BRAIN
ORIGINAL ARTICLE

Recovery from aphasia in the first year after stroke

Stephen M. Wilson,[1,2] Jillian L. Entrup,[1] Sarah M. Schneck,[1] Caitlin F. Onuscheck,[1] Deborah F. Levy,[1] Maysaa Rahman,[1] Emma Willey,[1] Marianne Casilio,[1] Melodie Yen,[1] Alexandra C. Brito,[3] Wayneho Kam,[4,5] L. Taylor Davis,[2] Michael de Riesthal[1] and Howard S. Kirshner[4,5]

Most individuals who experience aphasia after a stroke recover to some extent, with the majority of gains taking place in the first year. The nature and time course of this recovery process is only partially understood, especially its dependence on lesion location and extent, which are the most important determinants of outcome. The aim of this study was to provide a comprehensive description of patterns of recovery from aphasia in the first year after stroke. We recruited 334 patients with acute left hemisphere supratentorial ischaemic or haemorrhagic stroke and evaluated their speech and language function within 5 days using the Quick Aphasia Battery (QAB). At this initial time point, 218 patients presented with aphasia. Individuals with aphasia were followed longitudinally, with follow-up evaluations of speech and language at 1 month, 3 months, and 1 year post-stroke, wherever possible. Lesions were manually delineated based on acute clinical MRI or CT imaging. Patients with and without aphasia were divided into 13 groups of individuals with similar, commonly occurring patterns of brain damage. Trajectories of recovery were then investigated as a function of group (i.e. lesion location and extent) and speech/language domain (overall language function, word comprehension, sentence comprehension, word finding, grammatical construction, phonological encoding, speech motor programming, speech motor execution, and reading).

We found that aphasia is dynamic, multidimensional, and gradated, with little explanatory role for aphasia subtypes or binary concepts such as fluency. Patients with circumscribed frontal lesions recovered well, consistent with some previous observations. More surprisingly, most patients with larger frontal lesions extending into the parietal or temporal lobes also recovered well, as did patients with relatively circumscribed temporal, temporoparietal, or parietal lesions. Persistent moderate or severe deficits were common only in patients with extensive damage throughout the middle cerebral artery distribution or extensive temporoparietal damage. There were striking differences between speech/language domains in their rates of recovery and relationships to overall language function, suggesting that specific domains differ in the extent to which they are redundantly represented throughout the language network, as opposed to depending on specialized cortical substrates.

Our findings have an immediate clinical application in that they will enable clinicians to estimate the likely course of recovery for individual patients, as well as the uncertainty of these predictions, based on acutely observable neurological factors.

1 Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA
2 Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA
3 School of Medicine, Vanderbilt University, Nashville, TN 37232, USA
4 Department of Neurology, Vanderbilt University Medical Center, Nashville, TN 37232, USA
5 Vanderbilt Stroke and Cerebrovascular Center, Vanderbilt University Medical Center, Nashville, TN 37232, USA

Correspondence to: Stephen M. Wilson
Department of Hearing and Speech Sciences, Vanderbilt University Medical Center
1215 21st Ave S, MCE 8310, Nashville, TN 37232, USA
E-mail: s.m.wilson@vanderbilt.edu

Introduction

Aphasia is one of the most common and debilitating consequences of stroke. Fortunately, most stroke patients experience some degree of recovery of speech and language function over time. Recovery has a decelerating time course, with the greatest gains taking place early, and the slope of change then decreasing.[1–6] Several decades of work have established that the primary predictors of extent of recovery, or outcome, are lesion location and lesion extent, and especially lesion extent in key posterior perisylvian regions.[7–18] In contrast, demographic factors such as age, sex, handedness, and education have minimal predictive value.[5,19]

Despite much productive research on recovery from aphasia after stroke, it remains challenging to translate what has been learned into clinical practice, specifically for accurate prognostication, which is important for educating patients and caregivers and planning rehabilitation services. The overarching goal of the present study was to provide a clear and comprehensive description of patterns of recovery that will allow clinicians to answer a pressing question that arises every day in stroke units all over the world: for a given pattern of brain damage, to what extent can speech and language deficits be expected to resolve over time?

This question is difficult to answer in practice for a number of reasons. Most studies that have compared trajectories of recovery between different groups of individuals with aphasia have subdivided patients by initial severity[12,20–23] or by initial aphasia subtype.[4,24–28] However, initial severity can be strongly influenced by medical factors and the precise timing of initial testing, which is often confounded with these factors.[29,30] Aphasia subtypes are not natural kinds,[31] and in part because of this can be highly dynamic, especially in the early post-stroke period.[26] Initial severity and subtype are certainly dependent on lesion location,[11,32,33] but these relationships are far from straightforward.[34] Therefore, initial severity and subtype are not optimal organizing principles, given that the biological determinant of recovery potential is ultimately neuroanatomical.

Some studies have investigated recovery in cohorts of individuals with aphasia that were defined in whole or in part by lesion location.[7,11,12,35–42] These studies have been very informative, but all have had relatively small sample sizes, and most have investigated patients with just one or two different lesion sites, with no studies providing a comprehensive and comparative description of a range of different patterns of damage. Other studies have investigated the predictive value of damage to specific brain regions, especially posterior perisylvian regions.[9–13,15,17,18] While these studies have firmly established the critical role of posterior perisylvian cortex in determining aphasia outcome, which concords with other sources of information,[43,44] the data have not been presented in such a way as to allow clinicians to determine likely outcomes for patients with particular lesion locations.

Another limitation of work to date is that the nature of aphasia as it changes over time has rarely been characterized in detail. Most studies have reported outcome measures of overall aphasia severity,[4,18,20,28,45,46] which in some of the largest studies have been quite crude.[21,47,48] Global measures can be misleading, since they fail to capture important differences between individuals.[49] Some studies have characterized recovery in terms of transitions between aphasia subtypes,[26,27] which is limited because there is great variability within subtypes.[50,51] Other studies have reported recovery of expressive and receptive language functions separately,[3,25,52] sometimes cross-cut by written and spoken modalities.[2,53] Relatively few studies have described changing patterns of performance in specific language domains, generally in relatively small samples,[13,24] for specific types of patients,[11,12,41,54] for single aspects of language function,[9,10] or without respect to lesion location.[6,23] Comprehension has rarely been subdivided into word comprehension and sentence comprehension (but *cf.* Selnes *et al.*[9,10]), and apraxia of speech (AoS) and dysarthria have almost never been characterized and considered in relation to other speech production deficits in studies of recovery (but *cf.* Hybbinette *et al.*[55]).

The present study aimed to fill these gaps by providing a comprehensive description of patterns of recovery in a large and representative cohort of individuals with post-stroke aphasia, organized in terms of the fundamental determinants of outcome: lesion location and extent. We characterized speech and language function at four time points: 1–5 days, 1 month, 3 months, and 1 year after stroke. Trajectories of recovery across multiple speech and language domains are described for 13 commonly occurring patterns of anatomical damage, 11 of which are associated with aphasia. We then discuss implications of our findings for understanding the underlying mechanisms of recovery.

Materials and methods

Participants

A total of 334 individuals participated in the study: 218 with aphasia and 116 without aphasia (Fig. 1A). Demographic and key medical history data are shown in Table 1. Over a 3.3-year period from late 2016 to early 2020, we considered for inclusion all patients presenting at the Vanderbilt Stroke and Cerebrovascular Center at Vanderbilt University Medical Center (VUMC), by weekday attendance of a 'stroke huddle' and weekday review of electronic medical records. Our inclusion criteria were: (i) acute ischaemic or haemorrhagic stroke predominantly confined to left hemisphere supratentorial regions, or right hemisphere stroke with aphasia clearly indicating right hemisphere language dominance; (ii) age 18–90 years; and (iii) infarct at least 1 cm^3, except (a) thalamic infarcts were included regardless of extent; and (b) starting after ~21 months of data collection, basal ganglia and/or subcortical white matter infarcts were included only if they exceeded ~6 cm^3. Our exclusion criteria were: (i) unconscious with grave prognosis; (ii) not fluent
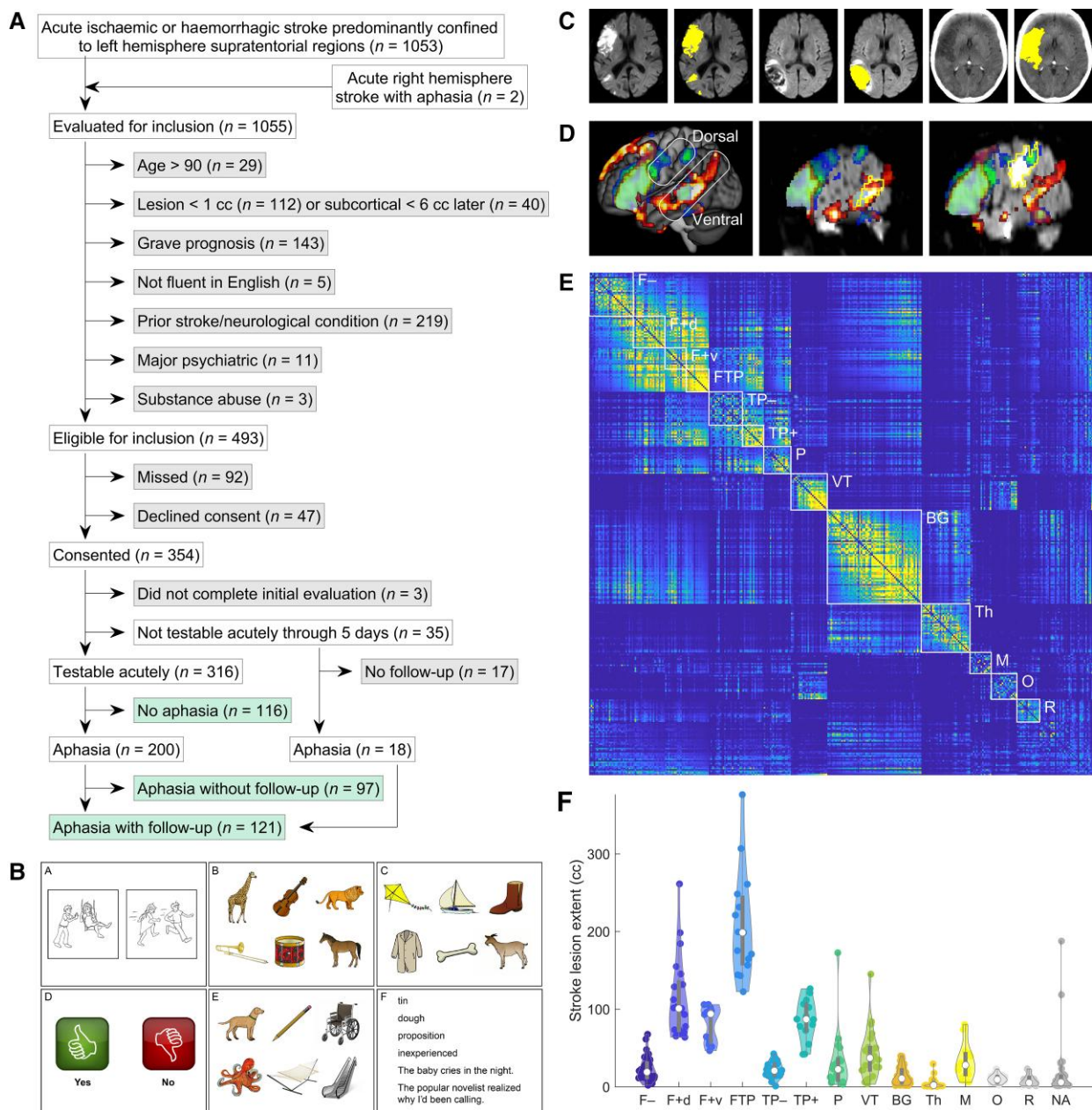
Figure 1 Methodological details. (**A**) Study cohort. Included groups of patients are shown in green, and excluded groups are shown in grey. (**B**) Example slides from the QAB; Wilson *et al.*,[56] used with permission. (**C**) Examples of manual lesion delineation on DWI, FLAIR, and CT images. (**D**) Activation maps for semantic decision (hot colours)[65] and rhyming decision (cool colours)[66] paradigms. Data are from reference[66] ($n = 16$, voxelwise threshold: $P < 0.005$; corrected at $P < 0.05$ based on cluster extent using permutation analysis). Based on these data, temporal and parietal language regions are subdivided into dorsal and ventral streams; note that the angular gyrus is a ventral stream region even though it is located in the parietal lobe. (**E**) Pairwise similarity matrix of all lesion images. Patients are ordered by final group assignments and then by lesion extent (ascending). Groups are indicated with white rectangles. F– = Frontal, less extensive; F+d = Frontal, more extensive, but sparing ventral stream; F+v = Frontal, more extensive, impacting ventral stream; FTP = Complete or near-complete perisylvian; TP– = Temporal or temporoparietal, less extensive; TP+ = Temporal or temporoparietal, more extensive; P = Parietal; VT = Ventral temporal; BG = basal ganglia; M = Midline; Th = Thalamus; O = Occipital; R = Rolandic; NA = none of the above. (**F**) Violin plot of lesion extent by group. White circles = medians; thick lines = interquartile range (IQR); thin lines = 1.5 × IQR.

in English premorbidly (note that non-native speakers were included so long as they were fluent in English); (iii) prior symptomatic stroke significantly impacting language regions or homotopic regions, neurodegenerative disease, or any other neurological condition impacting language or cognition; (iv) major psychiatric disorder; and (v) substance abuse serious enough to interfere with study participation. As detailed in Fig. 1A, 1055 patients met the first inclusion criterion and were evaluated for

inclusion, and 493 patients met all inclusion and exclusion criteria. Of these 493 patients, 401 were approached by a speech-language pathologist (J.L.E., S.M.S., or C.F.O.) at the bedside to request consent; the remaining 92 were missed, mostly due to weekends, holidays or rapid discharge of very mild cases. Of the 401 patients approached, 354 patients or their legally authorized representatives (88%) gave written informed consent to participate in the study. Our study was conducted in

Table 1 Demographic and medical history data

	Aphasia (*n* = 218)	No aphasia (*n* = 116)
Age	62.6 ± 13.6 (21–90) years	61.6 ± 16.0 (21–90) years
Sex	117 M; 101 F	54 M; 62 F
Handedness	193 R; 20 L; 5 Am	100 R; 13 L; 3 Am
Education	12.9 ± 3.2 (0–20) years	13.5 ± 2.6 (7–20) years
Race	182 W; 34 B; 1 A; 1 H	100 W; 16 B
TOAST stroke type		
Large-artery atherosclerosis	41 (19%)	17 (15%)
Cardioembolism	65 (30%)	35 (30%)
Small-vessel occlusion	6 (3%)	10 (9%)
Undetermined aetiology	63 (29%)	29 (25%)
Haemorrhage	43 (20%)	25 (22%)
Haemorrhagic transformation	28 (13%)	7 (6%)
Lesion extent	54.5 ± 61.7 (0.6–376.4) cm^3	9.6 ± 10.6 (0.5–61.9) cm^3
Hypertension	174 (80%)	88 (76%)
Diabetes Mellitus	65 (30%)	34 (29%)
Hyperlipidaemia	133 (61%)	59 (51%)
Coronary artery disease	43 (20%)	17 (15%)
Cardiac arrhythmia	28 (13%)	19 (16%)
Atrial fibrillation	47 (22%)	23 (20%)
Myocardial infarction	22 (10%)	6 (5%)

A = Asian; Am = Ambidextrous; B = Black; F = Female; H = Native Hawaiian; L = Left-handed; M = Male; R = Right-handed; W = White.

accordance with the principles of the 1964 Declaration of Helsinki and was approved by the Institutional Review Board at VUMC.

A total of 316 patients were able to be tested within the first 5 days after stroke (Time 1, T1), with 'testable' defined as 'able to stay awake, maintain attention, and attempt to follow commands, potentially with some lapses'. The 316 patients who were tested were all included in the study: per clinical judgement, 200 were aphasic and 116 were not, the threshold for diagnosing aphasia being that the speech-language pathologist assessed that deficits were present to a degree that would justify a recommendation for speech-language services on discharge. Three patients declined to complete the initial evaluation after consenting and were excluded from the study, while 35 patients remained untestable through the first 5 days. Of these, 18 were later tested and were included in the study, while 17 were never tested and were therefore excluded.

We sought to obtain follow-up language evaluations at 1 month (Time 2, T2), 3 months (Time 3, T3), and 1 year (Time 4, T4) for all patients with aphasia on initial evaluation and all untestable patients (who were presumed to be likely to have aphasia). We obtained one or more follow-up time points for a total of 121 patients: 103 of 200 of the patients with initial aphasia (52%) and 18 of 35 of the initially untestable patients (51%); these proportions did not differ.

A total of 589 evaluations were obtained. There were 316 T1 evaluations obtained on median day 3 (mean 2.7 ± 1.3 SD days; mode 2; range 0–7 days, but note that only five evaluations were obtained outside the range of 1–5 days); 101 T2 evaluations obtained on median day 35 (mean 35.5 ± 8.2 days; range 12–55 days); 98 T3 evaluations obtained on median day 97.5 (mean 100.7 ± 11.8 days; range 84–141 days); and 74 T4 evaluations obtained on median day 381 (mean 410.8 ± 71.6 days; range 360–665 days; note that 12 T4 evaluations were obtained later than 15 months due to the Covid-19 pandemic).

Of the 235 patients who had aphasia on initial evaluation or who were initially untestable, 78 patients participated throughout the first year after stroke (this does not necessarily mean that all four time points were obtained; some time points were missing for situational reasons). The remaining patients were discontinued either after one or more follow-up time points were acquired (43 patients) or without the acquisition of any follow-up time points (114 patients), for the following reasons: 35 patients (15%) lived too far away (more than a 2 h drive), 50 patients (21%) were no longer contactable at some point, 33 patients (14%) declined further participation at some point, 15 patients (6%) passed away, 19 patients (8%) were unable to participate further because they had major health problems or were transferred to hospice, two patients (1%) had new strokes affecting language function, and three patients (1%) were discontinued for other reasons.

Speech/language evaluation

Speech and language were evaluated at each time point with the QAB (Fig. 1B), which we designed specifically for this study.[56] In brief, the QAB aims to provide a reliable and multidimensional assessment of language function in about a quarter of an hour, bridging the gap between comprehensive batteries that are time-consuming to administer, and rapid screening instruments that provide limited detail regarding individual profiles of deficits. The QAB is made up of eight subtests, each comprising sets of items that probe different language domains, vary in difficulty, and are scored with a graded system to maximize the informativeness of each item. The QAB has three equivalent forms to minimize repetition of items when studying individuals longitudinally.

From the eight subtests, a number of summary measures are derived, which constitute a multidimensional profile of language function, quantifying strengths and weaknesses across core language domains. In this study, we report nine summary measures: overall language function, word comprehension, sentence comprehension, word finding, grammatical construction, phonological encoding, speech motor programming, speech motor execution, and reading. Most have been described in detail previously,[56] but we added the summary measures of phonological encoding and speech motor execution in the present study. Phonological

encoding was operationalized in terms of the prevalence and severity of phonological paraphasias in connected speech, which was rated on a 0–4 scale (0 = severe, 1 = marked, 2 = moderate, 3 = mild, 4 = normal) as part of the rating of the connected speech sample. Speech motor execution was operationalized in terms of the extent of dysarthria, which was rated as part of the motor speech subtest. We do not report one QAB summary measure—Repetition—because it lacks specificity, being impacted by deficits at many stages.

Several summary measures could be missing in patients with limited output, namely phonological encoding, speech motor programming, and speech motor execution. These measures were treated as 'missing' rather than zero when they could not be scored. Reading could not be assessed in four individuals with limited baseline reading ability, in which case QAB overall scores were calculated based on spoken language measures only.

All QAB summary measures described previously[56] exhibit strong concurrent validity with respect to the widely used Western Aphasia Battery–Revised (WAB–R),[50] and all have excellent inter-rater reliability, with respect to Cicchetti's[57] guidelines. Test-retest reliability is good to excellent for all QAB summary measures.[56] The only two measures for which reliability is good rather than excellent are word comprehension and sentence comprehension. Therefore, for most follow-up evaluations (215 of 273, starting ~14 months after the onset of data collection), we doubled the number of word comprehension and sentence comprehension items by presenting items from two different forms. Based on psychometric calculations, this improves the reliability of these measures from good to excellent (the intraclass correlation coefficient for both measures increases to 0.83).

In narrative descriptions of our results, it is at times useful to describe different degrees of severity; when we do so, we use terms according to the following criteria: severe (score < 5.0); moderate (5.0 ≤ score < 7.5); and mild (7.5 ≤ score < 8.9). These ranges are based on the use of these terms for the WAB–R Aphasia Quotient.[50] Scores ≥ 8.9 are described as 'recovered', but this is not intended to imply that there is no residual deficit at all; rather, it reflects the cutoff associated with optimal sensitivity and specificity for distinguishing individuals with and without aphasia on the QAB.[56]

All language evaluations were administered by certified speech-language pathologists (J.L.E., S.M.S., or C.F.O.). All acute evaluations were performed at the bedside, and most follow-up evaluations were performed at VUMC, inpatient rehabilitation facilities, or individuals' homes. Each session was recorded with a Marantz PMD661 Professional Portable Flash Field Recorder and videotaped with a GoPro HERO3+ or HERO6. After the onset of the Covid-19 pandemic, 17 follow-up evaluations were carried out via remote videoconferencing using Zoom.

Sessions were transcribed and scored from the recordings. The transcription and scoring of every patient's speech and language evaluations were comprehensively reviewed and edited in consensus meetings of four to six authors, always including the first four authors. These consensus meetings took place after all the data were acquired. The researchers involved had knowledge of the patients' longitudinal trajectories as well as their lesion locations; however, every effort was made to score each evaluation objectively and capture whatever changes did or did not take place. We chose this approach because blinding to time point would have been difficult to achieve, given the use of video data recorded in different physical locations at each time point (e.g. bedside, home). Also, we found that by considering each evaluation in its

longitudinal context, we were able to gain a greater understanding of each individual's evolving presentation, and hence to score their evaluations more accurately.

Neuroimaging

MRI or CT imaging was obtained in the course of routine clinical care. For each of the 334 patients included, a study was chosen on which to delineate the lesion according to the following descending order of preference: (i) first MRI obtained at VUMC; (ii) first MRI obtained at an outside hospital; (iii) first CT obtained at VUMC where the lesion was visible; and (iv) first CT obtained at an outside hospital where the lesion was visible. MRI was used for 270 patients and CT for 62 patients, while one patient's lesion could not be seen clearly on any scan, and one patient did not have any imaging performed.

All imaging obtained within the first 30 days after stroke was reviewed to determine if the lesion extent increased or if there were any new lesions subsequent to the initial imaging study. This was the case for 12 patients. For three of these patients, extension of the initial lesion occurred after the language evaluation had been obtained, and there were no follow-up data; therefore, the initial lesion, prior to the subsequent extension, was used in our analyses. For the other nine patients, review of the clinical course suggested that the extension of the lesion took place prior to the language evaluation, so the extended lesion was used in our analyses.

Lesions were manually delineated using ITK-SNAP version 3.6.0 running on Linux workstations (Fig. 1C). A neuroradiologist (L.T.D.) provided training to the researchers drawing the lesions and was consulted for input on specific cases as necessary. Ischaemic strokes were generally drawn on diffusion weighted imaging (DWI), but apparent diffusion coefficient (ADC) images were consulted to ensure that hyperintense signal on DWI was associated with restricted diffusion, and fluid attenuated inversion recovery (FLAIR), T_2^*-, and T_2-weighted images were also considered. Haemorrhages were generally drawn on FLAIR; surrounding oedema was not included. Again, other modalities were considered as needed, especially T_2^*.

DWI images were coregistered to FLAIR images using a 12-parameter affine transformation in SPM12; this yielded better results than the standard six-parameter approach, because it better accounted for warping of the diffusion weighted images. FLAIR images were segmented and warped to Montreal Neurological Institute (MNI) space using Unified Segmentation[58] implemented in SPM12. Prior to normalization, the lesion was replaced with healthy tissue from the opposite hemisphere (enantiomorphic normalization[59]), implemented with in-house code in MATLAB R2019a. CT images were normalized with the same approach; good normalizations were obtained with the Unified Segmentation algorithm after shifting values into the positive range, setting the origin to the centre of mass of the image, and in some cases, stripping the skull. There were many patients for whom coregistration and/or normalization gave poor results or failed entirely with the clinical images used in this study; these were handled with various strategies on a case by case basis, and ultimately all images were adequately coregistered and normalized.

Lesion-based grouping

To investigate how trajectories of recovery depend on lesion location and extent, we divided the patients into 13 groups based solely

on their lesions, without reference to speech and language data. The objective was to define groups of individuals with similar lesions reflecting common distributions of ischaemic or haemorrhagic stroke damage.

First, we calculated a similarity metric s between each pair of patient lesion images x and y as follows:

$$s_{xy} = \sum \min(v_{xi}, v_{yi}) / \sum \max(v_{xi}, v_{yi})$$
$$\cdot \left(\sum \min(1 - v_{xi}, 1 - v_{yi}) / \sum \max(1 - v_{xi}, 1 - v_{yi}) \right)^{10} \quad (1)$$

The first term is the weighted Jaccard similarity coefficient across all voxels of the smoothed lesion images. The second term is the weighted Jaccard similarity coefficient of the inverse lesion images, raised to the 10th power; multiplication by this term has the effect of penalizing non-overlap among patients with larger lesions, which yields a more intuitively satisfactory similarity metric.

We next used agglomerative hierarchical clustering as implemented in the MATLAB function 'linkage' to identify representative clusters of lesions, using the unweighted pair group method with arithmetic mean (other algorithms were also explored). However, no automated methods were found to yield intuitively coherent groups of patients distinguished with respect to critical anatomical features. Therefore, we used the hierarchical cluster tree as a starting point, but then manually sorted individuals into groups, taking into account the functional anatomy of the language network.

In particular, we were guided by the concept of dorsal and ventral streams,[60–63] which have been shown to have explanatory value in characterizing individuals with aphasia.[64] Our prior functional imaging studies, in line with much previous work, have shown a core frontal language region in the inferior frontal gyrus (IFG), and a core temporoparietal language region along the length of the superior temporal sulcus (STS).[65,66] This temporoparietal language region belongs to the ventral stream, including its extension into the angular gyrus (Fig. 1D).[67] Additionally, there are language regions implicated in phonological encoding for speech production in the supramarginal gyrus and ventral precentral gyrus[66]; these regions belong to the dorsal stream (Fig. 1D).

Lesion-based groups of patients were thus defined with reference to this model (Table 2). For example, patients with lesions to the ventral stream temporoparietal regions were distinguished from those with damage to dorsal stream parietal regions

(Fig. 1D). The anatomical criteria for the 13 groups are described in more detail in the 'Results' section. Of the 332 patients with lesion images, 297 patients (89%) were assigned to one of the 13 groups, while 35 patients had idiosyncratic patterns of damage that did not meet criteria for any group.

The similarity matrix between all individuals' lesion images, ordered by their group assignments and then by lesion extent (ascending) is shown in Fig. 1E. The coherence of each of the 13 groups can readily be observed. The distribution of lesion extent in each group is shown in Fig. 1F. Note that in cases where groups were distinguished based on lesion extent, there is some overlap of ranges; this is because grouping assignments were made based on extent of damage to the relevant language regions, not simply by total lesion extent.

For the purpose of illustrating representative individuals from each group, four objectively representative patients were identified from each group, by evaluating all possible sets of four patients belonging to the group, and choosing the set of four who maximized the median of the similarity between each patient who was not selected, and whichever of the four they were most similar to.

Statistical analysis

Model fitting

Initial scores, obtained at the acute time point, for overall language function and for each of the eight subscores characterizing different speech and language domains were modelled separately. We fit linear models using the 'fitlm' function in MATLAB. In each model, the dependent variable was the initial score, and the independent variables were lesion location (i.e. one of the 13 lesion-defined groups, or none of them), lesion extent (specified as a polynomial of degree 2), stroke type (ischaemic or haemorrhagic), age, sex, handedness, and education.

Recovery was also modelled separately for overall language function and for each of eight domain-specific subscores. We modelled change over time by fitting linear mixed-effects models using the 'fitlme' function in MATLAB. The observations consisted of pairs of consecutive time points, with participant identity included as a random effect. The dependent variable was the difference between scores obtained at consecutive time points. The independent variables were lesion location, lesion extent (specified as a polynomial of degree 2), preceding score (i.e. the score at the earlier of the two time points making up the pair), which was specified as a polynomial of degree 2, time point pair (i.e. T1-T2, T2-T3, or T3-T4), stroke type, age, sex, handedness, and education.

In the recovery models, the dependent variables (differences between scores) were transformed using an inverse hyperbolic sine function as follows:

$$\Delta y' = \text{asinh}(\Delta y \cdot \lambda) / \lambda \quad \text{with } \lambda = 4 \quad (2)$$

The rationale for this transformation, which essentially suppresses high values, was to reduce heteroscedasticity, because there was more variance when starting scores were lower, since lower starting scores entail more room for improvement. Note that we took the approach of modelling changes in scores rather than modelling scores directly because of the strong dependence of change on preceding scores, which could not readily be modelled by random slopes in models with multiple time points.

When patients were untestable at T1 (and occasionally T2), QAB overall was imputed to be zero, while all subscores were treated as

Table 2 Lesion-based groups

Label	Definition	n
F−	Frontal, less extensive	29
F+d	Frontal, more extensive, but sparing ventral stream	22
F+v	Frontal, more extensive, impacting ventral stream	14
FTP	Complete or near-complete perisylvian	15
TP−	Temporal or temporoparietal, less extensive	22
TP+	Temporal or temporoparietal, more extensive	14
P	Parietal (dorsal stream only)	18
VT	Ventral temporal	24
BG	Basal ganglia	62
Th	Thalamus	32
M	Midline (medial frontal/parietal and/or cingulate gyrus)	14
O	Occipital	17
R	Rolandic (superior to mouth/laryngeal sensorimotor regions)	15
NA	None of the above	35

missing. When time points T2 or T3 were missing for situational reasons, and were bookended by acquired time points, these time points were imputed prior to model fitting, by setting the value of the missing time point to the mean relative distance between the bookends calculated for all patients with all three relevant time points acquired. In the occipital and Rolandic groups, there were very few patients with aphasia, and only one with follow up data, so this patient was excluded from the model.

We constructed separate models that included the extent of speech-language therapy that patients had received between time points as an additional independent variable. The potential effect of speech-language therapy was evaluated using separate models, because we only obtained this information in 215 of 273 follow-up evaluations (starting ~14 months after the onset of data collection).

For models of initial scores and models of recovery, statistical significance was assessed with likelihood ratio tests comparing full and reduced models. The variance explained by different models was also estimated using likelihood ratio statistics.[68]

Trajectories by group

Trajectories of recovery were plotted for each speech/language measure as a function of the lesion-defined group. Although only data actually obtained are plotted, to estimate means and variances, it was necessary to impute missing time points. This was done by using the fitted models to project forward in time for patients who were discontinued at any time point. Similar models were constructed to predict preceding recovery between time points (i.e. using the earlier score as the dependent variable and the later score as an independent variable) and were used to project back in time to model earlier time points in cases where patients were untestable early in the course of their recovery. Means were estimated from the imputed data. Variances were estimated by creating 100 imputed datasets that included random prediction errors and then obtaining the median variance across the 100 iterations. Patients who were initially untestable were counted twice in calculations of means and variances, in order to simulate the contributions of the approximately half of initially untestable but presumably aphasic patients who were excluded from the study because no data were ever obtained.

Effects of lesion location or extent within groups

We fit within-group models to determine whether initial scores, final scores, or proportional recovery, differed within lesion groups as a function of lesion location or extent. Analyses of final scores were based only on patients for whom at least one follow-up time point was obtained (T2, T3, or T4), which implies that these were patients with aphasia (because we did not obtain follow-up data from patients without initial aphasia). Analyses of proportional recovery were based only on patients for whom at least two testable time points were acquired (again implying that these were patients with aphasia), and for whom the initial score was ≤8 (because proportional recovery becomes meaningless as scores approach ceiling). Proportional recovery was calculated between the initial and final time points (T1, T4). For all three measures, imputed scores were used for all missing time points. Lesion location was reduced to a scalar variable by calculating the centre of mass of each patient's lesion, conducting a principal components analysis on these 3D centroids within each group, and then summarizing each patient's lesion location in terms of the first principal component, which captured the axis of most variance. For each group (11

groups, since two groups rarely presented with aphasia), speech-language domain, and measure (initial, final, proportional recovery), a linear model was fit to determine whether the measure in question was influenced by lesion location and/or lesion extent. The 297 P-values thus obtained were corrected for multiple comparisons using the false discovery rate procedure.[69]

Comparisons between speech/language domains

To explore how performance on specific speech/language domains related to overall language function, we plotted all patients' trajectories in 2D spaces defined by overall language function on one axis and each speech/language domain in turn on the other axis. While the summary scores for each speech/language domain each make a contribution to the overall scores, this analysis nevertheless provides insight into differences between domains in the extent to which they can dissociate from other domains.

Data availability

The data and MATLAB code for the main analyses are provided in the Supplementary material. The complete dataset, including transcribed QAB evaluations and imaging, will be made available in future at: https://langneurosci.org/recovery.

Results

Initial evaluation

Of the 334 individuals included in the study, 218 presented with aphasia, comprising 200 patients who were tested in the immediate post-stroke period and 18 patients who were untestable acutely but were later found to be aphasic on follow-up. Among the individuals with aphasia, there was a wide range of overall initial severity (QAB overall mean = 5.8 ± 2.7, range = 0–9.8). Among patients who were testable acutely, there was no difference in initial aphasia severity between patients for whom follow-up data were obtained (mean overall score = 5.6 ± 2.7) versus not obtained (mean overall score = 6.1 ± 2.7, $t(198) = 1.31$, $P = 0.19$). This suggests that the longitudinal dataset is a representative sample of patients meeting the inclusion and exclusion criteria.

The remaining 116 individuals did not have aphasia per clinical impression and generally performed close to ceiling (QAB overall mean = 9.3 ± 0.5, range 7.4–10.0). Note that some non-aphasic individuals made errors on certain subtests that reflected dysarthria or non-specific post-stroke cognitive factors.

A linear model of initial severity (QAB overall) as a function of lesion location, lesion extent, stroke type, and demographic factors explained 59.5% of the variance (Supplementary Table 1). This variance was almost entirely accounted for by lesion location [$\chi^2(13) = 59.1$, $P < 0.0001$] and extent [$\chi^2(2) = 26.3$, $P < 0.0001$]; in combination, these two lesion factors explained an additional 58.0% of the variance relative to a model with no lesion variables (which explained just 1.5% of the variance). Initial scores were lowest for the F+d, F+v, FTP, and TP+ groups, and highest for the O and R groups (see Table 2 for group label definitions), and larger lesions were associated with lower initial scores through much of the range of the data. There was a modest negative effect of age [$\beta = -0.03$ per year, $\chi^2(1) = 14.2$, $P = 0.0002$, $\Delta r^2 = 1.8\%$] and a modest negative effect of haemorrhagic stroke [$\beta = -0.97$, $\chi^2(1) = 10.7$, $P = 0.0011$, $\Delta r^2 = 1.3\%$]. There were no effects of sex ($P = 0.26$), handedness ($P = 0.27$), or education ($P = 0.18$).

For the eight domain-specific subscores (Supplementary Table 2), lesion location and extent similarly accounted for most

of the explainable variance and there were modest or no effects of the other factors. Of note, among the demographic factors, the word comprehension and sentence comprehension subscores were most impacted by age, while the sentence comprehension and reading subscores were most dependent on education.

Trajectories of recovery

Longitudinal data were obtained for 121 individuals with aphasia. A linear mixed-effects model of overall recovery (QAB overall) as a function of lesion location, lesion extent, preceding score, time point, stroke type, and demographic factors explained 59.2% of the variance (Supplementary Table 3). Recovery was strongly predicted by preceding score [$\chi^2(2) = 133.6$, $P < 0.0001$], with larger improvements taking place when preceding scores were lower, and by time point [$\chi^2(2) = 13.3$, $P = 0.0013$]: recovery was decelerating, as has been observed many times previously, with the greatest recovery between 1–5 days and 1 month, followed by between 1 month and 3 months, and then between 3 months and 1 year, even though these intervals were progressively longer. Together, these two factors explained an additional 44.9% of the variance compared to a model excluding these factors. However, there were also significant contributions from lesion location [$\chi^2(11) = 42.2$, $P < 0.0001$] and extent [$\chi^2(2) = 5.38$, $P = 0.043$]. Taken together, these two lesion variables explained an additional 12.1% of the variance relative to a model with no lesion variables. The F−, F+d, F+v, P, and M groups made the greatest gains; the FTP and TP+ groups made the least gains (see Table 2 for group label definitions); and smaller lesions were associated with greater gains. Recovery was not dependent on stroke type ($P = 0.12$), age ($P = 0.20$), sex ($P = 0.71$), handedness ($P = 0.22$), or education ($P = 0.067$).

For the eight domain-specific subscores (Supplementary Table 4), preceding score and time point similarly explained much of the variance, but lesion location and extent also contributed to explaining recovery of most of the subscores (except for speech motor execution), while the other factors examined showed modest or no effects.

Patients received 107.4 ± 108.8 min (range 0–360 min) of speech-language therapy per week between 1–5 days and 1 month, 80.5 ± 77.6 min (range 0–300 min) between 1 month and 3 months, and 33.6 ± 50.3 min (range 0–205 min) between 3 months and 1 year. The amount of speech-language therapy received was not predictive of recovery [$\chi^2(1) = 0.12$; $P = 0.73$].

Patterns of recovery by lesion location and extent

Recovery trajectories for each of the 13 lesion-defined groups will now be described. Note that there were very few 'within-group' effects of lesion location or extent on initial scores, final scores, or proportional recovery, for any speech/language domain. These effects were all non-significant except where stated otherwise.

Frontal, less extensive: F−

There were 29 patients with relatively circumscribed frontal lesions (Fig. 2A). Maximal lesion overlap (18 of 29 patients) was observed in the anterior ventral precentral gyrus. While most patients had at least some involvement of the IFG, it was clear that these smaller frontal lesions were much more likely to be localized to the ventral precentral region posterior to Broca's area than to Broca's area itself. Of the 29 patients, 23 had aphasia on initial evaluation. At the initial time point, there was a wide range of severity, but recovery was invariably rapid, and by 1 month, all

but two patients scored in the mild to recovered range. Some patients had early word comprehension deficits, but all were close to ceiling by 1 month. Sentence comprehension, word finding, and grammatical construction recovered in line with overall language function. Phonemic paraphasias were typically absent, but when present were usually mild and often recovered. AoS was present acutely in 12 patients (41%); another was unscorable on AoS due to limited output. Follow-up data were acquired for 6 of these 12 patients; three showed improvements in AoS but still had mild AoS at 1 year, one recovered completely at 1 year, and two maintained a mild rating throughout the year. Dysarthria was acutely present in nine patients (another was unscorable on dysarthria due to AoS), but by 1 month, dysarthria persisted in just one patient. Reading aloud recovered in line with overall language function.

Frontal, more extensive, but sparing ventral stream: F+d

This group of 22 patients had large frontal lesions that mostly extended posteriorly into the anterior parietal lobe, often impacting dorsal language regions involved in phonological encoding (Fig. 2B). Critically, in this group, the ventral stream language areas of the STS and angular gyrus were largely spared. Maximal lesion overlap was again observed in the anterior ventral precentral gyrus, but in this group, 21 of 22 patients had overlapping lesions there. All 22 patients had aphasia acutely (although three were untestable), mostly severe, with a few in the moderate range. Recovery was less rapid than in the previous group, but was steady, and by 1 year, most patients' language function had recovered to the mild to moderate range. Word comprehension was variable acutely but close to ceiling by 1 month except in one patient. Sentence comprehension, word finding, and grammatical construction recovered in line with overall language function. About half of the patients made phonemic paraphasias, varying widely in severity, and these tended to recover only minimally. AoS was present acutely in eight patients, but usually improved over time, resolving to mild or completely within the year in the six patients for whom follow-up data were obtained. In another eight patients, AoS could not be assessed due to limited speech output or patients being untestable. Of these eight patients, five were followed up, all had AoS, and three still had moderate AoS at 1 year. Dysarthria was acutely present in seven patients, with another seven untestable or unscorable. Dysarthria usually resolved quickly and by 1 year, and only two patients had mild dysarthria. Reading aloud recovered in line with overall language function.

Frontal, more extensive, impacting ventral stream: F+v

This group of 14 patients had large frontal lesions that extended beyond the frontal lobe to significantly impact ventral stream language regions in the STS and/or angular gyrus (Fig. 2C). Maximal lesion overlap (12 of 14 patients) was in the insula and adjacent frontal operculum. Patterns of recovery were similar in many respects to the F+d group. All 14 patients had aphasia acutely (though two were untestable), and all but one were severe. Recovery was steady, and by 1 year, most patients' aphasia was in the mild to moderate range. Word comprehension impairments were somewhat more persistent, with three patients showing deficits at 1 month, but all were close to ceiling by 1 year. Sentence comprehension, word finding, and grammatical construction recovered in line with overall language function. All patients made phonemic paraphasias, which tended to recover only minimally. AoS was present acutely in seven patients, and in these patients tended to improve
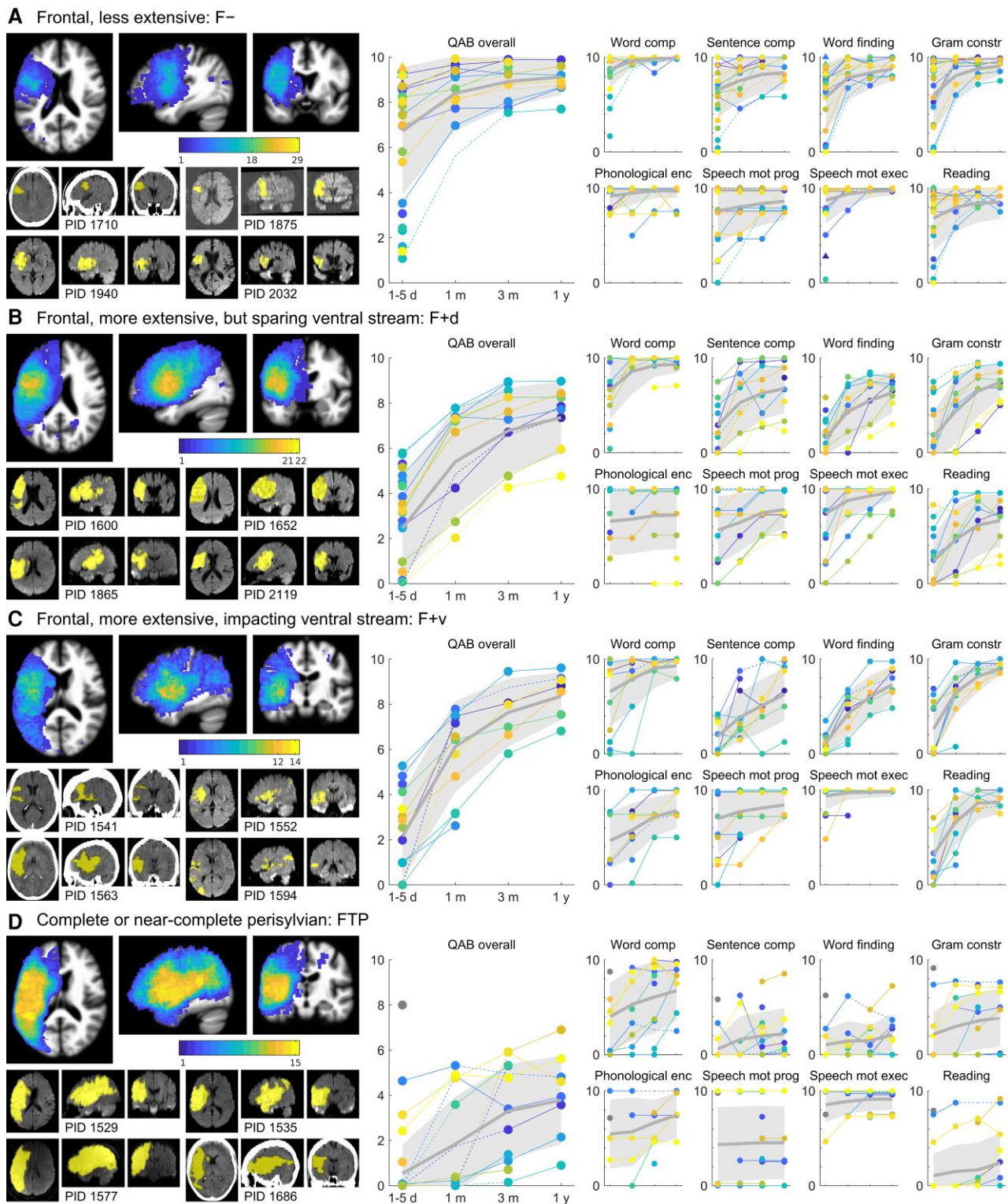
Figure 2 Trajectories of recovery for patients with frontal lobe damage. (**A**) F− group patients (*n* = 29). *Left*: A lesion overlay is shown; the maximum of the colour scale corresponds to the number of patients in the group. Below this, four representative individual clinical scans are shown, with manually delineated lesions indicated in yellow. These four individuals are objectively representative of the group as described in the text. *Right*: Observed QAB scores are presented: first the overall scores, and then subscores for word comprehension ('word comp'), sentence comprehension ('sentence comp'), word finding, grammatical construction ('gram constr'), phonological encoding ('phonological enc'), speech motor programming ('speech mot prog', i.e. the absence of AoS), speech motor execution ('speech mot exec', i.e. the absence of dysarthria), and reading. Circles indicate patients with aphasia at the initial time point, and triangles indicate patients without aphasia at the initial time point, per clinical judgment. Colours are arbitrary. Solid lines are plotted between temporally adjacent observations, and dotted lines between observations where intervening data were imputed. Thick grey lines indicate the group mean, and shaded grey areas indicate one standard deviation each side of the mean, i.e. ~68% of patients in this group would be expected to lie in the shaded area. To enhance readability, scores are randomly slightly jittered for phonological encoding, speech motor programming, and speech motor execution, since these measures were scored on five-point scales. (**B**) F+d group patients (*n* = 22). (**C**) F+v group patients (*n* = 14). (**D**) FTP group patients (*n* = 15). One patient with presumptive right hemisphere dominance for language is shown with a grey marker and does not contribute to the mean or variance estimates.

but did not resolve completely in the patients who were followed. Another four patients were untestable or unscorable; of these, three were followed up at 1 month and had no AoS. Only three patients had dysarthria acutely; the two untestable patients did not have dysarthria when followed up at 1 month. Reading aloud recovered in line with overall language function.

Complete perisylvian: FTP

These 15 patients had extensive lesions that substantially impacted frontal, temporal, and parietal perisylvian cortex (Fig. 2D). Lesion overlap was complete in the frontal operculum and posterior superior temporal gyrus. All 15 patients had aphasia acutely (seven untestable, three also untestable at 1 month), and all were initially severe with the exception of one left-handed patient with complete destruction of the middle cerebral artery territory yet only mild aphasia, which could only be explained by right hemisphere language dominance. Setting aside this right-lateralized patient, recovery was only modest in this group, with all but two patients remaining severe at 1 year. Even word comprehension remained impaired, with five patients showing significant deficits at 3 months or 1 year. Gains in sentence comprehension, word finding, and grammatical construction were modest at best. Most patients made phonemic paraphasias, which only sometimes improved. It should be noted that five patients never produced a scorable speech sample, so were not scored for phonemic paraphasias. Of the seven patients in whom AoS could be assessed acutely, there was a bimodal distribution, with three patients presenting with severe AoS, and four having no AoS at all (including the presumably right-lateralized patient). Of the eight patients where AoS could not be assessed, seven were followed up, and six had AoS at follow-up, varying in severity. There was no improvement of AoS in any patient at any time on the five-point scale we used, but we did observe very modest improvements in most patients that did not cross a score boundary. Dysarthria was generally mild but persistent, or absent. Reading aloud usually remained severely impaired but there were several patients who were relatively spared in their ability to read aloud.

Temporal or temporoparietal, less extensive: TP−

There were 22 patients who had relatively focal lesions within the ventral stream language regions that extend along the length of the STS from the temporal pole to the angular gyrus (Fig. 3A). Lesion overlap was maximal in the deep posterior STS, but only 12 of 22 patients' lesions overlapped there, as expected given that this group had circumscribed lesions distributed over a wide swath of cortex. A total of 19 of the 22 patients had aphasia acutely; one of the patients without aphasia was subsequently shown with functional MRI to have right hemisphere dominance for language, as described previously.[70] Initial severity was usually mild to moderate, and recovery was generally slow but steady, so by 1 year, all but one of the patients scored in the mild or recovered range. Word comprehension was variable initially but always close to ceiling by one month. Sentence comprehension was variable but tended to be a relative weakness. Word finding and grammatical construction recovered in line with overall language function. The majority of patients made phonemic paraphasias. In most patients these were mild, but they were also usually persistent. Neither AoS nor dysarthria were ever present. Reading aloud recovered in line with overall language function.

Temporal or temporoparietal, more extensive: TP+

There were 14 patients with extensive damage to the ventral stream temporoparietal language regions (Fig. 3B). Lesion overlap was complete in the posterior superior temporal gyrus. All patients had severe aphasia on initial evaluation (two were untestable). Recovery was modest, with patients generally still in the moderate to severe range at 1 year. Most patients had initial word comprehension deficits to various degrees, which sometimes recovered quickly but often recovered more slowly. Sentence comprehension remained significantly impaired in all patients throughout the year. Word finding and grammatical construction recovered in line with overall language function. Phonemic paraphasias were present in almost all patients, varied in severity, and showed minimal improvement. Neither AoS nor dysarthria were ever present. Recovery of reading aloud was largely in line with overall language function but was more variable.

Parietal, dorsal stream only: P

There were 18 patients with lesions primarily restricted to the parietal lobe, and largely sparing the ventral stream language region in the angular gyrus (Fig. 3C). Maximal lesion overlap (12 of 18 patients) was in the dorsal part of the angular gyrus, superior to the ventral stream language region. On initial evaluation, 14 of 18 patients presented with aphasia. Initial severity was most often mild to moderate, and improvement was slow and steady, with all patients reaching the mild to recovered range by 1 year. Some patients had word comprehension deficits early, but these resolved quickly in the limited follow-up data that we obtained from those patients. Sentence comprehension was variable acutely, but deficits resolved to mild or better within 1 month in all but one patient. Word finding recovered in line with overall language function. Grammatical construction was a domain of relative strength. About half of these patients made phonemic paraphasias, which tended to improve slowly over time. AoS was never present. A few patients had dysarthria, which improved over time. Reading aloud was an area of relative weakness, with many patients showing persistent deficits.

Ventral temporal: VT

A total of 24 patients had lesions to ventral and medial temporal lobe regions (Fig. 3D). Many of these lesions also involved occipital cortex. Maximal lesion overlap (20 of 24 patients) was observed in the white matter underlying the parahippocampal gyrus. On initial evaluation, 20 patients presented with aphasia. Initial severity varied widely, from mild to severe. Word comprehension was variable acutely, and usually but not always recovered quickly. There were sometimes validity concerns in the assessment of word comprehension in this group, due to frequent visual field and/or object recognition deficits. Sentence comprehension was highly variable initially, but quickly resolved to mild or better in most patients. Word finding recovered in line with overall language function. Grammatical construction was a domain of relative strength. Only 5 of 24 patients made phonemic paraphasias, which were mild with one exception, and recovered completely in three patients. AoS was never observed. A few patients had dysarthria, always associated with damage extending to the corticobulbar tract. Reading aloud was highly variable acutely, and recovered well in some patients, but remained severely impaired throughout the year in others. Visual field deficits sometimes complicated the assessment of reading impairments in these patients.
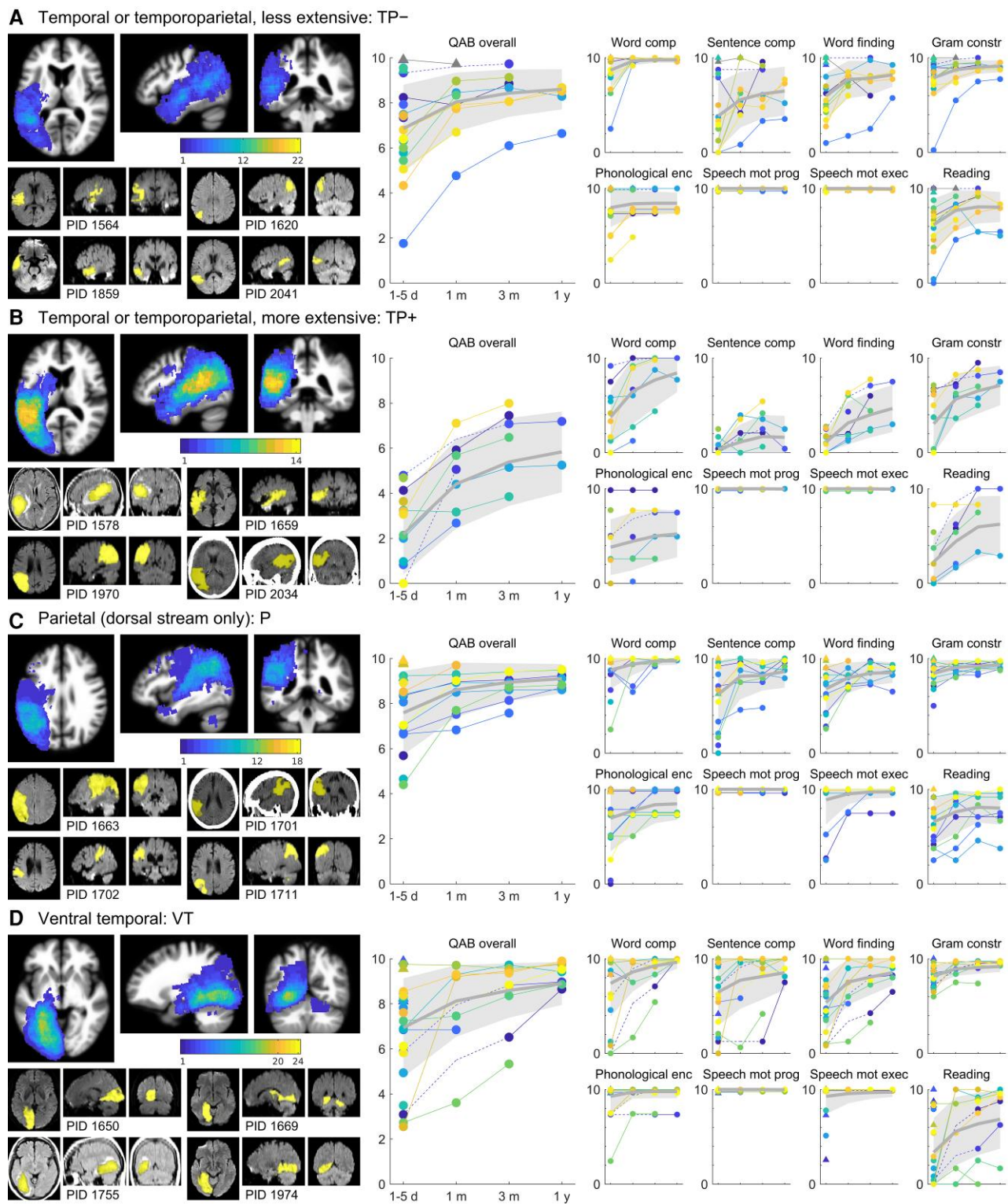
Figure 3 Trajectories of recovery for patients with temporal and/or parietal lobe damage. (**A**) TP− group patients (*n* = 22). One patient with confirmed right hemisphere language, and no aphasia, is shown with a grey marker and does not contribute to the mean or variance estimates. (**B**) TP+ group patients (*n* = 14). (**C**) P group patients (*n* = 18). (**D**) VT group patients (*n* = 24). See Fig. 2 legend for additional details.

Other lesion locations

The remaining five groups (Fig. 4) will be described only briefly. On initial evaluation, aphasia was observed in 34 of 62 basal ganglia patients (55%), 16 of 32 thalamic patients (50%), and 8 of 14 midline patients (57%). In these three groups, initial aphasias were most often mild or moderate, rarely severe, and by 1 year were mild or recovered in all cases. Most speech/language domains recovered in
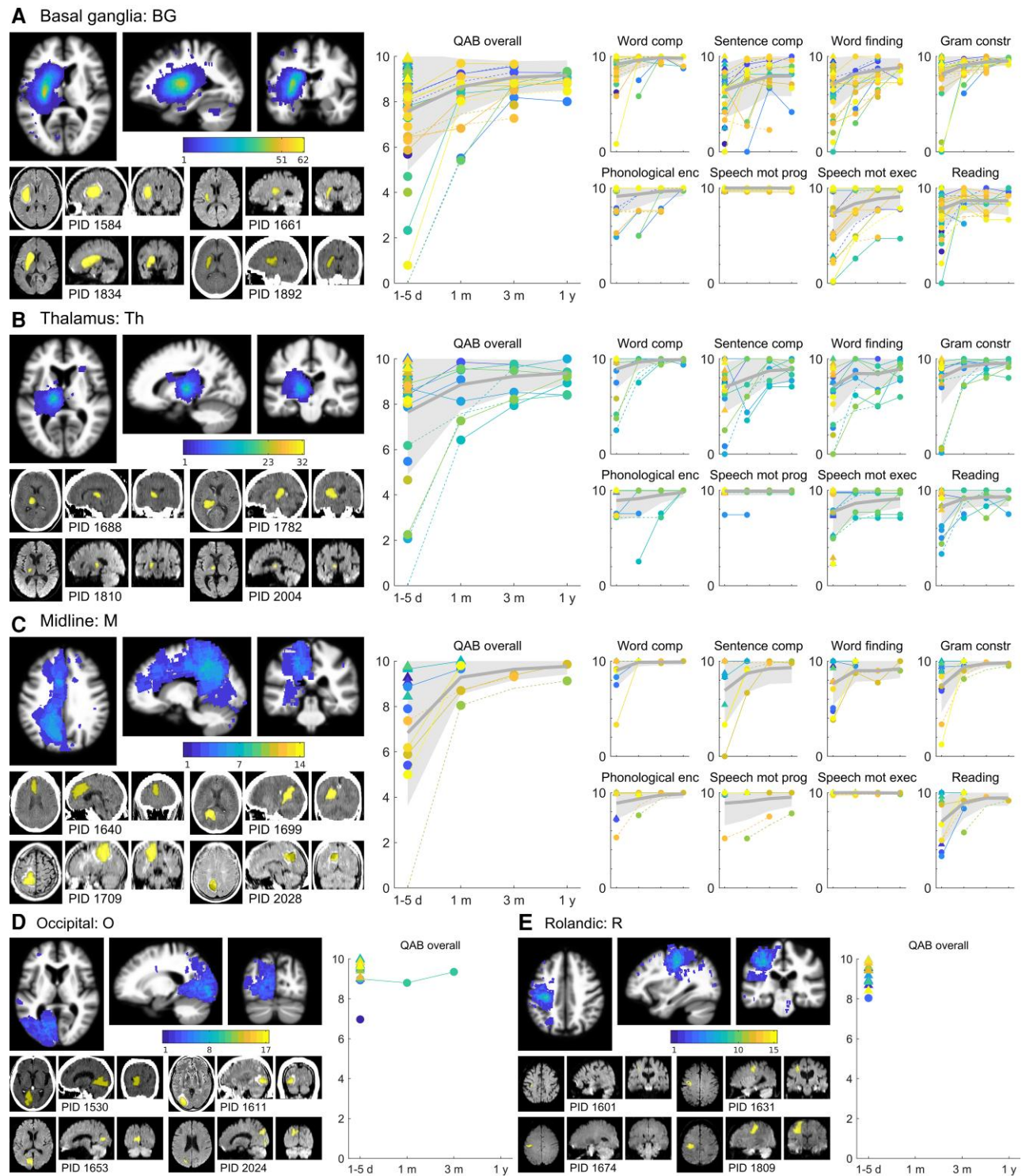
Figure 4 Trajectories of recovery for patients with damage to the basal ganglia, thalamus, midline, occipital, and Rolandic regions. (**A**) BG group patients (*n* = 62). (**B**) Th group patients (*n* = 32). (**C**) M group patients (*n* = 14). (**D**) O group patients (*n* = 17). (**E**) R group patients (*n* = 15). See Fig. 2 legend for additional details.

line with overall language severity, but AoS was never observed in basal ganglia patients, and only noted in one thalamic patient and two midline patients, in all of whom the motor speech deficits were unusual and did not resemble the apraxias encountered with frontal damage. Dysarthria was very common in basal ganglia patients, including in patients without aphasia.

The basal ganglia and thalamic groups were the only groups in which significant within-group associations were observed between lesion extent or location and any speech/language measures. In basal ganglia patients, larger lesions were associated with lower initial scores for overall language function, word comprehension, sentence comprehension, word finding, and reading

aloud, and larger and more posterior lesions were associated with lower final scores for grammatical construction. In thalamic patients, larger lesions were associated with lower initial scores for overall language function, word comprehension, grammatical construction, and phonological encoding, and larger and more posterior lesions were associated with lower final scores for sentence comprehension.

Aphasias were rare after occipital and Rolandic lesions. In the occipital group, 3 of 17 patients were judged to have aphasia, but two of these were borderline cases with very mild language deficits. In the Rolandic group, only 1 of 15 patients had a mild aphasia.

Comparisons between speech/language domains

To explore how performance on specific speech/language domains related to overall language function, we plotted all patients' trajectories, in 2D spaces defined by overall language function (*x*-axes) and each speech/language domain in turn (*y*-axes) (Fig. 5).

This analysis revealed striking differences between speech/language domains in the extent to which they tracked overall language function, or could be selectively spared or impaired. Four patterns were observed among the eight speech/language domains investigated: (i) word finding and grammatical construction generally recovered approximately in line with overall language function; (ii) word comprehension was often spared relative to other domains, and/or recovered much more quickly, but was rarely impaired relative to other language functions, as evidenced by many points in the upper left quadrant, and few in the lower right quadrant, in the word comprehension plot; (iii) sentence comprehension and reading aloud were often in line with overall function, but if not, were more often relatively impaired than relatively spared, as evidenced by the preponderance of observations in the lower right quadrants for these domains; and (iv) phonological encoding, speech motor programming, and speech motor execution were often dissociated in either direction relative to overall language function, as evidenced by the numerous observations in both the top left quadrants and bottom right quadrants for these domains.

Almost all changes were in the direction of improvement. We did not observe any convincing declines in language function. The few numerical declines that were observed were considered more likely to represent day-to-day variability in performance, or measurement error, rather than deterioration of function. Note that the study sample comprised patients with an uncomplicated recovery, because patients who experienced new strokes affecting language function or who developed major health problems were discontinued from the study.

Discussion

We have documented patterns of recovery of overall language function, as well as multiple distinct speech/language domains, in a large and representative cohort of stroke patients, who we divided into groups based on commonly occurring patterns of anatomical damage. It is hoped that the data presented will be of considerable practical value to clinicians. Clinicians who see stroke patients at different stages (e.g. stroke neurologists, inpatient rehabilitation facility staff, outpatient speech-language pathologists) generally have limited information about the whole course of recovery. For example, stroke neurologists in acute care settings often do not have the opportunity to observe how patients progress after they are discharged, since many patients do not return for follow-up visits or follow up at facilities other than the hospital providing acute

care. Because of this, treating clinicians are often not well positioned to confidently discuss the range of likely outcomes with patients, their loved ones, and other health care providers.

While a number of studies have established the dependence of speech and language outcomes on lesion location and extent,[7,9–15,17,18,35,38–40] this information has not been consolidated in a way that would readily facilitate real-life estimates of likely course, including quantification of uncertainty. The data we have presented make it straightforward for clinicians to make inferences about recovery based on empirical data, which can inform communication with individuals with aphasia and their loved ones and can be taken into account when planning provision of services.

The nature and severity of acute aphasia severity was strongly determined by lesion location and extent. Much of the variance in subsequent trajectories of recovery could be explained by initial scores and the effect of time, but lesion location and extent also made a major contribution to predicting the extent of recovery. In certain cases, knowledge of lesion location and extent can lead to very different predictions regarding outcome. For example, for a hypothetical individual with an initial QAB overall score of 2.0, a model without lesion factors provides a point estimate of 6.5 (a moderate aphasia) for the QAB overall score at the 1-year time point. However, the model incorporating lesion information yields different estimates depending on lesion location and extent. Our model predicts, for a patient with an F– lesion of average extent: 8.4 (a mild aphasia); for a patient with a TP+ lesion of average extent: 5.8 (a moderate aphasia); and for a patient with an FTP lesion of average extent: 3.6 (a severe aphasia). With some training in neuroanatomy, it should be feasible for clinicians to determine what group a patient belongs to based on examining their clinical MRI or CT images, and we do not envisage that the time-consuming lesion delineation and normalization procedures that we carried out would be necessary in clinical practice.

Compared to lesion location and extent, other potential explanatory factors made minimal contributions to explaining recovery. Older age and haemorrhagic stroke were modestly associated with lower initial scores, while none of the factors examined (age, sex, handedness, education, and stroke type) were predictive of the extent of recovery. The amount of speech-language therapy received was also not predictive of recovery, though this finding should be interpreted cautiously since this was an observational study, provision of therapy and compliance with therapy may have been related to other factors predictive of recovery, and the nature and quality of therapy varied widely and was not quantified.

Dynamics of recovery

Our data indicate that aphasia is dynamic, multidimensional, and gradated. Speech and language deficits occur on continua; for instance, patients may be able to retrieve frequent words but not infrequent words, or they may readily repeat single words and short sentences while struggling with longer sentences. There is no binary distinction between fluency and non-fluency; the factors that contribute to fluency—word finding, grammatical construction, and speech motor programming[51]—all tend to improve gradually over time, and there is no magical moment when an individual goes from being non-fluent to being fluent. Aphasia subtypes, although of undeniable historical and theoretical importance, have minimal explanatory value in the real-life context of recovery from aphasia after stroke. Rather, each individual presents with a constellation of deficits in different domains, which recover to
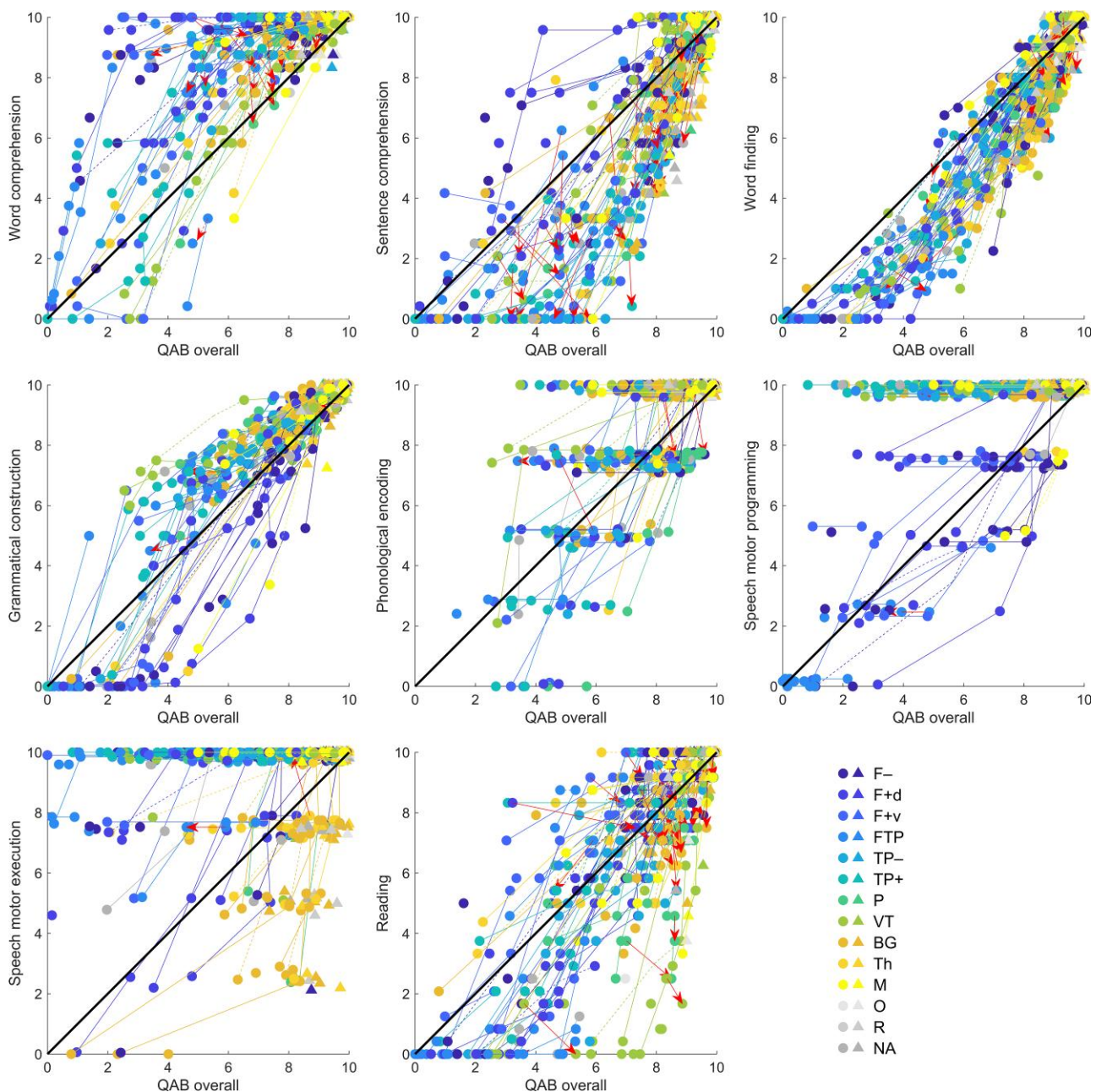
Figure 5 Comparisons between speech/language domains. Trajectories in spaces defined by each speech/language domain in relation to overall language function. Circles indicate patients with acute aphasia and triangles indicate patients without acute aphasia. Solid lines are plotted between temporally adjacent observations, and dotted lines between observations where intervening data were imputed. Most changes were positive; changes where either the QAB overall score or the relevant subscore declined by more than 0.5 points are plotted in red with arrowheads showing the direction of change. To enhance readability, scores are randomly slightly jittered for phonological encoding, speech motor programming, and speech motor execution, since these measures were scored on 5-point scales.

various extents, at different rates. Whether patients thereby cross a line between subtype diagnoses, as defined in any given scheme, is of little practical consequence.

Similar or related points have been made throughout the history of the study of aphasia.[71–75] Indeed, the transient and changeable nature of aphasias, especially in the early post-stroke period, was salient even to the pioneers of aphasia classification, whose case descriptions clearly documented theoretically important symptom-complexes that quickly resolved to more amorphous patterns of deficits.[60,61]

Only a few studies have reported trajectories of recovery in groups of patients defined by neuroanatomy, such as we have done here.[7,11,12,35,36,38–41] In a foundational study, Mohr[35] described the initial presentation and the nature of recovery in patients with damage restricted to Broca's area (see also Mohr *et al.*[36]). Mohr described a wide clinical spectrum acutely, but rapid recovery, such that within a few days to a few months, most of these patients 'pass for normal' in most circumstances. Apraxic deficits were held to be most persistent. Our group of patients with less extensive frontal lesions (F–) was quite similar to Mohr's cohort, as can be

seen by comparing the lesion overlay in our Fig. 2A to Fig. 2 in the study by Mohr.[35] Our data largely confirm Mohr's observations that patients with these lesions recover quickly and nearly completely, though our data suggest that the persistence of aphasic deficits is modestly understated by Mohr and colleagues.[36] The greater sensitivity of our study may reflect its prospective nature and the use of a validated language battery. On the other hand, in line with Mohr's conclusions, our data clearly refute the traditional association of Broca's area with Broca's aphasia, consistent with several other recent studies.[76,77]

Mohr[35] also investigated the lesions of patients who actually did experience persistent Broca's aphasia, and reported that these patients usually had lesions that far exceeded Broca's area and typically involved 'the bulk of the territory of supply of the upper division' of the middle cerebral artery. The temporal lobe was spared. These patients closely resemble our group of patients with extensive frontal lesions sparing the ventral stream (F+d); *cf.* our Fig. 2B to Fig. 3 in the study by Mohr.[35] In contrast to Mohr,[35] we found that these patients generally recovered quite well, with most in the mild to moderate range a year after their stroke. This discrepancy may reflect the way that patients were identified: unlike Mohr's[35] first group, which was defined based on lesion location, his second group was defined based on behavioural profiles. In contrast, we studied a representative sample of patients with damage to the regions in question. Our findings suggest that although most patients with persistent Broca's aphasia will have extensive upper division damage,[78] the converse is not true: many patients with extensive upper division damage recover quite well.

In patients with temporoparietal lesions, lesion extent is also an important factor for prognosis, while the roles of the various brain regions within this area, such as the posterior superior temporal gyrus, supramarginal gyrus, angular gyrus, and middle temporal gyrus have been debated.[9–12] We found that patients with less extensive temporal or temporoparietal lesions (TP−) recovered much better than those with more extensive temporal or temporoparietal lesions (TP+). Our data suggest that the cortical regions in the vicinity of the STS are critical, since patients with parietal lesions and those with ventral temporal lesions (VT) recovered more quickly and completely. It is clear that the posterior temporoparietal language region, which is the point of origin of both the dorsal and the ventral streams,[60–63] is the most important language region of the brain, giving rise to more severe, persistent aphasias when extensively damaged in the TP+ and FTP groups, consistent with prior observations.[43,44,73]

Mechanisms of recovery

The substantial extent of recovery that we observed in most individuals clearly supports the concept of neural plasticity, that is, functional reorganization of surviving brain regions to support new or expanded roles in speech and language function. Numerous functional imaging studies have investigated the nature of this process.[79,80] Many different claims have been made, with little consistency in the literature, due in large part to methodological challenges.[80] However, a broad outline of the recovery process has started to emerge. There is little evidence for dramatic shifts of language function to the right hemisphere; most individuals with aphasia continue to process language predominantly in the left hemisphere.[65,80,81] Moreover, there is scant evidence for recruitment of new left hemisphere regions, and minimal evidence for increased dependence on potentially compensatory systems such as the multiple demand network.[80]

Rather, recent work has converged on a central concept that aphasia is a network disorder.[64,79,82–84] In this view, recovery depends primarily on the surviving nodes of a large-scale network of temporal, frontal, and parietal regions,[85] which are predominantly left-lateralized. A 'weak shadow' of homotopic regions in the right hemisphere may also be considered part of the language network.[86]

The network concept of aphasia readily explains the most fundamental and salient aspect of our data, which is the relatively good recovery that was observed in almost all groups of patients, despite core language regions being damaged in many groups. This potential for recovery follows from the fact that all but the largest lesions leave enough of the large scale network intact to provide a substrate for recovery. However, our findings also serve to refine and qualify this emerging model in three important ways.

First, our data support a middle ground between the idea that the language network is functionally homogeneous[87,88] and the functional segregation that is implied by most historical[60,61] and contemporary[62,67,89–91] models. The evidence that the network is not functionally homogeneous is simply that distinct patterns of deficits were observed with damage to different parts of the network. On the other hand, if the language network were really a mosaic of distinct language regions and pathways with specific functions, as implied by most models, then even small lesions should result in persistent and specific deficits. The fact that this was generally not the case suggests that there is considerable redundancy within the network. While each speech or language function may have canonical neural substrates, when these are damaged, in many cases, alternative regions and pathways have the capacity to support the function, in whole or in part.[92,93] This process is not instantaneous, but appears to require a 'retuning' of surviving regions, which takes place on the time scale of weeks to months.

Second, our findings indicate that speech and language domains differ dramatically in the extent to which their neural substrates are distributed. At one extreme, word comprehension is the most distributed and redundantly represented function. Acutely, performance was highly variable and significant deficits were common, even in patients with damage to regions other than Wernicke's area, regardless of how broadly that is defined. But most patients performed close to ceiling by one month, with significant deficits persisting only in a minority of patients, mostly with complete perisylvian (FTP) or extensive temporoparietal (TP+) lesions. This is broadly consistent with a prior study that showed good recovery of word comprehension deficits between 1 and 6 months in most but not all patients,[10] and is consistent with the observation that word comprehension is generally a relative strength in chronic aphasia.[94,95] The relative preservation of word comprehension in aphasia has sometimes been interpreted in support of the view that the right hemisphere has an independent capacity not only for early stages of speech perception (i.e. phonemic discrimination), but also for mapping these phonological representations onto the lexicon (i.e. auditory word comprehension).[62,95,96] We concur that the 'weak shadow' language network nodes in the right hemisphere are one of the possible substrates that can contribute to supporting word comprehension. But the capacity of any one region, or of the right hemisphere in general, should not be overestimated. If the right hemisphere truly had an independent capacity for mapping phonological representations onto the lexicon, then word comprehension deficits would not be observed acutely, and would never persist as they do in some patients, because an independent right hemisphere would be capable of

comprehending single words regardless of which left hemisphere regions are damaged (for converging evidence, see also Gazzaniga[97] and Risse *et al.*[98]).

At the other extreme, sentence comprehension is a language domain that is strongly dependent on specific neural substrates, specifically, left posterior temporoparietal cortex. In the face of extensive damage to this region, as seen in the TP+ and FTP groups, sentence comprehension deficits were ubiquitous, severe, and persistent. This posterior perisylvian localization of sentence comprehension is consistent with some previous findings,[9,99,100] but contrasts with many other claims that have been made (for review, see Matchin and Hickok[91] and Wilson[101]). This is not to imply that other brain regions, such as the IFG, play no role in sentence comprehension. But only the posterior temporoparietal region is truly indispensable.

Another domain with specialized neural substrates is speech motor programming, in which deficits manifest as AoS. In one sense, this function is even more tightly localized than sentence comprehension, because AoS was almost never seen with lesions outside of left frontal cortex: we did not observe a single individual with a lesion confined to temporal and/or parietal cortex who presented with AoS (even though many of these patients had phonological encoding deficits). AoS often resolved well in patients with less extensive frontal lesions (F−), presumably reflecting distributed substrates within left frontal regions, while AoS recovered only modestly in patients with larger frontal lesions (F+d, F+v), and showed almost no improvement in patients with the most extensive lesions (FTP). However, in another sense, speech motor programming was less tightly localized than sentence comprehension, because unlike sentence comprehension, which followed from temporoparietal damage inexorably and without exception, there were many patients with extensive frontal damage yet no AoS.

This paradoxical observation implies a third qualification of the network concept of aphasia, which is that there must be considerable variability between individuals in the extent to which various regions have premorbid capacity to support different speech/language functions. In the case of speech motor programming, the fact that a significant minority of individuals exhibit no AoS immediately after complete destruction of all plausible left frontal substrates indicates that other regions can support speech motor programming premorbidly. Most plausibly, this may be an instance in which the right hemisphere has an independent premorbid capacity in some individuals.[102] To give another example, some patients with extensive left ventral temporal damage experienced severe and persistent alexia, while others recovered well. This is likely to reflect individual differences in the capacity of other regions to compensate. Another possible explanation for such interindividual variability is that there were critical differences in lesion locations between patients (within the same lesion group), but this explanation is less likely because we did not generally observe any significant within-group relationships between lesion location or extent and any speech/language measures, except in the case of subcortical lesions.

Two studies have provided functional imaging evidence that individual differences in right hemisphere temporal lobe activation are associated with individual differences in language outcomes after stroke.[103,104] However, we envisage that individual differences in the capacity of various regions to support particular speech and language functions is a property of left hemisphere language regions as well. Indeed, the premorbid capacity of left hemisphere regions to support a wide range of speech/language functions, besides their most canonical functions, may be responsible for

much of the residual variance of recovery in all but the most severe aphasias, given the strong lateralization of the language network and the fact that language processing usually remains in the left hemisphere in people with aphasia.

Limitations

Our study had several notable limitations. First, although we recruited a large number of individuals acutely, because we divided patients in groups based on lesion location and extent, there were ultimately only one to two dozen patients in most of these groups. Moreover, there were many missing data points. Longitudinal data were acquired for only 121 (56%) of the individuals with aphasia. This was inevitable when attempting to study a representative cohort of individuals with acute post-stroke aphasia over time. Indeed, our cohort was highly representative: 88% of eligible individuals who were approached consented to take part in the study, and the patients who were followed longitudinally did not differ in severity from those who were not. We were able to use several different imputation strategies in different analysis contexts to infer general patterns in the face of missing data. However, future studies with larger samples will be required to corroborate our findings.

Second, lesions were identified based only on acute clinical imaging. The majority of patients had MRI scans available, on which lesions were clearly delineated, but restricted diffusion is not always predictive of final infarct.[105,106] Moreover, some patients had only CT scans, on which lesion boundaries were not always easy to discern. These limitations entail that reconstructed lesions may not reflect the exact location and extent of irreversible tissue damage in all patients.

Third, the QAB[56] is a relatively brief battery and does not probe every aspect of language function. In particular, we did not quantify written word comprehension or writing acutely (these functions were characterized at follow-up in most patients, but those data are not included in the present article), and we did not quantify cognitive or executive deficits, which may account for some aspects of performance and may have explanatory value in understanding recovery.[107] Moreover, deficits in the domains examined were not characterized to the depth that would be possible with a more comprehensive battery of tests. However, the brevity of the QAB was a necessary compromise to acquire acute data at the bedside and to retain patients in the study over time.

Conclusion

The longitudinal data presented in this study will enable clinicians to make realistic predictions about recovery from aphasia, its dependence on lesion location and extent, and its variability. This information will be important for preparing patients and their loved ones for the road ahead and planning rehabilitation services. In future work, we plan to investigate the recovery of each specific speech/language domain in more detail. We are also using functional MRI to study potential reorganization of language networks in a subset of the individuals described in this study.

Acknowledgements

Funding

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Culton GL. Spontaneous recovery from aphasia. *J Speech Hear Res*. 1969;12:825–832.
2. Basso A, Capitani E, Vignolo LA. Influence of rehabilitation on language skills in aphasic patients: a controlled study. *Arch Neurol*. 1979;36:190–196.
3. Sarno MT, Levita E. Recovery in treated aphasia in the first year post-stroke. *Stroke*. 1979;10:663–670.
4. Kertesz A, McCabe P. Recovery patterns and prognosis in aphasia. *Brain*. 1977;100:1–18.
5. Basso A. Prognostic factors in aphasia. *Aphasiology*. 1992;6:337–348.
6. REhabilitation and recovery of peopLE with Aphasia after StrokE (RELEASE) Collaborators. Predictors of poststroke aphasia recovery: A systematic review-informed individual participant data meta-analysis. *Stroke*. 2021;52:1778–1787.
7. Kertesz A, Harlock W, Coates R. Computer tomographic localization, lesion size, and prognosis in aphasia and nonverbal impairment. *Brain Lang*. 1979;8:34–50.
8. Mazzocchi F, Vignolo LA. Localisation of lesions in aphasia: clinical-CT scan correlations in stroke patients. *Cortex*. 1979;15:627–653.
9. Selnes OA, Knopman DS, Niccum N, Rubens AB, Larson D. Computed tomographic scan correlates of auditory comprehension deficits in aphasia: a prospective recovery study. *Ann Neurol*. 1983;13:558–566.
10. Selnes OA, Niccum N, Knopman DS, Rubens AB. Recovery of single word comprehension: CT-scan correlates. *Brain Lang*. 1984;21:72–84.
11. Naeser MA, Helm-Estabrooks N, Haas G, Auerbach S, Srinivasan M. Relationship between lesion extent in 'Wernicke's area' on computed tomographic scan and predicting recovery of comprehension in Wernicke's aphasia. *Arch Neurol*. 1987;44:73–82.
12. Kertesz A, Lau WK, Polk M. The structural determinants of recovery in Wernicke's aphasia. *Brain Lang*. 1993;44:153–164.
13. Goldenberg G, Spatt J. Influence of size and site of cerebral lesions on spontaneous recovery of aphasia and on success of language therapy. *Brain Lang*. 1994;47:684–698.
14. Hope TMH, Seghier ML, Leff AP, Price CJ. Predicting outcome and recovery after stroke with lesions extracted from MRI images. *Neuroimage Clin*. 2013;2:424–433.
15. 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.
16. Hillis AE, Beh YY, Sebastian R, *et al*. Predicting recovery in acute poststroke aphasia. *Ann Neurol*. 2018;83:612–622.
17. Benghanem S, Rosso C, Arbizu C, *et al*. Aphasia outcome: the interactions between initial severity, lesion size and location. *J Neurol*. 2019;266:1303–1309.
18. Nakagawa Y, Sano Y, Funayama M, Kato M. Prognostic factors for long-term improvement from stroke-related aphasia with adequate linguistic rehabilitation. *Neurol Sci*. 2019;40:2141–2146.
19. Watila MM, Balarabe SA. Factors predicting post-stroke aphasia recovery. *J Neurol Sci*. 2015;352:12–18.
20. Porch BE, Collins M, Wertz RT, Friden TP. Statistical prediction of change in aphasia. *J Speech Hear Res*. 1980;23:312–321.
21. Pedersen PM, Jørgensen HS, Nakayama H, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants, and recovery. *Ann Neurol*. 1995;38:659–666.
22. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc Dis*. 2004;17:35–43.
23. Swinburn K, Porter G, Howard D. *Comprehensive aphasia test*. Psychology Press; 2004.
24. Lomas J, Kertesz A. Patterns of spontaneous recovery in aphasic groups: a study of adult stroke patients. *Brain Lang*. 1978;5:388–401.
25. Demeurisse G, Demol O, Derouck M, de Beuckelaer R, Coekaerts M, Capon A. Quantitative study of the rate of recovery from aphasia due to ischemic stroke. *Stroke*. 1980;11:455–458.
26. Pashek G, Holland A. Evolution of aphasia in the first year postonset. *Cortex*. 1988;24:411–423.
27. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med*. 2001;249:413–422.
28. Bakheit AMO, Shaw S, Carrington S, Griffiths S. The rate and extent of improvement with therapy from the different types of aphasia in the first year after stroke. *Clin Rehabil*. 2007;21:941–949.
29. Holland A, Fridriksson J. Aphasia management during the early phases of recovery following stroke. *Am J Speech Lang Pathol*. 2001;10:19–28.
30. Wilson SM, Eriksson DK, Brandt TH, *et al*. Patterns of recovery from aphasia in the first 2 weeks after stroke. *J Speech Lang Hear Res*. 2019;62:723–732.
31. Wertz R, Robinson A, Deal J. Classifying the aphasias: a comparison of the Boston diagnostic aphasia examination and the western aphasia battery. *Clin Aphasiology Proc Conf*. 1984;14:40–47.
32. Yang ZH, Zhao XQ, Wang CX, Chen HY, Zhang YM. Neuroanatomic correlation of the post-stroke aphasias studied with imaging. *Neurol Res*. 2008;30:356–360.
33. Kümmerer D, Hartwigsen G, Kellmeyer P, *et al*. Damage to ventral and dorsal language pathways in acute aphasia. *Brain*. 2013;136:619–629.
34. Hillis AE, Wityk RJ, Barker PB, *et al*. Subcortical aphasia and neglect in acute stroke: the role of cortical hypoperfusion. *Brain*. 2002;125:1094–1104.
35. Mohr JP. Broca's area and Broca's aphasia. In: Whitaker H, Whitaker HA, eds. *Studies in neurolinguistics*. Vol 1. Academic Press; 1976:201–236.
36. Mohr JP, Pessin MS, Finkelstein S, Funkenstein HH, Duncan GW, Davis KR. Broca aphasia: pathologic and clinical. *Neurology*. 1978;28:311–324.

37. Heiss WD, Kessler J, Thiel A, Ghaemi M, Karbe H. Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. *Ann Neurol.* 1999;45:430–438.

38. Liang CL, Chang HW, Lu K, *et al.* Early prediction of aphasia outcome in left basal ganglia hemorrhage: early prediction of aphasia outcome. *Acta Neurol Scand.* 2001;103:148–152.

39. Flamand-Roze C, Cauquil-Michon C, Roze E, *et al.* Aphasia in border-zone infarcts has a specific initial pattern and good long-term prognosis: aphasia associated with watershed infarcts. *Eur J Neurol.* 2011;18:1397–1401.

40. Komiya K, Sakai Y, Horikoshi T, Naganuma H. Recovery process and prognosis of aphasic patients with left putaminal hemorrhage: relationship between hematoma type and language modalities. *J Stroke Cerebrovasc Dis.* 2013;22:132–142.

41. Robson H, Griffiths TD, Grube M, Woollams AM. Auditory, phonological, and semantic factors in the recovery from Wernicke's aphasia poststroke: predictive value and implications for rehabilitation. *Neurorehabil Neural Repair.* 2019;33:800–812.

42. Stockert A, Wawrzyniak M, Klingbeil J, *et al.* Dynamics of language reorganization after left temporo-parietal and frontal stroke. *Brain.* 2020;143:844–861.

43. Penfield W, Roberts L. *Speech and brain-mechanisms.* Princeton University Press; 1959.

44. Metter EJ, Hanson WR, Jackson CA, *et al.* Temporoparietal cortex in aphasia: evidence from positron emission tomography. *Arch Neurol.* 1990;47:1235–1238.

45. Lincoln NB, McGuirk E, Mulley GP, Lendrem W, Jones AC, Mitchell JR. Effectiveness of speech therapy for aphasic stroke patients. A randomised controlled trial. *Lancet.* 1984;323:1197–1200.

46. Kim KA, Lee JS, Chang WH, *et al.* Changes in language function and recovery-related prognostic factors in first-ever left hemispheric ischemic stroke. *Ann Rehabil Med.* 2019;43:625–634.

47. Brust JC, Shafer SQ, Richter RW, Bruun B. Aphasia in acute stroke. *Stroke.* 1976;7:167–174.

48. Wade DT, Hewer RL, David RM, Enderby PM. Aphasia after stroke: natural history and associated deficits. *J Neurol Neurosurg Psychiatry.* 1986;49:11–16.

49. Turkeltaub PE. A taxonomy of brain-behavior relationships after stroke. *J Speech Lang Hear Res.* 2019;62:3907–3922.

50. Kertesz A. *Western aphasia battery-revised.* Grune and Stratton; 2006.

51. Casilio M, Rising K, Beeson PM, Bunton K, Wilson SM. Auditory-perceptual rating of connected speech in aphasia. *Am J Speech Lang Pathol.* 2019;28:550–568.

52. Prins RS, Snow CE, Wagenaar E. Recovery from aphasia: spontaneous speech versus language comprehension. *Brain Lang.* 1978;6:192–211.

53. Basso A, Capitani E, Zanobio ME. Pattern of recovery of oral and written expression and comprehension in aphasic patients. *Behav Brain Res.* 1982;6:115–128.

54. Nicholas M, Helm-Estabrooks N, Ward-Lonergan J, Morgan A. Evolution of severe aphasia in the first two years post onset. *Arch Phys Med Rehabil.* 1993;74:830–836.

55. Hybbinette H, Schalling E, Plantin J, *et al.* Recovery of apraxia of speech and aphasia in patients with hand motor impairment after stroke. *Front Neurol.* 2021;12:634065.

56. Wilson SM, Eriksson DK, Schneck SM, Lucanie JM. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One.* 2018;13:e0192773.

57. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6:284–290.

58. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage.* 2005;26:839–851.

59. Nachev P, Coulthard E, Jäger HR, Kennard C, Husain M. Enantiomorphic normalization of focally lesioned brains. *Neuroimage.* 2008;39:1215–1226.

60. Wernicke C. *Der aphasische symptomencomplex.* Cohn and Weigert; 1874.

61. Lichtheim L. On aphasia. *Brain.* 1885;7:433–484.

62. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci.* 2007;8:393–402.

63. Saur D, Kreher BW, Schnell S, *et al.* Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A.* 2008;105:18035–18040.

64. Fridriksson J, Yourganov G, Bonilha L, Basilakos A, Den Ouden DB, Rorden C. Revealing the dual streams of speech processing. *Proc Natl Acad Sci U S A.* 2016;113:15108–15113.

65. Wilson SM, Yen M, Eriksson DK. An adaptive semantic matching paradigm for reliable and valid language mapping in individuals with aphasia. *Hum Brain Mapp.* 2018;39:3285–3307.

66. Yen M, DeMarco AT, Wilson SM. Adaptive paradigms for mapping phonological regions in individual participants. *Neuroimage.* 2019;189:368–379.

67. Wilson SM, Bautista A, McCarron A. Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage.* 2018;171:62–74.

68. Magee L. R2 measures based on Wald and likelihood ratio joint significance tests. *Am Stat.* 1990;44:250–253.

69. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57:289–300.

70. Schneck SM, Entrup JL, Duff MC, Wilson SM. Unexpected absence of aphasia following left temporal hemorrhage: a case study with functional neuroimaging to characterize the nature of atypical language localization. *Neurocase.* 2021;27:97–105.

71. Jackson JH. On the nature of the duality of the brain. *Med Press Circ.* 1874;1:19–25, 41–49, 63–70.

72. Freud S. *On aphasia: a critical study.* Imago; 1891.

73. Marie P. La troisième circonvolution frontal gauche ne joue aucun rôle spécial dans la fonction du langage. *Sem Médicale.* 1906;26:241–247.

74. Goldstein K. *Language and language disturbances.* Grune and Stratton; 1948.

75. Bates E, Saygin AP, Moineau S, Marangolo P, Pizzamiglio L. Analyzing aphasia data in a multidimensional symptom space. *Brain Lang.* 2005;92:106–116.

76. Fridriksson J, Guo D, Fillmore P, Holland A, Rorden C. Damage to the anterior arcuate fasciculus predicts non-fluent speech production in aphasia. *Brain.* 2013;136:3451–3460.

77. Gajardo-Vidal A, Lorca-Puls DL; Team PLORAS, *et al.* Damage to Broca's area does not contribute to long-term speech production outcome after stroke. *Brain.* 2021;144:817–832.

78. Fridriksson J, Fillmore P, Guo D, Rorden C. Chronic Broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cereb Cortex.* 2015;25:4689–4696.

79. Hartwigsen G, Saur D. Neuroimaging of stroke recovery from aphasia—insights into plasticity of the human language network. *Neuroimage.* 2019;190:14–31.

80. Wilson SM, Schneck SM. Neuroplasticity in post-stroke aphasia: a systematic review and meta-analysis of functional imaging studies of reorganization of language processing. *Neurobiol Lang.* 2021;2:22–82.

81. Saur D, Lange R, Baumgaertner A, *et al.* Dynamics of language reorganization after stroke. *Brain.* 2006;129:1371–1384.

82. Corbetta M, Ramsey L, Callejas A, *et al.* Common behavioral clusters and subcortical anatomy in stroke. *Neuron.* 2015;85:927–941.

83. Geranmayeh F, Leech R, Wise RJS. Network dysfunction predicts speech production after left hemisphere stroke. *Neurology*. 2016;86:1296–1305.

84. Siegel JS, Ramsey LE, Snyder AZ, *et al*. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proc Natl Acad Sci U S A*. 2016;113: E4367-E4376.

85. Siegel JS, Seitzman BA, Ramsey LE, *et al*. Re-emergence of modular brain networks in stroke recovery. *Cortex*. 2018;101: 44–59.

86. Martin KC, Seydell-Greenwald A, Berl MM, *et al*. A weak shadow of early life language processing persists in the right hemisphere of the mature brain. *Neurobiol Lang*. 2022;3:364–385.

87. Dick F, Bates E, Wulfeck B, Utman JA, Dronkers N, Gernsbacher MA. Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychol Rev*. 2001; 108:759–788.

88. Blank I, Balewski Z, Mahowald K, Fedorenko E. Syntactic processing is distributed across the language system. *Neuroimage*. 2016;127:307–323.

89. Ueno T, Saito S, Rogers TT, Lambon Ralph MA. Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*. 2011;72:385–396.

90. Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. 2012;62:816–847.

91. Matchin W, Hickok G. The cortical organization of syntax. *Cereb Cortex*. 2020;30:1481–1498.

92. Price CJ, Friston KJ. Degeneracy and cognitive anatomy. *Trends Cogn Sci*. 2002;6:416–421.

93. Seghier ML, Price CJ. Interpreting and utilising intersubject variability in brain function. *Trends Cogn Sci*. 2018;22:517–530.

94. Goodglass H. *Understanding aphasia*. Academic Press; 1993.

95. Rogalsky C, Basilakos A, Rorden C, *et al*. The neuroanatomy of speech processing: a large-scale lesion study. *J Cogn Neurosci* 2022;34:1355–1375.

96. Rogalsky C, Pitz E, Hillis AE, Hickok G. Auditory word comprehension impairment in acute stroke: relative contribution of phonemic versus semantic factors. *Brain Lang*. 2008;107:167–169.

97. Gazzaniga MS. Right hemisphere language following brain bisection: a 20-year perspective. *Am Psychol*. 1983;38:525–537.

98. Risse GL, Gates JR, Fangman MC. A reconsideration of bilateral language representation based on the intracarotid amobarbital procedure. *Brain Cogn*. 1997;33:118–132.

99. Thothathiri M, Kimberg DY, Schwartz MF. The neural basis of reversible sentence comprehension: evidence from voxel-based lesion symptom mapping in aphasia. *J Cogn Neurosci*. 2012;24:212–222.

100. Pillay SB, Binder JR, Humphries C, Gross WL, Book DS. Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*. 2017;88:970–975.

101. Wilson SM. Lesion-symptom mapping in the study of spoken language understanding. *Lang Cogn Neurosci*. 2017;32:891–899.

102. Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B. Sensory-motor transformations for speech occur bilaterally. *Nature*. 2014:507:94–98.

103. Crinion J, Price CJ. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*. 2005;128:2858–2871.

104. Tyler LK, Marslen-Wilson WD, Randall B, *et al*. Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain*. 2011;134: 415–431.

105. Kidwell CS, Saver JL, Mattiello J, *et al*. Thrombolytic reversal of acute human cerebral ischemic injury shown by diffusion/perfusion magnetic resonance imaging. *Ann Neurol*. 2000;47:462–469.

106. Fiehler J, Knudsen K, Kucinski T, *et al*. Predictors of apparent diffusion coefficient normalization in stroke patients. *Stroke*. 2004;35:514–519.

107. Schumacher R, Halai AD, Lambon Ralph MA. Assessing and mapping language, attention and executive multidimensional deficits in stroke aphasia. *Brain*. 2019;142:3202–3216.