Commentary

Understanding the Nature and Treatment of Psychopathology: Can the Data Guide the Way?

Deanna M. Barch

Over the past 10 years, there has been a growing availability of sophisticated computational resources and algorithms that allow data-driven exploration of a wide range of data types in the service of enhancing our understanding of the structure, causes, course, and treatment of psychopathology. These approaches involve a whole family of related computational techniques, including both supervised and unsupervised machine learning approaches, such as support vector machines, random forest algorithms, and deep learning, to name a few, as well as more traditional data-driven clustering and factor analytic approaches. These methods rely on high-performance computational resources and the availability of sufficiently large data sets to drive robust learning and results. Such advances have generated much excitement about a potential revolution in our understanding of mental illness, with the hope that the outcomes of these innovations will lead to better prevention and treatment of a range of mental health conditions. There are a variety of applications that have been the target of data-driven exploration, including 1) generating novel understanding of the structure of psychopathology (e.g., new dimensions or categories that differ from traditional conceptions); 2) subtyping existing categories; 3) predicting risk for the development of mental health conditions; 4) predicting a course or outcome once an individual has been identified as having a mental illness; 5) predicting treatment response; 6) determining informative features that might guide further theory development; and 7) aggregating or integrating data across levels of analyses (i.e., multimodal data integration).

While there has been huge energy around the potential advances afforded by the availability of these techniques, it is not yet clear to what extent they are generating new and robust data that are reshaping our understanding of mental illness. Thus, the goal of this special issue of *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* is to review the progress that has (or has not) been made in a number of domains using data-driven approaches to better understand the structure of psychopathology, etiological mechanisms, risk factors, and treatment outcomes. In addition, the goal was to ask authors to identify best practices that will hopefully further spur robust and replicable innovative findings, as well as to identify barriers to progress and considerations for future work that will drive additional advancements.

Nosological Organization. Given the goals outlined above, it is fitting that the first article is this special issue outlines the progress to date on the program of research being conducted as part of the B-SNIP (Bipolar–Schizophrenia Network for Intermediate Phenotypes) consortium, designed to identify novel biotypes among individuals experiencing forms of mental illness that involve psychotic features (1). This program of research has arguably been the most successful to date in identifying a novel categorical structure for individuals experiencing mental illness in the psychosis spectrum domain (2), though the HiTOP (Hierarchical Taxonomy of Psychopathology) consortium, among others, has also had good success in identifying a hierarchical dimensional structure of psychopathology that includes psychosis (3,4). The B-SNIP consortium has used a deep phenotyping approach coupled with data-driven clustering of cognitive and electrophysiological measures to identify 3 clusters (biotypes) that cut across diagnoses of schizophrenia, schizoaffective disorder, and bipolar disorder, with follow-up work using a variety of neural and behavioral measures to validate these biotypes (5). In their contribution to this special issue, Clementz *et al.* (1) provide novel data by analyzing the clinical characteristics of the biotypes in relationship to traditional diagnoses, providing evidence that the clinical features of these biotypes are dissociable from those of traditional diagnostic categories. Further, they use the information about the clinical features that distinguish the biotypes to drive novel predictions for future research on personalized treatment selection and response, and they outline additional research directions that are needed to test these hypotheses. For example, they point out that individuals in Biotype 1 seemed to be characterized by impaired cognitive function coupled with altered sensory responses to salient stimuli. They hypothesize that such individuals may specifically benefit from treatment focused on sensory processing, and as such, biotype classification may be a useful predictor of, or selection criterion for, targeted interventions.

In service to a similar goal, Rokham *et al.* (6) present a novel data-driven approach for identifying inaccurate diagnostic "labels" using structural imaging data as an example. These authors also used data from the B-SNIP consortium and used support vector machine learning to identify individuals whose given diagnosis did not seem well-captured by the structural imaging data. They in turn used the information from the structural imaging data to suggest different diagnostic labels (i.e., schizophrenia rather than bipolar disorder) that created more homogeneous groups. Critically, by examining the features used to identify noisy labels and that predicted more "accurate" membership in diagnostic groups, one can isolate key neurobiological features that are relevant to identify neurobiological characteristics that might inform clearer nosological boundaries or even novel structure or subtyping.

719

Subtyping. In addition to attempts to more fundamentally revise our overall nosological frameworks, data-driven analyses can also be used in the service of identifying meaningful subtypes within existing diagnostic categories. The contribution by Nigg *et al.* (7) in this special issue outlines the growing body of work attempting to do so in the domain of attention-deficit/hyperactivity disorder. Nigg *et al.* point out the many computational and conceptual approaches that have been taken to attempt robust subtyping, ranging from using traditional symptom domains to cognitive function, temperament, and other trait-level characteristics, to neurobiological characteristics, to evidence about etiological mechanisms (e.g., genes, environmental exposures). They highlight the success and promising evidence for broadly replicable subtypes but also outline important limitations in the existing efforts and key directions for next steps and enhancements.

Prediction. Another key domain that holds promise for data-driven approaches is predicting future outcomes, whether that be development of a disorder, course or outcome, or response to treatment. The contribution by Worthington *et al.* (8) in this special issue reviews the evidence in this regard for predicting illness onset specifically in the domain of conversion from clinical high risk for psychosis to full-blown psychosis. Outside of predicting Alzheimer's disease, this is the domain in which the most conversion prediction work has been done, particularly in terms of attempts to replicate prediction across samples. These efforts have been supported by large-scale consortia and transnational efforts that are generating the type of data that allow for integration across studies and replication. As reviewed by Worthington *et al.* (8), to date the most accurate prediction algorithms at the individual level are those include features such as symptoms, family history, social function, and functional decline. Worthington *et al.* (8) also review efforts to determine whether brain features can add predictive utility. However, these authors note that most of these efforts have focused on group-level discrimination and relatively few have provided evidence of individual-level prediction, though they do review several promising efforts, such as those focused on using brain age discrepancies as predictors (9). The contribution by Yip *et al.* (10) reviews progress in data-driven prediction for treatment outcome, focusing in particular on addiction and response to substance abuse treatment. Yip *et al.* review distinctions among different key aspects of treatment response, including whether an individual completes treatment, his or her response to treatment, and whether he or she relapses posttreatment. Interestingly, Yip *et al.* (10) found that for most studies, neural variables was comparable to or even better than clinical variables in predicting these various components of treatment response. At the same time, Yip *et al.* (10) point out the many limitations of the existing work, including small sample sizes, the need for additional validation, and the challenges of interpreting the meaningfulness of predictive features.

Novel Features for Use in Prediction. Much of the work described in this special issue uses either clinical features or brain imaging. However, there are a range of data types that are amenable to data-driven analyses. Some examples of these additional types of data are provided in the contributions by Washington *et al.* (11) and Corcoran and Cecchi (12). In addition to outlining sophisticated approaches to using machine learning to identify salient features in diagnostic interviews, Washington *et al.* (11) provide intriguing examples of other types of data, such as the computer vision approach to analyze video data sets, eye-tracking metrics that may reflect attention to characteristics of other humans and in the environment, the use of crowdsourcing for labeling data for use in training algorithms, and the use of mobile artificial intelligence to help administer treatments and track outcomes. Corcoran and Cecchi (12) review the use of natural language processing approaches for diagnosis of a range of disorders, using schizophrenia as a more in-depth review case. They highlight the array of methodologies that can be used that fall within the domain of natural language processing but also highlight some of the challenges with standardization and harmonization.

Testing Etiological Theories and Identifying Neurobiological Mechanisms. As noted above, another important use for data-driven analytic approaches is to identify novel features that better inform our understanding of the etiology of mental illness. The contribution to the special issue by Zhang *et al.* (13) reviews advances in network science that are beginning to contribute to our understanding of psychopathology in this way, as well as being able to inform nosology. For example, they highlight the potential utility of generative network modeling (GNM) as way of organizing our understanding of structural and functional brain network disruptions in mental illness. GNM provides formalisms that instantiate hypotheses about the development and function of brain networks based on proposed rules about how brain regions are connected and communicate with each other, providing an in silico experimental platform for testing computational hypotheses based on empirical data. They also outline the potential utility of network control theory, an advance derived from dynamical systems models arising out of engineering. Like GNM, network control theory models can be used to develop, instantiate, and test hypotheses about how dynamic brain systems operate and the ways in which they may be disrupted in various forms of mental illness. While these approaches have not yet seen widespread use in the field, Zhang *et al.* (13) provide several examples of recent innovations in their application to schizophrenia, autism, traumatic brain injury, and response to neurostimulation treatment.

Best Practices, Challenges, Opportunities, and Future Directions. One of the repeated themes across many contributions to this special issue is the need to use best practices in data-driven analyses that promote robustness and replicability and that facilitate substantive interpretation of results. The contribution by Nielsen *et al.* (14) is an explicit outline of such best practices, including the need to identify useful applications that move the field forward, key issues with regard to ensuring generalizability through the use of both within-sample (e.g., k-fold cross-validation) and out-of-sample replication (i.e., testing performance on an untouched sample), and suggestions regarding how to assess the robustness of

prediction performance. Further, Nielsen et al. (14) also high-light the need to use best practices that facilitate interpretation of the results of data-driven analyses, which can be challenging when using a high-dimensional feature set. For example, they discuss methods for determining whether there are feature components that play a unique or important role in prediction, the conundrum of interpreting individual features in the context of results that involve complex interactions among features, and the challenges of determining when machine learning approaches actually provide added utility over and above more traditional univariate approaches.

These best practice suggestions are echoed in several other papers in the special issue, in particular in the contributions of Yip et al. (10) and Nigg et al. (7). These authors also emphasize the importance of both in-sample and out-of-sample replication, but also emphasize the need to balance between theory-driven and data-driven approaches. One of the advantages of data-driven approaches is that it allows for a broad exploration of a space of features that might have utility rather than focusing on only a specific set of features specified a priori by a particular theory. At the same time, the choice of the features to include in any data-driven analysis is always theory-driven to some extent, even if that theory is quite general (i.e., the brain is important!). Further, more explicit theoretical frameworks can significantly enhance the utility and generalizability of data-driven results, such as the use of GNM outlined by Zhang et al. (13) as a means to test hypotheses about neural organization. In addition, in the absence of any theory, it is often quite difficult to interpret the results of data-driven analyses. Thus, ideally progress in the domain of data-driven analyses will involve a cyclic interplay between bottom-up and top-down approaches that are mutually informative, with the strengths of one approach complementing the weaknesses of the other.

As noted above, the best practice suggestions consistently highlight the need to establish out-of-sample replication. Many papers in the field have primarily focused on within-sample replication, with the assumption that robust practices such as k-fold cross-validation and analytic approaches that penalize for complexity will help avoid capitalizing on sample-specific variance that might not facilitate replication. However, the contribution by Boeke et al. (15) to this special issue is a cautionary tale of why out-of-sample replication is so critical to establish generalizability. These authors used machine learning to identify biomarkers for trait anxiety, first in a discovery sample and then with attempted replication in an unseen test sample. They found robust results in the discovery sample, but the prediction model failed to replicate in the holdout sample. Such null replication results are important to publish, as they provide further evidence of the need to assess generalizability before making stronger inferences about findings based only on within-sample replication.

While it is easy to say that best practice is to test any data-driven analysis in an unseen sample, in practical reality this is often hard. Robust generation of a predictive model requires large sample sizes, and relatively few researchers have access to samples that are sufficiently big enough to allow them to hold out an "unseen" test set that affords an adequate test of out-of-sample generalization. Boeke et al. (15) were fortunate to have a sufficiently large sample to allow a reasonable size for the discovery data set as well as the unseen test sample,

but their data set used a relatively limited set of features (functional connectivity and structural data). The more complex and multidimensional the data set, the harder it is to collect an ample number of participants to conduct both within-sample and out-of-sample generalization. Further, it is often hard to convince funders of the need to collect a second large data set that is essentially the same as a previously collected one, though that is exactly what is needed to rigorously test replication. As such, it will be important for funders and the field to value the collection of such replication data sets in the service of validating the generalizability and clinical utility of novel results identified through data-driven analyses. This requires us to respect the need to establish replicability as highly as innovation, since a priority only on the latter can make establishing generalizability more challenging. More recently, there are several studies that are recruiting sample sizes large enough to allow robust replication, such as the ABCD (Adolescent Brain and Cognitive Development) study and the UK Biobank, but such data sets are still relatively rare. Efforts to support replication are also an area in which promoting and facilitating the use of common measures across studies is highly valuable because this affords the ability to use different studies as mutually informative train and test data sets (16). For example, the Human Connectome Project Development explicitly adopted a number of the same measures as the ABCD study to facilitate using the two data sets to establish replicability of data-driven analyses. Overall, however, it has been challenging to establish widespread adoption of common measures across studies in ways that facilitate harmonization, as many scientists also prioritize innovation over replication. Both are critically important, and both are needed to move the field forward in ways that translate to clinical useful measures. Hopefully the advances and promising findings reported in the articles in this special issue convince readers that it is worth the effort, and will help to promote the use of best practices in new studies employing the range of increasingly sophisticated computational approaches available to investigators in their pursuit of understanding the etiology, structure, course, and treatment of mental illness.

Article Information

From the Departments of Psychological and Brain Sciences, Psychiatry, and Radiology, Washington University in St. Louis, St. Louis, Missouri.

Address correspondence to Deanna M. Barch, Ph.D., Department of Psychological and Brain Sciences, Washington University, Box 1125, One Brookings Dr, St. Louis, MO 63130; E-mail: dbarch@wustl.edu.

References

1. Clementz BA, Trotti RL, Pearlson GD, Keshavan MS, Gershon ES, Keedy SK, et al. (2020): Testing psychosis phenotypes from Bipolar–Schizophrenia Network for Intermediate Phenotypes for clinical application: Biotype characteristics and targets. Biol Psychiatry Cogn Neurosci Neuroimaging 5:808–818.

2. 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.
3. Kotov R, Jonas KG, Carpenter WT, Dretsch MN, Eaton NR, Forbes MK, *et al.* (2020): Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. World Psychiatry 19:151–172.
4. 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.
5. Ivleva EI, Clementz BA, Dutcher AM, Arnold SJ, Jeon-Slaughter H, Aslan S, *et al.* (2016): Brain structure biomarkers in the psychosis biotypes: Findings from the Bipolar–Schizophrenia Network for Intermediate Phenotypes. Biol Psychiatry 82:26–39.
6. Rokham H, Pearlson G, Abrol A, Falakshahi H, Plis SM, Calhoun VD (2020): Addressing inaccurate nosology in mental health: A multilabel data cleansing approach for detecting label noise from structural magnetic resonance imaging data in mood and psychosis disorders. Biol Psychiatry Cogn Neurosci Neuroimaging 5:819–832.
7. Nigg JT, Karalunas SL, Feczko E, Fair DA (2020): Toward a revised nosology for attention-deficit/hyperactivity disorder heterogeneity. Biol Psychiatry Cogn Neurosci Neuroimaging 5:726–737.
8. Worthington MA, Cao H, Cannon TD (2020): Discovery and validation of prediction algorithms for psychosis in youths at clinical high risk. Biol Psychiatry Cogn Neurosci Neuroimaging 5:738–747.
9. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, *et al.* (2018): Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. JAMA Psychiatry 75:960–968.
10. Yip SW, Kiluk B, Scheinost D (2020): Toward addiction prediction: An overview of cross-validated predictive modeling findings and considerations for future neuroimaging research. Biol Psychiatry Cogn Neurosci Neuroimaging 5:748–758.
11. Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, *et al.* (2020): Data-driven diagnostics and the potential of mobile artifical intelligence for digital therapeutic phenotyping in computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 5:759–769.
12. Corcoran CM, Cecchi GA (2020): Using language processing and speech analysis for the identification of psychosis and other disorders. Biol Psychiatry Cogn Neurosci Neuroimaging 5:770–779.
13. Zhang X, Braun U, Tost H, Bassett DS (2020): Data-driven approaches to neuroimaging analysis to enhance psychiatric diagnosis and therapy. Biol Psychiatry Cogn Neurosci Neuroimaging 5:780–790.
14. Nielsen AN, Barch DM, Petersen SE, Schlaggar BL, Greene DJ (2020): Machine learning with neuroimaging: Evaluating its applications in psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 5:791–798.
15. Boeke EA, Holmes AJ, Phelps EA (2020): Toward robust anxiety biomarkers: A machine learning approach in a large-scale sample. Biol Psychiatry Cogn Neurosci Neuroimaging 5:799–807.
16. Barch DM, Gotlib IH, Bilder RM, Pine DS, Smoller JW, Brown CH, *et al.* (2016): Common measures for National Institute of Mental Health funded research. Biol Psychiatry 79:e91–e96.