

Predicting Language Recovery in Post-Stroke Aphasia using Behavior and Functional MRI

Michael Iorga (✉ michael.iorga@northwestern.edu)

Northwestern University Feinberg School of Medicine <https://orcid.org/0000-0002-9885-6602>

James Higgins

Northwestern University Feinberg School of Medicine

David Caplan

Massachusetts General Hospital

Richard Zinbarg

Northwestern University Department of Psychology

Swathi Kiran

Boston University College of Health and Rehabilitation Sciences Sargent College

Cynthia K. Thompson

Northwestern University School of Communication

Brenda Rapp

Johns Hopkins University Zanvyl Krieger School of Arts and Sciences

Todd B. Parrish

Northwestern University Feinberg School of Medicine

Research

Keywords: Aphasia, Language Therapy, rsfMRI, fALFF, Group ICA, Recovery Prediction

Posted Date: September 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-75485/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Language outcomes after speech and language therapy in post-stroke aphasia are challenging to predict. This study examines behavioral language measures and resting state fMRI (rsfMRI) as prognostics for response to language therapy.

Methods: Seventy patients with chronic aphasia were recruited and treated for one of three deficits: anomia, agrammatism, or dysgraphia. Treatment effect was measured by performance on a treatment-specific language measure, assessed before and after three months of language therapy. Each patient also underwent an additional 27 language assessments and an fMRI scan at baseline. Patient scans were decomposed into 20 components by group independent component analysis, and each component time series was summarized by its fractional amplitude of low-frequency fluctuations (fALFF).

Results: Treatment effects were modelled with elastic net regression, using clinical language measures and fALFF imaging predictors independently. Correlation analyses showed high performance for language measures in anomia ($r = 0.958$, $n = 30$) and for fALFF predictors in agrammatism ($r = 0.940$, $n = 11$) and dysgraphia ($r = 0.925$, $n = 18$). These models are state-of-the-art for aphasia recovery prediction.

Conclusion: Predicting aphasia recovery with rsfMRI features may outperform predictions from clinical language measures in some patient populations. This suggests rsfMRI may have prognostic value for chronic aphasia patients undergoing language therapy. Differentiating patients who respond to therapy from those who do not is a first step towards personalized treatment in post-stroke aphasia.

1. Background

Post-stroke aphasia is an impairment in language communication or understanding that affects one third of stroke survivors (1–3). Aphasia is managed with speech and language therapy (SLT), which addresses patient-specific language deficits through targeted training in order to improve functional communication (4, 5). Although SLT is effective overall, a majority of patients will continue to experience chronic aphasia both acutely and chronically (6, 7). Aphasia significantly lowers functional independence and health-related quality of life, necessitating an improved approach to aphasia management (2, 8, 9). While numerous alternatives to SLT and modifications of SLT have been tested, there is currently not enough evidence to recommend one form over another (4). Optimization of SLT paradigms is limited by high variability in patient response - some patients fully recover while others experience little benefit (10–13). By better understanding the patient-level factors which drive response variability, it may be possible to suggest optimal SLT parameters for each patient, and this personalized approach to SLT may ultimately lead to a more robust treatment approach (14).

Several diagnostic and clinical variables have been associated with the treatment response: particularly the aphasia type, aphasia severity, stroke lesion location, and stroke lesion volume (9, 15). As a result, several attempts have been made to predict individual patient outcomes using these variables. The SPEAK model combined several behavioral and clinical features to predict performance on the Aphasia Severity Rating Scale in 131 patients at 1-year after stroke ($R^2 = 0.56$) (16). A combination of baseline lesion variables and initial aphasia severity has been applied to predict sub-acute recovery in 20 patients ($R^2 = 0.73$) (17). In order to augment the predictive power of lesion variables, Halai et. al. investigate the overlap of patient lesions with known core language areas to build a model of 21 behavioral measures of aphasia severity in 70 patients (mean $R^2 = 0.48$) (18). This lesion-based approach to predicting multidimensional aphasia outcomes is supported by earlier work showing that a profile of aphasia type and severity can be inferred from lesion location and extent (19, 20). However, there remains considerable variance in patient outcomes that is incompletely explained by anatomical and behavioral variables alone (21, 22).

Although functional neuroimaging has been used extensively to study aphasia, there is limited understanding of how including functional neuroimaging variables in models of treatment response affects performance. Interpreting neuropsychological measures alongside data-driven and/or multi-modal neuroimaging features has yielded effective baseline and longitudinal models of stroke aphasia severity (23–26). Resting-state functional MRI (rsfMRI) has specifically demonstrated potential as an assessment tool, as patients with aphasia have observable differences in their resting functional connectivity and resting networks as compared to healthy controls (27–29). Furthermore, the initial aphasia severity profile can be inferred from resting network activity, and changes in global network activity tracks the extent of language recovery (30–33). These findings suggest that variables observed through functional neuroimaging modalities may be complementary to anatomical and behavioral variables, and further study of predictive models include functional variables may yield superior results (18).

In this study, we build predictive models of individual patient responses to SLT across three distinct aphasia treatments using behavioral and rsfMRI features. We then investigate the relative contribution of specific behavioral and rsfMRI features to each model. Our first goal is to improve the state-of-the-art in modelling therapeutic response, as a step towards a clinically viable optimization model for SLT. Second, while many aphasia assessments have been developed, their collective prognostic utility remains unknown. By identifying specific assessments which are useful to each aphasia type, we seek to optimize the prognostic assessment battery and decrease testing burden on patients and practitioners. Lastly, the utility of rsfMRI in the prognosis of aphasia is understudied, and successful rsfMRI-based models may motivate further study of this imaging modality as a clinical assessment tool.

2. Methods

2.1 Recruitment and Assessment

Patients with chronic aphasia were recruited from the local metropolitan area of three sites: Boston University (BU), Northwestern University (NU), and Johns Hopkins University (JHU). All patients presented with aphasia resulting from a single left-hemisphere thromboembolic or hemorrhagic stroke, were at least one-year post-stroke, and had no other impairments that impacted the ability to complete the behavioral or neural tasks (e.g. vision and hearing was within normal limits). All were monolingual English-speaking, had at least a high school education, and completed a written consent form approved by every site's

Institutional Review Board (IRB). Patients were independently recruited, diagnosed and treated for one aphasia type at each site: anomia at BU (N = 30), agrammatism at NU (N = 16), and dysgraphia at JHU (N = 24).

Aphasia is primarily diagnosed and assessed through language assessments, such as the Western Aphasia Battery (WAB) (34). However, the WAB lacks sensitivity to lexical-semantic, sentence processing, and spelling deficits, and supplementary language measures must also be performed to capture the full range of aphasic deficits (35–37). Although many language measures have been developed for this purpose, relatively few have been assessed for psychometric validity (38–41). As a result, there is a lack of consensus on the optimal aphasia assessment battery, and the prognostic properties of existing measures are largely unknown. In order to better understand the relative utility of aphasia assessments, we collected a broad range of language and cognitive measures (see Table 1). Each patient was also assessed on one of three treatment-specific measures (TSM), which served as the primary metric for evaluating the baseline aphasia severity and the resulting treatment response. Patients also underwent comprehensive multi-modal imaging assessment, including T1 structural MRI, perfusion, diffusion, task-based and resting-state fMRI. The present study only examines the rsfMRI data.

Table 1

Language/Cognitive Assessment Battery. All measures and submeasures are assigned an acronym for reference in later figures and analyses.

Measure (Acronym)	Submeasures (Acronym)
Western Aphasia Battery (WAB) (34)	Information Content (IC) Fluency (FL) Comprehension (CO) Repetition (RE) Naming (NA)
Northwestern Naming Battery (NNB) (63)	Noun Comprehension (NC) Verb Comprehension (VC) Noun Production (NP) Verb Production (VP)
Northwestern Assessment of Verbs and Sentences (NAVS) (64)	Canonical Sentence Comprehension Test (SCT-C) Noncanonical Sentence Comprehension Test (SCT-N) Canonical Sentence Production Priming Test (SPPT-C) Noncanonical Sentence Production Priming Test (SPPT-N)
Psycholinguistic assessments of language processing in aphasia (PALPA) 1 (65)	Phonological Discrimination
PALPA 35	Reading Regular (RE) Reading Exception (EX)
PALPA 40	Spelling High Frequency Words (HF) Spelling Low Frequency Words (LF)
PALPA 51	Semantic Association: High Imageability (HI) Semantic Association: Low Imageability (LI)
Pyramids and Palm Trees (PPT) (66)	Semantic Association
Doors and People (D&P) (67)	Explicit Memory
Cinderella Story (CIND) (68)	Words per Minute (WPM) Mean Length of Utterance - Words (MLW) Mean Length of Utterance - Morphemes (MLM)
Digit Span (DS)	Forwards (FOR) Backwards (BAC)
Treatment-specific Measure (TSM)	Object Naming <i>or</i> Sentence Comprehension & Production <i>or</i> Spelling Words

<Table 1 goes here>

2.2 Speech and Language Therapy

Following baseline testing, each patient received a three month course of deficit-specific SLT. The detailed treatment protocols have been described previously: anomia patients underwent a typicality-based semantic treatment (42); Agrammatism patients received sentence comprehension and production treatment through a Treatment of Underlying Forms program (43); dysgraphia patients underwent a spell-study-spell treatment protocol (44). The TSM was both before and after completion of the treatment protocol in order to estimate the treatment response.

2.3 Image Acquisition

MRI scans were acquired using 3.0 T scanners (Siemens Skyra at BU, Siemens Trio/Prisma at NU, and Philips Intera at JHU). Imaging protocols were harmonized across the sites to provide similar quality and timing. Structural images were collected using a 3D T1-weighted sequence (TR = 2300 ms, TE = 2.91 ms, flip angle = 9°, resolution = 1 mm³ isotropic). Whole brain functional images were collected using a gradient-echo T2*-weighted sequence (TR = 2 or 2.4 sec, TE = 20 ms, flip angle = 90°, resolution = 1.72 × 1.72 × 3 mm, 210 or 175 volumes). Initial studies (first 5 NU subjects) used a 2 second TR, but additional coverage was required to obtain whole brain data so TR was increased to 2.4 seconds. While NU and BU had one scan of 210 volumes, JHU subjects received 2 runs of 175 volumes each, and only the scan with the highest temporal signal-to-noise ratio (tSNR) was included for analysis.

2.4 Image Preprocessing

All images were archived on NUNDA (Northwestern University Neuroimaging Data Archive, <https://nunda.northwestern.edu>) for storage and data analysis. Upon arrival in the archive, image quality assurance (QA) was performed using automatic pipelines for functional and structural data. For fMRI data, a slice-wise tSNR was calculated as the ratio of the mean signal to the standard deviation of the time course data from each slice, weighted by the number of brain voxels in the slice. Poor-quality scans (tSNR < 100) were repeated or excluded from analysis.

Image preprocessing was performed using the NUNDA “Robust fMRI preprocessing pipeline”, which employs custom scripts built upon functions from AFNI, FSL, and SPM software (45–47). First, the fMRI time series were despiked (AFNI 3dDespike) and coregistered to the mean image (AFNI 3dvolreg). Normalization to standard MNI space was performed in a concatenated two-step procedure. A transformation aligning the first image in the fMRI time series to the T1 was created using boundary based registration (FSL BBR) (48). This was combined with the nonlinear warp of the T1 to an MNI template of 2 × 2 × 2 mm resolution (SPM Dartel Toolbox) (49). Structural images were corrected using enantiomorphic lesion transplant (SPM Clinical Toolbox) to minimize distortion effects caused by warping brains with lesions (50, 51). Using the lesion mask as a reference, right hemisphere homologous tissue was mirrored into the lesioned space to create a lesion-corrected left hemisphere. The optimal transform, calculated using the lesion-corrected brain, was then applied to the native brain.

2.5 rsfMRI Analysis

The rsfMRI features which predict aphasia recovery are currently unknown, encouraging a data-driven approach. Group independent component analysis (GICA) is a data-driven method of decomposing signals into underlying spatial and temporal components. Applied to rsfMRI, GICA may identify statistically independent patterns of brain activity across subjects. These patterns may then be compared across subjects, permitting analysis of network activity within and across groups. GICA also offers intrinsic noise filtering, as noise which is statistically independent of the signal filters into its own component. These components can be filtered out to perform group artifact removal, which has been demonstrated to be more reliable than artifact removal at the subject level (52). Subjects from all sites were processed together to attenuate the impact of single-scanner artifacts on final components.

GICA was performed through the GIFT toolbox for MATLAB (53). Data were decomposed using default parameters (20 components, InfoMax algorithm). Component projections clustered strongly across 100 random parameter initializations, indicating the chosen parameters were stable in our analysis (54). GICA components were then backprojected, producing 20 spatial components and 20 corresponding time series for each subject. Only right-handed patients were used for imaging analyses and imaging-based models due to alternate functional lateralization of language networks which would interfere with GICA-based modelling.

Measurement of low-frequency oscillations is broadly used as a generalized activity metric in fMRI analyses, including studies in both stroke (55, 56) and aphasia (57, 58). The fractional amplitude of low-frequency fluctuations (fALFF) is the ratio of power in the 0.01 Hz – 0.08 Hz band to the total power (59). The fALFF was calculated for each time series of each component for each subject, then standardized such that the sum of a subject's 20 fALFF values equals one. This permits comparison of relative component power within a subject, and normalizes activity ranges prior to regression analyses. We combined fALFF with GICA as a data reduction technique, whereby a series of functional volumes is summarized by an activity measure of 20 components. This approach helps avoid overfitting by reducing problem dimensionality, and makes regression more feasible with the given sample sizes.

2.6 Model Construction and Validation

Prediction of post-treatment primary dependent measures was modeled using elastic net regression. This model was chosen because linear models are relatively robust to lower sample sizes, and the elastic net penalty combats overfitting by combining L1 and L2 regularization (60). The relative contributions of L1 and L2 regularization were controlled by coefficient hyperparameters determined by leave-one-out cross-validation (LOOCV). Three separate models were built for each deficit type using pre-treatment data: (1) baseline TSM alone, (2) all language measures (including the TSM), or (3) fALFF alone. Model training and validation was performed with the caret package in RStudio. Model performance was estimated through the correlation and median absolute deviation (MAD) of the predicted post-treatment TSM with the actual post-treatment TSM.

2.7 Variable Importance

Backwards feature elimination was used to estimate the importance of individual measures due to high variable multicollinearity (Fig. 1, see below), which can impede strict interpretation of linear model coefficients. Instead, the median model correlation after LOOCV and 100 missing data imputations was computed after removing one variable at a time. The variable which, when discarded, maximizes model performance is then removed from the pool and the

process is repeated. The variable importance is then the difference in the model R^2 with vs. without the removed variable. We interpret this as the percentage of variability in the post-Tx TSM which is uniquely accounted for by that variable. Under this interpretation, variables with low or negative importance are not strictly poor prognostic measures, but rather there exists another measure which contributes the same information more clearly.

3. Results

3.1 Participants

Demographics and aphasia severity for the 70 recruited participants are shown in Table 2. Of the participants who entered the study, one dropped immediately and one passed away before completing testing. Both of these participants were completely excluded from analyses. Of the remaining 68 participants, one dropped out and one suffered a hematoma during therapy. For these participants the baseline language measures were used, and the post-treatment dependent measures were imputed. All baseline TSM data were collected, and 2.8% of the other baseline language measurements were missing due to incomplete testing and/or lack of follow-up. Missing data were imputed by random forests (randomForest Package, R programming language) (61). Imputation parameters were selected to minimize output variance. For subsequent analyses, model metrics are estimated across multiple imputations. In addition, two agrammatism participants were scanned with a different sequence, and these subjects were removed from fALFF-based analyses.

Table 2
Subject Demographics and WABAQ by Language-Specific Deficit.

Attribute	Anomia (N = 30)	Agrammatism (N = 16)	Dysgraphia (N = 24)	p
Gender	F: 10 M: 20	F: 5 M: 11	F: 9 M: 15	0.8995
Age	62.5 +/- 11.1	51 +/- 5.2	61.5 +/- 10.4	0.0005
Handedness	L: 2 R: 28	L: 3 R: 13	L: 5 R: 19	0.2992
Education (Years)	16 +/- 1.5	16.5 +/- 2.2	16 +/- 3.0	0.0418
Months Post Stroke	27.5 +/- 22.2	31 +/- 19.3	63.5 +/- 48.2	0.0186
Aphasia Severity (WABAQ)	62.2 +/- 32.5	74.85 +/- 14.1	83 +/- 15.4	0.0055

For categorical variables, counts by attribute and deficit type are shown. Categorical variable p-values are calculated with a two-sided Fisher's Exact Test. For continuous variables, median +/- median absolute deviation is shown. Continuous variable p-values are calculated with a Kruskal-Wallis one-way analysis of variance. Significant p-values are bolded ($p < 0.05$).

Significant differences were found to exist across the deficit groups in age, years of education, months post-stroke, and overall aphasia severity. Relationships between subject demographics and language measures or fALFF may be present across the aphasia subtypes. However, each aphasia subtype treatment is modelled separately, limiting the confounding effect of group differences. Furthermore, demographic variables were not included in the model, since they have not been found to be robust predictors of aphasia recovery (9).

3.2 Language/Cognitive Assessments

Correlations between the measures included in the language/cognitive battery are displayed in Fig. 1A. Hierarchical clustering was performed, using one minus the Kendall's Tau-b correlation as a distance metric between tests (Fig. 1B). Almost all measures correlated positively with all other measures, except for the PALPA 1 and Doors & People measures. These measures clustered independently in the dendrogram. Correlations are especially high within a language measure group (i.e. NAVS), and submeasures cluster together. The observed multicollinearity across measures is expected, since nearly all measures test an aspect of language ability.

Figure 1: Multicollinearity across Baseline Language Measures. 1A: A shaded color-plot of the correlation matrix is shown. Due to imbalance in sample sizes, correlations were first calculated within each language deficit, and then averaged to yield the matrix shown here. Box colors correspond to pairwise Kendall's Tau-b values (red is positive, blue is negative correlation). Only pairwise complete observations were used (no imputation). Measures are generally correlated with one another (mean pairwise correlation = 0.328). 1B: An association dendrogram of language measures is shown. Correlation distance is one minus the absolute pairwise Kendall's Tau correlation. The dendrogram formed by hierarchical clustering of the correlation distances is shown (Unweighted Pair Group Method with Arithmetic Mean).

3.3 GICA Results

Averaged components from across subjects created through the GIFT toolbox are displayed in Fig. 2. The observed maps are a mix of known networks, regions consisting of grey matter, regions known to be sensitive to physiologic motion, and the ventricles. This reflects our minimal image preprocessing and inclusion of all voxels within the brain. Some grey matter activation may be due to partial volume effects from cerebrospinal fluid motion. The identical regression analyses were repeated with more heavily preprocessed data that accounts for motion and physiologic noise, however the prediction results could

not be replicated. This implies that non-grey matter voxels may be driving the predictive power of the subsequent models, and is further interpreted in the discussion.

Figure 2: Aggregate GICA Components. Backprojected spatial components were averaged across subjects to create aggregate component maps (shown). Sagittal, axial, transverse planes are displayed about each component's peak point. Color corresponds to the z-score of the voxel coefficient (1-sided t test). Colors are scaled to each component's range ($3 \leq |z| \leq |z_{\text{peak}}|$). Red voxels are component correlated while blue voxels are component anticorrelated.

3.4 Treatment Response Modeling

Patient performance on the TSM increased significantly over the course of treatment for all therapy groups: Anomia ($p = 9.32E-9$), Agrammatism ($p = 3.05E-5$), Dysgraphia ($p = 1.19E-7$) (one-sided Exact Wilcoxon Signed Rank Test, 1000 imputations). First, we predicted post-Tx TSM using only the pre-Tx TSM as a performance reference point for other models (Fig. 3). The error, correlation, and p-value are calculated for each model (one-sided t-test for Pearson correlation). Post-Tx TSM after anomia treatment had a strong linear relationship to pre-Tx score ($R = 0.902$, $N = 30$, $p = 5.0e-12$, 95% CI: (0.833, 0.944), $MAD = 0.113$). In the agrammatism group, pre-Tx TSM was a borderline-significant predictor of post-Tx TSM ($R = 0.385$, $N = 16$, $p = 0.070$, 95% CI: (-0.075, 0.724), $MAD = 0.165$). The dysgraphia therapy had no linear relationship between pre-Tx and post-Tx TSM scores ($R = 0.142$, $N = 22$, $p = 0.264$, 95% CI: (-0.390, 0.571), $MAD = 0.061$). This is expected since the dysgraphia treatment trains individuals to a performance threshold. Therefore the post-Tx score was largely independent of the pre-Tx score, and this pattern was reflected in our model performance.

Figure 3: Predicting Post-Tx Language Outcome with Baseline TSM. Linear models which predict deficit-specific measure after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

3.4.1 Modeling with Language Measures

Next, we predicted post-Tx score on the deficit-specific measure using all baseline scores on the cognitive/language assessment battery (28 predictors total, Fig. 4). The effect of anomia treatment on the TSM again demonstrated a strong linear relationship ($R = 0.958$, $N = 30$, $p = 5.0e-17$, 95% CI: (0.922, 0.979), $MAD = 0.075$). This relationship was significantly stronger than when only the pre-Tx TSM is used ($p < 0.05$). The agrammatism model demonstrated a statistically significant linear relationship ($R = 0.589$, $N = 16$, $p = 0.0082$, 95% CI: (0.245, 0.836), $MAD = 0.132$), but was not significantly better than the TSM alone ($p > 0.05$). The dysgraphia model also showed a statistically significant relationship ($R = 0.456$, $N = 22$, $p = 0.016$, 95% CI: (0.125, 0.741), $MAD: 0.105$). However it misses the general trend of the data and produces predictions with very high error. The magnitude of this error is not entirely captured by the MAD.

Figure 4: Predicting Post-Tx Language Outcome with Baseline Language Measures. Linear models which predict the TSM after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

3.4.2 Relative Importance of Language Predictors

Next we identified which cognitive/language predictors contributed most to the predictive models performance (Table 3). Individual behavioral measure importances were typically small - only 12% of the variables individually explained more than 5% of the outcome variance across all the aphasia types. This is expected due to high multicollinearity and therefore redundant information across measures. For the anomia model, the baseline deficit-specific measure was the most important variable by far, explaining nearly 80% of outcome variability. SCT-N and PALPA40-HF followed in importance, which although small, likely drive the significant difference between the baseline deficit measure and the full behavioral model. The agrammatism model was primarily driven by SPPT-C, an agrammatism measure, and then the baseline TSM. Top variables in the dysgraphia model were PALPA35-EX and PALPA1. However the dysgraphia model had very high error, limiting interpretability for important variables. Interestingly, D&P has high importance in both the agrammatism and dysgraphia models. A follow-up analysis demonstrated positive regression coefficients for all important tests in their respective models, with exception of SCT-N in the anomia model.

Table 3
Relative Importance of Behavioral Measures.

Language Measure	Anomia (N = 30)	Agrammatism (N = 16)	Dysgraphia (N = 22)
WAB-IC	-0.25	-0.72	-0.45
WAB-FL	1.24	-2.25	-0.33
WAB-CO	-0.35	1.64	-0.99
WAB-RE	-0.37	0.80	-1.05
WAB-NA	-0.21	-2.40	-0.67
NNB-NC	-0.23	0.16	-0.26
NNB-VC	-0.16	-1.13	-2.30
NNB-NP	-0.27	-0.46	-0.78
NNB-VP	-0.25	-4.11	-1.81
SCT-C	0.28	-2.16	-0.43
SCT-N	6.34	-1.59	-0.67
SPPT-C	-0.30	40.67	0.36
SPPT-N	0.77	-2.16	1.07
PALPA1	0.81	-2.32	12.40
PPT	0.01	-0.15	-0.65
PALPA35-RE	-0.18	-0.28	-2.15
PALPA35-EX	1.84	2.47	16.51
PALPA40-HF	3.42	-0.85	-1.70
PALPA40-LF	-0.24	-3.38	-1.91
PALPA51-HI	-0.69	-3.32	1.22
PALPA51-LI	0.22	-11.37	-1.78
CIND-WPM	-0.23	-1.09	-0.27
CIND-MLW	0.26	9.79	-2.17
CIND-MLM	-0.16	-2.34	-0.75
D&P	-0.17	11.24	9.51
DS-FOR	-0.18	-1.27	-0.73
DS-BAC	0.82	-1.49	-0.46
TSM	79.89	12.92	2.01

Variable importance within the predictive model was calculated by backwards feature elimination. As the TSM is collected at baseline by default, it was always retained by the model. Importance is the percentage of variability in the post-Tx TSM explained by each baseline language measure. Variables with higher importance contribute more to model performance, while low and negative values suggest a better alternative to that measure is available.

3.4.3 Modeling with fALFF Predictors

We next built prognostic models using only fALFF values of independent components (Fig. 5). There was a significant drop in performance for the anomia model ($R = 0.366$, $N = 28$, $p = 0.028$, 95% CI: (0.107, 0.650), $MAD = 0.299$). The model missed the overall trend of the data, but had a marginally significant statistical relationship. However, the fALFF agrammatism model was significantly better than the behavioral model ($R = 0.940$, $N = 11$, $p = 8.5e-6$, 95% CI: (0.730, 0.995), $MAD = 0.030$). Similar performance gains were seen in the dysgraphia model ($R = 0.925$, $N = 18$, $p = 2.0e-8$, 95% CI: (0.819, 0.973), $MAD = 0.016$).

Figure 5: Predicting Post-Tx Language Outcome with Baseline fALFF. Linear models which predict deficit-specific measure after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

3.4.4 Relative Importance of fALFF Predictors

Identifying the most important components in fALFF-based models highlights the components which drive performance. For direct comparison to the behavioral variable importance, we calculated importance of component fALFF with backwards feature elimination (Table 4). The anomia model was primarily driven by components 2 and 16. Component 2 was unexpected since it has a large ventricular presence, although there is some insular activation as well. This model has very high error so the significance of the variable importances is hard to interpret. Component 1 accounts for almost 72% of the agrammatism variability, and along with component 5 form the bulk of the model's predictive power. The dysgraphia model importances are more distributed, with components 9 and 18 scoring highest.

Table 4
Relative Variable Importance by Backwards Elimination.
 Variable importance within the predictive model was calculated by backwards feature elimination. Importance is the percentage of variability in the post-Tx TSM explained by each component. Components with higher importance contribute more to model performance. Negative values suggest a better alternative to that component is available.

GICA Component	Anomia (N = 28)	Agrammatism (N = 11)	Dysgraphia (N = 18)
1	-1.90	71.65	0.88
2	16.30	-0.41	0.00
3	-3.61	-0.29	-2.44
4	0.41	4.82	5.94
5	0.02	11.84	2.14
6	-2.03	0.06	-1.01
7	0.20	0.11	-0.80
8	3.57	0.73	8.07
9	-5.95	0.15	15.38
10	-0.44	-0.83	5.58
11	3.22	3.67	1.61
12	-3.85	1.62	-12.53
13	-1.10	3.34	6.79
14	-0.17	0.85	-0.68
15	-1.20	-0.25	8.66
16	11.48	-0.04	4.30
17	1.15	-2.19	10.76
18	0.45	1.31	22.56
19	0.83	-0.84	5.05
20	-3.18	-6.87	5.09

4. Discussion

We have presented several prognostic models for language recovery in stroke aphasia. Models based only on the treatment-specific measure (TSM) performed surprisingly well in the anomia group. Adding the SCT-N (sentence comprehension) and PALPA40-HF (spelling) measures to the model significantly improved performance of the anomia model ($r = 0.958$ vs. $r = 0.902$). Although neither of these are anomia measures, this is expected because language sub-measures are highly correlated (Fig. 1) and including anomic measures alongside the TSM may not contribute sufficient independent information. For the agrammatism and dysgraphia groups, D&P was a surprisingly important and positive predictor of recovery. D&P is primarily a recall test, and correlates comparatively weakly with the other tests. It is possible this test represents general cognitive abilities which generalize to aphasia recovery. Neither the dysgraphia or agrammatism language-measure models performed particularly well, suggesting that general cognitive ability is the most important predictor of recovery in the absence of more specific predictors.

Models based on the fractional amplitude of low-frequency fluctuations (fALFF) had strong performances in the agrammatism and dysgraphia groups. The agrammatism model benefited most from the inclusion of component 1, while the dysgraphia model's predictive power was more distributed across variables. The fALFF model performed especially poorly in the anomia group. Anomia is one of the most common forms of post-stroke aphasia, and the brain lesions which cause anomia are considerably less centralized than in other aphasic disorders (62). Prior work with our anomic participants has shown that relative nonresponders have altered patterns of skill transfer (transitioning non-anomic language abilities into anomic recovery over the course of treatment) when compared to responsive participants (42). The decentralization of brain lesions and varied patterns of skill transfer suggest that nonlinearities may be present

in the neuroimaging features which exceed the capacity of the linear model. These limitations were not present in the behavioral models, as performance on language measures account for deficiencies across a language network. Splitting the patient population into anomia subtypes may improve fALFF model performance in this group.

Several limitations were present in this study. First, model selection was constrained by sample size, to where linear models were the only practical choice. The LOWESS curves had a sigmoid shape in a number of our models, suggesting a nonlinear relationship between the predicted and actual post-treatment TSMs that could be rectified by a more complex model. Second, performing regression with more variables than subjects necessitated a linear model with heavy regularization. While regularization improves model generalization, more accurate models could be realized with more samples. Inclusion of more subjects in the study would permit creation of independent test and validation sets and a more accurate error estimate than can be achieved with cross-validation. Subject recruitment was limited in part by the extensive testing which required several hours for each subject to complete. Group independent component analysis (GICA) yielded several image components which lack clear interpretation as known functional brain networks. GICA components are averaged across subject-level back-projected networks, creating spatially diffuse patterns that are hard to recognize. Despite this, components which predict therapeutic response in our cohort more closely resemble physiologic motion than brain networks. When the data were preprocessed with correction for motion and physiologic noise, the resulting components were not predictive. This implies that patterns of physiologic motion measured through rsfMRI may have prognostic value for stroke aphasia recovery. The vessels and sinuses of the brain form a densely coupled fluid dynamic system which is perturbed by a stroke. It is possible that aspects of neurovascular pathology are reflected as alterations in physiologic motion. In this case, our GICA components may be operating as proxy variables for measures of neurological health.

5. Conclusions

High-performance predictive models of individual response to therapy were trained for each aphasia deficit type ($r = 0.925\text{--}0.958$). These models are the state-of-the-art at the time of writing, and are among the first to extend associations between patient outcomes and variables to predictive model. Inclusion of rsfMRI features resulted in models which far outperform language measures alone (SPEAK model, $r = 0.75$) (16). A prior model by Pustina et. al. applied both lesion information and functional connectivity to create predictive models of four aphasia scores in 53 patients, with model correlations ranging between 0.79–0.88 (26). This indicates that language measures and rsfMRI should be applied together in order to determine outcomes for chronic aphasia patients, and prompts further inquiry into the clinical utility of rsfMRI in post-stroke aphasia.

rsfMRI has several advantages over conventional language measures in aphasia assessment and prognosis. Language assessments have a limited range of deficit severity to which they are sensitive. In some cases, this limits assessment of changes in patients with severe or minor deficits. These ceiling, floor, and discretization errors were observed in our data, limiting the sensitivity of assessment. rsfMRI offers a more continuous measure of brain activity, with potential to equitably evaluate a wider range of patients. Creating an aphasia profile from rsfMRI may be easier in practice than refining a battery from existing language measures. However there are challenges facing clinical adoption of rsfMRI for aphasia. Imaging is relatively expensive, and accessibility to quality scanning varies across patient populations. It is challenging to image patients who are claustrophobic or unable to keep still. Non-right-handed subjects were excluded from image-based analyses in our study since few were recruited. Larger studies will be needed to overcome the sample imbalance in handedness, and this may prove problematic for imaging-based measures. Finally, rsfMRI relies on neurovascular coupling to detect changes in activity related to connectivity. In persons with stroke there may be additional variability in sensitivity due to the physiological limitations of the rsfMRI scan.

In addition to predicting language outcomes for clinical purposes, prognostic models may help improve patient outcomes. Large matched cohorts or randomized control trials are not practical to run in the exploratory phase of treatment optimization. An alternative is to model patient response to all existing therapies, and match the patient to the therapy with highest expected performance. Repeating this analysis on other aphasia therapies will improve our understanding of patient-level factors which are predictive of treatment response, and open the door to personalized aphasia therapy.

Declarations

Ethics approval and consent to participate: All participants provided written informed consent according to Institutional Review Board policies at Boston University, Johns Hopkins University, and Northwestern university.

Consent for publication: Not applicable.

Availability of data and materials: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. Source code for the methods applied in this work is available at <https://github.com/miorga7> (repository: aphasia_prediction, R code).

Competing interests: The authors have no competing interests.

Funding: This work was supported by the NIH-NIDCD, Grant P50DC012283 (recipient Cynthia K. Thompson); and the NIH-NIHMS, Grant T32GM008152 (recipient Northwestern University).

Author's Contributions: Michael Iorga designed and implemented the modelling approaches. James Higgins preprocessed imaging data. Todd Parrish provided guidance on the image analysis approach. Richard Zinbarg provided guidance on statistical analysis. Cynthia Thompson, Brenda Rapp, Swathi Kiran, David Caplan oversaw patient recruitment, testing, and treatment. All authors jointly interpreted the results. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

1. Goodglass H. Understanding Aphasia. 1993. 297p.
2. Lazar RM, Boehme AK. Aphasia As a Predictor of Stroke Outcome. *Curr Neurol Neurosci Rep*. 2017 Sep 19;17(11):83.
3. Engelter ST, Gostynski M, Papa S, Frei M, Born C, Ajdacic-Gross V, et al. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006 Jun;37(6):1379–84.
4. Winstein CJ, Stein J, Arena R, Bates B, Cherney LR, Cramer SC, et al. Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*. 2016 Jun;47(6):e98–169.
5. Thiel A, Zumbansen A. Recent advances in the treatment of post-stroke aphasia. *Neural Regeneration Res*. 2014;9(7):703.
6. Laska AC, Hellblom A, Murray V, Kahant T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med*. 2001;249(5):413–22.
7. Lazar RM, Minzer B, Antoniello D, Festa JR, Krakauer JW, Marshall RS. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke*. 2010 Jul;41(7):1485–8.
8. Hilari K, Byng S. Health-related quality of life in people with severe aphasia. *Int J Lang Commun Disord*. 2009 Mar;44(2):193–205.
9. M. m. W, Watila MM, Balarabe SA. Factors predicting post-stroke aphasia recovery. *J Neurol Sci*. 2015;352(1–2):12–8.
10. 10.1002/14651858.cd000425.pub4
Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev* [Internet]. 2016; Available from: <http://dx.doi.org/10.1002/14651858.cd000425.pub4>
11. Charidimou A, Kasselimis D, Varkanitsa M, Selai C, Potagas C, Evdokimidis I. Why is it difficult to predict language impairment and outcome in patients with aphasia after stroke? *J Clin Neurol*. 2014 Apr;10(2):75–83.
12. Seghier ML, Patel E, Prejawa S, Ramsden S, Selmer A, Lim L, et al. The PLORAS Database: A Data Repository for Predicting Language Outcome and Recovery After Stroke. *J Clin Neurol*. 2018 Mar;18(12):90–100.
13. 10.12688/f1000research.11122.1
Tippett DC, Hillis AE. Where are aphasia theory and management “headed”? *F1000Res* [Internet]. 2017 Jul 3;6. Available from: <http://dx.doi.org/10.12688/f1000research.11122.1>
14. Doogan C, Dignam J, Copland D, Leff A. Aphasia Recovery: When, How and Who to Treat? *Curr Neurol Neurosci Rep*. 2018 Oct 15;18(12):90.
15. Benganem S, Rosso C, Arbizu C, Moulton E, Dormont D, Leger A, et al. Aphasia outcome: the interactions between initial severity, lesion size and location. *J Neurol*. 2019 Jun;266(6):1303–9.
16. El Hachoui H, Lingsma HF, van de Sandt-Koenderman MWME, Dippel DWJ, Koudstaal PJ, Visch-Brink EG. Long-term prognosis of aphasia after stroke. *J Neurol Neurosurg Psychiatry*. 2013 Mar;84(3):310–5.
17. Osa Garcia A, Brambati SM, Brisebois A, Désilets-Barnabé M, Houzé B, Bedetti C, et al. Predicting Early Post-stroke Aphasia Outcome From Initial Aphasia Severity. *Front Neurol*. 2020 Feb 21;11:120.
18. Halai AD, Woollams AM, Lambon Ralph MA. Predicting the pattern and severity of chronic post-stroke language deficits from functionally-partitioned structural lesions. *Neuroimage Clin*. 2018 Mar 16;19:1–13.
19. Schumacher R, Halai AD, Lambon Ralph MA. Assessing and mapping language, attention and executive multidimensional deficits in stroke aphasia. *Brain*. 2019 Oct 1;142(10):3202–16.
20. Sul B, Lee KB, Hong BY, Kim JS, Kim J, Hwang WS, et al. Association of Lesion Location With Long-Term Recovery in Post-stroke Aphasia and Language Deficits. *Front Neurol*. 2019 Jul 24;10:776.
21. Price CJ, Seghier ML, Leff AP. Predicting language outcome and recovery after stroke: the PLORAS system. *Nat Rev Neurol*. 2010 Apr;6(4):202–10.
22. Harvey RL. Predictors of Functional Outcome Following Stroke. *Phys Med Rehabil Clin N Am*. 2015;26(4):583–98.
23. Tochadse M, Halai AD, Lambon Ralph MA, Abels S. Unification of behavioural, computational and neural accounts of word production errors in post-stroke aphasia. *Neuroimage Clin*. 2018 Mar 27;18:952–62.
24. Halai AD, Woollams AM, Lambon Ralph MA. Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *Neuroimage Clin*. 2018;17:465–73.
25. Halai AD, Woollams AM, Lambon Ralph MA. Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex*. 2017 Jan;86:275–89.
26. Pustina D, Coslett HB, Ungar L, Faseyitan OK, Medaglia JD, Avants B, et al. Enhanced estimations of post-stroke aphasia severity using stacked multimodal predictions. *Hum Brain Mapp*. 2017 Nov;38(11):5603–15.
27. Yang M, Yang P, Fan Y-S, Li J, Yao D, Liao W, et al. Altered Structure and Intrinsic Functional Connectivity in Post-stroke Aphasia. *Brain Topogr*. 2018 Mar;31(2):300–10.
28. Sandberg CW. Hypoconnectivity of Resting-State Networks in Persons with Aphasia Compared with Healthy Age-Matched Adults. *Front Hum Neurosci*. 2017 Feb 28;11:91.
29. Balaev V, Petrushevsky A, Martynova O. Changes in Functional Connectivity of Default Mode Network with Auditory and Right Frontoparietal Networks in Poststroke Aphasia. *Brain Connect*. 2016;6(9):714–23.
30. Baliki MN, Babbitt EM, Cherney LR. Brain network topology influences response to intensive comprehensive aphasia treatment. *NeuroRehabilitation*. 2018;43(1):63–76.

31. Siegel JS, Seitzman BA, Ramsey LE, Ortega M, Gordon EM, Dosenbach NUF, et al. Re-emergence of modular brain networks in stroke recovery. *Cortex*. 2018 Apr;101:44–59.
32. Nair VA, Young BM, La C, Reiter P, Nadkarni TN, Song J, et al. Functional connectivity changes in the language network during stroke recovery. *Ann Clin Transl Neurol*. 2015 Feb;2(2):185–95.
33. Zhao Y, Lambon Ralph MA, Halai AD. Relating resting-state hemodynamic changes to the variable language profiles in post-stroke aphasia. *Neuroimage Clin*. 2018 Aug 21;20:611–9.
34. 10.1037/t15168-000
Kertesz A. Western Aphasia Battery – Revised [Internet]. *PsycTESTS Dataset*. 2006. Available from: <http://dx.doi.org/10.1037/t15168-000>
35. 10.1016/j.apmr.2018.08.177
Gilmore N, Dwyer M, Kiran S. Benchmark of significant change after aphasia rehabilitation. *Arch Phys Med Rehabil* [Internet]. 2018 Sep 18; Available from: <http://dx.doi.org/10.1016/j.apmr.2018.08.177>
36. Martin N, Minkina I, Kohen FP, Kalinyak-Fliszar M. Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *J Neurolinguistics*. 2018 Nov;48:199–225.
37. Fromm D, Forbes M, Holland A, Dalton SG, Richardson J, MacWhinney B. Discourse Characteristics in Aphasia Beyond the Western Aphasia Battery Cutoff. *Am J Speech Lang Pathol*. 2017 Aug 15;26(3):762–8.
38. Rohde A, Worrall L, Godecke E, O'Halloran R, Farrell A, Massey M. Diagnosis of aphasia in stroke populations: A systematic review of language tests. *PLoS One*. 2018 Mar 29;13(3):e0192773.
39. El Hachoui H, Visch-Brink EG, de Lau LML, van de Sandt-Koenderman MWME, Nouwens F, Koudstaal PJ, et al. Screening tests for aphasia in patients with stroke: a systematic review. *J Neurol*. 2017 Feb;264(2):211–20.
40. 10.1111/1460-6984.12420
Pritchard M, Hilari K, Cocks N, Dipper L. Psychometric properties of discourse measures in aphasia: acceptability, reliability, and validity. *Int J Lang Commun Disord* [Internet]. 2018; Available from: <http://dx.doi.org/10.1111/1460-6984.12420>
41. Wilson SM, Eriksson DK, Schneck SM, Lucanie JM. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One*. 2018 Feb 9;13(2):e0192773.
42. Gilmore N, Meier EL, Johnson JP, Kiran S. Typicality-based semantic treatment for anomia results in multiple levels of generalisation. *Neuropsychol Rehabil*. 2018 Jul 20;1–27.
43. Thompson CK, Shapiro LP. Treating agrammatic aphasia within a linguistic framework: Treatment of Underlying Forms. *Aphasiology*. 2005 Nov;19(10–11):1021–36.
44. Rapp B, Kane A. Remediation of deficits affecting different components of the spelling process. *Aphasiology*. 2002;16(4–6):439–54.
45. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier; 2011. 656p.
46. Cox RW. AFNI: What a long strange trip it's been. *Neuroimage*. 2012;62(2):743–7.
47. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012;62(2):782–90.
48. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*. 2009 Oct 15;48(1):63–72.
49. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage*. 2007 Oct 15;38(1):95–113.
50. Nachev P, Coulthard E, Jäger HR, Kennard C, Husain M. Enantiomorphic normalization of focally lesioned brains. *Neuroimage*. 2008 Feb 1;39(3):1215–26.
51. Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath H-O. Age-specific CT and MRI templates for spatial normalization. *Neuroimage*. 2012 Jul 16;61(4):957–65.
52. Du Y, Allen EA, He H, Sui J, Wu L, Calhoun VD. Artifact removal in the context of group ICA: A comparison of single-subject and group approaches. *Hum Brain Mapp*. 2015;37(3):1005–25.
53. Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp*. 2001 Nov;14(3):140–51.
54. 10.1109/nnspp.2003.1318025
Himberg J, Hyvarinen A. ICA: software for investigating the reliability of ICA estimates by clustering and visualization. In: 2003 IEEE III Workshop on Neural Networks [Internet]. Available from: <http://dx.doi.org/10.1109/nnspp.2003.1318025>
55. La C, Mossahebi P, Nair VA, Young BM, Stamm J, Birn R, et al. Differing Patterns of Altered Slow-5 Oscillations in Healthy Aging and Ischemic Stroke. *Front Hum Neurosci*. 2016 Apr 13;10:156.
56. Egorova N, Veldsman M, Cumming T, Brodtmann A. Fractional amplitude of low-frequency fluctuations (fALFF) in post-stroke depression. *Neuroimage Clin*. 2017 Jul 18;16:116–24.
57. van Hees S, McMahon K, Angwin A, de Zubicaray G, Read S, Copland DA. A functional MRI study of the relationship between naming treatment outcomes and resting state functional connectivity in post-stroke aphasia. *Hum Brain Mapp*. 2014 Aug;35(8):3919–31.
58. Li J, Du D, Gao W, Sun X, Xie H, Zhang G, et al. The regional neuronal activity in left posterior middle temporal gyrus is correlated with the severity of chronic aphasia. *Neuropsychiatr Dis Treat*. 2017 Jul 20;13:1937–45.
59. Zou Q-H, Zhu C-Z, Yang Y, Zuo X-N, Long X-Y, Cao Q-J, et al. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *J Neurosci Methods*. 2008 Jul 15;172(1):137–41.
60. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301–20.
61. Breiman L. *Random Forests* [Internet]. Vol. 45, *Machine Learning*. 2001. p. 5–32. Available from: <http://link.springer.com/10.1023/A:1010933404324>
62. Yourganov G, Smith KG, Fridriksson J, Rorden C. Predicting aphasia type from brain damage measured with structural MRI. *Cortex*. 2015 Dec;73:203–15.

63. ThompsonCK,LukicS,KingMC,MesulamMM,WeintraubS. Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery().*Aphasiology*.2012May1;26(5):632–55.
64. Cho-ReyesS,ThompsonCK. Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS).*Aphasiology*.2012;26(10):1250–77.
65. KayJ,LesserR,ColtheartM. Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction.*Aphasiology*.1996;10(2):159–80.
66. KleinLA,BuchananJA. Psychometric properties of the Pyramids and Palm Trees Test.*J Clin Exp Neuropsychol*.2009Oct;31(7):803–8.
67. BaddeleyAD. Doors and People: A Test of Visual and Verbal Recall and Recognition.2006.
68. MacWhinneyB,FrommD,HollandA,ForbesM,WrightH. Automated analysis of the Cinderella story.*Aphasiology*. 2010 Jun1;24(6–8):856.

Figures

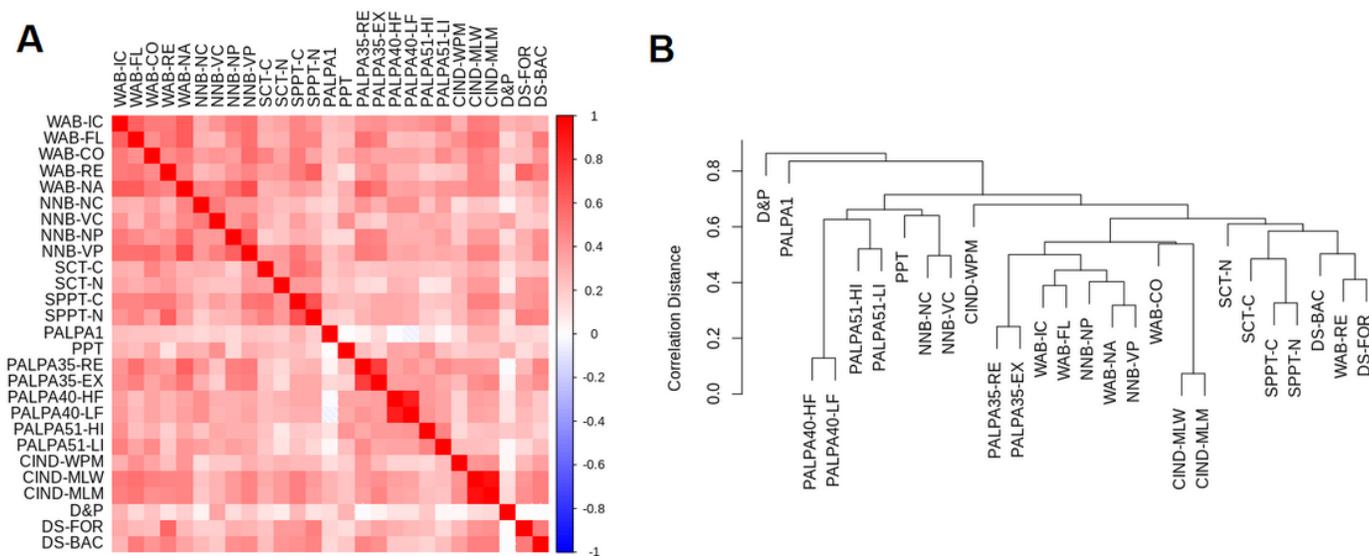


Figure 1

Multicollinearity across Baseline Language Measures. 1A: A shaded color-plot of the correlation matrix is shown. Due to imbalance in sample sizes, correlations were first calculated within each language deficit, and then averaged to yield the matrix shown here. Box colors correspond to pairwise Kendall's Tau-b values (red is positive, blue is negative correlation). Only pairwise complete observations were used (no imputation). Measures are generally correlated with one another (mean pairwise correlation = 0.328). 1B: An association dendrogram of language measures is shown. Correlation distance is one minus the absolute pairwise Kendall's Tau correlation. The dendrogram formed by hierarchical clustering of the correlation distances is shown (Unweighted Pair Group Method with Arithmetic Mean).

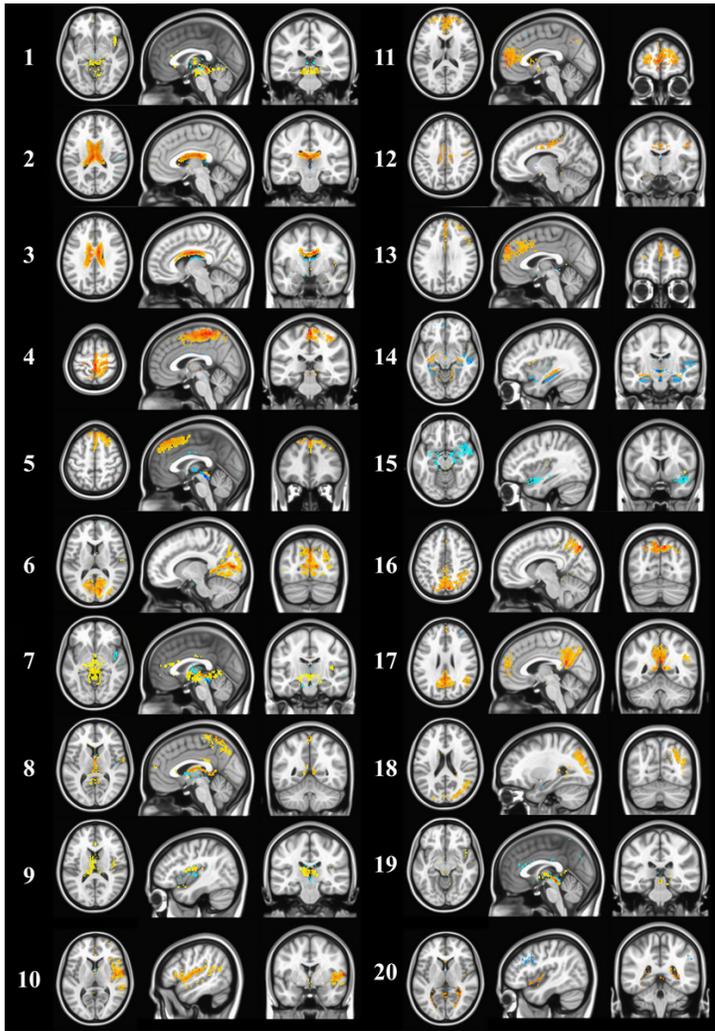


Figure 2
 Aggregate GICA Components. Backprojected spatial components were averaged across subjects to create aggregate component maps (shown). Sagittal, axial, transverse planes are displayed about each component's peak point. Color corresponds to the z-score of the voxel coefficient (1-sided t test). Colors are scaled to each component's range ($3 < |z| < |z_{peak}|$). Red voxels are component correlated while blue voxels are component anticorrelated.

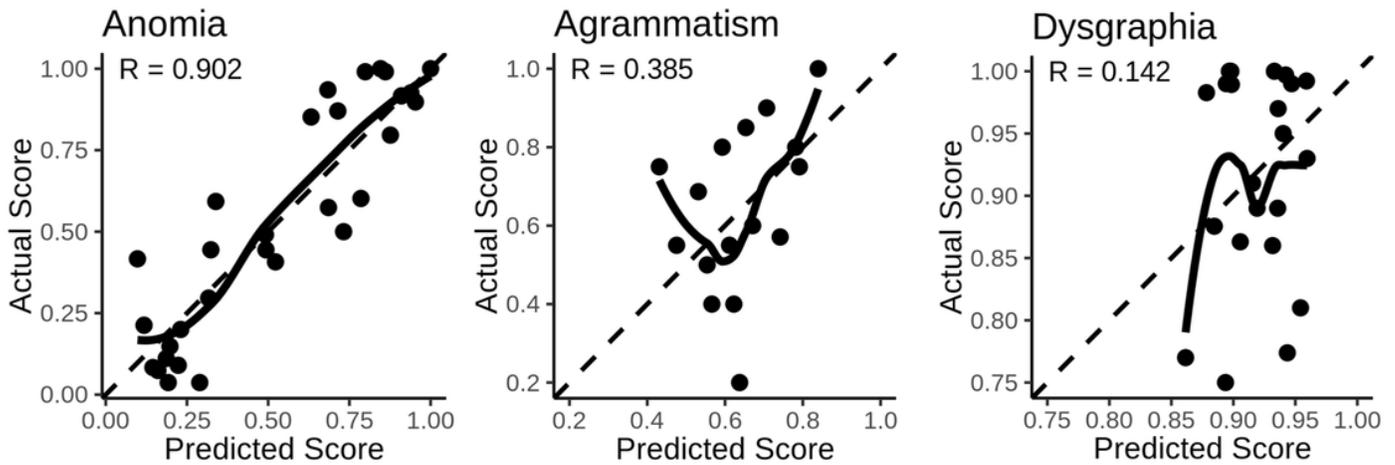


Figure 3
 Predicting Post-Tx Language Outcome with Baseline TSM. Linear models which predict deficit-specific measure after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows

smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

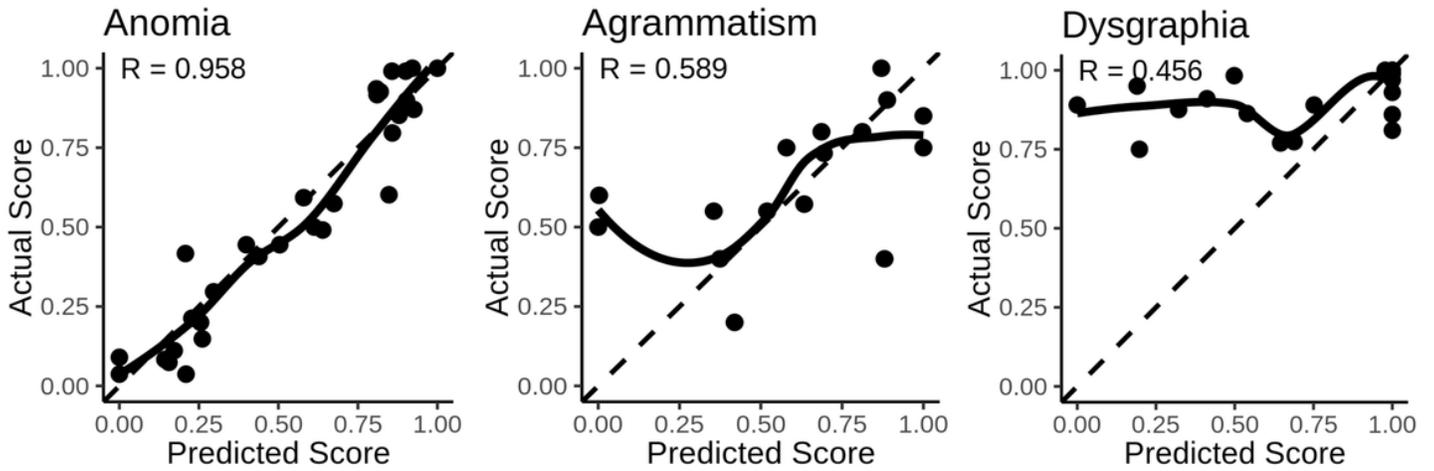


Figure 4

Predicting Post-Tx Language Outcome with Baseline Language Measures. Linear models which predict the TSM after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

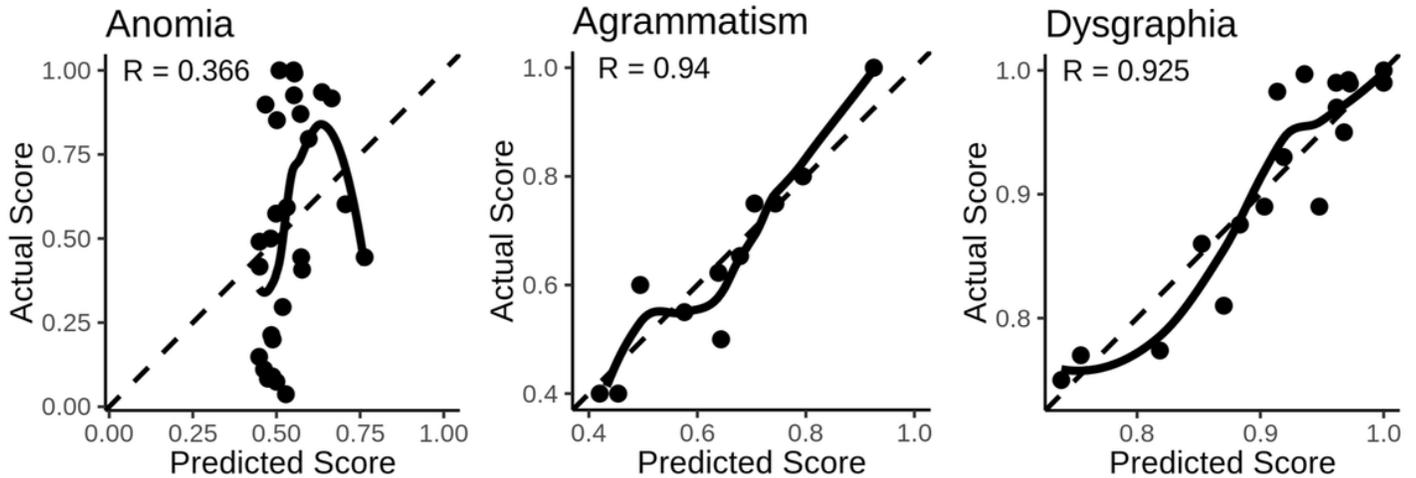


Figure 5

Predicting Post-Tx Language Outcome with Baseline fALFF. Linear models which predict deficit-specific measure after therapy were constructed for each aphasia type. Black circles show the predicted score for each patient during LOOCV. Where multiple values exist due to imputation, the median value and prediction are shown. The dashed line represents a perfect prediction. The LOWESS curve (locally-weighted polynomial regression, solid black line) shows smoothed median predictions across 100 imputations. For each model, the median correlation across imputations is shown (convergence data available in supplement).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFigure2.png](#)

- [SupplementalFigure1.png](#)