms). For example, internalizing (inward behaviors such as disparaging oneself) and externalizing (outward behaviors such as disparaging another) are common traits due to extensive replication and validation. Between-factor correlations (Figure 2A, dotted gray line) may be modeled if factors interact. Often statistical fits estimate model accuracy (55,56). Similarly correlated factor models map low-dimensional input features to psychopathology traits and are assessed by statistical fits. However, an additional orthogonal general factor (black ellipse) is modeled; the factor correlates with all features.

The framework enables investigators to test whether identified biomarkers, such as polygenic risk scores, reflect the specific condition (e.g., ADHD/schizophrenia) or p, and two studies examined this question. From 3650 children's genetic and clinical data within the Avon Longitudinal Study of Parents and Children cohort, investigators found that schizophrenia and neuroticism polygenic risk scores were more associated with p factor than specific factors such as anxiety (38). Another clinical/genetic study of 11,551 children within the Child and Adolescent Twin Study in Sweden found that ADHD polygenic risk scores were more associated with p than hyperactivity/impulsivity (37). Taken together, these studies show how bifactor models help validate and derive inferences from potential biomarkers.

Though most p factor studies use a bifactor framework, this framework has limits. Because it is a supervised approach, the bifactor framework requires researchers to specify specific factors, such as internalizing and externalizing. Therefore, the researcher cannot discover new ones (18). Simulations show that bifactor models tend to be robust to misspecified factors (57) and, therefore, always have fits superior to those of correlated factor models. If one were to incorrectly model latent factors, models' fits would remain superior despite misspecification. Other empirical studies employing the bifactor model have failed to find evidence for p (17) or have found that a more specific factor fit just as well (58). Therefore, a bifactor model's statistical fits limit inferences regarding heterogeneity and comorbidity.

HYBRID MODELING CAN INCORPORATE BOTH PROBLEMS IN A STATISTICAL FRAMEWORK

Although there are limits to the bifactor framework, the studies we discussed above clearly demonstrate its utility. Newer hybrid models, such as normative modeling (19,59–61) or the functional random forest (FRF) (18,62), may overcome these limitations and enable new inferences from psychopathology studies.

Normative Modeling Is One Framework That Address Heterogeneity and Comorbidity Problems

Normative models (59) enable researchers to identify common (i.e., comorbid) mechanisms, identify potential subtypes or subfactors tied to an outcome of interest, and make predictions on individual cases. Constructing a normative model requires 4 steps (Figure 3A). First, one selects a reference cohort, clinical outcomes, and external measures. The external measures could be measures of brain function or structure, such as functional connectivity or cortical thickness, or the external data could be genetic or even behavioral. Using the reference cohort, the model maps the set of clinical outcomes to the external measures, and independent validation assesses performance. Careful selection of outcomes, measures, and cohort can address the comorbidity problem. Theoretically, one could even select a bifactor model within the framework. If performance exceeds a predefined threshold, the model is applied to another cohort to identify outliers (Figure 3B, scatterplot). To identify subtypes for further testing, one extracts features from the model, representing clinical/nonclinical associations and identifies subtypes from the features (Figure 3B, histograms).

Several studies incorporated normative models to address comorbidity and heterogeneity in clinical and nonclinical cohorts (19,60,61). Using data from 491 subjects within the Human Connectome Project, investigators constructed normative models to identify impulsivity subtypes tied to self-reported hyperactive symptoms (19). The investigators found one subtype, where dorsolateral prefrontal activity predicted high hyperactive symptoms and another subtype unrelated to such activity. To characterize heterogeneity of schizophrenia and bipolar disorder nested within age and sex, investigators applied normative modeling to 474 participants from the Thematically Organized Psychosis study (61). The investigators created normative models where age and sex predicted structural magnetic resonance imaging measures and then examined patient variation by diagnosis. They found that patients with the same or no diagnosis varied dramatically under the normative model, suggesting that bipolar/schizophrenia brain structural heterogeneity is nested within typical heterogeneity. Another study characterized ASD heterogeneity tied to brain structural development using 527 participants from the Longitudinal European Autism Project (60). Individuals with ASD were more variable; however, ASD heterogeneity in brain structure development remained largely nested within typical heterogeneity. Although a good future direction, limited sample sizes prevented the investigators from subtyping using neuroimaging data.

FRF Is Another Framework That Can Address Both Problems

More recently, we proposed the FRF to identify subtypes tied to an outcome (18,62). The FRF comprises several approaches, which have been explicitly detailed elsewhere (18). Here we will focus on the hybrid model (Figure 4A). In the hybrid model, a random forest model (63) is constructed and validated to predict an outcome from a set of predictors (Figures 4B, C); for example, using brain structural measures to predict a comorbid measure such as p . If the model performance exceeds predefined thresholds, then at least one brain structure may be associated with p . To test p 's heterogeneity, a community detection algorithm called infomap (64) is used to identify subtypes (Figure 4D) from the random forest model itself. Because the model links to the broader question, the identified subtypes likely reflect p -associated subtypes. The subtypes are assessed statistically via a permutation test assessing within-subtype compared with random-subtype similarity (65).

The FRF can identify trajectory subtypes from longitudinal data, such as p subtypes united or tied to ADHD outcomes. Here, functional data analysis (66) is used to extract each individual's trajectory, such as p symptom trajectories. Such trajectories can then be entered into the hybrid model (Figure 4A) to identify trajectory subtypes tied to an outcome, such as ADHD. An alternative approach uses an unsupervised algorithm, where a correlation matrix is derived from every participant's trajectory, where each participant pair represents the correlation between the pair's trajectories. Infomap is used to identify communities from the correlation matrix. Unlike the hybrid approach, such subtypes are less tied to any outcome and require greater validation.

Hybrid Frameworks Can Be Flexibly Applied to Address Comorbid and/or Heterogeneity Problems

By combining supervised and unsupervised learning, hybrid statistical frameworks, such as FRF and normative modeling, enable investigators to test biological underpinnings of common outcomes and whether subtypes underlie such mapping. The supervised component ensures that the biological measures are actually related to the outcome of interest. Through the unsupervised component, one can discover unknown subtypes tied to the model and therefore the outcome of interest. Out-of-sample validation is built in to both frameworks, which aids in deriving inferences from comorbidity and heterogeneity problems.

Despite our enthusiasm for such hybrid frameworks, no approach solves both problems. Both methods outlined here are limited, specializing them toward specific questions. For example, the normative model assumes that clinical features are all Gaussian and therefore has difficulty incorporating categorical information into the model. The longitudinal FRF struggles to identify individual trajectories when missing the first or last time point, while tweaking community detection algorithms may affect identified subtypes. More critically, both frameworks require large datasets. Therefore, the hybrid frameworks would benefit from further investigation using large-scale datasets and direct comparisons against unified, multifactor, and bifactor models. Such studies help determine best practices for the comorbidity and heterogeneity problems in future studies.

BOTH CLINICAL AND NONCLINICAL TRIAL RESEARCH SHOULD TEST THE HOMOGENEITY ASSUMPTION

The homogeneity assumption drives a comorbid and a heterogeneity problem. Research studies that tackle the comorbid and heterogeneity problem may identify new insights into pathophysiological processes and their biological links. However, one must adopt frameworks to overcome both problems, such as normative models (19), the FRF (18), and bifactor models (27,28). Each framework has strengths and limitations. Further study that compares the tools in new use cases may uncover new insights that improve both future usage of the tools and the tools themselves.

For example, our current tools to tackle the homogeneity assumption may not help address the treatment-effect heterogeneity problem (67–69) in research clinical trials (RCTs), where multiple factors may drive treatment-response variation. Most RCTs follow a parallel design, where each individual gets randomly assigned to a different treatment group, called an arm (69). In such designs, one can measure the variation between treatments; however, the variation between patients, clusters, or cluster/patient-treatment interactions cannot be measured. More complex RCTs, such as repeated cross-over designs (69), where patients are randomly assigned to different treatments at different periods throughout the study, can measure patient/cluster-treatment interactions. Variance between different samples can be tested to determine whether treatment or treatment-by-cluster effects are heterogeneous (67). Such complex RCTs combined with newer statistical methods may better tackle the heterogeneity problem with regard to treatment effects.

Ultimately, we need to continue to advance current statistical techniques and study designs tackling heterogeneity and comorbidity in our samples. Such improvements in clinical and nonclinical trial research will be required to refine treatments and truly expand our understanding of the biological phenomenon associated with various psychopathologies.

ACKNOWLEDGMENTS AND DISCLOSURES

This research was supported by DeStefano Family Foundation (to DAF); the National Library of Medicine Grant No. T15 LM007088 (to EF); and the National Institute of Mental Health Grant Nos. R01 MH096773, R01 MH091238, R01 MH096773-03S1, R01 MH 096773-05, R01 MH086654, R01 MH086654, R01 MH59107 (to DAF).

REFERENCES

1. Kendler KS (2009): An historical framework for psychiatric nosology. *Psychol Med* 39:1935–1941. [PubMed: 19368761]
2. Nigg JT (2006): Temperament and developmental psychopathology. *J Child Psychol Psychiatry* 47:395–422. [PubMed: 16492265]
3. Mason D, Hsin H (2018): “A more perfect arrangement of plants”: The botanical model in psychiatric nosology, 1676 to the present day. *Hist Psychiatry* 29:131–146. [PubMed: 29480060]
4. Wang T, Zhang X, Li A, Zhu M, Liu S, Qin W, et al. (2017): Polygenic risk for five psychiatric disorders and cross-disorder and disorder-specific neural connectivity in two independent populations. *Neuroimage Clin* 14:441–449.

5. Cross-Disorder Group of the Psychiatric Genomics Consortium, Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, et al. (2013): Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* 381:1371–1379. [PubMed: 23453885]
6. Sabuncu MR, Konukoglu E, Alzheimer's Disease Neuroimaging Initiative (2015): Clinical prediction from structural brain MRI scans: A large-scale empirical study. *Neuroinformatics* 13:31–46. [PubMed: 25048627]
7. Katuwal GJ, Baum SA, Cahill ND, Michael AM (2016): Divide and conquer: Sub-grouping of ASD improves ASD detection based on brain morphometry. *PLoS One* 11:1–24.
8. Chen CP, Keown CL, Jahedi A, Nair A, Pflieger ME, Bailey BA, Müller R-A (2015): Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual re- gions in autism. *Neuroimage Clin* 8:238–245. [PubMed: 26106547]
9. Sen B, Borle NC, Greiner R, Brown MRG (2018): A general prediction model for the detection of ADHD and autism using structural and functional MRI. *PLoS One* 13:e0194856. [PubMed: 29664902]
10. Moffitt TE, Harrington H, Caspi A, Kim-Cohen J, Goldberg D, Gregory AM, Poulton R (2007): Depression and generalized anxiety disorder: Cumulative and sequential comorbidity in a birth cohort followed prospectively to age 32 years. *Arch Gen Psychiatry* 64:651–660. [PubMed: 17548747]
11. Copeland W, Shanahan L, Costello EJ, Angold A (2011): Cumulative prevalence of psychiatric disorders by young adulthood: A prospective cohort analysis from the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry* 50:252–261. [PubMed: 21334565]
12. Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ, et al. (2011): Development of lifetime comorbidity in the World Health Organization world mental health surveys. *Arch Gen Psychiatry* 68:90–100. [PubMed: 21199968]
13. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009): Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. *Lancet* 373:234–239. [PubMed: 19150704]
14. Sartor CE, Grant JD, Bucholz KK, Madden PAF, Heath AC, Agrawal A, et al. (2010): Common genetic contributions to alcohol and cannabis use and dependence symptomatology. *Alcohol Clin Exp Res* 34:545–554. [PubMed: 20028363]
15. Guo W, Samuels JF, Wang Y, Cao H, Ritter M, Nestadt PS, et al. (2017): Polygenic risk score and heritability estimates reveals a genetic relationship between ASD and OCD. *Eur Neuropsychopharmacol* 27:657–666. [PubMed: 28641744]
16. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. (2014): Most genetic risk for autism resides with common variation. *Nat Genet* 46:881–885. [PubMed: 25038753]
17. Murray AL, Eisner M, Ribeaud D (2016): The development of the general factor of psychopathology “p factor” through childhood and adolescence. *J Abnorm Child Psychol* 44:1573–1586. [PubMed: 26846993]
18. Feczko E, Miranda-Domínguez Ó, Marr M, Graham AM, Nigg JT, Fair DA (2019): The heterogeneity problem: Approaches to identify psychiatric subtypes. *Trends Cogn Sci* 23:584–601. [PubMed: 31153774]
19. Marquand AF, Rezek I, Buitelaar J, Beckmann CF (2016): Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biol Psychiatry* 80:552–561. [PubMed: 26927419]
20. Brainstorm Consortium, Anttila V, Bulik-Sullivan, Finucane HK, Walters RK, Bras, et al. (2018): Analysis of shared heritability in common disorders of the brain. *Science* 360:eaap8757.
21. Constantino JN, Charman T (2016): Diagnosis of autism spectrum disorder: Reconciling the syndrome, its diverse origins, and variation in expression. *Lancet Neurol* 15:279–291. [PubMed: 26497771]
22. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ (2013): DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *Am J Psychiatry* 170:59–70. [PubMed: 23111466]
23. Fried EI (2015): Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Front Psychol* 6:1–11. [PubMed: 25688217]

24. Uddin LQ, Dajani DR, Voorhies W, Bednarz H, Kana RK (2017): Progress and roadblocks in the search for brain-based biomarkers of autism and attention-deficit/hyperactivity disorder. *Transl Psychiatry* 7:e1218. [PubMed: 28892073]
25. Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017): Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* 145:137–165. [PubMed: 27012503]
26. Friston KJ, Redish AD, Gordon JA (2017): Computational nosology and precision psychiatry. *Comput Psychiatry* 1:2–23.
27. Caspi A, Moffitt TE (2018): All for one and one for all: Mental disorders in one dimension. *Am J Psychiatry* 175:831–844. [PubMed: 29621902]
28. Neumann A, Pappa I, Lahey BB, Verhulst FC, Medina-Gomez C, Jaddoe VW, et al. (2016): Single nucleotide polymorphism heritability of a p factor in children. *J Am Acad Child Adolesc Psychiatry* 55:1038–1045.e4. [PubMed: 27871638]
29. Pruett JR, Povinelli DJ (2016): Commentary—Autism spectrum disorder: Spectrum or cluster? *Autism Res* 9:1237–1240. [PubMed: 27333214]
30. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. (2017): The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* 126:454–477. [PubMed: 28333488]
31. 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]
32. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. (2014): The p factor. *Clin Psychol Sci* 2:119–137. [PubMed: 25360393]
33. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ (2012): Is there a general factor of prevalent psychopathology during adulthood? *J Abnorm Psychol* 121:971–977. [PubMed: 22845652]
34. Kaczkurkin 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]
35. Kaczkurkin AN, Park SS, Sotiras A, Moore TM, Calkins ME, Cieslak M, et al. (2019): Evidence for dissociable linkage of dimensions of psychopathology to brain structure in youths. *Am J Psychiatry* 176:1000–1009. [PubMed: 31230463]
36. 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]
37. Brikell I, Larsson H, Lu Y, Pettersson E, Chen Q, Kuja-Halkola R, et al. (2018): The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology [published online ahead of print Jun 22]. *Mol Psychiatry*.
38. Jones HJ, Heron J, Hammerton G, Stochl J, Jones PB, Cannon M, et al. (2018): Investigating the genetic architecture of general and specific psychopathology in adolescence. *Transl Psychiatry* 8:145. [PubMed: 30089819]
39. Tackett JL, Lahey BB, van Hulle C, Waldman I, Krueger RF, Rathouz PJ (2013): Common genetic influences on negative emotionality and a p factor in childhood and adolescence. *J Abnorm Psychol* 122:1142–1153. [PubMed: 24364617]
40. McDougall GJ (1990): A review of screening instruments for assessing cognition and mental status in older adults. *Nurse Pract* 15:18–28. [PubMed: 2255423]
41. Moore TM, Calkins ME, Satterthwaite TD, Roalf DR, Rosen AFG, Gur RC, Gur RE (2019): Development of a computerized adaptive screening tool for overall psychopathology (“p”). *J Psychiatr Res* 116:26–33. [PubMed: 31176109]
42. Oparil S, Schmieder RE (2015): New approaches in the treatment of hypertension. *Circ Res* 116:1074–1095. [PubMed: 25767291]
43. Hamrahian SM (2017): Management of hypertension in patients with chronic kidney disease. *Curr Hypertens Rep* 19:43. [PubMed: 28451853]

44. Liu M-Y, Li N, Li WA, Khan H (2017): Association between psycho- social stress and hypertension: A systematic review and meta-anal- ysis. *Neurol Res* 39:573–580. [PubMed: 28415916]
45. Appel LJ (2017): The effects of dietary factors on blood pressure. *Cardiol Clin* 35:197–212. [PubMed: 28411894]
46. Cornelissen VA, Smart NA (2013): Exercise training for blood pressure: A systematic review and meta-analysis. *J Am Heart Assoc* 2:e004473. [PubMed: 23525435]
47. Rossi GP, Ceolotto G, Caroccia B, Lenzini L (2017): Genetic screening in arterial hypertension. *Nat Rev Endocrinol* 13:289–298. [PubMed: 28059156]
48. Insel TR (2014): The NIMH Research Domain Criteria (RDoC) Project: Precision medicine for psychiatry. *Am J Psychiatry* 171:395–397. [PubMed: 24687194]
49. Easson AK, Fatima Z, McIntosh AR (2019): Functional connectivity- based subtypes of individuals with and without autism spectrum dis- order. *Netw Neurosci* 3:344–362. [PubMed: 30793086]
50. Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, Nigg JT (2014): Subtyping attention-deficit/ hyperactivity disorder using temperament dimensions. *JAMA Psychiatry* 71:1015. [PubMed: 25006969]
51. Vaidya CJ, You X, Mostofsky S, Pereira F, Berl MM, Kenworthy L (2019): Data-driven identification of subtypes of executive function across typical development, attention deficit hyperactivity disorder, and autism spectrum disorders. *J Child Psychol Psychiatry* 61:51–61. [PubMed: 31509248]
52. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. (2017): Resting-state connectivity biomarkers define neuro- physiological subtypes of depression. *Nat Med* 23:28–38. [PubMed: 27918562]
53. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, et al. (2017): Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive- behavioral therapy or antidepressant medication for major depressive disorder. *Am J Psychiatry* 174:533–545. [PubMed: 28335622]
54. Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, et al. (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *Neuroimage Clin* 22:101796. [PubMed: 30935858]
55. Wagenmakers E, Farrell S (2004): AIC model selection using Akaike weights. *Psychon Bull Rev* 11:192–196. [PubMed: 15117008]
56. Akaike H (1973): Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. *Proceedings of the Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, 267–281.
57. Greene AL, Eaton NR, Li K, Forbes MK, Krueger RF, Markon KE, et al. (2019): Are fit indices used to test psychopathology structure biased? A simulation study. *J Abnorm Psychol* 128:740–764. [PubMed: 31318246]
58. Haltigan JD, Aitken M, Skilling T, Henderson J, Hawke L, Battaglia M, et al. (2018): “P” and “DP:” Examining symptom-level bifactor models of psychopathology and dysregulation in clinically referred children and adolescents. *J Am Acad Child Adolesc Psychiatry* 57:384–396. [PubMed: 29859554]
59. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF (2019): Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry* 24:1415–1424. [PubMed: 31201374]
60. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. (2019): Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4:567–578. [PubMed: 30799285]
61. Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, et al. (2018): Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry* 75:1146. [PubMed: 30304337]

62. Feczko E, Balba N, Miranda-Dominguez O, Cordova M, Karalunas SL, Irwin L, et al. (2018): Subtyping cognitive profiles in autism spectrum disorder using a random forest algorithm. *Neuroimage* 172:674–688. [PubMed: 29274502]
63. Breiman LEO (2001): Random forests. *Mach Learn* 45:5–32.
64. Rovall M, Bergstrom C (2008): Maps of random walks on complex network reveal community structure. *Proc Natl Acad Sci U S A* 105:1118–1123. [PubMed: 18216267]
65. Newman MEJ (2006): Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103:8577–8582. [PubMed: 16723398]
66. Ramsay T (2002): Spline smoothing over difficult regions. *J R Stat Soc Ser B Stat Methodol* 64:307–319.
67. Cortés J, González JA, Medina MN, Vogler M, Vilaró M, Elmore M, et al. (2019): Does evidence support the high expectations placed in precision medicine? A bibliographic review. *F1000Research* 7:30.
68. Senn S, Julious S (2009): Measurement in clinical trials: A neglected issue for statisticians? *Stat Med* 28:3189–3209. [PubMed: 19455540]
69. Senn S (2016): Mastering variation: Variance components and personalised medicine. *Stat Med* 35:966–977. [PubMed: 26415869]

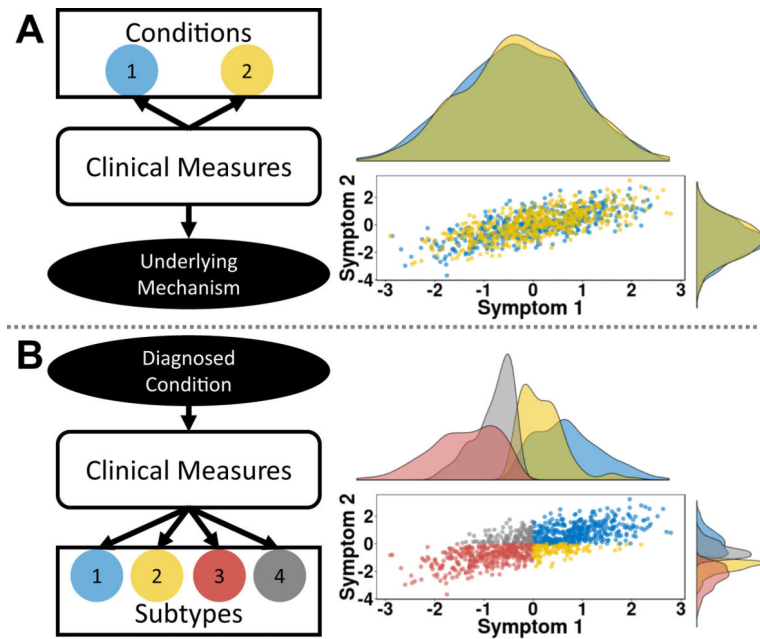


Figure 1.

Illustration of two problems that emerge from the homogeneity assumption. (A) The comorbidity problem occurs when a common mechanism may underlie two different conditions and can be inferred from the relationship between measures. The interaction between two symptoms is plotted on the left and colored by different diagnosed conditions (blue and yellow). Axes represent symptom z scores for simplicity. The histograms show that the symptoms do not distinguish between the conditions, suggesting that the observed clinical features may reflect a common underlying mechanism. (B) The heterogeneity problem occurs when multiple underlying mechanisms may drive a given condition. The interaction between two symptoms is plotted on the left. Axes represent symptom z scores. Here, different subtypes (blue, yellow, red, gray) can emerge that may vary independent of the interaction between measures. The histograms show that the subtypes relate to each other in different ways. Subtypes 1 and 2 show overlapping distributions for symptom 1, as do subtypes 3 and 4. However, for symptom 2, subtypes 1 and 4 overlap, as do subtypes 2 and 3.

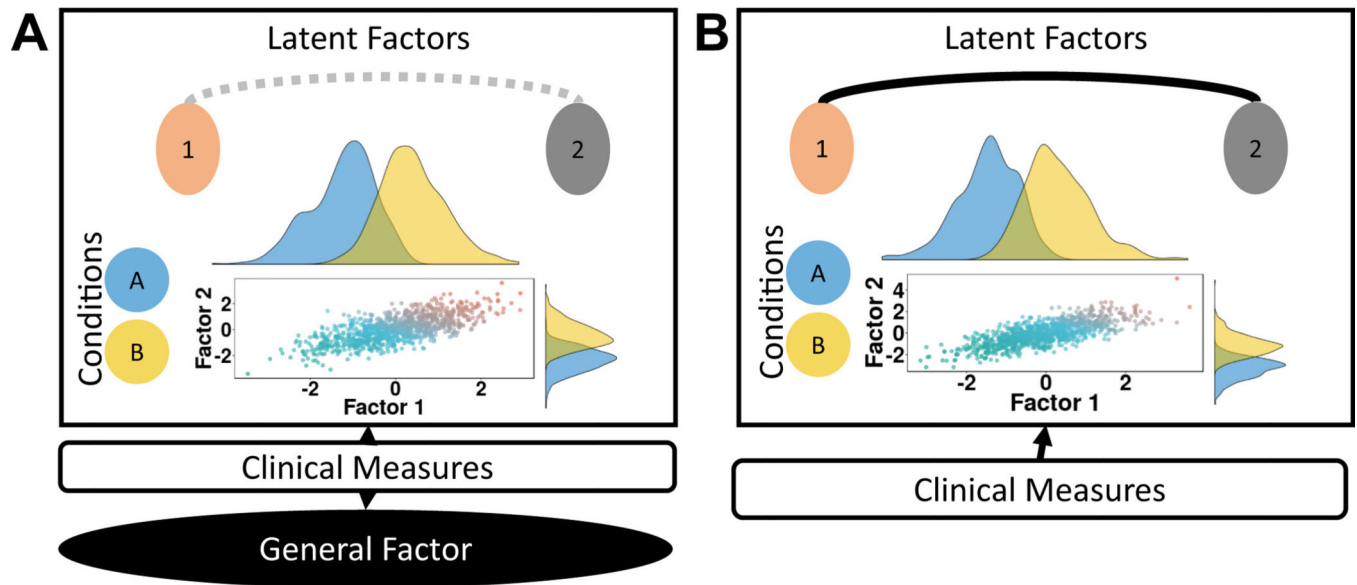


Figure 2.

Comparison of bifactor (A) and correlated models (B). (A) Clinical measures (middle box) are modeled to a general factor (bottom black circle) and to latent factors (beige and gray), which distinguish between two conditions (blue and yellow). The scatterplot shows the relationship between the two factors, which are correlated (gray dotted line) here. The histograms show that each factor helps to distinguish between conditions, such that lower values in each factor correspond more to the blue than the yellow condition. Therefore, while the general factor may capture a shared mechanism across diagnosis, the latent factors may capture mechanisms specific to each diagnosis. (B) In correlated models, clinical measures are modeled onto specific factors, but do not include a general factor. The scatterplot and histograms show that the identified factors help distinguish between conditions. However, no common mechanism can be captured with this model.

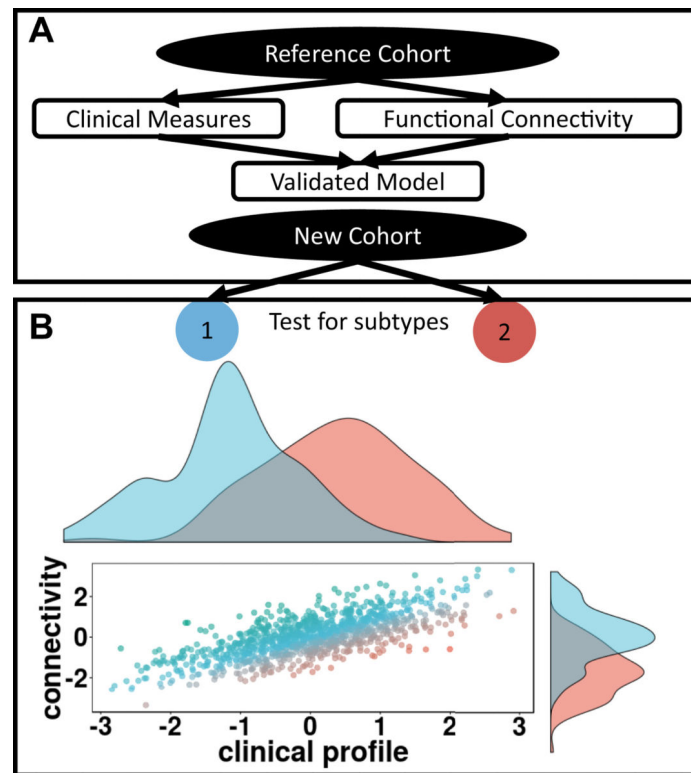


Figure 3.

(A) Overview of normative modeling. A reference cohort (top black ellipse) is selected and a model is generated fitting external data such as functional connectivity and clinical severity scores. Once validated, the model can be applied to an independent cohort to identify potential outliers. Through proper cohort, feature selection, and model validation, the normative model accounts for the comorbidity problem. (B) From an independent cohort, features can be extracted from the normative model. Independent algorithms can then identify subtypes (black rectangle) controlling for the heterogeneity problem. The scatterplot shows outliers from the relationship between functional connectivity and clinical profile. Axes represent connectivity and clinical severity z scores for simplicity. Individuals are color-coded by how far they deviate from the fitted model. The histograms show two possible subtypes identified from the model residuals (blue and red). These subtypes differ in ways that are opposite from the common mechanism. Subtype 1 shows a lower clinical scores but greater connectivity, while subtype 2 shows greater clinical scores but lower connectivity.

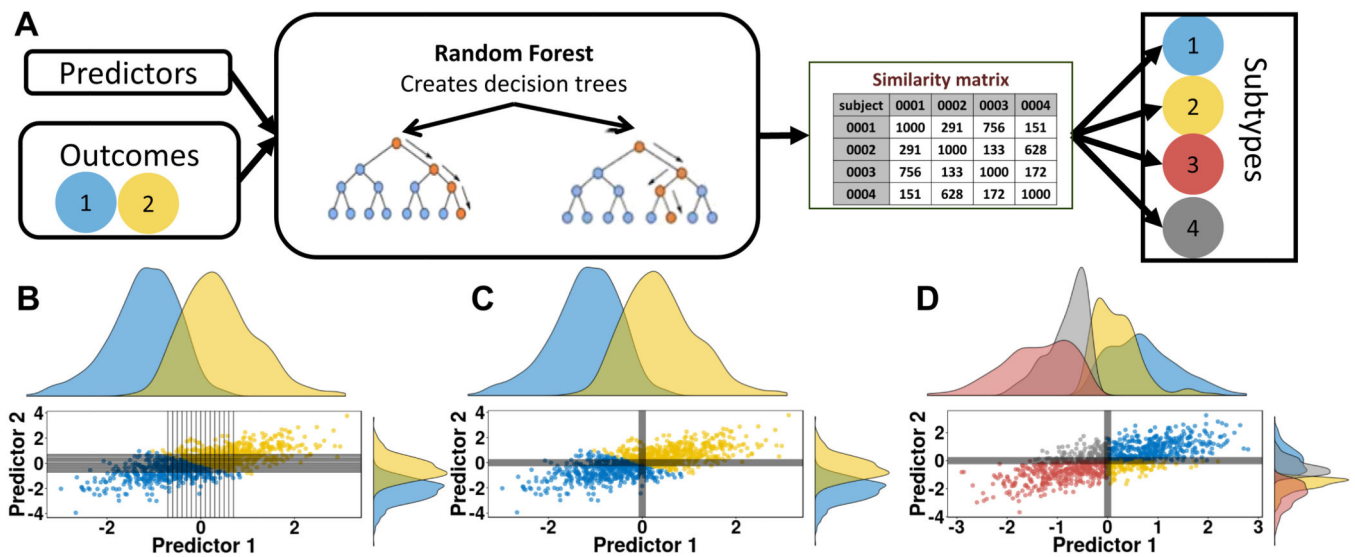


Figure 4.

(A) Schematic of the functional random forest approach. A random forest model is used to predict outcomes. Depending on the question (e.g., by selecting transdiagnostic outcomes and/or clinical measures) the random forest model itself can test for a common underlying mechanism and, therefore, account for comorbidity. If the model is valid, a similarity matrix is generated from the random forest model and can be used to identify subtypes in the same or an independent dataset. (B) A scatter visualization of two predictors represented as z scores. Color reflects outcome. Each tree within the forest creates a rule for each predictor, shown in black lines. (C) The same scatter shows the average decision made across the random forest, which separates the two outcomes. Participants within the same quadrant are more similar than in different quadrants. (D) A scatter of predictors from an independent cohort, where participants are coded by identified subtype.