NeuroImage: Clinical 19 (2018) 22–29

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Predicting language outcomes after stroke: Is structural disconnection a useful predictor?

Thomas M.H. Hope[a,*], Alex P. Leff[b,c], Cathy J. Price[a]

[a] Wellcome Centre for Human Neuroimaging, University College London, UK
[b] Institute of Cognitive Neuroscience, University College London, UK
[c] Department of Brain, Repair and Rehabilitation, Institute of Neurology, University College London, UK

ARTICLE INFO

Keywords:
Stroke
Language
Aphasia
Outcomes
MRI
Connectomics
White matter

ABSTRACT

For many years, researchers have sought to understand whether and when stroke survivors with acquired language impairment (aphasia) will recover. There is broad agreement that lesion location information should play some role in these predictions, but still no consensus on the best or right way to encode that information. Here, we address the emerging emphasis on the structural connectome in this work – specifically the claim that disrupted white matter connectivity conveys important, unique prognostic information for stroke survivors with aphasia.

Our sample included 818 stroke patients extracted from the PLORAS database, which associates structural MRI from stroke patients with language assessment scores from the Comprehensive Aphasia Test (CAT) and basic demographic. Patients were excluded when their lesions were too diffuse or small (< 1 cm3) to be detected by the Automatic Lesion Identification toolbox, which we used to encode patients' lesions as binary lesion images in standard space. Lesions were encoded using the 116 regions defined by the Automatic Anatomical Labelling atlas. We examined prognostic models driven by both "lesion load" in these regions (i.e. the proportion of each region destroyed by each patient's lesion), and by the disconnection of the white matter connections between them which was calculated via the Network Modification toolbox. Using these data, we build a series of prognostic models to predict first one ("naming"), and then all of the language scores defined by the CAT.

We found no consistent evidence that connectivity disruption data in these models improved our ability to predict any language score. This may be because the connectivity disruption variables are strongly correlated with the lesion load variables: correlations which we measure both between pairs of variables in their original form, and between principal components of both datasets. Our conclusion is that, while both types of structural brain data do convey useful, prognostic information in this domain, they also appear to convey largely the same variance. We conclude that connectivity disruption variables do not help us to predict patients' language skills more accurately than lesion location (load) data alone.

1. Introduction

For many years, researchers have tried to understand and predict whether and when stroke survivors will recover lost speech and language abilities (Bang et al., 2005; Cloutman et al., 2009; Crinion and Price, 2005; Hope et al., 2017; Hope et al., 2013; Konig et al., 2008; Lazar et al., 2008; Lendrem and Lincoln, 1985; Marshall and Phillips, 1983; Payabvash et al., 2010; Pedersen et al., 1995; Seghier et al., 2016; Tilling et al., 2001; Ween et al., 2000). There is broad agreement that lesion location information should play some role in this work (Plowman et al., 2012), but still no consensus on the best or right way to encode that information (Forkel et al., 2014; Hope et al., 2013; Mah et al., 2014; Price et al., 2017; Zhang et al., 2014). An emerging emphasis on structural (i.e. white matter) connectivity in studies of language has naturally encouraged the same attention in studies of aphasia (Agosta et al., 2010; Epelbaum et al., 2008; Fridriksson et al., 2013; Hope et al., 2016; Olsen et al., 2015; Ripamonti et al., 2014). As many studies have shown that disrupted connectivity contributes to language impairments and their recovery (Forkel et al., 2014; Hope et al., 2016; Kuceyeski et al., 2015a; Pani et al., 2016; Wu et al., 2015; Yourganov et al., 2016), it is natural to presume that connectivity disruption data should be pivotal when predicting language outcomes after stroke.

However, lesion distributions are highly structured (Inoue et al., 2014; Mah et al., 2014). If one brain region is damaged, neighbouring

* Corresponding author at: Wellcome Centre for Human Neuroimaging, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK.
E-mail address: t.hope@ucl.ac.uk (T.M.H. Hope).

https://doi.org/10.1016/j.nicl.2018.03.037
Received 2 March 2018; Received in revised form 22 March 2018; Accepted 28 March 2018
Available online 30 March 2018

regions are often damaged too, and white matter disruption will tend to be highly correlated with cortical damage. So even if connectivity disruption is the causal mechanism for some post-stroke cognitive symptoms, it may be that lesion location can serve as a reliable proxy in prognostic models. We might find that the addition of connectivity disruption data adds little, unique prognostic value to our models of post-stroke outcomes. Or to put the point another way, mechanistic importance is no guarantee of clinical importance, in this domain. In what follows, we test the clinical importance of connectivity disruption data in a very large sample stroke patients.

2. Methods

2.1. Patient data

Our patient data were extracted from our *PLORAS* database (Seghier et al., 2016), which associates stroke patients, tested over a broad range of times post stroke, with demographic data, behavioural test scores from the Comprehensive Aphasia Test (Swinburn et al., 2004), and high resolution T1-weighted MRI brain scans. Patients are excluded from the PLORAS database when there is evidence they have other neurological conditions (e.g. dementia, multiple sclerosis), contraindications to MRI scanning, are unable to see or hear the stimuli required to assess their language abilities, or have insufficient comprehension of the purpose of the study to provide consent for their participation. We included all patients whose data was available at the time, irrespective of their: age at stroke onset; sex; premorbid handedness; or native language. Patients were only excluded if their lesions were too diffuse or small ($< 1cm^3$) to be detected by our Automatic Lesion Identification (ALI) toolbox (Seghier et al., 2008).

2.2. Structural brain imaging data

Imaging data were collected using sequences described elsewhere (Hope et al., 2015). Data from different scanners were combined after conversion to quantitative probabilistic estimates of grey matter density. Pre-processed with Statistical Parametric Mapping software (SPM, 2012), these images were spatially normalised into Montreal Neurological Institute (MNI) space using a modified version of the unified segmentation algorithm (Ashburner and Friston, 2005) that has been optimized for use in patients with focal brain lesions (Seghier et al., 2008). We used the ALI toolbox (Seghier et al., 2008) to index the degree of abnormality at each voxel in each patient image (in relation to the same type of images in healthy controls), combining the grey and white matter outputs to generate a single thresholded (i.e. binary) image that shows the presence or absence of a lesion at each voxel. Lesion volume is calculated as the sum of those voxels where lesions were deemed to be present.

Following the approach taken by Yourganov and colleagues (Yourganov et al., 2016), in a recent study which demonstrates that connectivity disruption data can drive useful predictions for language outcomes after stroke, we encoded our lesion images using the 116 grey-matter regions defined by the Automatic Anatomical Labelling atlas (Tzourio-Mazoyer et al., 2002a). We examined models driven by both lesion load in these regions (i.e. the proportion of each region destroyed by each patient's lesion), and by the disconnection of the white matter connections between them. Disconnection was calculated via the Network Modification toolbox (Kuceyeski et al., 2013), which generates the mean disconnection implied by each lesion, using structural connectomes defined for a separate sample of 73 neurologically normal controls. This toolbox has been used to successfully predict both network atrophy (Kuceyeski et al., 2014) and cognitive outcomes (Kuceyeski et al., 2016) after stroke, and has also been successfully employed in studies of longitudinal patterns of atrophy in Alzheimer's patients (Raj et al., 2015), the spread of Progressive Supranuclear Palsy (Pandya et al., 2017), cortical atrophy in temporal lobe epilepsy (Abdelnour et al., 2015), and early Multiple Sclerosis (Kuceyeski et al., 2015b).

2.3. Behavioural data

Every patient was assessed using the Comprehensive Aphasia Test (CAT) (Swinburn et al., 2004). For ease of comparison across tasks, task scores are expressed as T-scores, representing each patient's assessed skill on each task (e.g., describing a picture; reading non-words) relative to a reference population of 113 aphasic patients. The threshold for impairment is defined relative to a separate population of 27 neurologically normal controls such that performance below threshold would place the patient in the bottom 5% of the normal population (Swinburn et al., 2004). Lower scores indicate poorer performance. The CAT yields 34 separate scores, though six refer to non-linguistic skills such as line bisection, arithmetic and memory. Here, we focus initially on scores in naming (i.e. of visually presented pictures), before widening the analysis to include all of the other 27 language scores. Detailed descriptions of the tasks are given in the CAT manual (Swinburn et al., 2004).

2.4. The baseline model

Our aim here was to measure what the introduction of structural (dis)connection variables buys us, in terms of improved predictive accuracy. Our baseline for this comparison, is a model driven by variables whose prognostic relevance is already supported by prior evidence: (i) basic demographic data including time post-stroke (Hope et al., 2017; Hope et al., 2013), age at stroke (Ramsey et al., 2017), pre-stroke handedness (Knecht et al., 2000), and bilingualism (Hope et al., 2015); (ii) lesion volume (Plowman et al., 2012); and (iii) lesion location (Hope et al., 2013; Plowman et al., 2012; Yourganov et al., 2016), which is calculated as described above. We use the term ‘lesion load variables’ to refer to variables representing the proportion of each of a series of anatomically defined regions, which is destroyed or encroached upon by each patient's lesion(s). We use the term ‘lesion load model’ to refer to models driven by the combination of: (a) demographic and lesion volume variables, as described above; and (b) lesion load variables.

2.5. Structural connectivity models

To measure whether structural connectivity variables add prognostic information over and above that already conveyed by lesion-load models, we compare the predictions made by lesion-load models to those made using models which either: (a) replace the lesion load variables with structural connectivity variables, or (b) add structural connectivity variables to the lesion-load model, or (c) stack lesion-load and connectivity models together. Like the lesion load model, all of these models also include basic demographic data and lesion volume. For the sake of brevity, we refer to them as: “connectivity models”, ‘lesion load plus connectivity models’, and “stacked models” in what follows.

Stacking starts by training component models separately (e.g. a lesion-load model and a structural connectivity model), and using those models to predict the language scores under study via cross-validation. The resulting predictions are then used as input to a new model, also trained to predict the same language scores. This new, higher level model is also assessed in cross-validation, using the same folds as employed to generate the predictions from the component models. Our use of this approach is motivated by recent work which employs stacking to apparently good effect in this domain (Pustina et al., 2017), reporting modest but significant improvements in predictive power over what was possible with any component model alone. More generally, stacking is thought to be useful when – as here – we want to combine inferences made from datasets containing very unequal numbers of