



# City Research Online

## City St George's, University of London

**Citation:** Hope, T., Nardo, D., Holland, R., Ondobaka, S., Akkad, H., Price, C. J., Leff, A. & Crinion, J. (2021). Lesion site and therapy time predict responses to a therapy for anomia after stroke: a prognostic model development study. *Scientific Reports*, 11(1), 18572. doi: 10.1038/s41598-021-97916-x

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/26830/>

**Link to published version:** <https://doi.org/10.1038/s41598-021-97916-x>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).



OPEN

## Lesion site and therapy time predict responses to a therapy for anomia after stroke: a prognostic model development study

Thomas M. H. Hope<sup>1,2,✉</sup>, Davide Nardo<sup>1,3</sup>, Rachel Holland<sup>4</sup>, Sasha Ondobaka<sup>1</sup>, Haya Akkad<sup>1</sup>, Cathy J. Price<sup>2</sup>, Alexander P. Leff<sup>1,5</sup> & Jenny Crinion<sup>1</sup>

Stroke is a leading cause of disability, and language impairments (aphasia) after stroke are both common and particularly feared. Most stroke survivors with aphasia exhibit anomia (difficulties with naming common objects), but while many therapeutic interventions for anomia have been proposed, treatment effects are typically much larger in some patients than others. Here, we asked whether that variation might be more systematic, and even predictable, than previously thought. 18 patients, each at least 6 months after left hemisphere stroke, engaged in a computerised treatment for their anomia over a 6-week period. Using only: (a) the patients' initial accuracy when naming (to-be) trained items; (b) the hours of therapy that they devoted to the therapy; and (c) whole-brain lesion location data, derived from structural MRI; we developed Partial Least Squares regression models to predict the patients' improvements on treated items, and tested them in cross-validation. Somewhat surprisingly, the best model included only lesion location data and the hours of therapy undertaken. In cross-validation, this model significantly out-performed the null model, in which the prediction for each patient was simply the mean treatment effect of the group. This model also made promisingly accurate predictions in absolute terms: the correlation between empirical and predicted treatment response was 0.62 (95% CI 0.27, 0.95). Our results indicate that individuals' variation in response to anomia treatment are, at least somewhat, systematic and predictable, from the interaction between where and how much lesion damage they have suffered, and the time they devoted to the therapy.

Stroke is a leading cause of disability<sup>1</sup>, and language impairments (aphasia) after stroke are both common<sup>2</sup> and particularly feared<sup>3</sup>. Most stroke survivors with aphasia exhibit anomia, a difficulty finding words when naming common objects<sup>4</sup>, but while many therapeutic interventions for anomia have been proposed, treatment effects typically vary, substantially, from patient to patient<sup>5</sup>. Inter-individual variation in treatment responses is ubiquitous in medicine (e.g. in psychiatry and pharmacology, respectively<sup>6,7</sup>), but emerging evidence suggests that variation in responses to therapy for aphasia after stroke might be more systematic than previously thought<sup>5,8,9</sup>. To the extent that this is true, the implication is that we can use pre-treatment (e.g. behavioural and/or brain imaging) data to explain and even predict patients' likely treatment responses.

For example, we recently showed that a model derived from: (a) pre-treatment scores on standardised cognitive and language tasks; and (b) lesion location data, derived from pre-treatment structural MRI, could be used to predict 23 aphasic stroke patients' responses to a treatment for acquired reading impairments (central alexia)<sup>8</sup>. Like the treatment considered here, for anomia, this earlier treatment for alexia was a computerised application designed to engage participants in massed practice of trained items at home, over a period of weeks. Using stepwise forward feature selection, we selected specific predictors from the pre-treatment data for entry

<sup>1</sup>Institute of Cognitive Neuroscience, University College London, 17-19 Queen Square, London WC1N 3AR, UK. <sup>2</sup>Wellcome Centre for Human Neuroimaging, University College London, London, UK. <sup>3</sup>MRC Cognition and Brain Sciences Unit, Cambridge University, London, UK. <sup>4</sup>Division of Language and Communication Science, City University of London, London, UK. <sup>5</sup>UCL Queen Square Institute of Neurology, London, UK. ✉email: t.hope@ucl.ac.uk

into a multiple linear regression model, which explained over 90% of the variance in patients' empirical treatment responses. This result is biased by over-fitting, because all of the patients' data were used to select features, but even when the feature selection was nested within each fold of a leave-one-out cross-validation process (i.e. removing the bias), the resulting predictions were significantly correlated with patients' empirical treatment responses ( $r = 0.48$ ,  $p < 0.05$ )<sup>8</sup>.

In what follows, we add to the evidence that responses to aphasia treatment might be predictable from patients' pre-treatment data. Here, we focus on a computerised treatment for naming difficulties (anomia) that is the most common language impairment after stroke<sup>4</sup>. This treatment's effectiveness at the group level has already been verified<sup>10</sup>; here, we attempt to explain and predict the same treatment effects at the individual level. Encouraged by the apparent importance of lesion site, both in language outcomes after stroke irrespective of treatment<sup>11–13</sup>, and in responses to a treatment for reading impairments after aphasic stroke<sup>8</sup>, our main hypothesis was that the individual patients' responses to the treatment are systematic and predictable given the site and extent of lesion damage. We tested this by comparing predictions made by models derived from those data (alone or in combination with their pre-treatment anomia severity, demographic data and the hours of therapy completed), to the predictions made by a 'null' model, which simply predicted the mean treatment response of its training sample. If (any of) the former are significantly more accurate than the latter, the implication is that individual variation in responses to this treatment is systematic and predictable, at least to some extent.

## Methods

The current analysis employs pre-treatment: (a) demographic data (age at stroke onset, time post-stroke at assessment, and sex); (b) patients' initial impairment severity (i.e. their accuracies when naming to-be-treated items, pre-treatment); (c) the hours of therapy actually undertaken; and (d) structural MRI, which we use to extract lesion location data. We used these data to predict and explain patients' responses to therapy, measured as the absolute change in naming accuracy, from pre- to post-treatment, for 'trained' items (i.e. items practiced during the therapy).

The therapy was designed to engage participants in massed practice of object naming, over a 6-week period at home. A variety of phonemic cues (e.g. an audio recording of the object's name, or of the name's first phoneme) were presented concurrently with the picture to be named during treatment to encourage error-reducing learning<sup>14</sup>. The approach was both effective and specific to spoken word production, significantly improving patients' overall object naming accuracy and reaction time immediately post-treatment (unstandardized effect size: 29% and 17%, respectively; Cohen's  $d$ : 3.45 and 1.83). Longer term gains in naming were maintained three months later, though in this study we focus only on the immediate gains made for items trained during the therapy.

**Participants.** The study participants were 18 right-handed native English-speakers, with normal hearing, no history of psychiatric disease and no prior history of neurological disorder prior to a left-hemisphere stroke, causing language impairment (aphasia). Participants were recruited either from an aphasia clinic, run by JC, or via the Predicting Language Outcomes After Stroke (PLORAS) study, run by CJP, between 2009 and 2012. The study size was arrived at via a power calculation based on the expected effect size of the treatment considered.

Participants were only included if they had: (i) naming difficulties (anomia), as assessed via the Boston Naming Test (cut-off  $< 56$ ); (ii) relatively preserved single-word comprehension as assessed via the Comprehensive Aphasia Test (CAT)<sup>15</sup>; (iii) good mono-syllabic word repetition as assessed via the Psycholinguistic Assessments of Language Processing in Aphasia<sup>16</sup>; (iv) no speech apraxia as determined by the Apraxia Battery for Adults<sup>17</sup>; and, (v) at least partially spared left inferior frontal cortex (thought to support speech re-learning<sup>10</sup>). All gave written informed consent to take part in the study, which was approved by the Central London Research Ethics Committee and conducted in accordance with the ethical principles stated by the Declaration of Helsinki. A table of the participants' key characteristics, reproduced from<sup>10</sup>, is included in supplementary material.

**Stimuli and procedure.** The procedure for the treatment study<sup>10</sup> involved behavioural assessments and neuroimaging data acquisition both pre- and post-treatment. Here, we use pre-treatment data only, to predict treatment response, calculated as the change in the number of trained items that patients could name correctly.

Stimuli were drawn from a pool consisting of 299 black and white line drawings of objects adapted from the International Picture-Naming Project (<https://crl.ucsd.edu/experiments/ipnp/>). All object names were monosyllabic, with a consonant–vowel–consonant structure and high name agreement (e.g. 'car'). The treatment employed 150 of the 299 stimuli: i.e. for each patient, there were 150 treated items and 149 untreated items. 54/150 to-be-trained items and 53/149 un-trained items were kept common across all patients (for use in an fMRI experiment<sup>10</sup>, which we do not report here). The remaining items (96/150 to be trained; 95/150 to be untrained) were determined for each patient on the basis of their individual pre-treatment naming performance (accuracy) on the 299 items. Performance on the treated and untreated lists was matched pre-treatment.

After baseline assessment and pre-treatment structural MRI, patients were given a laptop and asked to complete a minimum of two hours of naming practice five days a week, over a six-week period. The pictures and auditory cues were presented using the 'StepByStep' aphasia treatment software (<http://www.aphasia-software.com>). The naming practice was designed to be completed in an error-reducing manner<sup>18</sup>. For example, in naming a picture of a car the patient was asked to name it three times: (i) after a whole word auditory cue /ka:r/; (ii) after an initial phonemic cue /ka/; (iii) after a whole word cue again. Only then would the patient proceed to the next item to be named. Patients completed, on average, a total of 73 ( $\pm 25$ ) hours of naming practice: i.e., within one standard deviation of the mean therapy dose found, in a meta-analysis of aphasia treatment studies<sup>19</sup>, to improve patients' communicative ability. After the six-week period, patients were reassessed using the same test

battery as at baseline. Naming accuracy was scored according to the standardized Comprehensive Aphasia Test guidelines<sup>15</sup>. Reported analyses are purposefully limited to absolute change in naming accuracy (from pre- to post-treatment) on the 150 treated items.

**Imaging acquisition and analysis.** The same scanner and hardware were used for the acquisition of all images. Whole-brain imaging was performed on a 3 T Siemens TIM-Trio system (Siemens) at the Wellcome Centre for Human Neuroimaging. Lesion images were derived from structural MRI using the Automatic Lesion Identification toolbox<sup>20</sup>, and then double-checked for accuracy by a researcher experienced in manual lesion-tracking (DN), reviewing individual axial slices.

Lesion data were then encoded as lesion load in a series of 398 anatomically defined regions of interest, derived from four publicly available atlases (two focused on grey matter and two focused on white matter)<sup>21–24</sup>. Where regions were represented in probabilistic format, they were re-encoded as binary images at a 50% threshold. For each region, lesion load was calculated as the number of (2mm<sup>3</sup>) voxels shared by the lesion and the region, divided by the total number of voxels in that region. Notably, there was significant overlap between these regions, across atlases. Rather than deciding a priori what the best or most useful atlas might be, our goal was simply to reduce the dimensionality of the lesion data in a manner that retained an explicit link with familiar brain regions and/or tracts.

**Modelling methods.** Our primary aim was to assess whether individual patients' treatment responses could be predicted from pre-treatment data alone. Here, we define 'treatment responses' as the absolute change in patients' naming accuracy from pre- to post-treatment.

Treatment studies in this domain are resource intensive and typically involve massed practice, so take time to complete. Like most others in the field, our sample is therefore smaller (n = 18) than is usually desirable when building predictive models, increasing the risk of over-fitting. This risk is further increased because we have extensive pre-treatment data to consider, including behavioural data, and lesion data derived from structural MRI.

One way to mitigate this risk is via feature selection, as we employed in similar, previous work<sup>8</sup>. Although successful, that work still revealed significant over-fitting, because our in-sample results (using the whole dataset to select features) were so much stronger than our out-of-sample results (i.e. nesting feature selection in cross-validation):  $R^2$  (predicted response, empirical response) = 0.94 (in-sample); 0.23 (out-of-sample). Accordingly, we took a simpler approach in this work by using dimensionality reduction, rather than feature selection, to manage the high dimensionality of the pre-treatment (behavioural and brain imaging) predictors.

**Predictive models.** We used Partial Least Squares (PLS) regression, as implemented in Matlab 2019a, to develop our models, using either: (a) demographic variables, including age at stroke onset, sex and time post-stroke; (b) pre-treatment naming accuracy (i.e. measuring the initial severity of each patient's anomia); and/or (c) lesion data, derived from pre-treatment structural MRI. We additionally considered one further variable, both singly and in combination with the other data: the hours of therapy actually completed by each patient. There were no missing data for any patient for any of these variables. All predictor variables were standardised (z-scored) prior to entry into models.

PLS regression is appropriate here because it employs dimensionality reduction analogous to, but more efficient than, that implemented by Principal Components Analysis (PCA): i.e., where PCA identifies latent variables which explain maximal variance in the predictors, PLS regression identifies variables that explain maximal variance in the response variable(s). PLS regression thus allows us to build potentially effective models that are (at least somewhat) robust to irrelevant predictors, rather than excluding those predictors explicitly.

The behavioural model employed 28 predictors: i.e. scores on our battery of pre-treatment language and cognitive assessments (as described in detail in<sup>10</sup>). The lesion data were encoded as described previously, into 398 lesion load variables: however, all patients had left-hemisphere lesions, and in fact all patients had zero lesion load in 220/398 regions. These were removed from the analysis (leaving 178 variables), but their removal had no substantive effect on the results. We trained models employing predictors derived from each data type separately, and all higher order combinations of data types.

**Model assessment and model comparison.** Predictive performance was assessed with cross-validation. We report results using 1000 times tenfold cross-validation here, but analyses employing different types of cross-validation were substantially similar. Absolute measures of predictive performance are suspect in small samples, so we assessed our models in relative terms, by comparing them to an empty, or baseline (i.e. null) model, which simply predicts the average treatment response for all patients in the group. This model reflects our null hypothesis, that treatment responses are *not predictable at the individual level*, leaving group-level averages as the only recourse. When our empirical models outperform the empty model, we reject the null hypothesis, concluding that individual treatment responses are predictable, at least to some extent. We compare models by recording the Root Mean Squared Error (MSE) for predictions made in each of 1000 repetitions of a tenfold cross-validation. These folds are kept identical across models, so the MSE values can be compared pair-wise.

Traditional paired tests are not appropriate on their own here because different partitions of the data will create overlapping training datasets, which are therefore not independent of each other. Accordingly, while we use the traditional, paired, non-parametric Wilcoxon signed rank test to compare MSEs across models, we further threshold those statistics with paired permutation test. The test construes the two vectors to be compared as having labels, reflecting the models used to generate them. The null hypothesis is that those labels are arbitrary, because the models' performance do not differ except by chance. We therefore create a null distribution of paired

Data types	Median/IQR MSE
Null	303/16
Hrs (Therapy)	300/16
Initial (severity)	396/31
Demographics	321/24
Lesions	205/30
Initial + Hrs	364/30
Demographics + Hrs	316/26
<b><u>Lesions + Hrs</u></b>	<b><u>182/21</u></b>
Initial + Demographics	364/30
Initial + Lesions	253/29
Lesions + Demographics	267/21
Hrs + Initial + Demographics	355/30
Hrs + Initial + Lesions	220/32
Hrs + Demographics + Lesions	253/21
Initial + Demographics + Lesions	274/23
Hrs + Initial + Demographics + Lesions	261/23

**Table 1.** Data configurations and predictive performance, as assessed across the same 1000 tenfold cross-validation runs. *MSE* mean squared errors of the model predictions; *IQR* inter-quartile range of the model predictions. These quantities are employed in preference to mean and standard deviation because MSEs typically have a Poisson distribution rather than a normal distribution. Lower MSEs imply more accurate predictions. The best model configuration is underlined (Hrs + Lesions): the most accurate predictions are derived from these data.

test statistics by randomly permuting those labels *within each pair*, and repeating the original paired (signed rank) test. If the original statistic is extreme relative to the null distribution, we conclude that the performance difference between the models is significant ( $p < 0.05$ ) after a correction for FamilyWise Error (FWE).

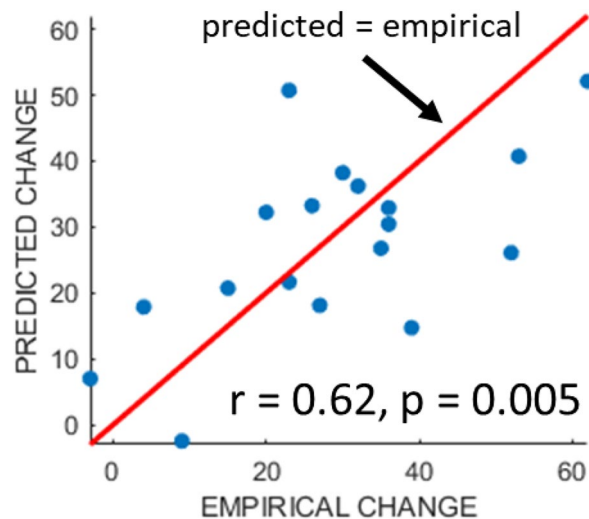
**Model interpretation.** PLS regression models can be interpreted by examining the weights of each of their components on each of the original variables. However, this approach can be challenging when there are multiple components to consider, and because the sign of each component is arbitrary: i.e. positive weights on a given component do not necessarily imply a positive relationship between the highly weighted independent variables and the dependent variable(s). We circumvent these issues with ‘data perturbation’.

The data perturbation procedure involves permuting random subsamples of the empirical independent variables and recording the effect of the perturbation on the model’s predictions. The PLS model beta weights themselves are fixed based on the original empirical data: our goal is not to fit further models, but rather to better understand the relationships that have already been encoded. We do this by: (a) permuting random subsets of the independent variables; (b) observing the effect on the models’ predictions; and (c) relating perturbed variable values to the resulting predictions with mass univariate correlation analyses. The resultant correlation coefficients approximate the influence that each variable has on the model’s predictions. We ran 1000 iterations of the process per model, yielding a total sample size of 18 (patients) \* 1000 (iterations), including both perturbed independent variables and the resultant, predicted dependent variable (treatment response). Repeated analyses with this number of iterations yielded very consistent coefficients for all of the models we report across ten repetitions of 1000 iterations of this process, all pairwise correlations between derived weights on behavioural and lesion variables were  $> 0.99$ .

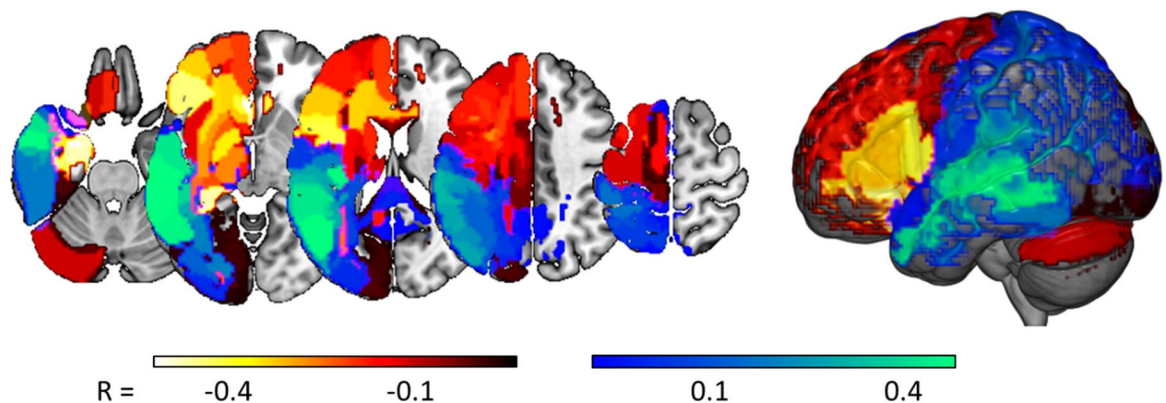
**Ethical approval.** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Results

**Predictive performance.** Table 1 reports predictive performances (median and inter-quartile ranges of Mean Square Errors, or MSEs) of models driven by all combinations of the data we considered. All but one of the models that out-performed the null model, with lower MSEs, included lesion data. The exception was a model including hours of therapy only, with a median MSE of 300 (IQR = 16): i.e. a very small difference relative to the null model, albeit a significant one (FWE adjusted  $p < 0.05$ ). The best combination was hours of therapy plus lesion data (MSE median/inter-quartile range = 182/21), and indeed this was the only combination which improved upon lesion data alone: see Table 1. The mean predictions of that best model, across the 1000 repetitions, were strongly and significantly correlated with empirical treatment responses ( $r = 0.62$ ,  $p = 0.006$ , 95% CI 0.27, 0.95): see Fig. 1.



**Figure 1.** Predicted responses versus empirical responses, for the best model identified in Table 1 (lesion load variables + hours of therapy undertaken).

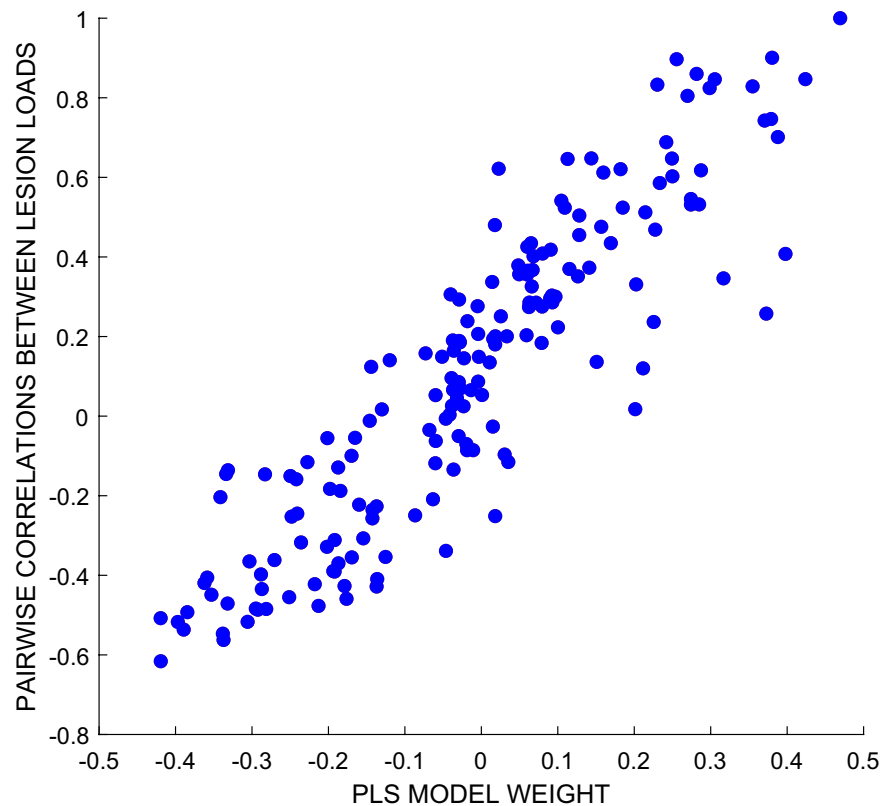


**Figure 2.** Relating lesion locations to predicted treatment responses. Correlation coefficients, derived via data perturbation, relating the lesion load in each of 177 regions, to treatment responses predicted by our best model (appending lesion data to the hours of therapy actually undertaken).

**Interpreting the best model.** Variable weights for the best models predicting change on treated items, using a combination of the hours of therapy undertaken and lesion location data, were calculated via data perturbation, as described in the Methods.

First, as expected, the best model predicted better improvement when patients devoted more hours to practice ( $r = 0.33$ ). Regional weights for the lesion data in this model (i.e. taking therapy hours into account) are displayed in Fig. 2, with the most negative weights (predicting lesser treatment benefit with more damage) in and around the left inferior frontal gyrus, and positive weights (predicting greater treatment benefit with more damage) in the middle, superior and anterior temporal lobe regions. Where voxels appear in two overlapping regions with different weights (e.g. we had one region covering the whole of the hippocampus and others covering only its cornu ammonis and dentate gyrus subfields), the most extreme of those two weights is displayed.

Notably, weights in many brain areas, including the auditory cortex and the superior, middle and anterior temporal lobes, are positive. The potentially curious implication here, is that more damage predicts larger treatment responses. Instead, we suggest that these positive regions are driven by the contingent distribution of the patients' lesions: more damage in those positively weighted regions implies less damage in the negatively weighted regions (where the latter make the more intuitive association between more damage and smaller treatment responses). As an illustration of this relationship, we considered area TE11 of the primary auditory cortex, where the strongest, positive weight was observed (0.47). Pairwise correlations, between lesion loads in this region and lesion loads in each of the other (177) regions under consideration, were very strongly correlated with the weights displayed in Fig. 2, which were assigned to those regions by our best PLS regression model ( $r = 0.90$ ): see Fig. 3.



**Figure 3.** Scatter plot relating: (i) coefficients of the pairwise correlations between lesion load values in primary auditory cortex area TE11, and lesion loads in all of the 177 brain regions that we considered (y-axis); to (ii) the weights assigned to each of those same brain regions by our best PLS regression model, as derived via data perturbation (described in the Methods). The strong correlation between these two quantities implies that lesser lesion load in primary auditory cortex area TE11 serves as a proxy for greater lesion load in areas where that extra damage most strongly predicts poorer treatment responses.

## Discussion

Recent results suggest that individual stroke patients' responses to aphasia treatment are to some extent systematic and predictable<sup>8</sup>. Our results add to this evidence, showing that responses to a behavioural treatment for anomia are at least somewhat predictable at the individual level. We assessed models derived from demographic variables, and from pre-treatment behavioural and lesion data. Models were derived via PLS regression, drawing on efficient predictor dimensionality reduction and thus obviating the need for either algorithmic or a priori feature selection. These models provide a sound way to establish at an individual level whether pre-treatment data include signals that might be used to predict treatment responses.

Many of the models we tested made significantly better predictions than those of a baseline model, in which each patient's prediction was simply the mean response of the group (see Table 1). However, the best model employed lesion location data, derived from MRI, plus the hours of therapy undertaken by each patient. As hours of therapy alone has very little predictive power, the results suggest that the benefit of increased therapy depends on lesion location. This may explain why detecting therapy dose effects has been so challenging<sup>25</sup>. Notably, we could not predict training effort, as indexed by hours of therapy undertaken at the individual level, from any of the other data considered here.

Our best prognostic model, including lesion data and hours of therapy, is broadly sensible. The negative weights assigned to the left inferior frontal gyrus, the hippocampus and the cerebellum (more damage = less improvement) are consistent with prior work emphasising the importance of the preservation of these regions in the response to aphasia therapy (e.g.<sup>26–28</sup>). It might be tempting to explain the positive (counter-intuitive) lesion weights as reflecting areas where more damage more severely impairs naming, allowing more room for improvement. However, since pre-treatment naming performance was not useful in predicting treatment responses in our sample (see Table 1), we prefer an alternative explanation emphasising the contingent distribution of the patients' lesions. Specifically, as illustrated in Fig. 3, greater damage in regions with positive (counter-intuitive) weights also predicted greater *preservation* of regions with negative weights, and vice versa.

This explanation predicts that we might, in principal, disregard those counter-intuitive weights. Notably, we cannot simply remove them from the models reported here, because our models' weights (both intuitive and counter-intuitive) are all optimised together, so that in a sense every weight depends on every other weight. But future work might usefully explore more elegant methods for achieving this result, perhaps using Bayesian priors

to constrain the signs of the weights that we are prepared to accept. Paradoxical lesion effects<sup>29</sup> aside, we do have strong priors in this domain (about the relationship between brain damage and behavioural performance), which might simplify both the model fitting process and the interpretation of the results. Our proposed explanation for our models' counter-intuitive weights would predict that some sensibly constrained alternative model would be at least as good as any of those reported here if we have strong prior assumptions about the directionalities of the weights we expect to find in this domain—which they cannot be removed from the current models because those models' weights are optimised across all variables together.

PLS regression assigns non-zero weights to every variable in each model, and in this case these models appear to us to be capturing this better preservation of the regions exhibiting 'counter-intuitive' lesion-symptom associations (more damage predicts better treatment responses) was strongly associated with greater damage at locations where damage predicted poorer treatment responses.

And the positive weights may best be explained as emphasising those regions where more extensive damage predicts better preservation of the regions that appear to support better responses to treatment (see Fig. 3). Note that the natural alternative explanation here, that these counter-intuitive.

Notably, we did not employ any feature selection in this work: i.e. we did not attempt to select the subset of lesion location variables that might best explain the patients' treatment responses. This is a limitation of the current work, made necessary because feature selection encourages over-fitting in small samples<sup>8</sup>: the only general way to circumvent this issue is via external validation: testing the best model from this study in a second, completely independent sample. But this is no simple endeavour, because the time and effort required to run these studies is substantial, and we do not yet know how similar such a study would need to be to that reported here. Does the treatment have to stay exactly the same? How much can the inclusion criteria vary? Work to address these questions, by measuring how prognostic models generalise across independent samples (e.g. as in<sup>30</sup>) and different therapy studies, is ongoing.

Perhaps surprisingly, our models did not benefit from the addition, either of the initial severity or the demographic data that we considered – suggesting that this treatment's efficacy did not depend on the patients' ages, sex, time post-stroke or pre-treatment impairment severity (at least once lesion location had been taken into account). The null finding for initial impairment severity is perhaps particularly surprising, given the apparently important role that it plays in other studies of anomia treatment (e.g.<sup>31</sup>). Notably, this is not just a case of the treatment response variance being better explained by lesions than by initial impairment severity, because the model using just the latter was significantly *worse* than the null model. Instead, we speculate that, in selecting patients with relatively preserved auditory comprehension skills, we also implicitly reduced the variance in their initial symptom severity that might otherwise have better explained their eventual treatment responses.

Whether these null results generalise in larger samples, is a question for future work. But our results do suggest that pre-treatment structural neuroimaging (lesion data), in combination with treatment dose, can be used to predict individual patients' therapeutic anomia intervention response. This is consistent with prior results, suggesting that the individual responses to treatment for aphasic stroke might interact with where and how much lesion damage individual patients have suffered<sup>8</sup>. We hope that these results will encourage further attempts to explain and predict inter-individual differences in treatment responses, with pre-treatment data, opening the way for a more positive and personalised treatment approach for aphasia.

## Data availability

The data described in this study is available to accredited researchers from JC, on request.

Received: 19 April 2021; Accepted: 14 July 2021

Published online: 17 September 2021

## References

- Corraini, P. *et al.* Comorbidity and the increased mortality after hospitalization for stroke: a population-based cohort study. *J. Thromb. Haemost.* **16**(2), 242–252 (2018).
- Engelter, S. T. *et al.* Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke* **37**(6), 1379–1384 (2006).
- Lam, J. M. & Wodchis, W. P. The relationship of 60 disease diagnoses and 15 conditions to preference-based health-related quality of life in Ontario hospital-based long-term care residents. *Med Care* **48**(4), 380–387 (2010).
- Laine, M. & Martin, N. *Anomia: Theoretical and clinical aspects* (Psychology Press, 2013).
- Crinion, J. T. & Leff, A. P. Using functional imaging to understand therapeutic effects in poststroke aphasia. *Curr. Opin. Neurol.* **28**(4), 330–337 (2015).
- Hellidin, L. *et al.* Neurocognitive variability in schizophrenia spectrum disorders: relationship to real-world functioning. *Schizophr. Res. Cognit.* **20**, 100172 (2020).
- Kadiev, E. *et al.* Role of pharmacogenetics in variable response to drugs: focus on opioids. *Expert Opin. Drug Metab. Toxicol.* **4**(1), 77–91 (2008).
- Aguilar, O. M. *et al.* Lesion-site-dependent responses to therapy after aphasic stroke. *J. Neurol. Neurosurg. Psychiatry* **89**, 1352–1354 (2018).
- Lambon Ralph, M. A. *et al.* Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. *Neuropsychol. Rehabil.* **20**(2), 289–305 (2010).
- Nardo, D. *et al.* Less is more: neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain* **140**(11), 3039–3054 (2017).
- Hope, T. M. *et al.* Comparing language outcomes in monolingual and bilingual stroke patients. *Brain* **138**(Pt 4), 1070–1083 (2015).
- Hope, T. M., Leff, A. P. & Price, C. J. Predicting language outcomes after stroke: Is structural disconnection a useful predictor?. *NeuroImage Clin.* **19**, 22–29 (2018).
- Hope, T. M. H. *et al.* Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage Clin.* **2**, 424–433 (2013).

14. Fillingham, J. K. *et al.* The application of errorless learning to aphasic disorders: a review of theory and practice. *Neuropsychol. Rehabil.* **13**(3), 337–363 (2003).
15. Swinburn, K., Porter, G. & Howard, D. *Comprehensive Aphasia Test* (Psychology Press, 2004).
16. Kay, J., Lesser, R. & Coltheart, M. *Psycholinguistic Assessments of Language Processing in Aphasia (PALPA)* (Hove, Laurence Erlbaum Associates, 1992).
17. Dabul, *Apraxia battery for adults (second ed.)* (2000) Austin, Texas: Pro-Ed.
18. Fillingham, J. K., Sage, K. & Lambon Ralph, M. A. The treatment of anomia using errorless learning. *Neuropsychol. Rehabil.* **16**(2), 129–154 (2006).
19. Bhogal, S. K., Teasell, R. & Speechley, M. Intensity of aphasia therapy, impact on recovery. *Stroke* **34**(4), 987–993 (2003).
20. Seghier, M. L. *et al.* Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage* **41**(4), 1253–1266 (2008).
21. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**(1), 273–289 (2002).
22. Hua, K. *et al.* Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* **39**(1), 336–347 (2008).
23. Mori, S., Wakana, S., Van Zijl, P. C. & Nagae-Poetscher, L. M. *MRI Atlas of Human White Matter* Vol. 2 (Elsevier, 2011).
24. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**(4), 1325–1335 (2005).
25. Harvey, S., *et al.* Dose effects in behavioural treatment of post-stroke aphasia: a systematic review and meta-analysis. *Disabil. Rehabil.* 1–12, online version ahead of print <https://doi.org/10.1080/09638288.2020.1843079>. (2020).
26. Sebastian, R. *et al.* Cerebellar tDCS: a novel approach to augment language treatment post-stroke. *Front. Hum. Neurosci.* **10**, 695–695 (2017).
27. Mattioli, F. *et al.* Early aphasia rehabilitation is associated with functional reactivation of the left inferior frontal gyrus. *Stroke* **45**(2), 545–552 (2014).
28. Meinzer, M. *et al.* Integrity of the hippocampus and surrounding white matter is correlated with language training success in aphasia. *Neuroimage* **53**(1), 283–290 (2010).
29. Zeiler, S. R. *et al.* Paradoxical motor recovery from a first stroke after induction of a second stroke: reopening a postischemic sensitive period. *Neurorehabil. Neural Repair* **30**(8), 794–800 (2015).
30. Loughnan, R. *et al.* Generalizing post-stroke prognoses from research data to clinical data. *NeuroImage Clin.* **24**, 102005 (2019).
31. Bonilha, L. *et al.* Success of anomia treatment in aphasia is associated with preserved architecture of global and left temporal lobe structural networks. *Neurorehabil. Neural Repair* **30**(3), 266–279 (2016).

### Author contributions

J.C., A.L., C.J.P. and T.H. conceived the study; D.N., S.O., H.A. and T.H. prepared the data, which was acquired by C.J.P., D.N. and S.O. TH and JC wrote the manuscript, and all authors reviewed and edited it.

### Funding

This study was supported by Wellcome (203147/Z/16/Z; 205103/Z/16/Z; 106161/Z/14/Z), MRC (G0701888), the Stroke Association (TSA PDF 2017/02) and NIHR (RP-2015-06-012). The funders had no participation in the design and results of this study.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97916-x>.

**Correspondence** and requests for materials should be addressed to T.M.H.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021