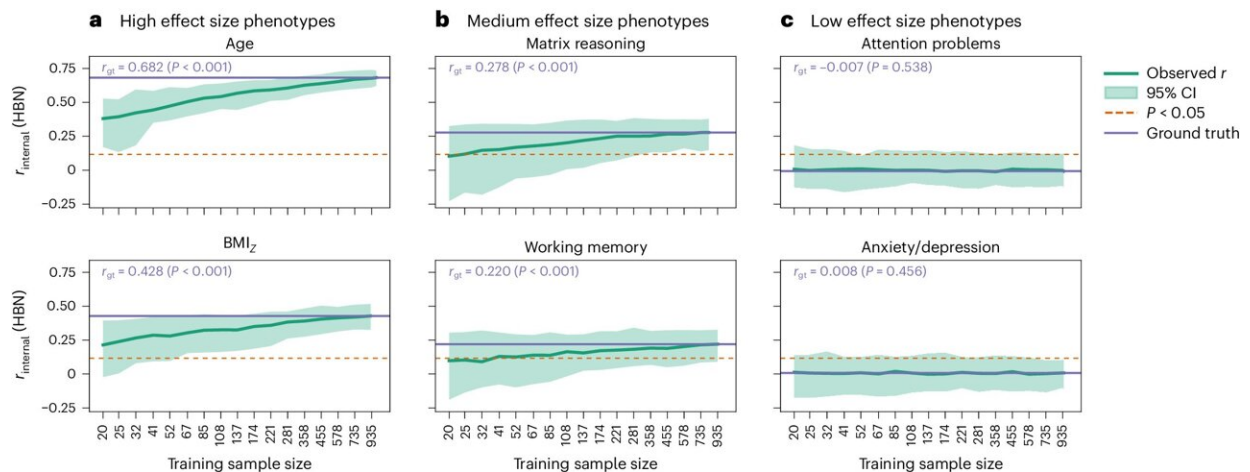


In probing brain-behavior nexus, big datasets are better

August 7 2024, by Mallory Locklear



Internal validation performance in HBN. a–c, Prediction performance in HBN for (a) high (age, body mass index), (b) medium (matrix reasoning, working memory) and (c) low (attention problems, anxiety/depression symptoms) effect size phenotypes. Credit: *Nature Human Behaviour* (2024). DOI: 10.1038/s41562-024-01931-7

When designing machine learning models, researchers first train the models to recognize data patterns and then test their effectiveness. But if the datasets used to train and test aren't sufficiently large, models may appear to be less capable than they actually are, a new Yale study reports.

When it comes to models that identify patterns between the brain and behavior, this could have implications for future research, contribute to the replication crisis affecting psychological research, and hamper understanding of the human brain, researchers say.

The findings were [published](#) July 31 in the journal *Nature Human Behavior*.

Researchers increasingly use machine learning models to uncover patterns that link brain structure or function to, say, cognitive attributes like attention or symptoms of depression. Making these links allows researchers to better understand how the brain contributes to these attributes (and vice versa) and potentially enables them to predict who might be at risk for certain cognitive challenges based on brain imaging alone.

But models are only useful if they're accurate across the general population, not just among the people included in the [training data](#).

Often, researchers will split one dataset into a larger portion on which they train the [model](#) and a smaller portion used to test the model's ability (since collecting two separate sets of data requires greater resources). A growing number of studies, however, have subjected machine learning models to a more rigorous test in order to evaluate their generalizability, testing them on an entirely different dataset made available by other researchers.

"And that's good," said Matthew Rosenblatt, lead author of the study and a graduate student in the lab of Dustin Scheinost, associate professor of radiology and biomedical imaging at Yale School of Medicine. "If you can show something works in a totally different dataset, then it's probably a robust brain-behavior relationship."

Adding another dataset into the mix, however, comes with its own complications—namely, in regard to a study's "[power](#)." Statistical power is the probability that a research study will detect an effect if one exists. For example, a child's height is closely related to their age. If a study is adequately powered, then that relationship will be observed. If the study is "low-powered," on the other hand, there's a higher risk of overlooking the link between age and height.

There are two important aspects to statistical power—the size of the dataset (also known as the sample size) and the effect size. And the smaller that one of those aspects is, the larger the other needs to be. The link between age and height is strong, meaning the effect size is large; one can observe that relationship in even a small dataset. But when the relationship between two factors is more subtle—like, say, age and how well one can sense through touch—researchers would need to collect data from more people to uncover that connection.

While there are equations that can calculate how big a dataset should be to achieve enough power, there aren't any to easily calculate how large two datasets—one training and one testing—should be.

To understand how training and testing dataset sizes affect study power, researchers in the new study used data from six neuroimaging studies and resampled that data over and over, changing the dataset sizes to see how that affected statistical power.

"We showed that statistical power requires relatively large sample sizes for both training and external testing datasets," said Rosenblatt. "When we looked at published studies in the field that use this approach—testing models on a second dataset—we found most of their datasets were too small, underpowering their studies."

Among already published studies, the researchers found that the median

sizes for training and testing datasets were 129 and 108 participants, respectively. For measures with large effect sizes, like age, those dataset sizes were big enough to achieve adequate power. But for measures with medium effect sizes, such as working memory, datasets of those sizes resulted in a 51% chance that the study would not detect a relationship between brain structure and the measure; for measures with low effect sizes, like attention problems, those odds increased to 91%.

"For these measures with smaller effect sizes, researchers may need datasets of hundreds to thousands of people," said Rosenblatt.

As more neuroimaging datasets become available, Rosenblatt and his colleagues expect more researchers will opt to test their models on separate datasets.

"That's a move in the right direction," said Scheinost. "Especially with reproducibility being the problem it is, validating a model on a second, external dataset is one solution. But we want people to think about their dataset sizes. Researchers must do what they can with the data they have, but as more data becomes available, we should all aim to test externally and make sure those test datasets are large."

More information: Matthew Rosenblatt et al, Power and reproducibility in the external validation of brain-phenotype predictions, *Nature Human Behaviour* (2024). [DOI: 10.1038/s41562-024-01931-7](https://doi.org/10.1038/s41562-024-01931-7)

Provided by Yale University

Citation: In probing brain-behavior nexus, big datasets are better (2024, August 7) retrieved 7 August 2024 from <https://medicalxpress.com/news/2024-08-probing-brain-behavior-nexus-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.