NEUROLOGY

Predicting post-stroke aphasia from brain imaging

Stroke can lead to debilitating consequences, including loss of language. An important goal of stroke research is to use machine learning to predict outcomes and response to therapy. A new study compares different approaches to predicting post-stroke outcomes and highlights the need for systematic optimization and validation to ultimately translate scientific insights to clinical settings.

Monica D. Rosenberg and Hayoung Song

After suffering a stroke, approximately 30% of patients experience persistent language deficits, or chronic aphasia[1]. Communication impairments can be devastating. Even when controlling for physical abilities and social support, patients with post-stroke aphasia report worse quality of life than patients without aphasia[2]. Intensive speech and language therapy shows promise for improving outcomes in aphasia[3]. Doctors, however, aren't always able to anticipate which patients could benefit most—that is, who will go on to suffer the worst and longest-lasting impairments or show the pattern of symptoms most amendable to treatment. Thus, an important goal of stroke research is to determine what markers of brain damage best predict patients' symptom constellation, potentially informing treatment outcomes. A new study by Halai and colleagues[4] in *Nature Human Behaviour* works toward this goal by comparing approaches to predicting post-stroke aphasia from magnetic resonance imaging (MRI) measures.

Neuroimaging techniques, including MRI, are potentially valuable tools for personalized medicine because they can provide information about each patient's unique pattern of brain structure, function and connections. Initial efforts to predict post-stroke aphasia from MRI data found that models based on measures of patients' brain structure predicted their aphasia subtype and language function[5,6]. Results aligned with previous findings that lesion size and location, as well as aphasia severity, are particularly important for communication outcomes[7]. Subsequent work from Halai and colleagues identified four orthogonal dimensions of language impairments following stroke[8] and used structural MRI data from 70 patients with chronic post-stroke aphasia to predict aspects of these deficits[9]. Despite these advances, it is not yet clear whether MRI can improve post-stroke aphasia prediction in real-world clinical settings. In other words, can different types of brain scans help clinicians predict whether—and how severely—a patient will go on to suffer different forms of language impairments?

The questions raised by Halai et al. in their new study are important for advancing the clinical efficacy of brain-based models in post-stroke aphasia. Specifically, the authors ask: what features of an individual's brain structure and connections best predict their language impairments following stroke? (For example, do measures of grey and white matter integrity predict post-stroke outcomes? Does adding additional measures of neural integrity improve predictions?) What machine learning algorithm generates the most accurate predictions?

To address these questions, Halai and colleagues analysed structural MRI and diffusion-weighted imaging data from the 70 patients included in their prior work. Structural MRI scans provide information about different types of brain tissue and stroke-related tissue damage, whereas diffusion scans provide information about anatomical brain connections. The researchers built models to predict each of the four language dimensions they introduced previously: phonology, semantics, executive or cognitive skills, and speech fluency. They compared the success of different model 'recipes' built using combinations of six MRI-derived brain measures from multiple brain regions with four machine learning algorithms. Models were defined using ten-fold cross-validation, meaning that they were trained using data from 90% of patients and applied to predict language outcomes in the held-out 10% of patients.

Models with many different recipes predicted patients' aphasia outcomes. Although no single recipe won the prediction contest, some consistent patterns emerged. First, the authors found that while measures of grey and white matter integrity predicted outcomes, adding diffusion data did not improve models' predictive power. In addition, they observed that, across comparisons, Gaussian process regression significantly outperformed the other machine learning algorithms. When moving to a more detailed level of analysis, however, it is important to note that some individual model recipes may have outperformed others due to chance because of the large number of models tested (24 combinations of 6 brain features × 5 brain parcellation schemes × 4 prediction algorithms = 480 model recipes per language component).

The predictive models for post-stroke aphasia described in Halai et al. and in previous work represent intriguing advances. However, all models described here were internally validated, which means that they were trained and tested on subsets of data from the same patient sample. Their generalizability to new patient groups—that is, their external validity—is unknown. Because models that only predict outcomes in one group of patients using data from one site or MRI scanner are less practically useful, external validity is a hurdle that must be cleared before models can be translated from the bench to the bedside. Resources such as the Predict Language Outcome and Recovery After Stroke (PLORAS) Database, which includes MRI data and language measures from 750 patients with aphasia[10], provide valuable opportunities to test whether models generalize to predict outcomes in new datasets. As Halai et al. note in the limitations section of their article, the current models were not externally validated because of a lack of large-scale datasets including all of the brain scan types tested. Future model validations can address a key question untested by

Halai and colleagues: do the top-performing models in their sample also best predict language function in completely new individuals? As data become more widely available, adopting cross-dataset validation as standard practice in brain-based predictive modelling is critical for increasing the clinical utility of MRI.

A central goal of translational neuroscience is building brain-based models that generate robust, reliable predictions of symptoms and behaviour. Research that systematically evaluates model features, algorithms and generalizability represents an important step towards developing a standardized tool-kit of predictive modelling approaches in human neuroimaging. Optimizing predictive models not only improves our basic understanding of brain-to-behaviour relationships, but may also increase the feasibility of applying models in real-world settings. For example, knowing which type of data best predicts post-stroke aphasia could help doctors decide what behavioural and MRI data to collect (and not collect), minimizing costs to hospitals and patients. The work by Halai and colleagues highlights the importance of optimizing brain-based predictive models and evaluating their cross-dataset generalizability for improving basic scientific understanding and patient care. ❐

Monica D. Rosenberg ✉ and Hayoung Song

Department of Psychology, University of Chicago, Chicago, IL, USA.

✉*e-mail: mdrosenberg@uchicago.edu*

Published online: 19 June 2020
https://doi.org/10.1038/s41562-020-0902-1

References

1. Engelter, S. T. et al. *Stroke* **37**, 1379–1384 (2006).
2. Hilari, K. *Disabil. Rehabil.* **33**, 211–218 (2011).
3. Breitenstein, C. et al. *Lancet* **389**, 1528–1538 (2017).
4. Halai, A. D. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-020-0854-5 (2020).
5. Hope, T. M. H., Seghier, M. L., Leff, A. P. & Price, C. J. *Neuroimage Clin* **2**, 424–433 (2013).
6. Yourganov, G., Smith, K. G., Fridriksson, J. & Rorden, C. *Cortex* **73**, 203–215 (2015).
7. Plowman, E., Hentz, B. & Ellis, C. Jr. *J. Eval. Clin. Pract* **18**, 689–694 (2012).
8. Halai, A. D., Woollams, A. M. & Lambon Ralph, M. A. *Cortex* **86**, 275–289 (2017).
9. Halai, A. D., Woollams, A. M. & Lambon Ralph, M. A. *Neuroimage Clin* **19**, 1–13 (2018).
10. Seghier, M. L. et al. *Neuroimage* **124 Pt B**, 1208–1212 (2016).

Competing interests

The authors declare no competing interests.