Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Serena Song April 12, 2022

Predicting Rehabilitation Responses in Persons with Aphasia Undergoing Language Therapy Using Baseline Whole-Brain Task fMRI in a Multivariate Method

by

Serena Song

Venkatagiri Krishnamurthy, PhD.

Advisor

Neuroscience and Behavioral Biology

Venkatagiri Krishnamurthy, PhD.

Advisor

Amy Rodriguez, PhD.

Committee Member

Malavika Murugan, PhD.

Committee Member

2022

Predicting Rehabilitation Responses in Persons with Aphasia Undergoing Language Therapy Using Baseline Whole-Brain Task fMRI in a Multivariate Method

by

Serena Song

Venkatagiri Krishnamurthy, PhD.

Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience and Behavioral Biology

2022

Abstract

Predicting Rehabilitation Responses in Persons with Aphasia Undergoing Language Therapy Using Baseline Whole-Brain Task fMRI in a Multivariate Method

by Serena Song

Background: With the diversity in aphasia profiles as well as a small window of plasticity, it is imperative to deliver the most effective rehabilitation treatment to patients soon after their incidence of stroke. Voxel-based lesion-symptom mapping (VLSM) is a widely used approach for identifying predictors of treatment outcomes, but the lack of network-level information in VLSM may omit the holistic details needed to create valid language-network prediction models. As task-fMRI offers neurobiological inspection of lesion impact on the whole language network, this study aims to develop a multivariate analysis inclusive of whole-brain task-fMRI data to more holistically predict treatment outcomes in persons with aphasia (PWA) undergoing language therapy.

Methods: This thesis will be a correlational study focused on analyzing the relationship between patterns of pretreatment brain activity across multiple PWA and treatment-induced language improvements over time to predict post-treatment outcomes. Initially, multivariate methods will be optimized to handle whole-brain task fMRI data by inserting a customized mask representative of the whole-brain as well as testing a series of k-fold cross-validations in order to find the optimal sparseness value. This will then be tested systematically for reliability, extended utility, and overall sensitivity.

Results: Multivariate optimizations included the modification of k-fold CVs, brain masks, and lesion size correction. With these changes, the multivariate analysis was able to feasibly predict post-treatment outcomes at both whole-level and group-level analyses. Additionally, predictive biomarkers such as the right extrastriate area and right fusiform gyrus were discovered during brain-behavior analysis of semantic fluency. When analyzing the extended utility of multivariate methods, the left medial frontal gyrus was found to predict post-treatment responses to more sentence-level language functions such as grammar.

Conclusion: Ultimately, the goal of this study is to create an optimized prediction model that is better able to capture network-level information for functions such as language. The use of such information would help clinicians predict rehabilitation responses in PWA to develop a more effective rehabilitation regimen for aphasia patients.

Predicting Rehabilitation Responses in Persons with Aphasia Undergoing Language Therapy Using Baseline Whole-Brain Task fMRI in a Multivariate Method

by

Serena Song

Venkatagiri Krishnamurthy, PhD.

Advisor

An abstract of a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience and Behavioral Biology

2022

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Venkatagiri Krishnamurthy for his endless support and guidance throughout my research experience. This work has been made possible because of his encouraging and motivating advice as well as his shared extensive knowledge. In addition to my advisor, I would also like to thank all members of my committee, Dr. Amy Rodriguez and Dr. Malavika Murugan for their words of encouragement and thorough insights regarding the details of my thesis. All members of my committee have provided valuable feedback and strong suggestions that have bolstered my thesis accordingly. Without all members of my committee, this thesis would not have been possible, and I have learned tremendously from everyone. Additionally, I would like to thank my honors department advisor Dr. Leah Roesch for her considerable guidance throughout my undergraduate research experience. In all aspects of research, she has dedicated her time and energy to help me grow as a researcher. I truly express my warmest thanks to everyone for their contribution and support!

My sincerest thanks also goes to Dr. Lisa Krishnamurthy who has helped me tremendously throughout my research journey. She has helped me navigate various computational and scientific methods and has always been very supportive throughout this process. I would also like to thank Dr. Bruce Crosson for initially giving me the opportunity to work with Dr. Venkatagiri Krishnamurthy and for his expertise and guidance regarding the data utilized in this study. These members have played a vital part in not only my thesis but also my growth overall as a researcher.

I would also like to thank my fellow lab mates Joo Hee Han and Clara Glassman for their help in various aspects of my thesis. They have provided considerable contributions to this paper by guiding me through a multitude of computational methods. As peers, they both have also provided immense support mentally and emotionally. I appreciate all of the time, energy, and support that they have offered throughout this journey.

Last but not least, I would like to thank my family: my parents Chang Hoon Song and Jung Suk Song, for believing in me and for continuously providing support with the warmest heart and the strongest faith. Every step of not only this thesis but also my life would not have been possible without their guidance, love, and trust!

Table of Contents

Introduction

Aphasia

Aphasia is an acquired language impairment most commonly caused by stroke or other forms of brain trauma (Damasio 1992). This disorder affects the ability to produce and comprehend language both orally, auditorily, and visually (Sidman 1972, Le et al. 2016). Abilities in speech, reading, and writing are therefore affected, with specific impairments noted in semantic, grapheme, phonetic, imagery, as well as syntax language skills (Yu et al. 2017). Given the complexity of language, many types of aphasia exist where a certain aspect of language is impaired. For example, Broca's aphasia involves impairments in speech production whereas Wernicke's aphasia involves impairments in speech comprehension (Kertesz 1993). The symptoms of these types of aphasia are thus drastically different and these differences are attributed to differences in lesion location. For example, left-lateralized frontal damage (Keller et al. 2009) may give rise to speech production deficits while more temporal damage (Dewitt and Rauschecker 2013) gives rise to speech comprehension deficits. Additionally, the size of the lesion has also been found to be associated with not only aphasia severity, but also the extent of recovery (Plowman et al. 2011). Considering these variances, the lesion location, size, as well as type of brain trauma (i.e. ischemic stroke, infection, neurodegeneration) affect how language impairments may manifest in a person with aphasia (PWA) to differing degrees in terms of function and severity.

As one can imagine, language impairments can be devastating as a bulk of our day-to-day involves verbal communication. As social beings, the ability to reciprocally communicate is a vital part of our well-being and is highly correlated with quality of life and self-efficacy (Cruice et al. 2010). Studies have shown that the inability to effectively communicate can have great

detriments on not only an individual's standard of living but also their overall health (Babbitt and Cherney 2015). This leads to consequences such as increased healthcare costs and higher mortality rates (Ellis et al. 2012). Oftentimes, PWA also report psychiatric comorbidities due to the underlying frustrations and difficulties in expressing one's needs and wants (Morrow et al. 2022). Perhaps most notably, people with aphasia are often diagnosed with depression, generalized anxiety, and other psychiatric disabilities which sprouts from increasing self-doubt and feelings of incompetence (Baker et al. 2019, Code and Herrmann 2003). For this reason, PWA must be properly treated in order to minimize the negative impacts of communication impairments.

Given that aphasia occurs as a result of head trauma, the brain is in a critical stage of recovery where many restorative mechanisms are in action. During recovery, neuronal representations within the brain are undergoing relateralization and overall plasticity (Thompson et al. 2010). Studies have shown, however, that there is a "window of plasticity" that dictates the extent to which neuronal functions will be restored after the onset of brain trauma. Oftentimes aphasia severity and type are greatly altered within the first year after onset (Pederson et al. 2004). After this time frame, it is difficult to observe noticeable differences in aphasia severity and overall language recovery. Given that treatments for language recovery are often a long-term process involving multiple hours of repeated sessions, this limited time frame can pose problems if treatment plans are not effective. Moreover, studies have highlighted that the amount of individual variability in response to treatment interventions depend on the type of aphasia (Bakheit et al., 2007). The combination of diversity in aphasia, small "window of plasticity", and uncertainty in treatment response exacerbates the need to deliver the most effective treatment

plan that has been individualized to a patient's specific needs soon after the incidence of stroke so that the optimal level of recovery can be achieved.

Diversity of Language Treatments Available

Due to the extent of aphasia types as well as the multiplicity of underlying mechanisms involved in language, treatment interventions for aphasia are diverse and typically cater towards a specific aspect of language impairment. Typically, the treatment interventions for a person with deficits in speech production will greatly differ from someone who has deficits in speech comprehension ultimately perpetuating the need for specialized, impairment-specific therapies. As language is a complicated and multi-faceted function, both speech production and comprehension are needed to communicate effectively.

Aphasia can be treated with methods involving biological and behavioral approaches or even adjuvant methods. Biological approaches involve neural remediation through methods such as neural stimulation (Breining and Sebastian 2020), cell transplantation, and neurotrophins (Small and Llano 2009). These approaches are often paired with behavioral therapies as the brain is incredibly plastic and receptive to environmental stimuli, especially in the context of recovery (Horgn and Sur 2006). The most commonly used behavioral method to treat aphasia involves speech and language therapy, however, a variety of other methods exist. For example, music intonation therapy has been widely used to treat patients with Broca's aphasia as the brain recruits different areas of the brain for singing as opposed to speech (Norton et al. 2009). The act of singing however greatly mimics the motor activities involved in speech, which may facilitate the reorganization of neural mechanisms for language production within the brain. Other examples of language interventions involve reading treatments (Cherney et al., 1986), technology-based support (Nichol et al., 2022), word retrieval treatments (Raymer and Cohen,

2006), and many more. PWA may also undergo treatments targeting more domain-general functions that support language such as attention and working memory. Given the extensive utility and diversity of treatments for aphasia, it may be hard to find the right treatment plans for optimal recovery, and PWA may end up wasting time and money to find an effective treatment on a trial-and-error basis. A potential solution that could help clinicians mitigate negative treatment responses could be to discover biomarkers that could potentially predict treatment responses in PWA utilizing neuroimaging and computational techniques.

Imaging-Based Prediction Models for Aphasia

Advancements in neuroimaging have made it possible to understand and observe underlying brain neurophysiology and functional activity. For example, functional magnetic resonance imaging (fMRI) during a language task provides valuable information regarding the location of brain areas involved in language functions (Harnish et al. 2015). Additionally, for mechanisms such as language, this functional brain activity can be a marker for neuroplasticity by revealing information about brain reorganization within longitudinal studies (Meinzer et al. 2013). Recent advancements have also utilized resting-state fMRI as a measure of functional connectivity changes between various brain regions (Meszlenyi et al. 2017). These examples highlight how clusters of brain activity, revealed in neuroimaging methods, can be valuable in understanding long-term neural changes. Considering PWA whose brain exists in an altered state, this information is imperative in not only understanding but also predicting language function over time. Altogether, the current literature points to the promising uses of neuroimaging techniques as a method to understand and potentially predict responses to treatment in PWA. This may aid the filtration of interventions so that the most effective treatment regimen may be administered for optimal recovery.

A variety of predictive functional neuroimaging methods have been developed over the past couple decades. One such method is voxel-based lesion-symptom mapping (VLSM) which discovers relationships between behavioral deficits and lesion sites to identify predictive biomarkers of language recovery (Bates et al. 2003). VLSM utilizes a mass-univariate approach to independently analyze correlations between brain activity and behavior at a voxel level. One study utilizing VLSM discovered that lesions within the inferior and middle frontal gyri predicted a positive response to speech entrainment therapy in PWA (Fridriksson et al., 2015). A caveat, however, to this approach is the isolated nature of the analysis. VLSM involves lesion exclusive analysis in creating brain-behavior relationships. Language, however, involves multiple areas of the brain in a network fashion due to the inherent complexity and attentional demands of speech. Not only does VLSM not consider network-level information, but brain areas around the lesion are often compromised due to low blood flow as a result of head trauma (Crinion et al., 2013). Given that functional imaging is the most common way to obtain brain data for predictive analysis, this decrease in blood flow may not effectively reflect BOLD changes despite the presence of neural activation (Altamura et al., 2009). Combining this lack of network information and lack of sensitivity, VLSM omits the holistic details needed to create valid language-network prediction models. Overall, these limitations pose problems in the efficacy of previously discovered predictive biomarkers, and efforts towards developing holistic analyses to predict treatment response are needed.

Computationally, VLSM mass-univariate approaches pose further limitations. For example, the possibility of obtaining false positives increases as thousands of regression analyses are being performed in this voxel-by-voxel analysis (Mirman et al. 2017). Additionally, mass-univariate approaches do not account for voxel-to-voxel relationships and instead treat each

voxel as an independent unit (Wilson 2018). Neurobiological, however, adjacent regions in the brain are not purely isolated and instead are context-dependent where voxel-to-voxel interactions hold significance (Baldassano et al. 2012). For this reason, multivariate methods have been developed to account for relative voxel interactions. This analysis computes brain-behavior correlations for each voxel simultaneously which allows for voxel-to-voxel relationships to be accounted for (Ivanova et al. 2021). Overall, multivariate analysis overcomes the limitations of mass univariate analysis by sensitizing for more network-level associations within neuroimaging data.

As functional task fMRI offers neurobiological inspection of lesion impact on the whole language network, a potential solution to the VLSM approach could be to utilize whole-brain task-fMRI data within the brain-behavior relationship analysis. Therefore, the goal of this project is to utilize whole-brain task fMRI data within a prediction model in order to capture holistic network-level brain activation using a multivariate methodology. Given the computational demand as well as limited approaches involved in conducting multivariate analysis, various optimizations will be made in order to feasibly utilize whole-brain task fMRI brain data in the prediction model. The prediction design will be a correlational analysis focused on analyzing the relationship between patterns of pretreatment brain activity across multiple PWA and improvements in language behavior over time. The language treatment involves a standard language therapy approach utilizing picture naming and category exemplar generation coupled with either an intention treatment or control treatment. Overall, the goal is to help clinicians predict rehabilitation responses in PWA to ultimately aid the development of effective rehabilitation regimens for aphasia patients.

Hypothesis

Aim 1: To develop and optimize multivariate-based predictors in order to analyze significant relationships between whole-brain task fMRI and post-treatment behavioral data.

Aim 2: To assess the reliability of multivariate analysis by analyzing how well it converges with conventionally used predictors such as mass univariate analysis.

Aim 3: To validate both multivariate and mass univariate analyses using another behavioral measure.

Aim 4: To sensitize the multivariate method to predict treatment-specific language changes.

Hypothesis: Multivariate analysis of whole-brain task fMRI provides the feasibility to extract brain-behavior relationships that are predictive of treatment-induced changes.

Methods

Participants

A total of 19 participants were recruited for the study. These participants underwent a pretreatment screening procedure to determine eligibility. Inclusion criteria were the ability to follow one-step commands, English as the first language, as well as right-handedness which was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Aphasia severity was also assessed and participants were required to score between 4 to 45 correct items on the Boston Naming Test (BNT), and fall below <93.8 (the threshold for normal language function (Kertesz 2006)) on the Western Aphasia Battery Aphasia Quotient (WAB-AQ) (Shewan & Kertesz, 1980). Exclusion factors included the presence of other neurological or psychiatric disorders, the presence of learning disorders such as dyslexia, the inability to be scanned in an MRI, and alcohol/drug abuse within the past year. Five participants did not pass the pretreatment screening; therefore, only 14 participants participated in the study. All participants provided written informed consent according to the procedures outlined by the University of Florida Health Science Institutional Review Board. All participant data were collected >10 years ago and are de-identified for post-analysis to be conducted by institutions such as the Center for Visual and Neurocognitive Rehabilitation at the Atlanta VA Medical Center. Participant demographics are outlined in Table 1. Further MRI scans and medical records indicated that all subjects were six months post-stroke with brain lesions located in the left hemisphere concentrated in both cortical and subcortical regions. All subjects were randomly assigned to either the intention treatment (INT) or control treatment (CT) group. There were no significant differences in age, gender, education, BNT, and WAB-AQ scores between the two groups. The INT group consisted of 5 female and 2 male participants with an average age of 72.1 (standard deviation = 10.5), average years of education of 14.9 (standard deviation = 2.5), a 5/2 ischemic to hemorrhagic

cerebrovascular accident type ratio, an average BNT score of 24.7 (standard deviation = 13.4) and an average WAB-AQ score of 65.5 (standard deviation = 8.3). The CT group consisted of 1 female and 6 male participants with an average age of 63.0 (standard deviation = 9.2), average years of education of 12.9 (standard deviation = 1.1), a 6/1 ischemic to hemorrhagic cerebrovascular accident type ratio, an average BNT score of 30.9 (standard deviation = 6.3) and an average WAB-AQ score of 71.9 (standard deviation = 11.8).

	ID	Age	Gender	Education (Years)	CVA Type	Lesion Location	BNT	WAB-AQ	WAB Classification
Intention Group (Gesture)	S01	79	Female	12	H	Subcortical	8	51.2	Broca's
	S03	92	Female	12	I	Parietal	29	69.1	Conduction
	S05	73	Male	14	I	Insula	38	76.4	Anomic
	S06	62	Female	18	H	Subcortical	16	57.8	Conduction
	S11	68	Male	14	I	Parietal, Temporal	28	67.8	Conduction
	S12	63	Female	18	I	Insula, Temporal	11	66.6	Conduction
	S15	68	Female	16	I	Parietal	43	69.4	Broca's
Control Group (No Gesture)	S04	53	Male	12	I	Parietal, Temporal	27	65.8	Conduction
	S07	80	Female	14	I	Insula	40	78.0	Anomic
	S08	55	Male	14	I	Frontal, Parietal, Subcortical, Temporal	31	67.1	Transcortical Motor
	S10	69	Male	12	I	Frontal, Parietal, Subcortical	32	52.2	Broca's
	S14	61	Male	12	I	Subcortical	30	75.8	Anomic
	S16	59	Male	14	H	Frontal, Insula	30	74.5	Anomic
	S19	64	Male	12	I	Insula	37	90.0	Anomic

Table 1. Participant Demographics

fMRI Acquisition

A Philips 3 T Achieva scanner utilizing a body coil for radio frequency (RF) transmission and an 8-channel head coil for RF receiving was used to obtain the fMRI data. Foam pads were used to stabilize the participants' heads to minimize motion during scans. High resolution T1-weighted anatomical images were acquired using turbo field echo acquisition. An echo time (TE) of 3.7 ms, repetition time (TR) of 8.1 ms, field of view (FOV) of 240x240 mm^2, flip angle (FA)

of 8°, and matrix size of 240x240 was used for this acquisition. These structural images were obtained for spatial normalization to the Montreal Neurological Institute (MNI) template space. Functional fMRI images were acquired to capture brain activity, represented by changes in the blood-oxygen-level-dependent (BOLD) signal, during a semantic fluency task. Six runs of the task fMRI time course consisting of BOLD weighted echo-planar (EPI) sequences were acquired with a slice thickness of 4mm, FOV of 240x240 mm^2, matrix size of 64x64, TR of 1700 ms, TE of 30 ms, FA of 70°, and a total of 186 volumes per run.

Task Design

Participants underwent pretreatment task fMRI to collect brain activity during word retrieval. Category exemplar generation (CEG), a semantic fluency task, was chosen because it most closely parallels word selection demands in conversation. During the scan, participants were visually and auditorily presented a category (e.g., "color) in which they were instructed to orally generate an examplar (e.g., "blue). Each task trial length was 6.8 seconds and six runs each containing 10 trials were conducted for a total of 60 trials. During intertrial intervals (ITI) that alternated between 13.6, 15.3, and 17.0 seconds (Benjamin et al., 2014), participants were instructed to remain still and not speak while viewing a "+" symbol on a screen.

Treatment

Baseline Probes and Treatment

Before treatment, participants were asked to name 400 black-and-white line drawings of objects as well as generate exemplars for 120 categories (Henry et al. 2008, Howard et al. 1985, Marshall et al. 1990). The most frequently missed items (120 pictures and 60 categories) were selected as treatment items. Each treatment set included 75% consistently incorrect items as well as 25% consistently correct items to prevent participant discouragement throughout the sessions.

Thirty hours of treatment was administered in 2 hours/day, 5 days/week for 3 weeks. The treatment was delivered in 3 phases consisting of one phase per week. The sessions were at least half an hour apart. The first two phases of treatment consisted of picture naming sessions. During these sessions, the participants were instructed to name a series of black-and-white line drawings of objects. The last phase consisted of CEG sessions where participants were instructed to orally cite exemplars for a given category, similar to the fMRI task. Altogether, the use of picture naming and CEG sessions was motivated by restitutive treatments that are designed to rehabilitate language mechanisms by utilizing activities involving semantic and/or phonological representations of words (Maher and Raymer 2004, McClung et al. 2010).

Intention and Control Groups

Participants were split into two groups where one group experienced INT while the other group did not. Participants within the INT group had to lift the lid of a box and press a red button with their left hand to start the treatment trial during phase 1 and phase 2 (Figure 1A). A primary difference between phase 1 and phase 2 is that there is a flashing start with a beep in phase 1. Contrastingly, phase 3 is initiated by only a left-hand gesture. If the participant provided an incorrect response, the subjects were instructed to create a nonmeaningful circular gesture with their left hand while orally stating the correct answer (Figure 1B). This complex left-hand movement coupled with naming was designed to activate the intention mechanism within the contralateral hemisphere of the lesion (Crosson, 2008). Thus, the right-hemispheric activation initiated by intention is created to aid the activation of frontal cortical regions involved in language to ultimately guide the right lateralization of language functions. On the other hand, the participants within the CT group did not have to undergo these left-hand actions. Instead, their correction procedure only involved the oral repetition of correct answers.

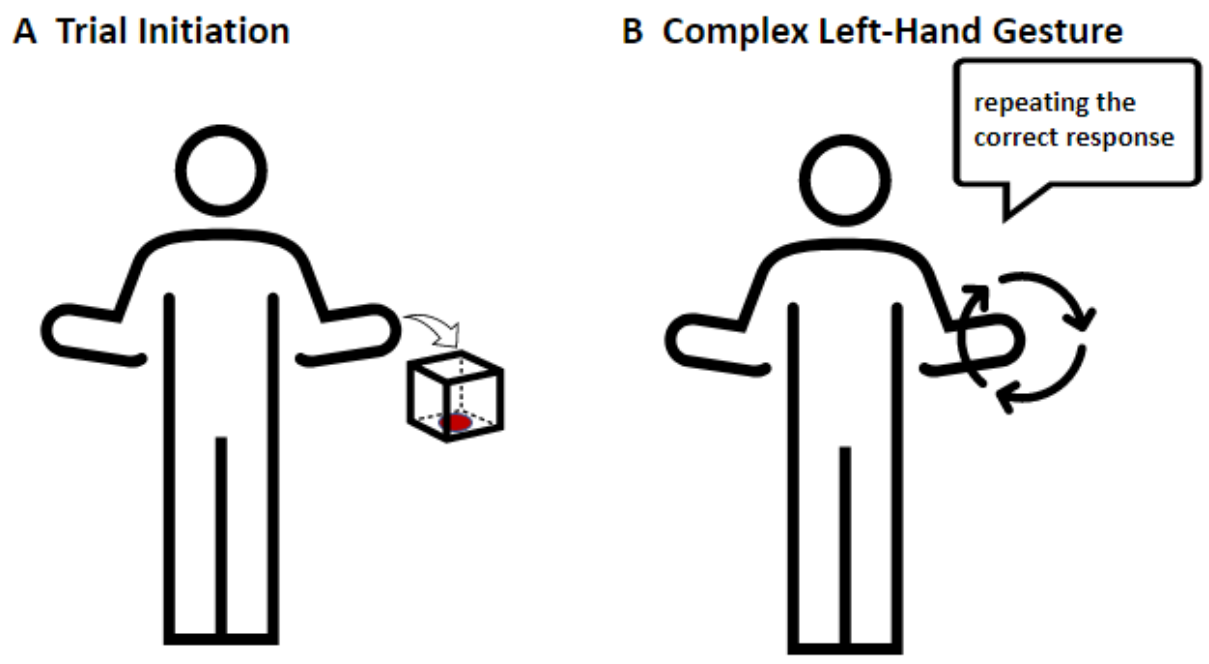


Figure 1. A visual representation of the INT treatment. (A) The participants were instructed to lift the lid of a box and press a button with their left hand in order to start each trial during phase 1 and phase 2. (B) The participants were instructed to make a nonmeaningful circular gesture with their left hand when orally repeating correct responses to trials.

Behavioral Assessments

Before treatment as well as two weeks after the last treatment session, all subjects underwent behavioral examinations of language function. Multiple behavioral examinations were conducted including a reexamination of picture naming and CEG, language discourse assessments, functional outcomes questionnaire (Glueckauf et al., 2003), etc. Only CEG, however, was selected as the primary measurement of behavioral change for the prediction analysis since it was also used as the task given during the fMRI scan. Specifically, the CEG z scores obtained from Tyron's C statistics were used in the prediction model. Tryon's C statistic provides information regarding the presence of a differential trend in, for example, a treatment series collected after baseline compared to the baseline data. This statistic accounts for changes

between baseline and treatment measurements, and a significant result indicates that the treatment series differs meaningfully from the baseline series (Tryon 1982). The statistical significance of the C statistic is given by the z score which is defined by the ratio of the C statistic to the standard error (Tryon 1982). Overall, the z scores for CEG from baseline to two weeks post-treatment (Figure 2A) were collected for all subjects and used as the primary behavioral measurement in the prediction model.

An additional behavioral measure that differed from the fMRI task was utilized to assess if treatment responses involving other language functions can be predicted. For this measurement, language discourse scores were chosen as this behavior models a more holistic approach in defining language functions by analyzing the connectivity and cohesiveness of speech. During the language discourse examination, participants were asked to describe a given picture as well as answer open-ended questions. Within their response, a combination of quantitative (number of words (W), verbs (V), nouns (N), etc.) and qualitative (number of correct information units (CIU), new information (NI), grammatically correct phrases (GRAM), etc.) linguistic characteristics indicative of connected speech was recorded. These assessments were conducted before as well as 2 weeks after treatment (Figure 2B). Unlike the behavioral scores for CEG, Tyron's C statistics and z-scores were not utilized, and instead percent change values from pre to two weeks post-treatment were calculated. The behavioral scores yielding the greatest percent change values were analyzed in order to find meaningful changes from the raw scores. Such behavioral data included discourse CIU, GRAM, N, and V. In order to assess how task fMRI data may provide more network-level information within the prediction models, discourse GRAM was chosen as the second behavioral measure as grammar involves complex understandings of not only semantics but also word order, tense, plurality, possessiveness, etc.

Grammatically correct phrases that were positively scored were based on the rules of subject-verb agreement. Many studies have shown that grammar and lexical semantics are closely tied (Bates 2003) and patients with aphasia who undergo grammar-focused treatments show significant improvements in semantics with a high level of generalization to linguistically similar lexicons and sentence types (Helm-Estabrooks and Ramsberger 1986, Thompson and Shapiro 2005). As grammar is a complex language function, multiple areas of the brain are involved during this process in a network-like fashion (Friederici and Gierhan 2013). Thus, the ability of network-level baseline brain activity to predict network-level language functions such as grammar in multivariate methods can be assessed by utilizing discourse GRAM.

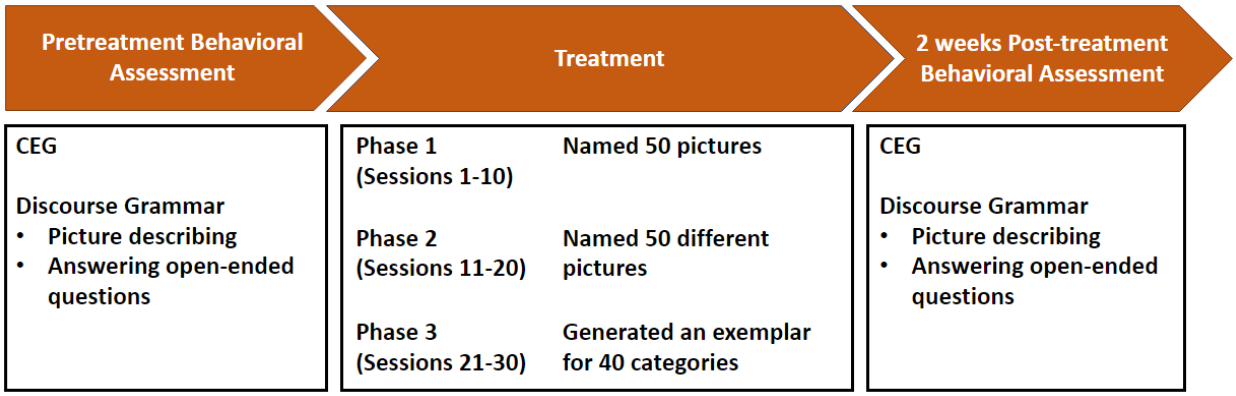


Figure 2. A schematic representation of the treatment and behavioral assessment timeline.

Image Processing

Structural and Task-fMRI Preprocessing

The T1 weighted structural images are denoised using an optimized nonlocal means (ONLM) filter and then bias field corrected. A binary lesion mask was then estimated utilizing an automated segmentation algorithm called Lesion Identification with Neighborhood Data Analysis (LINDA) (Pustina et al., 2016). The skull was then stripped from the images by applying a binary brain mask generated by an optimized brain extraction script (Lutkenhoff et

al., 2014) to the T1 weighted image. To assure complete removal of meninges, the brain mask was manually edited using ITK Snap, a software that allows users to manually manipulate the inclusion or exclusion of certain voxels, before application to the structural image. Lastly, chimera spatial normalization was conducted to increase the efficacy of MNI transformation.

The functional EPI images for each subject were initially slice-time corrected and head motion-corrected using various in-house scripts and programs such as AFNI, a software for processing and displaying functional MRI data, to quality check each step. The BOLD data was then denoised for task correlated motion (TCM) involving not only spatial artifacts created due to movements related to the task but also temporal artifacts present within the task-induced hemodynamic response function (Gopinath et al. 2009). In this case, any speech-related motions such as facial/temporal movements would be classified. This classification occurred using independent component analysis (ICA) and additional methods iterated in Krishnamurthy et al., 2021. The denoised images were then co-registered to the T1 weighted images using a software consisting of tools for fMRI brain imaging called FMRIB Software Library (FSL) using a boundary-based registration algorithm. The images were then warped to MNI space with a voxel size of 2 mm^3 using tools in FSL such as FLIRT which involves linear registrations and FNIRT which involves nonlinear registrations for more fine-tuned warping. Additionally, a binarized cerebrospinal fluid (CSF) mask inclusive of the ventricles and lesioned brain area, all in MNI space, was generated. This mask was applied when the BOLD images were spatially smoothed (Gaussian kernel size = 6 mm) to minimize CSF contamination. Images were then resampled using an AFNI command to a voxel size of 4x3.75x3.7 mm^3. The BOLD time course was then scaled to the initial baseline condition, consisting of 7 TRs lasting 11.9 seconds, and then deconvolved via AFNI's 3dDeconvolve command. During deconvolution, general linear model

(GLM) regression analyses are used to best estimate the hemodynamic response function (HRF) within a set duration characterized by TENTs. In this case, the HRF duration was set as 18.7 which is equivalent to 12 TRs or in other words 11 beta weights. The area under the curve for the estimated HRF for each participant was then computed and Z-transformed (ZAUC) in order to normalize the distribution. This normalization was conducted using in-house scripts. The ZAUC was then extracted for all subjects in order to represent the average brain activity for a given cluster of voxels within the correlational analysis.

Multivariate Analysis

The multivariate analysis used in this study consisted of Sparse Canonical Correlations Analysis for Neuroimaging (SCCAN) through an R package called LESYMAP (Pustina et al., 2018) to analyze brain-behavior relationships. The benefit of using SCCAN is that it gives information regarding the level of contribution each voxel has towards a correlational analysis. In this case, the relationship between the task fMRI data and behavioral data is being assessed and the amount of contribution a voxel has in this relationship is given by a numerical value (i.e. "weight"). Additionally, an optimal sparseness value that determines the amount of neuronal representation a cluster of voxels has within a spectrum ranging from local/sparse representation (Figure 3A) to distributed representation (Figure 3B) is found through a series of cross-validation analyses (Quiroga and Kreiman 2010). The optimal sparseness, therefore, dictates how extensive the results should be by influencing the number of available significant voxels (Avants et al. 2014). Considering these factors, multivariate analysis utilizing SCCAN determines meaningful weights for a voxel where the level of significance is based on an optimal sparseness value.

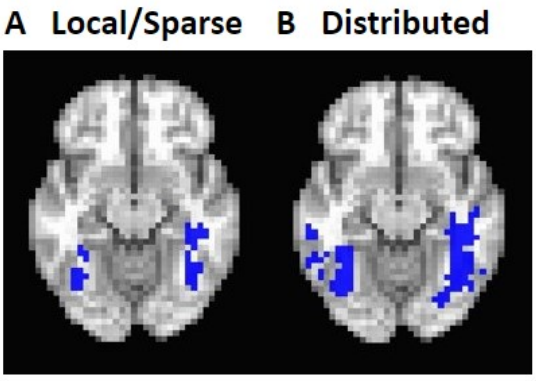


Figure 3. More local/sparse representation involves (A) less significant clusters while more distributed representation involves (B) more significant clusters.

To find the optimal sparseness, LESYMAP runs an internal k-fold cross-validation (CV) technique which splits the dataset into a training and test model. The total amount of training and test models are based on the number of folds specified by the user. By default, LESYMAP is set as a 4-fold CV where the dataset is split into four sets with each set containing a random but evenly distributed number of subjects (Figure 4A). One of these sets is allocated as the test set and the other three become the training set. The training sets identify the weights of the voxels based on a sparseness value randomly determined by SCCAN, while the test set is used to create a brain-behavior prediction model using those weights. This brain-behavior prediction model is then tested for accuracy by comparing the model's expected predictions to the actual behavioral scores corresponding to the test set (Figure 4B). Furthermore, the p-value is calculated and maintained at a certain threshold (default is 0.05) and the solution is dropped if this value falls below the threshold at any point returning an empty brain map with no significant voxels. Assuming the p-value stays under the threshold, LESYMAP repeats the analysis while alternating the allocation of the sets' roles until every combination of the test set and training set has been computed. These repeated analyses are conducted using different sparseness values that are randomly assigned by LESYMAP if it is not pre-determined by the user. From this series of

computations, the most optimal sparseness value is determined from the brain-behavior prediction model with the most accuracy (Figure 4C). The optimal sparseness value is then used in a final SCCAN run, and a raw weights image representing voxels with significant correlations is obtained (Figure 5A-B). This image is then normalized by the absolute greatest raw weight value in order to obtain a raw weights range between -1 and 1 since the values of the raw weights are very small (i.e. 0.000000003) (Figure 5C). Then voxels with small weight values within the range of -0.1 to 0.1 are removed. The normalization and thresholding are done internally in the LESYMAP algorithm and a final statistical image is created (Figure 5D). This image highlights meaningful voxels that have been normalized and thresholded for significance.

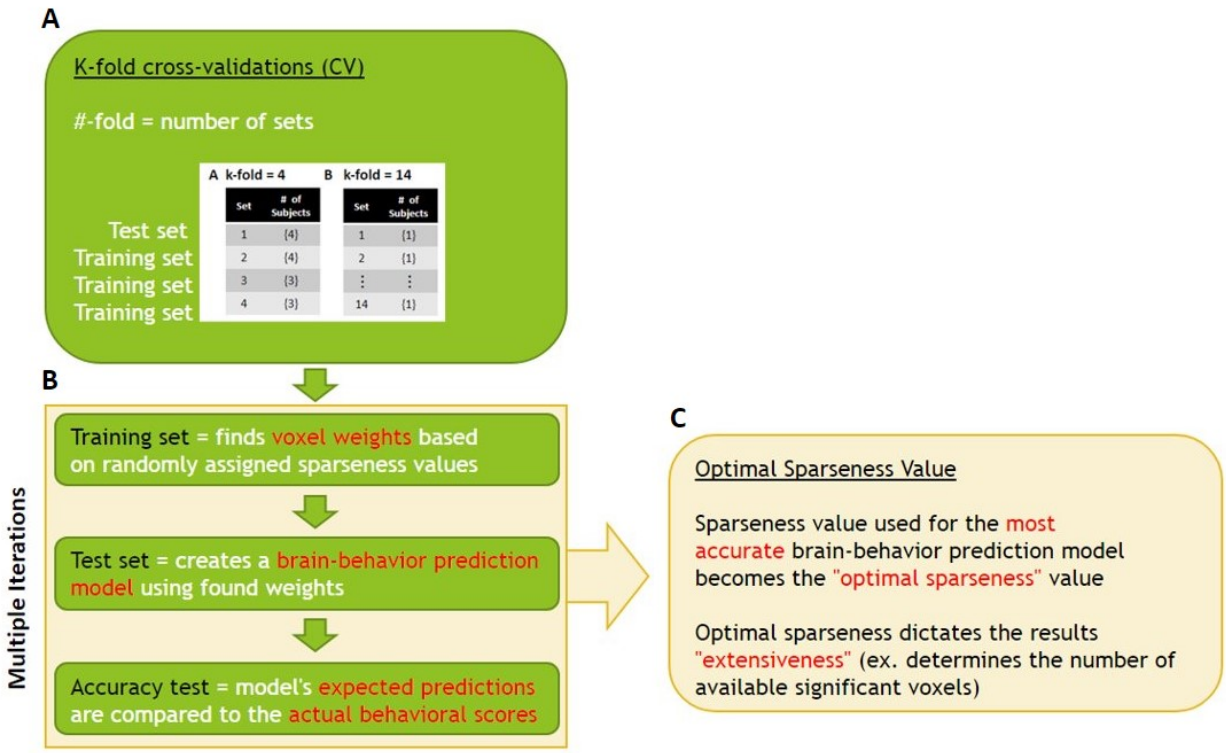


Figure 4. A schematic representation of how the optimal sparseness value is obtained

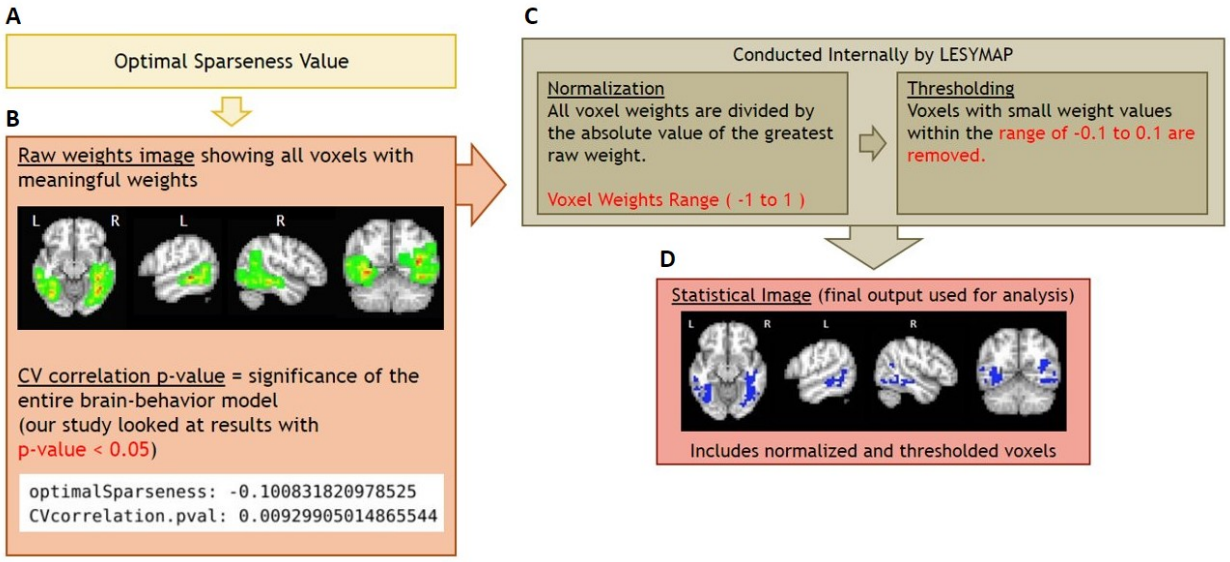


Figure 5. A schematic representation of how the statistical image is obtained

A caveat to k-fold CV is that the stability of the solution greatly varies depending on the number of folds. Typically, a smaller number of folds is used for datasets that are larger to decrease computation time. However, due to the small number of participants in this study, all possible k-fold combinations from 2-fold to 14-fold were conducted and analyzed to find the k-fold that offered the most robust prediction model. In doing so, identical LESYMAP runs were conducted three times each for the different k-fold values for a total of 36 runs. The number of significant voxels within the statistical image as well as the optimal sparseness value tended to vary during each run despite having the same k-fold setting. The only k-fold, however, where this variance was not seen was when a k-fold equivalent to the total number of subjects in the analysis was used. Thus, a k-fold of 14 yielded (Figure 6B) the most robust data for whole data analysis as the number of significant voxels within the statistical map (Figure 7A) as well as the optimal sparseness value (Figure 7B) was consistently the same for all three runs. For group-level analysis investigating the prediction model in INT versus CT groups, a 7-fold CV was used during multivariate analysis. Although the computation time may increase, these k-fold settings

were found to produce the most statistically robust results, given our small dataset.

K-fold Setting

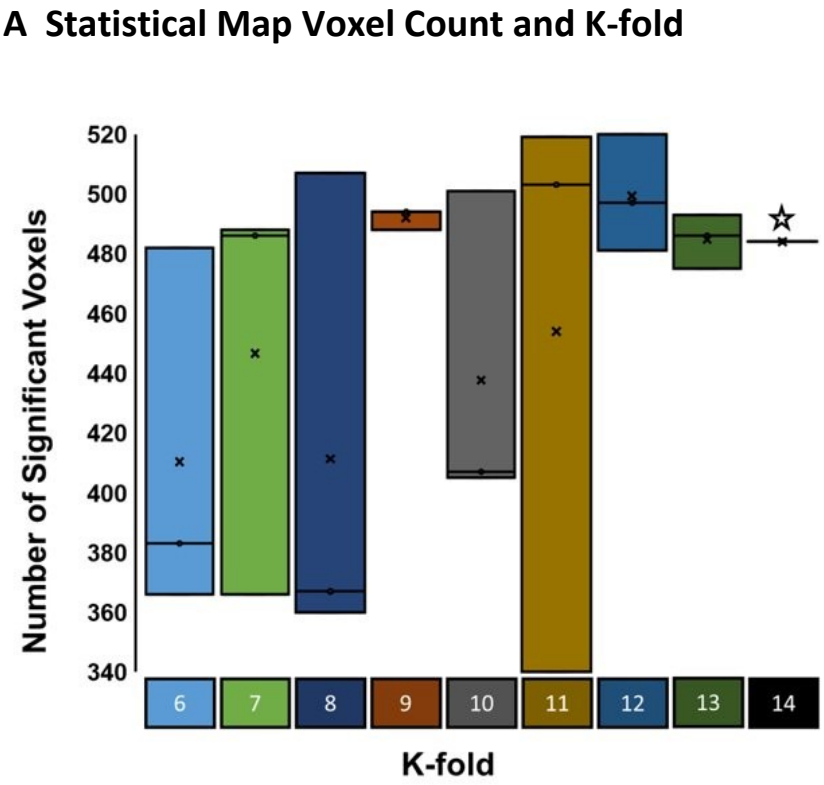
A k-fold = 4

Set	# of Subjects
1	{4}
2	{4}
3	{3}
4	{3}

B k-fold = 14

Set	# of Subjects
1	{1}
2	{1}
⋮	⋮
14	{1}

Figure 6. A visual representation of default versus optimal k-fold analyses. (A) The default k-fold setting for LESYMAP was 4-fold. (B) The optimal k-fold setting for our study was a k-fold equal to the number of subjects used within the analysis (i.e. 14 for whole data analysis).



B Optimal Sparseness and K-fold

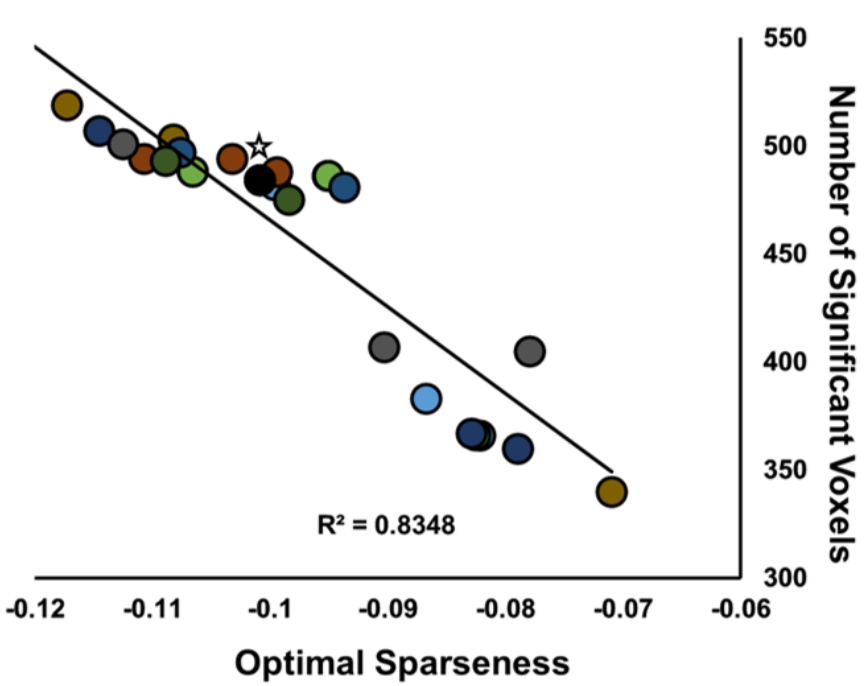


Figure 7. 14-fold CV (during whole group analysis of the primary behavioral measure) yielded the most robust data. (A) The number of significant voxels presented within the statistical image varied across three runs for all k-folds except for when the k-fold was equal to the number of subject data inserted into the analysis (i.e. 14-fold CV for whole data analysis, 7-fold CV for group-level analysis). This lack of variance is highlighted with a star. (B) The optimal sparseness also did not vary for a 14-fold CV (indicated by the star above the black circle). All dots are color-coordinated according to the k-fold from figure 1A.

An additional parameter contributing to multivariate analysis is the brain mask. This mask is applied to the images and ultimately defines the voxels of interest. For the purposes of this study, a brain mask encapsulating the whole brain instead of a singular cluster is needed. Originally, if a brain mask is not input into LESYMAP, a default brain mask will be automatically computed and applied. When this occurred, the generated default brain mask was

not inclusive of the entire brain, and instead included spurious patches of empty voxels throughout the image (Figure 8A). Thus, a structural brain mask including voxels representing at least four subjects was created using ITKsnap and used for the analysis. A representation of at least four subjects was chosen on the basis of balancing whole-brain coverage with statistical power. As all fourteen participants are aphasic, a bulk of the left hemisphere is lesioned (Figure 8B). This poses a problem as some subjects will contain empty voxels within the left hemisphere. Thus, a brain mask with voxels of greater representation (i.e. at least 14 subjects per voxel (Figure 8C)) yielded too many empty voxels within the left hemisphere. Alternatively, a brain mask with lower representation (i.e. at least 2 subjects per voxel) yielded insignificant results due to low statistical power. Thus, a brain mask with voxels representing at least four subjects was the most optimal for our analysis (Figure 8D). Compared to the default brain mask offered by LESYMAP, this customized mask better represents the whole brain while also mitigating problems with low statistical power in the computational analysis.

After configuring all multivariate optimizations, final clusters within the statistical map were viewed in AFNI and clusterized. All ROI clusters within the statistical image were corrected for multiple comparisons ($p<0.05$, cluster size=50). Altogether, the various inputs required for SCCAN analysis through LESYMAP involve the whole-brain task fMRI data, behavioral data, determined k-fold setting, a brain template, as well as a brain mask (Figure 9A). Certain inputs were manipulated according to our analysis, and the final inputs into our multivariate model are highlighted in figure 9B.

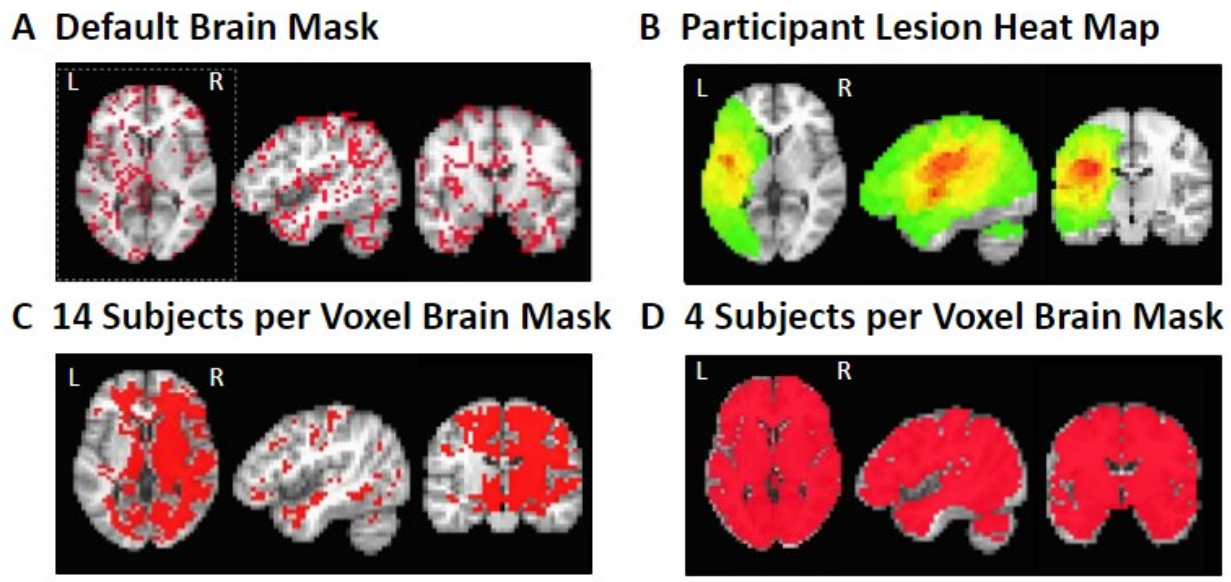


Figure 8. Brain Mask Optimizations (A) The default LESYMAP brain mask was not inclusive of the whole-brain (red = voxels that are counted within the mask). (B) A heat map representing the location of participant lesions reveals the extent to which empty voxels are present as well as the left-lateralization of participant lesions (red = more representation, green = less representation). (C) A customized brain mask with voxels representative of at least fourteen subjects is not inclusive of the whole brain (red = voxels that are counted within the mask). (D) A customized brain mask with voxels representing at least four subjects is more inclusive of the whole brain without lacking statistical power (red = voxels that are counted within the mask).

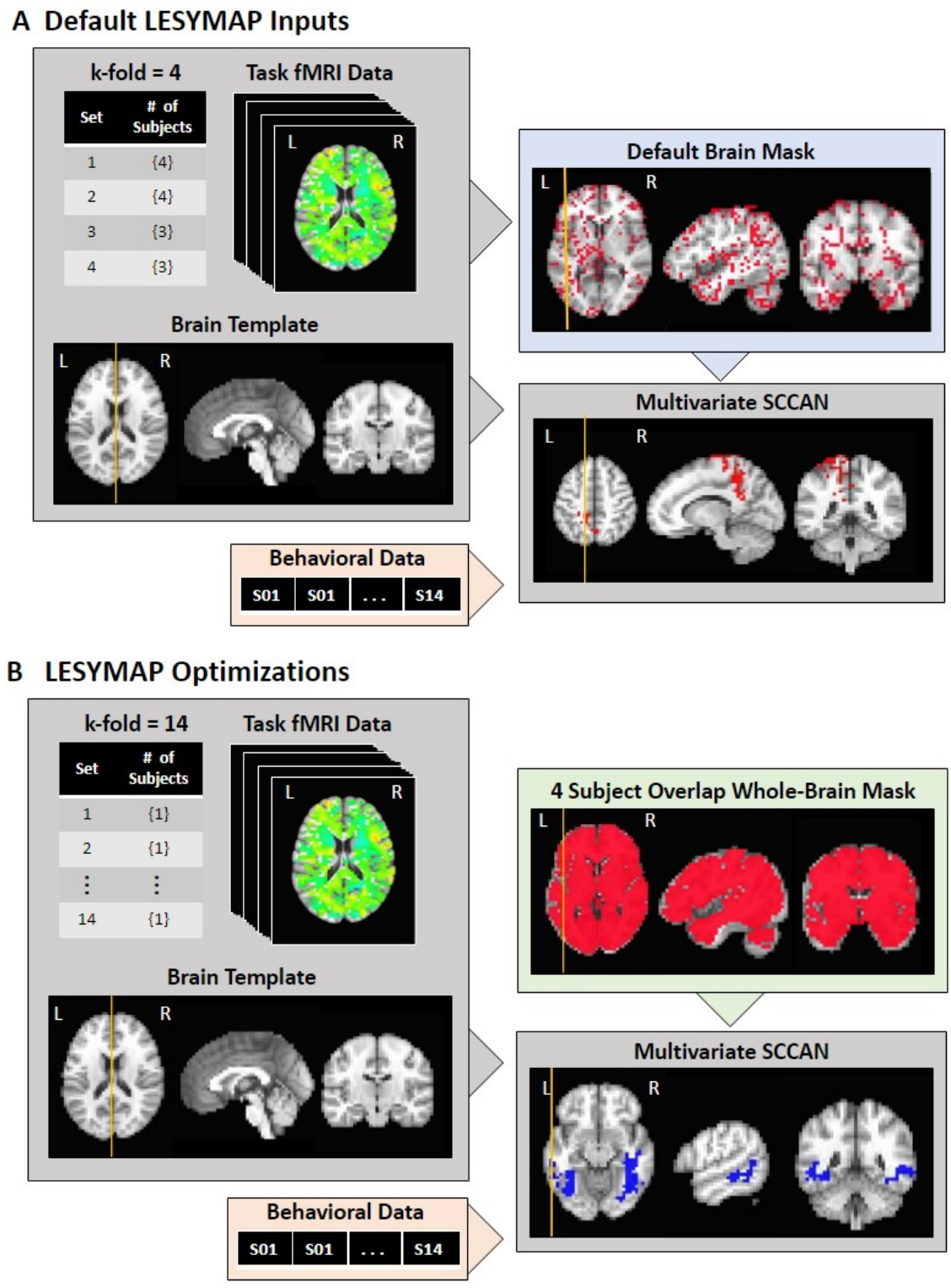


Figure 9. A schematic representation of the various inputs fed into the LESYMAP R package for multivariate analysis. (A) The default LESYMAP parameters consisted of a k-fold of 4 and an

automatically generated default brain mask. When inserting the task fMRI brain data and behavioral data, a multivariate SCCAN statistical map is obtained. (B) LESYMAP optimizations included a k-fold of 14 and a customized brain mask exclusively consisting of voxels that a representative of at least four subjects. These changes result in a multivariate SCCAN statistical map with different spatial clusters compared to the default setting.

Mass Univariate Analysis

In order to assess the reliability of the multivariate-based predictors, the brain-behavior relationships were compared to those obtained from more conventional prediction analyses such as mass univariate analysis. This method utilizes simple linear regressions to assess the relationship between BOLD activity, represented as ZAUC, and behavior on an independently assessed voxel-by-voxel basis. The independent variable in this analysis consisted of the brain data and the dependent variable consisted of the behavioral data. Mass univariate analysis was conducted using MATLAB and a mask inclusive of voxels representative of at least four participants was applied in congruence with the multivariate method. These results were observed spatially using AFNI and thresholded so that only voxels with a p-value ≤ 0.05 are displayed. This thresholding was conducted by setting an R^2 value in AFNI where $R^2 = 0.29$ for whole-data analysis (i.e. fourteen subjects) and $R^2 = 0.577$ for group-level analysis (i.e. seven subjects). Additionally, an overlay image representing the slopes of the linear regression is set so that positive or negative correlations between the brain and behavioral data can be visualized for each voxel across the brain. Clusters were corrected for multiple comparisons ($p<0.05$, cluster size = 50), and significant ROIs were isolated using a 5 mm sphere. Linear regressions specific to those ROIs were visualized by plotting the extracted ZAUC values to the behavioral data for all fourteen subjects.

Correcting for Lesion Size

Patients with aphasia often contain lesions of varying sizes. Studies have shown that lesion size is often correlated with aphasia severity measured by WAB-AQ and object naming (Thye and Mirman 2018). To account for the effects that lesion size may have in influencing the prediction models, both multivariate and mass univariate approaches were modified so that the influence of lesion size is regressed out of the brain-behavior prediction analysis. Initially, the lesion size for each subject was identified and recorded.

Within the multivariate approach, a function called "CorrectByLesSize" was modified from "none" to "both" in order to account for lesion size. The purpose of this function is to covariate out the effects of lesion size. The option "both" performs a regression analysis on both the brain activity data as well as the behavioral data in which they are residualized to account for lesion size. A caveat, however, is that LESYMAP is originally a lesion-to-symptom mapping software involving relationships between lesional data and behavior. Our analysis, however, is accounting for the entire brain without the lesion. The "lesion size" in our case is therefore equivalent to the entire brain without the lesion size which is inversely related as functional brain size decreases when lesion size increases. Thus, the in-build LESYMAP "lesion" correction would account for the inverse between lesion size and brain activity relationships as well as lesion size and behavioral data relationships for this study. In order to obtain brain-behavior relationships that have accurately been corrected for lesion size, the sign of the weights within the multivariate output must therefore be inversed.

Within the mass univariate approach, linear regressions between each variable (i.e. BOLD activity and behavior) and lesion size were assessed. For example, linear regressions between the BOLD activities within each voxel for all subjects and lesion sizes for all subjects

were calculated using MATLAB where lesion size was equivalent to the independent variable and BOLD activity represented the dependent variable. A trend line representing the relationship between these variables was found and used to calculate an estimated BOLD signal based on the independent variable (i.e. lesion size). The estimated BOLD signal is then subtracted from the actual BOLD signal to obtain the newly lesion corrected BOLD signal for each voxel. Thus, the final output from this analysis is fourteen images with voxels representing lesion corrected BOLD signal. The same correction was conducted for the behavioral data in excel where estimated behavioral data was subtracted from the actual behavioral data. Mass univariate analysis was then conducted using the lesion corrected BOLD signal and behavioral data.

Results

Aim 1: To develop and optimize multivariate based predictors in order to analyze significant relationships between whole-brain task fMRI and post-treatment behavioral data

As the primary behavior of interest is CEG, multivariate optimizations were performed using CEG z-scores in the prediction model. Given a small sample of subjects in this study, k-fold optimizations revealed that a k-fold equivalent to the total number of subjects within the analysis yielded the most robust data. With this setting, optimal sparseness values, as well as the number of significant voxels within the statistical map, did not vary between runs. Further optimizations involved customizing a brain mask that best encapsulated the entire brain without losing statistical power. Multivariate analysis with the optimized brain mask yielded drastically different results compared to the output from the default mask (Figure 10). Perhaps most notably, the areas of activation that are correlated to post-treatment responses varied. When the default brain mask is utilized, regions within the left posterior cingulate and the primary sensory cortex are found to significantly predict post-treatment outcomes (Figure 10A). This varies greatly from the regions highlighted in the optimized analysis involving voxels located in the bilateral middle temporal area (Figure 10B). Within this cluster, the right extrastriate area and the right fusiform gyrus are indicated as potential biomarkers in predicting the recovery of semantic fluency. Ultimately, these optimizations have made it possible to predict post-treatment responses from whole-brain task fMRI data.

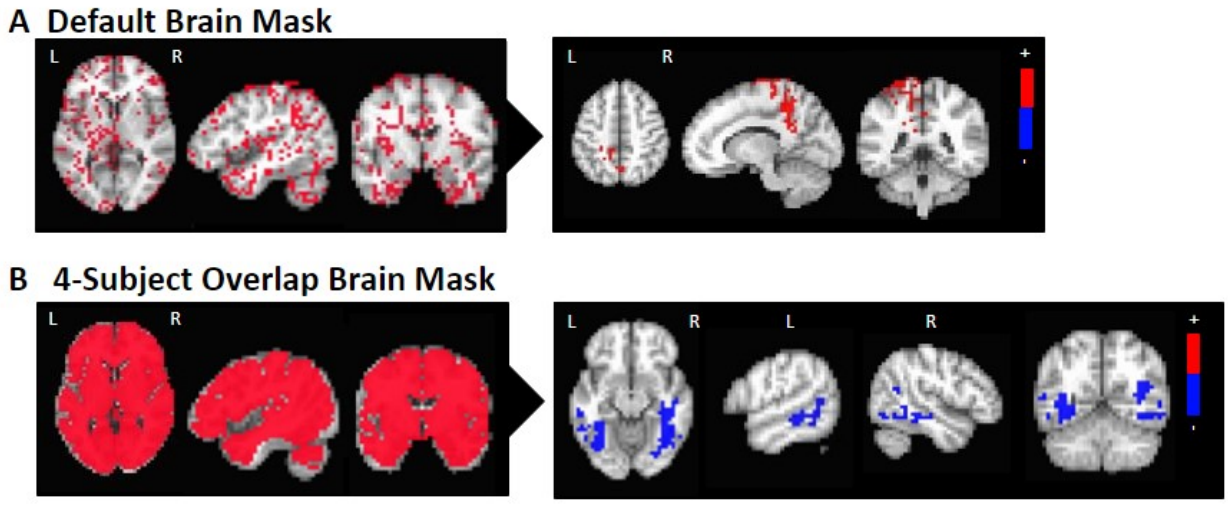


Figure 10. LESYMAP brain mask (shown in red on the left side images) optimizations involved changing the (A) default mask to a (B) 4-subject overlap brain mask that was more inclusive of the whole-brain. With this newly optimized mask, significant clusters ($p<0.05$) within the statistical image greatly differed in terms of location as well as correlational association (positive (red)/negative (blue)).

Aim 2: To assess the reliability of multivariate analysis by analyzing how well it converges with conventionally used predictors such as mass univariate analysis

In order to assess the reliability of multivariate-based predictors, mass univariate analysis was also conducted. When analyzing the relationship between baseline task fMRI and post-treatment CEG scores, mass univariate analysis yielded clusters of activation in regions overlapping with multivariate clusters (Figure 11). Multivariate analysis, however, revealed more significant voxels bordering the clusters highlighted in the mass univariate analysis. Both analyses, however, revealed negative correlations between baseline brain activity and CEG behavior.

Altogether, brain-behavior relationships between the task fMRI ZAUC and CEG behavioral scores were analyzed in distinct ROIs present within the multivariate analysis as well

as overlapping regions. Within the right extrastriate area, a multivariate found biomarker, brain activity negatively correlated with CEG z-scores (p-value = 0.049*) (Figure 12). Within the right fusiform gyrus, a biomarker found in both multivariate and mass univariate analysis, brain activity was also negatively correlated with CEG z-scores (p-value = 0.125), however, only marginally (Figure 12).

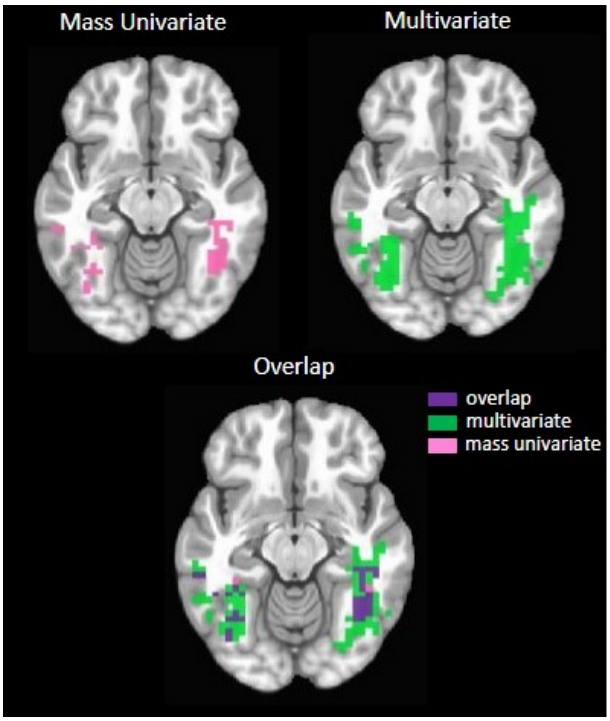


Figure 11. Mass univariate and multivariate predictors for CEG display spatial overlap. Mass univariate clusters are more conservative than multivariate clusters.

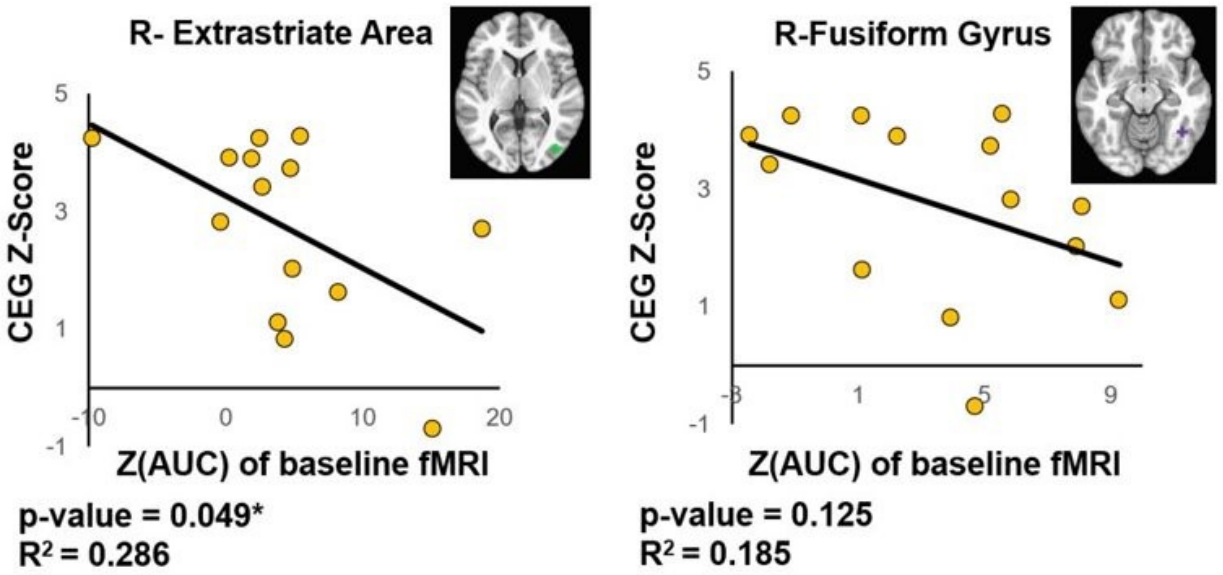


Figure 12. Distinct ROIs predict brain-behavior relationships for CEG.

Aim 3: To validate both multivariate and mass univariate analyses using another behavioral measure

The sensitivity and expanded utility of multivariate analysis was assessed by using another behavioral measure within the prediction model. Specifically, post-treatment discourse GRAM rehabilitation was assessed and results revealed that mass univariate predictors also converged spatially with the multivariate predictors (Figure 13). Both analyses revealed positive correlations between baseline brain activity and discourse GRAM behavior. Altogether, brain-behavior relationships between the task fMRI ZAUC and discourse GRAM behavioral scores were analyzed in distinct ROIs. Within the left medial frontal gyrus, a biomarker found in both multivariate and mass univariate analysis, brain activity was positively correlated with discourse GRAM scores (p-value =0.09), however, only marginally (Figure 14).

Additionally, various k-fold settings within the multivariate analysis involving discourse GRAM were also analyzed. Similar to the k-fold value that yielded the most robust results when predicting CEG outcomes, a 14-fold CV also yielded the most robust results. With this setting,

the number of significant voxels within the statistical map (Figure 15A), as well as the optimal sparseness values (Figure 15B), did not vary between runs. Other k-folds such as 10, 12, and 13 also yielded robust results.

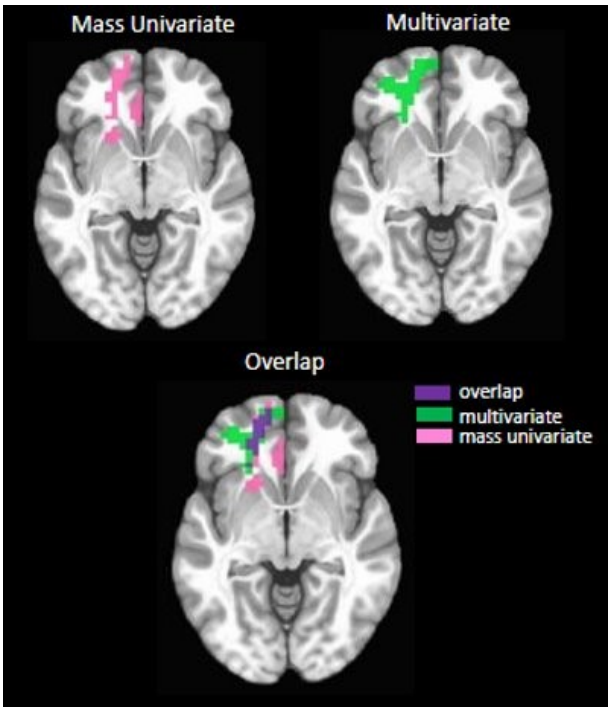


Figure 13. Mass univariate and multivariate predictors for discourse GRAM display spatial overlap.

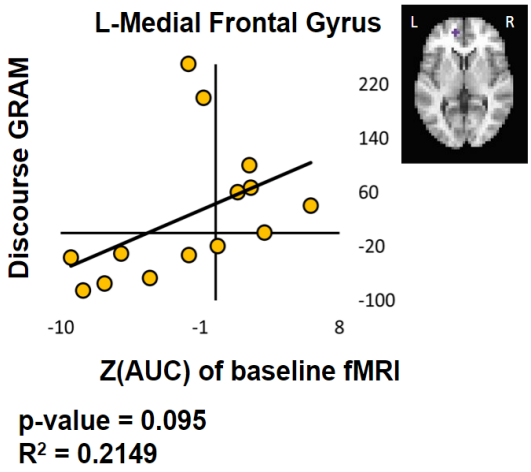


Figure 14. Distinct ROIs predict brain-behavior relationships for discourse GRAM.

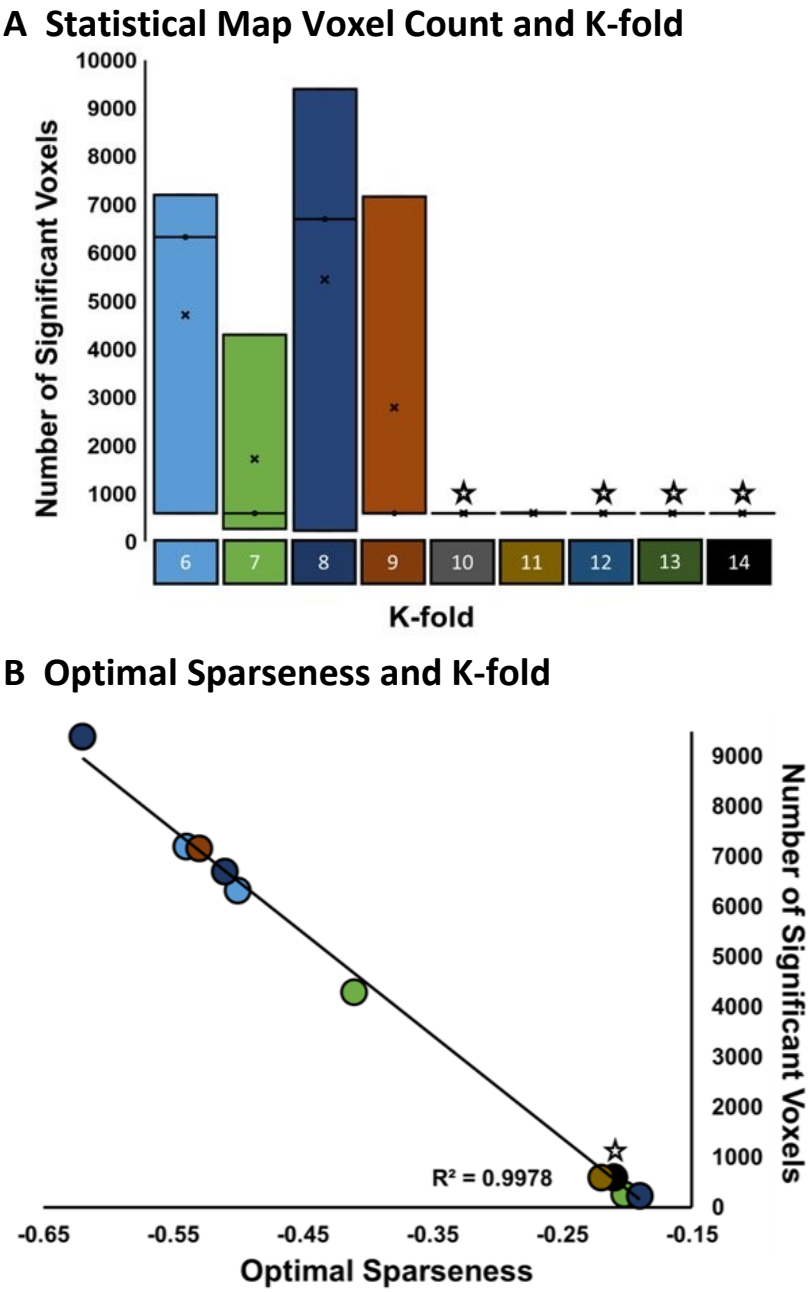
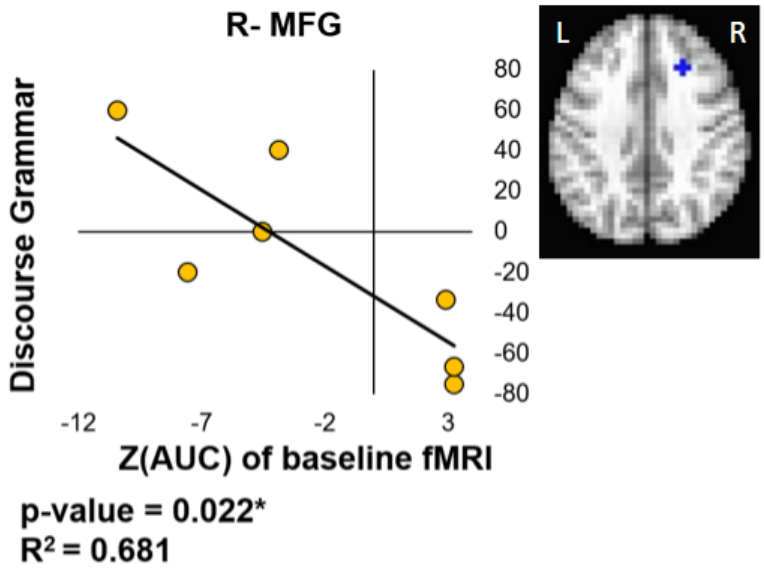


Figure 15. A k-fold of 14, again, yielded robust data within the multivariate analysis for discourse GRAM predictors. (A) The number of significant voxels presented within the statistical image varied across three runs for all k-folds except for k-folds 10, 12, 13, and 14. This lack of variance is highlighted with a star. (B) The optimal sparseness also did not vary for k-folds 10, 12, 13, and 14 (indicated by the star). All dots are color-coordinated according to the k-fold from figure 1A.

Aim 4: To sensitive the multivariate method to predict treatment-specific language changes

Multivariate and mass univariate group-level analysis (CT and INT) was conducted for both behavioral measures in order to assess if these prediction models are sensitive in detecting treatment-induced changes. When analyzing brain-behavior relationships between BOLD activity and CEG, no significant predictive biomarkers were found in both multivariate and mass univariate analyses. However, when analyzing brain-behavior relationships between BOLD activity and discourse GRAM, only multivariate analysis, as opposed to mass univariate analysis, revealed significant predictive biomarkers. Multivariate control-specific predictors included the right middle frontal gyrus (R-MFG), where brain activity and discourse GRAM were negatively associated (Figure 16A). Multivariate intention-specific predictors included the left middle frontal gyrus (L-MFG), where brain activity and discourse GRAM were positively associated (Figure 16B).

A Control Specific Predictor



B Intention Specific Predictor

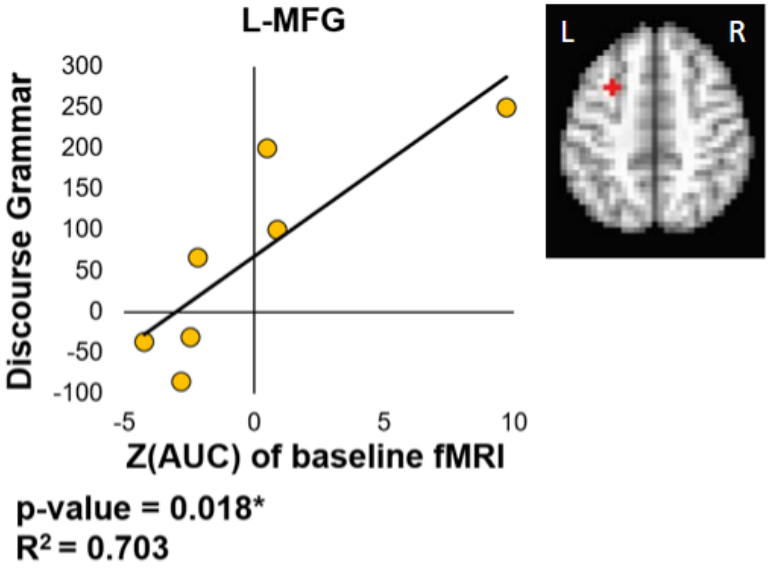


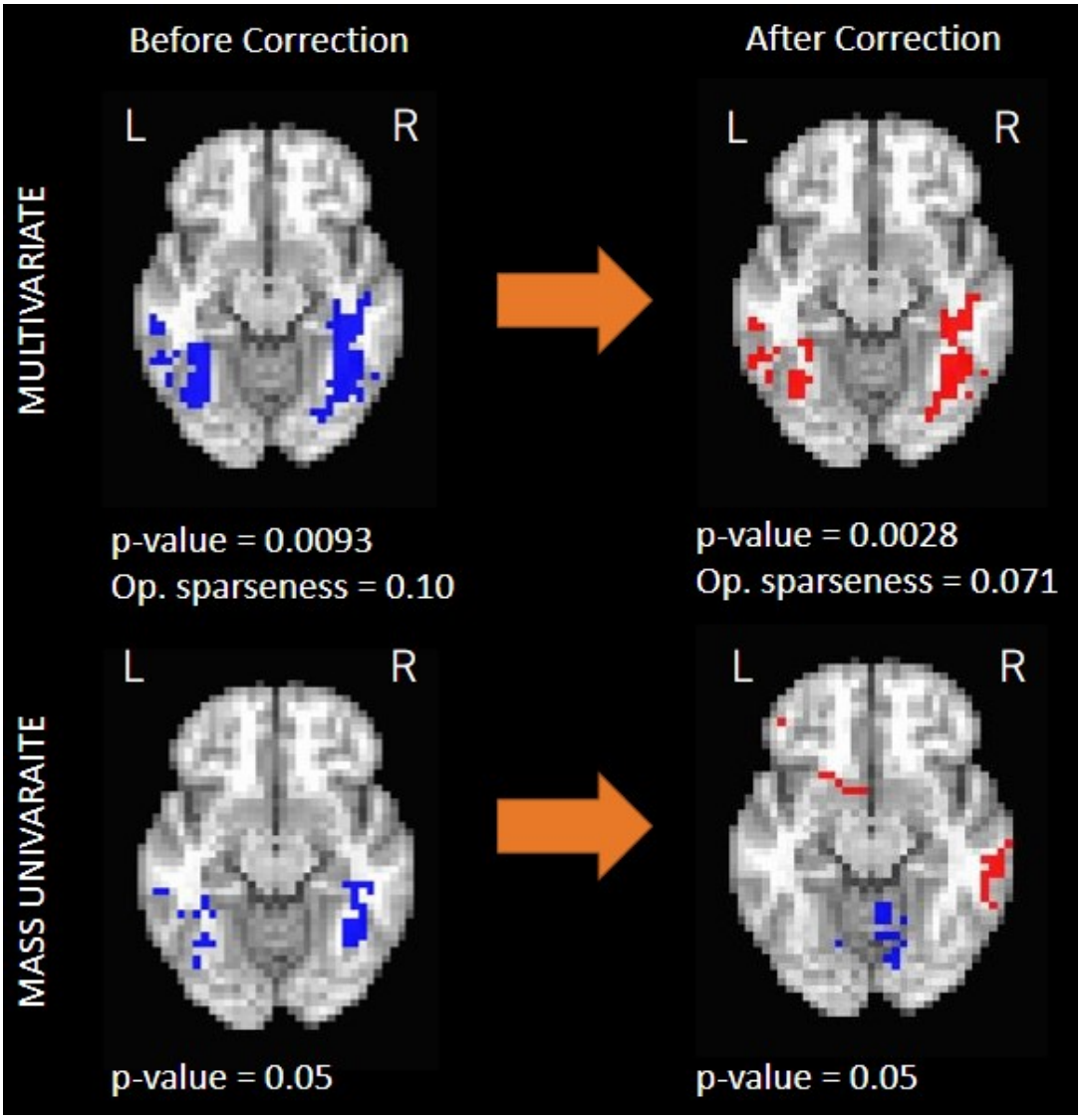
Figure 16. Multivariate analysis revealed predictive biomarkers during group-level analysis. (A) For the control group, activation within the R-MFG was found to be negatively associated with discourse GRAM scores (p-value = 0.022*, R^2 = 0.681). (B) For the intention group, activation within the L-MFG was found to be positively associated with discourse GRAM scores (p-value = 0.018*, R^2 = 0.703).

Correcting for lesion size is feasible for both multivariate and mass univariate analysis

This project has demonstrated the feasibility of correcting for lesion size in both analyses. In terms of the brain-behavior relationships for CEG, multivariate analysis results (p-value = 0.024*, optimal sparseness = 0.51) display spatial overlap with previously non-correct results, however, the sign of the voxel weights are inversed (Figure 17A). The reason behind this inverse has been iterated in the methods. Additionally, the lesion size corrected multivariate results are more conservative compared to the previous analysis. Contrastingly, mass univariate analysis results displayed significant clusters (p-value $\leq$ 0.05) that spatially diverged from the non-corrected results (Figure 17A). In terms of the brain-behavior relationships for discourse GRAM,

both multivariate and mass univariate results display spatial overlap with minor changes in the number of significant voxels (Figure 17B).

A CEG Lesion Size Correction



B Discourse GRAM Lesion Size Correction

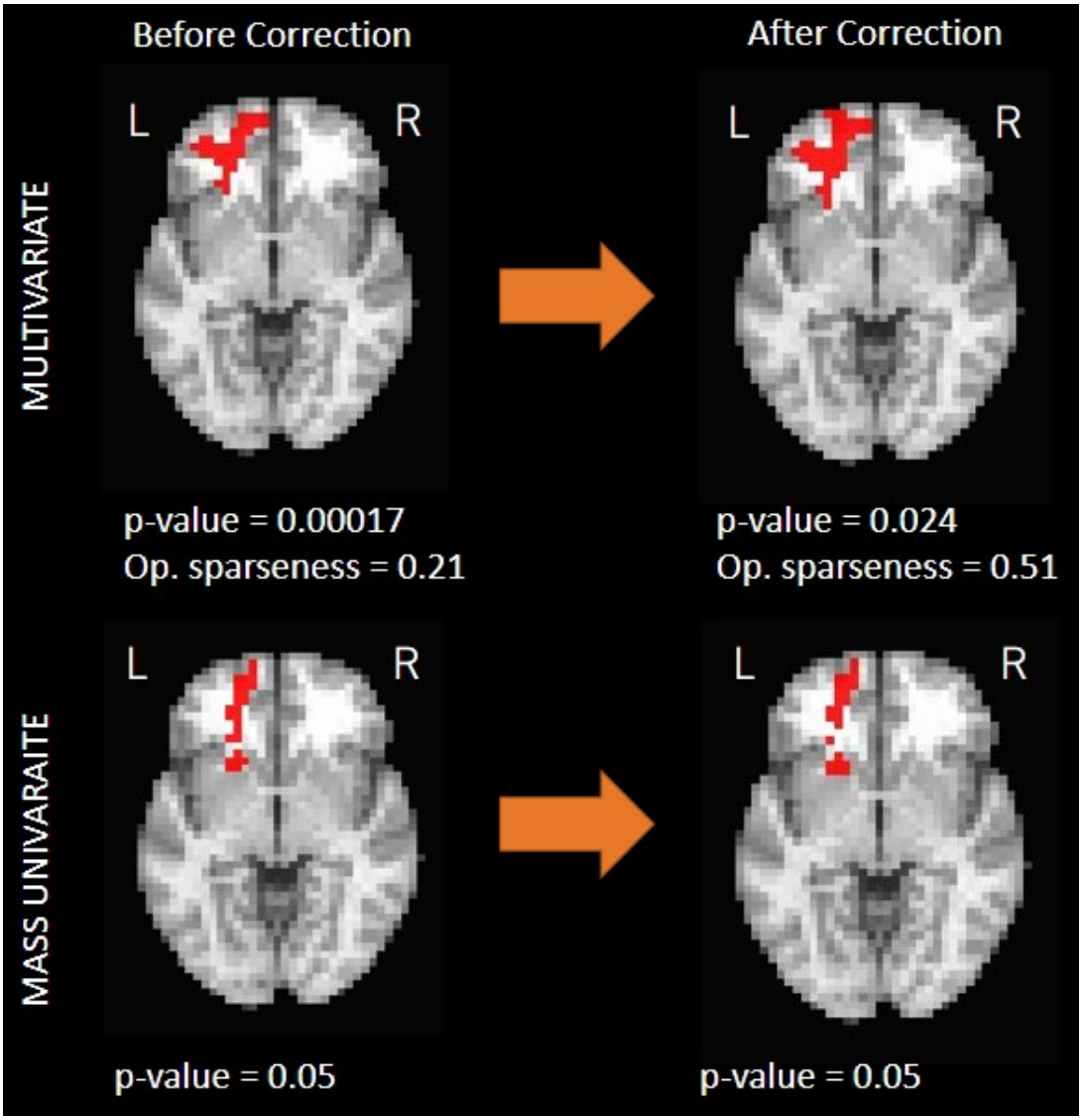


Figure 17. Multivariate and mass univariate approaches can be optimized to account for the effects of lesion size. (A) Multivariate results for CEG predictors display some spatial overlap while mass univariate results for CEG predictors display no spatial overlap. (B) Multivariate and mass univariate results for discourse GRAM predictors display some spatial overlap.

Discussion

This study focused on optimizing a multivariate methodology of identifying significant brain-behavior relationships to predict treatment outcomes in PWA. Conventionally used methods of predicting treatment outcomes present limitations that have been mitigated in this study. Specifically, a multivariate approach has been utilized in order to accurately capture context-driven voxel interactions. Additionally, whole-brain task fMRI data was utilized in order to more holistically capture network-level information that may meaningfully contribute to the brain-behavior relationship. Given, however, that current lesion-symptom mapping approaches have been optimized for lesion-specific analysis, the LESYMAP R package had to initially be modified in order to feasibly analyze whole-brain task fMRI data. K-fold CV modifications revealed that utilizing a k-fold equivalent to the number of samples yielded the most robust data. Additionally, brain mask modifications were made to account for the whole brain while maintaining statistical power. Lastly, the influence of lesion size on the predictive analysis was able to be regressed. Once the multivariate method was fully optimized, the reliability of this technique was assessed by analyzing the level of convergence with other, more conventionally used predictive tools such as mass univariate analysis. Results showed that both methods displayed significant spatial overlap, however, interestingly multivariate analysis uniquely discovered a predictive biomarker within the right extrastriate area. Once reliability was assessed, the expanded utility of the multivariate method was assessed by utilizing another behavioral measure. Notably, previous multivariate optimizations complied in this new analysis, and convergence with mass univariate approaches was also observed. Interestingly, multivariate analysis was able to identify predictive biomarkers of recovery for a more complex function that did not identically overlap with the function involved in the task fMRI. This suggests the

potential utilities of multivariate analysis in assessing treatment influence on not only task-related functions but also more network-level functions. Moreover, assessments regarding the sensitivity of multivariate analysis in discovering treatment-specific (i.e. INT versus CT) biomarkers were conducted. Results showed that, as opposed to the mass-univariate approach, multivariate analysis was able to discover meaningful predictors of treatment outcomes. All in all, this study assesses the efficacy of whole-brain task fMRI multivariate analysis from multiple angles including optimization, reliability, expanded utility, as well as sensitivity in discovering predictive biomarkers. In accordance with the proposed hypothesis, this study has therefore methodically revealed the promising uses of a novel brain-behavior prognostic tool.

Multivariate analysis can predict post-treatment responses to CEG from whole-brain task-fMRI data

When identifying predictive biomarkers for CEG, mass univariate and multivariate results displayed spatial overlap of significant clusters. For example, activation within the right fusiform gyrus was found to predict no response to treatment-induced changes in CEG for both analyses. Despite this overlap, mass univariate clusters contained fewer voxels and overall were more conservative. Multivariate clusters on the other hand contained more voxels and encapsulated unique biomarkers of CEG such as the right extrastriate area despite inputting the same data. This could be explained by the multifaceted nature of SCCAN as voxels are tested simultaneously in group comparisons as opposed to independent voxel-by-voxel analysis. Thus, SCCAN's ability to concurrently compute relationships between voxels allows network-level information to be considered since voxel-to-voxel interactions will be accounted for. Overall, this makes multivariate analysis better equipped to predict treatment-induced changes for complex functions such as language. Considering, however, that mass univariate clusters still

overlapped with multivariate clusters, this method is still viable to find localized biomarkers of brain-behavior relationships. As mass univariate analysis is heavily data-driven, more subjects may be needed in order to increase statistical power and ultimately mitigate the absence of significant voxels.

The right fusiform gyrus predicted a null treatment response to CEG, however, insignificantly. Contrastingly, activation within the right extrastriate area predicted that the participant's did not respond to treatment and their CEG skills did not improve. This region is known to be involved in the visual processing of categories such as faces, bodies, and objects (Pitcher et al. 2009). This categorization of object stimuli is involved within the CEG task and post-treatment behavioral assessment as participants must orally state examples of a given category that are presented orthographically and auditorily. Evidence has shown that the extrastriate area selectively responds to visually presented words (Baker et al., 2007) in a categorical fashion suggesting the prevalent role of this region within the task. Given that activation within this region is negatively correlated to CEG measures, a category-specific visual attention deficit could explain this relationship. During a CEG session, domain-general aspects of language such as visual attention are heavily involved. Oftentimes this attention is selective depending on the stimuli and may require different neural mechanisms. Specifically, it has been revealed that attention may enhance or suppress activity in category-selective areas of the extrastriate depending on the stimulus (Johnson and Johnson 2009). This balanced interplay between activation and suppression within the extrastriate area during the onset of certain categories may be influencing the performance of a CEG task where, at baseline, a net increase in activity within the right extrastriate area could be hindering the ability to effectively alter attention to category-specific orthographic stimuli. Given, however, that the categories within the

Johnson and Johnson study utilized faces and scenes instead of semantic categories, a follow-up study to test this theory may be necessary to verify the presence of extrastriate modulations in response to different semantic category-specific stimuli.

Multivariate analysis can predict network-level language rehabilitation in response to treatment

When investigating the expanded utility of multivariate analysis in predicting treatment effects on another behavioral measure, activation within the left medial frontal gyrus predicted a positive response to treatment in discourse GRAM, however, only marginally. Despite the marginal significance of this result, the feasibility of utilizing whole-brain task-fMRI in predicting indirect outcomes is reflected. Specifically, the ability of whole-brain activity information during a semantic fluency task to predict more complex functions of language such as grammar highlights the potential viability of predicting network-level brain activity and thus performance. Despite the more complex nature of grammar, this language function is not exclusively isolated from the language functions involved within CEG. For example, as the formation of grammatically correct phrases is determined by subject-verb agreement, an understanding of the meaning and thus semantics of various nouns and verbs is needed. Treatment at the single word level (CEG/Picture Naming), therefore, may influence the understanding of words to aid in connecting speech at the sentence level.

Multivariate analysis is sensitive in detecting treatment-induced changes

When investigating the sensitivity of multivariate analysis in identifying treatment-specific biomarkers of rehabilitation, both the L-MFG and R-MFG were found to significantly predict changes in discourse GRAM in response to the INT and CT correction procedure respectively. Interestingly, the mass univariate approach was unable to identify significant brain-

behavior relationships during group-level analysis which highlights the statistical limitations of this study. Given a small sample of seven subjects, the multivariate analysis may therefore be more attuned to discovering predictive biomarkers despite smaller datasets. Similar to the reasoning behind mass univariate's more conservative nature in identifying significant voxels, the data-driven approach to this method may pose a hindrance for group-level analysis.

Multivariate analysis revealed that baseline activity within the L-MFG for INT subjects significantly predicts positive response to treatment in discourse GRAM. This region has been shown to be involved in a variety of language functions for both healthy populations and PWA. For example, evidence has shown that activity within the L-MFG is positively associated with literacy (Koyama et al., 2017). Additionally, an fMRI study in people with a lesioned Broca's area revealed that the L-MFG was heavily involved during a verbal fluency task (Dong et al., 2016). Lastly, despite the intended purpose of INT treatment in relateralizing language functions to the right hemisphere, residual language knowledge within the left hemisphere is necessary for language rehabilitation (Crosson 2009). Considering the consensus within the field, the L-MFG is a prominent role in not only generalized language functions, but also language recovery. Thus, baseline activation within this region may suggest the availability of the neural resources necessary to rehabilitate more generalized language functions such as discourse GRAM.

Contrastingly, multivariate analysis revealed that baseline activity within the R-MFG for CT subjects predicted that the participants did not respond to treatment and their discourse grammar skills did not improve. In PWA, this region has been known to be involved in aiding the relateralization of neural mechanisms involved in language. For example, in a study on PWA, language recovery was present when participants had left-hemisphere lesions, however, this recovery stopped after the onset of right-hemisphere lesions (Barlow 1877). Further evidence

reveals that the involvement of the right hemisphere during language recovery is more likely to occur in structures that are homologous to damaged left hemisphere structures (Crosson 2009). As a bulk of the participants within the CT group had left frontal lesions, right frontal areas would be recruited during rehabilitation. Although the CT group did not undergo the INT treatment that was designed to motivate lateralization of language functions, all participants underwent the same standard therapy involving picture naming and CEG. Thus, language rehabilitation is still involved in the standard therapy and recruitment of right hemisphere regions would be necessary for optimal recovery. The timing in which the right hemisphere is involved during a PWA's treatment timeline, however, has been shown to be imperative. For example, certain studies have shown that the initial reliance on the right hemisphere before treatment in aphasia patients resulted in less language recovery (Abel et al. 2015). Such findings may suggest that pre-treatment reliance of right hemisphere regions may reflect less availability for neuronal plasticity of newly recovered language functions to occur. This could explain why baseline activation within the R-MFG is negatively correlated to the recovery of network-level language functions like grammar. Altogether, the results highlight the diverging predictors of language recovery for INT and CT treatment. Not only in terms of spatial location but also correlation, the predictors for treatment-specific groups were discovered by multivariate analysis to be oppositely related. Specifically, the predictive biomarkers for participants within the CT group were negatively associated with post-treatment gains while the INT predictors were positively associated with the post-treatment gains. This could potentially suggest the impact that INT treatment may have in overcoming the hindrances apparent when the right hemisphere is initially relied on before treatment interventions take place. The emphasis on left-hand gestures and the initiation procedure may therefore aid the manifestation of right hemisphere relateralization by

enhancing the mechanisms involved in neural plasticity.

Limitations and Future Works

Despite the promising utilities of whole-brain task fMRI multivariate analysis, there are some limitations to this study. Perhaps most notably, this study utilized a small number of subjects mainly due to screening procedures as well as the difficulty of recruiting long-term PWA participants. As a result, this study lacks statistical power which may explain why the multivariate analysis was not able to sufficiently obtain certain. For example, this approach was unable to find meaningful INT and CT-specific biomarkers for CEG. Another limitation to this study is that the maintenance of language functions was not assessed. Only behavioral data collected 2-weeks post-treatment was used in the model, therefore, the predictive biomarkers discovered in this study may not be meaningful in terms of long-term recovery. Due to this limitation, a follow-up study utilizing behavioral measures obtained much later (i.e. 3-months post-treatment) may be necessary to solidify the results of this study. Moreover, due to technical problems and time constraints, the lesion size corrected results have not yet been interpreted. Additionally, the task fMRI-based approach to study brain-behavior relationships has primarily involved cortical and subcortical analysis of gray matter. Stroke, however, may affect not only gray matter but also white matter tracts which have been shown to be involved in language by connecting nodes within the language network such as the Broca's area and Wernicke's area (Catani and Mesulam, 2008). Adjacent approaches in understanding brain-behavior relationships involving diffusion tensor imaging (DTI) may therefore be necessary to fully grasp whole-brain analysis. Lastly, this study has focused on utilizing task fMRI data in order to predict treatment responses, however, resting state fMRI (rsfMRI) data may also provide meaningful information needed to create accurate prediction models. Studies have shown that resting network activity

can be used to assess changes in connectivity (Nair et al. 2015) as well as aphasia severity profiles that may give insights into the extent of language recovery (Baliki et al. 2018). rsfMRI may therefore be a beneficial contributor to the rehabilitation prediction model by providing complementary hemodynamic changes within the resting network (Zhao et al. 2018), and future studies utilizing both rsfMRI and task-fMRI data should by conducted. Considering these limitations, future works may involve testing multivariate analysis on datasets involving more subjects, utilizing longitudinal behavioral measures, and using DTI as well as rsfMRI data in parallel with task fMRI data. These future works could potentially reveal meaningful information regarding language treatment-induced neuroplasticity that may open new doors for individualized treatment planning and optimal rehabilitation in clinical settings for PWA.

Conclusion

Overall, this study has established the feasibility of utilizing whole-brain task-fMRI to compute a brain-behavior prediction model using multivariate analysis. By fully optimizing and assessing the reliability, extended utility, as well as sensitivity of multivariate predictors, this study has methodically provided the groundwork for future prediction analyses. The discovery of certain predictive biomarkers such as the right extrastriate area has shed light on how network-level information may present not only domain-specific but also domain-general predictors. Additionally, the discovery of INT and CT-specific biomarkers highlights the attuned sensitivity of multivariate analysis when whole-brain network-level information is utilized. Altogether, these findings reveal the promising potential of utilizing this novel method to ultimately aid patient triaging and the individualization of treatment planning by offering a more precise prediction of treatment-induced language recovery.

References

Abel, S., Weiller, C., Huber, W., Willmes, K., & Specht, K. (2015). Therapy-induced brain reorganization patterns in aphasia. Brain, 138(Pt 4), 1097-1112.

Altamura, C., Reinhard, M., Vry, M. S., Kaller, C. P., Hamzei, F., Vernieri, F., et al. (2009). The longitudinal changes of BOLD response and cerebral hemodynamics from acute to subacute stroke. A fMRI and TCD study. BMC Neurosci, 10, 151.

Altmann, L. J., Hazamy, A. A., Carvajal, P. J., Benjamin, M., Rosenbek, J. C., & Crosson, B. (2014). Delayed Stimulus-Specific Improvements in Discourse Following Anomia Treatment Using an Intentional Gesture. J Speech Lang Hear Res, 57(2), 439-454.

Avants, B. B., Libon, D. J., Rascovsky, K., Boller, A., McMillan, C. T., Massimo, L., et al. (2014). Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. 84, 698-711.

Babbitt, E. M., & Cherney, L. R. (2010). Communication Confidence in Persons with Aphasia. Topics in Stroke Rehabilitation, 17(3), 214-223.

Baker, C., Worrall, L., Rose, M., & Ryan, B. (2020). 'It was really dark': the experiences and preferences of people with aphasia to manage mood changes and depression. Aphasiology, 34(1), 19-46.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc Natl Acad Sci U S A, 104(21), 9087-9092.

Bakheit, A. M., Shaw, S., Carrington, S., & Griffiths, S. (2007). The rate and extent of improvement with therapy from the different types of aphasia in the first year after stroke. Clin Rehabil, 21(10), 941-949.

Baldassano, C., Iordan, M. C., Beck, D. M., & Fei-Fei, L. (2012). Voxel-level functional connectivity using spatial regularization. Neuroimage, 63(3), 1099-1106.

Baliki, M. N., Babbitt, E. M., & Cherney, L. R. (2018). Brain network topology influences response to intensive comprehensive aphasia treatment. NeuroRehabilitation, 43(1), 63-76.

Barlow, T. (1877). On a Case of Double Hemiplegia, with Cerebral Symmetrical Lesions. Br Med J, 2(865), 103-104.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion–symptom mapping. 6(5), 448-450.

Benjamin, M. L., Towler, S., Garcia, A., Park, H., Sudhyadhom, A., Harnish, S., et al. (2014). A Behavioral Manipulation Engages Right Frontal Cortex During Aphasia Therapy. Neurorehabil Neural Repair, 28(6), 545-553.

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. Am J Speech Lang Pathol, 13(3), 236-249.

Breining, B. L., & Sebastian, R. (2020). Neuromodulation in post-stroke aphasia treatment. Curr Phys Med Rehabil Rep, 8(2), 44-56.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. Cortex, 44(8), 953-961.

Cherney, L. R., Merbitz, C. T., & Grip, J. C. (1986). Efficacy of oral reading in aphasia treatment outcome. Rehabilitation Literature, 47(5-6), 112-118.

Code, C., & Herrmann, M. (2003). The relevance of emotional and psychosocial factors in aphasia to rehabilitation. Neuropsychol Rehabil, 13(1-2), 109-132.

Crinion, J., Holland, A. L., Copland, D. A., Thompson, C. K., & Hillis, A. E. (2013). Neuroimaging in aphasia treatment research: quantifying brain lesions after stroke. Neuroimage, 73, 208-214.

Crosson, B. (2008). An intention manipulation to change lateralization of word production in nonfluent aphasia: current status. Semin Speech Lang, 29(3), 188-200; quiz C-184.

Crosson, B., Rodriguez, A. D., Copland, D., Fridriksson, J., Krishnamurthy, L. C., Meinzer, M., et al. (2019). Neuroplasticity and aphasia treatments: new approaches for an old problem. Journal of Neurology, Neurosurgery & Psychiatry, 90(10), 1147-1155.

Cruice, M., Worrall, L., Hickson, L., & Murison, R. (2003). Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being. Aphasiology, 17(4), 333-353.

Damasio, A. R. (1992). Aphasia. N Engl J Med, 326(8), 531-539.

David Pitcher and Lucie Charles and Joseph, T. D. a. V. W. a. B. D. (2009). Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. Current Biology, 19(4), 319-324.

Dong, J. W., Brennan, N. M., Izzo, G., Peck, K. K., & Holodny, A. I. (2016). fMRI activation in the middle frontal gyrus as an indicator of hemispheric dominance for language in brain tumor patients: a comparison with Broca's area. Neuroradiology, 58(5), 513-520.

Ellis, C., Simpson, A. N., Bonilha, H., Mauldin, P. D., & Simpson, K. N. (2012). The one-year attributable cost of poststroke aphasia. Stroke, 43(5), 1429-1431.

Fridriksson, J., Basilakos, A., Hickok, G., Bonilha, L., & Rorden, C. (2015). Speech entrainment compensates for Broca's area damage. Cortex, 69, 68-75.

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. Macrocircuits, 23(2), 250-254.

Glueckauf, R. L., Blonder, L. X., Ecklund-Johnson, E., Maher, L., Crosson, B., & Gonzalez-Rothi, L. (2003). Functional Outcome Questionnaire for Aphasia: overview and preliminary psychometric evaluation. NeuroRehabilitation, 18(4), 281-290.

Goodman, E. B. J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. Language and Cognitive Processes, 12(5-6), 507-584.

Gopinath, K., Crosson, B., McGregor, K., Peck, K., Chang, Y. L., Moore, A., et al. (2009). Selective detrending method for reducing task-correlated motion artifact during speech in event-related FMRI. Hum Brain Mapp, 30(4), 1105-1119.

Harnish, S. M., Neils-Strunjas, J., Lamy, M., & Eliassen, J. (2008). Use of fMRI in the Study of Chronic Aphasia Recovery After Therapy: A Case Study. Topics in Stroke Rehabilitation, 15(5), 468-483.

Helm-Estabrooks, N., & Ramsberger, G. (1986). Treatment of agrammatism in long-term Broca's aphasia. British Journal of Disorders of Communication, 21(1), 39-45.

Henry, M. L., Beeson, P. M., & Rapcsak, S. Z. (2008). Treatment for lexical retrieval in progressive aphasia. Aphasiology, 22(7-8), 826-838.

Horng, S. H., & Sur, M. (2006). Visual activity and cortical rewiring: activity-dependent plasticity of cortical networks. Prog Brain Res, 157, 3-11.

Howard, D., Patterson, K., Franklin, S., Orchard-lisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. Cognitive Neuropsychology, 2(1), 49-80.

Iain DeWitt and Josef, P. R. (2013). Wernicke’s area revisited: Parallel streams and word processing. Brain and Language, 127(2), 181-191.

Ivanova, M. V., Herron, T. J., Dronkers, N. F., & Baldo, J. V. (2021). An empirical comparison of univariate versus multivariate methods for the analysis of brain-behavior mapping. Hum Brain Mapp, 42(4), 1070-1101.

Johnson, M. R., & Johnson, M. K. (2009). Top-down enhancement and suppression of activity in category-selective extrastriate cortex from an act of reflective attention. J Cogn Neurosci, 21(12), 2320-2327.

Kertesz, A. (1993). Clinical forms of aphasia. Acta Neurochir Suppl (Wien), 56, 52-58.

Koyama, M. S., O'Connor, D., Shehzad, Z., & Milham, M. P. (2017). Differential contributions of the middle frontal gyrus functional connectivity to literacy and numeracy. Sci Rep, 7(1), 17548.

Krishnamurthy, V., Krishnamurthy, L. C., Meadows, M. L., Gale, M. K., Ji, B., Gopinath, K., et al. (2021). A method to mitigate spatio-temporally varying task-correlated motion artifacts from overt-speech fMRI paradigms in aphasia. Hum Brain Mapp, 42(4), 1116-1129.

Kuriki, S., Takeuchi, F., & Hirata, Y. (1998). Neural processing of words in the human extrastriate visual cortex. Brain Res Cogn Brain Res, 6(3), 193-203.

Le, D. a. L. K. a. P. C. a. P. E. M. (2016). Automatic Assessment of Speech Intelligibility for Individuals With Aphasia. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2187-2199.

Lutkenhoff, E. S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J. D., Owen, A. M., et al. (2014). Optimized brain extraction for pathological brains (optiBET). PLoS One, 9(12), e115551.

Maher, L. M., & Raymer, A. M. (2004). Management of anomia. Top Stroke Rehabil, 11(1), 10-21.

Marshall, J., Pound, C., White-thomson, M., & Pring, T. (1990). The use of picture/word matching tasks to assist word retrieval in aphasic patients. Aphasiology, 4(2), 167-184.

McClung, J. S., Rothi, L. J., & Nadeau, S. E. (2010). Ambient experience in restitutive treatment of aphasia. Front Hum Neurosci, 4, 183.

Meinzer, M., Beeson, P. M., Cappa, S., Crinion, J., Kiran, S., Saur, D., et al. (2013). Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis. 73, 215-224.

Mirman, D., Landrigan, J. F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. Neuropsychologia, 115, 112-123.

Murray, S. (1972). THE BEHAVIORAL ANALYSIS OF APHASIA**This research was supported by Grant NS 03535 from the National Institute of Neurological Diseases and Stroke, and by the Joseph P. Kennedy, Jr., Laboratories of the Neurology Service, Massachusetts General Hospital. Colleagues in these studies have been J. Leicester, J. P. Mohr, P. B. Rosenberger and L. T. Stoddard. In J. V. B. a. W. J. H. NAUTA (Ed.), Principles, Practices, and Positions in Neuropsychiatric Research (pp. 413-422): Pergamon.

Nair, V. A., Young, B. M., La, C., Reiter, P., Nadkarni, T. N., Song, J., et al. (2015). Functional connectivity changes in the language network during stroke recovery. Ann Clin Transl Neurol, 2(2), 185-195.

Nichol, L., Pitt, R., Wallace, S. J., Rodriguez, A. D., & Hill, A. J. (2022). "There are endless areas that they can use it for": speech-language pathologist perspectives of technology support for aphasia self-management. Disabil Rehabil Assist Technol, 1-16.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy: shared insights on how it is done and why it might help. Ann N Y Acad Sci, 1169, 431-436.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. Curr Biol, 19(4), 319-324.

Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. Journal of Evaluation in Clinical Practice, 18(3), 689-694.

Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. Neuropsychologia, 115, 154-166.

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., & Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. Hum Brain Mapp, 37(4), 1405-1421.

Quian Quiroga, R., & Kreiman, G. (2010). Measuring sparseness in the brain: comment on Bowers (2009). Psychol Rev, 117(1), 291-297.

Raymer, A., & Kohen, F. (2006). Word-retrieval treatment in aphasia: Effects of sentence context. J Rehabil Res Dev, 43(3), 367-378.

Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). J Speech Hear Disord, 45(3), 308-324.

Simon, S. K. a. T. C. a. A. F. a. K. A. a. N. R. (2009). Broca's area: Nomenclature, anatomy, typology and asymmetry. Brain and Language, 109(1), 29-48.

Small, S. L., & Llano, D. A. (2009). Biological approaches to aphasia treatment. Curr Neurol Neurosci Rep, 9(6), 443-450.

Thompson, C., & Shapiro, L. (2005). Treating agrammatic aphasia within a linguistic framework: Treatment of Underlying Forms. Aphasiology, 19(10-11), 1021-1036.

Thompson, C. K., den Ouden, D.-B., Bonakdarpour, B., Garibaldi, K., & Parrish, T. B. (2010). Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. 48(11), 3211-3227.

Thye, M., & Mirman, D. (2018). Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. Neuroimage Clin, 20, 1129-1138.

Wilson, S. M. (2017). Lesion-symptom mapping in the study of spoken language understanding. Lang Cogn Neurosci, 32(7), 891-899.

Yu, Z. Z., Jiang, S. J., Jia, Z. S., Xiao, H. Y., & Zhou, M. Q. (2017). Study on Language Rehabilitation for Aphasia. Chin Med J (Engl), 130(12), 1491-1497.

Zhao, Y., Lambon Ralph, M. A., & Halai, A. D. (2018). Relating resting-state hemodynamic changes to the variable language profiles in post-stroke aphasia. Neuroimage Clin, 20, 611-619.