JAMA Psychiatry | Review

Clinical Promise of Brain-Phenotype Modeling
A Review

Abigail S. Greene, PhD; R. Todd Constable, PhD

IMPORTANCE Assessing the link between whole-brain activity and individual differences in cognition and behavior has the potential to offer insights into psychiatric disorder etiology and change the practice of psychiatry, from diagnostic clarification to intervention. To this end, recent application of predictive modeling to link brain activity to phenotype has generated significant excitement, but clinical applications have largely not been realized. This Review explores explanations for the as yet limited practical utility of brain-phenotype modeling and proposes a path forward to fulfill this clinical potential.

OBSERVATIONS Clinical applications of brain-phenotype models are proposed and will require coordinated collaboration across the relatively siloed fields of psychometrics and computational neuroscience. Such interdisciplinary work will maximize the reliability and validity of modeled phenotypic measures, ensuring that resulting brain-based models are interpretable and useful. The models, in turn, may shed additional light on the neurobiological systems into which each phenotypic measure taps, permitting further phenotype refinement.

CONCLUSIONS AND RELEVANCE Together, these observations reflect an opportunity: bridging the divide between phenotypic measure development and validation and measure end use for brain-phenotype modeling holds the promise that each may inform the other, yielding more precise and useful brain-phenotype models. Such models can in turn be used to reveal the macroscale neural bases of a given phenotype, advancing basic neuroscientific understanding and identifying circuits that can be targeted (eg, via closed-loop neurofeedback or brain stimulation) to slow, reverse, or even prevent functional impairment.

JAMA Psychiatry. 2023;80(8):848-854. doi:10.1001/jamapsychiatry.2023.1419
Published online June 14, 2023.

Viewpoint page 765

Author Affiliations: Interdepartmental Neuroscience Program, Yale School of Medicine, New Haven, Connecticut (Greene, Constable); MD/PhD program, Yale School of Medicine, New Haven, Connecticut (Greene); Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut (Constable); Department of Biomedical Engineering, Yale School of Engineering and Applied Science, New Haven, Connecticut (Constable); Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut (Constable).

Corresponding Authors: Abigail S. Greene, PhD (abigail.greene@yale.edu) and R. Todd Constable, PhD (todd.constable@yale.edu), Yale University, The Anlyan Center, 300 Cedar St, New Haven, CT 06520.

How we think, feel, and behave—the manifestation of fundamental cognitive processes[1]—makes us who we are and creates the axes along which individuals traverse health and psychiatric disease. A recognition of the importance of such processes (eg, working memory, inhibitory control) to disease, and in particular the notion that each such process can be transdiagnostically mapped to behavior and neural circuitry, has motivated recent reconceptualizations of psychopathology, such as the Research Domain Criteria.[1] Assessing the association of phenotypic measures that reflect these processes with the brain thus presents an opportunity for basic and translational research; such work promises to reveal insights into the pathophysiology of psychiatric disease that complement symptom-driven, categorical diagnostic frameworks and that identify novel targets for intervention. Advances in modeling techniques to associate brain with phenotype have correspondingly generated much excitement, particularly in their potential for clinical utility.[2,3] However, this potential has largely not borne out.[4]

Herein, we propose that a significant contributor to the as yet limited practical utility of brain-phenotype modeling is the siloed nature of 2 fields, psychometrics and computational, cognitive neuroscience, which has yielded brain-based models that reflect phenotypic measure bias.[5] That is, much work has been devoted to psychometric characterization of a wide range of phenotypic measures and separately to the association of existing phenotypic measures with patterns of brain activity, but these literatures have largely remained separate. First, we briefly review work in each domain and then explore how these literatures can inform each other to most usefully associate phenotype with the brain.

How Do We Operationalize Phenotype?

Phenotypic measures are tremendously diverse, capturing observable characteristics that range from the physiological to the morphological to the behavioral and clinical. We focus our discussion on measures that reflect fundamental cognitive processes relevant to health and disease, as does the majority of brain-phenotype modeling work. This category alone is quite broad, encompassing self-report tools, neuropsychological tests, and cognitive tasks. We use examples of each to illustrate concepts critical to brain-phenotype modeling, with reference to more comprehensive treatments where relevant.

Many phenotypic measures have extensive psychometric literatures, which explore their reliability and validity. Both are crucial to a measure's utility: reliability, or the extent to which a tool yields

consistent results (across test items, raters, administrations, or instruments), places an upper limit on measure validity[6-8] and provides a quantification of measurement error; validity, or the degree to which a tool measures what it is intended to measure, is critical to ensure that the modeled outcome and its corresponding brain activity pattern are interpretable. Imagine, for example, a study that seeks differences in brain activity patterns in tall and short children from early childhood to adolescence. Age is, of course, causally related to height in children, so without adequately controlling for age, a brain activity pattern found to be related to height is likely, in part or full, a pattern relevant to development, not to height per se. Similarly, an unreliable stadiometer may yield height estimates so inaccurate that they cannot be meaningfully related to brain activity. Both reliability and validity are thus important for brain-phenotype modeling, although reliability need not always track predictive utility.[6] We turn next to a brief discussion of each of these psychometric concepts.

Reliability and Reproducibility

In response to concerns about a crisis of reproducibility[9] in psychology and human neuroscience,[10] attention to measure reliability is growing, with respect to neuroimaging-derived measures[7,11,12] and to phenotypic measures (eg, reflecting psychological phenomena[10]). The topic has long been seen as crucial to evaluation of neuropsychological instruments.[13,14] Much work has sought to explain cases of low reliability, focusing on such factors as research practices and publication incentives.[9,10,15]

Indeed, discussions of the appropriateness of a given measure for brain-phenotype modeling generally focus on its reliability, as, for example, in recent work exploring the relative reliability of self-report and behavioral task measures of self-regulation.[16] However, while measure reliability places an upper limit on predictive utility, reliability and predictive utility are dissociable,[6] and decreased between-individual variability may in some cases enhance individual differences of interest.[17,18] Finally, a perfectly reliable measure may still have low predictive utility if it does not meet modeling distributional assumptions[19] or if it does not measure the construct of interest. We turn next to this related topic of measure validity.

Validity

Many open questions regarding validity of phenotypic measures have important implications for their use in brain-phenotype modeling analyses. First, constructs defined independently of in vivo brain measurements may not map directly or meaningfully onto neurobiology,[20] an issue that has motivated recent efforts to restructure neuropsychological and clinical ontologies, such as the Research Domain Criteria.[1,21] Further, even a neurobiologically meaningful construct may prove problematic if its operationalization differs across fields, yielding low criterion validity. Such differences may be observed across measurement approaches (eg, self-report–, neuropsychological test–, and task performance–based measures of attentional control[22] or self-report and interview administrations[23]) or even within an administration method (eg, on a questionnaire, such factors as question order,[24] mode of administration,[23,25] culture,[26] and gender[27] can dramatically affect results). Conversely, independent development of constructs in siloed fields may cause issues of construct redundancy that compromise validity.[28]

Second, many tools for phenotypic measurement are accompanied by norms and score transforms intended to classify ranges of normal and abnormal function and to facilitate comparison across studies. However, this approach often encourages the inappropriate pathologization of variance,[29] and there is no consensus on how to operationalize normal function across populations and time. Efforts to do so are further complicated by the enormity of the task; globally representative or population-specific norms require large quantities of normative data to be acquired in all populations. In practice, many effects are demonstrated and tests normed in WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations, which are among the least representative in the world.[30] More inclusive and generalizable norm development is an ongoing effort[8,14] that is not without controversy. For example, some have suggested the use of race-based norms, raising concerns that such an approach may be used to justify differential treatment of race-based groups. Further, race-specific norms may obscure the variables for which race is a bad proxy, suggesting that race, a nonscientific, nonbiological, nonspecific social construct,[31] is itself the cause of racial disparities in phenotypic measures.[32]

Such disparities and evidence of cultural bias in neuropsychological testing are well documented,[14,33,34] leading to the development of a cross-cultural neuropsychology movement[35] and a third issue of validity: tests may only measure the construct of interest in the group in which the test was developed and normed. Test performance is affected by such factors as acculturation,[36] language,[37] education quality,[38] and poverty and socioeconomic status (SES).[39] These relationships are complex, and these factors in many cases represent substantial diversity (eg, SES is itself a proxy for such factors as environmental exposures, nutrition, and stress),[29] further complicating efforts to identify the drivers of test performance in a given group. The construct itself may also differ meaningfully across groups.[33]

In sum, such work highlights the importance of asking the question, what does this tool measure in the individual sitting before me? Assumptions that tests reflect unitary cognitive constructs, stable traits, or innate ability will frequently mislead. Rather, we must view phenotypic measures as complex and continuously evolving as populations change, more inclusive normative data are collected, and more mechanistic understandings of environmental influences on construct measurement are developed. Such a framing is crucial to the productive link of phenotype to brain, a topic to which we turn later on. First, however, we offer a brief review of methods currently in use to link brain to phenotype.

How Do We Link Existing Phenotypic Measures to the Brain?

The goal of relating phenotypic measures to the brain is by no means a new one. Indeed, the earliest neuropsychological tests were developed for this purpose, revealing stereotyped functional deficits associated with specific brain lesions.[14] While such an approach has at times been used pseudoscientifically to promote harmful prejudices, it has, when used responsibly, also offered crucial insights into the modularity of functional brain organization.[40] Recent advances in computing and statistical tools have added to this understanding, nuancing the idea of a 1-to-1 mapping from brain region to

function. Even within regions that demonstrate preferential activation to a particular stimulus type, such as the fusiform face area,[41] a diverse array of face and object stimuli is represented, suggesting that their neural representations are broadly distributed, overlapping, and distinct even in areas that do not maximally respond to their presentation.[42] This finding led to the development of multivoxel pattern analysis,[43] work questioning the stability of regional boundaries,[44] and a move away from the dogma that each brain region can be ascribed a unitary function. Further, complex phenotypes are increasingly understood to be represented not by focally specialized regions alone, but rather by networks distributed across the brain.[45,46] Whole-brain functional connectivity (FC), a network-based approach, is thus well suited for studies assessing the link between brain activity and complex, continuous phenotypes. Further, patterns of FC are distinct across individuals and stable within individuals, crucial to their utility for phenotypic association.[47,48] For these reasons, we focus primarily on work studying the association of FC-based brain measures with neurocognitive phenotypes.

While many approaches have been taken to study the association of brain with phenotype, one that has been particularly fruitful is brain-based predictive modeling of phenotypic measures.[47,49] We here define prediction as any analysis that tests a model on unseen data, and thus, unlike explanatory analyses that train and test a model on the same data, reveals overfitting that limits model generalizability.[50] While increasing availability of large, publicly available data sets has offered some protection against overfitting, it remains a common problem in FC analysis given the number of pairwise brain interactions (usually tens of thousands) relative to the sample size[50] of even the largest consortium data sets (hundreds to thousands,[51] with the notable exception of the UK Biobank[52]),[53] making prediction particularly useful for FC-phenotype modeling. Indeed, predictive modeling has identified, in a data-driven fashion, macroscale neural circuitry associated with a range of traits,[47,54,55] behaviors,[56-58] psychopathology,[59-61] clinical risk factors,[62] and treatment outcomes.[63,64]

Algorithmically, prediction can take many forms. Models can be linear or nonlinear, although common neuroimaging sample sizes as well as the increased computational tractability, increased interpretability, and largely comparable performance of linear models[65-67] have made them a popular choice. While prediction can be used to associate brain activity with categorical or continuous phenotypic measures,[19] it is well suited to the latter, promoting the kinds of dimensional, transdiagnostic analyses that are necessary to validate and refine the Research Domain Criteria[21] and related conceptualizations of psychopathology. Models can be validated internally (ie, cross-validation) or externally, performance can be evaluated in various ways, and many approaches exist to investigate what is driving the resulting models, mapping predictive utility back to brain features[19,43,68]; despite these differences, predictive modeling approaches are unified by their strict separation of training and test data, offering protection against relationship overestimation.[69,70]

As a family of approaches, prediction is not without limitations. First, model performance may face an upper limit defined by irreducible error; that is, the blood oxygenation level–dependent signal may not contain all of the information required to predict a given outcome.[71] Predictive models may therefore be useful but should not be expected to completely explain a given outcome. Further, a successful predictive model may prove challenging to interpret; predictors are often many, weak, and selected algorithmically, in contrast to traditional statistical approaches that have preference to a smaller set of important, often knowledge-guided,[69] covariates.[72] That such weak predictors can often be variably combined to yield comparably high prediction accuracy underscores that improved prediction accuracy may not track degree of biological insight. Indeed, systematic comparisons of inference and prediction have demonstrated that variables found to be important by null hypothesis testing may not yield accurate outcome predictions, particularly at common sample sizes.[69,73] Such results have been used to suggest that explanatory and predictive analyses may be complementary, with the former optimized to generate pathophysiological insight across individuals, and the latter to generate accurate, pathophysiology-agnostic predictions for a given individual.[73] We note that this distinction is in many cases a practical one; predictive and biologically relevant variables may be expected to demonstrate increasing overlap with methodological and technical advances, such as the growing sample sizes[73] made possible by publicly available data sets, and with intentional study design (eg, prioritizing model interpretability or comparing full and partial models).[50] While predictive modeling may also, and more imminently, be used to predict clinical end points without biological interpretation, it may thus be possible to use the benefits of prediction to reveal more robust, generalizable brain-phenotype relationships that shed light on neurobiology and pathophysiological mechanism. We focus here on this potential but argue that models may not always be predicting what we think they are predicting, presenting another important limitation to predictive modeling approaches. That is, phenotypic measures used as outcomes in these analyses are often confounded, biased, and differentially valid across groups. These properties affect their interpretation and the interpretation and utility of corresponding brain-based models, a topic to which we turn next.

Reaching Across Literatures to Increase the Accuracy, Precision, and Utility of Brain-Phenotype Models

The past several decades have witnessed tremendous advances in phenotypic measurement and evaluation (including the decomposition of psychological and philosophical concepts into concrete, theory-based elemental units[74]), as well as in modeling techniques to relate brain to phenotype. This progress has yielded many new measures to characterize phenotype, and increasing concern about reproducibility has supported the development of a robust corresponding psychometric literature. In parallel, an exciting movement encouraging result transparency and sharing[11,12,51] has increased the accessibility of neuroimaging and phenotypic data from large samples and of tools to relate them via predictive modeling.

But despite these advances, or perhaps in part because of them, the psychometric and neuroscientific communities remain relatively siloed. For example, predictive modeling articles often limit discussion of potential confounds to in-scanner head motion and straightforward demographic variables such as sex and age[19,75] that are widely understood to affect brain measures.[76,77] Questions of phenotypic measure appropriateness, whether due to reliability, validity, or both, for brain-phenotype modeling are only beginning to be asked[16] and are not routinely used to guide dependent variable

choices in such modeling. This is problematic because measure replication failures abound[10] and biased, confounded tests remain in common use.[29,33]

What, then, are the consequences of relating such measures to the brain? The case of an unreliable measure is relatively straightforward and has already been addressed here; assuming adequate test validity, high reliability should often be sought to maximize measure utility for individual differences research.[6,16,70,78] Often, validity, particularly construct validity, is low, the construct itself has variable validity or interpretations across groups, or the construct is associated with different neural circuits across groups. We explore each of these possibilities below.

We begin with the question of overall measure validity. As described earlier, biases in commonly used phenotypic measures are well documented. In recent work, we demonstrated that brain-based predictive models are predicting not unitary cognitive constructs, but rather neurocognitive measures intertwined with constellations of sociodemographic and clinical variables.[5] Across 3 data sets and a range of phenotypic measures, predictive models systematically failed in individuals who defy stereotypical profiles of high and low scorers. These results demonstrate limited model generalizability and that bias in phenotypic measures may be reflected in corresponding brain models, consistent with the findings that variables such as SES are associated with functional network organization and cortical structure into adulthood,[79-81] and that brain measures mediate associations between SES and various outcomes, such as negative mood.[82] A related issue is that phenotypic measures, often designed to capture complex behaviors, may not map onto single cognitive constructs and corresponding neurobiological systems.[20]

That covariates may track outcome in some individuals but not others demonstrates the second possibility of differential measure validity or interpretation across groups. The construct of verbal memory offers an illustrative example. While the articulation rate of subvocal rehearsal tracks digit and word span performance in native English speakers, this association is attenuated in native Mandarin speakers, suggesting an important difference in the cognitive processes associated with task execution across groups.[83] The interpretation of performance, and thus of the brain activity pattern associated with it, must therefore depend on the native language of the sample, underscoring the need for group-specific brain-phenotype models.

Such group-specific models may also be required in the case of possibility 3: the same construct corresponds to group-specific neural circuitry, perhaps reflecting different task strategies or cognitive styles (consistent with the idea of multiple realizability[84]). For example, individuals under acute stress demonstrate distinct neural activity patterns when faced with a navigation planning task,[85] a finding with clear implications for patients experiencing traumatic sequelae. Such group differences may be compensatory, as suggested by the finding that older adults with high SES engage a distinct neural signature during a contextual memory task and are spared age-related decline in task performance.[86]

Overall, the implications of modeling phenotypic measures with low or variable reliability and validity are 2-fold: (1) limited model generalizability and (2) unhelpful or, at worst, harmful model interpretation. Limited phenotypic measure generalizability yields a corresponding model that may not work for a given individual, and this poses substantial limitations to its practical, and particularly clinical, applications. Second, and even more concerningly, a model that is based on a biased or imprecisely defined phenotypic measure may perpetuate harmful biases, as has been described in diverse applications of machine learning algorithms, from criminal justice[87,88] to health care,[89,90] and recently to brain-phenotype modeling.[5] These concerns, limited generalizability and bias, represent substantial obstacles to transitioning brain-phenotype models to applied contexts.

Practical Recommendations and Implications

To overcome these obstacles will require an approach that acknowledges and addresses the properties of phenotypic measures of interest. Specifically, we propose 4 key processes to be incorporated into predictive modeling pipelines:

1. Carefully select phenotypes and corresponding outcome measures to maximize relevance and minimize bias. Future brain-phenotype modelers must first, with the guidance of clinicians and neuropsychologists, select phenotypes of interest. They must then select corresponding outcome measures, using relevant psychometric literatures as a guide. Standardized resources to assess the risk of bias (eg, PROBAST [Prediction Model Risk of Bias Assessment Tool][91]) may be used, test administration practices should minimize bias,[8,33,34] and, when generating novel measures, recommendations to minimize bias should be followed.[8,33] Modelers must then seek to understand the stereotypical profile for a given score in their sample to refine the battery and understand what the model is predicting. This will likely require the collection of more comprehensive demographic and clinical information than is typical (for general guidelines, see Hughes et al[92]). A move toward deep phenotyping in large, publicly available population data sets will further support this goal.[93]
2. Deconfound phenotypic measures. When it is not possible to choose a less biased measure, biases should be identified and corrected. Recent work provides systematic assessments of confound correction approaches and should be used to guide data cleaning, particularly for cases in which confounds exist for only a subset of the sample.[94-97]
3. Interpret brain-based models more precisely and use resulting models to advance phenotypic measure development. It will likely be impossible to eliminate or correct all possible confounds.[98] Thus, researchers must acknowledge that the model is predicting a composite outcome and interpret resulting brain models accordingly. Relatedly, interrogating resulting brain-based models may further our understanding of the links between measures and (the likely multiple) relevant latent neurobiological processes,[99] yielding additional insights into phenotypic measure ontology and encouraging an iterative cycle of measure and model adjustment.
4. Build and use more precise, group-specific models. Having eliminated confounds to the extent possible and defined the profile that is being predicted, modelers must characterize brain-phenotype relationships in those who defy this profile. Practically, these individuals will, as demonstrated in Greene et al,[5] require a distinct model. Developing subtype-specific models will

ensure that a given model applies to a given individual, with the potential to explain observed disease heterogeneity[100] and inform both diagnosis and treatment. Studying these subtype-specific models may yield insights into alternative cognitive processes that are recruited in subsets of individuals,[84,101] mechanisms of resilience, and neuromarkers of specific obstacles to performance.

Conclusions

By accounting for known and uncovered biases and leveraging resulting group-specific models, such work holds the promise of yielding more complete, precise, individualized, and useful brain-phenotype relationships. This will permit brain-based predictive modeling to transition to clinical applications, which may include using brain models as targets for intervention (eg, via neurofeedback[102,103] or brain stimulation[104] or as proxies for treatment progress or effect[105,106]), or prospectively using brain activity to predict disease progression[107] or treatment response.[64] These are exciting ideas, but they will only be possible if we develop models of the constructs of interest and ensure that they apply to the patient sitting before us. The framework proposed here to integrate psychometrics and cognitive, computational neuroscience lays the foundation for such a future in which the neural bases of a given phenotype are known and can be targeted to reverse, slow, or even prevent functional impairment.

ARTICLE INFORMATION

Accepted for Publication: March 27, 2023.

Published Online: June 14, 2023. doi:10.1001/jamapsychiatry.2023.1419

Conflict of Interest Disclosures: Dr Greene reported grants from the National Institutes of Health during the conduct of the study. No other disclosures were reported.

Funding/Support: Preparation, review, or approval of the manuscript was supported by funding from the National Institutes of Health (grants GM007205 and TR001864 to Dr Greene and MH121095 to Dr Constable).

Role of the Funder/Sponsor: The funder had no role in the preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Insel TR. The NIMH Research Domain Criteria (RDoC) project: precision medicine for psychiatry. *Am J Psychiatry*. 2014;171(4):395-397. doi:10.1176/appi.ajp.2014.14020138

2. Gabrieli JDE, Ghosh SS, Whitfield-Gabrieli S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*. 2015;85(1):11-26. doi:10.1016/j.neuron.2014.10.047

3. Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 2017;20(3):365-377. doi:10.1038/nn.4478

4. Zhang J, Kucyi A, Raya J, et al. What have we really learned from functional connectivity in clinical populations? *Neuroimage*. 2021;242(August):118466. doi:10.1016/j.neuroimage.2021.118466

5. Greene AS, Shen X, Noble S, et al. Brain-phenotype models fail for individuals who defy sample stereotypes. *Nature*. 2022;609(7925):109-118. doi:10.1038/s41586-022-05118-w

6. Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. Influences on the test–retest reliability of functional connectivity MRI and its relationship with behavioral utility. *Cereb Cortex*. 2017;27(11):5415-5429. doi:10.1093/cercor/bhx230

7. Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *Neuroimage*. 2019;203:116157. doi:10.1016/j.neuroimage.2019.116157

8. *The Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014.

9. Baker M. 1,500 Scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452-454. doi:10.1038/533452a

10. Aarts AA, Anderson JE, Anderson CJ, et al; Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):aac4716. doi:10.1126/science.aac4716

11. Poldrack RA, Baker CI, Durnez J, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci*. 2017;18(2):115-126. doi:10.1038/nrn.2016.167

12. Nichols TE, Das S, Eickhoff SB, et al. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat Neurosci*. 2017;20(3):299-303. doi:10.1038/nn.4500

13. Calamia M, Markon K, Tranel D. The robust reliability of neuropsychological measures: meta-analyses of test-retest correlations. *Clin Neuropsychol*. 2013;27(7):1077-1105. doi:10.1080/13854046.2013.809795

14. Casaletto KB, Heaton RK. Neuropsychological assessment: past and future. *J Int Neuropsychol Soc*. 2017;23(9-10):778-790. doi:10.1017/S1355617717001060

15. Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. *Sci Adv*. 2021;7(21):eabd1705. doi:10.1126/sciadv.abd1705

16. Enkavi AZ, Eisenberg IW, Bissett PG, et al. Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc Natl Acad Sci U S A*. 2019;116(12):5472-5477. doi:10.1073/pnas.1818430116

17. Finn ES, Scheinost D, Finn DM, Shen X, Papademetris X, Constable RT. Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage*. 2017;160:140-151. doi:10.1016/j.neuroimage.2017.03.064

18. Greene AS, Gao S, Noble S, Scheinost D, Constable RT. How tasks change whole-brain functional organization to reveal brain-phenotype relationships. *Cell Rep*. 2020;32(8):108066. doi:10.1016/j.celrep.2020.108066

19. Scheinost D, Noble S, Horien C, et al. Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage*. 2019;193:35-45. doi:10.1016/j.neuroimage.2019.02.057

20. Buzsáki G. The brain-cognitive behavior problem: a retrospective. *eNeuro*. 2020;7(4):1-8. doi:10.1523/ENEURO.0069-20.2020

21. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748-751. doi:10.1176/appi.ajp.2010.09091379

22. Williams PG, Rau HK, Suchy Y, Thorgusen SR, Smith TW. On the validity of self-report assessment of cognitive abilities: attentional control scale associations with cognitive performance, emotional adjustment, and personality. *Psychol Assess*. 2017;29(5):519-530. doi:10.1037/pas0000361

23. Cook C. Mode of administration bias. *J Man Manip Ther*. 2010;18(2):61-63. doi:10.1179/106698110X12640740712617

24. Schuman H, Presser S. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. SAGE Publications; 1996.

25. Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health (Oxf)*. 2005;27(3):281-291. doi:10.1093/pubmed/fdi031

26. Bardwell WA, Dimsdale JE. The impact of ethnicity and response bias on the self-report of negative affect. *J Appl Biobehav Res*. 2001;6(1):27-38. doi:10.1111/j.1751-9861.2001.tb00105.x

27. Hebert JR, Ma Y, Clemow L, et al. Gender differences in social desirability and social approval bias in dietary self-report. *Am J Epidemiol*. 1997;146(12):1046-1055. doi:10.1093/oxfordjournals.aje.a009233

28. Hodson G. Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *J Theor Soc Psychol*. 2021;5(4):576-590. doi:10.1002/jts5.120

29. Pitts-Taylor V. Neurobiologically poor? Brain phenotypes, inequality, and biosocial determinism. *Technol Hum Values*. 2019;44(4):660-685. doi:10.1177/0162243919841695

30. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? *Behav Brain Sci*. 2010;33(2-3):61-83. doi:10.1017/S0140525X0999152X

31. Williams DR. The concept of race in health services research: 1966 to 1990. *Health Serv Res*. 1994;29(3):261-274.

32. Gasquoine PG. Race-norming of neuropsychological tests. *Neuropsychol Rev*. 2009; 19(2):250-262. doi:10.1007/s11065-009-9090-5

33. Fernández AL, Abe J. Bias in cross-cultural neuropsychological testing: problems and possible solutions. *Cult Brain*. 2018;6:1-35. doi:10.1007/s40167-017-0050-2

34. Manly JJ. Critical issues in cultural neuropsychology: profit from diversity. *Neuropsychol Rev*. 2008;18(3):179-183. doi:10.1007/s11065-008-9068-8

35. Fletcher-Janzen E, Strickland TL, Reynolds CR. *Handbook of Cross-Cultural Neuropsychology*. Kluwer Academic/Plenum Publishers; 2000. doi:10.1007/978-1-4615-4219-3

36. Manly JJ, Miller SW, Heaton RK, et al; The HIV Neurobehavioral Research Center (HNRC) Group. The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. *J Int Neuropsychol Soc*. 1998;4(3):291-302. doi:10.1017/S1355617798002914

37. Flores I, Casaletto KB, Marquine MJ, et al. Performance of Hispanics and Non-Hispanic Whites on the NIH Toolbox Cognition Battery: the roles of ethnicity and language backgrounds. *Clin Neuropsychol*. 2017;31(4):783-797. doi:10.1080/13854046.2016.1276216

38. Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and White elders. *J Int Neuropsychol Soc*. 2002;8(3):341-348. doi:10.1017/S1355617702813157

39. Farah MJ. The neuroscience of socioeconomic status: correlates, causes, and consequences. *Neuron*. 2017;96(1):56-71. doi:10.1016/j.neuron.2017.08.034

40. Greenblatt SH. Phrenology in the science and culture of the 19th century. *Neurosurgery*. 1995; 37(4):790-804. doi:10.1227/00006123-199510000-00025

41. Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 1997;17(11):4302-4311. doi:10.1523/JNEUROSCI.17-11-04302.1997

42. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 2001;293(5539):2425-2430. doi:10.1126/science.1063736

43. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*. 2006;10(9):424-430. doi:10.1016/j.tics.2006.07.005

44. Salehi M, Greene AS, Karbasi A, Shen X, Scheinost D, Constable RT. There is no single functional atlas even for a single individual: functional parcel definitions change with task. *Neuroimage*. 2020;208:116366. doi:10.1016/j.neuroimage.2019.116366

45. Turk-Browne NB. Functional interactions as big data in the human brain. *Science*. 2013;342(6158): 580-584. doi:10.1126/science.1238409

46. Goodkind M, Eickhoff SB, Oathes DJ, et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*. 2015;72(4):305-315. doi:10.1001/jamapsychiatry.2014.2206

47. Finn ES, Shen X, Scheinost D, et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015;18(11):1664-1671. doi:10.1038/nn.4135

48. Horien C, Shen X, Scheinost D, Constable RT. The individual functional connectome is unique and stable over months to years. *Neuroimage*. 2019; 189:676-687. doi:10.1016/j.neuroimage.2019.02.002

49. Shen X, Finn ES, Scheinost D, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc*. 2017;12(3):506-518. doi:10.1038/nprot.2016.178

50. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect Psychol Sci*. 2017;12(6):1100-1122. doi:10.1177/1745691617693393

51. Horien C, Noble S, Greene AS, et al. A hitchhiker's guide to working with large, open-source neuroimaging datasets. *Nat Hum Behav*. 2021;5(2):185-193. doi:10.1038/s41562-020-01005-4

52. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

53. Whelan R, Garavan H. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol Psychiatry*. 2014;75(9):746-748. doi:10.1016/j.biopsych.2013.05.014

54. Dubois J, Galdi P, Paul LK, Adolphs R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1756):20170284. doi:10.1098/rstb.2017.0284

55. Hsu W-T, Rosenberg MD, Scheinost D, Constable RT, Chun MM. Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Soc Cogn Affect Neurosci*. 2018; 13(2):224-232. doi:10.1093/scan/nsy002

56. Rosenberg MD, Finn ES, Scheinost D, et al. A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci*. 2016;19(1):165-171. doi:10.1038/nn.4179

57. Avery EW, Yoo K, Rosenberg MD, et al. Distributed patterns of functional connectivity predict working memory performance in novel healthy and memory-impaired individuals. *J Cogn Neurosci*. 2020;32(2):241-255. doi:10.1162/jocn_a_01487

58. Stark GF, Avery EW, Rosenberg MD, et al. Using functional connectivity models to characterize relationships between working and episodic memory. *Brain Behav*. 2021;11(8):e02105. doi:10.1002/brb3.2105

59. Barron DS, Gao S, Dadashkarimi J, et al. Transdiagnostic, connectome-based prediction of memory constructs across psychiatric disorders. *Cereb Cortex*. 2021;31(5):2523-2533. doi:10.1093/cercor/bhaa371

60. Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. *N Engl J Med*. 2013;368 (15):1388-1397. doi:10.1056/NEJMoa1204471

61. Mihalik A, Ferreira FS, Rosa MJ, et al; NSPN Consortium. Brain-behaviour modes of covariation in healthy and clinically depressed young people. *Sci Rep*. 2019;9(1):11536. doi:10.1038/s41598-019-47277-3

62. Rapuano KM, Rosenberg MD, Maza MT, et al. Behavioral and brain signatures of substance use vulnerability in childhood. *Dev Cogn Neurosci*. 2020;46:100878. doi:10.1016/j.dcn.2020.100878

63. Yip SW, Scheinost D, Potenza MN, Carroll KM. Connectome-based prediction of cocaine abstinence. *Am J Psychiatry*. 2019;176(2):156-164. doi:10.1176/appi.ajp.2018.17101147

64. Drysdale AT, Grosenick L, Downar J, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*. 2017;23(1):28-38. doi:10.1038/nm.4246

65. He T, Kong R, Holmes AJ, et al. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *Neuroimage*. 2020;206:116276. doi:10.1016/j.neuroimage.2019.116276

66. Schulz M-A, Yeo BTT, Vogelstein JT, et al. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nat Commun*. 2020;11(1):4238. doi:10.1038/s41467-020-18037-z

67. Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*. 2003;19(2 Pt 1):261-270. doi:10.1016/S1053-8119(03)00049-1

68. Dadi K, Rahim M, Abraham A, et al; Alzheimer's Disease Neuroimaging Initiative. Benchmarking functional connectome-based predictive models for resting-state fMRI. *Neuroimage*. 2019;192:115-134. doi:10.1016/j.neuroimage.2019.02.062

69. Bzdok D, Ioannidis JPA. Exploration, inference, and prediction in neuroscience and biomedicine. *Trends Neurosci*. 2019;42(4):251-262. doi:10.1016/j.tins.2019.02.001

70. Rosenberg MD, Finn ES. How to establish robust brain-behavior relationships without thousands of individuals. *Nat Neurosci*. 2022;25(7): 835-837. doi:10.1038/s41593-022-01110-9

71. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning with Applications in R*. 2nd ed. Springer Link; 2021. doi:10.1007/978-1-0716-1418-1

72. Efron B. Prediction, estimation, and attribution. *Int Stat Rev*. 2020;88(S1):S28-S59. doi:10.1111/insr.12409

73. Bzdok D, Engemann D, Thirion B. Inference and prediction diverge in biomedicine. *Patterns (N Y)*. 2020;1(8):100119. doi:10.1016/j.patter.2020.100119

74. Poeppel D, Adolfi F. Against the epistemological primacy of the hardware: the brain from inside out, turned upside down. *eNeuro*. 2020;7(4):1-8. doi:10.1523/ENEURO.0215-20.2020

75. Greene AS, Gao S, Scheinost D, Constable RT. Task-induced brain state manipulation improves prediction of individual traits. *Nat Commun*. 2018;9 (1):2807. doi:10.1038/s41467-018-04920-3

76. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*. 2012;59(3):2142-2154. doi:10.1016/j.neuroimage.2011.10.018

77. Scheinost D, Finn ES, Tokoglu F, et al. Sex differences in normal age trajectories of functional brain networks. *Hum Brain Mapp*. 2015;36(4):1524-1535. doi:10.1002/hbm.22720

78. Dubois J, Galdi P, Han Y, Paul LK, Adolphs R. Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personal Neurosci*. 2018;1:e6. doi:10.1017/pen.2018.8

79. Chan MY, Na J, Agres PF, Savalia NK, Park DC, Wig GS. Socioeconomic status moderates age-related differences in the brain's functional network organization and anatomy across the adult lifespan. *Proc Natl Acad Sci U S A*. 2018;115(22):E5144-E5153. doi:10.1073/pnas.1714021115

80. Kweon H, Aydogan G, Dagher A, et al. Human brain anatomy reflects separable genetic and environmental components of socioeconomic status. *Sci Adv*. 2022;8(20):eabm2923. doi:10.1126/sciadv.abm2923

81. Poeppl TB, Dimas E, Sakreida K, et al. Pattern learning reveals brain asymmetry to be linked to socioeconomic status. *Cereb Cortex Commun*. 2022;3(2):tgac020. doi:10.1093/texcom/tgac020

82. Barch D, Pagliaccio D, Belden A, et al. Effect of hippocampal and amygdala connectivity on the relationship between preschool poverty and school-age depression. *Am J Psychiatry*. 2016;173(6):625-634. doi:10.1176/appi.ajp.2015.15081014

83. Mattys SL, Baddeley A, Trenkic D. Is the superior verbal memory span of Mandarin speakers due to faster rehearsal? *Mem Cognit*. 2018;46(3):361-369. doi:10.3758/s13421-017-0770-8

84. Pitts-Taylor V. *The Brain's Body*. Duke University Press; 2016.

85. Brown TI, Gagnon SA, Wagner AD. Stress disrupts human hippocampal-prefrontal function during prospective spatial navigation and hinders flexible behavior. *Curr Biol*. 2020;30(10):1821-1833.e8. doi:10.1016/j.cub.2020.03.006

86. Czernochowski D, Fabiani M, Friedman D. Use it or lose it? SES mitigates age-related decline in a recency/recognition task. *Neurobiol Aging*. 2008;29(6):945-958. doi:10.1016/j.neurobiolaging.2006.12.017

87. Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. *Sci Adv*. 2018;4(1):eaao5580. doi:10.1126/sciadv.aao5580

88. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica. Published online May 23, 2016. Accessed May 8, 2023. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

89. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell*. 2021;3:199-217. doi:10.1038/s42256-021-00307-0

90. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342

91. Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58. doi:10.7326/M18-1376

92. Hughes JL, Camden AA, Yangchen T, College AS. Rethinking and updating demographic questions: guidance to improve descriptions of research samples. *Psi Chi J Psychol Res*. 2016;21(3):138-151. doi:10.24839/2164-8204.JN21.3.138

93. Kopal J, Bzdok D. Endorsing complexity through diversity: computational psychiatry meets big data analytics. *Biol Psychiatry*. 2023;93(8):655-657. doi:10.1016/j.biopsych.2022.07.023

94. Dinga R, Schmaal L, Penninx BWJH, Veltman DJ, Marquand AF. Controlling for effects of confounding variables on machine learning predictions. *bioRxiv*. August 2020. doi:10.1101/2020.08.17.255034

95. Linn KA, Gaonkar B, Doshi J, Davatzikos C, Shinohara RT. Addressing confounding in predictive models with an application to neuroimaging. *Int J Biostat*. 2016;12(1):31-44. doi:10.1515/ijb-2015-0030

96. Zhao Q, Adeli E, Pohl KM. Training confounder-free deep learning models for medical applications. *Nat Commun*. 2020;11(1):6010. doi:10.1038/s41467-020-19784-9

97. Alfaro-Almagro F, McCarthy P, Afyouni S, et al. Confound modelling in UK Biobank brain imaging. *Neuroimage*. 2021;224:117002. doi:10.1016/j.neuroimage.2020.117002

98. Rao A, Monteiro JM, Mourao-Miranda J, Initiative AD; Alzheimer's Disease Initiative. Predictive modelling using neuroimaging data in the presence of confounds. *Neuroimage*. 2017;150:23-49. doi:10.1016/j.neuroimage.2017.01.066

99. Poldrack RA, Kittur A, Kalar D, et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front Neuroinform*. 2011;5:17. doi:10.3389/fninf.2011.00017

100. Lynch CJ, Gunning FM, Liston C. Causes and consequences of diagnostic heterogeneity in depression: paths to discovering novel biological depression subtypes. *Biol Psychiatry*. 2020;88(1):83-94. doi:10.1016/j.biopsych.2020.01.012

101. Schulz M-A, Chapman-Rounds M, Verma M, Bzdok D, Georgatzis K. Inferring disease subtypes from clusters in explanation space. *Sci Rep*. 2020;10(1):12900. doi:10.1038/s41598-020-68858-7

102. deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. *Nat Neurosci*. 2015;18(3):470-475. doi:10.1038/nn.3940

103. Scheinost D, Hsu TW, Avery EW, et al. Connectome-based neurofeedback: a pilot study to improve sustained attention. *Neuroimage*. 2020;212:116684. doi:10.1016/j.neuroimage.2020.116684

104. Siddiqi SH, Kording KP, Parvizi J, Fox MD. Causal mapping of human brain function. *Nat Rev Neurosci*. 2022;23(6):361-375. doi:10.1038/s41583-022-00583-8

105. Rosenberg MD, Zhang S, Hsu W-T, et al. Methylphenidate modulates functional network connectivity to enhance attention. *J Neurosci*. 2016;36(37):9547-9557. doi:10.1523/JNEUROSCI.1746-16.2016

106. Karim HT, Andreescu C, Tudorascu D, et al. Intrinsic functional connectivity in late-life depression: trajectories over the course of pharmacotherapy in remitters and non-remitters. *Mol Psychiatry*. 2017;22(3):450-457. doi:10.1038/mp.2016.55

107. Olde Dubbelink KTE, Stoffers D, Deijen JB, et al. Resting-state functional connectivity as a marker of disease progression in Parkinson's disease: a longitudinal MEG study. *Neuroimage Clin*. 2013;2(1):612-619. doi:10.1016/j.nicl.2013.04.003