

Post-stroke outcomes predicted from multivariate lesion-behaviour and lesion network mapping

 Mark Bowren Jr,¹ Joel Bruss,² Kenneth Manzel,² Dylan Edwards,^{3,4} Charles Liu,^{5,6} Maurizio Corbetta,⁷ Daniel Tranel^{1,2,†} and  Aaron D. Boes^{8,†}

[†]These authors contributed equally to this work.

See Gray and Nachev (<https://doi.org/10.1093/brain/awac100>) for a scientific commentary on this article.

Clinicians and scientists alike have long sought to predict the course and severity of chronic post-stroke cognitive and motor outcomes, as the ability to do so would inform treatment and rehabilitation strategies. However, it remains difficult to make accurate predictions about chronic post-stroke outcomes due, in large part, to high inter-individual variability in recovery and a reliance on clinical heuristics rather than empirical methods. The neuroanatomical location of a stroke is a key variable associated with long-term outcomes, and because lesion location can be derived from routinely collected clinical neuroimaging data there is an opportunity to use this information to make empirically based predictions about post-stroke deficits. For example, lesion location can be compared to statistically weighted multivariate lesion-behaviour maps of neuroanatomical regions that, when damaged, are associated with specific deficits based on aggregated outcome data from large cohorts.

Here, our goal was to evaluate whether we can leverage lesion-behaviour maps based on data from two large cohorts of individuals with focal brain lesions to make predictions of 12-month cognitive and motor outcomes in an independent sample of stroke patients. Further, we evaluated whether we could augment these predictions by estimating the structural and functional networks disrupted in association with each lesion-behaviour map through the use of structural and functional lesion network mapping, which use normative structural and functional connectivity data from neurologically healthy individuals to elucidate lesion-associated networks. We derived these brain network maps using the anatomical regions with the strongest association with impairment for each cognitive and motor outcome based on lesion-behaviour map results. These peak regional findings became the ‘seeds’ to generate networks, an approach that offers potentially greater precision compared to previously used single-lesion approaches. Next, in an independent sample, we quantified the overlap of each lesion location with the lesion-behaviour maps and structural and functional lesion network mapping and evaluated how much variance each could explain in 12-month behavioural outcomes using a latent growth curve statistical model.

We found that each lesion-deficit mapping modality was able to predict a statistically significant amount of variance in cognitive and motor outcomes. Both structural and functional lesion network maps were able to predict variance in 12-month outcomes beyond lesion-behaviour mapping. Functional lesion network mapping performed best for the prediction of language deficits, and structural lesion network mapping performed best for the prediction of motor deficits. Altogether, these results support the notion that lesion location and lesion network mapping can be combined to improve the prediction of post-stroke deficits at 12-months.

1 Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA

2 Department of Neurology, Carver College of Medicine, Iowa City, IA 52242, USA

3 Moss Rehabilitation Research Institute, Elkins Park, PA 19027, USA

4 Edith Cowan University, Joondalup, WA 6027, Australia

5 Neurorestoration Center and Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

6 Rancho Los Amigos National Rehabilitation Center, Downey, CA, USA

7 Department of Neuroscience, Venetian Institute of Molecular Medicine and Padova Neuroscience Center, University of Padua, Padova, PD 32122, Italy

8 Departments of Neurology, Psychiatry, and Pediatrics, Carver College of Medicine, Iowa City, IA 52242, USA

Correspondence to: Mark Bowren, Jr

G60 Psychological and Brain Sciences Building

Iowa City, IA 52242, USA

E-mail: markbowrenjr@gmail.com

Keywords: stroke; lesion-behaviour mapping; lesion network mapping; functional connectivity; brain networks

Abbreviations: LBM = lesion-behaviour mapping; LNM = lesion network mapping; LGC = latent growth curve; LL = lesion load; PCA = principal components analysis

Introduction

Patients with stroke typically present with diverse cognitive and motor deficits; while some resolve in the days and weeks following stroke, others persist and become permanent disabilities.^{1,2} Clinicians and scientists alike have long sought to predict the course and severity of chronic deficits, as accurate predictions would inform treatment and rehabilitation strategies, and improve the design of clinical trials. Currently, it remains difficult to make accurate predictions due to high inter-individual variability in recovery and a reliance on clinical judgment rather than quantitative data.^{3,4} Thus, the development of empirical methods to aid clinical predictions of post-stroke deficits remains an important goal for neurology and clinical neuroscience.

Lesion location has a well-established association with post-stroke outcomes.^{5–9} Because the neuroanatomical location of a stroke can be derived from structural neuroimaging that is routinely acquired for clinical purposes in the evaluation of stroke, there is an opportunity to leverage this information to inform outcome predictions. Importantly, several decades of research have refined the methods for generating accurate and reliable statistical associations between lesion location anatomy and behavioural deficits, an approach often referred to as lesion-behaviour mapping (LBM).^{10–12} LBM uses lesion location and behavioural data from groups of individuals with focal brain damage to produce statistically weighted maps of the neuroanatomical regions that, when damaged, are associated with specific deficits.¹³ Previous research has demonstrated that LBM can be used to predict the severity of post-stroke hemiparesis,^{14,15} as well as deficits in domain-specific⁹ and domain-general cognitive ability across patient cohorts.¹⁶ Here, our goal is to evaluate whether we can leverage LBMs derived from large cohorts of individuals with focal brain lesions tested in the chronic epoch to make longitudinal predictions of 12-month motor and cognitive outcomes in an independent sample of patients with imaging acquired within 2-weeks of a new-onset stroke.

In addition to lesion location, post-stroke deficits can be understood in light of the disruption of anatomically distributed brain networks that extend beyond the boundaries of the lesion. The importance of brain networks for cognitive and motor functions has been recognized for well over a century,¹⁷ and this topic has garnered more attention in recent years with the development of new imaging tools to investigate the remote

effects of focal lesions.^{18–21} Several studies have performed advanced imaging of subjects with focal brain lesions to reveal associations between behavioural deficits and the remote network effects of a lesion, whether measured by PET, functional MRI, or diffusion tractography.^{22–24} Because these advanced imaging modalities are not routinely collected for clinical purposes, we and others have proposed that the network effects of a focal lesion can be inferred using large, high quality normative connectome data, an approach we term lesion network mapping (LNM). Recent work has evaluated the connectivity associated with a lesion location using normative resting-state functional MRI data in order to localize neurological deficits to functional brain networks, referred to here as functional LNM.^{25,26} Similarly, diffusion-based tractography has been used to associate deficits with damage to structural brain networks (i.e. white matter tracts), an approach we refer to as structural LNM.^{27–29} However, it is unclear whether these inferred network effects of brain lesions derived from healthy subjects add value to the prediction of post-stroke deficits, with mixed results in prior attempts.^{27,30–32} A recent study that evaluated the predictive power of lesion-associated networks showed better predictive power for LBM and structural LNM, and quite limited predictive value for functional LNM, except for visual deficits.³¹ One possible explanation for this finding is the use of large lesions to seed functional connectivity analyses, which may contain several functionally distinct regions such that signal averaging of the blood oxygen level-dependent signal within the lesion mask is problematic.³³ Thus, it remains unknown whether there are alternative approaches that could be used to combine lesion location, structural LNM, and functional LNM to maximize outcome predictions in patients with stroke.

Here, we sought to evaluate the utility of both lesion location and lesion-associated networks for the prediction of post-stroke cognitive and motor outcomes. Specifically, we derived lesion-behaviour maps for different cognitive and motor outcomes from two large cohorts of individuals with focal brain lesions. Structural and functional networks were then derived from these lesion behaviour maps using normative data,^{16,34} which is a departure from prior attempts using the lesion location from each individual subject to seed network analyses. This LBM-guided approach to LNM avoids the ‘large lesion’ confounds described above for functional LNM and has the potential to offer greater precision for identifying the networks

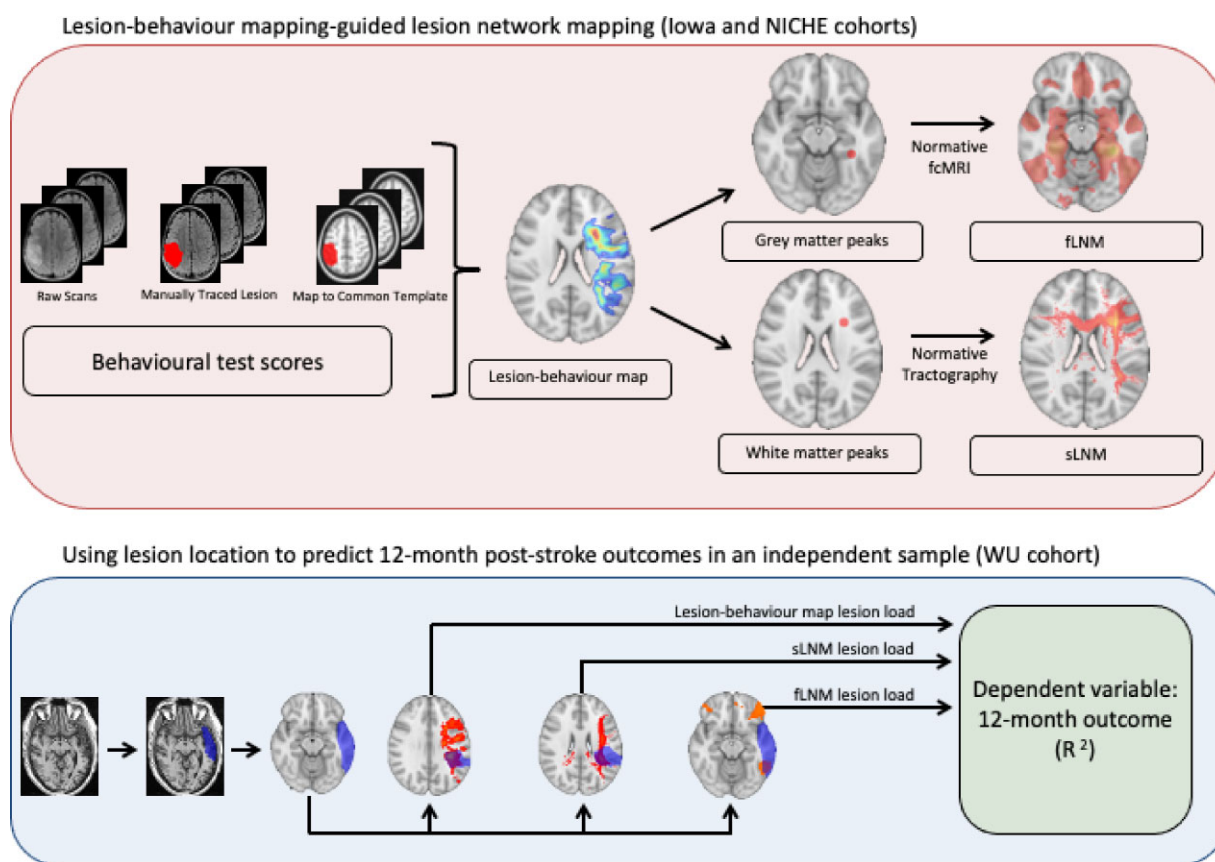


Figure 1 Schematic of approach. Multivariate LBMs were generated from sparse canonical correlation analysis based on lesion locations (i.e. voxel lesion status) and behavioural measurements. Data from the Iowa and NICHE cohorts were used to generate LBMs for cognitive and motor functions, respectively. These statistically-weighted maps were used to identify seed regions-of-interest for structural and functional lesion network mapping (LNM), which explore the structural and functional brain networks associated with regional peaks in the white and grey matter from each LBM, respectively. Then, for each patient in the WU cohort, we summed the voxel intensities from the LBM and structural and functional LNM results that also appeared in the patient's lesion mask to generate lesion load scores. The 12-month post-stroke outcome for each domain in the WU cohort was predicted from the lesion load scores. We examined the separate and combined effects of these predictors.

most causally associated with neurological deficits.³³ We tested the predictive utility of the LBMs and the structural and functional networks derived from each LBM by examining the variance that each could explain in the 12-month post-stroke behavioural outcomes of an independent sample (see Fig. 1 for overview of the study design).

Materials and methods

Participants, behavioural measures, and neuroimaging data

All participants gave written informed consent to participate in the studies described below, which were approved by the Institutional Review Boards of the participating institutions. Three separate cohorts were included. The Iowa cohort was used to derive LBMs of cognitive outcomes. The NICHE cohort was used to derive an LBM of motor outcomes, as motor outcomes were not collected for the Iowa sample. Both Iowa and NICHE cohorts were tested and imaged in the chronic epoch, 3 months or more after the lesion onset. Lesion-behaviour mapping results from the Iowa and NICHE cohorts were then used to make predictions of 12-month outcomes in the Washington University (WU) cohort, which

included individuals with stroke who had imaging performed within 2 weeks of stroke, and motor and cognitive testing performed longitudinally within 2 weeks after stroke, 3 months post-stroke and 12 months post-stroke.

Iowa cohort

The Iowa Neurological Patient Registry contains cognitive test data from individuals with stable, focal brain lesions.³⁵ Participants in the Iowa Registry completed a range of standardized neuropsychological tests and neuroimaging three or more months after lesion onset.³⁶ We focused on the neuropsychological tests that measure a cognitive domain that was also measured in the WU cohort: the Boston Naming Test (a test of expressive language and naming), the Token Test (a test of receptive language), and the delayed recall trial from the Rey Auditory Verbal Learning Test (a test of anterograde verbal memory). When applicable, normative corrections were applied using meta-normative data.³⁷ When a test was administered more than once to the same participant, we used the score from the administration that was the most contemporaneous with the date of the participant's neuroimaging scan. See Table 1 for demographic information from the Iowa cohort by neuropsychological test.

Table 1 Demographic information

Sample by Test	Age (SD)	Sex	Education (SD)	Handedness	Lesion chronicity (IQR)	Lesion laterality (R/L/B)
Iowa: BNT (n = 432)	56.38 (13.60)	205M/227F	13.47 (2.84)	386R/29L/17Mi	16.9 months (49.1 months)	133R/217L/82B
Iowa: Token Test (n = 210)	55.00 (12.87)	110M/100F	13.46 (2.79)	187R/13L/10Mi	15.4 months (46.8 months)	32R/143L/35B
Iowa: RAVLT (n = 477)	56.51 (13.62)	249M/228F	13.26 (2.82)	427R/31L/19Mi	17.5 months (46.0 months)	169R/209L/99B
NICHE: Fugl-Meyer (n = 103)	60.77 (11.15)	70M/33F	NA	91R/8L/3Mi	9.76 months (5.0 months)	49R/54L/0B
WU: BNT (n = 135)	53.61 (10.77)	72M/63F	13.16 (2.53)	124R/11L/0Mi	13 days (6 days)	69R/66L/0B
WU: CIM (n = 135)	53.61 (10.77)	72M/63F	13.16 (2.53)	124R/11L/0Mi	13 days (6 days)	69L/66R/0B
WU: HVLTL (n = 103)	53.29 (11.22)	58M/45F	13.33 (2.65)	94R/9L/0Mi	13 days (6 days)	47L/56R/0B
WU: ARAT (n = 133)	53.55 (10.83)	71M/62F	13.14 (2.53)	122R/11L/0Mi	13 days (6 days)	68L/65R/0B

Lesion chronicity for the Iowa and NICHE cohorts represents the number of months between lesion onset and the behavioural data collection; lesion chronicity for the WU cohort represents the number of days between lesion onset and the date of neuroimaging data collection; years of education were not collected for the NICHE cohort. ARAT = Action Research Arm Test; B = bilateral; BNT = Boston Naming Test; CIM = Complex Ideational Material; F = female; HVLTL = Hopkins Verbal Learning Test; IQR = interquartile range; L = left; M = male; Mi = mixed; R = right; RAVLT = Rey Auditory Verbal Learning Test; SD = standard deviation.

NICHE cohort

The NICHE cohort included data from two sites collected as part of the NICHE motor rehabilitation clinical trial (<https://clinicaltrials.gov>): the Burke Medical Research Institute and the Rancho Los Amigos National Rehabilitation Center. Inclusion criteria included: having a stroke that resulted in hemiparesis in the 3–12 months prior to enrollment (prior stroke with complete recovery was allowed), and the ability to follow instructions. Exclusion criteria were: fixed joint contracture at the upper limb, a history of seizures, other comorbid neurological conditions, and implanted devices. Additional details of the cohort have been described previously.³⁸ Prior to any rehabilitation intervention each participant underwent baseline motor evaluation performed by a trained clinician using the upper extremity Fugl-Meyer assessment.³⁹ These baseline Fugl-Meyer scores of the affected extremity were used for LBM. A structural MRI scan was acquired that included a magnetization-prepared rapid gradient-echo sequence with 1 mm isotropic voxel size. Demographic data are reported in Table 1.

Washington University cohort

The WU cohort consisted of participants with stroke who completed a range of cognitive and motor tests. We focused on those tests that measured a domain that was also captured in either the Iowa cohort or the NICHE cohort: the short form of the Boston Naming Test (a test of expressive language and naming), the Complex Ideational Material Test (a test of receptive language), the delayed recall trial from the Hopkins Verbal Learning Test—Revised (a test of anterograde verbal memory), and the Action Research Arm Test for the limb contralateral to the lesion (a test of upper extremity motor function). Cognitive test scores were normed as described previously;⁴⁰ the raw scores were used for the delayed recall trial of the Hopkins Verbal Memory Test due to the bimodal distribution that was created after norming. Imaging was performed within the first two weeks after the stroke as described previously.⁴⁰ Demographic data are reported in Table 1.

Lesion segmentation

Lesions in each cohort were manually segmented in three dimensions using MRI by a rater blind to behavioural outcome data. In rare cases for the Iowa cohort, a CT was used if MR contraindications were present. A total of 36, 14, and 37 CT scans

were included in the Iowa samples for the Boston Naming Test, the Token Test, and the Rey Auditory-Verbal Learning Test, respectively. The anatomical accuracy of each lesion tracing was reviewed by a neurologist in both native space and upon transformation to the MNI152 1 mm template brain using a combination of linear and non-linear registration techniques. Additional details of lesion segmentation are provided as [Supplementary material](#).

Multivariate lesion-behaviour mapping

Multivariate LBM was performed using the LESYMAP package available in R.⁴¹ Specifically, LESYMAP's sparse canonical correlation analysis was used to generate a brain-wide map of voxel weights that indicate the association between regions of brain damage and behavioural outcome data.¹⁰ The sparse canonical correlation analysis method uses an optimization procedure to derive the pattern of voxel weights that maximizes the multivariate correlation between voxel values and behavioural scores. The validity of the map is determined through a 4-fold within-sample cross validation; a model is built using 75% of the data and then applied to the remaining 25% to generate predicted scores, where the correlation between predicted and observed scores is used to evaluate the 'significance' of the overall map. By testing the significance of the entire map at once, rather than of individual voxels, this approach avoids some pitfalls associated with mass-univariate methods, such as inflated rates of false positive errors and erroneous maps due to functions associated with distributed brain regions.⁴² Prior validation studies support the accuracy of the sparse canonical correlation analysis method over mass univariate methods, along with improved ability to identify multiple foci of damage associated with behavioural deficits.¹⁰ Moreover, multivariate LBM approaches such as sparse canonical correlation analysis are better able to mitigate the influence of spatial distortion that is occurs due to collinear and stereotyped spatial patterns of damage, which simulation studies have shown can significantly affect the local and distributed lesion-behaviour patterns.^{43,44} We excluded voxels with minimal coverage (fewer than three lesions) as performed previously.^{16,45}

LBM was performed for each neuropsychological test from the Iowa cohort and for the Fugl-Meyer test from the NICHE cohort. Each map was then used to generate predictions of behavioural outcomes in the WU cohort (described further below).

The LBMs from the Iowa and NICHE cohorts were also used to generate seed regions-of-interest for structural and functional

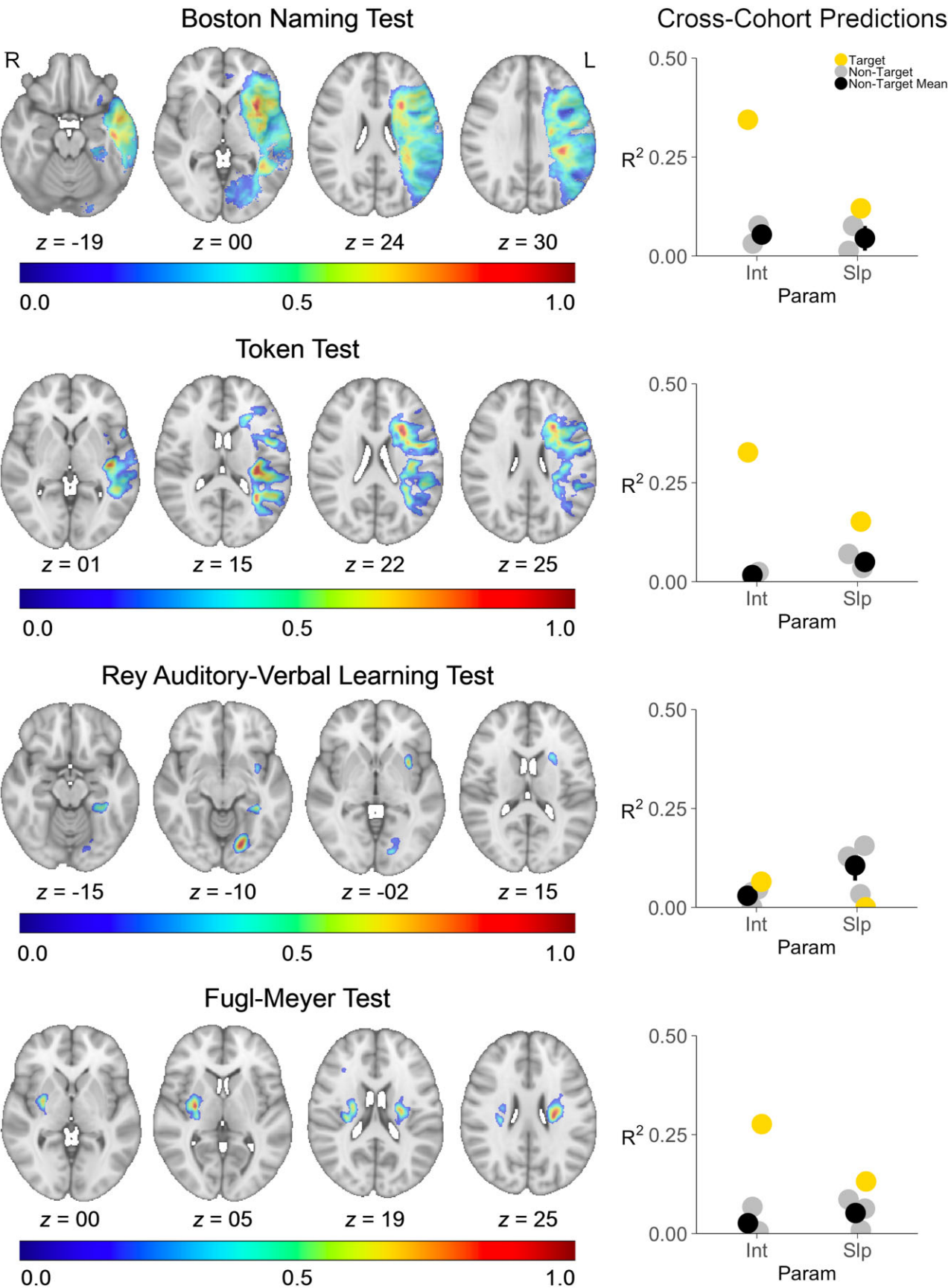


Figure 2 LBM results and LBM-LL predictions of recovery trajectories. LBM-LL scores were calculated for each LBM and used to predict the intercept (which represents performance at 12 months post-stroke) of the recovery trajectories of the cognitive and motor functions measured in the WU cohort.

(Continued)

LNM analyses. First, we used FSL's cluster function to derive the coordinates of the local maxima within each LBM. To avoid issues with mixing grey and white matter signals,^{33,46} we separated the coordinates into atlas-defined grey and white matter.⁴⁷ Five millimetre spherical regions of interest were placed at local maxima located within grey matter for functional LNM analyses and within white matter for structural LNM analyses.

Structural lesion network mapping

The resulting regions of interest in white matter were used to 'seed' deterministic tractography analyses using LEAD-DBS software. In deterministic tractography, the local directions of white matter streamlines are estimated for a sample of healthy individuals, and local directions are combined to reconstruct white matter tracts. We used the lead mapper function to determine the number of white matter streamlines projecting from each seed region based on diffusion MRI data collected from neurologically healthy individuals included in the Human Connectome Project's MGH 32-fold group connectome, as performed previously.^{16,48} The directions of the streamlines were not constrained to any other region-of-interest beyond the LBM-derived seed in question (i.e. there was no predefined end point). Thus, the resulting voxel-wise maps represented the number of streamlines that pass through both a given voxel and LBM-derived seed in question. We used these voxel-wise maps to predict stroke outcomes in the WU cohort, as described below.

Functional lesion network mapping

Grey matter regions of interest were used to 'seed' resting-state functional connectivity MRI (rs-fcMRI) analyses based on data from a large cohort of 303 healthy adults included in the Human Connectome Project. Processing of the rs-fcMRI data is described in detail elsewhere.⁴⁹ Briefly, participants completed two 6.2 min rs-fcMRI scans during which they were asked to rest in the scanner (3 T, Siemens) with their eyes open (TR = 3000 ms, TE = 30 ms, FA = 85°, 3 mm voxel size, FOV = 216, 47 axial slices with interleaved acquisition and no gap). We first calculated the temporal correlation between the average resting-state BOLD signal within the seed region-of-interest and the BOLD signal of every voxel within the brain. Correlation coefficients were converted to normally distributed Z-scores using the Fisher r-to-Z transformation. The resulting three-dimensional voxel-wise network maps were used in generating outcome predictions in the WU cohort, as described below, using both thresholded (95th percentile Z-score) and unthresholded maps.^{50,51}

Principal components analysis

When there were three or more seeds derived from a single LBM, we used principal components analysis (PCA) to reduce the LNM results to their primary axes of variation. The use of PCA served to reduce the number of separate maps that would be used to predict stroke outcomes in the WU cohort, which helped to protect against

possible issues with collinear predictors and over-fitting. We performed PCA analyses separately for the structural and functional LNM results. In both cases, the network maps were read into R using ANTSR tools (<https://github.com/ANTsX/ANTsR>). Each network map derived from a given LBM was converted into a vector and placed into a data frame. We then submitted that data frame to a PCA using the `prcomp` function in base R. To determine the number of principal components to retain, we used parallel analysis as implemented in the `psych` package in R,⁵² which compares the eigenvalues of the observed data used in the PCA to those of a random data matrix of equivalent size. The number of observed eigenvalues greater than those from the random data indicated the number of principal components to retain. We then extracted the principal component scores for each voxel using the `predict` function in base R, and back-projected them into MNI152 space. We refer to these output as the principal component maps. In order to facilitate comparisons between the structural and functional LNM PCAs, the principal component maps from the functional LNM results were resampled to 1 mm resolution using ANTSR tools.

Quantifying overlap of Washington University cohort lesions with lesion-derived maps

The 3D lesion volume from each WU participant was used to predict 12-month outcomes for each domain based on each lesion's overlap with the LBMs and structural and functional LNMs derived from Iowa and NICHE data. For the LBMs, this was performed according to LESYMAP's `lesymap.predict` function (<https://github.com/dorianps/LESYMAP>), which multiplies a matrix of the raw voxel weights from the lesion-behaviour map by a matrix of the lesion masks from the WU cohort, and then by the eigenvalue derived from the LBM analysis in question. In keeping with previous literature, we termed these scores LBM lesion load (LBM-LL). The lesion's intersection with each LNM was also quantified as the sum of voxel intensities from a lesion network map that also appeared within the boundaries of a given WU patient's lesion mask. For the network maps there was no equivalent eigenvalue used in the LBMs. The resulting 'lesion load' values are referred to as structural LNM-LL for the structural LNMs, and functional LNM-LL for the functional LNMs (Fig. 1). We focused on lesion load values for the PC maps where possible. For PC maps with positive and negative values, we created two separate lesion load values in order to avoid potential problems with summing both positive and negative voxels into a single score. In order to address our research question regarding the out-of-sample predictive utility of LNM over LBM, we evaluated the combined effect of structural LNM-LL and/or functional LNM-LL values derived from the PC maps within a given domain on the corresponding 12-month behavioural measurement in WU cohort. Thus, the number of independent variables for analyses involving LNM variables was equal to the number of network maps in the results of the LNM analyses. Dependent variables for these analyses are described below.

Figure 2 Continued

Predictions were based on the LBMs of the Boston Naming Test (A), the Token Test (B), the delayed recall of the Rey Auditory-Verbal Learning Test (C), and the Action Research Arm Test (D). For each LBM, the 'target test' was the test in the WU cohort that measured the function on which the LBM was based, and the non-target tests were all other tests. For example, the target test for the LBM of the Token Test was the WU cohort's Complex Ideational Material, as both tests measure receptive language abilities. In each panel, we graphically depicted the variance accounted for (i.e. intercept R^2 within the model) by the target map to the average variance accounted for across the non-target maps. Error bars for the mean non-target predictions represent the standard error of the mean.

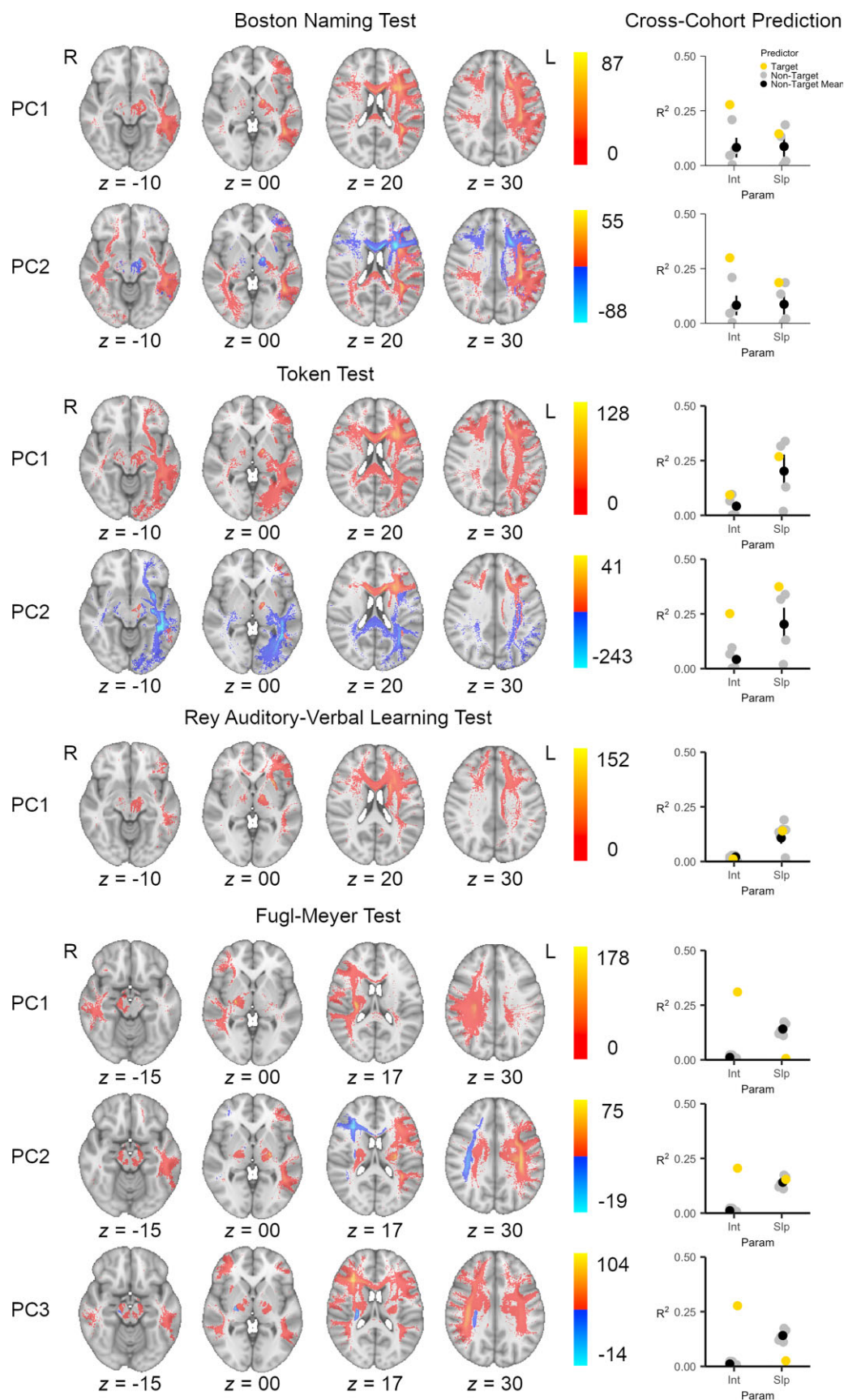


Figure 3 Structural lesion network mapping. White matter seeds were derived from the LBMs and submitted as seed regions-of-interest in a deterministic fiber tractography analysis. Each LBM was associated with three or more seeds. The results were calculated for each seed and then

(Continued)

Latent growth curve modelling

We investigated the associations between the lesion-derived measures and 12-month outcomes using latent growth curve (LGC) modelling of each behavioural domain in the WU cohort, estimated using the lavaan library's growth function.⁵³ LGC models are flexible applications of structural equation modelling that allow for the estimation of subject-specific recovery trajectories, characterized in terms of an intercept (set as the 12-month chronic outcome) and a slope (recovery trajectory) each represented by latent variables.⁵⁴ The latent variables for the intercept and slope can be flexibly defined by the specification of factor loadings of the observed variables (i.e. the repeated behavioural measurements), such that the latent variables can be used as dependent variables to explore the amount of variance in these outcomes that can be accounted for by lesion-derived variables. There are at least three advantages to LGC models that favored their inclusion here: (i) modelling chronic outcome and recovery trajectory separately allows for potentially unique associations with lesion variables, despite these variables often being correlated; (ii) LGC models evaluate and potentially incorporate random effects to account for inter-individual differences in recovery parameters; and (iii) LGC models provide a robust framework from which to impute missing longitudinal data. The total proportion of longitudinal data missing from WU participants for each test was above 5% and well below 40% missing data, which has been suggested as the range of percent missing data in which data imputation techniques should be for clinical trials research.⁵⁵ The percent of missing data by test was: Boston Naming Test (19.3%), Complex Ideational Material Test (19.0%), Hopkins Verbal Learning Test Delayed Recall (18.8%), and the Action Research Arm Test (19.0%). Additional details regarding the LGC modelling are provided in the [Supplementary Material](#).

Comparing LBM- and LNM-based predictions of 12-month behavioural outcomes

We evaluated the relative value of LBM-LL, structural LNM-LL, and functional LNM-LL for predicting 12-month outcomes in the WU cohort, which was represented as the intercept in the LGC models. Because the structural neuroimaging data from which lesion location was derived in the WU cohort was acquired within two weeks of the stroke and temporally preceded the 12-month behavioural outcomes, we use the word 'prediction' to refer to the association between lesion load variables (derived from quantifying lesion location relative to out-of-sample lesion-behaviour maps) and outcomes. However, the data used for the analysis was collected previously and the term prediction is not meant to imply that this was a prospective study.

Importantly, the following analyses are not an effort to cross-validate statistical models from the Iowa cohort, but rather, the WU cohort provides an out-of-sample cohort with which to test the relative predictive value of the different lesion load variables. First, we compared these predictors using three separate models, one for each lesion-deficit mapping modality (LBM-LL, structural LNM-LL, and functional LNM-LL). In addition, we compared these three models to a fourth model in which all lesion load predictors were entered simultaneously into the model. We compared these models in terms of the variance in the intercept accounted for by the predictors (R^2), and the root mean square error of approximation (RMSEA) and Akaike Information Criterion (AIC) associated with each model. Whereas the R^2 and RMSEA values provided indices of effect size and absolute model fit, respectively, the AIC values provided an index with which to compare relative model fit while penalizing for model complexity. AIC can be particularly useful when comparing individual versus combined models with multiple predictors. A more complex model could be associated with a higher R^2 simply due to the large number of predictors in the model; however, such a model would suffer in terms of AIC (relative to other models) if the model is unnecessarily complex.

In addition, we explored whether the LNM modalities predicted unique or overlapping variance with LBM-LL and with one another. To do this, we first created null models in which we entered all relevant variables into the model and set the regression parameters associated with the predictors of interest (either structural LNM-LL or functional LNM-LL) to zero. In an alternative model, we allowed all regression parameters to be freely estimated. The statistical significance of each effect was evaluated using the change in model fit from the null model to alternative model, as indicated by the chi-squared difference test and by an inspection of the individual regression parameters (i.e. the local fit).

Evaluating the added value of lesion information for the prediction of chronic outcomes beyond acute behavioural measurements

Because the acute behavioural measurement was already set to load onto the intercept in the LGC models described above, we first extracted the intercept values for each participant and then used these values in a standard linear regression performed using base R functions. Specifically, we used hierarchical linear regressions to explore whether the lesion-deficit mapping techniques (i.e. LBM-LL, structural LNM-LL, and functional LNM-LL) could predict variance in 12-month post-stroke outcomes beyond that accounted for by acute behavioural measurements alone. Acute behavioural measurements were entered in at the first step, and the lesion-deficit mapping data in

Figure 3 Continued

submitted to a PCA. The seeds associated with the map of the Boston Naming Test identified the white matter tracts between the left frontal, parietal, and temporal cortices as critical for expressive language and naming. The seeds associated with the map of the Token Test identified tracts between the left frontal, parietal, and temporal cortices as being critical for receptive language. The seeds associated with the map of the Rey Auditory-Verbal Learning Test Delayed Recall identified tracts within the left inferior temporal cortex and the left frontal lobe as being critical for anterograde verbal memory. The seeds associated with the map of the Fugl-Meyer Test identified tracts descending from the bilateral primary motor cortices as being critical for upper extremity motor function. Colour scales for the principal component maps indicate the corresponding principal component score at each voxel. For each map, we also depict the amount of variance that can be predicted (R^2) in the WU cohort's corresponding behavioural test's latent growth curve intercept. We also present the average variance explained in each WU cohort test when predicted from the non-target functional LNM maps (i.e. the principal component structural LNM maps from the other behavioural domains). Error bars represent standard error. ARAT = Action Research Arm Test; BNT = Boston Naming Test; CIM = Complex Ideational Material Test; HVLT = Delayed Recall from the Hopkins Verbal Learning Test; Int = Intercept for the WU Cohort Latent Growth Curve Model; Param = Type of WU Cohort Latent Growth Curve Parameter; Slp = Slope for the WU Cohort Latent Growth Curve Model.

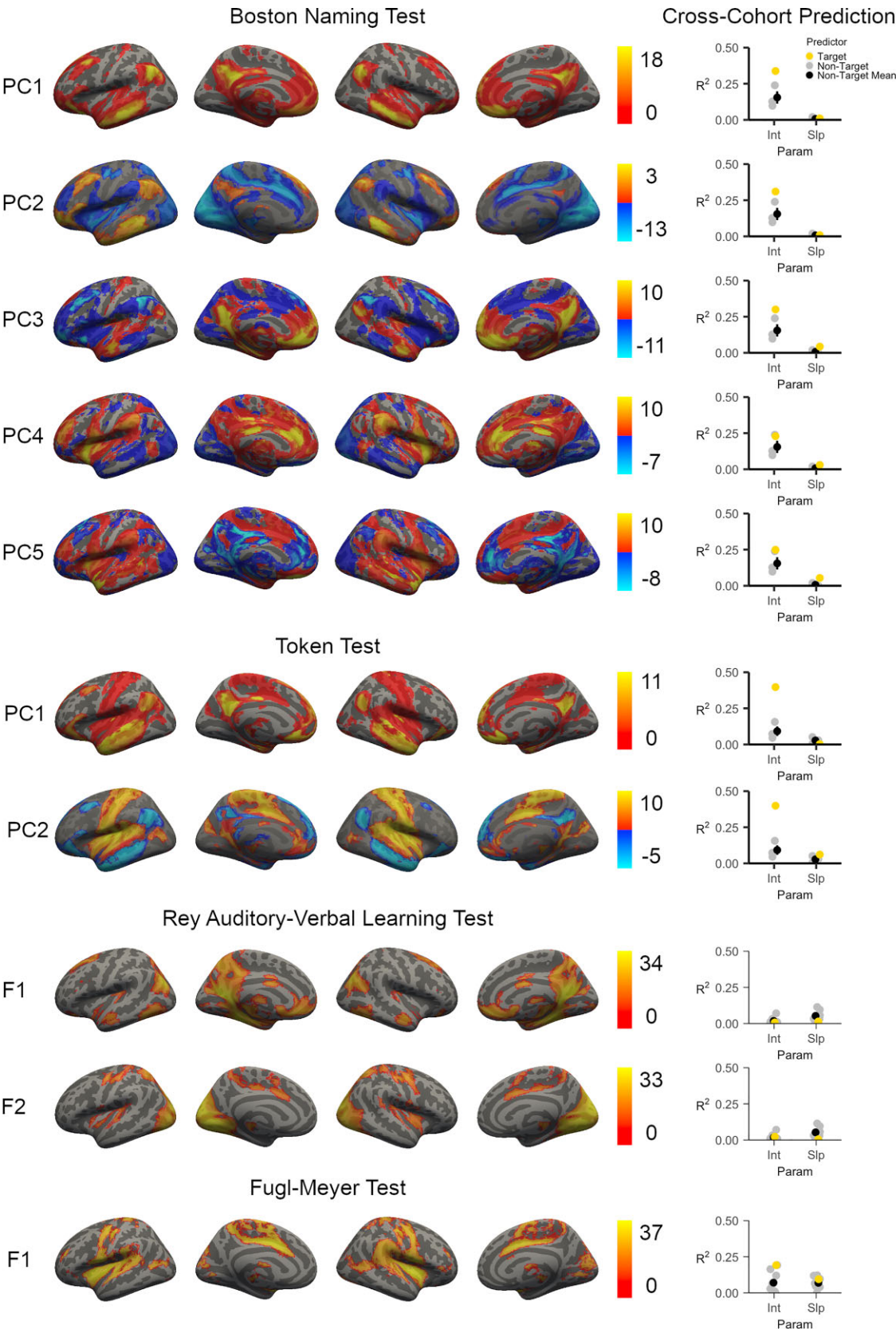


Figure 4 Functional lesion network mapping. Grey matter seeds were derived from the LBMs of each domain and submitted as seed regions-of-interest for lesion network mapping analyses. Results based on the LBM of the Boston Naming Test and the Token Test were reduced using PCA, which describes major axes of variation across the individual functional lesion network maps. Principal component maps are ordered in terms of the amount of variance they captured across the individual functional lesion networks maps (F). Principal components with positive and

(Continued)

the second step. We evaluated the statistical significance of the value added using the change in R^2 in the chronic measurements as applied by the modelCompare function in the R package lmSupport (<https://github.com/cran/lmSupport>). Because it may be of clinical interest, we also report the total variance explained by the combination of acute behavioural data and lesion-deficit mapping information.

Data availability

The data that support the findings of this study are available upon reasonable request.

Results

Demographic data are reported in Table 1. The distribution of lesions for each sample is shown in Supplementary Figs 1 and 2.

Lesion-behaviour and lesion network mapping results

LBM results were robust and statistically significant for each behavioural measure (Fig. 2). Expressive language was measured with the Boston Naming Test and was left-lateralized with the strongest regional peaks within the left anterior insula, the left frontal operculum, and the left arcuate fasciculus. This map is shown in Fig. 2A and had a within-sample cross-validation correlation of $r=0.617$ ($P<0.001$, optimal sparseness=0.627). Receptive language was assessed with the Token Test and was similarly left-lateralized with the strongest regional peaks in the left superior longitudinal fasciculus, the left frontal aslant tract, the left parietal operculum, and the left posterior insula (cross-validation correlation=0.555, $P<0.001$, optimal sparseness=−0.509). Anterograde verbal memory tested with the Rey Auditory-Verbal Learning Test localized to the left fusiform and parahippocampal cortices, deep left frontal white matter, and the left sub-insula/claustrum (cross-validation correlation=0.344, $P<0.001$, optimal sparseness=−0.061). Finally, motor deficits tested with the Fugl-Meyer Test included peaks within the descending corticospinal tracts of the corona radiata, the right somatomotor region of the putamen and posterior limb of the internal capsule (cross-validation correlation=0.357, $P<0.001$, optimal sparseness=0.313).⁵⁶ Each regional peak in the white and grey matter was used to seed structural LNM and functional LNM results, respectively. The coordinates for each peak are included in Supplementary Table 1. The structural and functional lesion network maps derived from these regions are depicted in Figs 3 and 4, respectively.

Washington University cohort latent growth curve models

Latent growth curve model path diagrams and individual-level latent growth curves are depicted in Fig. 5. The final models for each domain included: a random intercept, a random quadratic slope, and equality constraints on the unique variances of the observed variables. Model fit statistics for the final models (Supplementary Table 3) indicated acceptable to excellent model fit to the observed data across the four models. Importantly, the superiority of random slopes and intercepts over fixed slopes and intercepts within these models suggested that individual patients demonstrate their own unique rates of change and 12-month outcomes, such that single group-level estimates of these variables (such as those used in standard linear regression) are less accurate in modelling stroke recovery across individuals. Additionally, the superiority of a quadratic slope over a linear slope in all models suggested that most recovery occurred earlier rather than later in the time course of stroke recovery. The latent slopes and intercepts were not statistically significantly correlated within any of the models (r values ranged between −0.123 and 0.135; P values ranged between 0.380 and 0.927), suggesting that one's rate of recovery is not associated with one's 12-month outcome. For the model of expressive language and naming (measured by the Boston Naming Test), the best-fitting model indicated that the unique variances of the measurements were not equivalent; however, subsequent models in which predictors of the intercept and slope were included failed to converge in the presence of freely estimated unique variances. Thus, the final model for this domain utilized an equality constraint on the unique variances, which allowed subsequent models to converge normally.

Can LBM-LL predict 12-month outcomes?

For each domain, we first evaluated whether lesion location in the WU cohort relative to the LBMs was able to predict 12-month outcomes. To assess the specificity of these predictions we also made these same predictions with each of the non-target LBMs (e.g. using LBMs of receptive language, expressive language and verbal memory scores to predict motor outcomes would each be an example of non-target predictions). Given the well-established association between different aspects of language, we did not include receptive language maps in the non-target predictions for the expressive language map and vice versa. The LBM of the Boston Naming Test predicted 34.4% of the variance in WU Boston Naming Test performance at 12-months, compared to the average variance explained by the non-target maps at 5.4%. The Token Test LBM explained 32.9% of the variance in receptive

Figure 4 Continued

negative voxels indicate that the distinction between the positive and negative voxels captures a distinction present across the maps. Five principal components were identified for the functional networks linked to performance on the Boston Naming Test, with most results converging on a fronto-parieto-temporal network (PC1). Two principal components were linked to the Token Test, and these seeds also converged primarily on a fronto-parieto-temporal network (PC1). Results based on the delayed recall trial from the Rey Auditory Verbal Learning Test identified only two networks: a lateral occipital-precuneate network, and a network spanning primary and secondary visual cortices. The results based on the Fugl-Meyer Test were linked to a single rolandic-insular-opercular functional network. Colour scales for the PC maps indicate the principal component score at each voxel; otherwise, colour scales represent the Z-value associated with the statistical test of functional connectivity at each voxel. The ability of each map to predict variance (R^2) in the corresponding behavioural test's latent growth curve intercept (Int) for the WU cohort is presented alongside each map. We also present the average variance explained in each corresponding WU cohort test when predicted from the non-target functional LNM maps (i.e. the functional LNM maps and/or PC functional LNM maps from the other behavioural domains). Error bars represent standard error. Int=Intercept for the WU Cohort Latent Growth Curve Model; Param=Type of WU Cohort Latent Growth Curve Parameter; Slp=Slope for the WU Cohort Latent Growth Curve Model.

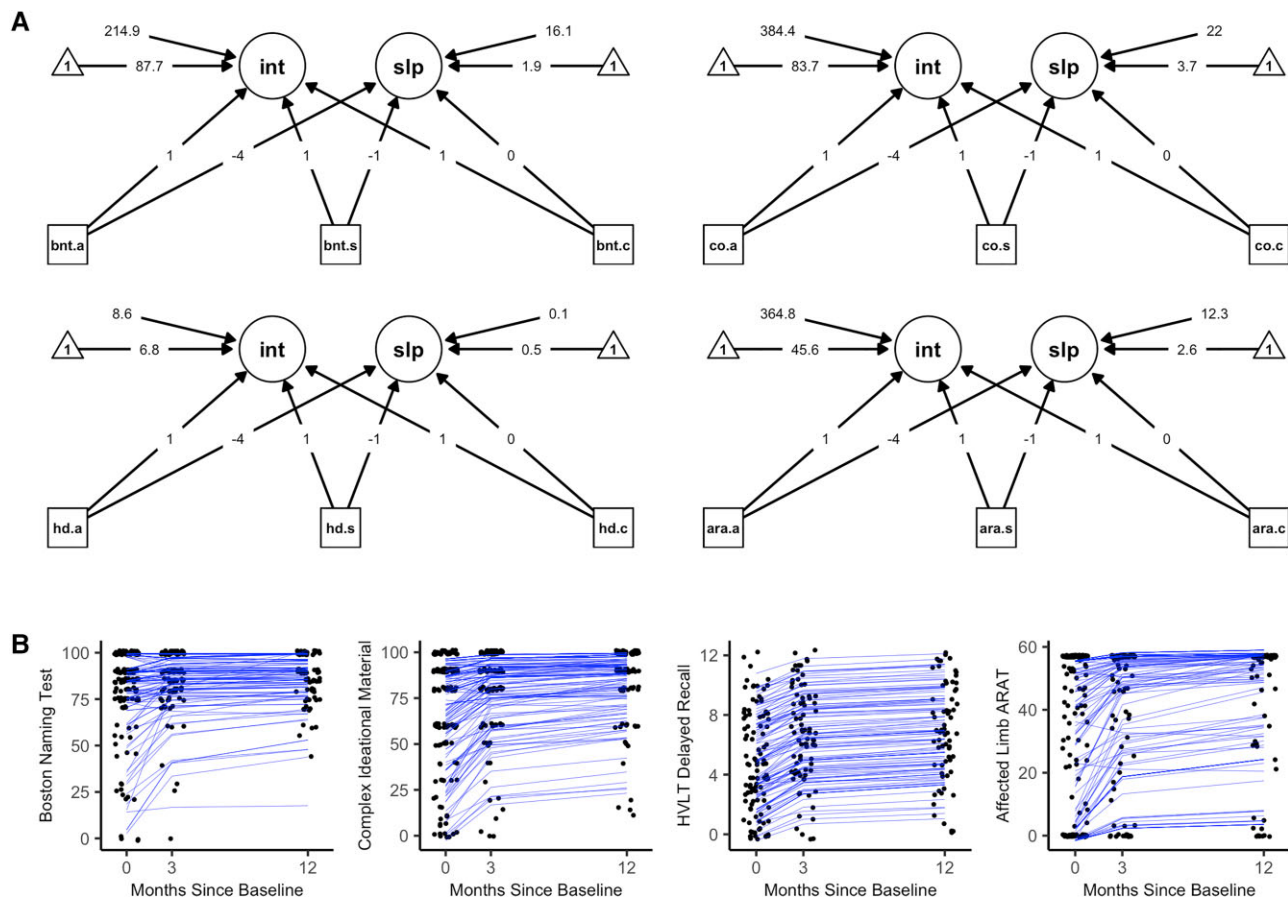


Figure 5 Latent growth curve models. (A) The path diagrams for the latent growth curve models. (B) A depiction of the individual-level latent growth curves using jittered data-points and blue subject-specific growth curves (i.e. model-based recovery trajectories), each with their own intercept at 12-months post-baseline and their own slope of recovery that vary around the mean of the latent intercept (int) and slope (slp) variables. Missing data due to participant attrition at the 3- and 12-month measurements were imputed using Full Information Maximum Likelihood. Arrow from observed variables (squares) to latent variables (circles) represent loadings specified from the Latent Growth Curve analyses. Loadings onto the slope were determined through comparisons of model fit to the data for linear and quadratic modeled change over time. Loadings onto the intercept were all set to one, as is standard in Latent Growth Curve analyses. Arrows from triangles represent the mean of the latent variables. Arrows from free floating numbers to the latent variables represent the variance of the latent variables. Model fit to the observed data is presented in [Supplementary Table 2](#). .a = Acute Measurement; ara = Action Research Arm Test; bnt = Boston Naming Test; .c = Chronic Measurement; co = Complex Ideational Material Test; hd = Hopkins Verbal Learning Test Delayed Recall Trial; int = Lateng Growth Curve Intercept (12-month time point); .s = Subacute Measurement; slp = Lateng Growth Curve Slope.

language relative to 1.5% of non-target predictions. The LBM of delayed recall trial from the Rey Auditory-Verbal Learning Test accounted for 6.5% of the variance of anterograde verbal memory performance compared to 2.9% of non-target predictions. The Fugl-Meyer LBM accounted for 27.7% of the variance in the upper extremity motor function compared to 2.5% of non-target predictions.

For each LBM the percent variance explained in the WU cohort was similar to values seen in the 'within-sample' cross-validation. The only exception was for verbal memory, with 12% of the variance explained based on within-sample cross-validation correlation versus 6.5% out-of-sample.

Do brain networks add value to the prediction of 12-month outcomes beyond lesion-behaviour maps?

To address our primary research question, we first evaluated whether the network maps added value to the cross-cohort predictions beyond variance explained by LBM alone. For the expressive and receptive language domains, the addition of LNM

variables helped to explain statistically significantly more variance than was explained with LBM alone ([Table 2](#)). In the motor domain, only structural LNM-LL added statistically significant value beyond LBM. Network variables did not statistically significantly contribute to the prediction of anterograde verbal memory outcomes. The association between the final structural and functional LNM results and the WU cohort 12-month intercept and slope from the WU cohort are presented in [Figs 4 and 5](#), respectively.

Comparisons among lesion-derived measures

Next, we compared LBM-LL, structural LNM-LL, and functional LNM-LL using a head-to-head comparison of model fit indices ([Table 3](#)). Functional LNM-LL provided the strongest prediction of the 12-month outcomes for the language domains, whereas structural LNM-LL provided the strongest prediction of outcomes in motor domain. The variance in 12-month outcomes explained by all variables combined (LBM, structural LNM, functional LNM) was 46.3%, 50.8%, 8.1%, and 37.9% for the maps based on the

Table 2 Relative contributions of structural LNM-LL and functional LNM-LL to the prediction of latent growth curve intercept and slope parameters beyond LBM-LL and one another

Model comparison	Chi-squared difference	ΔR^2 12-month intercept	ΔR^2 Slope
Boston Naming Test			
sLNM-LL Over LBM-LL	19.12**	0.021	0.182
fLNM-LL Over LBM-LL	29.65*	0.123	0.127
sLNM-LL Over fLNM-LL	36.00***	0.042	0.177
fLNM-LL Over sLNM-LL	30.47*	0.129	0.064
Complex Ideational Material Test			
sLNM-LL Over LBM-LL	24.9***	0.117	0.29
fLNM-LL Over LBM-LL	22.00**	0.141	0.158
sLNM-LL Over fLNM-LL	33.63***	0.038	0.347
fLNM-LL Over sLNM-LL	30.79***	0.206	0.035
Hopkins Verbal Learning Test Delayed Recall			
sLNM-LL Over LBM-LL	4.33	−0.006	0.144
fLNM-LL Over LBM-LL	2.8	0.019	0.02
sLNM-LL Over fLNM-LL	8.59*	0.021	0.142
fLNM-LL Over sLNM-LL	5.16	0.04	0.014
Action Research Arm Test			
sLNM-LL Over LBM-LL	18.31*	0.092	0.064
fLNM-LL Over LBM-LL	0.053	0.000	0.000
sLNM-LL Over fLNM-LL	42.99***	0.19	0.088
fLNM-LL Over sLNM-LL	3.78	0.012	0.003

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

ΔR^2 = Change in R^2 from null model to alternative model; fLNM = functional LNM; sLNM = structural LNM.

Boston Naming Test, the Complex Ideational Material Test, the Hopkins Verbal Learning Test, and the Action Research Arm Test, respectively. For each test, the R^2 and RMSEA values were superior for the model with all of these predictors in the context of marginally higher AIC values. Overall, this suggests that the combination of these variables was superior to the use of any one variable alone.

The association between lesion load for each individual LNM (prior to PCA) and the WU cohort 12-month intercept is presented in [Supplementary Table 4](#). These same associations after controlling for age, sex, and lesion volume are presented in [Supplementary Table 5](#). For functional LNM, comparison of thresholded and unthresholded functional LNM yielded similar results, with slightly greater variance explained using thresholded results. Similarly, predictions were comparable whether negative correlations (anticorrelations) were included in the functional LNM results or not.

Does lesion location predict one's rate of recovery (slope)?

In addition to 12-month outcomes, we also performed the same analyses to evaluate the role of lesion location and lesion-associated networks in predicting the rate of recovery, which was represented as the slope in the LGC models. Across our analyses, structural LNM-LL was the most robust predictor of the recovery slope ([Fig. 3](#) and [Tables 2](#) and [3](#)). Functional LNM-LL was generally not predictive of the recovery slopes across behavioural domains ([Fig. 4](#) and [Tables 2](#) and [3](#)). Predictive models that combined all lesion load variables (LBM-LL, structural LNM-LL, and functional LNM-LL) accounted for 36.6%, 45.9%, 18.5%, and 19.7% of the variance in the slope for the Boston Naming Test, the Complex Ideational Material Test, the Hopkins Verbal Learning Test, and the Action Research Arm Test, respectively.

Table 3 Comparing LBM-LL, structural LNM-LL, functional LNM-LL and combined models

Predictors	12-Month intercept R^2	Slope R^2	RMSEA	AIC
Boston Naming Test				
LBM-LL	0.344	0.120	0.241	2584.93
sLNM-LL	0.305	0.287	0.205	2585.19
fLNM-LL	0.398	0.178	0.152	2613.26
All	0.463	0.366	0.126	2589.86
Complex Ideational Material Test				
LBM-LL	0.342	0.149	0.077	2887.93
sLNM-LL	0.289	0.419	0.070	2893.59
fLNM-LL	0.457	0.107	0.040	2896.42
All	0.508	0.459	0.049	2875.02
Hopkins Verbal Learning Test Delayed Recall				
LBM-LL	0.065	0.000	0.15	1169.32
sLNM-LL	0.011	0.140	0.055	1168.94
fLNM-LL	0.030	0.012	0.075	1176.37
All	0.081	0.185	0.122	1172.55
Action Research Arm Test				
LBM-LL	0.277	0.131	0.026	2602.79
sLNM-LL	0.370	0.180	0.014	2602.74
fLNM-LL	0.192	0.095	0.047	2625.95
All	0.379	0.197	0.00	2605.49

AIC = Akaike Information Criterion; fLNM = functional LNM; RMSEA = root mean square error of approximation; sLNM = structural LNM.

Can lesion information add value to the prediction of chronic outcomes beyond acute behavioural data?

Lastly, we explored the combined predictive power of all lesion location metrics for 12-month outcomes (LBM-LL, structural LNM-LL, and functional LNM-LL) relative to that of acute behavioural measurements alone, which is known to have a robust association with chronic outcomes.^{40,57,58} When used in combination, the lesion-deficit mapping metrics were able to account for a statistically significant amount of variance in the domain of expressive language ($\Delta R^2 = 0.088$, $P = 0.005$) and receptive language ($\Delta R^2 = 0.075$, $P < 0.001$) beyond acute behavioural scores. In contrast, acute lesion metrics could not add meaningful value to the prediction of anterograde verbal memory ($\Delta R^2 = 0.007$, $P = 0.467$) or upper extremity motor function ($\Delta R^2 = 0.027$, $P = 0.109$) beyond acute behavioural data. Altogether, the lesion-deficit mapping metrics and acute behavioural data combined to predict well over half of the variance in chronic outcomes of expressive language (67.2%), receptive language (71.7%), anterograde verbal memory (80.7%), and upper extremity motor outcomes (72.6%).

Comparison of structural lesion network mapping approaches

As a final post hoc analysis we also compared the current approach for structural LNM to an existing method that uses each individual lesion mask to 'seed' a tractography network. While we avoid this approach for functional LNM due to the 'large lesion confound' discussed earlier, this averaging of signal within a large lesion

does not apply to tractography data.³³ To perform this analysis, each lesion from the Iowa and NICHE cohorts was used to seed a deterministic tractography analysis using the LEAD DBS software as described above. These resulting maps were binarized and used in place of lesion masks in LESYMAP to produce a single structural LNM map for each functional domain. The predictive value of this approach was compared to the LBM-derived structural LNM approach. The results were similar between the two methods. The individual lesion-based seeding strategy predicted more variance in language, but less in motor, with similar results for memory. The individual lesion strategy produced higher RMSEA values (indicating worse model fit), and the AIC values marginally lower (indicating better fit), for all models compared to the LBM-based approach. Overall, there was inconclusive evidence regarding the superiority of one approach to structural LNM over the other.

Discussion

The neuroanatomical localization of symptoms has been a foundational principle in neurology and the clinical neurosciences for more than a century. It involves identifying a set of neurological signs and symptoms and pinpointing a likely site of pathology in the nervous system that would account for the symptoms. This process guides diagnostic testing and informs management decisions. Much of the pioneering work in localization of symptoms was done in the 19th century before brain imaging was available. Using modern approaches with neuroimaging and advanced analytical techniques, we are now poised to also proceed in the reverse direction of classical lesion localization. That is, by starting with the lesion location derived from clinically acquired imaging, we may be able to make personalized predictions of symptoms and long-term prognosis, guiding more focused cognitive assessments and informing personalized rehabilitation strategies. The current analysis is a step towards this goal of incorporating new computational technology and brain mapping to invert the classical progression of mapping symptoms-to-anatomy, and instead mapping anatomy to predict symptoms, which can then be evaluated and confirmed with subsequent, targeted behavioural examinations.

Here, we utilized large-scale lesion-behaviour mapping to more precisely identify specific brain regions that are maximally associated with motor and cognitive deficits. We used these maps to derive the associated structural and functional networks of these regions using normative data. We then evaluated the role of these maps in predicting post-stroke deficits in an independent cohort of patients with stroke. To our knowledge, no previous study has evaluated the additive value of LNM above and beyond LBM in this way. In addition, we used a Latent Growth Curve (LGC) modelling framework to simultaneously evaluate the associations between lesion-based metrics, and individual differences in 12-month outcomes and rates of recovery. Our results demonstrated that lesion location relative to LBM provides a robust predictor of 12-month functional outcomes. Moreover, we found that functional and structural LNM predicted variance in outcomes and/or recovery rates beyond those provided by LBM. Though the best individual predictor varied by domain, the combination of LBM and structural and functional LNM consistently performed the best. Finally, we found that lesion-based metrics explained unique variance in 12-month language outcomes beyond what could be predicted from acute

behavioural measurements. Our results support the conclusion that LBM and structural and functional LNM can be used to predict post-stroke outcomes, and that each method captures distinct information.

Our results demonstrated that behavioural domains are differentially associated with structural and functional brain network data. We found support for the additive utility of structural and functional brain networks for predicting expressive and receptive language deficits. We also observed that motor outcomes were best predicted by structural connectivity data, and receptive language outcomes by functional connectivity data. One potential explanation for this divergence across functional domains is that structural network data might be better suited to the prediction of deficits linked to networks with a final common pathway, such as the corticospinal tract.^{7,59,60} By contrast, functional connectivity data may be better able to predict deficits associated with damage to complex, bidirectional, and dynamic networks, such as those classically associated with language,^{61–63} or networks that depend on short-range, local connectivity rather than long-distance projection fibers.⁶⁴ Overall, a consistent finding across the functional domains was that the greatest strength of prediction was observed when combining lesion location with both structural and functional network data. We suggest that it remains important, both clinically and experimentally, for future studies to investigate brain-behaviour relationships using multiple methodologies so as to further elucidate the respective contributions of specific structural disconnections and functional brain network alterations to lesion-associated deficits.^{31,65,66}

Previous work has demonstrated that comprehensive neuropsychological and motor evaluations in the acute phase of stroke can predict a substantial amount of variance in chronic post-stroke outcomes,^{40,57,58} and it is acute behavioural scores that serve as a relative gold standard for outcome predictions. However, the type of formal behavioural assessments conducted at discrete time periods after stroke that aid in outcome predictions are not routinely acquired outside of research studies. Here, we show that lesion-derived measures that can be easily acquired from clinically acquired structural imaging predict a similar amount of variance in 12-month cognitive and motor outcomes. Moreover, we demonstrate evidence for the added value of lesion information for predicting some aspects of behavioural outcomes beyond acute behavioural measurements. It is likely that behavioural and lesion data can be combined to enhance predictions. This type of combined approach may represent an improvement over a strict reliance on proportional recovery projections that may lead to over-estimates about the proportion of variance in chronic outcomes that can be explained by acute measurements alone.^{67–69} However, further research will be required to identify the best combination of predictors for the different behavioural domains of interest.

The present study also highlights important anatomical correlates of several cognitive-behavioural functions. Broadly, we found that multiple structural and functional brain networks were associated with each domain. We identified both the anterior and posterior aspects of the arcuate fasciculus, as a critical anatomical correlate of both expressive and receptive language.^{70,71} Both domains of language were also well predicted by the functional connectivity among frontal, parietal, and anterior temporal cortices. One explanation for the observation of shared networks for expressive and receptive language is that these functions both involve aspects of cognitive control and lexical retrieval, associated with the fronto-parietal and anterior temporal cortices,

respectively.^{16,72–77} However, further work will be needed to determine the necessary contributions of these functions to the clinical measurement of expressive and receptive language.^{78,79}

A potential limitation of our study is the limited lesion coverage in the WU cohort of the regions identified as important for anterograde verbal memory in the Iowa cohort. It may be the case the lesion data were weakly predictive for this domain because the WU cohort was largely composed of individuals with strokes involving the MCA territory. However, previous work suggests this type of deficit is best captured by within-subject functional connectivity measures. Thus, we and others may have failed to observe predictions for verbal memory due to the use of normative datasets for network analyses.^{31,80} Another important caveat to our findings is that our analyses were based on research-quality MRI data; for these findings to translate to clinical practice, our approach would need to be integrated with a fast and reliable method for lesion segmentation based on clinically acquired imaging data. Also, the exact test used to measure cognitive and motor domains differed between cohorts. For example, the different cohorts in our study used the Fugl-Meyer or the ARAT to assess upper extremity motor performance. It is possible that the variance in outcomes predicted here would be higher if the same tests were used across cohorts, though ultimately it may be viewed as a strength that localization is robust to the underlying function and not limited to a specific type of measurement. Importantly, several methodological factors limit direct comparisons of the current analyses to previously published work using the WU cohort data, namely differences in lesion etiology, lesion chronicity, sample size, and lesion spatial distribution between the WU, Iowa, and NICHE cohorts. It will be important for future work to explore whether the variance in behavioural outcomes accounted for by previous studies using the WU data is the same or different than the variance accounted for in the present study. A number of alternative methods for quantifying network damage exist that were not evaluated here,^{27,81,82} and it will be important to compare these methods with those used here to potentially further optimize predictive accuracy. Finally, future studies could seek to explore the sample sizes and lesion distributions necessary for multivariate LBM alone to detect distributed network-behaviour dependencies. LNM likely adds more value to LBMs based on relatively small or anatomically constrained samples. In contrast, LBMs based on a sufficiently large and anatomically diverse sample should be able to elucidate network-behaviour relationships without LNM. The precise parameters at which this occurs remain unknown.

In summary, the present study provides evidence for the importance of lesion location and lesion-associated brain networks in the prediction of post-stroke deficits. Moreover, we demonstrate evidence for the validity of large-scale lesion-behaviour mapping for the identification of seed regions-of-interest for connectivity analyses. Our work adds to a growing body of evidence regarding brain network-behaviour relationships and represents a step towards clinical translation of basic tools in cognitive neuroscience.

Acknowledgements

We would like to thank Joseph Griffis for reviewing the manuscript prior to submission. We would also like to thank all of the patients that participated in the research that contributed to this manuscript.

Funding

This study was supported by the National Institute of General Medical Sciences (T32GM108540), the National Institutes of Mental Health (1 P50 MH094258; 1 R21 MH120441-01), the Kiwanis Foundation, FC-Neuro University of Padua, the National Institute of Neurological Disease and Stroke (1 R01 NS114405-01; NS095741), FLAG-ERA JTC, Neuro-DiP: Progetto Dipartimenti di Eccellenza Italian Ministry of Research (MIUR), CARIPARO Foundation Padova. This work was conducted on an the National Institutes of Health instrument funded by 1S10RR028821-01.

Competing interests

The authors declare no competing interests.

Supplementary material

[Supplementary material](#) is available at *Brain* online.

References

- Lai S-M, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the stroke impact scale. *Stroke*. 2002;33(7):1840–1844.
- Gottesman RF, Hillis AE. Predictors and assessment of cognitive dysfunction resulting from ischaemic stroke. *Lancet Neurol*. 2010;9(9):895–905.
- Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243(4899):1668–1674.
- Counsell C, Dennis M, McDowall M. Predicting functional outcome in acute stroke: comparison of a simple six variable model with other predictive systems and informal clinical prediction. *J Neurol Neurosurg Psychiatry*. 2004;75(3):401–405.
- Forkel SJ, Thiebaut de Schotten M, Dell'Acqua F, et al. Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks. *Brain*. 2014; 137(Pt 7):2027–2039.
- Zhao L, Biesbroek JM, Shi L, et al. Strategic infarct location for post-stroke cognitive impairment: a multivariate lesion-symptom mapping study. *J Cereb Blood Flow Metab*. 2018; 38(8):1299–1311.
- Rondina JM, Park C-H, Ward NS. Brain regions important for recovery after severe post-stroke upper limb paresis. *J Neurol Neurosurg Psychiatry*. 2017;88(9):737–743.
- Puy L, Barbay M, Roussel M, et al. Neuroimaging determinants of poststroke cognitive performance. *Stroke*. 2018;49(11):2666–2673.
- Weaver NA, Kuijf HJ, Aben HP, et al. Strategic infarct locations for post-stroke cognitive impairment: a pooled analysis of individual patient data from 12 acute ischaemic stroke cohorts. *Lancet Neurol*. 2021;20(6):448–459.
- Pustina D, Avants B, Faseyitan OK, Medaglia JD, Coslett HB. Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*. 2018;115:154–166.
- Bates E, Wilson SM, Saygin AP, et al. Voxel-based lesion-symptom mapping. *Nat Neurosci*. 2003;6(5):448–450.
- Damasio H, Damasio AR. *Lesion analysis in neuropsychology*. Oxford University Press; 1989.
- de Haan B, Karnath H-O. A hitchhiker's guide to lesion-behaviour mapping. *Neuropsychologia*. 2018;115:5–16.

14. Lindenberg R, Renga V, Zhu LL, Betzler F, Alsop D, Schlaug G. Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. *Neurology*. 2010;74(4):280–287.
15. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. *Neuroimage Clin*. 2016;12:372–380.
16. Bowren M Jr, Adolphs R, Bruss J, et al. Multivariate lesion-behavior mapping of general cognitive ability and its psychometric constituents. *J Neurosci*. 2020;40(46):8924–8937.
17. Franz EA, Gillett G, John Hughlings Jackson's evolutionary neurology: a unifying framework for cognitive neuroscience. *Brain*. 2011;134(Pt 10):3114–3120.
18. Park H-J, Friston K. Structural and functional brain networks: from connections to cognition. *Science*. 2013;342(6158):1238411.
19. He BJ, Snyder AZ, Vincent JL, Epstein A, Shulman GL, Corbetta M. Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*. 2007;53(6):905–918.
20. Grefkes C, Fink GR. Connectivity-based approaches in stroke and recovery of function. *Lancet Neurol*. 2014;13(2):206–216.
21. Baldassarre A, Ramsey LE, Siegel JS, Shulman GL, Corbetta M. Brain connectivity and neurological disorders after stroke. *Curr Opin Neurol*. 2016;29(6):706–713.
22. Catani M, Ffytche DH. The rises and falls of disconnection syndromes. *Brain*. 2005;128(Pt 10):2224–2239.
23. Carrera E, Tononi G. Diaschisis: past, present, future. *Brain*. 2014;137(Pt 9):2408–2422.
24. Thiebaut de Schotten M, Dell'Acqua F, Ratiu P, et al. From phineas gage and monsieur leborgne to H.M.: revisiting disconnection syndromes. *Cereb Cortex*. 2015;25(12):4812–4827.
25. Boes AD, Prasad S, Liu H, et al. Network localization of neurological symptoms from focal brain lesions. *Brain*. 2015;138(Pt 10):3061–3075.
26. Fox MD. Mapping symptoms to brain networks with the human connectome. *N Engl J Med*. 2018;379(23):2237–2245.
27. Kuceyeski A, Navi BB, Kamel H, et al. Structural connectome disruption at baseline predicts 6-months post-stroke outcome. *Hum Brain Mapp*. 2016;37(7):2587–2601.
28. Del Gaizo J, Fridriksson J, Yourganov G, et al. Mapping language networks using the structural and dynamic brain connectomes. *eNeuro*. 2017;4(5):ENEURO.0204-17.2017.
29. de Schotten MT, Foulon C, Nachev P. Brain disconnections link structural connectivity with function and behaviour. *Nat Commun*. 2020;11(1):5094.
30. Hope TMH, Leff AP, Price CJ. Predicting language outcomes after stroke: is structural disconnection a useful predictor? *Neuroimage Clin*. 2018;19:22–29.
31. Salvalaggio A, De Grazia MDF, Zorzi M, de Schotten MT, Corbetta M. Post-stroke deficit prediction from lesion and indirect structural and functional disconnection. *Brain*. 2020;143(7):2173–2188.
32. Lopes R, Bournonville C, Kuchcinski G, et al. Prediction of long-term cognitive function after minor stroke using functional connectivity. *Neurology*. 2021;96(8):e1167–e1179.
33. Boes AD. Lesion network mapping: where do we go from here? *Brain*. 2021;144(1):e5.
34. Albazron FM, Bruss J, Jones RM, et al. Pediatric postoperative cerebellar cognitive affective syndrome follows outflow pathway lesions. *Neurology*. 2019;93(16):e1561–e1571.
35. Frank RJ, Damasio H, Grabowski TJ. Brainvox: an interactive, multimodal visualization and analysis system for neuroanatomical imaging. *NeuroImage*. 1997;5:13–30.
36. Tranel D. Theories of clinical neuropsychology and brain behavior relationships: Luria and beyond. In: Morgan JE, Ricker JH, eds. *Textbook of clinical neuropsychology*. Vol. 1027. Studies on Neuropsychology, Neurology and Cognition. Psychology Press; 2008:25–37.
37. Schmidt M. *Rey auditory verbal learning test: a handbook*. Western Psychological Services; 1996.
38. Harvey RL, Edwards D, Dunning K, et al. Randomized sham-controlled trial of navigated repetitive transcranial magnetic stimulation for motor recovery in stroke. *Stroke*. 2018;49(9):2138–2146.
39. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Stegling S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13–31.
40. Corbetta M, Ramsey L, Callejas A, et al. Common behavioral clusters and subcortical anatomy in stroke. *Neuron*. 2015;85(5):927–941.
41. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing; 2020.
42. Gajardo-Vidal A, Lorca-Puls DL, Crinion JT, et al. How distributed processing produces false negatives in voxel-based lesion-deficit analyses. *Neuropsychologia*. 2018;115:124–133.
43. Mah Y-H, Husain M, Rees G, Nachev P. Human brain lesion-deficit inference remapped. *Brain*. 2014;137(Pt 9):2522–2531.
44. Inoue K, Madhyastha T, Rudrauf D, Mehta S, Grabowski T. What affects detectability of lesion-deficit relationships in lesion studies? *Neuroimage Clin*. 2014;6:388–397.
45. Hindman J, Bowren MD, Bruss J, Wright B, Geerling JC, Boes AD. Thalamic strokes that severely impair arousal extend into the brainstem. *Ann Neurol*. 2018;84(6):926–930.
46. Salvalaggio A, Pini L, De Grazia M DF, De Schotten M T, Zorzi M, Corbetta M. Reply: Lesion network mapping: where do we go from here? *Brain*. 2021;144(1):e6.
47. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016;536(7615):171–178.
48. Horn A, Neumann W-J, Degen K, Schneider G-H, Kühn AA. Toward an electrophysiological 'sweet spot' for deep brain stimulation in the subthalamic nucleus. *Hum Brain Mapp*. 2017;38(7):3377–3390.
49. Holmes AJ, Hollinshead MO, O'Keefe TM, et al. Brain genomics superstruct project initial data release with structural, functional, and behavioral measures. *Sci Data*. 2015;2:150031.
50. Cohen AL, Fox MD. Reply: the influence of sample size and arbitrary statistical thresholds in lesion-network mapping. *Brain*. 2020;143(5):e41.
51. Sperber C, Dadashi A. The influence of sample size and arbitrary statistical thresholds in lesion-network mapping. *Brain*. 2020;143(5):e40.
52. Revelle WR. *psych: Procedures for personality and psychological research*. Published online 2017. <https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research>
53. Rosseel Y. lavaan: An R Package for structural equation modeling. *J Stat Softw*. 2012;48(2):1–36.
54. Duncan TE, Duncan SC. The ABC's of LGM: an introductory guide to latent variable growth curve modeling. *Soc Personal Psychol Compass*. 2009;3(6):979–991.
55. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):162.

56. Choi EY, Yeo BTT, Buckner RL. The organization of the human striatum estimated by intrinsic functional connectivity. *J Neurophysiol.* 2012;108(8):2242–2263.
57. Duncan PW, Goldstein LB, Matchar D, Divine GW, Feussner J. Measurement of motor recovery after stroke. Outcome assessment and sample size requirements. *Stroke.* 1992;23(8):1084–1089.
58. Ramsey LE, Siegel JS, Lang CE, Strube M, Shulman GL, Corbetta M. Behavioural clusters and predictors of performance during recovery from stroke. *Nat Hum Behav.* 2017;1:0038.
59. Lindenberg R, Zhu LL, Rüber T, Schlaug G. Predicting functional motor potential in chronic stroke patients using diffusion tensor imaging. *Hum Brain Mapp.* 2012;33(5):1040–1051.
60. Feng W, Wang J, Chhatbar PY, et al. Corticospinal tract lesion load: an imaging biomarker for stroke motor outcomes. *Ann Neurol.* 2015;78(6):860–870.
61. Matsumoto R, Nair DR, LaPresto E, et al. Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain.* 2004;127(Pt 10):2316–2330.
62. Catani M, Jones DK, Ffytche DH. Perisylvian language networks of the human brain. *Ann Neurol.* 2005;57(1):8–16.
63. Stockert A, Wawrzyniak M, Klingbeil J, et al. Dynamics of language reorganization after left temporo-parietal and frontal stroke. *Brain.* 2020;143(3):844–861.
64. Tomasi D, Volkow ND. Resting functional connectivity of language networks: characterization and reproducibility. *Mol Psychiatry.* 2012;17(8):841–854.
65. Hope TMH, Seghier ML, Prejawa S, Leff AP, Price CJ. Distinguishing the effect of lesion load from tract disconnection in the arcuate and uncinate fasciculi. *Neuroimage.* 2016;125:1169–1173.
66. Reber J, Hwang K, Bowren M, et al. Cognitive impairment after focal brain lesions is better predicted by damage to structural than functional network hubs. *Proc Natl Acad Sci USA.* 2021;118(19):e2018784118.
67. Hope TMH, Friston K, Price CJ, Leff AP, Rotshtein P, Bowman H. Recovery after stroke: not so proportional after all? *Brain.* 2019;142(1):15–22.
68. Kundert R, Goldsmith J, Veerbeek JM, Krakauer JW, Luft AR. What the proportional recovery rule is (and is not): methodological and statistical considerations. *Neurorehabil Neural Repair.* 2019;33(11):876–887.
69. Bonkhoff AK, Hope T, Bzdok D, et al. Bringing proportional recovery into proportion: bayesian modelling of post-stroke motor impairment. *Brain.* 2020;143(7):2189–2206.
70. Gajardo-Vidal A, Lorca-Puls DL, Team P, et al. Damage to Broca's area does not contribute to long-term speech production outcome after stroke. *Brain.* 2021;144(3):817–832.
71. Griffis JC, Nenert R, Allendorfer JB, Szaflarski JP. Damage to white matter bottlenecks contributes to language impairments after left hemispheric stroke. *Neuroimage Clin.* 2017;14:552–565.
72. Woolgar A, Parr A, Cusack R, et al. Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proc Natl Acad Sci USA.* 2010;107(33):14899–14902.
73. Mesulam M-M, Wieneke C, Hurley R, et al. Words and objects at the tip of the left temporal lobe in primary progressive aphasia. *Brain.* 2013;136(2):601–618.
74. Abel TJ, Rhone AE, Nourski KV, et al. Direct physiologic evidence of a heteromodal convergence region for proper naming in human left anterior temporal lobe. *J Neurosci.* 2015;35(4):1513–1520.
75. Mehta S, Inoue K, Rudrauf D, Damasio H, Tranel D, Grabowski T. Segregation of anterior temporal regions critical for retrieving names of unique and non-unique entities reflects underlying long-range connectivity. *Cortex.* 2016;75:1–19.
76. Fridriksson J, den Ouden D-B, Hillis AE, et al. Anatomy of aphasia revisited. *Brain.* 2018;141(3):848–862.
77. Assem M, Blank IA, Mineroff Z, Ademoğlu A, Fedorenko E. Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex.* 2020;131:1–16.
78. Griffis JC, Nenert R, Allendorfer JB, Szaflarski JP. Linking left hemispheric tissue preservation to fMRI language task activation in chronic stroke patients. *Cortex.* 2017;96:1–18.
79. Diachek E, Blank I, Siegelman M, Affourtit J, Fedorenko E. The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. *J Neurosci.* 2020;40(23):4536–4550.
80. Siegel JS, Ramsey LE, Snyder AZ, et al. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proc Natl Acad Sci USA.* 2016;113(30):E4367–E4376.
81. Foulon C, Cerliani L, Kinkingnéhun S, et al. Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. *Gigascience.* 2018;7(3):1–17.
82. Griffis JC, Metcalf NV, Corbetta M, Shulman GL. Lesion Quantification Toolkit: A MATLAB software tool for estimating grey matter damage and white matter disconnections in patients with focal brain lesions. *Neuroimage Clin.* 2021;30:102639.