Anatomical predictors of aphasia recovery: a tractography study of bilateral perisylvian language networks

Stephanie J. Forkel, Michel Thiebaut de Schotten, Flavio Dell’Acqua, Lalit Kalra, Declan G. M. Murphy, Steven C. R. Williams and Marco Catani

1 Research Department of Clinical, Educational, and Health Psychology (RDCEHP), Division of Psychology and Language Sciences, Faculty of Brain Sciences, University College London, UK
2 Natbrainlab, Department of Neuroimaging, Institute of Psychiatry, King’s College London, UK
3 Natbrainlab, Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King’s College London, London, UK
4 Inserm U1127; UPMC-Paris6, UMR_S 1127; CNRS UMR 7225, Centre de Recherche de l’Institut du Cerveau et de la Moelle épinière, Groupe Hospitalier Pitié-Salpêtrière, 75013 Paris, France
5 NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King’s College London, London, UK
6 Department of Clinical Neuroscience, Institute of Psychiatry, King’s College London, London, UK
7 Sackler Institute for Translational Neurodevelopment and Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King’s College London, UK

Correspondence to: Stephanie Forkel, Natbrainlab, PO89, Institute of Psychiatry, De Crespigny Park, SE5 8AF, London, UK
E-mail: s.forkel@ucl.ac.uk

Correspondence may also be addressed to Marco Catani, Natbrainlab, Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King’s College London, UK, E-mail: m.catani@iop.kcl.ac.uk

Stroke-induced aphasia is associated with adverse effects on quality of life and the ability to return to work. For patients and clinicians the possibility of relying on valid predictors of recovery is an important asset in the clinical management of stroke-related impairment. Age, level of education, type and severity of initial symptoms are established predictors of recovery. However, anatomical predictors are still poorly understood. In this prospective longitudinal study, we intended to assess anatomical predictors of recovery derived from diffusion tractography of the perisylvian language networks. Our study focused on the arcuate fasciculus, a language pathway composed of three segments connecting Wernicke’s to Broca’s region (i.e. long segment), Wernicke’s to Geschwind’s region (i.e. posterior segment) and Broca’s to Geschwind’s region (i.e. anterior segment). In our study we were particularly interested in understanding how lateralization of the arcuate fasciculus impacts on severity of symptoms and their recovery. Sixteen patients (10 males; mean age 60 ± 17 years, range 28–87 years) underwent post stroke language assessment with the Revised Western Aphasia Battery and neuroimaging scanning within a fortnight from symptoms onset. Language assessment was repeated at 6 months. Backward elimination analysis identified a subset of predictor variables (age, sex, lesion size) to be introduced to further regression analyses. A hierarchical regression was conducted with the longitudinal aphasia severity as the dependent variable. The first model included the subset of variables as previously defined. The second model additionally introduced the volume of the right long segment to the first-level model increased the overall predictive power of the model from 28% to 57% [F(1,11) = 7.46, P = 0.02]. These findings suggest that...
Introduction

Every year 15 million people worldwide suffer a stroke, among which nearly 6 million patients die and 5 million survivors are left permanently disabled (Mackay and Mensah, 2014). About one-third of all stroke patients are affected by language problems, collectively referred to as aphasia (Pedersen et al., 2004). Aphasia has an adverse effect on functional outcome, mood, quality of life and the ability to return to work (Ferro and Madureira, 1997).

Clinically, established predictors of language recovery include age, lesion characteristics, education, and possibly sex (Eslinger and Damasio, 1981; McDermott et al., 1996; Ferro and Madureira, 1997; Laska et al., 2001; Pedersen et al., 2004). However, taken together these factors only explain ~40% of the variance (Pedersen et al., 2004); hence other factors may contribute to recovery.

The use of structural and functional neuroimaging applied to stroke patients has helped to better understand anatomical and metabolic factors associated with aphasia recovery. These can be distinguished into ipsilateral and contralateral factors (Hillis, 2005, 2006). Established ipsilateral factors include lesion location at the cortical level (Ferro et al., 1999), the extension of the damage to subcortical structures (Naeser et al., 1982; Vallar et al., 1988; Ferro et al., 1999), the extent of the area of ischaemic penumbra and the degree of reperfusion (Croquelois et al., 2003; Hillis et al., 2001, 2006). Hillis et al. (2001) for example, reported that in patients with aphasia, improved perfusion of Wernicke’s area within 3 days post onset is associated with better language outcomes. This study highlighted the importance of mapping those areas that show significant diffusion-perfusion mismatch (i.e. perfusion delay but normal diffusion-weighted imaging signal) as these areas may exhibit functional improvement if blood flow can be restored. Likewise, the magnitude of perfusion delay provides a proxy of the degree of functional impairment. Moreover, the authors suggest that dysfunction of other language-related areas could be due to a diascis mechanism affecting areas not directly damaged by the stroke (Hillis et al., 2001).

Recovery may also depend on contralateral factors. Previous functional imaging and case report studies suggested that language recovery may involve the recruitment of right hemispheric homologues of language areas (Weiller et al., 1995; Musso et al., 1999; Rosen et al., 2000; Leff et al., 2002; Sharp et al., 2004; Crinion and Price, 2005). Early studies indicated a compensatory role of the right hemisphere in patients who recovered their language after a left-hemispheric stroke and later became aphasic again, following a right-hemispheric stroke (Nielsen, 1946) or as a consequence of temporary right-hemispheric anaesthesia (i.e. Wada test) (Kinsbourne, 1971; Wada, 1975; Ohyama et al., 1996; Cappa et al., 1997; Thulborn et al., 1999). A more recent functional MRI study investigated the evolution of language recovery after stroke and proposed a three-stage model whereby: (i) the acute stage is characterized by reduced activation in intact perilesional left hemispheric language areas; (ii) during the subacute stage a bilateral upregulation occurs and the peak activation in right homologues areas significantly correlates with language improvement; and (iii) a re-shift of peak activation to the left hemisphere in the chronic stage (Saur et al., 2006). The authors suggested that a shift to the right facilitates recovery only if transient and followed by a return of the peak activation to the left hemisphere. Indeed, sustained right hemisphere activation at 1 year has been associated with poor language performances, which could reflect a maladaptive compensatory mechanism in chronic stroke patients (Szaflarski et al., 2013).

Overall, these and other studies suggest that recovery of language after stroke is a dynamic process in which the right hemisphere is important for longitudinal outcomes (Ohyma et al., 1996; Cappa et al., 1997; Thulborn et al., 1999; Thiel et al., 2001, 2006; Leff et al., 2002; Saur et al., 2006).

The individual variability observed in functional imaging studies may also be related to underlying anatomical patterns of lateralization of language networks (Knecht et al., 2002; Baynes and Long, 2007). This structural variability has been studied both at the cortical and subcortical level. The planum temporale, part of Wernicke’s region, is an area of the superior temporal gyrus located behind the primary auditory cortex and includes associative cortex involved in auditory comprehension. The planum temporale is consistently reported to be larger in the left compared to the right hemisphere (Geschwind and Levitsky, 1968; Steinmetz, 1996; Shapleske et al., 1999; for a review see Chance and Crow, 2007). Geschwind and Levitsky (1968) hypothesized that this asymmetry of the planum temporale is associated with functional lateralization. Initial studies that were supportive of an association between functional lateralization and asymmetry of the planum temporale have not been replicated in a recent study where Dorsaint-Pierre et al. (2006) used MRI scans and sodium amytal procedure. Similarly, studies using functional MRI for single word comprehension and dichotic listening tasks found no correspondence between left lateralization of language and leftward asymmetry of the planum temporale (Eckert et al., 2006; Sequeira et al., 2006). The lack of structure-function correlation may explain the absence of evidence for the influence of planum temporale asymmetries in aphasia recovery after stroke.

Other structural asymmetries underlying functional dominance have been proposed. Leftwards asymmetry has been reported for Broca’s region using cytoarchitectonic analysis (Brodman area 44 in the pars opercularis) (Amunts et al., 1999). Significant leftward asymmetry in different predictors of recovery are at play in the left and right hemisphere. The right hemisphere language network seems to be important in aphasia recovery after left hemispheric stroke.

Keywords: diffusion tensor imaging tractography; language recovery; aphasia; stroke; arcuate network

Abbreviations: AICc = corrected Akaike Information Criterion; WAB-R = Western Aphasia Battery Revised
the volume of the pars opercularis has also been reported based on in vivo MRI-based measurements (Keller et al., 2007). Grey matter concentration differences in the pars opercularis have been found to correlate with language dominance assessed by the sodium amytal procedure (Dorsaint-Pierre et al., 2006); however, there is no evidence that this asymmetry is related to recovery of language after stroke.

Recent methodological advances in diffusion imaging tractography permitted to study the in vivo asymmetries of language networks, including the arcuate fasciculus connecting temporal, parietal, and frontal language regions (Catani et al., 2005; Powell et al., 2006; Glasser and Rilling, 2008). This tract is composed of a direct long segment between classical Broca’s and Wernicke’s regions and an indirect pathway through the inferior parietal lobe (i.e. Geschwind’s region) that includes the anterior fronto-parietal and the posterior temporal-parietal segments (Catani et al., 2005). The pattern of lateralization for the individual segments of the arcuate fasciculus shows a significant inter-individual variability among the healthy population (Catani et al., 2007; Thiebaut de Schotten et al., 2011b). This is particularly evident for the direct long segment, which is bilateral in 40% of the healthy population and extremely left lateralized in the remaining 60%, where the segment is either absent, or very small in the right hemisphere (Catani et al., 2007). There is some preliminary evidence that this individual variability in the pattern of lateralization of the arcuate fasciculus could be related to differences in functional dominance and behavioral performances. A study combining the Wada test and diffusion tensor imaging tractography reported a good concordance between language dominance and anatomical lateralization of perisylvian networks (Matsumoto et al., 2008). In particular 95% of individuals with an anatomical left lateralization of the long segment of the arcuate fasciculus were also functionally left dominant on the Wada test. Conversely, two-thirds of the subjects with functional right dominance on the Wada test presented with a right lateralization of the long segment of the arcuate fasciculus. Furthermore, a more bilateral pattern of lateralization of the long segment has been found to correlate with better performances on auditory verbal learning tasks in healthy subjects (Catani et al., 2007). In conclusion, structural and functional studies suggest a possible role of the right hemisphere in language recovery.

In this study, we investigated prospectively whether tractography-based measurements of different segments of the arcuate fasciculus could help to predict language recovery at 6 months in left hemispheric stroke patients.

Materials and methods

Patients

Patients with left hemisphere stroke and language impairment were consecutively recruited from the hyperacute stroke unit at King’s College Hospital, London between 2009 and 2012. Patients were initially assessed within 3 days of admission using the revised Western Aphasia Battery (WAB-R) bedside screening (Kertesz, 2007). Inclusion criteria were: (i) right-handedness based on Edinburgh Handedness Inventory (Oldfield, 1971); (ii) first ever left middle cerebral artery infarct; (iii) presence of aphasia; (iv) no previous neurological or psychiatric diagnoses; (v) medically stable to tolerate ambulance transport; (vi) no MRI contraindications; and (vii) native English speaker. After the initial screening, a total of 18 eligible patients underwent comprehensive WAB-R assessment within 10 days of aphasia onset (mean 5 ± 5 days) and research MRI within 2 weeks, except for two patients who were scanned at day 20 and 24 after admission (overall mean 10 ± 6 days). WAB-R provides a global measure of aphasia severity, the aphasia quotient (AQ), which ranges from 0 to 100. A score ≥ 93.8 is indicative of normal language function (Swindell et al., 1984; Pedersen et al., 2004).

Six months after admission patients were re-invited for a repeated WAB-R assessment (mean 200 ± 55 days). Sixteen patients completed the follow-up assessment (10 males; mean age 60 ± 17 years; age range 28–87 years). Of these 16 patients, 11 had received thrombolysis as a clinical intervention on admission to hospital. Unfortunately, information regarding the nature or intensity of community-based speech and language therapy amongst the patients who received such intervention was not available at the completion of the study. This limited the possibility of analysing the impact of therapy on recovery in our cohort.

For this study, all patients or their next of kin gave written informed consent. The study was approved by the Wandsworth Ethical Research Committee (09/H0803/95) and the local review board (KCH1700).

Neuroradiological acquisition and processing

MRI data were acquired using a 3 T HDx GE scanner (General Electric) equipped with an 8-channel radio frequency receiver head coil. For each subject a high-resolution structural T1-weighted volume (1 × 1 × 1 mm) of the whole brain was acquired. Diffusion tensor imaging data were acquired using a spin echo, single shot EPI pulse sequence optimized for subjects with high risk of movement during the scan. This consisted of two different scans of 30 diffusion-weighted directions (b-value 1500 mm²/s) combined together for a total of 60 directions and seven non-diffusion weighted volumes. Matrix size was 128 × 128 × 60 and voxel size was 2.4 × 2.4 × 2.4 mm. Peripheral gating was applied to avoid brain pulsation artefacts.

Diffusion tensor imaging data were preprocessed using ExploreDTI (www.exploreti.org) and corrected for eddy current and motion artefacts through iterative correction to the seven non-diffusion weighted volumes. For each subject data quality was visually inspected. In compliance with the study protocol, participants who generated corrupted images on more than two diffusion-weighted imaging volumes would have been excluded. No participant had to be excluded after inspection. RESTORE was also run to exclude outliers (Chang et al., 2005). The diffusion tensor was estimated using a non-linear least squares approach (Jones and Basser, 2004).

Whole brain tractography was performed from all brain voxels with fractional anisotropy > 0.2. Streamlines were propagated with a step-size of 1 mm, using Euler integration and b-spline interpolation of the diffusion tensor field (Basser et al., 2000). Where fractional anisotropy was < 0.2 or when the angle between two consecutive tractography steps > 45°, streamline propagation was stopped.

Tractography dissections were obtained using a three regions of interest approach as previously described (Catani et al., 2005). Regions of interest for both hemispheres were defined on fractional anisotropy images in the patients’ native space (Fig. 1A). The frontal region of interest was defined anterior to the central sulcus to encompass the white matter of the posterior region of the inferior and middle frontal gyri and the inferior part of the precentral gyrus. The temporal
region of interest was defined in the white matter of the posterior part of the superior and middle temporal gyri. The parietal region of interest included the white matter of the supramarginal and angular gyri. All streamlines passing through both frontal and temporal regions of interest were considered as belonging to the long segment of the arcuate fasciculus. All streamlines between temporal and parietal regions of interest were classified as posterior segment of the arcuate fasciculus connecting Wernicke’s and Geschwind’s region. Finally, streamlines between parietal and frontal regions of interest were labelled as anterior segment of the arcuate fasciculus connecting Broca’s and Geschwind’s region (Fig. 1B).

Two anatomists used TrackVis (www.trackvis.org) for virtual dissections and volume measurements of the three segments of the arcuate fasciculus in all participants and showed a high inter-rater reliability ($r = 0.635; P < 0.001$). The second rater was blinded to any information on symptoms. The volume was calculated as the number of voxels intersected by the streamlines of each segment. This measure indicates the space occupied by the reconstructed streamlines but its relationship to the underlying anatomy of the fibres (e.g. axonal diameter, myelination and density) is not fully established (Beaulieu, 2002).

To control for the possibility that hemisphere size might be driving the volume of the arcuate segments (i.e. larger hemisphere means larger arcuate fasciculus) the tract volume was normalized by the hemisphere volume (segment volume/hemisphere volume). The hemispheric volume was obtained using FMRIB Software Library package (FSL, http://www.fmrib.ox.ac.uk/fsl/). The normalized segment volume was then used for further analysis.

Ischemic lesions were manually delineated on $T_1$-weighted images and the volume of interest (number of voxels) was extracted using FSL and subsequently normalized to MNI space. Lesion overlay percentage maps were obtained by binarizing (i.e. assigning a value of 0 or 1 to each voxel) the normalized lesion volumes and calculating percentage overlaying voxels.

**Statistical analysis**

In this study, our primary goal was to identify anatomical predictors of aphasia recovery. Hence, we limited our model to factors that could directly influence the anatomy of the three segments of the arcuate fasciculus. Previous studies have shown that amongst these factors are age (i.e. smaller arcuate in older patients; Bava et al., 2011; Lebel and Beaulieu, 2011; Lebel et al., 2012), sex (i.e. larger right long segment in females; Catani et al., 2007; Hsu et al., 2008; Inano et al., 2011; Kanaan et al., 2012), lesion size (i.e. smaller arcuate in larger lesions;
Goldberg and Ransom, 2003; Johansen-Berg and Behrens, 2006; Ciccarelli et al., 2008), and level of education (Thiebaut de Schotten et al., 2014). Other studies have also shown baseline aphasia severity to be predictive of longitudinal outcome (Kertesz, 1988b; Ferro et al., 1999; Laska et al., 2001; Pedersen et al., 2004; Lazar et al., 2008), which we were not able to include in our analysis. Previous studies show that baseline measurement fluctuations occur within a fortnight of symptom onset due to clinical-physiological processes and the influence of psychodynamic mechanisms (Hillis et al., 2001; Hackett et al., 2005; Lazar et al., 2008; Gottesman and Hills, 2010). For this reason, baseline language measures in relation to longitudinal therapeutic goals are usually obtained 2 weeks after symptom onset to allow sufficient time for acute processes to settle. In the current study, baseline measures were obtained on admission (mean 5 ± 5 days) and therefore are not reliable indices of language impairment. Indeed, in our sample, severity of aphasia at baseline is not correlated with the language score obtained 6 months after the stroke ($r(16) = 0.49, P > 0.05$).

To determine in our data the most relevant variables among those chosen for the analysis (i.e. age, sex, lesion size, level of education) and to identify those explaining the dependent variable (i.e. longitudinal aphasia severity measured by the AQ) all variables were introduced to a backward elimination analysis. This method first places all possible variables, as identified from the literature or driven by a specific hypothesis, in the model and calculates the contribution of each of them. Their contribution is then compared against a removal criterion (here we used a probability value for the test statistic of $P > 0.001$). The variable with the least contribution to the model is subsequently removed and the reduced model re-estimated for the remaining variables. The contribution of the remaining variables is reassessed in an iterative way until the model reaches statistical significance.

The resulting subset of variables was introduced into subsequent regression analyses. The primary analysis employed a hierarchical linear regression. In this analysis, two models were defined with the longitudinal aphasia severity (AQ) at 6 months as the dependent variable. The first-level model included the variables identified by the backward elimination, namely age, gender and lesion size. In the second-level of the model the normalised volumes of the three segments of the arcuate fasciculus were separately added to the model.

Where both models were significant, the fit of each model was estimated by calculating the corrected Akaike Information Criterion (AICc) for small sample sizes (Akaike, 1974; Hurvich and Tsai, 1989). The AICc is a goodness of fit measure corrected for model complexity (i.e. penalizing increasing number of predictors). We used this analysis to compare both levels of the regression models and to verify that the increase in predictive power of the second-level model is not merely driven by a higher number of predictors.

The secondary analysis addressed group differences between sex, fluent versus non-fluent aphasia types (defined according to a cutoff of 4 on the WAB-R fluency scale), and thrombolysis groups (thrombolysed versus non thrombolysed).

Statistical analyses were performed using R 2.15.1 software (www.R-project.org). Power analysis was conducted using the software package G*Power (http://wwwpsycho.uni-duesseldorf.de/abteilungen/aap/gpower3/).

### Results

Patient details and demographics are available in Table 1. A total of 18 patients were included at baseline, whereof 16 were followed-up 6 months post stroke. Eleven of these 16 patients had...
received thrombolysis treatment as part of their standard clinical intervention.

Identification of previously established predictors of symptoms severity

Based on previous studies we considered age, sex, lesion size and education as potential predictors of aphasia severity at 6 months (Eslinger and Damasio, 1981; McDermott et al., 1996; Ferro and Madureira, 1997; Laska et al., 2001). In addition these factors can also influence the anatomy of the white matter tracts (Catani et al., 2007; Lebel et al., 2008; Thiebaut de Schotten et al., 2014). To confirm whether these factors significantly influence language outcomes in our data set, a backward regression analysis was conducted. This method introduces all potential variables into the model at once before subsequently eliminating each independent variable, starting with the variable with the smallest partial correlation coefficient. The analysis shows that when all variables are included, the model is not significant for predicting longitudinal language outcomes \(R^2 = 0.605, F(4,9) = 3.448, P = 0.06\). The subsequent analysis removed education from the variables and the model becomes significant with the inclusion of age, sex, and lesion size only \(R^2 = 0.596, F(3,10) = 4.921, P = 0.024\). This result is in line with the literature reporting that education has no strong evidence of being an independent predictor of long-term recovery (Ferro et al., 1999). Based on these findings age, sex and lesion size can be considered as good predictors of recovery in our data set. This subset of variables was therefore taken into account for subsequent regression analyses.

Percentage lesion overlay maps

As a first investigation we computed percentage lesion overlay maps. The group comparison lesion analysis identified the inferior frontal gyrus, the internal and external/extreme capsules, the claustrum, the putamen, the medial thalamic nuclei, and the perisylvian white matter as areas most commonly damaged in our sample of stroke patients (Fig. 2).

Diffusion tensor imaging tractography

The volume, defined as the number of voxels intersected by the streamlines of each segment of the arcuate fasciculus, was extracted for the left and right hemispheres for each patient (Table 2).

Left hemisphere

The hierarchical regression analysis showed that a model including age, sex and lesion size was predictive of longitudinal aphasia severity \(R^2 = 0.502, F(3,11) = 3.689, P = 0.047\). The predictive value of the model improved when the volume of the left long segment was included in addition to age, sex and lesion size \(R^2 = 0.623, F(4,10) = 4.138, P = 0.031\); \(R^2\) change: \(F(1,10) = 3.235, P = 0.102\) although the change was not statistically significant. Among the four variables entered in the analysis only lesion size was an independent predictor \([\beta = -0.630, t(-3.129), P = 0.011]\) of longitudinal aphasia severity.

The same analysis was repeated for the left anterior and posterior segments of the arcuate fasciculus and both models were not predictive \([\text{anterior segment index size: } R^2 = 0.541, F(4,10) = 2.943, P = 0.076; \text{ posterior segment index size: } R^2 = 0.577, F(4,10) = 3.411, P = 0.053]\). The result indicates that by taking into account all three predictors, the left hemispheric model can explain approximately 50% of the variability in language recovery. By adding the volume of the left long segment the model can explain 62% of the variability, which represents only a 12% increase in predicting value. Furthermore this analysis indicates that in the left hemisphere the only independent predictor of longitudinal aphasia is lesion size. It should be taken into consideration, however, that the volume measurements of the left and right arcuate segments reflect two different anatomical properties of these fibres. In the right hemisphere the volume of the tracts reflects the anatomy of a pre-existing tract, which is...
Predictors of aphasia recovery

Table 2  Diffusion tensor imaging measurements (number of voxels) for each segment in the left and right hemisphere

<table>
<thead>
<tr>
<th></th>
<th>Left hemisphere</th>
<th></th>
<th>Right hemisphere</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume three segments</td>
<td></td>
<td>Volume three segments</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>LS AS PS</td>
<td></td>
<td>LS AS PS</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>533 283 529</td>
<td></td>
<td>920 360 381</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1138 475 775</td>
<td></td>
<td>666 637 692</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1191 610 894</td>
<td></td>
<td>545 633 778</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1705 0 517</td>
<td></td>
<td>742 678 545</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>356 605 310</td>
<td></td>
<td>445 388 627</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>489 347 284</td>
<td></td>
<td>830 346 269</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1234 104 327</td>
<td></td>
<td>285 765 378</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>861 851 811</td>
<td></td>
<td>432 626 750</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>736 552 1014</td>
<td></td>
<td>563 557 1015</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>488 397 796</td>
<td></td>
<td>0 720 374</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0 0 699</td>
<td></td>
<td>672 482 300</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>959 786 581</td>
<td></td>
<td>439 743 660</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>880 660 440</td>
<td></td>
<td>161 399 0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>830 0 589</td>
<td></td>
<td>385 347 339</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>780 242 669</td>
<td></td>
<td>380 643 448</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>659 232 765</td>
<td></td>
<td>0 378 112</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>532 287 623</td>
<td></td>
<td>582 841 293</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>622 496 569</td>
<td></td>
<td>206 1971 628</td>
<td></td>
</tr>
</tbody>
</table>

AS = anterior segment; ID = patient identification; LS = long segment; PS = posterior segment. Volume is defined as the number of voxels intersected by the streamlines of each segment.

unaffected by the stroke. By contrast, in the left hemisphere, tract volume measurements are indicative of the residual fibres whose quantity depends on the amount of damage occurred in the middle cerebral artery territory. Hence, while the right tract measurements reflects the anatomical volume of the pre-existing arcuate the left tract volume is an indirect measure of lesion load specifically to the arcuate fasciculus. We therefore investigated if the lesion size and the volume of the left arcuate segments correlate, which was not the case for the left long segment nor the sum of the three segments in the left hemisphere [left long segment: \( r(18) = 0.224, P = 0.372 \); sum of left three segments: \( r(18) = 0.078, P = 0.760 \)]. Also when adding the sum of the three segment as a predictor to the regression model, the model did not explain the observed data [level 1 (age, sex, lesion size): \( R^2 = 0.275, F(3,12) = 1.52, P = 0.260 \); level 2 (addition of the left three segments): \( R^2 = 0.305, F(4,11) = 1.206, P = 0.362 \); \( R^2 \) change: \( R^2 = 0.030, F(1,11) = 0.467, P = 0.508 \).  

Right hemisphere

The hierarchical regression analysis showed that a model including age, sex and lesion size was not predictive of longitudinal aphasia severity [\( R^2 = 0.275, F(3,12) = 1.52, P = 0.260 \)]. The predictive value of the model improved significantly when the volume of the right long segment was added to age, sex and lesion size [\( R^2 = 0.568, F(4,11) = 3.62, P = 0.041 \); \( R^2 \) change: \( F(1,11) = 7.462, P = 0.020 \)]. Of the four predictors only age [\( \beta = -0.678, t(3.087), P = 0.010 \)] and the right long segment [\( \beta = 0.730, t(2.732), P = 0.020 \)] were independent predictors. Gender [\( \beta = 0.505, t(1.920), P = 0.081 \)] and lesion size [\( \beta = -0.441, t(-2.04), P = 0.066 \)] were marginally significant. For the right hemisphere the model with age, sex and lesion size explained approximately 30% of the variance of language performances at 6 months [\( R^2 = 0.275, F(3,12) = 1.520, P = 0.260 \)]. By adding the volume of the right long segment the model increased to 57%, which represents a statistically significant increase in predictive value (Fig. 3).

Cross-validation of the model

The AICc was used to estimate the fit of each model. The AICc is used for small sample size data to calculate the goodness of fit measure corrected for model complexity (i.e. increased number of predictors). Given that the predictive power of the model did not improve for left-hemispheric analysis the AICc model comparison applies only to the right hemispheric models. The AICc showed that the three-predictor model has an AICc of 121.13 and the four-predictor model has an AICc of 114.84. The model with the lower AICc, hence the four-predictor model, is the preferred model. This indicates that the increased predictive power of the four-predictor model is not simply due to the benefits of an added variable per se, but it is specific to the right long segment.

Power calculation analysis

A sensitivity analysis was conducted using the software package G*power. The result of this analysis demonstrated that our study was able to detect an effect size (\( f^2 \)) of 1.14 with a power of \( 1 - \beta = 0.80 \) given the sample size (\( n = 16 \)) and a specified \( \alpha = 0.05 \). This indicates that a minimum effect size of 1.14 is necessary to reach sufficient sensitivity. The two-level hierarchical regression model (i.e. hierarchical analysis of the three- and four-predictor model) has an effect size of \( f^2 = 1.88 \). This effect size is larger than the critical value (\( f^2 = 1.44 \)) determined in the previous sensitivity analysis and we could therefore assume that this analysis is sufficiently powered to detect a genuine effect with at least an 80% chance (see Supplementary material for more details).

Changes in symptom severity

Among the 16 patients who returned after 6 months, 81.25% were still aphasic according to the WAB-R and three patients (18.75%) showed full recovery. As a group, there was significant overall improvement in language functioning 6 months after stroke onset [Aphasia Quotient \( \text{baseline} = 43.48 \pm 28 \), Aphasia Quotient \( \text{6 months} = 85.68 \pm 8.44 \); \( t(15) = -6.759, P < 0.001 \); Fig. 4A]. Irrespective of the severity of aphasia, improvements were seen in all aphasic patients, however, with different recovery slopes (Fig. 4B). None of our patients deteriorated between the two assessments.

At baseline, the fluent group presented with less severe aphasia (as estimated by the Aphasia Quotient score) compared to the non-fluent group [Aphasia Quotient \( \text{fluuent} = 64.4 \pm 25.2 \), Aphasia Quotient \( \text{non-fluent} = 32.1 \pm 21.9 \); \( t(14) = -2.85, P < 0.01 \)]. This difference was
lost at 6 months [Aphasia Quotient_{fluent} = 88.9 \pm 9.1, \text{Aphasia Quotient}_{non-fluent} = 84.2 \pm 8.1; t(14) = -1.03, P > 0.05]. No sex differences were observed for aphasia severity at baseline and after 6 months [see Table 3 for mean and deviations, t_{Baseline}(16) = -0.087, P > 0.05; t_{6 \text{ months}}(14) = -0.510, P > 0.05]. Thrombolysed patients (11/16) did not differ in their baseline severity or language recovery compared to non-thrombolysed patients [t_{Baseline}(16) = 0.808, P > 0.05; t_{6 \text{ months}}(14) = -0.153, P > 0.05; see Supplementary material for details]. Lesion size correlated negatively with baseline severity [r(18) = -0.637, P < 0.004]. This significant negative correlation implies that bigger lesions are associated with a lower score on the WAB-R (i.e. more severe language impairments). The same test statistic applied to the follow-up aphasia severity was not significant [r(16) = -0.276, P = 0.30].

**Tract-specific measurements of fractional anisotropy**

Previous studies have shown that the asymmetry of the long segment is evident only for the volume measurements and not for fractional anisotropy (Thiebaut de Schotten et al., 2011b). Nonetheless, we additionally investigated if fractional anisotropy measures for each arcuate segment are associated with the aphasia severity (aphasia quotient) at baseline and at follow-up.

---

**Figure 3** Anatomical variability in perisylvian white matter anatomy and its relation to post-stroke language recovery. The right long segment of the arcuate fasciculus is shown for three patients presenting with different degrees of language recovery at 6 months. These patients are indicated in the regression plot of the volume of the right long segment plotted against the longitudinal aphasia quotient (AQ) (corrected for age, sex and lesion size).
Table 3  Baseline and longitudinal AQ group comparisons stratified by sex, fluency, and thrombosis

<table>
<thead>
<tr>
<th>Time point</th>
<th>Groups</th>
<th>n</th>
<th>Mean ± SD</th>
<th>t</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement over time</td>
<td>AQ_baseline</td>
<td>16</td>
<td>43.48 ± 28.0</td>
<td>6.76</td>
<td>15</td>
<td>0.00***</td>
</tr>
<tr>
<td>(Paired t-test)</td>
<td>AQ_longitudinal</td>
<td>16</td>
<td>85.68 ± 8.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex differences (Independent t-test)</td>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>6</td>
<td>42.18 ± 33.1</td>
<td>-0.087</td>
<td>16</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>12</td>
<td>43.4 ± 25.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>Female</td>
<td>6</td>
<td>84.3 ± 10.4</td>
<td>-0.510</td>
<td>14</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>10</td>
<td>86.5 ± 7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency differences (Independent t-test)</td>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fluent</td>
<td>6</td>
<td>64.84 ± 25.2</td>
<td>-2.85</td>
<td>16</td>
<td>0.01**</td>
</tr>
<tr>
<td></td>
<td>Non-fluent</td>
<td>12</td>
<td>32.1 ± 21.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>Fluent</td>
<td>5</td>
<td>88.9 ± 9.1</td>
<td>-1.03</td>
<td>14</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Non-fluent</td>
<td>11</td>
<td>84.2 ± 8.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thrombolysis (Independent t-test)</td>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thrombolysed</td>
<td>13</td>
<td>51.5 ± 23.6</td>
<td>0.808</td>
<td>16</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Non-thrombolysed</td>
<td>5</td>
<td>39.7 ± 28.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>Thrombolysed</td>
<td>11</td>
<td>85.9 ± 8.8</td>
<td>-0.153</td>
<td>14</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Non-thrombolysed</td>
<td>5</td>
<td>85.2 ± 8.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***Significance assume at the $P \leq 0.001$.
**Significance assumed on the $P \leq 0.01$.
AQ = aphasia quotient; n = number of subjects within the group; SD = standard deviation.

Figure 4  Mean (A) and individual (B) language function curves (total possible = 100) at baseline and follow-up ($n = 16$). Values 1–5 correspond to the severity classification according to the Western Aphasia Battery Revised manual. The blue dotted line signals the cut-off for normal language function (Aphasia Quotient > 93.8, category 5 according to the severity scoring). Only three patients reached the cut-off for normal language function at follow-up.
No correlations were found between fractional anisotropy measures in the three segments in either hemisphere and aphasia severity at either time point (Table 4).

**Discussion**

This longitudinal study in acute stroke is the first to prospectively examine anatomical predictors of language recovery with diffusion tensor imaging tractography. We observed that the volume of the long segment of the arcuate fasciculus in the right hemisphere (contralateral to the lesion) is an important predictive factor for recovery of language after stroke. The volume of the other segments in the right and left hemisphere was not correlated with recovery. In addition we have confirmed the importance of other predictive factors, including age, sex and lesion size.

In particular, in our sample, lesion size in the left hemisphere is the strongest predictor of post stroke aphasia recovery after 6 months. Lesion size has previously been shown to play an important role in recovery through a variety of different mechanisms, including tissue neuronal repair, reperfusion of stroke tissue, and recruitment of perilesional spared areas (Heiss et al., 1999; Warburton et al., 1999; Croquelois et al., 2003). In larger lesions, compensation could also occur after recruitment of ipsilateral circuits that were not involved with language before the stroke (Code, 2001; Hillis and Heidler, 2002; Crosson et al., 2007).

An original finding of our study is the predictive value of tractography-derived measurements of tract volume in the right hemisphere. Specifically, the volume of the long segment connecting posterior temporal and inferior frontal regions is a good predictor of longitudinal recovery of aphasia after stroke.

This result indicates that different mechanisms might be at play in the two hemispheres. As inferred from the left hemisphere regression model, lesion size is predictive of long-term outcome regardless of added diffusion tensor imaging measures. However, in the right hemisphere model the classical predictors do not sufficiently predict recovery, but when diffusion tensor imaging measures are added the model significantly improves and can explain up to nearly 60% of the variance in language performances at follow-up.

Recent tractography imaging studies show that this tract is involved in auditory memory (Catani et al., 2007; López-Barrosso et al., 2013) and may have a role in recovery even long after the stroke (Tuomiranta et al., 2013). López-Barrosso et al. (2013) reported in healthy subjects a correlation between higher performances in an auditory memory task for pseudo-words and strength of connectivity of the left long segment measured with both tractography (i.e. radial diffusivity) and functional connectivity (i.e. temporal correlation of blood oxygen level-dependent response between the three perisylvian regions connected by the long segment of the arcuate fasciculus). The right long segments has also a role in auditory memory tasks based on semantic clustering strategies, where a larger volume of this segment is related to better performances (Catani et al., 2007). Furthermore, a recent single case study reported a woman with stroke affecting the left arcuate fasciculus and resulting aphasia. After rehabilitation the patient was able to recover the ability to learn novel active vocabulary and the authors speculate that this was due to the presence of compensatory pathways in the right hemisphere (Tuomiranta et al., 2013).

This hypothesis is in line with previous PET and functional MRI studies that indicated an important role of the right hemisphere for aphasia recovery after stroke (Karbe et al., 1998; Cappa, 2000; Saur et al., 2006). Two possible mechanisms have been suggested: unmasking of previously ready language capacities or reorganization of right hemisphere language areas (Cappa, 2000). In both cases the presence of a larger right long segment could facilitate direct cross-talk between right hemisphere homologues of Broca’s and Wernicke’s regions. Additionally, these studies revealed a dynamic shift of activation to the contralosional hemisphere during recovery (Saur et al., 2006). This shift seems to be only advantageous for recovery of if of temporary nature (Hillis, 2006; Saur et al., 2006; Szaflarski et al., 2013). Nonetheless, in such cases the right hemisphere may be capable of temporarily adopting linguistic competence beyond the non-verbal aspects which have already been assigned to the non-dominant hemisphere (e.g. prosody, intonation, and affective content) (Ross et al., 1988; Turkeltaub et al., 2012). Our study suggests that these right-hemispheric language functions could be mediated by a specific segment of the arcuate fasciculus.

Currently, it remains difficult to determine how a pre-existing right arcuate could facilitate functional recovery. In our study, we used the number of voxels visited by the streamlines as a surrogate of tract volume. The determinants of right hemisphere tract size are currently unknown, but most likely depend on several factors, including the degree of fibre myelination, axonal number and diameter, and organization of fibres (Beaulieu, 2002).

### Table 4 Correlations between AQ and fractional anisotropy of the three arcuate segments stratified by hemisphere

<table>
<thead>
<tr>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional anisotropy three segments</td>
<td>Fractional anisotropy three segments</td>
</tr>
<tr>
<td>LS</td>
<td>AS</td>
</tr>
<tr>
<td>AQ baseline</td>
<td>$r = 0.268$</td>
</tr>
<tr>
<td></td>
<td>$P = 0.29$</td>
</tr>
<tr>
<td>AQ 6 months</td>
<td>$r = -0.017$</td>
</tr>
<tr>
<td></td>
<td>$P = 0.95$</td>
</tr>
</tbody>
</table>

AQ = aphasia quotient; AS = anterior segment; LS = long segment; PS = posterior segment.
All these biological characteristics of white matter are correlated with the speed of signal propagation and therefore could influence efficiency of signal processing between distant areas (Thiebaut de Schotten et al., 2011a; López-Barroso et al., 2013). Hence, we hypothesize that in people with smaller long segment in the right hemisphere communication between Broca’s and Wernicke’s right homologues would mainly rely on an indirect multi-synaptic pathway mediated by the anterior and posterior segments. The absence or a smaller size of the direct long segment could, therefore, affect the efficiency of communication between distant areas and hinder right hemisphere mechanisms of language recovery.

An alternative hypothesis of the specific contribution of the right long segment to recovery could be related to the functional specialization of the right anterior and posterior segments. These segments connect to the right inferior parietal lobe, a region involved in visuo-spatial attention tasks and often damaged in patients with visuo-spatial neglect. This suggests that while both direct and indirect pathways on the left hemisphere are involved in language functions (Fridriksson et al., 2013), in the right hemisphere the two pathways may have a different role (Corbetta et al., 2000; Kaplan et al., 2008; Thiebaut de Schotten et al., 2011a). Future studies combining structural and functional imaging will be able to elucidate the relationship between white matter anatomy, cortical plasticity and functional activation.

The study benefitted from cutting edge diffusion methodology applied to a clinical setting and our results encourage the use of tractography as part of the clinical routine. However, several limitations should also be acknowledged. Current diffusion tensor imaging algorithms are prone to implicit limitations, including generating the presence of false positives (i.e. non-existing tracts) and false negatives (i.e. absence of truly existing tracts) (Basser et al., 2000; Ciccarelli et al., 2008; Dell’Acqua and Catani, 2012; Dell’Acqua et al., 2013). These obstacles have to be considered with caution, especially when trying to extract quantitative measures within the lesioned hemisphere. In addition, the current study focused on the perisylvian pathways of the right hemisphere, but it cannot be excluded that other tracts might play an important role in recovery or redistribution of residual language capacities. Also, our observations reflect results derived from a selected group of patients that were able to tolerate a 45-min MRI scanning session in the acute stage of stroke. This patient group can be considered otherwise healthy before the stroke. Arguably, a high potential for plasticity can be assumed in these patients, which may not be as efficient in wider cohorts with known comorbidities and/or multiple strokes. Finally, the inclusion criteria for this study can be considered restrictive, but did allow us to exclude various potential nuisance variables that would have otherwise reduced the power.

In conclusion, tractography-based measurements of the arcuate fasciculus provide a good estimate of the presence of pre-existing right hemisphere networks that may serve as vehicle for functional compensation after stroke. Our findings, if confirmed in a larger cohort, could represent a step forward in our understanding of brain mechanisms underlying language recovery following stroke and may provide clinically useful predictive biomarkers.

Acknowledgements

Many thanks to our reviewers and contributing colleagues, Dr Matthew Howard, Dr Andre Marquand and Wasim Khan. We are deeply grateful to all patients and families who participated in this study as well as to the Friend’s Stroke Unit Team and the Department of Neuroimaging at King’s College London.

Funding

This study was supported by Guy’s and St.Thomas’ Charity (R080511) and Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, King’s College London and the ANR (project PHENOTYPES, no. ANR-13-JSV4-0001-01). The funding sources had no influence on the writing of the manuscript or the decision to submit it for publication.

Supplementary material

Supplementary material is available at Brain online.

References


