C-STAR – Center for the Study of Aphasia Recovery Modeling Treated Recovery from Aphasia

NCT03416738
Unique Protocol ID: Pro00053559
U.S. NIH Grant/Contract Award # P50DC014664

Study Protocol (Version 14) and Research Plan
April 20, 2020

Human Subjects Research Proposal
Study Title: Center for the Study of Aphasia Recovery (C-STAR)
Project 001(032)[1]: Modeling Treated Recovery from Aphasia
Project 003(038): Combining Behavior, Neuroimaging and Biomarkers to Predict Language Deficits after Stroke
Principle Investigators: Julius Fridriksson, Ph.D., Christopher Rorden, Ph.D.

Table of Contents:

[1] Note, project times here are consistent with project names in attached grant submission

A. Abstract

Stroke is the leading cause of serious adult disability in the United States. One of the most devastating impairments resulting from stroke is aphasia, a language impairment caused by left hemisphere damage involving cortical language areas. Although it is generally accepted that behavioral aphasia treatment is effective, different patients experience very different degrees of benefit from aphasia treatment, and the relationship between patient factors and treatment response is poorly understood. Moreover, few reliable prognostic indicators have been identified. Therefore, the purpose of the current project is to investigate a variety of factors (e.g., biographical, cognitive/linguistic, and patterns of brain damage) that predict an individual's response to aphasia treatment. To this end, the current submission is one part of a large, multi-center collaboration, with this specific project focused on assessment and treatment of 120 post-stroke individuals in the chronic phase of stroke (>12 months post-onset). In a crossover design, two different treatments will be studied, over the course of a 39-week period (including rest periods). Following the completion of treatments, we will model participants' response to therapy based on biographical, cognitive/linguistic variables, as well as patterns of brain damage. An additional 30 individuals will be recruited for evaluation, but will not complete treatment. In the protocol to follow, we have outlined human subjects information relevant to testing that will take place at the University of South Carolina or the Medical University of South Carolina.

Please note that this constitutes the first phase of the overall project (NIH submitted grant is included in this submission), and that additional projects will stem from the results found here (and in turn, will likely require either amendments to the current IRB or entirely new IRB submissions at later dates).

B. Background and Significance

Background

Speech/language therapy for the management of aphasic impairments is generally shown to be effective. However, the reasons that certain treatments may work for some individuals, and not others, and why some individuals do not respond to treatment is largely unknown. Two recent Cochrane reviews of behavioral aphasia treatment suggest that although aphasia treatment may be effective, no specific treatments have emerged as *standards of care* (1, 2). After decades of research, no general consensus has been reached regarding which aspects of language or communication abilities should be treated, how treatment should proceed, or which factors are important for predicting aphasia treatment success. As a result, individuals with aphasia receive treatment relying on an assortment of approaches, often with very minimal scientific data to support their utility. Additionally, prediction of long-term outcomes typically takes a “wait and see” approach. Accordingly, clinicians who routinely treat patients with aphasia have very limited empirical data to guide their treatment and predict outcome. Whereas biomarker studies are common in medicine, almost no studies have been carried out to understand the relationship between patient factors and aphasia rehabilitation potential in stroke (3-5). This is a major problem, as valuable resources are probably wasted because time and funds available to rehabilitate individual patients are likely limited. Whereas single case or small-group studies are helpful for testing different aspects of aphasia treatment, the time is ripe for larger scale studies that rely on “big data” approaches to answer general questions regarding what should be treated and who is most likely to respond to treatment. Over time, such studies should be

transformative for the field as *standards of care* may be established based on the patient characteristics that are most important for determining the aspects of speech and language that should be targeted in treatment. In addition, large well-controlled studies have the potential to provide credibility for the field and influence public policies regarding appropriations of rehabilitation funds.

Existing evidences suggests that aphasia severity is one of the few factors that reliably predicts performance in speech and language therapy (herein referred to as SALT). It is generally accepted that more severe aphasia is associated with poorer treatment outcomes. However, aphasia severity is a multidimensional construct, and patients with similar overall severity scores might demonstrate very different language impairment profiles.

Over the past decade, our group has sought to identify factors that predict post-stroke impairment and response to treatment. In a recent pilot study, data from our group has shown that different patterns of brain damage may influence treatment outcomes (6). In this study, treatment approaches were theoretically driven by the dual stream model (DS model; (7)). The dual stream model proposes that the brain has two "streams" for processing speech/language - a dorsal stream in charge of production, and a ventral stream in charge of semantic processing. Based on this model, it was predicted that participants with dorsal stream damage would respond to dorsal treatments (i.e., phonological/production based treatments) and those with ventral stream damage would respond better to ventral stream treatments (i.e., semantically based treatments). Participants underwent two types of treatment - one that was based on phonological processes (sounds of words) and one that was based on semantic processes (meanings of words). Results of this work (6) show a wide range in treatment responses across the patients, but generally, our results showed that responses to different treatment approaches may differ based on patterns of brain damage. However, there are a number of other variables that may influence an individual's response to treatment, which were not investigated. For example, several studies suggest that factors such as aphasia severity (8-11), lesion location (9, 12), and time post-stroke (13) are related to aphasia treatment outcome.

Therefore, to better understand how language impairment relates to treatment outcomes, the current project will again consult the DS model. Specifically, we will test whether measures of proportional damage to the cortical areas that comprise the DS model improve prediction of aphasia treatment response, beyond biographical and cognitive/linguistic factors. Accordingly, we will investigate whether the DS improves prediction of treatment response better than a "default" model (i.e., cognitive/linguistic factors alone). To understand whether our predictive model can be generalized across different kinds of treatment foci, each participant will undergo two types of treatment devoted to phonological stimulation, and a separate treatment phase focusing on semantic stimulation.

Relying on a large participant sample, we plan to identify and model the interaction of biographical, cognitive, and linguistic factors that predict aphasia treatment outcome. This model will provide clinicians with crucial information to make more accurate prognoses regarding participants' treatment outcomes. Whereas identifying biomarkers of positive aphasia treatment may be important, we believe that such an endeavor may be improved if it takes models of speech and language processing into account. That is, cortical damage that affects specific speech and language processes may not only have a particularly negative effect on communication, but also may affect aphasia treatment success. Furthermore, prior research in aphasiology has shown that theory-driven treatments are more effective, but unfortunately, few treatments are indeed motivated by well-supported theories. Hence, the current project will test the utility of the DS model, arguably the most influential

contemporary model of speech and language processing to guide the clinical management of aphasia.

In addition to identifying patterns of brain damage that predict treatment outcomes, this study will investigate the role of learning potential on treatment success. Learning potential is a concept that has received relatively little attention in the aphasia rehabilitation literature. Although treatment-assisted improvements in speech and language processing probably reflect learning (or relearning), the relationship between participants' capacity to learn and response to aphasia treatment is poorly understood. The relatively recent emphasis on individual differences rather than group averages has highlighted very clear differences in learning capacity and potential in the normal population (14-17). We will investigate whether participants' ability to learn at baseline predicts long-term success in aphasia treatment. If this study yields positive results, it will further improve clinicians' ability to guide aphasia management.

In sum, there is a great need for prognostic indicators of aphasia treatment response. At the completion of our research, we will understand why some participants respond better to aphasia treatment than others. We have selected treatment approaches that are routinely used in clinical practice, allowing for immediate translation of the findings directly into participant management. The current project will yield a vast dataset that will be made publicly available allowing others to study further aphasia treatment response in relation to cognitive/linguistic and lesion factors (please see *Data Analyses* section for human subjects protection related to this available dataset).

Significance

An important contribution of this project to the literature of aphasiology will be results from a large sample of participants. In the most downloaded article in the history of the Public Library of Science, "Why Most Published Research Findings Are False," Ioannidis (18) noted that when all factors are equal, in fields where smaller studies are the norm, the chances of research findings being false is much higher than in fields where larger studies are typical. The norm in aphasia treatment studies, including our own previous studies, is to rely on single-subject design or very small group studies. It is straightforward to understand why this is the state of affairs. Treatment studies take a relatively long time and a disproportionate level of resources compared to most other kinds of research in aphasia. Also, subject recruitment is often an issue, especially in studies that target participants with specific cognitive/linguistic impairment profiles. Nevertheless, our field needs larger treatment studies to reveal specific trends that can be generalized to the larger population of participants, not just to a narrow number of cases. We believe that our past success, institutional environment, and collaborative spirit puts us in a position to conduct a relatively large aphasia study that has the potential to exert a sustained impact on the field of clinical aphasiology. In the context of previous aphasia treatment studies, project 1 entails a relatively large sample size of treated participants (N=120; rational for sample size is described in the *Human Subjects Protection* section) and extensive cognitive/linguistic testing, including assessment of learning potential and a neuroimaging workup. As we contend that recent meta-analyses studies have demonstrated the value of aphasia treatment, the purpose of this project is not to test the efficacy of treatment, but rather to determine why some participants respond better to treatment than others and whether treatment response relates to which language domains are targeted in treatment. The overall significance of this work points to the fact that the clinicians who provide most of

the aphasia therapy in the United States could far better serve their patients if they were armed with prognostic factors that would indicate whether a given patient was likely to respond to direct SALT**,** and what focus that treatment should assume for maximum recovery.

Finally, the significance of this project lies in the multi-center collaboration as outlined in the P50 submission attached. In addition to testing the hypotheses outlined here, this data will be used in concert with data gathered from Project 2 (PI: Argye Hillis, M.A., M.D., Johns Hopkins University) to fuel Projects 3 (PI: Rorden; USC) and 4 (PI: Hickok; University of California, Irvine).

C. Specific Aims and Hypotheses

To test whether contemporary models of speech processing can be used to predict response to aphasia treatment in chronic stroke patients. The goal is to develop a model that includes biographical and cognitive/linguistic factors to predict who is likely to respond to aphasia treatment and what kinds of stimuli (phonological vs. semantic) should be emphasized in treatment. The specific aims are as follows:

1. To identify biographical and cognitive/linguistic factors that predict aphasia treatment outcome. Construct a predictive model of aphasia treatment success (herein referred to as the "default model") and make that model available on-line to clinicians.

2. To test whether the DS model (as an anatomical model of speech and language processing) improves outcome prediction beyond what can be inferred based on biographical and cognitive/linguistic factors. Essentially, this aim will test whether the DS model provides greater improvement in outcome prediction compared to the default model and classical models based on localizationist approaches to brain-language relationships (i.e., the classical Wernicke-Lichtheim-Geschwind model; WLG(19-22)).

    To test whether measures of learning potential predict overall treatment outcome. Three learning tasks will be administered by the end of treatment phase one – two will be linguistic learning, whereas the third will be a nonlinguistic learning task.

Specific Aims 1 and 2: Models of Treated Aphasia Recovery

Aim 1 will rely on participant and caregiver questionnaires, direct behavioral testing, and MRI scanning. Specific details for each of these components are presented in the *Research Activities, Protection of Human Subjects* section and the Appendix. As we explain in Data Analysis section, this information will be used to determine treatment outcomes based on these different variables. We will model response to treatment based on: 1. *Biographical factors:* Age (23, 24); education (25, 26); gender (27); handedness (28); age at stroke, type of stroke, number of strokes, and time post first stroke that caused the aphasia (29); 2. *Cognitive/linguistic factors:* Clinical assessment of aphasia (10, 11), motor speech production (30), grammatical processing (31), semantic processing of verbs and nouns (32), phonological processing (33), speech discrimination (34), speech repetition (35), executive functioning (36), and verbal short-term memory (37). **Once we have built our final data-driven model(s), it will be possible to examine the specific interactions between these factors that make up each model.**

Hypotheses Aims 1 and 2. Along with several other groups, we have demonstrated that functional brain changes in the residual language areas of the left hemisphere support treated improvement in naming in aphasia (21, 38). Based thereon, we could conclude that strengthening of the residual left-hemisphere language network, the areas that remain intact after stroke, leads to improved language processing in aphasia. However, as far as we can tell, no research has examined whether this kind of network compensation relies on overall network integrity or survival of specific sub-components of the network. We suggest that the DS model offers a principled approach to answer such questions. For example, for cases of damage to the ventral stream, and concomitant semantic impairment, treatment might focus specifically on semantic stimulation. Here, one could assume that preservation of the ventral stream would be related to treatment success, so that patients with greater damage to the ventral stream are less likely to respond compared to others whose ventral stream is relatively more preserved. This view would contend that successful rehabilitation relies on the survival of the network that is being stimulated. However, an opposing view could argue that what really matters is cross-domain compensation where the dorsal stream is forced to become more strongly involved to support improvements in language. Along these lines, we propose that reliance on the DS model may improve our understanding of why some patients show greater improvements than others and to relate the findings in neuroanatomy.

To test this, the data collected from the treatment and assessment of the individuals enrolled in this project will inform further projects (i.e., specifically Dr. Rorden's project, Project 3). This project will rely on the same behavioral data collected here, but it will focus on developing sophisticated neuroimaging analyses to understand the relationship between localized brain damage and aphasic impairment/recovery. The benefit of this setup is that each project may yield complementary results that can better explain differences in treatment response among participants.

Specific Aim 3:

Unlike the measures included in Aims 1 and 2 that mostly represent static factors (e.g., performance on cognitive/linguistic tests, proportional damage to specific brain regions), learning potential represents dynamic change, something that could be assessed at baseline to understand if prolonged aphasia rehabilitation is likely to be effective. Aim 3 therefore investigates the predictive value of learning potential for aphasia treatment outcome. One way to estimate learning is to measure changes in contextual support (e.g. cueing) needed to evoke the correct response (124). In a previous study (N=30; 10, 85) our group investigated the level of cueing needed to elicit correct naming response to stimulus presentations across multiple treatment sessions. Results revealed that changes in cueing strength during the initial treatment sessions (for the phonological and semantic treatments) strongly predict overall treatment outcome. These results suggest that participants whose level of cueing strength needed to elicit correct naming decreases during the initial treatment session are far more likely to respond to the overall treatment, regardless of initial aphasia severity. We recognize that decreased reliance on cues during naming may not be a classical learning task and that what we deem "learning" in this context could also be called "stimulability" or priming. Regardless, performance on the cueing hierarchy task by our 30 participants was a highly robust predictor of outcome and no other factor – behavioral or neuroimaging – came even close in terms of predictive power. From a theoretical perspective, it is a caveat that we do not know whether the changes in cueing

level represent a general learning mechanism or something that is language specific. Therefore, in addition to measuring potential changes in cueing strength during a baseline session that uses cueing hierarchies to elicit picture naming, all participants will also complete a separate learning task that does not explicitly rely on language processing. As a result, it will be possible to determine if initial success in treatment is related to language independent learning or is language specific.

Hypothesis Aim 3. Learning is driven by neuroplasticity – the ability of the brain to continuously adapt its structure and function based on internal and external environmental changes. Neuroplasticity occurs due to changes in neuronal morphology, glia, and vascular and metabolic processes and can be stimulated by a variety of sensory-motor experiences (5, 39, 40). Learning potential is a common concept in the education literature (41, 42). It assumes that not all individuals have the same ability to acquire new information and probably reflects flexibility of neural systems to adapt to internal changes or external stimulation. Although most clinicians probably assume that some aphasic participants have greater rehabilitation potential than others, it is not clear whether participants' learning potential is related to aphasia treatment outcome. Nevertheless, we predict that learning potential will be related to treatment outcomes.

D. Data Collection

Comparing Biomarkers as a Function of Treatment Foci

One of the challenges to interpreting the value of aphasia treatment biomarkers is the generalizability across different kinds of treatment types and foci. For that reason, this project will rely on a crossover design where each participant completes one treatment phase that focuses on semantic stimulation and another treatment phase where the focus is on phonological stimulation. Accordingly, we will be able to compare biomarkers of treatment outcome and determine if the same factors predict success, regardless of treatment focus, or whether biomarkers are dependent on treatment focus.

Research Design and Methods and Data Analysis

Our plan is to treat 120 participants with chronic aphasia caused by damage to the left hemisphere. Utilizing a crossover design, each participant will receive 45 minutes of aphasia treatment per weekday (5x/week) focusing on treatment tasks that emphasize semantic or phonological stimulation (Table 1). Half of the participants will receive treatment in the following order: a. 15 treatment sessions utilizing phonological stimulation tasks; b. 4-week inter-treatment interval; c. 15 treatment sessions with semantic stimulation oriented tasks. The remaining half of participants will receive treatment administered in the opposite sequence. Participants will be randomized to receive treatment in either order.

During the first baseline visit (referred to as W0 here, but W1 in the consent form), participants' biographical and testing information is entered into online modules in WebDCUTM (see P50 *Administrative Core*). All subsequent participant data, including testing and treatment scores, MRI related information, and treatment fidelity information, will also be entered and stored using WebDCUTM. To establish treatment baselines and for the purpose of in-depth characterization of the participant sample, each participant will undergo extensive testing during the week prior to treatment.

To assess treatment related changes in speech and language processing, testing will occur during the weeks immediately before and after each treatment phase. Follow-up testing will occur

4-weeks and 6 months (at or around week 39) after completing the second treatment phase. To assess damage to the cortical areas included in the WLG and DS models, participants will undergo MRI (see P50 *Neuroimaging Core*) at baseline during the week before treatment is initiated. Additionally, participants will undergo MRI testing following each treatment phase and at 6-months post treatment completion to assess functional brain changes. Approximately two-thirds of the participants will be enrolled at the University of South Carolina. The remaining one- third will be enrolled at the Medical University of South Carolina under the supervision of Dr. Bonilha, a stroke neurologist with ample experience managing aphasia, and a long-time collaborator of the PI. Individuals will be asked to complete an anonymous participant satisfaction survey at the week 15 visit to provide feedback on their clinical research experience.

Timeline for testing and treatment. G1 notes a group of participants who first undergo aphasia treatment utilizing tasks that target phonological stimulation. Then, during the second treatment phase, these participants will cross over to receive treatment focused on semantic stimulation. G2 (group 2) will undergo treatment using the opposite order of treatment phases. W=Weeks; Eval=Evaluation of performance (testing); Phon. Tx.=Phonological processing focused treatment; Sem. Tx.=Semantic processing focused treatment.

Group	W1	W2-4	W5	W6-7	W8	W9-11	W12	W15*	W 39#
G1	MRI & Eval 2-3 days	15 Phon. Tx	MRI & Eval 1-2 days	Rest	Eval 1-2 days	15 Sem. Tx	MRI & Eval 1-2 days	Eval 1-2 days	MRI & Eval 1-2 days
G2	MRI & Eval 2-3 days	15 Sem. Tx	MRI & Eval 1-2 days	Rest	Eval 1-2 days	15 Phon. Tx	MRI & Eval 1-2 days	Eval 1-2 days	MRI & Eval 1-2 days
*W15 denotes 4-week follow-up testing #W39 denotes 6-month follow-up testing (this testing should take place at or around week 39)									

Additional Follow-Up

A subset of participants will be called back for a long-term follow-up, at least six months after the 6 month follow-up (i.e., one-year post-treatment). The purpose is to extend the follow-up window for the POLAR study, to one-year post aphasia treatment. There is evidence to suggest that declining performance in stroke-induced aphasia is attributed to pathologic changes in the brain's white matter, referred to as leukoaraiosis (Basilakos et al., in press). Leukoaraiosis is common in stroke survivors, and is often attributed to age, as well as cardiovascular health and other health concerns. In the general population, it has been associated with cognitive decline and loss of functional ability. By extending follow-up interval for the POLAR study, this study will address a new aim of evaluating the progression of leukoaraiosis and how it affects overall cognitive-linguistic ability in stroke survivors. To this end, participants will undergo a subset of the same speech/language and cognitive testing administered at study baseline, as well as testing with the NIH Toolbox. Participants will also undergo a magnetic resonance imaging (MRI) protocol (about one hour in duration). The extent of behavioral change will be evaluated and related to neuroanatomical changes. The role of cognitive reserve and other personal factors in mediating the effects of leukoaraiosis will also be explored. The Appendix outlines additional measures that will be obtained at this testing time point.

Recruitment

Identification of eligible participants. For **Project 1**, the Clinical Core team will focus recruitment efforts on referrals from rehabilitation professionals as outlined in our

Patient/Participant Referral Program (PRP). This program provides a monetary incentive of $100 to any rehabilitation professional who refers a patient who comes in for an evaluation and is deemed eligible for this study. We will continue to use online and social media outlets as well as community activities to reach out potential participants.

Aphasia Treatment

Since our primary goal is to assess predictors of treatment that can be applicable to existing practices, we selected treatments that are typically employed. Neither semantic nor phonological processing exists independently of each other in oral language use. Accordingly, the treatment approaches to be used here do not focus exclusively on either semantic or phonological processing; they each involve both, but place different emphasis on semantics versus phonological processing. In fact, even when semantic processing is the treatment emphasis, these treatments also affect phonological processing and vice versa. We stress that the main purpose here is not to assess whether one treatment focus is more potent than the other for improving language processing. Rather, we only include the two treatment foci to understand if factors that predict outcome with one treatment approach also predict success using the other treatment approach. In addition, we include two semantic and phonological treatment foci to explore whether damage to the dorsal and ventral streams differentially predicts response to phonological and semantic treatment.

Semantically focused treatment tasks. Three types of tasks will be featured. Please see the Appendix for scripts and examples for each treatment task. Tasks are as follows: 1) Semantic feature analysis (SFA) has a relatively long history as a treatment approach for aphasia (43-45). For each pictured stimulus the participant is prompted to name the picture. Then, he or she is encouraged to produce semantically related words that represent features similar to the target word (e.g. superordinate category, use, action, physical properties, location, and association). For example, to elicit a location feature, a clinician might say "where do you typically find this object?" If the participant is not able to name the target item once each word feature has been produced, the clinician will say target word. Regardless of naming accuracy on the last item, treatment continues on to the next stimulus item. Because both nouns and verbs (46, 47) have been used for SFA focused activities, stimuli for SFA tasks here will utilize both. 2) The second semantic treatment approach is the semantic barrier task. This approach includes features of the Promoting Aphasics' Communication Effectiveness (PACE) approach (48-51) and has also been included as part of constraint-induced language therapy (52). It relies on a stack of picturable stimuli, which are split between the participant and clinician and placed face up on a table. A visual barrier is placed between the clinician and the participant so they cannot see each other's pictures. The goal of the task is for one participant (e.g., participant) to describe each card so that the other participant (e.g., clinician) can guess the picture on the card. Participants are only allowed to describe the semantic features of the target and the clinician models the kinds of cues that are allowed. The clinician and participant take turns describing pictures. 3) The third approach, Verb network strengthening treatment (VNeST), is a semantic treatment approach that targets lexical retrieval of verbs and their thematic nouns (53). The objective of VNeST is for the participant to generate verb-noun associates with the purpose of strengthening the connections between the verb and its thematic roles. VNeST can be modified to fit participants with very limited speech output (e.g., using sentence completion).

Phonologically focused treatment tasks. Three approaches, also with sound research pedigrees, will be used. 1) The first is the phonological components analysis task (54), which was modeled after semantic feature analysis. It requires the participant to first name a given picture and then to identify the phonological features of the target words (e.g., first sound, syllables, last

sound, association, and rhyme). Once the features have been identified, participants are required to attempt to name the picture again. Then, the treatment moves on to the next item in a stack of targeted imageable nouns and verbs. 2) Phonological production task is a phonologically based therapy that focuses on the identification of phonological features using a stack of targeted imageable nouns and verbs. It requires the participant to first sort the stack of picture stimuli based on the number of syllables by tapping out each syllable. Once the participant has sorted the targeted words into two stacks, the treatment moves on to identifying the following hierarchy of phonological features using a pair of the targeted imageable nouns/verbs: a. first syllable-first syllable; b. first syllable-last syllable; c. last syllable-last syllable; d. last syllable-first syllable; e. first syllable-first sound; f. last syllable-last sound; g. first syllable-last sound; h. last syllable-first sound. Once each targeted feature is identified for the pair of words, the participant is required to blend the syllables/sounds together. 3) The phonological judgment task relies on computerized presentation of verbs and nouns where participants are required to judge whether pairs of words include similar phonological features. The task is comprised of five conditions that entail determining if a set of words: a. Includes the same number of syllables; b. Includes the same initial syllable; c. Includes the same final syllable; d. Which word has more syllables; and e. rhyme. Participants respond to each condition by pressing one of two response buttons depending on the task requirements and instructions. **Additional information regarding all treatment tasks can be found in Appendix A.**

Biographical Factors and Cognitive/linguistic Assessments

Biographical factors. For the purpose of recording data on biographical factors, each participant (and/or caregiver) will be required to fill out a case history form and a questionnaire on previous medical events, years of education, etc.

Cognitive/linguistic testing. The following tests and tasks will be utilized to assess cognitive/linguistic status at baseline, before any treatment starts: 1. The revised version of the WAB will be used to assess overall aphasia severity and type (55). Although the WAB does not provide in-depth analysis of speech and language impairment, it is a test that is commonly used in clinical practice as well as a reasonable measure of overall aphasia severity; 2. The Apraxia of Speech Rating Scale (ASRS; (56)) will be utilized to rate the presence and severity of apraxia of speech AOS (note: the ASRS constitutes a major component of project 3); 3. To assess grammatical processing (agrammatism), we will rely on the Northwestern Assessment of Verbs and Sentences (57). The NAVS was designed for participants with aphasia and allows for detailed examination of verb processing (e.g. verb naming) as well as production and comprehension of canonical and non-canonical sentences; 4. The Pyramids and Palm Trees Test (PPTT; (58)) and the Kissing and Dancing Test (KDT; (59)) will be used to assess amodal semantic processing of nouns and verbs, respectively; 5. To assess phonological processing, several sub-tests from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; (60)) will be utilized (a. Auditory discrimination of nonword minimal pairs; b. Auditory discrimination of word minimal pairs; c. Rhyme Judgments for pictures and words; d. Phonological Segmentation of initial and final sounds); 6. To assess speech repetition ability we will rely on the Philadelphia Repetition Test (PRT), a low-imageability word repetition test (61), and the non-word speech repetition sub-test from the PALPA (65); 7. For analysis of cognitive status, the Matrix test on the Wechsler Adult Intelligence Scale (WAIS; (62)) will be administered; 8. To assess verbal short-term memory, we will use tasks described by Martin et al. (63). Their tasks allow for detailed evaluation of lexical-semantic performance in relation of verbal short-term memory load. All participants will be screened for visual field cut and visual neglect during the neurological examination as a part of the NIH Stroke Scale. To minimize the influence of test fatigue, the assessment battery will be spread

over three sessions administered on three separate days during the week before the first treatment phase starts. The motivation for the extensive behavioral test battery included here is to provide a comprehensive set of factors used to predict treatment outcome. In addition, the behavioral data collected here will also feed into projects 3 and 4 (e.g. non-word repetition and PRT). **For details on test reliability, validity, and clinician training, please see the Clinical Core section in the overall P50 Grant Application (relevant pages beginning on p. 264). Additionally, scripts that will be used for each treatment approach are presented in Appendix B.**

Learning tasks for specific Aim 3

All participants will complete three learning tasks; two linguistic based learning tasks (phonological and semantic) and one nonlinguistic learning task (spatial). These will be administered in random order for each participant and will be completed by the end of treatment phase one. The tasks are set-up similar to the classical computer game "PONG."

Linguistic learning is completed by a game-like interface, where four words fall at randomly generated speeds from the top of a computer screen. Participants are asked to mouse over the words that fit the category of the game, such as 'animal' or 'has the "s" sound', and to avoid words that are not targets but may be semantically or phonologically related or unrelated.

Nonlinguistic learning will be tested by the same game-like interface. The task involves shape-patterns where participants are required to identify identical shapes to the target shape and to avoid distractor shapes. In this way, we circumvent internal language representations of task stimuli. Before beginning the task, the participant is shown a target shape and told that they will be presented with a computer screen where a variety of different shapes will fall from the top of their screen. The participant's objective is to mouse over the target shape and to avoid the other shapes (distractors).

All learning tasks will be conducted 5 times consecutively for 2 minutes each time. The accuracy score will be recorded for each session. The participant's learning ability is summarized as the increase in accuracy between baseline and final session. The study will compare linguistically based task vs nonlinguistic learning tasks (each to be completed by the end of treatment phase 1) to determine whether learning potential relates directly to prognosis.

Treatment Outcome Measures

Consistent with previous work in our lab and Project 2, change in correct naming on the Philadelphia Naming Test (PNT; (64)) will be assessed as an outcome measure. Confrontation naming is included on most comprehensive aphasia tests and is commonly assessed in aphasia treatment research, making it possible to compare the findings in the current project to other studies. The PNT will be administered twice at baseline and then once during the week before and the week after each 3-week treatment phases. For follow-up testing, the PNT will be administered again at four weeks and six months after completion of the second treatment phase. The primary outcome measure, however, is specifically defined as change in correct naming immediately before and after each treatment phase. The PNT is a computer-based assessment of naming for persons with aphasia and includes 175 pictures representing mid- and high-frequency nouns from a word frequency list compiled by Francis and Kucera (65). Participants are videotaped during assessment and responses are scored offline. In addition to assessing changes in correct naming, potential treatment-related changes in naming errors will also be scrutinized. The Treated Naming 40 is a computer-based assessment created for this study to assess changes in correct naming of

targeted nouns and verbs from treatment. There are a total of 40 pictures representing targeted nouns and verbs from treatment (20 from phonological and 20 from semantic). Of those 20 from each deck they were equally matched by nouns and verbs. The nouns included are matched by equal number of high frequency and low frequency words as well as by syllable length. The verbs are matched only according to syllable length. Additionally, any words that appear on the PNT were excluded. Unlike the PNT, the Treated Naming 40 is administered only once during each assessment time point: baseline, before and after each treatment phase, and at 4 weeks and 6 months post completion of treatment. For a more detailed description of PNT and Treated Naming 40 scoring and training of clinicians, see Clinical Core.

To evaluate the effects of treatment on discourse abilities, participants will complete a discourse protocol included in AphasiaBank, an archival database of discourse samples funded by the NIDCD since 2007. Three discourse tasks will be administered before and after each treatment phase: 1. Broken Window picture sequence description (66); 2. Cinderella story telling (67); and 3. A procedural description of how to make a peanut butter and jelly sandwich (first introduced by Davis, 1901 (68), as referenced in Lau, 2013 (69). Transcription will be accomplished through the Clinical Core using transcription (CHAT) and the automated coding analysis (CLAN) systems available through AphasiaBank. We have already used this setup to assess outcome in aphasia treatment (30). As in the CATES trial, the analyses used for discourse samples will focus on lexical and syntactic aspects of spoken discourse: 1. Content words/min (excludes repetitions and nonwords); 2. Propositional density (70, 71); 3. Verbs/utterance. Each of these three measures is already included in AphasiaBank and it is a particularly nice feature that our data can be compared against the 440 aphasic participants included in AphasiaBank. This feature is especially important for determining variance across participants. In addition, we will assess potential changes in participants' ability to generate main concepts (72). Finally, to assess the effects of treatment on functional communication ability, all participants will be administered the Adaptive Aphasia Communication Outcome Measure (CAT-ACOM; (73)). The CAT-ACOM test adaptively administers 12 items from the 59 item ACOM item bank. A content-balancing strategy is used to select items from each domain (talking, writing, number use, comprehension, and naming) to insure that the content balance of each CAT-ACOM administration is reflective of the content balance of the full 59 item bank. The test items administered probe different aspects of real life communication situations and can be administered with a short administration time. The ACOM will only be administered at baseline and at six months post treatment completion.

Neuroimaging

MRI data will be collected at USC and MUSC. Both sites are equipped with a Siemens 3T Trio and an established quality assurance protocol is already in place to ensure that the same MRI sequences yield comparable data on both scanners. The MRI scanner at USC is housed at the McCausland Center for Brain Imaging (MCBI), a facility that is co-directed by Fridriksson and Rorden. At MUSC, MRI scanning will take place at the Center for Biomedical Imaging, a facility Bonilha routinely uses for MRI data collection. Once MRI scanning is complete, all data are transferred to and stored on a secure server housed at University Technology Services (UTS), USC. Extensive data backup and data mirroring is included as part of UTS services. All preprocessing of MRI data will occur at USC under the direction of Dr. Rorden as part of the Neuroimaging Core. We have already constructed a state-of-the-art data pipeline for preprocessing various MRI data modalities as well as a new software package to relate various types of MRI data to aphasia impairment and recovery.

DNA Analyses

All participants will be asked to provide a saliva sample for the purpose of genotyping for BDNF (brain derived neurotrophic factor) and the *FOXP2* gene (forkhead box p2).

BDNF is a growth factor that has been shown to promote plasticity. Some individuals are carriers of the VAL66MET polymorphism, which leads to reduced BDNF secretion, and potentially decreased outcomes in aphasia therapy (74).

The FOXP2 gene is the first gene that has been implicated in speech-language impairment. Specifically, the FOXP2 polymorphism rs17137124. This gene has been implicated in neurogenesis and synaptogenesis, but little is known about is role in post-stroke recovery in humans. However, in a mouse model, FOXP2 expression has been related to recovery of vocalization (75).

We will investigate the extent that the presence of these polymorphisms affects therapy outcomes. To do so, a DNA Genotek Oragene-DNA-OG500 kit will be used. Each kit contains a small tube that collects saliva. Participants will be asked to spit into the tube until 2 milliliters of saliva has been acquired. Participants will be provided with written/picture instructions (see attached Oragene DNA User Instructions). The process of delivering a saliva sample takes approximately 2-5 minutes. Sample tubes will be labeled with participant number and date of acquisition. For participants who have already completed the study, we will request that they either: i) come to the Aphasia Lab to supply a sample, ii) allow us to meet them at their home to obtain sample, or iii) for participants not living in the Columbia area, we will ask them to complete the sample by mail.

If participants are to complete the sample by mail, they will be asked to complete and return the provided consent form with their sample. We will ask that they call one of the treating clinicians if they need help providing the sample, and a SLP will guide them through the instructions. We will provide all postage for participants to mail their sample to the lab. Also, we will use the Aphasia Lab's address as the return address to ensure confidentiality.

All samples will be stored in DNA Genotek DNA storage boxes and locked in a storage cabinet in the lab. Samples are stable at room temperature for up to 5 years. All samples will be labeled with participant numbers.

E. Data Analyses

Specific Aim 1

To construct the default model of recovery, both biographical and cognitive/linguistic factors are entered into a step-wise regression analysis using a leave-one-out method. This approach narrows down the number of factors that provide independent prediction of treatment outcome. The leave-one-out method controls for selection bias where the same subjects are not used to select the factors and test the model. The factors that survive in the final step-wise regression model are then entered into the support vector regression (SVR) analysis to determine how well the default model predicts actual outcome in participants. SVR is a more robust alternative to multiple linear regression. Both methods use a linear model to characterize the relationship between the predictors and the target variable (e.g, change in correct naming); that is, the target variable is modeled as a weighted sum of predictors, plus the offset term. The difference between the two methods is in the estimation of the model parameters (the weights and the offset). In multiple linear regression, it is done by computing the Moore-Penrose pseudo inverse of the predictor matrix; this minimizes the prediction errors in the least-squares sense, but leads to highly unstable estimation if there is correlation between the predictors (76). Also, the pseudo inverse cannot be computed if the number of predictors is greater than the number of observations. In support vector regression, we do not attempt to minimize the prediction errors; rather, we allow for a certain amount of errors, and set an upper bound on permissible prediction errors (77). In this formulation, the stability of estimation is not affected by correlations between predictors, and by the number of predictors relative to number of observations. The superiority of support vector regression over linear regression models has been demonstrated by Drucker et al. (78). We expect that the final default model will include a combination of several factors that together contribute to outcome prediction. As not every factor will contribute equally to the

proportional reduction in error, beta weights (converted to Z-scores) will be calculated so that clinicians can better understand the relationship between predictors.

Specific Aim 2

To compare the default model to previous models of speech/language processing (the WLG model and the DS model) we will also rely on SVR. Here, all factors that comprise each model are entered into an SVR analysis and the R2 for the correlation between the actual scores (e.g., change in correct naming or propositional density) and the predicted scores is calculated. Then, a statistical comparison is made between the R2 yielded by the different prediction models. The model with statistically significantly highest R2 compared to other models will be determined as the best prediction model. Separate analyses will be conducted for each treatment focus (semantic vs. phonological) and overall treatment outcome.

Specific Aim 3

Separate SVR analyses will be conducted to test whether change in cueing in a single naming session or performance on the hit-and-avoid task predicts aphasia treatment outcome. The learning curve (slope of the regression line) will be calculated for each participant. The same SVR analyses planned for Aim 1 will be run again but now with the learning factors also included to assess whether the default model can be improved by including assessments of learning potential. Potential concerns: Without collecting further data, it is impossible to estimate the necessary length of the learning paradigms. This is a potential concern as administering an hour-long learning task to participants in clinical care is probably not feasible. Thus, we will calculate learning curves for progressively longer increments (e.g. for the first 10, 20, 30, or 40 minutes of the task) to appreciate the necessary length of each learning task. For example, it could be the case that improvement in performance over the first 20 minutes of the language-learning task is sufficient to predict treatment outcome.

Making the Default Model Available On-line

Once the default model (i.e., response to treatment based on biographical and cognitive/linguistic factors) has been constructed, we will make the algorithm available to clinicians who can then assess their own participants and plug in the results to estimate treatment response relying on either semantically or phonologically focused treatment. For this purpose, we will design a website that includes detailed instructions regarding what factors need to be assessed and the strengths and limitations of this approach. Similarly, we will provide guidance regarding assessment fidelity, threats to assessment reliability and validity, as well as disclaimers about potential abuse of the model (e.g., denying participants treatment based on model prediction of poor outcome). Importantly, when datasets from this project are made available, no individually identifying information will be shared.

F. Protection of Human Subjects

Human Subjects Involvement and Characteristics

Approximately 150 participants with left hemisphere ischemic or hemorrhagic stroke will be included in this study (120 will receive aphasia treatment). For the purposes of future projects (outlined in the P50 plan, Projects 3 and 4), we will also test (but not treat) additional 30 participants with chronic left stroke who do not present with aphasia. We will not exclude participants with recurrent stroke unless one or more strokes included the right hemisphere. In general, aphasia recovery associated with spontaneous healing (i.e., recovery not attributed to specific external intervention) is thought to mostly subside between 6 and 12 months post-stroke. Therefore, to establish a stable behavioral baseline, only participants who are at least 12 months

post-stroke will be enrolled in this research. The age range will be set at 21-80 years. The lower limit is based on evidence suggesting that brain-language development is not complete until in the late teens (79). The upper limit is set in an attempt to exclude participants with dementia. To increase the chance that all participants can perform the selected treatment tasks, potential participants will be excluded based on aphasia severity if their spontaneous speech score on the WAB is lower than 2 or if their auditory comprehension sub-score on the WAB is lower than 2. Participants who have contraindications for MRI will be excluded, as we will not be able to determine the extent of damage to the cortical areas that comprise the DS model or the WLG model. Participants will not be excluded based on AOS or dysarthria severity unless it is clinically determined by the PIs that speech intelligibility is too poor for accurate assessment of naming performance on the primary outcome measure (PNT). Each participant will undergo in-depth assessment of speech and language abilities allowing for detailed description of the participant sample

Rationale for Sample Size

For Aim 1, the primary goal is to construct a model that predicts aphasia treatment success. We have used results from prior work (80) to conduct a power analysis for the current project. We used the factors from this data to compute exhaustive leave-one-out regression values, resulting in the more conservative R2=.088, R2=.27 and R2=.36 for phonological, semantic, and overall treatment effects (rather than .22, .44 and .46). This suggests that a sample size of 120 individuals provides 0.95 power at $p < 0.05$, and 0.85 power at $p < 0.01$. Deriving power estimates for Aim 2 is more speculative: rather than comparing against a static null hypothesis (e.g., treatment has no influence) we are comparing against different dynamic models that will each benefit from a larger sample size. Conventional analyses suggest that a sample size of 120 individuals will provide 0.92 power at $p < 0.05$ for supporting the weakest effect, shown in Figure 3 (default model's ability to predict phonological treatment benefits) and 0.72 power at $p < 0.05$ for distinguishing the difference between the default model and the Default+DS model. We note that a null result for the latter in the presence of a robust result in the former could still be clinically meaningful, potentially suggesting that initial behavioral tests are sufficient for treatment planning.

Research activities

Participant involvement will take approximately eight weeks (excluding time between treatment phases) with an inter-treatment interval of four weeks and follow-up testing scheduled at four weeks and six months after treatment completion. Each participant will undergo neuropsychological testing, including baseline testing of naming ability, and complete three learning tasks, as well as an MRI session during the first week before treatment starts. The next three weeks will be devoted to behavioral aphasia treatment (with five treatment sessions administered on Monday-Friday each week) followed by a 4-week rest period, and, finally, another 3- week treatment period. Before and after each of the two treatment phases, participants will be tested with the primary and secondary outcome measures. Participants will also undergo a follow-up MRI session after each of the two treatment phases and at 6-months post treatment completion.

Over the past 15 years, we have established a large pool of participants with chronic aphasia (N=145) from which we will recruit for the current project. The Midlands area of South Carolina, where the University of South Carolina is located, has one of the highest stroke rates in the United States. As a result, the number of stroke participants with chronic aphasia that live in Columbia and the surrounding areas is relatively high. In an ongoing unrelated study, we are testing participants with left hemisphere stroke in acute care, providing us with a pool of participants with and without aphasia. Many of these participants end up coming to our weekly

aphasia support groups (current enrollment=37) and are eager to participate in research, especially treatment studies. Based on our previous success in recruiting participants for our treatment studies as well as a major recruitment effort that will reach South Carolina and neighboring states, we see no reason why we would not be able to enroll and treat 120 participants with chronic left hemisphere stroke.

Eligibility

Participants with chronic aphasia (>12 months from stroke) as a result of left hemisphere stroke (ischemic or hemorrhagic) will be included. Persons who incur more than one stroke will be excluded only if the subsequent stroke(s) includes the right hemisphere. **Inclusion/exclusion criteria are summarized as follows.**

Inclusion Criteria

1. Participants must have incurred a left hemisphere ischemic or hemorrhagic stroke.
2. Participants must be greater than 12-months post-stroke onset.
3. Participants must between 21 and 80 years of age.
4. Participants must have spoken English as their primary language for the last 20 years.
5. Participants must be willing and able to provide informed written or verbal consent.
6. Participants must be MRI-compatible (e.g. no metal implants, not claustrophobic, etc.)

Exclusion Criteria

1. Severely limited verbal output (WAB Spontaneous Speech rating scale score of 0-1)
2. Severely impaired auditory comprehension (WAB Comprehension score of 0-1).
3. Bilateral stroke

Source Materials

Medical information will be obtained from study subjects. Results of assessments obtained during each subject contact will be recorded in the research record. In the event that the participant's prior medical record is required to document medical history, an authorization for records release will be obtained from the subject. Each subject that signs the consent form and meets the study eligibility criteria will have a unique study identifier assigned to them by the WebDCUTM system upon enrollment into the study. The DCU computers and servers will not store any personal health identifiers (i.e., name, medical record number); rather, the subject will be tracked during the study period through the assigned unique identifier. Only the appropriate local site study personnel will have access to a participant's personal identifying information. All paperwork that includes participant identifiers (e.g., medical records) will be kept in a locked cabinet in the PI's office. Once a participant's baseline information has been reported in WebDCUTM, all subsequent testing and treatment data are entered online using customized case report forms (CRF) that are specifically designed with the specific needs of project 1 in mind. In addition to the use of passwords and other security measures, all documents containing identifying information on individuals or physicians are considered confidential materials and are safeguarded to the greatest possible extent. No information that identifies a specific person, hospital, or physician is released to or discussed with anyone other than study staff members. All employees are informed of this policy before being hired and required to sign a pledge to maintain the confidentiality of all information. In addition, a set of guidelines are developed for the staff dealing with the specific office and data processing procedures to assure the confidentiality and security of all data.

Consent

Informed consent will be obtained from participants during their initial assessment

session. The speech-language pathologist in charge of testing will obtain the informed consent. For participants whose auditory comprehension is severely impaired, every effort will be made to convey what study enrollment entails and caregivers will be included in the process. An additional consent form will be completed to allow participants and caregivers (if applicable) to choose whether or not they give consent for media release of video files or pictures for use of educational purposes and/or shared on social media.

Potential Risks and Discomforts

For the current research, three areas are of paramount interest with regard to the protection of human subjects: 1) Potential adverse effects associated with MRI scanning; 2) Potential adverse effects associated with behavioral testing and treatment; and 3) Maintenance of participant confidentiality. We will address each of these below on point-by-point basis.

1. MRI Scanning. For the purpose of determining the extent of cortical damage, each participant will undergo one pre-treatment MRI session as well as three additional MRI sessions. Additional MRI scans will be completed following each treatment phase and at 6-months post treatment completion. This will allow us to assess potential treatment related changes in functional brain activation, functional brain connectivity, and structural brain connectivity. We have conducted extensive research using MRI in stroke. Overall, the number of stroke participants who have received MRI in our studies exceeds 270 (includes both acute and chronic stroke). Among participants with chronic stroke who have participated in our research, we have occasionally had to exclude participants because of factors contraindicative to MRI. We find that most participants who have reported negative reaction to past MRI experiences do not have problems with our MRI setup. The MRI scanners used for this research at USC and MUSC are short-bore magnets (Siemens 3T Trio) with a wider bore than most older generation scanners. This allows the participant rather easily to see through either end of the scanner bore without shifting head position.

All participants will undergo thorough screening to check for factors counter-indicative for MRI scanning (e.g., metal implants, pregnancy [women who are of child bearing age will be required to take a pregnancy test], pacemaker, severe claustrophobia, etc.). We will follow the same protocol that we have used with past participants. First, each person fills out a questionnaire in collaboration with a clinician and, when needed, a caregiver. Then, we give each participant ample time to get familiar with the scanner (we show them the stimulus presentation hardware, the computer console, the headphone/microphone used to record speech) before the actual scanning session starts. We find that taking this extra time to familiarize participants and, when appropriate, family members, with the MRI setup, participants usually feel comfortable with the MRI session.

2. Participant Testing/Treatment. In our experience, occasional participants will have a negative reaction to neuropsychological testing. For example, this can be caused when participants realize that their performance on a given test is far worse than they would have expected. We always make sure that ample time is allotted for neuropsychological testing and that participants are allowed breaks as needed. So far, we have never had a participant discontinue study participation because of adverse reaction to study testing. Almost without exception, participants who participate in our treatment studies wish to continue study participation even when the treatment period has expired as it allows them access to free aphasia treatment. With regard to the behavioral treatment, the only adverse reaction that we occasionally see is fatigue. This is something that we warn participants about ahead of time.

3. Participant Confidentiality. All study investigators participating in the current project must ensure that the confidentiality of personal identity and all personal medical information of study participants will be maintained at all times. Additionally, the clinical sites are to follow privacy obligations to study participants under the Health Insurance Portability and Accountability Act (HIPAA). Because the DCU uses a web-based system, source documents and CRFs will remain at the site. The study database and any study documents submitted to the DCU will only identify study subjects by unique study identification codes. All data will be stored in a manner that is HIPAA compliant, without the ability to track the information back to a specific subject except through a password protected system (all collected information about a subject will be stored by a unique identification code). All DCU personnel are certified by the NIH Office of Human Subjects Research in the Protection of Human Research Subjects course.

Potential Benefits

The participants enrolled in this research will receive behavioral aphasia treatment, gratis, something that few participants with chronic aphasia receive through third-payer insurance.

Importance of Knowledge to be Gained

Aphasia is common in stroke, the leading cause of adult disability in the United States. Although gradual, long-lasting aphasia recovery is probably common, most participants with chronic aphasia remain aphasic for the rest of their lives. Aphasia therapy can improve aphasia recovery; however, very little is known about what kind of participants benefit from different kinds of treatment. The current research will inform the relationship between language impairment profiles and responses to treatment.

Data Sharing Across Research Projects, Service Cores, and Ongoing Study Trials:

Because the current submission is part of a larger multi-site collaboration, all data collection for the entirety of the collaboration will occur as part of the treatment studies planned in projects 1 (PI: Fridrikson) and 2 (PI: Hillis). The same data will also feed into projects 3 (PI: Rorden) and 4 (PI: Hickok) to predict treatment outcome. Project cores will be managed to support the overall project (Clinical Core, Neuroimaging Core, and the Administrative Core). The two service cores – the Clinical Core (PI: Hillis) and Neuroimaging Core (PI: Rorden) – will support data collection and analyses. In addition, we will incorporate data from two ongoing NIDCD funded trials: 1. tDCS to Treat Aphasia: A phase II trial (CATES trial; DC011739; PI-Fridriksson); and 2. Neural Bases of Language and Cognitive Deficits in Acute Stroke and Recovery (Early Aphasia Recovery; DC005375; PI–Hillis). The CATES trial will provide outcome data for a behavioral treatment that combines semantic and phonological stimulation, which will be compared to outcomes from treatments that assume only a semantic or phonological treatment focus (Project 1). Data from Dr. Hillis' Early Aphasia Recovery project will be utilized in project 3 to predict treatment outcome using neuroimaging factors. The Early Aphasia Recovery project uses DWI and PWI data to predict recovery from aphasia within the first year of stroke onset. This project is in its 15th year of funding. We propose that the analyses planned in project 3 can further bolster the impact of the Early Aphasia Recovery project by applying the state-of-the-art neuroimaging analyses designed as part of analysis programs created by Dr. Rorden (i.e., NiiStat software: https://www.nitrc.org/projects/niistat/). In addition, project 3 will use baseline data collected for projects 1 and 2 to predict treatment outcome. Project 4 will primarily rely on treatment data collected in projects 1 and 2 to refine and integrate the DS model and other contemporary models to better understand the underlying mechanisms that give rise to

speech errors in aphasia as well as predict recovery. The vast data collection planned here also provides an opportunity for others affiliated with the project (Bonilha, Den Ouden, and Desai) to further study language impairment and aphasia recovery. It should be stressed that any data shared will be de-identified; all participants will be referred to by an assigned study number and all documents shared will use this number to identify participants.

In addition to the above collaborations, Dr. Guorong Wu at the University of Chapel Hill will assist with processing of the MR images. Dr. Wu is an expert in image segmentation, registration, and normalization, and will assist with the processing of the longitudinal data that are acquired. He will be provided with access to MRI scans and de-identified behavioral data. This data will be shared with him via the HIPPA compliant Dropbox.

Dr. Brielle Stark at Indiana University will also have access to data for the purpose of analysis and interpretation. There is an MTA agreement with IU and USC to allow for data sharing (dated 12/18/18) of "Audiovisual data of neurophysiological testing, including tasks evaluating language." Data will be shared via the HIPPA compliant Dropbox. She will have access to MRI data and behavioral testing.

G. **References:**

1. Brady MC, Kelly H, Godwin J, Enderby P. Speech and language therapy for aphasia following stroke. Cochrane Database Syst Rev. 2014;5(CD000425). doi: 10.1002/14651858.CD000425.pub2.
2. Kelly H, Brady MC, Enderby P. Speech and language therapy for aphasia following stroke. 2010. 2010;5(CD000425). doi: 10.1002/14651858.CD000425.pub2.
3. Burke E, Cramer SC. Biomarkers and predictors of restorative therapy effects after stroke. Curr Neurol Neurosci Rep. 2013;13(2):329. doi: 10.1007/s11910-012-0329-9. PubMed PMID: 23299824; PMCID: 3580200.
4. Floel A, Cohen LG. Recovery of function in humans: cortical stimulation and pharmacological treatments after stroke. Neurobiol Dis. 2010;37(2):243-51. doi: 10.1016/j.nbd.2009.05.027. PubMed PMID: 19520165.
5. Milot MH, Cramer SC. Biomarkers of recovery after stroke. Current opinion in neurology. 2008;21(6):654-9. doi: 10.1097/WCO.0b013e3283186f96. PubMed PMID: 18989108; PMCID: 2882885.
6. Thors H, Richardson JD, Fridriksson J, editors. Dual stream model guided treatment of aphasia. Society for the Neurobiology of Language; 2015; Chicago, IL.
7. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007;8(5):393-402. doi: 10.1038/nrn2113. PubMed PMID: 17431404.
8. Baird AE, Dambrosia J, Janket S, Eichbaum Q, Chaves C, Silver B, Barber PA, Parsons M, Darby D, Davis S, Caplan LR, Edelman RE, Warach S. A three-item scale for the early prediction of stroke recovery. Lancet. 2001;357(9274):2095-9. PubMed PMID: 11445104.
9. Kertesz A. Recovery from aphasia. Adv Neurol. 1984;42:23-39. PubMed PMID: 6209949.
10. Pedersen PM, Jorgensen HS, Kammersgaard LP, Nakayama H, Raaschou HO, Olsen TS. Manual and oral apraxia in acute stroke, frequency and influence on functional outcome: The Copenhagen Stroke Study. Am J Phys Med Rehabil. 2001;80(9):685-92. PubMed PMID: 11523971.
11. Pedersen PM, Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants, and recovery. Ann Neurol. 1995;38(4):659-66. doi: 10.1002/ana.410380416. PubMed PMID: 7574464.

12. Fridriksson J, Bonilha L, Baker JM, Moser D, Rorden C. Activity in preserved left hemisphere regions predicts anomia severity in aphasia. Cereb Cortex. 2010;20(5):1013-9. doi: 10.1093/cercor/bhp160. PubMed PMID: 19687294; PMCID: 2852500.
13. Kertesz A. Western aphasia battery test manual: Psychological Corp; 1982.
14. Kanai R, Rees G. The structural basis of inter-individual differences in human behaviour and cognition. Nature Reviews Neuroscience. 2011;12(4):231-42.
15. Lövdén M, Wenger E, Mårtensson J, Lindenberger U, Bäckman L. Structural brain plasticity in adult learning and development. Neuroscience & Biobehavioral Reviews. 2013;37(9):2296-310.
16. May A. Experience-dependent structural plasticity in the adult human brain. Trends in cognitive sciences. 2011;15(10):475-82.

17. Wig GS, Grafton ST, Demos KE, Wolford GL, Petersen SE, Kelley WM. Medial temporal lobe BOLD activity at rest predicts individual differences in memory ability in healthy young adults. Proceedings of the National Academy of Sciences. 2008;105(47):18555-60.
18. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
19. Wernicke C. Der aphasische symptomenkomplex. Berlin Heidelberg: Springer; 1974.
20. Lichteim L. On aphasia. Brain. 1885;7(4):433-84.
21. Geschwind N. Disconnexion syndromes in animals and man. II. Brain. 1965;88(3):585-644. PubMed PMID: 5318824.
22. Geschwind N. Language and the brain. Sci Am. 1972;226(4):76-83. PubMed PMID: 5014017.
23. Holland AL, Wertz RT. Measuring aphasia treatment effects: large-group, small-group, and single-subject studies. Res Publ Assoc Res Nerv Ment Dis. 1988;66:267-73. PubMed PMID: 2451851.
24. Wertz RT, Dronkers N. Effects of age on aphasia. In: Cherow E, editor. Proceedings of the research symposium on communication sciences and disorders of aging. Rockville, MD: ASHA; 1990. p. 88-98.
25. Sarno M, Levita E. Natural course of recovery in severe aphasia. Archives of physical medicine and rehabilitation. 1971;52(4):175.
26. Smith A. Objective indices of severity of chronic aphasia in stroke patients. Journal of Speech and Hearing Disorders. 1971;36(2):167-207.
27. Lomas J, Kertesz A. Patterns of spontaneous recovery in aphasic groups: a study of adult stroke patients. Brain Lang. 1978;5(3):388-401. PubMed PMID: 656906.
28. Subirana A. The prognosis in aphasia in relation to cerebral dominance and handedness. Brain. 1958;81(3):415-25.
29. Price CJ, Seghier ML, Leff AP. Predicting language outcome and recovery after stroke: the PLORAS system. Nature Reviews Neurology. 2010;6(4):202-10.
30. Fridriksson J, Hubbard HI, Hudspeth SG, Holland AL, Bonilha L, Fromm D, Rorden C. Speech entrainment enables patients with Broca's aphasia to produce fluent speech. Brain. 2012;135(Pt 12):3815-29. doi: 10.1093/brain/aws301. PubMed PMID: 23250889; PMCID: 3525061.
31. Shapiro K, Shelton J, Caramazza A. Grammatical class in lexical production and morhpological processing: Evidence from a case of fluent aphasia. Cognitive Neuropsychology. 2000;17(8):665-82.
32. Berndt RS, Mitchum CC, Haendiges AN, Sandson J. Verb retrieval in aphasia. 1. Characterizing single word impairments. Brain and Language. 1997;56(1):68-106.
33. Milberg W, Blumstein S, Dworetzky B. Phonological processing and lexical access in aphasia. Brain and Language. 1988;34(2):279-93.
34. Caramazza A, Papagno C, Ruml W. The selective impairment of phonological processing in speech production. Brain and language. 2000;75(3):428-50.

35. Schlaug G, Marchina S, Norton A. From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. Music perception. 2008;25(4):315.
36. Lambon Ralph MA, Snell C, Fillingham JK, Conroy P, Sage K. Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. Neuropsychol Rehabil. 2010;20(2):289-305. doi: 10.1080/09602010903237875. PubMed PMID: 20077315.
37. Caspari I, Parkinson SR, LaPointe LL, Katz RC. Working memory and aphasia. Brain and cognition. 1998;37(2):205-23.
38. Geschwind N. The organization of language and the brain. Science. 1970;170(3961):940-4. PubMed PMID: 5475022.

39. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res. 2008;51(1):S225-39. doi: 10.1044/1092-4388(2008/018). PubMed PMID: 18230848.
40. Cramer SC. Repairing the human brain after stroke. II. Restorative therapies. Ann Neurol. 2008;63(5):549-60. doi: 10.1002/ana.21412. PubMed PMID: 18481291.
41. Grigorenko EL. Dynamic assessment and response to intervention: two sides of one coin. J Learn Disabil. 2009;42(2):111-32. doi: 10.1177/0022219408326207. PubMed PMID: 19073895; PMCID: 3575109.
42. Sternberg RJ, Grigorenko EL. Dynamic testing: The nature and measurement of learning potential: Cambridge university press; 2002.
43. Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. Am J Speech Lang Pathol. 2004;13(3):236-49. doi: 10.1044/1058-0360(2004/025). PubMed PMID: 15339233.
44. Boyle M. Test-retest stability of word retrieval in aphasic discourse. J Speech Lang Hear Res. 2014;57(3):966-78. doi: 10.1044/2014_JSLHR-L-13-0171. PubMed PMID: 24686776.
45. Boyle M, Coelho CA. Application of semantic feature analysis as a treatment for aphasic dysnomia. American Journal of Speech-Language Pathology. 1995;4(4):94-8.
46. Boo M, Rose ML. The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. Aphasiology. 2011;25(2):154-75.
47. Wambaugh JL, Wright S, Nessler C, Mauszycki SC. Combined Aphasia and Apraxia of Speech Treatment (CAAST): Effects of a Novel Therapy. Journal of Speech, Language, and Hearing Research. 2014;57(6):2191-207.
48. Davis GA. PACE revisited. Aphasiology. 2005;19(1):21-38.
49. Davis GA. Aphasiology: Disorders and Clinical Practice. Boston, MA: Pearson/Allyn and Bacon; 2007.
50. Davis GA, Wilcox MJ. Adult aphasia rehabilitation: Applied pragmatics. San Diego, CA: Singular; 1985.
51. Li EC. Treatment for naming impairment. In: Wallace GJ, editor. Adult aphasia rehabilitation. Boston, MA: Butterworth-Heinemann; 1996. p. 229-42.
52. Pulvermuller F, Neininger B, Elbert T, Mohr B, Rockstroh B, Koebbel P, Taub E. Constraint-induced therapy of chronic aphasia after stroke. Stroke. 2001;32(7):1621-6. PubMed PMID: 11441210.
53. Edmonds LA, Nadeau SE, Kiran S. Effect of Verb Network Strengthening Treatment (VNeST) on Lexical Retrieval of Content Words in Sentences in Persons with Aphasia. Aphasiology. 2009;23(3):402-24. doi: 10.1080/02687030802291339. PubMed PMID: 19763227; PMCID: 2744980.
54. Leonard C, Rochon E, Laird L. Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. Aphasiology. 2008;22(9):923-47.

55. Kertesz A. Western Aphasia Battery-Revised. San Antionio, TX: Pearson; 2007.
56. Strand EA, Duffy JR, Clark HM, Josephs K. The apraxia of speech rating scale: a tool for diagnosis and description of apraxia of speech. Journal of communication disorders. 2014;51:43-50. doi: 10.1016/j.jcomdis.2014.06.008. PubMed PMID: 25092638; PMCID: PMC4254321.
57. Cho-Reyes S, Thompson CK. Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). Aphasiology. 2012;26(10):1250-77.
58. Howard D, Patterson K. The Pyramids and Palm Trees Test: A test of semantic access from words and pictures. Cambridge: Thames Valley Test Company; 1992.
59. Bak TH, Hodge JR. Kissing and dancing-a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. Journal of Neurolinguistics. 2003;16(2):169-81.

60. Kay J, Lesser R, Coltheart M. PALPA: Psycholinguistic assessments of language processing in aphasia. New York, NY: Psychology Press; 2009.
61. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychol Rev. 1997;104(4):801-38. PubMed PMID: 9337631.
62. Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio, TX: NCS Pearson. 2008.
63. Martin N, Kohen F, Kalinyak-Fliszar M, Soveri A, Laine M. Effects of working memory load on processing of sounds and meanings of words in aphasia. Aphasiology. 2012;26(3-4):462-93. doi: 10.1080/02687038.2011.619516. PubMed PMID: 22544993; PMCID: 3335394.
64. Roach A, Schwartz MF, Martin N, Grewal RS, Brecher A. The Philadelphia naming test: Scoring and rationale. Clinical Aphasiology. 1996;24:121-34.
65. Francis WN, Kucera H. Frequency analysis of English usage: Lexicon and grammar. Boston, MA: Houghton Mifflin; 1982.
66. Menn L, Ramsberger G, Estabrooks NH. A linguistic communication measure for aphasic narratives. Aphasiology. 1994;8(4):343-59.
67. Grimes N. Walt Disney's Cindarella. New York, NT: Random House; 2005.
68. Davis JD. The Boston Cooking School Magazine of Culinary Science and Domestic Economics. Boston, MA: Boston Cooking-School Magazine; 1901.
69. Lau M. Who Made That? New York Times Magazine. 2013 June 7, 2013.
70. Brown C, Snodgrass T, Kemper SJ, Herman R, Covington MA. Automatic measurement of propositional idea density from part-of-speech tagging. Behavior research methods. 2008;40(2):540-5.
71. Kemper S, Rash S, Kynette D, Norman S. Telling stories: The structure of adults' narratives. European journal of cognitive psychology. 1990;2(3):205-28.
72. Dalton SG, Richardson JD. Core-Lexicon and Main-Concept Production During Picture-Sequence Description in Adults Without Brain Damage and Adults With Aphasia. American Journal of Speech-Language Pathology. 2015;24(4):S923-S38.
73. Doyle PJ, Hula WD, Hula SNA, Stone CA, Wambaugh JL, Ross KB, Schumacher JG. Self-and surrogate-reported communication functioning in aphasia. Quality of Life Research. 2013;22(5):957-67.
74. de Boer RG, Spielmann K, Heijenbrok-Kal MH, van der Vliet R, Ribbers GM, van de Sandt-Koenderman WME. The role of the BDNF Val66Met polymorphism in recovery of aphasia after stroke. Neurorehabilitation and neural repair. 2017;31(9):851-7.
75. Doran SJ, Trammel C, Benashaski SE, Venna VR, McCullough LD. Ultrasonic vocalization changes and FOXP2 expression after experimental stroke. Behavioural brain research. 2015;283:154-61.
76. Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences.: Routledge; 2003.

77. Smola AJ, Scholkopf B. Bayesian Kernel Method. Advanced Lectures on Machine Learning. 2003;2600:65-117. doi: 10.1007/3-540-36434-X_3.
78. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Advances in neural information processing systems. 1997;9:155-61.
79. Chugani HT. A critical period of brain development: studies of cerebral glucose utilization with PET. Prev Med. 1998;27(2):184-8. doi: 10.1006/pmed.1998.0274. PubMed PMID: 9578992.
80. Fridriksson J. Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. J Neurosci. 2010;30(35):11558-64. doi: 10.1523/JNEUROSCI.2227-10.2010. PubMed PMID: 20810877; PMCID: 2938788.

Appendix A
Study Protocol

1. SELECTION OF PARTICIPANTS

All participants in this study will have chronic left hemisphere stroke. Diagnostic evaluations will be conducted during the participants' initial visit to confirm aphasia diagnosis for treatment inclusion. Participants with aphasia diagnosis that are currently receiving speech and language treatment (apart from aphasia support groups) will be required to temporarily discontinue involvement until study completion. A total of 120 subjects with aphasia will be enrolled. A total of 30 subjects without aphasia will be enrolled.

1.1 Participant Inclusion Criteria

1. Participants must have incurred a left hemisphere ischemic or hemorrhagic stroke.
2. Participants must be greater than 12-months post-stroke onset.
3. Participants must between 21 and 80 years of age.
4. Participants must have spoken English as their primary language for the last 20 years.
5. Participants must be willing and able to provide informed written or verbal consent.
6. Participants must be MRI-compatible (e.g. no metal implants, not claustrophobic, etc.)

1.2 Participant Exclusion Criteria

1. Severely limited verbal output. (WAB Spontaneous Speech rating scale score of 0-1)
2. Severely impaired auditory comprehension (WAB Comprehension score of 0-1).
3. Bilateral stroke

2. STUDY PROCEDURES

2.1 Study Procedures Overview

After informed consent is received, a neurological examination utilizing the NIH Stroke Scale will be performed and multiple screening assessments will be conducted including a MRI safety screening. If the participants pass the initial screening portion, multiple speech and language diagnostic tests will be conducted. Participants will also undergo two baseline assessments of naming- and discourse ability (primary and secondary outcome measures) and structural/functional MRI examination. During the next three weeks the participants with aphasia will receive 45 minutes of treatment every day, Monday through Friday (weekend treatment/assessment may take place with schedule conflicts). After the first treatment phase concludes, participants will be tested, on one-two consecutive days, with the primary and secondary outcome measures and undergo a MRI examination. After that there will be a 4-week break between treatment phases. In the week before starting the next treatment phase there will be one-two baseline testing sessions to test the primary and secondary outcome measures. During the next three weeks the participants will receive 45 minutes of treatment every day, Monday through Friday (weekend treatment/assessment may take place with schedule conflicts). After the second treatment phase is completed the participants will be tested again, on one-two consecutive days, with the primary and secondary outcome measures and undergo a MRI examination. Follow-up testing will occur at 4-weeks and 6-months (at or around week 39) after completion of the second treatment phase. Participants will complete an anonymous satisfaction survey at the 4-week follow-up time point. During testing at 6-months post treatment completion, participants will undergo a final MRI examination.

2.1.1 Procedures for Screening (Visit 1)

The following procedures will be performed:

1. Obtain written informed consent:

- A signed and dated informed consent form will be obtained from each participant before conducting any screening procedures. Participants will then be assigned an identification number by the WebDCU™ for purposes of confidentiality.
- A media release form will be completed with each participant and caregiver if applicable.
- All research staff authorized to obtain informed consent will have completed the Miami CITI course in the Responsible Conduct of Research and Protection of Human Subjects prior to their involvement with the study. Furthermore, they will be oriented to the study and trained by the study PI and study co-investigators who have extensive training and experience in the ethical and practical aspects of informed consent procedures.

2. Review inclusion/exclusion criteria.
3. Obtain medical history.
4. Conduct neurological examination.
5. Administer the MRI safety screening.

2.1.2 Procedures for MRI Examination (Week 1, 5, 12, 39, and 52+ (if eligible for the one-year post treatment follow-up))

1. Run the participant on high-resolution anatomical MRI scans during visit 1:
   - The participant will complete the fMRI Naming 40 where they will be instructed to name the pictures presented on the screen. If they see an abstract picture they are instructed to be silent. Administration time is approximately 10-15 minutes.
   - Following the fMRI Naming 40, the participant is instructed to lay still.
   - Complete additional functional and structural scans per follow-up

2.1.3 Procedures for Diagnostic Testing (Collected only once. Baseline may be split over three days (Visit 1, 2, and 3)

1. Administer the *Western Aphasia Battery-Revised* (WAB-R).
   - The WAB-R will characterize a participant's overall language impairment through the evaluation of the main clinical aspects of language functioning, including speech content, speech fluency, auditory comprehension, repetition, naming, and reading. The WAB-R allows for the differentiation of these specific language abilities, as well as the classification of aphasia type. The WAB-R also yields a composite score, the Aphasia Quotient, which provides an overall measure of severity, in which lower scores denote more severe aphasia (Kertesz, 2007). The reading section of the WAB-R: Part 2 will be completed to measure participants' oral reading ability and his or her reading comprehension of words and sentences. Speech- language pathologists (SLPs) will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will range between 30-45 minutes.

2. Administer the *Pyramids and Palm Trees Test* (PPTT).
   - The PPTT is a test of semantic processing. This test assesses the degree to which a participant can access meaning from pictures and words. Information from the test will help determine whether a participant's difficulty in naming or pointing to a named picture is due to a difficulty in retrieving semantic information from pictures, or a difficulty in retrieving semantic information from words, or, in the case of a naming failure, a difficulty in retrieving the appropriate spoken form of the word (Howard & Patterson, 1992). SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will range between 10-20 minutes.

3. Administer the *Kissing and Dancing Test*.
   The Kissing and Dancing Test is a short test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation (Bak & Hodges, 2003). SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will range between 10-20 minutes.

4. Administer sub-tests 1, 2, 8, 14, 15, 16 and 17 from the *Psycholinguistic Assessments of Language Processing in Aphasia* (PALPA):
   1. *Same-Different Discrimination Using Nonword Minimal Pairs.*
      - All the materials in this task are monosyllabic real words with a CVC structure and are designed for use in conjunction with word minimal pairs (see below). Good discrimination of nonwords implies that Auditory Phonological Analysis is intact. Poor scores may be due either to impaired hearing or to impaired phonemic perception. Stimulus pairs are minimally different according to voice, manner or place of articulation. Differences occur either in initial or final positions of pairs or in pairs that are metathetically related (in which the order of the sounds is reversed).
   2. *Same-Different Discrimination Using Word Minimal Pairs.*
      - All the materials in this task are monosyllabic real words with a CVC structure and are designed for use in conjunction with nonword minimal pairs (see above). By comparing results on this task with that using nonword minimal pairs, one can look firstly for any impairment in phoneme perception and secondly for whether impairment is reduced by lexical information.

   8. Nonword repetition.
      - All the materials in this task are nonwords. The purpose of the task is to test the integrity of sub-lexical acoustic –phonological conversion. It examines ability to repeat unfamiliar yet word-like sound forms in which length of utterance is systematically from one to three syllables. Although syllable length is manipulated, phoneme length is constant across the items.

   14. Rhyme Judgment Requiring Picture Selection.
      - This task assesses whether a subject can select a picture whose name rhymes with a given stimulus picture. The task requires that the subject is fist able to retrieve picture-names from the phonological output lexicon.

   15. Word Rhyme Judgments: Auditory Version.
      - This task aims to find out if the subject can detect whether a pair of words rhyme. For the task to be carried out successfully, the two items must be held in storage temporarily while the relevant parts of the word are segmented and compared. The task therefore tests the integrity of phonological short-term storage systems (Monsell, 1987), as well as input processing abilities and segmentation skills.

   16. Phonological Segmentation of Initial Sounds.
      - All the materials in this task are monosyllabic words and nonwords of CVC structures. This task provides information about a person's ability to segment at least the initial sounds of a heard string. By including nonwords as well as words, auditory analysis skills can be tested independently of lexical phonological processes.

   17. Phonological Segmentation of Final Sounds.
      - All the materials in this task are monosyllabic words and nonwords of CVC structures. This task provides information about a person's ability to segment at least the final sounds of a heard string. By including nonwords as well as words, auditory analysis skills can be tested independently of lexical phonological processes.

SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures.

5. Administer the *Northwestern Assessment of Verbs and Sentences* (NAVS)
   - The NAVS was designed for participants with aphasia and allows for detailed examination of verb processing (e.g. verb naming) as well as production and comprehension of canonical and non-canonical sentences (Cho-Reyes and Thompson, 2012).
   - SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will range between 50-60 minutes.
6. Administer subtest 7 (Matrix Reasoning) on the *Wechsler Adult Intelligence Scale* (WAIS)
   - This subtest is composed of four types of nonverbal reasoning tasks: pattern completion, classification, analogy and serial reasoning. The examinee looks at a matrix from which a section is missing and either identifies by number or points to one of five response options that complete the matrix. This test will allow for analysis of the cognitive status of the participants (Wechsler, 2008).
   - SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures.

7. Administer tasks used for scoring the *Apraxia of Speech Rating Scale* (ASRS)
   - This rating scale will be utilized to rate the presence and severity of apraxia of speech (AOS). It quantifies the frequency and severity of the characteristics associated with apraxia of speech (Strand et al., 2014).
   - SLPs will refer to (Strand et al., 2014) for explicit instructions regarding administration and scoring procedures.
   - To score the ASRS (offline), the following tasks should be administered:
     - DDK rates (alternating and sequential motion rates)
     - Sustained phonation (do 3x, record duration of “ah”)
     - Mayo Clinic “supplemental tasks for the ASRS." This includes word and sentence repetition tasks.
     - Cursory oral mechanism evaluation (ROM/strength/coordination of lips, tongue, jaw)

8. Administer *Philadelphia Repetition Test* (PRT)
   - This test was designed to test speech repetition ability. It uses the same items as the Philadelphia Naming Test, but instead of showing the pictures to the participants, the names are spoken to them, once only, and participants must repeat the names back to the experimenter (Dell et al., 1997).

9. Administer TALSA Battery (short term memory tasks) from Martin et al. (2012)
   - These tests assess verbal short-term memory.
   - TALSA Word Rhyming Triplet Judgment. This task assesses the ability to identify two rhyming words out of a selection of three. The non-rhyming foil will overlap phonologically with one of the rhyming words in one of three ways: same initial phoneme, same stressed vowel, or same final phoneme.
   - TALSA Rhyming Judgments. This task assesses the ability to identify if the words or nonwords presented rhyme with a 5 second delay presented and additionally a five second delay with a distraction presented. The nonwords are derived from the real words by altering one or two phonemes. The words and nonwords are presented separately.

- SLPs will refer to Martin et al. (2012) for explicit instructions regarding administration and scoring procedures.

10. Administer the Hit and Avoid Learning Tasks (Phonological, Semantic, and Spatial)
    - All participants will complete three learning tasks before initiation of treatment; two linguistic based learning tasks (phonological and semantic) and one nonlinguistic learning task (spatial). These will be administered in random order for each participant and will be completed by the end of treatment phase one. The tasks are set-up similar to the classical computer game "PONG."
    - Linguistic learning is completed by a game-like interface, where four words fall at randomly generated speeds from the top of a computer screen. Participants are asked to mouse over the words that fit the category of the game, such as 'animal' or 'has the "s" sound', and to avoid words that are not targets but may be semantically or phonologically related or unrelated.
    - Nonlinguistic learning will be tested by the same game-like interface. The task involves shape-patterns where participants are required to identify identical shapes to the target shape and to avoid distractor shapes. In this way, we circumvent internal language representations of task stimuli. Before beginning the task, the participant is shown a target shape and told that they will be presented with a computer screen where a variety of different shapes will fall from the top of their screen. The participant's objective is to mouse over the target shape and to avoid the other shapes (distractors).
    - All learning tasks will be conducted 5 times consecutively for 2 minutes each time. The accuracy score will be recorded for each session. The participant's learning ability is summarized as the increase in accuracy between baseline and final session. The study will compare linguistically based task vs nonlinguistic learning tasks (each to be completed by the end of treatment phase 1 to determine whether learning potential relates directly to prognosis.

11. Obtain *Arterial Blood Pressure and Stiffness*
    - Arterial stiffness plays a role in outcomes following cardiovascular disease and stroke, especially in individuals with hypertension. Accordingly, we will obtain measures of arterial stiffness using the SphygmoCor device. We will use data from this measue to inform our MRI analyses pertaining to brain health and its relationship with overall health.
    - SphygmoCor XCEL PWA (pulse wave analysis) uses a standard cuff to measure systolic and diastolic pressures, and capture a brachial waveform. The brachial waveform is then analyzed by SphygmoCor to provide a central aortic waveform. Central blood pressure measurements such as central aortic systolic blood pressure, and central pulse pressure. Aortic stiffness is approximated with non-invasive measurement of carotid-femoral pulse wave velocity, with improvements made with time to make the assessment procedure quicker and more user independent.
    - Participants will lay supine resting for 5 minutes prior to measurement. They will have a blood pressure cuff placed on their thigh. Trained study staff will make three key measurements (cm) from the placement of the cuff to the major physical landmarks. Study staff will locate the carotid and femoral arteries. Using the wand, gentle pressure will be placed on the neck of participants at the strongest pulse point and study staff will move the wand slowly on the neck area around the carotid

artery. Participants will feel pressure on their thigh, as is normal with having blood pressure taken. It will take approximately 15 minutes and participants will remain in the supine position resting throughout the testing.

2.1.4 Procedures for the Computerized Naming Assessments

Philadelphia Naming Test (Week 1, 5, 8, 12, 15, & 39 [Administered twice- Week 1])

1. Turn on the laptop computer and position in front of the participant
2. Open PNT application
    - Go to "Utilities" and "Start Webcam Preview" and stop when positioned correctly.
    - Enter participant number and session and click "Start."
3. Administer the *Philadelphia Naming Test* (PNT; Roach et al., 1996) on a laptop computer. The PNT is a picture-naming task consisting of 175 high- and medium-frequency nouns that vary in length from 1-4 syllables.
    - Instruct the participant to overtly name each picture as soon as it is displayed.
    - Trials will end following a response or after 30-seconds have elapsed.
4. Once completed, the video file will automatically save to Dropbox and you may exit the application.

Treated Naming 40 (Administered once on Week 1, 5, 8, 12, 15, & 39)

1. Turn on the laptop computer and position in front of the participant
2. Open Treated Naming 40 application
    - Go to "Utilities" and "Start Webcam Preview" and stop when positioned correctly.
    - Enter participant number and session and click "Start."

3. Administer the *Treated Naming 40*. The *Treated Naming 40* is a picture-naming task consisting of 40 target words including 20 target words from the Semantic treatment protocol and 20 from the phonological treatment protocol. The target word lists will include an equal number of nouns and verbs.
    - Instruct the participant to overtly name each picture as soon as it is displayed.

Trials will end following a response or after 30-seconds have elapsed.

4. Once completed, the video file will automatically save to Dropbox and you may exit the application.

2.1.5 Procedures for Discourse Assessments (Administered once on Week 1, 5, 8, 12, 15, & 39)

1. Turn on the laptop computer and position in front of the participant
2. Open Discourse application
    - Go to "Utilities" and "Start Webcam Preview" and stop when positioned correctly.
    - Enter participant number and session and click "Start."
3. Administer picture description task: Broken Window
    - Press esc when completed and file will save automatically
4. Administer Story Narrative: Cinderella Story
    - Press "Select to finish" button on screen when completed and file will save automatically
5. Administer Procedural Discourse Task: Making a peanut butter and jelly sandwich

- Press esc when completed and file will save automatically

6. Once completed, you may exit the application
7. SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will range between 20-30 minutes.

2.1.6 Procedures for the Aphasia Communication Outcome Measure (ACOM) (Week 1 and Week 39)

- Administer the Adaptive Aphasia Communication Outcome Measure (CAT-ACOM; Doyle et al., 2013).
- The ACOM consists of 12 items from the 59-item ACOM item bank using a content-balancing strategy from each domain (talking, writing, number use, comprehension, and naming) to insure that the content balance of each CAT-ACOM is reflective of the content balance of the full-length ACOM.
- SLPs will refer to the manual for explicit instructions regarding administration and scoring procedures. Administration time will be approximately 15 minutes.

2.1.7 Procedures for Treatment (Aphasia Participants Only) (Weeks 2-4 and 9-11)

Two different types of treatment will be used in this study; one primarily semantically focused and one primarily phonologically focused. Participants with aphasia will receive both types of treatments but in different order. The order of the sequence of treatments will be randomly assigned.

<u>Semantically focused treatment tasks</u>:

1. Semantic Feature Analysis. This task requires participants to actively analyze a concept and to attempt to generate its semantic features.

2. Semantic Barrier Task. Clinician and participant sit opposite to each other with a barrier between them so that they cannot see the picture the other one has in front of them. Both have a set of pictures and they take turns giving the other individual enough semantic information to identify the picture they have in front of them.

3. Verb Network Strengthening Treatment (VNeST). This treatment approach targets lexical retrieval of verbs and their thematic nouns. The objective is for the participant to generate verb-noun associates with the purpose of strengthening the connections between the verb and its thematic roles and ultimately improve sentence production.

<u>Phonologically focused treatment tasks:</u>

1. Phonological Components Analysis. This task requires participants to first name a given picture, and then identify the phonological features of the target word. Once the features have been identified, participants are required to attempt to name the picture again.

2. Phonological Production Task. This task requires participants to first sort a stack of targeted imageable nouns and verbs by one syllable and more than one syllable words

by tapping out each syllable. The treatment then continues to focus on the identification of a hierarchy of phonological features with a given pair of targeted words. Once each targeted feature is identified for the pair of words, the participant is then required to blend the syllables/sounds together.

3. Computerized Phonological Judgment Task. Participants will watch videos of a speaker producing words and nonwords, and will press a button (green for yes, red for no) to indicate whether successive stimuli rhyme, start with the same syllable, have the same number of syllables, etc.

Each treatment task will take about 15-minutes so the total treatment time will be 45 minutes per day.

2.1.8 Data Collection for Treatment Tasks

Semantic Feature Analysis:

- Target word is presented prior to beginning the feature analysis to determine if participant can independently name. Credit given if the participant can name correctly, or with >50% of the phonemes correct and being intelligible to an unfamiliar listener.

- Once the analysis is completed, and also reviewed, the participant is again asked to name the target word a second time. Same criteria for correct/incorrect is used.

Semantic Feature Barrier Task:

- Document **numbers of the target words** that were used in the task (ex: cards #21-28), and whether it was a closed set (where both clinician and participant have the exact **same** cards, per example both persons would have cards #21-28 in front of them) or an open set (where the clinician and participant have **different** cards in front of them, per example: clinician would have #21-24, and participant would have #25-28, each having only four cards in front of them)

Verb Network Strengthening Treatment (VNeST):

- After STEP 4 in VNeST protocol, Data is collected for Steps 5 and 6. The either do or do not receive credit for being able to: recall the target verb correctly, as well as 3 WHO's and 3 WHAT's. Credit can be given for any 3 correct WHO's and WHAT's even if they vary from the original 3 that were discussed during the VNeST activity.

Phonological Components Analysis:

- Target word is presented prior to beginning the feature analysis to determine if participant can independently name. Credit can only be given if participant names target correctly without any phonemic errors.

- Once the analysis is completed, and also reviewed, the participant is again asked at the end to name the target word a second time. Same criteria for correct/incorrect is used.

Phonological Production Task:

- Data is kept on which of the levels the participant participated and how many trials were completed in that day. Typically, 5 or more repetitions of one level are to be completed before moving on to the next combination; however, clinician can use judgment to move on to the next level if participant completed less than 5 in the time allowed, but was not exhibiting significant difficulty with the task.

- If participant goes through all of the levels, start over again at "First Syllable – First Syllable" and move through the cycle a second time.

Phonological Judgment Task:

- Data is kept on which of the 10 Levels the participant worked on, and what the % of accuracy was for the level(s) that day. Sometimes with time required for training the task there is only time for one attempt, but often time the task will be completed twice in one session. Clinician can use clinical judgement as to when to stay at a certain level even if the percentage was low, or when to move up even if the participant has not mastered that level at 80% (ex: if frustration level gets too high on a certain level, switching to a different task)

2.1.9 Procedures for Completion of Participant Satisfaction Survey

- At or around the 4-week follow-up testing date, all treated participants will be asked to complete an anonymous participant satisfaction survey. Participants may complete this online through a link that we email to them or will be given a paper copy of the survey with an addressed, stampled envelope to return the survey. A family member or friend will be asked to assist the participant in answering the questions on the aphasia-friendly survey.

2.1.10 Procedures for >one-year follow-up

- Participants who are at least one-year post-treatment will be called back for follow up testing to determine change in performance on a subset of the initial baseline measures or treatment outcomes. They will also be asked to complete the entire battery of MRI sequences. Participants will be paid $100 for this visit (in the form of a Visa giftcard). They will be provided lodging and roundtrip mileage reimbursement if they reside outside of the Columbia, SC metropolitan area (>35 miles travel distance).
- They will be asked to fill out a payment form and a W-9 so that the payment can be processed.
- Specific measures that will be administered at this time include the:
  - Western Aphasia Battery-Revised
  - Philadelphia Naming Test
  - Philadelphia Repetition Test
  - Pyramids and Palm Trees Test, Kissing and Dancing Test
  - Aphasia Bank Discourse Assessments
  - Updated case history form/health history form
  - PALPA subtests
  - National Institutes of Health Toolbox – Cognitive Battery

- Aphasia Communication Outcome Measure (ACOM)
- WAIS Matrix Reasoning Subtest
- Sphygmacor aterial stiffness measure

Appendix B
Scripts for Treatment tasks

Semantically focused treatment tasks

SEMANTIC FEATURE ANALYSIS (nouns & verbs)

- **We are going to look at some pictures. We will name the items in the pictures and also talk about many different features of these items.**
- *Get whiteboard*
- **I am going to place a picture in the middle of the board, and I would like for you to name the item in the picture if possible.** *(place in center)* **Can you tell me the name of this item?**
    - If they are able to correctly name, then **"Great, you said _____. Now let us review the different features."** *(write name of the item)*
    - If they approximate with an artic error, semantic or phonemic paraphasia, then **"Very close, you said _____. The name of the item is _____** *(and write the name of the item on the whiteboard).* **Now let us review the different features."**
    - If they are unable to name, then **"The name of the item is _____** *(and write the name of the item on the whiteboard).* **Now let us review the different features."**
    - *Note when reviewing features below – NOT EVERY FEATURE IS APPROPRIATE FOR EVERY TREATMENT STIMULUS. ONLY ELICIT THOSE FEATURES APPROPRIATE FOR THE STIMULUS ITEM.
- **The first feature is group or category membership.** (For very first session, **"For example, a triceratops is a dinosaur.") Your item** *(or just name it)* **is a member of which group?**
    - If produced, *write in box and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), write in box, move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **The second feature is use or utility.** (For very first session, **"For example, a telescope is used to look into outer space."**) **What is _____ used for?**
    - *If animal or person, *SKIP*
    - If produced, *write in box and move to next feature.*
    - If approximated*, shape to correct (model w/ carrier sentence), write in box, move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **The third feature is action.** (For very first session, **"For example, a triceratops eats plants and runs from the t-rex"**). **What does _____ do?**
    - If produced, *write in box and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), write in box, move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **The fourth feature is properties, or a description.** (For very first session, **"For example a triceratops has three horns and four legs"). What are some properties of _____?** (max 3)
    - If produced, *write in box and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), write in box, move to next feature.*

  - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **The fifth feature is location.** (For very first session, **"For example, a telescope is found near a window or on a rooftop"). Where do we find _____?**
  - If produced, *write in box and move to next feature.*
  - If approximated, *shape to correct (model w/ carrier sentence), write in box, move to next feature.*
  - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **The last feature is association.** (For very first session, **"For example, a triceratops reminds me of a rhinoceros"). What does _____ remind you of?**
  - If produced, *write in box and move to next feature.*
  - If approximated, *shape to correct (model w/ carrier sentence), write in box, move to next feature.*
  - If unable to produce, *provide answer (model w/ carrier sentence), write in box, move to next feature.*
- **To review the features again, try to say them with me**
  - **it is a _____** (group),
  - **it is used for _____, it _____** (action),
  - **it has/is _____,**
  - **it is found _____,**
  - **and it reminds you of _____.**
- **What is the name of this item?**
  - If produced, then **"Great, next item."**
  - If approximated, *shape to correct, have them repeat correct form after you,* and then **"Great, next item."**
  - If not produced, then **"Say this after me: _____"** OR **"Say it with me: _____" "Great work on that. Lets move to the next item."**

SEMANTIC FEATURE - BARRIER (nouns & verbs)

- **I am going to deal some cards out to the both of us. We will each have a matched deck of X* cards** (*5-10, depending on participant ability). **^ We will take turns describing the pictures to each other without saying the name of the pictures to try and get the other person to guess which picture we are describing.**
    - **^**If the participant is very good at this and/or as the treatment progresses, you may not need a matched deck. The matched deck is really to help narrow the field and give options for the participant to choose from.
- **Before we start giving clues to each other, I want to quickly review the semantic features that you may decide to use as clues.**
- *Get semantic barrier sheet*
- **For example, if you were holding a picture of _____** *(select sample card),* **you might provide a clue like** *(give an example sentence about 'group' feature)* **or** *(give an example clue about 'use' feature)*. **If I am not able to guess which picture you are describing, you could provide another clue like** *(give an example sentence about 'physical properties')* **or** *(give an example sentence about 'association')*.
- **If you have a hard time giving me clues, I will ask you to describe, draw, write, or gesture some features. If you are still having difficulty, I will ask you to describe or gesture the use or action. If we are still stuck, I will ask you to name the picture if possible and/or show me your card. We will review a few features together before moving to the next picture.**
- *Place barrier between you and client, give stack of cards*
- For very first session, clinician goes first to model
- **Okay, I am going to give you 2 clues.**
    - **First** – *(give group info)*, **and**
    - **Second** – *(give use info, gesturing when appropriate)*
- **Can you guess which picture I am describing?**
    - If correct guess, **"Yes, I was describing ___(target word)____.** (*summarize features).* **Now, your turn."**
    - If incorrect guess, *provide 2 additional clues/features appropriate for the stimulus item, gesturing when appropriate*. **"Can you guess which picture I am describing?"**
        - If correct guess, **"Yes, I was describing ___(target word)___.** (*Summarize features).* **Now, your turn."**
        - If incorrect guess, *review features you have already provided, provide 1-2 more if appropriate,* then **"I was describing ___(target word)___. Can you say that word with me? _____."**
- **All right, now it is time for you to give 2 clues. You can use your feature visual to help get you started.**
    - Hopefully they provide 2 clues. ******If they do, and if it is obvious, *guess correctly.* If they do, and if you are pretty confident you know what it is but if it is the least bit ambiguous, *guess incorrectly, choosing the alternative.* They would then need to provide more clues.
    - If they are only able to provide 1 clue, *ask them if they can give you more information. ***
    - If they are unable to provide clues, *ask them to describe, draw, write, or gesture 2 specific features (name the specific features on their chart, direct them to their visual aid). ***

        - If still having difficulty, *ask participant to describe or gesture the use or action.* **
        - If still having difficulty, *ask participant to name the picture if possible and/or show the card.*
    - Once clinician has correctly guessed or participant has named or shown picture, *review features*. **"You were describing ___(target word)___. Can you say that word with me? _____" Now, please fill in the blank with the features you provided:** *(review only those provided, unless unable to provide, in which case, select 2-3 features to review)*
        - **it is a _____ (group),**
        - **it is used for _____, it _____ (action),**
        - **it has/is _____,**
        - **it is found _____,**
        - **and it reminds you of _____.**
- *Take turns until time is up.*

For week 2, we can consider alternating between Clue-giving and 20 questions during the semantic barrier portion of the session.

- **Now we are going to talk about some verbs and try to make sentences using some cards.**
    - *Pick a card with a target verb for the session out of the 10 possible verbs and put it on the table in front of you.*

Step 1. Generation of 3 agents and patients for targeted verb.

1. **For this step I want to ask you to tell me who can __(target verb)__** (e.g. **'tell me who can measure'**) * The participant does NOT have to say the verb, only the nouns.
    - *Write down the word 'who' at the top left corner of a dry erase board.* For each word generated*, write the word underneath 'who' or place a card with the word written on it. Encourage the participant to provide personal responses.*
    - A minimum of 3 words is required before moving to next step. If correct, *move on to the next step.*
    - If participant cannot produce 3 words, *give the participant 2 cards with appropriate responses on them and 1 foil written on them.* ***"*Now I have given you some cards to choose from. Can you tell me which ones you think go with the verb and which ones do not?"**
    - If still incorrect*, provide the participant with the correct answer and get them to say it with you.*
2. **Now, I want to ask you to tell me what can be_(target verb)_** (e.g. **'what can be measured'.**
    - *Write down the word 'what' at the top right hand corner of the dry erase board.* For each word generated*, write the word underneath 'what' or place a card with the word written on it.*
    - If participant cannot produce 3 words, *give the participant 2 cards with appropriate responses on them and 1 foil written on them.* ***"*Now I have given you some cards to choose from. Please read the word pairs aloud and then tell me which ones you think go with the verb and which ones do not?"**
    - If correct, *move on to the next step.*
    - If incorrect*, provide the participant with the correct answer and get them to say it with you.*
    - *Once 3 agents and 3 patients have been produced, move on to next step.*

Step 2: Read aloud each triad. (it's not important to conjugate the verb or add any articles to the nouns (e.g. 'farmer drives tractor').

- **They may be read aloud independently, in unison with clinician, or repeated after the clinician.**

Step 3. Answer wh-questions about agent-patient pair

- **Now I want you to pick one pair of person and item that you want to discuss further.**
    - *Write or move the pair of words (agent-verb-patient) that they choose out of the list of agents and patients to focus attention. Write or place cards with 'where', 'why' and 'when' on the dry erase board.*
- **Now I am going to ask you to answer some questions about this pair of words.** (The participants do NOT have to respond with complete sentences). **A) Where does a _(agent-verb-patient)_?** (e.g., where does a chef measure sugar **B) Why does a _(agent-verb-patient)_?** (e.g., why does a chef measure sugar?) **C) When does a _(agent-verb-patient)_?** (e.g., when does a chef measure sugar?)

- If correct, *move on to the next step.*
- If incorrect, *give the participant options to choose from and say it with them.*

Step 4. Semantic judgment of sentences.

- *Remove all the cards from the table.*
- **Now I am going to read some sentences and I want you to tell me if the sentences make sense (are semantically correct) or not.**
  - *Read the 12 sentences containing the target verb* (3 correct, 3 with inappropriate patient, 3 with inappropriate agent, 3 with agent and patient switched).
  - After each sentence, ***"*Did this one make sense?"**
  - **If participant answers 5 out of 6 correctly, you may move on.**
  - **To challenge participant you may ask them to correct the sentence.**
  - Once successful, *move on to the next step*

Step 5. Produce target word independently

- Now I want you to tell me what verb we were working on today. If they are unable to produce independently then say the verb and have them repeat.

Step 6. Generation of 3 agent-participant pairs (repeat Steps 1-2)

- **Now we are going to go back to the first thing that we did today, but this time we are not going to use the cards.**
  - *Ask the exact same questions that were asked in Step 1, but without the cards. (i.e.,* **Tell me who** (or **what**, to be alternated throughout a session) **can __(target verb)__** (e.g. **'tell me who can measure'**) **(be _(target verb)_** (e.g. **'tell me what can be measured'**)**.)**
  - *Provide the participant with general feedback* (e.g. **"good job"** etc.)
- Once successful, *start over from the top using a different verb.*

Phonologically focused treatment tasks

PHONOLOGICAL COMPONENTS ANALYSIS

- **We are going to look at some pictures. We will talk about different sound features of the objects or actions in the picture.**
- *Get whiteboard*
- **I am going to place a picture in the middle of the board, and I would like for you to name the item in the picture if possible.** *(place card in center*). **Can you tell me the name of this item?**
    - If they are able to correctly name, then **"Great, you said _____. Now let us review the different features."**
    - If they approximate with an artic error, semantic or phonemic paraphasia, then **"Very close, you said _____. The name of the item is _____***)*. **Now let us review the different features."**
    - If they are unable to name, then **"The name of the item is _____ Now let us review the different features."**
- **The first feature is the first sound of the item you see.** (For very first session, **"For example, the word lotion starts with the sound 'l'.**) **What is the first sound of this item?**
    - If produced, *write in box, then read together** **"the first sound is ____"** *and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the first sound is ____"** *and move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the first sound is ____"** *and move to next feature.*
        - *for all features, the participant does NOT have to say the carrier phrase if they cannot or do not wish to, just try to encourage them to fill in the blank with you.
- **The second feature is the number of syllables.** (For very first session, **"For example, the word lotion has two syllables"***(tap out 2 syllables)*.) **How many syllables does this word have?**
    - If produced, *write in box, then read together* **"the number of syllables is _____** *(then tap), and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the number of syllables is _____** *(then tap), and move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the number of syllables is _____** *(then tap), and move to next feature.*
- **The third feature is the last or final sound of the item you see.** (For very first session, **"For example, the final sound of the word lotion is "n".**) **What is the final sound?**
    - If produced, *write in box, then read together* **"the last sound is ____"** *and move to next feature.*
    - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the last sound is ____"** *and move to next feature.*
    - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the last sound is ____"** *and move to next feature.*
- **The fourth feature requires us to think of or make up a word that rhymes with our item.** (For very first session, **"For example, a word that rhymes with lotion might be "ocean" or "motion".**) **What is a word that rhymes with _____?** (max 2)

  - If produced, *write in box, then read together* **"____(target)___ rhymes with ____(rhyming word)___"** *and move to next feature.*
  - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"____(target)___ rhymes with ____(rhyming word)___"** *and move to next feature.*
  - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"____(target)___ rhymes with ____(rhyming word)___"** *and move to next feature.*
- **The fifth feature is the vowel sound of the first syllable.** (For very first session, **"For example, the vowel sound of the first syllable of lotion is "o"**). **What is the vowel sound of the first syllable?** (If the word is monosyllabic, acknowledge that and feel free to change your script to **"the vowel sound of this word".**)
  - If produced, *write in box, then read together* **"the vowel sound of the first syllable (or word) is _____"** *and move to next feature.*
  - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the vowel sound of the first syllable (or word) is _____"** *and move to next feature.*
  - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the vowel sound of the first syllable (or word) is _____"** *and move to next feature.*
- **The last feature** (IF MORE THAN ONE SYLLABLE) **is the vowel sound of the last syllable.** (For very first session, **"For example, the vowel sound of the last syllable of lotion is "uh".**) **What is the vowel sound of the last syllable?**
  - If produced, *write in box, then read together* **"the vowel sound of the last syllable is _____"** *and move to next feature.*
  - If approximated, *shape to correct (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the vowel sound of the last syllable is _____"** *and move to next feature.*
  - If unable to produce, *provide answer (model w/ carrier sentence), prompt to repeat, write in box, then read together* **"the vowel sound of the last syllable is _____"** *and move to next feature.*
- **Great, we've reviewed all of the features. I want to go over them again, and I want you to say each feature with me.** (*provide carrier phrase*, they fill in with you with unison speech)
  - **The first sound is _____,**
  - **The number of syllables is ____#____**,*(and then tap out together)*
  - **The last sound is _____,**
  - **A rhyming word is _____,**
    - (if listed 2, **Another rhyming word is _____**),
  - **The vowel of the first syllable is _____,**
  - **The vowel of the last syllable is _____.**
- **What is the name of this item?**
  - If produced, then **"Great, next item."**
  - If approximated, *shape to correct, have them repeat correct form after you,* and then **"Great, next item."**
- If not produced, then **"Say this after me: _____"** OR **"Say it with me: _____" "Great work on that. Lets move to the next item."** *Go through all the steps using a different picture.*

PHONOLOGICAL PRODUCTION TASK

- **We need to sort the picture cards into words that have only 1 syllable and words that have 2 or more syllables.**
- **I'll do the first few.** (*Pick a card and tap the word while you say it.)* **That had X syllable(s) so it goes in this pile.** *(Do for a few more cards.)*
- **Okay, now we will work together. This word is "_____"** *(and tap at same time).* **How many syllables does it have?** OR **Can you tap it out for me?**
    - *Let them try independently first.*
    - *If incorrect,* "**Close, but watch me** – *(say the word and tap).* **X syllable(s)."**
- Complete entire stack or stop at **5 minutes**.

- **"Now we have a stack of single syllable words and a stack of multi syllable words. We are going to break these words down into smaller syllables and sounds and then mix them all up to make new words. These words will not make any sense most likely, but they will be new words we are creating."**

- First syllable – First syllable (with multisyllabic cards only)
    - **"First we will focus on the first syllable of words."**(*Place target picture on board.)*
        - **This word has X syllables.** *(tap all syllables).* **The first syllable of this word is "___"** *(write syllable on white board)*
        - Place next picture on board.
        - **This word has X syllables.** *(tap all syllables).* **The first syllable of this word is "___"***(write syllable on white board)*
        - **Now, let us make a new word with the first syllables of these words. "_____" plus "_____" makes "_____"** *(touch syllables while saying the isolated syllables, and then pull written syllables together to make the final nonword while you are talking)*
        - **"Now, let's review and say these together."**
            - **The first syllable of this word is "_____"** *(say together)*
            - **The first syllable of this word is "_____"** *(say together)*
            - **"_____" plus "_____" makes "_____"** *(say together, the syllables and nonword, or the nonword at the very least)*
- **Complete 5 trials.**
- ***Follow the same steps with the given hierarchy:***

➢ First syllable – First syllable (multisyllabic cards only)
➢ First syllable – Last syllable (multisyllabic cards only)
➢ Last syllable – Last syllable (multisyllabic cards only)
➢ Last syllable – First syllable (multisyllabic cards only)
➢ First syllable – First sound (multisyllabic cards for first, single syllable cards for second)
➢ Last syllable – Last sound (multisyllabic cards for first, single syllable cards for second)
➢ First syllable – Last sound (multisyllabic cards for first, single syllable cards for second)
➢ Last syllable – First sound (multisyllabic cards for first, single syllable cards for second)

DECISIONS

- Monitor what subtask they finish each day and start with next level on the hierarchy if 5 trials have been completed. If less than 5 trials were completed and participant is exhibiting significant difficulty, repeat current level. Use clinical judgment to move throughout each level.

COMPUTERIZED PHONOLOGICAL JUDGMENT TASK

- *Open Phonological Video Treatment application on the laptop*
  - *Enter participant number, session number and treatment level.*
  - *Press start.*

- *Read the instructions that come up on the computer screen out loud to the participant and make sure they understand what is being asked of them. Give them an example of the task, e.g. if they're doing the first task you give them an example of two words that have the same number of syllables and two that don't.*
  - "In this task, you will see two videos at a time. Each video is of a person's mouth saying a single word. Listen closely to both words"
  - Complete the practice first.
  - Once the practice is completed, press start to begin task

LEVEL 1: Same number of syllables?
- You will see 2 videos.
- Press GREEN if they have the SAME number of syllables.
- Press RED if they do not.

LEVEL 2: Which has more syllables?
- You will see 2 videos.
- Press GREEN if the FIRST word has more syllables.
- Press RED if the SECOND word has more syllables.

LEVEL 3, 4, & 5: Start with the same sound?
- You will see 2 videos.
- Press GREEN if they START with the SAME sound.
- Press RED if they do not.

LEVEL 6, 7, & 8: End with the same sound?
- You will see 2 videos.
- Press GREEN if they END with the SAME sound.
- Press RED if they do not.

LEVEL 9 & 10: Do they rhyme?
- You will see 2 videos.
- Press GREEN if they rhyme.
- Press RED if they do not.

Criteria for each level:
- If participant scores **<60%** the program will recommend they move down a level.
- If participant scores between **60%-80%** then they remain at the same level.
- If participant scores **80% or above**, they should move up to a harder level.
- If a participant were to achieve an **80% or above** on all levels prior to completion of treatment phase. Begin at level 1 and increase the target score to **90%** or higher for each level.

DECISIONS:
- Clinical judgment may be used to determine whether the participant moves up or down a level in certain circumstances. For example, if the participant has just moved on to a new level and they are not demonstrating understanding of the task and they score below 60%, clinical judgment can be used to override the protocol for dropping back a level. Another example would be if the participant is frustrated and stuck on a particular level for several attempts, clinical judgment can be used to allow them to try the next level up (particularly if it is a separate targeted skill).

Appendix C: Alternative Remote Protocol During COVID 19 Quarantine

A. Modifications for Remote Assessment

To continue to enroll participants during the Covid 19 quarantine, some procedures will be adapted to be able to administer assessments online. Changes are listed and explained below.

1. Online Platform – Individuals completing the assessments will be assigned a private, zoom account. This online platform allows researchers to connect to participants in their own homes via video-conferencing. Video-conferencing will be password-protected and any videorecordings made during the assessment will be saved to the researcher's local computer and then transferred to a HIPAA-compliant dropbox account. Participants will be asked to have a friend or family member (assistant) to be present throughout the assessment to assist with the technology.

2. Informed Consent – Potential participants will be mailed 2 copies of the informed consent document with a pre-addressed, stamped envelope in advance of the assessment session. The individual consenting the participant will review the informed consent via video-conferencing with his/her assistant who can assist with the technology as needed. The participant (or his/her assistant) will sign the consent and show it to the clinician via his/her webcam. The clinician will take a screen shot of the form for a digital copy. The participant will be asked to mail the original, signed form back to the Aphasia Lab in the stamped, pre-addressed envelope provided.

3. Behavioral Assessments – Behavioral assessments have been adapted to be administered online. A brief training session will take place before administration of assessments to make sure that the participant and his/her assistant can demonstrate how to use controls in the zoom platform to point to pictures on a shared screen. The assistant will also be instructed on how to alter the webcam to allow the clinician to view the participant completing different tasks.

4. MRI/fMRI Tasks – MRI scans and the fMRI tasks will be postponed for participants until the COVID 19 quarantine is over. At that time, participants will be contacted for an MRI/fMRI scanning appointment at MCBI.

B. Modifications for Remote Treatment

1. Teletherapy Kits – Each enrolled participant will be mailed a teletherapy kit to complete treatments. These kits will include:

- A laptop, touchscreen computer pre-loaded with the treatment apps, zoom for online videoconferencing with the SLP and Team Viewer which will allow the SLP to see the participant's computer to help with initial set-up and troubleshooting.
- A high quality headset with a microphone to maximize communication with the SLP online.
- A mouse for optional use.
- A mobile WiFi hotspot if a participant does not have adequate WiFi connection at home.

2. Treatment Apps – Participant laptops will be preloaded with both semantic and phonological treatment apps which wil allow the SLP to interact with the participant during treatments. The assistant will be asked to help with the treatment technology if necessary.

Appendix D. POLAR Research Plan

1. Significance.

Although some stroke patients experience significant spontaneous recovery in the first few days and weeks following onset, approximately 30-40% of patients are left with persistent aphasia affecting communication ability and life participation. Stroke is generally associated with older age, but a significant proportion of patients incur stroke in the fourth to sixth decade of life (15). Once spontaneous aphasia recovery has ceased, patients' greatest chance at significant communication improvement is by means of aphasia treatment. Although considerable evidence suggests that aphasia treatment is effective, the problem is that very little is known about what kind of treatment works best, and for whom (2, 3, 16, 17). As suggested in a recent paper by Fama and Turkeltaub (18), the *standard of care* for the treatment of aphasia in the United States is behavioral speech-language therapy. However, what exactly this therapy consists of probably varies widely - what might be common practice at one clinical facility might be very different than what is practiced somewhere else. As anyone who has extensive experience treating aphasia knows, it is simply not the case that all treatment approaches work equally well, not even for patients with similar impairments (19). Nevertheless, we believe that an appropriately weighted use of combined behavioral and neurophysiological biomarkers should improve our ability to predict treatment outcome, as well as aid in the choice of optimal treatment targets for individual patients with aphasia.

In the most downloaded article in the history of the Public Library of Science, "Why Most Published Research Findings Are False," Ioannidis (20) noted that when all factors are equal, in fields where smaller studies are the norm, the chances of research findings being false is much higher than in fields where larger studies are typical. The norm in aphasia treatment studies, including our own previous studies, is to rely on single-subject design or very small group studies. It is straightforward to understand why this is the state of affairs. Treatment studies take a relatively long time and a disproportionate level of resources compared to most other kinds of research in aphasia. Also, subject recruitment is often an issue, especially in studies that target patients with specific cognitive/linguistic impairment profiles. Nevertheless, our field needs larger treatment studies to reveal specific trends that can be generalized to the larger population of patients, not just to a narrow number of cases. We believe that our past success, institutional environment, and collaborative spirit puts us in a position to conduct a relatively large aphasia study that has the potential to exert a sustained impact on the field of clinical aphasiology. In the context of previous aphasia treatment studies, the project proposed here entails a relatively large sample size of treated patients (N=120) and extensive cognitive/linguistic testing, including assessment of learning potential and neuroimaging workup. As we contend that recent meta-analyses studies have demonstrated the value of aphasia treatment, the purpose of project 1 is not to test the efficacy of treatment, but rather to determine why some patients respond better to treatment than others and whether treatment response relates to which language domains are targeted in treatment. In addition to testing hypotheses in project 1, the data collected here will also fuel projects 3 (PI: Rorden) and 4 (PI: Hickok).

1.1 Biomarkers of Treated Aphasia Recovery. In 1972, Darley (21) suggested that factors such as age, education, intelligence, social status, health, time post-stroke, and aphasia type are related to success in aphasia therapy, but that very few studies had actually incorporated these factors as predictors of outcome. More than four decades later, the situation has not improved much, as limited research has established a strong relationship between specific patient factors and treated aphasia recovery. A few studies have related patient characteristics to spontaneous recovery (6, 7, 22-24); however, it is not known whether the same factors would predict aphasia treatment outcome. Accordingly, clinicians who routinely treat patients with aphasia have very limited empirical data to guide their treatment and predict outcome. Whereas biomarker studies are common in medicine, almost no studies have been carried out to understand the relationship between patient factors and aphasia rehabilitation potential in stroke (25-27). This may not be a particular problem in academic settings where extensive diagnostic behavioral tests can be administered. But in clinical settings such as rehabilitation hospitals, sub-acute nursing facilities, long-term care facilities, and home health care

settings, financial and personnel resources are limited and often allow for only one initial assessment session (28). The point here is that the clinicians who provide most of the aphasia therapy in the United States could far better serve their patients if they were armed with prognostic factors that would indicate whether a given patient was likely to respond to direct speech and language therapy (SALT), and what focus that treatment should assume for maximum recovery. In the spirit of Darley's argument, we believe that patient factors such as aphasia severity and patterns of cognitive/linguistic impairment may, indeed, provide clinicians with the necessary biomarkers regarding the type of treatment that should be utilized and the extent of expected improvement. Therefore, the major goal of Aim 1 is to build a predictive model of treated aphasia recovery based on biographical and cognitive/linguistic factors. Henceforth, we refer to this model as the "default model" of outcome prediction.

1.2. Models of Treated Aphasia Recovery. In a recent publication, Geranmayeh, Brownsett, and Wise (29) proposed a model to explain where in the brain plastic changes that support aphasia recovery occur. Their model is mostly based on neuroimaging studies conducted in the past two decades, but it provides far greater insight into spontaneous rather than treated aphasia recovery. As such, this model offers minimal guidance regarding the current project where the goal is to predict outcome based on biographical, cognitive/linguistic, and neuroimaging factors. It is widely assumed that a large number of factors predict how likely a given patient is to respond to aphasia treatment. However, we are not aware of comprehensive theories that attempt to predict aphasia treatment outcome. As a first step, we adopt a data-driven approach to explain response to aphasia treatment. By most accounts, the process of theory building includes the following steps: 1. Observation; 2. Description; 3. Categorization; and 4. Analysis and a statement about causality. For Aim 1, the first three steps are accomplished by data collection and data organization that rely on patient and caregiver questionnaires, direct behavioral testing, and MRI scanning. As we explain in section 3.10 (Data Analyses), the final step is achieved by first using step-wise multiple regression to identify the most important predictors and then rely on support vector regression analyses (SVR; See Neuroimaging Core for details) to compare different explanatory models. For the sake of clarity, we are not going to focus on the potentially very complex interactions between the environment (e.g., family support, access to healthcare, and residential setting) and treatment outcome, as doing so would probably constitute an increase in effort that far exceeds what can be supported in the current project. Instead, our default model focuses on two sets of factors that we propose are related to aphasia treatment response. 1. *Biographical factors:* Age (30, 31); education (32, 33);, gender (23, 34); handedness (35); age at stroke, type of stroke, number of strokes, and time post first stroke that caused the aphasia (36); 2. *Cognitive/linguistic factors:* Clinical assessment of aphasia (6, 7, 37), motor speech production (38), grammatical processing (39), semantic processing of verbs and nouns (40-42), phonological processing (43, 44), speech discrimination (45-47), speech repetition (48-50), executive functioning (51-56), and verbal short-term memory (57-60). We have chosen to cite research that supports the role of the aforementioned factors as predictors of aphasia treatment outcome; however, it is beyond the scope of the proposal to speculate how these factors might interact to support treated aphasia recovery. Once we have built our final data-driven model(s), it will be possible to examine the specific interactions between the factors that make up each model.

Aphasia severity is one of the very few factors that has been identified as a reliable predictor of aphasia treatment outcome, and it is generally accepted that more severe patients are less likely to respond (61, 62). It is a caveat, however, that aphasia severity is a multidimensional construct where different patients with severe aphasia might present with very different language impairment profiles. We believe that contemporary neuropsychological models of speech and language processing can help us better understand which aspects of language impairment predict response to aphasia treatment. This may seem like a straightforward notion. However, it is important to note that neuropsychological models of speech and language are typically designed to explain how specific processes are rooted and related in the brain, but not to account for dynamic changes in response to behavioral manipulation. Nevertheless, if speech and language impairment can be predicted based on lesion data (63-73), it seems reasonable to expect that the extent of brain damage and lesion location in relation to the cognitive processes they support may also aid in the prediction of aphasia treatment outcome. The dual stream (DS) model was conceived to explain how auditory information maps onto the phonological-

articulatory (dorsal stream) and conceptual systems (ventral stream) in the brain (1, 74, 75). Even though this model has been highly influential in speech and language research for over a decade, it has not gained much traction in the aphasia rehabilitation literature. In Aim 2, we will test whether proportional damage to the anatomically defined dorsal and ventral streams improves outcome prediction beyond what can be achieved with the default model in Aim 1. Specifically, by adding factors that comprise the default model and DS model factors (proportional damage to the dorsal and ventral streams) we will be able to predict treatment outcome with greater accuracy compared to the default model alone. We refer to this as the "default model + DS model" (D+DS model). Note, however, that damage to any area that is supplied by the middle cerebral artery in the left hemisphere is likely to improve prediction beyond the default model. Hence, we will compare the utility of the DS model to the Wernicke-Lichtheim-Geschwind anatomical model (76-78) to predict aphasia treatment outcome. Accordingly, the crucial questions we aim to answer are whether the default model can be improved by combining it with the DS model and whether the DS model fares better in comparison to its older counterpart, the Wernicke-Lichtheim-Geschwind model (D+WLG model). We certainly recognize that the WLG model has been fervently challenged for well over a century and that its power to explain individual patients may be limited. However, most published aphasia research (e.g., 79-82), including contemporary lesion studies (83, 84), continues to report aphasia types based on the WLG model and, in some ways, the model could be considered a simpler version of contemporary dual stream models of speech processing. It is widely recognized by clinicians in neuropsychology, behavioral neurology, and speech-language pathology. As such, the WLG model serves as a reasonable control to which to compare the DS model.

Figure 1. A schema depicting the three models that will be tested for predicting aphasia treatment outcome in Aims 1 and 2.

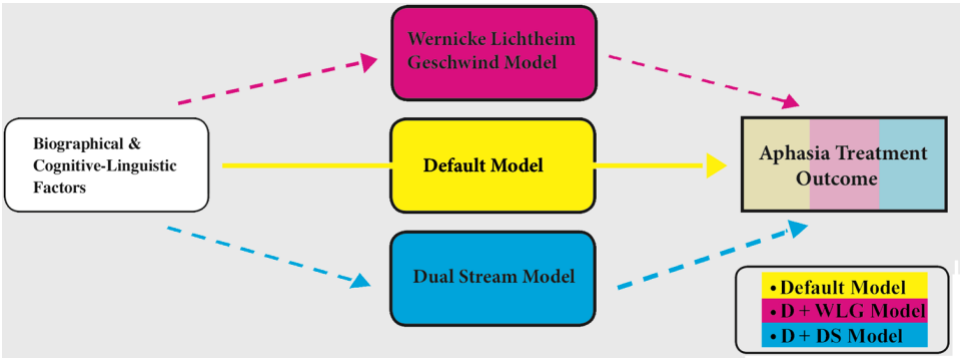


Figure 1 shows the relationship between the models that will be tested in Aim 1. The default model (yellow) includes biographical and cognitive/linguistic factors that are collected in the first few patient sessions. A different model assumes improved outcome prediction by combining the default model with proportional damage to the cortical areas that comprise the WLG model (red; D+WLG model). We hypothesize that the strongest prediction of outcome will be achieved by combining the default model and the DS model (blue; D+DS model).

Along with several other groups, we have demonstrated that functional brain changes in the residual language areas of the left hemisphere support treated improvement in naming in aphasia (77, 78, 85-87). Based thereon, we could conclude that strengthening of the residual left-hemisphere language network, the areas that remain intact after stroke, leads to improved language processing in aphasia. However, as far as we can tell, no research has examined whether this kind of network compensation relies on overall network integrity or survival of specific sub-components of the network. We suggest that the DS model offers a principled approach to answer such questions. For example, for cases of damage to the ventral stream, and concomitant semantic impairment (a component of the default model), treatment might focus specifically on semantic stimulation. Here, one could assume that preservation of the ventral stream would be related to treatment success, so that patients with greater damage to the ventral stream are less likely to respond compared to others whose ventral stream is relatively more preserved. This view would contend that successful rehabilitation relies on the survival of the network that is being stimulated. However, an opposing view could argue that what really matters

is cross-domain compensation where the dorsal stream is forced to become more strongly involved to support improvements in language. The same logic could be applied to the WLG model, but with less detail, as it is a relatively cruder model compared to the DS model. Hence, because the DS model provides a more detailed model of language organization in the brain, we hypothesize it will outperform the WLG model and provide better prediction of treatment outcome. Along these lines, we propose that reliance on the DS model can potentially improve our understanding of why some patients show greater improvements than others and relate the findings in neuroanatomy. As the DS model is a functional model that is grounded in neuroanatomy, it could be the case that measures of proportional damage to dorsal and ventral streams are redundant with the speech and language measures included in our test battery. In that case, the D+DS model will not outperform the Default model. Dr. Rorden's project (Project 3) will rely on the same behavioral data collected here, but unlike the current project, it focuses on developing sophisticated neuroimaging analyses to understand the relationship between localized brain damage and aphasic impairment/recovery. The benefit of this setup is that each project may yield complementary results that can better explain differences in treatment response among patients.

1.3. Comparing Biomarkers as a Function of Treatment Foci. One of the challenges to interpreting the value of aphasia treatment biomarkers is the generalizability across different kinds of treatment types and foci. For example, in one of our previous studies (38), the severity of apraxia of speech (AOS) was found to be a strong negative predictor of success in speech entrainment treatment, a training program that emphasizes real-time mimicking of audio-visual speech narratives by patients with non-fluent aphasia. Although AOS is generally recognized as a deterrent to aphasia treatment success (88), it is not guaranteed that the same relationship between AOS severity and patient improvement would have been realized had the treatment included a different approach such as Melodic Intonation Therapy (89, 90), a treatment program also specifically designed to improve speech production in non-fluent aphasia. For that reason, Project 1 will rely on a cross-over design where each patient completes one treatment phase that focuses on semantic stimulation and another treatment phase where the focus is on phonological stimulation. Accordingly, we will be able to compare biomarkers of treatment outcome and determine if the same factors predict success, regardless of treatment focus, or whether biomarkers are dependent on treatment focus. In one of our previous studies (10, 85), a group of 30 patients with aphasia underwent anomia treatment in a cross-over design where one half of the patients first received 15 hours of treatment using a phonological cueing hierarchy and then later received 15 hours of treatment with a semantic cueing hierarchy. The remaining half of the patients received treatment in the opposite order, semantic treatment first and then phonological treatment. Overall, there was a wide range in treatment responses across the patients, and treatment success with phonological stimulation was not correlated with success in the semantic treatment. Relying on data from this study, we conducted *post hoc* analyses to provide proof of concept for the current project (see Section 3.1 – Preliminary Studies). Specifically, we used a subset of the factors that comprise the models shown in Figure 1 (default model, D+DS model, and D+WLG model) to predict success in semantically and phonologically focused anomia treatments. Our decision to divide the treatment phases to target semantic versus phonological stimulation was based on the fact that most impairment based treatment approaches tend to assume a semantic or phonological focus (10, 91-99). However, despite the fact that aspects of semantic and phonological processing are thought to dissociate at both the behavioral (40, 91, 92, 100-104) and cortical level (105-107), it is practically impossible to dissociate semantic and phonological stimulation in aphasia treatment. Accordingly, we emphasize that we use treatment methods that stress phonological stimulation but will also induce semantic processing, and vice versa. The key here is *treatment focus* rather than complete dissociation of targeted processes in treatment.

In addition to relating patient factors to treatment outcome, Aim 1 will take advantage of data collection in the CATES trial ('Trans-cranial Direct Current Stimulation to Treat Aphasia: Phase II Trial' U01 DC011739). As we discuss in greater detail in section 3.1 (Preliminary Studies), the aphasia treatment in the CATES trial relies on a computerized training task that includes both phonological and semantic stimulation. Each of the two treatment phases used in the current project will emulate the timing components (timing of pre-post treatment testing, length and number of treatment sessions) of the treatment phase in the CATES trial. Accordingly, we will be able to compare potential predictors of aphasia treatment success across the current project and the CATES trial.

1.4 Learning Potential as a Biomarker of Aphasia Treatment Outcome. Learning is driven by neuroplasticity – the ability of the brain to continuously adapt its structure and function based on internal and external environmental changes. Neuroplasticity occurs due to changes in neuronal morphology, glia, and vascular and metabolic processes and can be stimulated by a variety of sensory-motor experiences (108-112). Learning potential is a common concept in the education literature (113, 114). It assumes that not all individuals have the same ability to acquire new information and probably reflects flexibility of neural systems to adapt to internal changes or external stimulation (112, 115-117). Although most clinicians probably assume that some aphasic patients have greater rehabilitation potential than others, it is not clear whether patients' learning potential is related to aphasia treatment outcome. Unlike the measures included in Aims 1 and 2 that mostly represent static factors (e.g., performance on cognitive/linguistic tests, proportional damage to specific brain regions), learning potential represents dynamic change, something that could be assessed at baseline to understand if prolonged aphasia rehabilitation is likely to be effective. Aim 3 is therefore to assess the predictive value of learning potential for aphasia treatment outcome.

2. Innovation.

2.1 Developing a Model to Predict Aphasia Treatment Outcome. Comparing the outcomes of treatments that assume either a semantic or phonological focus is not new in the aphasia literature (91, 93, 94, 96, 98, 99). However, most of this research has relied exclusively on single-subject design or very small group studies. To the best of our knowledge, no previous studies that included relatively large sample sizes have attempted to identify biomarkers of aphasia treatment outcome. Each patient enrolled here will undergo a comprehensive cognitive/linguistic test battery, including measures of learning potential. Relying on the same SVR approach used in Project 3, we will construct a computational model that includes the factors that strongly predict treatment success. Then, we will make that model available on-line in a web based format so that clinicians can find out which tests to administer. They can then plug in their test results to predict the likelihood that a given patient will respond to speech and language treatment. Furthermore, at the end of Project 2 (PI: Hillis) we will also be able to explore whether a default model predicts response to speech and language therapy (SALT) with a phonological-semantic focus.

2.2 Relating Dual Stream Damage to Aphasia Treatment Outcome. Since its inception, the dual stream model of speech processing (1, 74, 75) has been highly influential in the field of cognitive neuroscience and neurolinguistics. Published in 2007 (1), the most comprehensive article that describes the model has now been cited over 1,400 times. As far as we can tell, the current project is the first to study whether damage to the cortical areas that support the dorsal and ventral streams predicts outcome in aphasia therapy. We suggest that this is not only a highly innovative approach but also has theoretical and practical implications for understanding the factors that drive successful treatment outcomes in patients with aphasia. As we explained earlier, by relating damage in the dorsal and ventral streams to treatments with semantic versus phonological focus, it will be possible to judge whether outcomes depend on relative intactness of each processing stream. The initially broad focus of the phonological and semantic treatment approaches is necessary to relate ventral and dorsal stream damage to treatment outcome. Later, we can focus on what specific aspects of treatment contribute to aphasia recovery.

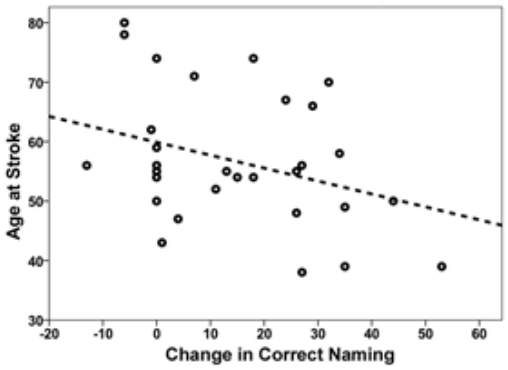
2.3 Measuring Whether Patients' Ability to Learn Predicts Treatment Success. Normal subjects' ability to learn new information has been studied extensively for many decades (11, 12, 118, 119). Whereas it is beyond the scope of the current project to summarize the findings, it is clear that the ability to learn varies quite significantly among the general population (11, 13, 14, 120-122). As recovery from aphasia requires a (broken) system to relearn information, it is plausible that learning ability, as measured at baseline, should be related to how well a given patient responds to treatment. Our approach to take learning potential into account is innovative, given that this is an area of study that has received very little attention in previous aphasia treatment research. By charting short-term improvement on two separate tasks – one that relies on language processing and one that does not – it will be possible to understand if learning potential as a predictor of aphasia treatment outcome is

domain general or domain specific. If Aim 3 yields positive results, it would indicate that learning tasks could be added as part of the baseline test battery to inform the default model.

3. Approach

3.1 Preliminary Studies. As we discussed earlier in section 1.3, one of our previous aphasia treatment studies (N=30 chronic patients with aphasia) related lesion and fMRI data to improvements in naming following 30 hours of anomia treatment (10, 85). Each patient completed two 15-hour treatment phases separated by one week where one treatment phase focused on a semantic cueing hierarchy and the other used a phonological cueing hierarchy (123). To provide support for Aim 1, we reanalyzed some of the data from this study to explore whether biographical factors (gender, chronological age, age at stroke, time post-stroke, and race) and cognitive/linguistic factors (Western Aphasia Battery subtests; WAIS Matrix Subtest; Pyramids and Palm Trees test [PPTT], Boston Naming Test [BNT]) predict treatment outcome. These factors would be a part of our "default model" of outcome prediction. As this was an exploratory analysis, we used multiple step-wise regression with a liberal p-value threshold for significant predictors (p<.10).

Figure 2. The relationship between age at stroke (y-axis) and treatment related changes in correct naming (x-axis).



Treatment outcome (change in correct naming before and after treatment) was predicted separately for the phonological treatment, the semantic treatment, and overall treatment (collapsing across the phonological and semantic treatments). Two statistically significant regression models predicted phonological treatment outcome, Model 1: Age at stroke (F=3.6, p=.06, R^2=.11); Model 2: Age at stroke + scores on the PPTT (F=3.9, p=.03, R^2=.22). Similarly, two models emerged as significant predictors of semantic treatment outcome, Model 1: BNT scores (F=16.2, p<.0001, R^2=.37); Model 2: BNT scores + Age at stroke (F=10.7, p<.0001, R^2=.44). Finally, two models predicted overall treatment outcome, Model 1: BNT scores (F=9.8, p<.0001, R^2=.26); Model 2: BNT scores + Age at stroke (F=11.5, p<.0001, R^2=.46). In relation to the current proposal, we find that these results are particularly important for three reasons: First, in a sample that only included 30 patients, one of our models (Model 2 for overall outcome: BNT scores + Age at stroke) accounted for almost half (46%) of all the variance in predicting which patients are likely to respond to treatment. Second, it appears that among the 30 patients who completed our study, the age when they had the stroke was a particularly important predictor of who was likely to respond to treatment (Figure 2). Finally, a combination of both biographical factors (Age at stroke) and cognitive/linguistic factors (PPTT and BNT) provided the best outcome prediction. Although the cognitive/linguistic test battery we used was not ideal for predicting outcome, we contend that these data provide strong support for Aim 1, which will rely on a more targeted cognitive/linguistic test battery and a far larger patient sample than our preliminary study.

To provide support for Aim 2, we further analyzed the dataset discussed above to test whether adding the WLG model (proportional damage to the arcuate fasciculus, Broca's area [*pars opercularis+pars triangularis+pars orbitalis*] and Wernicke's area [*STG*]) to the default model (based on the regression analysis above: Age at stroke; PPTT, & PNT) would improve outcome prediction. Similarly, we

combined the factors in the default model with the DS model (proportional damage to the dorsal stream [*pars triangularis+pars opercularis+posterior STG*] and ventral stream [*posterior MTG+posterior ITG+temporal pole*]) for outcome prediction. The results are summarized in figure 3. We compared the models using SVR (leave-one-out method) where the factors in each model were used to predict treatment related changes in correct naming. The output from this analysis is the correlation coefficient between the actual improvement scores and the predicted scores. For improvements in correct naming associated with semantic treatment, phonological treatment, or overall treatment (collapsing across phonological and semantic treatments), the D+DS model provided the strongest outcome prediction. In fact, the D+DS model accounted for approximately 30-40% better reduction in error (R^2) than the D+WLG model across the three treatment outcomes. Although our research only included 30 patients, the results strongly support our plan in Aim 2 to add behavioral and lesion factors to predict treatment success using SVR.

Figure 3. A comparison of the three models for predicting anomia treatment success. Color coding of the models corresponds to Figure 1.

Semantic

	r	R^2	*p*
Default	0.5	0.25	0.003
D+WLG	0.43	0.19	0.01
D+DS	0.52	0.27	0.002

Phonological

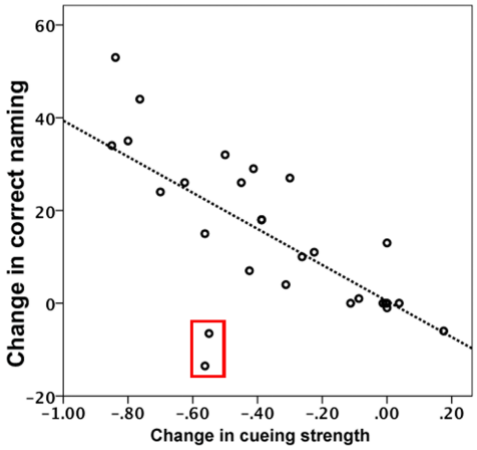
	r	R^2	p
Default	0.26	0.07	0.08
D+WLG	0.4	0.16	0.02
D+DS	0.47	0.22	0.006

Overall

	r	R^2	p
Default	0.58	0.34	0.0005
D+WLG	0.57	0.33	0.0006
D+DS	0.67	0.46	0.00003

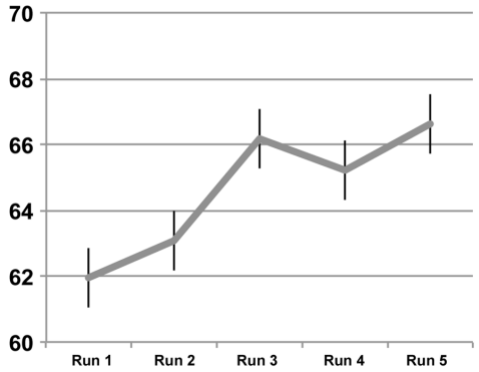
To provide proof of concept in support of Aim 3, we also analyzed data from the study described above. One way to estimate learning is to measure changes in contextual support (e.g. cueing) needed to evoke the correct response (124). The cueing hierarchies used in our study were ordered so that the weakest cues were presented first but proceeded to stronger cues if correct naming was not achieved. We recorded the level of cueing needed to elicit correct naming for each stimulus presentation across all treatment sessions. This allowed us to calculate change scores for the level of cueing needed to elicit correct naming during the beginning compared to the end of the first treatment session. Then, these change scores were used to predict overall improvement in correct naming following completion of all 30 hours of treatment (Figure 4).

Figure 4. Changes in the level of cueing in the initial phonological treatment session predict overall treatment outcome. Outliers are in red box.



The results revealed that changes in cueing strength during the initial phonological treatment session strongly predict overall treatment outcome, $F(1,29)=25.28$, $R^2=.50$, $p<.0001$. When two outliers were removed, these statistics changed to, $F(1,27)=89.34$, $R^2=.78$, $p<.0001$. Changes in cueing during the first semantic treatment session also predicted overall change in correct naming, $F(1,29)=21.3$, $R^2=.45$, $p<.0001$. When aphasia severity was included as a cofactor, both regression analyses remained statistically significant and the R^2 only showed a minimal decrease. These results suggest that patients whose level of cueing strength needed to elicit correct naming decreases during the initial treatment session are far more likely to respond to the overall treatment, regardless of initial aphasia severity. We recognize that decreased reliance on cues during naming may not be a classical learning task and that what we deem "learning" in this context could also be called "stimulability" or priming. Regardless, performance on the cueing hierarchy task by our 30 patients was a highly robust predictor of outcome and no other factor – behavioral or neuroimaging – came even close in terms of predictive power. From a theoretical perspective, it is a caveat that we do not know whether the changes in cueing level represent a general learning mechanism or something that is language specific. Therefore, in addition to measuring potential changes in cueing strength during a baseline session that uses cueing hierarchies to elicit picture naming, all patients will also complete a separate learning task that does not explicitly rely on language processing. As a result, it will be possible to determine if initial success in treatment is related to language independent learning or is language specific. For the purpose of control, we will use a hand-eye coordination task that is similar to the classical computer game "PONG." Available at USC and MUSC, the hardware used for the control task is a KINARM End-Point Lab with Dexterit-E software (BKIN Technologies, Kingston, Canada). Patients are seated in a modified wheelchair base designed for stroke survivors who grasp a handle with their left hand. The patients move their less-affected hand in the horizontal plane to interact with visual targets displayed in the same plane as their hand by a virtual reality system. The actual task is a two-minute 'object hit and avoid task' where the patients use a virtual paddle attached to left hand to hit away two kinds of visual targets (n=160) while trying to avoid other kinds of non-targets (N=80). The task is repeated five times and accuracy is collected for each run. Target shape, location and speed are randomly distributed, ensuring that every task repetition is distinct. Task difficulty increases slightly over time by slowly increasing the number of objects and their movement speed. To provide 'proof of concept' that aphasic patients show learning on this task, we had 11 patients complete this task five times each. Figure 5 shows that the overall accuracy (average number of target shapes hit and non-targets avoided) increased over the five runs.

Figure 5. Task accuracy (y-axis) by 11 aphasic patients on the control learning task over five runs (x-axis). Error bars denote standard error.



For Aim 3, our plan is to plot a learning curve over the five runs and use the slope of the curve to predict overall outcome in aphasia treatment. Along with many other groups, we have demonstrated that audio/visual (AV) speech perception modulates frontal areas typically implicated in motor speech output (125-127). Our studies were specifically conducted to inform aphasic impairment and to understand if AV speech processing has implications for aphasia treatment. In Fridriksson et al (2009; 125), we demonstrated that a training task that involves picture-word matching where the words are presented as AV stimuli (seeing just the mouth of a talker) yields greater improvement in naming compared to a treatment that relies on picture-word matching with audio-only stimuli. The treatment

task used in this study was fairly simple: Patients see a picture of an object on a computer screen (2 sec) followed by AV or audio-only presentation of a noun. The noun either matches the preceding picture or includes a phonemic, semantic, or an unrelated foil. The phonological relatedness of semantic and phonological foils to the target noun is parametrically varied. Patients must determine whether the picture and AV (or audio-only) stimulus match by pushing a green (correct match) or red (incorrect match) response button. Immediate feedback is provided in the form of a "smiley" face for correct responses and a "frowny" face for incorrect responses. We found that patients show greater improvement in picture naming following a training phase that includes AV stimuli compared to training that relies on audio-only stimuli. This method was used in two subsequent pilot studies using anodal tDCS to enhance aphasia treatment outcome (128-130). Moreover, we continue to rely on the AV version of this treatment approach in our current CATES trial and plan to use it in project 2 (PI: Hillis). Although there is limited consensus on what constitutes "typical" aphasia treatment, we suggest that our approach builds on some of the same treatment principles often incorporated in current aphasia treatment approaches. For example, it includes a parametric relationship between treatment targets and cues (92, 123) and response-contingent feedback (93, 131). Besides taking into account literature on AV speech processing, this approach incorporates both phonologically and semantically related distractors. Therefore, it relies on both phonological and semantic stimulation within the same treatment task. For the purpose of the current project, we simply refer to the approach as 'speech and language treatment' (SALT) for aphasia. We will examine whether the same factors that predict outcome in treatments that assume a phonological or semantic focus also predict outcome in the CATES trial.

3.2 Study Aims – Overall Strategy. Our plan is to treat 120 participants with chronic aphasia caused by damage to the left hemisphere (Appendix I: Study Protocol). Utilizing a cross-over design, each participant will receive 45 minutes of aphasia treatment per weekday (5x/week) focusing on treatment tasks that emphasize semantic or phonological stimulation (Table 1). Half of the patients will receive treatment in the following order: a. 15 treatment sessions utilizing phonological stimulation tasks; b. 4-week inter-treatment interval; c. 15 treatment sessions with semantic stimulation oriented tasks. The remaining half of patients will receive treatment administered in the opposite sequence. Patients will be randomized to receive treatment in either order. During the first baseline visit (W0), patients' biographical and testing information is entered into online modules in WebDCU™ (see Administrative Core). All subsequent patient data, including testing and treatment scores, MRI related information, and treatment fidelity information, will also be entered and stored using WebDCU™. To establish treatment baselines and for the purpose of in-depth characterization of the patient sample, each patient will undergo extensive testing during the week prior to treatment. Data from baseline testing will also be used for the purpose of projects 3 (PI: Rorden) and 4 (PI: Hickok). To assess treatment- related changes in speech and language processing, testing will occur during the weeks immediately before and after each treatment phase. Follow-up testing will occur 4-weeks and 6 months after completing the second treatment phase. To assess damage to the cortical areas included in the WLG and DS models, patients will undergo MRI (see Neuroimaging Core) at baseline during the week before treatment is initiated. Approximately two-thirds of the patients will be enrolled at the University of South Carolina. The remaining one- third will be enrolled at the Medical University of South Carolina under the supervision of Dr. Bonilha, a stroke neurologist with ample experience managing aphasia, and a long-time collaborator of the PI.

Table 1. Timeline for testing and treatment. G1 notes a group of patients who first undergo aphasia treatment utilizing tasks that target phonological stimulation. Then, during the second treatment phase, these patients will cross over to receive treatment focused on semantic stimulation. G2 (group 2) will undergo treatment using the opposite order of treatment phases. W=Weeks; Eval=Evaluation of performance (testing); Phon. Tx.=Phonological processing focused treatment; Sem. Tx.=Semantic processing focused treatment.

Group	W0	W1-3	W4-7	W8-10	W14*	W38#
G1		15 Phon. Tx.		15 Sem. Tx.		
G2		15 Sem. Tx.		15 Phon. Tx.		
G1 & G2	Eval x2	Eval before & after	Eval	Eval before & after	Eval x2	Eval x2
Scans	MRI	*W14 denotes 4-week follow-up testing #W38 denotes 6-month follow-up testing				

3.3 Sample Size. The data described in section 3.1 provide the basis for our power analysis. For Aim 1, the primary goal is to construct a model that predicts aphasia treatment success. We recognize that correlations based on small sample sizes can be influenced by outliers, and therefore using the full model results described in 3.1 could provide liberal estimates. Therefore, we used the factors described in 3.1 to compute exhaustive leave-one-out regression values, resulting in the more conservative R^2=.088, R^2=.27 and R^2=.36 for phonological, semantic, and overall treatment effects (rather than .22, .44 and .46). This suggests that a sample size of 120 individuals provides 0.95 power at $p < 0.05$, and 0.85 power at $p < 0.01$. Deriving power estimates for Aim 2 is more speculative: rather than comparing against a static null hypothesis (e.g., treatment has no influence) we are comparing against different dynamic models that will each benefit from a larger sample size (as we discuss in more detail in section 4.7 of project 3). Conventional analyses suggest that a sample size of 120 individuals will provide 0.92 power at $p < 0.05$ for supporting the weakest effect, shown in Figure 3 (default model's ability to predict phonological treatment benefits) and 0.72 power at $p < 0.05$ for distinguishing the difference between the default model and the D+DS model. We note that a null result for the latter in the presence of a robust result in the former could still be clinically meaningful, potentially suggesting that initial behavioral tests are sufficient for treatment planning (as we discuss in section 2 of project 3). For the purposes of projects 3 and 4, we will also test (but not treat) additional 30 patients with chronic left stroke who do not present with aphasia. Accordingly, the total number of patients enrolled here will be 120 for project 1 with additional 30 patients tested for projects 3 and 4 (total sample = 150 patients).

3.4 Patient Eligibility. Participants with chronic aphasia (>12 months from stroke) as a result of left hemisphere stroke (ischemic or hemorrhagic) will be included (specific inclusion/exclusion criteria are discussed in Protection of Human Subjects). Persons who incur more than one stroke will be excluded only if the subsequent stroke(s) includes the right hemisphere. The chance of recurrent stroke has been reported to be as high as nine times greater than having a first stroke and at least 30% of all stroke patients have had more than one stroke (132, 133). Thus, excluding patients with more than one stroke fails to reflect the general population of stroke patients thus limiting the generalization of the findings. However, projects 3 and 4 will rely only on data from patients with first event stroke, because these projects focus on making inferences about specific brain-behavior relationships rather than application to clinical practice. The number of strokes will be included as co-factors in the statistical analyses planned for each aim. Proportional damage to different cortical areas will be determined based on structural MRI scans.

Exclusion criteria are as follows: 1. Severely limited verbal output. As some of the treatment tasks rely on verbal output, it is pertinent to exclude patients who are essentially mute or whose speech only consists of stereotypies (judged as score of 0-1 on the Spontaneous Speech rating scale on the WAB); 2. Severely impaired auditory comprehension. Potential participants will be excluded if their Comprehension subscore on the WAB falls in the 0-1 range. Moreover, participants will be excluded if, as per clinical judgment, it is determined that they do not understand treatment task instructions; 3. Because damage to different cortical areas will be determined using MRI, patients with contraindications to MRI will be excluded.

3.5 Aphasia Treatment. Since our primary goal is to assess predictors of treatment that can be applicable to existing practices, we need to select treatments that are typically employed. Neither semantic nor phonological processing exists independently of each other in oral language use. Accordingly, the treatment approaches to be used here do not focus exclusively on either semantic or phonological processing; they each involve both, but place different emphasis on semantics versus phonological processing. In fact, even when semantic processing is the treatment emphasis, these treatments also affect phonological processing and vice versa (Appendix II: Scripts for treatment tasks). We stress that the main purpose here is not to assess whether one treatment focus is more potent than the other for improving language processing. Rather, we only include the two treatment foci to understand if factors that predict outcome with one treatment approach also predict success using the other treatment approach. In addition, we include two semantic and phonological treatment foci to explore whether damage to the dorsal and ventral streams differentially predicts response to phonological and semantic treatment. Project 4 (Hickok) will attempt to zero in on measurement of "computational" level deficits and assessment of the computational effects of treatment. As we discuss both semantically and phonologically focused treatments, this must be kept in mind.

Semantically focused treatment tasks: Three types of tasks will be featured. 1) Semantic feature analysis (SFA) has a relatively long history as a treatment approach for aphasia (134-136)\. For each pictured stimulus the patient is prompted to name the picture. Then, he or she is encouraged to produce semantically related words that represent features similar to the target word (e.g. superordinate category, use, action, physical properties, location, and association). For example, to elicit a location feature, a clinician might say "where do you typically find this object?" If the patient is not able to name the target item once each word feature has been produced, the clinician will say target word. Regardless of naming accuracy on the last item, treatment continues on to the next stimulus item. Because both nouns and verbs (137, 138) have been used for SFA- focused activities, stimuli for SFA tasks here will utilize both. 2) The second semantic treatment approach is the semantic barrier task. This approach includes features of the Promoting Aphasics' Communication Effectiveness (PACE) approach (139-142) and has also been included as part of constraint-induced language therapy(143). It relies on a stack of picturable stimuli, which are split between the patient and clinician and placed face up on a table. A visual barrier is placed between the clinician and the patient so they cannot see each other's pictures. The goal of the task is for one participant (e.g., patient) to describe each card so that the other participant (e.g., clinician) can guess the picture on the card. Participants are only allowed to describe the semantic features of the target and the clinician models the kinds of cues that are allowed. The clinician and patient take turns describing pictures. 3) The third approach, Verb network strengthening treatment (VNeST), is a semantic treatment approach that targets lexical retrieval of verbs and their thematic nouns (144, 145). The objective of VNeST is for the patient to generate verb-noun associates with the purpose of strengthening the connections between the verb and its thematic roles. VNeST can be modified to fit patients with very limited speech output (e.g., using sentence completion).

Phonologically focused treatment tasks: Three approaches, also with sound research pedigrees, will the used. 1) The first is the phonological components analysis task (96), which was modeled after semantic feature analysis. It requires the patient to first name a given picture and then to identify the phonological features of the

target words (e.g., first sound, syllables, last sound, association, and rhyme). Once the features have been identified, patients are required to attempt to name the picture again. Then, the treatment moves on to the next item in a stack of targeted imageable nouns. 2) Phonomotor treatment is a phonologic sequence-based therapy that includes two stages of training that targets speech perception and speech production (146-150). In Stage 1, the clinician uses multi-modality stimulation to practice perception and production of speech sounds in isolation. Stage 2 mirrors stage 1 but focuses on speech sounds in a variety of consonant-vowel sequences.

3) The phonological judgment task relies on computerized presentation of verbs and nouns where patients are required to judge whether pairs of words include similar phonological features (142, 151, 152). The task is comprised of five conditions that entail determining if a set of words: a. Includes the same number of syllables; b. Includes the same initial syllable; c. Includes the same final syllable; d. Which word has more syllables; and e. rhyme. Patients respond to each condition by pressing one of two response buttons depending on the task requirements and instructions.

3.6 Biographical Factors and Cognitive/linguistic Assessments. *Biographical factors:* For the purpose of recording data on biographical factors, each patient (and/or caregiver) will be required to fill out a case history form and a questionnaire on previous medical events, years of education, etc. *Cognitive/linguistic testing:* The following tests and tasks will be utilized to assess cognitive/linguistic status at baseline, before any treatment starts: 1. The revised version of the WAB will be used to assess overall aphasia severity and type (153). Although the WAB does not provide in-depth analysis of speech and language impairment, it is a test that is commonly used in clinical practice as well as a reasonable measure of overall aphasia severity; 2. The Apraxia of Speech Rating Scale (ASRS; 154) will be utilized to rate the presence and severity of apraxia of speech AOS (note: the ASRS constitutes a major component of project 3); 3. To assess grammatical processing (agrammatism), we will rely on the Northwestern Assessment of Verbs and Sentences (155). The NAVS was designed for patients with aphasia and allows for detailed examination of verb processing (e.g. verb naming) as well as production and comprehension of canonical and non-canonical sentences; 4. The Pyramids and Palm Trees Test (PPTT; 156) and the Kissing and Dancing Test (KDT;(157) will be used to assess amodal semantic processing of nouns and verbs, respectively; 5. To assess phonological processing, several sub-tests from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; 158) will be utilized (a. Auditory discrimination of nonword minimal pairs; b. Auditory discrimination of word minimal pairs; c. Rhyme Judgments for pictures and words; d. Phonological Segmentation of initial and final sounds); 6. To assess speech repetition ability we will rely on the Philadelphia Repetition Test (PRT), a low-imageability word repetition test (159), and the non-word speech repetition sub-test from the PALPA (158); 7. For analysis of cognitive status, the Matrix test on the Wechsler Adult Intelligence Scale (WAIS; 160)\ will be administered; 8. To assess verbal short-term memory, we will use tasks described by Martin et al. (161). Their tasks allow for detailed evaluation of lexical-semantic performance in relation of verbal short-term memory load. All patients will be screened for visual field cut and visual neglect using a short in-house test developed by Dr. Rorden. To minimize the influence of test fatigue, the assessment battery will be spread over two sessions administered on two separate days during the week before the first treatment phase starts. The motivation for the extensive behavioral test battery included here is to provide a comprehensive set of factors used to predict treatment outcome. In addition, the behavioral data collected here will also feed into projects 3 and 4 (e.g. non-word repetition and PRT). For details on test reliability, validity, and clinician training, see Clinical Core.

3.7 Learning tasks for Aim 3. All participants will complete two learning tasks before initiation of treatment: *Language learning* – We will rely on the same cueing hierarchy naming task used in Fridriksson (2010; 10) and Fridriksson et al (2012; 85). This task will be completed in a 1-hour session administered before the initiation of the initial treatment phase. Semantic and phonological cues will be randomized to apply to individual items and the measure of learning will be qualified as change in the level of cueing at the beginning of the session compared to the end of the session. Thus, it will be possible to separately calculate changes in naming that rely on semantic cues versus phonological cues. We have already demonstrated that improved performance on this task within a single session is a very strong predictor of anomia treatment success (see pilot data in Figure 4). *Hit-and-avoid task* – Each patient will complete the hit-and-avoid task described earlier (see pilot data in Figure 5). Mean accuracy for each of five runs will be plotted on patient-by-patient basis and the slope of the learning curve will be utilized as a predictor of aphasia treatment outcome. The total administration time of the hit-and-avoid task is approximately 15 minutes.

3.8 Treatment Outcome Measures. Consistent with the CATES trial and project 2, change in correct naming

on the Philadelphia Naming Test (PNT; 162) will be assessed as an outcome measure. Confrontation naming is included on most comprehensive aphasia tests and is commonly assessed in aphasia treatment research, making it possible to compare the findings in the current project to other studies. The PNT will be administered twice during the week before and twice during the week after each 3-week treatment phases (scores averaged to reduce variability). For follow-up testing, the PNT will be administered again at four weeks and six months after completion of the second treatment phase. The primary outcome measure, however, is specifically defined as change in correct naming immediately before and after each treatment phase. The PNT is a computer-based assessment of naming for persons with aphasia and includes 175 pictures representing mid- and high-frequency nouns from a word frequency list compiled by Francis and Kucera (163). Participants are videotaped during assessment and responses are scored offline. The clinicians who score the PNT will be blinded to when each assessment video was recorded, either before or after each of the two treatment phases. In addition to assessing changes in correct naming, potential treatment-related changes in naming errors will also be scrutinized. For a more detailed description of PNT scoring and training of clinicians, see Clinical Core.

To evaluate the effects of treatment on discourse abilities, patients will complete a discourse protocol included in AphasiaBank, an archival database of discourse samples funded by the NIDCD since 2007. Three discourse tasks will be administered before and after each treatment phase: 1. Broken Window picture sequence description (164); 2. Cinderella story telling (165); and 3. A procedural description of how to make a peanut butter and jelly sandwich (first introduced by Davis, 1901 (166), as referenced in Lau, 2013 (167)). Transcription will be accomplished through the Clinical Core using transcription (CHAT) and the automated coding analysis (CLAN) systems available through AphasiaBank. We have already used this setup to assess outcome in aphasia treatment (38). As in the CATES trial, the analyses used for discourse samples will focus on lexical and syntactic aspects of spoken discourse: 1. Content words/min (excludes repetitions and nonwords); 2. Propositional density (168, 169); 3. Verbs/utterance. Each of these three measures is already included in AphasiaBank and it is a particularly nice feature that our data can be compared against the 440 aphasic patients included in AphasiaBank. This feature is especially important for determining variance across participants. In addition, we will assess potential changes in patients' ability to generate main concepts (170). Finally, to assess the effects of treatment on functional communication ability, all participants will be administered the Aphasia Communication Outcome Measure (ACOM; 171). The ACOM consists of 59 items that probe different aspects of real life communication situations and can be administered in 15-30 minutes. Whereas the PNT and discourse measures will be administered during the weeks immediately before and after each treatment phase, the ACOM will only be administered at baseline before both treatment phases and again at four weeks after completion of each treatment phase. By only administering the ACOM four weeks after treatment completion it may allow participants more time to realize potential treatment effects on functional communication ability.

3.9 Neuroimaging. The effects of stroke on cortical areas that are described in the DS model (proportional damage to the dorsal stream [*pars triangularis+pars opercularis*+lateral precentral gyrus+area Spt] and ventral stream [posterior MTG+posterior inferior MTG+temporal pole]) and WLG model (Broca's area [*pars opercularis+pars triangularis+pars orbitalis*], arcuate fasciculus, and Wernicke's area [STS+STG]) will be qualified as proportional damage to each cortical region. Excluding area Spt (Sylvian fissure at the parietal- temporal boundary), the regions of interest (ROIs) for each cortical region will be selected using the Johns Hopkins University standardized brain atlas (172). We note that there is no standard way to determine what ROIs should be included in either model. The ROIs for the DS model were selected in consultation with Dr. Hickok, including an ROI for area Spt that was derived from a previous fMRI study (173). The ROIs for the WLG model were selected based on input from Drs. Hillis and Bonilha, both of whom have extensive history with lesion-symptom mapping research in aphasia. Note that in comparison to the current project, Project 3 will accomplish much more detailed analyses of our neuroimaging data in relation to treatment outcome.

3.10 Data Analyses *Aim 1* – To construct the default model, both biographical and cognitive/linguistic factors are entered into a step-wise regression analysis using a leave-one-out method. This approach narrows down the number of factors that provide independent prediction of treatment outcome. The leave-one-out method controls for selection bias where the same subjects are not used to select the factors and test the model. The factors that survive in the final step-wise regression model are then entered into the support vector regression (SVR) analysis to determine how well the default model predicts actual outcome in patients. SVR is a more robust alternative to multiple linear regression. Both methods use a linear model to characterize the relationship between the predictors and the target variable (e.g, change in correct naming); that is, the target variable is modeled as a weighted sum of predictors, plus the offset term. The difference between the two methods is in the estimation of the model parameters (the weights and the offset). In multiple linear regression, it is done by computing the Moore-Penrose pseudo inverse of the predictor matrix; this minimizes the prediction errors in the least-squares sense, but leads to highly unstable estimation if there is correlation between the predictors (174). Also, the pseudo inverse cannot be computed if the number of predictors is greater than the number of observations (175). In support vector regression, we do not attempt to minimize the prediction errors; rather, we allow for a certain amount of errors, and set an upper bound on permissible prediction errors (176). In this formulation, the stability of estimation is not affected by correlations between predictors, and by the number of predictors relative to number of observations. The superiority of support vector regression over linear regression models has been demonstrated by Drucker et al. (177). We expect that the final default model will include a combination of several factors that together contribute to outcome prediction. As not every factor will contribute equally to the proportional reduction in error, *beta weights* (converted to Z-scores) will be calculated so that clinicians can better understand the relationship between predictors. *Potential concerns:* Relying on a relatively large sample of patients, it is possible that many biographical and cognitive/linguistic factors may correlate with treatment success. This is potentially problematic for clinicians who may not have time to administer the test battery necessary to assess each factor. Therefore, we will calculate several separate models that include sub-sets of predictors. Then, it will be possible for clinicians to estimate success based on a more restricted version of the default model.

Aim 2 – To compare the default model to the D+WLG and D+DS models, we will also rely on SVR. Here, all factors that comprise each model (e.g., D+DS model includes the default model factors in addition to proportional damage to the dorsal and ventral streams) are entered into an SVR analysis and the R^2 for the correlation between the actual scores (e.g., change in correct naming or propositional density) and the predicted scores is calculated. Then, a statistical comparison is made between the R^2 yielded by the different prediction models. The model with statistically significantly highest R^2 compared to other models will be determined as the best prediction model. Separate analyses will be conducted for each treatment focus (semantic vs. phonological) and overall treatment outcome. *Potential concerns:* Although we have been highly successful in attracting patients to our treatment studies, we have never carried out a project that requires the sample size proposed here. If actual patient recruitment lags far behind planned recruitment at the University of South Carolina and the Medical University of South Carolina, we will initiate enrollment of chronic patients at Johns Hopkins University. In such a case, .5 SLP appointment will be moved from South Carolina to Johns Hopkins University. It is also a concern that patient recruitment for the current project (C-STAR) and the CATES trial will overlap by approximately 4-6 months. During this time, we will randomize patients to be enrolled in either C-STAR or the CATES trial. We emphasize that we have always had a waiting list for patient enrollment in the CATES trial and at the time when the current proposal was submitted, 52/74 patients had been enrolled in the CATES trial (nine more patients are scheduled to complete enrolment in summer 2015).

Aim 3 – Separate SVR analyses will be conducted to test whether change in cueing in a single naming session or performance on the hit-and-avoid task predicts aphasia treatment outcome. The learning curve (slope of the regression line) will be calculated for each patient. The same SVR analyses planned for Aim 1 will be run again but now with the learning factors also included to assess whether the default model can be improved by including assessments of learning potential. *Potential concerns:* Without collecting further data, it is impossible to estimate the necessary length of the learning paradigms. This

is a potential concern as administering an hour-long learning task to patients in clinical care is probably not feasible. Thus, we will calculate learning curves for progressively longer increments (e.g. for the first 10, 20, 30, or 40 minutes of the task) to appreciate the necessary length of each learning task. For example, it could be the case that improvement in performance over the first 20 minutes of the language-learning task is sufficient to predict treatment outcome.

3.11 Making the Default Model Available On-line. Once the default model has been constructed, we will make the algorithm available to clinicians who can then assess their own patients and plug in the results to estimate treatment response relying on either semantically or phonologically focused treatment. For this purpose, we will design a website that includes detailed instructions regarding what factors need to be assessed and the strengths and limitations of this approach. Similarly, we will provide guidance regarding assessment fidelity, threats to assessment reliability and validity, as well as disclaimers about potential abuse of the model (e.g., denying patients treatment based on model prediction of poor outcome).

Appendix E. Clinical Core Research Plan

1. Significance.

1.1 Background. Speech-language therapy to treat aphasia is supported by evidence - individuals with aphasia who participate in speech-language treatment demonstrate greater improvements than those who do not.[1-4] Nonetheless, individual response to treatment varies greatly and little is known about how factors related to assessment, treatment, and patient characteristics interact to engender a positive treatment outcome. Different assessment foci can contribute to the mixed results.[5-8] For example, if discrete language abilities are treated and assessed, improvement in those language abilities (relative to a control condition) may be considered to support treatment efficacy. However, an alternative view is that treatment is only efficacious if there is a demonstration of improvement in communication activities or life participation that is clearly associated with the treatment, and not the control, condition. Different assessment administration procedures, particularly those involving the patient response and resultant score (e.g., item time limits, multidimensional scoring, self-corrected responses counted as accurate, etc.), can also lead to a patient being labeled as a "responder" with one set of procedures and as a "nonresponder" with another. In addition, treatment factors (e.g., treatment elements, dose, etc.) are oft studied, but despite emphasis, results are not straightforward. While the "more treatment is better" view is generally supported,[1,3,9-11] which treatment elements and targets should be practiced more,[1,3,12] and which outcome domains (e.g., body functions, activity limitations, participation limitations) benefit from intensive treatment schedules[6], is unclear. Finally, numerous patient factors, including stroke-related (lesion site and size, type of stroke, etc.), demographic, and neurophysiological variables have been the focus of many investigations, often with conflicting results.[3,13-17]

This P50 grant application systematically investigates many of these factors. Some elements (e.g., patient variability) may also be thoroughly discussed in "Limitations" and "Future Directions" sections of numerous journal articles, and solutions to preserve the validity of the study may be readily located in research design references. However, in the speech-language or stroke rehabilitation literature, relatively little attention has been paid to other threats to a study's validity that relate to correct implementation of treatment, and also assessment, procedures. Statistical conclusion validity, referring to inferences about the presence of a relationship between two variables and the strength of that relationship, can be threatened by measurement error, unreliable treatment implementation, and other sources of variance introduced into the experimental setting, as well as low statistical power (which can also be a result of the random variance introduced by the forenamed threats).[18] These threats can increase the chance of Type I or Type II error, or the additional error ("Type III") of concluding significance or nonsignificance when in fact the tests or the treatments were not correctly implemented.[19-21] Internal validity, referring to whether or not causation can be inferred from the statistical conclusions, can be threatened by those same threats to statistical conclusion validity, patient selection and attrition, and various assessment factors.[18] (Note: All threats to statistical conclusion and internal validity are not listed here, only those most critical to implementation of clinical procedures, which will fall under the aegis of the Clinical Core.) In other words, an additional contributor to the historically mixed results in aphasia treatment research emphasized by **Project 1**, **Project 2**, and here in the Clinical Core could also be those variables related to fidelity, or adherence to the study procedures. In the absence of fidelity monitoring, investigators are unable to confidently determine whether or not results (significant or nonsignificant) were caused by the targeted independent variable or were due to other random factors introduced because the clinician "drifted" from the protocol, or "contaminated" the protocol by adding or omitting elements.[19]

1.2 Treatment fidelity. Treatment fidelity, or how well the essential treatment elements were delivered as intended and were distinguishable from comparison conditions,[19,20,22] has so far been the focus of most implementation discussions. In general psychology research, several meta-analyses demonstrate that studies taking steps to ensure treatment fidelity have larger effect sizes (two to three times higher) for treatment outcomes compared to those studies that do not.[23] School psychology researchers have revealed that as the degree of treatment fidelity increases, the rates of positive outcomes also

increase.[24] In addition, treatment fidelity (or "program integrity") affects a number of variables in abuse prevention research.[25,26] Thus, implementation can be a determinant of success. Hinckley and Douglas[20] recently examined the reporting of treatment fidelity in aphasia treatment studies (reviewing 149 aphasia treatment articles published between 2002-2011), revealing that while almost half of the studies reported treatment methods that would enable replication, only 21/149 (14%) of the studies specifically described methods for establishing or monitoring treatment fidelity. Twenty of those studies utilized only a *single* treatment fidelity method - monitoring of adherence to treatment protocol, supervising treatment sessions, utilizing a training manual, *or* pre-implementation role-playing. Only one study utilized more than one treatment fidelity method, combining the use of a training manual with monitoring of adherence to the treatment protocol.

The NIH has recently established a treatment fidelity workgroup within the Behavior Change Consortium (BCC) in recognition of poor monitoring of treatment fidelity in behavioral research. This workgroup is tasked with defining treatment fidelity and offering guidelines for researchers. According to the BCC workgroup, establishing treatment fidelity should address the following five components[19]: (1) study design – the study should be designed appropriately so that hypotheses can be tested and inferences are valid; (2) training – training procedures should be standardized across clinicians; (3) treatment delivery – treatment should be monitored to ensure it is delivered as intended; (4) treatment receipt – the patient should understand the treatment procedures and demonstrate utilization of them within experimental sessions; and (5) treatment enactment – the patient should utilize behaviors targeted in treatment in real-world settings. The intricacies and the defense of study design for **Projects 1 through 4** have been laid out in their corresponding research plans. The training, treatment delivery, and treatment receipt components of treatment fidelity will be addressed in the Approach section of this Clinical Core proposal. Consistent with Gearing et al.,[22] the Clinical Core will not address treatment enactment since it is more about generalization of behavior and less about adherence to clinical procedures.

1.3 Assessment fidelity. Compared to treatment fidelity, similar guidelines to ensure adherence to an assessment protocol have not been established. The general impression is that selection of tests with good reliability and validity, and perhaps performing rater reliability checks within the study, is enough to ensure assessment fidelity. However, just as there can be clinician drift from treatment procedures (e.g., cueing hierarchy not delivered as designed) and contamination of treatment procedures (e.g., clinician is incorporating cues from Treatment B into Treatment A), assessor and/or rater drift and contamination are just as likely to occur, especially when administering lengthy assessment batteries that include tests with different administration procedures (e.g., item time limits, assessor cueing) and complex scoring systems (e.g., discourse, Philadelphia Naming Test), and are given repeatedly over a long period of time. Borrowing from treatment fidelity guidelines, assessment fidelity components of training and assessment delivery will be addressed in the Approach section of this Clinical Core proposal.

1.4 Role of the Clinical Core. Efforts to optimize assessment and treatment fidelity are critical to the projects proposed by Fridriksson, Hillis, Rorden, and Hickok. As displayed in Table 1, 12 of 13 project aims require recruitment and retention of patients and collection of high quality data that can be entered into the WebDCU™ (Clinical Core Aim 1). Those same 12 project aims rely on data collected by an assessor using prescribed administration procedures and scored by a rater adhering to prescribed scoring procedures (Clinical Core Aim 2). Seven of 12 project aims rely on treatment (Clinical Core Aim 3).

Table 1. Contributions of the Clinical Core to Clinical Research Center Project Aims

Clinical Core	Project 1 Aims			Project 2 Aims			Project 3 Aims				Project 4 Aims		
	1	2	3	1	2	3	1	2	3	4	1	2	3
Aim 1	*	*	*	*	*	*	*	*		*	*	*	*
Aim 2	*	*	*	*	*	*	*	*		*	*	*	*
Aim 3	*	*	*	*	*	*		*					*

The assumption of the investigators is that assessors, raters, and clinicians adhere to well-defined study protocols. The Clinical Core will exist to make that a reality. By doing so, the Clinical Core will guard against threats to statistical conclusion validity and internal validity, increasing the power to detect effects and increasing investigator and consumer confidence in the results. By reducing the chances of making Type I, II, or III errors, we would decrease the probability that this ambitious endeavor (C-STAR) would result in research waste, the costs of which can only begin to be fathomed if one recognizes that the impact of those errors would not only be felt in this study, but throughout the entire research community.[19,27-29] This responsibility is not taken lightly and is why the Clinical Core is essential to the research activities of C-STAR.

2. Innovation.

The innovative projects proposed in this application are dependent upon clinical activities. For the correct inferences to be made, adherence to assessment and treatment protocols must be continually monitored. It is important that the Clinical Core also apply the most up-to-date and innovative procedures for ensuring collection of high quality data. The following methods will be implemented, all of which appear to be novel among aphasia treatment studies.

2.1 Centralized training of all personnel in contact with patients. The Clinical Core will be responsible for training all personnel who will have contact with patients and/or friends and family members to ensure that all exchanges are ethical, respectful, and employ successful communication strategies. For Aim 1, this will include training for communication during initial recruitment and patient identification, for the informed consent process, and for participant retention. For Aim 2, this will include training of assessment administration procedures, and for Aim 3, treatment procedures. This training will help to reduce variance introduced by clinicians communicating or carrying out procedures with different styles than others. As far as we can tell, training of personnel in this manner has not been reported so far in the aphasia treatment literature.

2.2 Monitoring participant recruitment and retention. To increase the efficiency of recruitment and to ensure resources are not wasted due to patient withdrawal from the project, the Clinical Core will monitor the rate of recruitment with a recently developed “recruitment index”. The Clinical Core has an adaptive plan for recruitment in place so that advertising and recruitment efforts will be modified if the recruitment index shows **Projects 1 and 2** are not on target for meeting proposed recruitment numbers. The recruitment index has not been used in the speech-language literature, and participant retention plans have not been reported upon in previous research in this area.

2.3 Optimizing treatment fidelity. No aphasia treatment studies to date have reported more than two tools or methods used to monitor or attempt to establish treatment fidelity. The aphasiology community has been slow to respond to the treatment fidelity guidelines set forth by the BCC workgroup. The Clinical Core will optimize treatment fidelity by monitoring multiple components, specifically clinician training, treatment delivery, and treatment receipt. This is an ongoing process, not only occurring at the onset of the project, but rather continually performed over the entire duration of the project to ensure

that treatment implementation on the first day of the first year of the project is the same as on the last day of the last year of the project.

2.4 Optimizing assessment fidelity. In the aphasia literature, relatively little attention has been paid to the training of assessment personnel, other than ensuring a minimum training standard (e.g., clinician with at least 3 years of experience, etc.) or presenting intra- and inter-rater reliability information. Given the importance of assessment to nearly all of the aims proposed in all projects, it is important that we take additional steps to ensure that data will be collected with appropriate administration procedures and will also be scored appropriately. This will entail training of both assessors, who will also score baseline tests, and blinded raters, who will score repeated measures. Borrowing from treatment fidelity guidelines, the Clinical Core will coordinate assessor and rater training, assessment delivery, and reliability of raters.

3. Approach

3.1 Capability of the Clinical Core team. Dr. Argye Hillis is a Professor of Neurology, Physical Medicine & Rehabilitation, and Cognitive Science, as well as Director of the Cerebrovascular Division of the Department of Neurology at Johns Hopkins Medical Institutions (JHMI). She has 10 years of experience as a speech- language pathologist and Director of Neurological Rehabilitation, prior to becoming a stroke neurologist. She directs a large, multidisciplinary Stroke Prevention and Recovery Center (SPARC), which has been certified by the American Heart Association as a Comprehensive Stroke Center. Hillis is experienced in organizing national and international multidisciplinary groups; for example, she was recently President of the World Federation of Neurology Research Group on Aphasia and Cognitive Disorders and is on the Board of Directors of the American Neurological Association. As PI of the Clinical Core, she will be responsible for overseeing all aspects of the core. She will work closely with the Clinical Core Coordinator (located at USC – *to be hired*) to carry out the plans described herein, and that the goals of the core are being met. To ensure that the ideal staff members are hired and trained appropriately, Dr. Jessica Richardson, Assistant Professor of Speech and Hearing Sciences, University of New Mexico, will serve as consultant and provide initial training to the Clinical Core Coordinator, speech-language pathologist (SLP) Clinical Core Project 1 Leader, and raters. Dr. Richardson is a certified SLP, with 13 years of experience providing (and supervising the provision of) diagnostic and therapeutic services to aphasic adults. Richardson is uniquely suited for this role given her previous contributions to some of the studies (published and pilot) that have motivated this P50 application, with roles ranging from treatment designer to assessor to clinician to trainer, including both behavioral and brain stimulation methods. Dr. Leonardo Bonilha is a neurologist and neurophysiologist at the Medical University of South Carolina (MUSC), where he is actively involved in the care of patients with stroke within the large MUSC stroke center. Besides his clinical expertise, Dr. Bonilha is a researcher with experience in behavioral and neuroimaging assessments of aphasia, being the leading neurologist at MUSC in the NIDCD supported clinical trial “Transcranial Direct Current Stimulation (tDCS) to Treat Aphasia: Phase II Trial” (CATES trial; DC011739, PI: Fridriksson) and the PI on the NIDCD supported project “Brain Connectivity Supporting Language Recovery in Aphasia” (NIH R01DC014021).

Aim 1: To supervise data collection and data management at each clinical site.

3.2 Rationale. Perhaps the most important aspect of data collection is ensuring that there are participants from whom to collect data. Our discussion of data collection and management thus begins with participant recruitment and is followed by plans for participant retention. The Clinical Core’s recruitment plans focus on pre-enrollment activities while retention plans include post-enrollment activities designed to promote study completion.[30] These plans will guard against selection bias and attrition threats to internal validity.[18] Next, plans for ensuring collection of high quality data are presented, as is information about data storage and security. Other processes involved in promoting data quality that are dependent on assessment or treatment implementation will be discussed in fidelity sections.

3.3 Baseline training. The Clinical Core will ensure that all study personnel having any contact with participants and/or the friends and family members of the participants will have a basic understanding of the communication needs and obstacles for this population, and also some understanding of their experience. While most study personnel will have had a background with adult neurogenic communication disorders, standardizing this baseline training for all personnel, from those who might only be escorting them to the elevator to those who will be spending hours with them in treatment sessions, will ensure that no erroneous assumptions about qualifications are made. This baseline training (see Appendix 1) comprises many components that are already implemented by Richardson's lab, with positive feedback and results. The Clinical Core will facilitate the training and will maintain compliance logs.

First, all study personnel will complete human subjects training (CITI) to understand ethical treatment of human research participants, which is essential training even in the earliest stages of a study (e.g., to avoid coercion and/or bias during advertising and recruitment efforts). Second, all personnel will independently read two articles by the Hilari research group discussing the impact of aphasia: 1) *"Aphasia blog talk: How does stroke and aphasia affect a person's social relationships?"* (Fotiadou, Northcott, Chatzidaki, & Hilari, 2014); and 2) *"Aphasia blog talk: How does stroke and aphasia affect the carer and their relationship with the person with aphasia."* (Winkler, Bedford, Northcott, & Hilari, 2014). Third, all personnel will independently view four videos available via AphasiaBank (http://talkbank.org/AphasiaBank/; participants consented to be videotaped by AphasiaBank) during which persons with aphasia discuss their stroke story. Two videotaped persons will have nonfluent aphasia (1 high, 1 low severity) and two will have fluent aphasia (1 high, 1 low severity). Finally, the Clinical Core will lead small group sessions that will include a brief lecture on *"How to communicate with individuals with aphasia",* a lecture used by Fridriksson when training speech-language pathology students to be aphasia group leaders. This lecture will be followed by an experiential learning session utilizing the program *"Experiencing Aphasia: An Inservice Model Demonstrating Aspects of Aphasia."*[31] The stated goals of the program are: (1) To give participants an opportunity to experience what it is like to be communicatively impaired; (2) To encourage participants to discuss their emotional responses to being treated as an impaired individual; and (3) To instruct participants about the different modalities that may be impaired as a result of stroke or head injury. This session will provide opportunities for personnel to participate in simulations of communicating with aphasia through receptive language, expressive language, reading, and writing stations. Richardson has used this for training her lab members and social workers (with Adult Protective Services).

3.4 Participant recruitment.

Advertising and identification of eligible participants – For **Project 1**, the Clinical Core team will focus advertising efforts on approaches that have yielded participants for our previous studies and are proven to be among the most effective strategies for recruiting elderly patients,[32-34] which forms the bulk of the target participant base (Table 2). We will depend on local and state news broadcasts and newspaper advertisements. If the yield is limited, radio advertisements and group presentations to stroke and aphasia support groups will be added into the advertising plan.[32,33] At the same time, the usual lab recruitment activities will continue, and include networking with colleagues, advertising studies through statewide directories (http://screserach.org/) and a lab website (http://www.mccauslandcenter.sc.edu/aphasia/), and referrals from well-established area contacts (community, in-state, out-of-state). For **Project 2**, these same advertising efforts are not required, as this project recruits participants with acute stroke directly from the inpatient stroke wards at JHU and Johns Hopkins Bayview Medical Center, and also the follow-up clinic, the Stroke Prevention and Recovery Center (SPARC) at JHMI. The stroke neurologist who is attending on the wards or in SPARC is approached by a study team member on a daily basis, and identifies potential participants who have agreed to be approached by the study team member about the research.

Table 2. Advertising outlets for recruitment for Project 1

Television News		Newspaper		Baseline Laboratory Advertising			Backup Advertising	
Columbia	Charleston	South Carolina State	North Carolina State	Networking	Referrals	Website	Presentations	Radio

As **Projects 1 and 2** address relevant health issues, we expect interest and willingness to participate from potential participants with aphasia once they are made aware of the projects.[34] Following successful advertising efforts, identification of eligible potential participants by screening for inclusion and exclusion criteria is completed by interview (phone or face-to-face) and/or case history review for **Project 1**, and by record review and interview for **Project 2**. Given the complexity of recruitment for these types of studies and in this population, it is essential that potential participants have a positive and informative interaction with the initial contact person.[34] The Clinical Core will work with project investigators to coordinate the hiring of study personnel, screening for excellent interpersonal skills and a background communicating with persons with communication impairment (speech-language pathologists, clinical coordinators with experience with this clinical population, speech-language pathology graduate students, etc.). The Clinical Core will ensure through training that anyone performing recruitment duties is well-informed about the study details (i.e., expectations, timelines, etc.), because individuals are likely to be more responsive if they are confident in the knowledge of the person they are speaking with.[34] Training will include independent reading of the full study protocol, the informed consent document, and any other related documents, followed by a review of the documents with a member of the Clinical Core team (Appendix 1). In addition, we will provide scripts to assist with communicating about the study via e-mail, telephone, or in-person. As discussed in **Project 1** (Research Plan, 4.4) and **Project 2** (Research Plan 4.4.9), we do not anticipate difficulty recruiting participants, but these efforts will only serve to enhance the already successful recruitment strategies employed by both project leaders (Fridriksson, Hillis) and the other Clinical Core team members (*to be hired* Clinical Core Coordinator, Bonilha) in their laboratories.

Informed Consent – For both projects, study explanation and informed consent can be a challenging process. The potential participants not only have limitations in producing and/or understanding language, but are also being approached after what is often the most devastating event so far in their lives. For **Project 2** in particular, the reality of this event has had little time to be fully realized. For these reasons, the person obtaining consent must be a skilled and supportive communicator who is aware of the realities faced by this clinical population and respectful of their decisions and requests for information, more time, etc. The person obtaining consent should also be skilled enough to recognize when the aphasic participant has a comprehension impairment (e.g., Wernicke's aphasia or Global Aphasia) so that they seek informed consent from the participant's spouse, adult child, legal representative, or identified decision maker, with accompanying assent from the aphasic individual. The person obtaining consent will be a Clinical Core leader (Hillis, Bonilha, *to be hired* Clinical Core Coordinator at USC) or another speech-language pathologist with informed consent training.[35,36] The required human subjects training will have already been completed as part of baseline training. Additionally, the Clinical Core will facilitate the following informed consent training procedures (Appendix 1) for both **Project 1 and Project 2:** (1) independent reading of informed consent articles *("Informed consent and aphasia: Evidence of pitfalls in the process"* [37] and *"Discrepancy between participants' understanding and desire to know in informed consent: Are they informed about what they really want to know?"* [38] with an additional reading for **Project 2** members (*"Informed consent: The rate-limiting step in acute stroke trials"* [39]); (2) observation of a minimum of three expert (Hillis, Richardson, or Bonilha) consent sessions; (3) supervised consent sessions with expert feedback; and (4) yearly review of consent information plus expert supervision of 1-2 sessions. This training will ensure that the informed consent process is informative, respectful, and ethical, and will positively influence retention[40,41].

Research suggests that comprehension of written materials is significantly improved in people with aphasia when the following modifications are made: language simplification, increased font size, increased spacing, increased white space, and use of pictures.[42,43] The Clinical Core will develop aphasia-friendly consent documents using these modifications, additionally incorporating feedback from persons with aphasia. Figure 1 highlights sections of modified documents created (with aid of consultants with aphasia) and used by Richardson. Though not to scale (font size is actually 14-point), aphasia-friendly features are clear: increased spacing between words, increased white space overall, and use of changes in typography, organizational tools, and pictures.

Figure 1. Sample elements from aphasia-friendly ICFs.

3.5 Participant retention. While efforts to attract participants to the study and then initiate their participation is essential, and challenging enough, efforts to ensure participation should not end there. It is vital to the majority of aims across all projects that participants complete all arms of study and that they are available for follow up assessments (see Table 1). Participant retention plans are therefore crucial.[44] Though we refer to participant retention activities here as those engaged after enrollment to ensure project completion (as in Blanton et al.[30]), pre-enrollment activities or factors, such as study design and the informed consent process, can also influence participant retention.[40,41] In regard to study design, studies that provide some benefit and make efforts to reduce participant burden (e.g., travel reimbursement), such as **Project 1** and **Project 2**, are likely to have higher retention rates.[40] **Project 1** and **Project 2** employ therapeutic approaches that have been shown to offer benefit, though, as emphasized in the informed consent process, benefit cannot be guaranteed. Even in the case of **Project 2**, which has a sham group, participants are likely to benefit from the computerized treatment alone.[45] Regarding informed consent, if full disclosure and adequate study explanation has occurred during informed consent, then the participant should have realistic expectations of the study (schedule, burdens, benefits), and is less likely to drop-out because they were surprised or caught off-guard by

expectations.[40] The participant will be given a copy of the consent form and will be encouraged to refer to it and ask questions throughout the study, as informed consent should be an ongoing process.[37]

After a participant is enrolled, several additional steps coordinated by the Clinical Core will assist with successful participation and completion, and concern the following: study personnel, study identity, contact and scheduling methods, and other visit characteristics.[40] First, as detailed in 3.3 and Appendix 1, all personnel will have baseline training in communication with persons with aphasia and their friends and family members and will be screened for characteristics important for working with this population. Each participant will be assigned one primary contact person with whom they will communicate,[40] reducing the confusion that can accompany communicating with multiple contacts through multiple outlets (e-mail, phone) and reducing the potential for mixed messages and lost information. Second, for **Project 1** and **Project 2,** a study logo, study name, and study-specific typography (font color, font style, etc.) (i.e., "study identity"[40]) will be used for all written and digital communication to promote easier discrimination of important study-related communications from others (e.g., junk mail, telemarketing calls, etc.) and a sense of study team membership. Third, once the aphasic individual has consented to participate, they will be mailed a packet that will include the following documentation marked with the study identity: aphasia-friendly study explanation with a personal message from the Clinical Core team and the project PI, aphasia-friendly calendar of scheduled sessions (time, place, etc.), and a contact sheet with pictures and contact information of relevant study personnel.[40] To promote successful follow-up after breaks from regular contact, the Clinical Core will also coordinate reminder phone calls and/or emails to remind of upcoming appointments. Fourth, in regard to both personnel and visit characteristics, the Clinical Core will train all study personnel to be responsive to the participant (and friends and family members of the participant), quickly returning phone calls or emails and never hesitating to answer any and all questions they may have.[30,40] Further, during the session, the participant will have the full attention of the study personnel and sessions will not be rushed or unorganized. Finally, already in place for the lab of **Project 1**, and to be initiated for **Project 2**, a newsletter will be sent frequently to everyone in database (see archived Aphasia Lab newsletters at http://www.mccauslandcenter.sc.edu/aphasia/links-resources/). At the end of the study, participants will be sent a *thank you* card as well as a report of the assessment and treatment results.

The Clinical Core will monitor the success of participant recruitment and retention activities throughout the entire study so that changes to recruitment and retention efforts are modified accordingly. We will utilize an index introduced by Rojavin[46] and endorsed by clinical trial investigators,[30] called the recruitment index (RI), which provides an estimate of the number of days required to recruit one analyzable (completed the study) participant. It is calculated by:

$$RI=[(LPFV–FPFV)xS] \div P$$

where *LPFV – FPFV* is the number of days between "*l*ast *p*atient *f*irst *v*isit" and "*f*irst *p*atient *f*irst *v*isit", *S* is the number of recruitment/study sites, and *P* is the number of participants who completed the study. The RI for each project will be calculated every 9 months to determine whether or not the target enrollment would be reached with the current recruitment rate, and recruitment plans will be adjusted if needed. The Clinical Core will also collect information related to participant attrition, to monitor whether or not it is systematic (e.g., drop- outs more likely to occur during Phase X than Phase Y) and therefore a threat to internal validity,[18] if it should occur.

3.6 Data capture and management. Data capture and management plans are recommended to ensure high data quality for eventual analysis.[44] It is critical to have high quality audio and video for Projects 1, 2, 3, and 4, as all projects rely on one or more of the following that will be performed offline, and most often by blinded raters: detailed analysis of naming performance and errors, detailed transcriptions of discourse, and judgments of motor speech behaviors that rely on motor speech and discourse tasks. It is also important that audiovisual recordings of assessment and treatment sessions are recorded for fidelity monitoring.

Assessor – The Clinical Core will ensure that the assessor-rater will have multiple checks in place to ensure recording of assessment sessions. (We refer to these individuals as assessor-raters because they are responsible for administering the assessment and also scoring or rating those measurements that are *not* repeated measures. Those scoring repeated measures are referred to as blinded raters, or raters. For the remainder of the document, we will refer to them as assessors and raters.) There will be two recording devices in use for each assessment session - for **Project 1,** a laptop camera and a standalone digital camera, and for **Project 2**, a laptop camera and an audio recording device. Quick A/V checks will be performed immediately before the session to ensure devices are working properly. Both recording devices will then be used to record assessment, and the assessor will be reminded to check the indicator lights to make sure capture is in progress. Immediately after the session, the assessor will log information about the session into an experiment manual, checking procedures off of a list developed by the Clinical Core and recording any observations regarding anything of note during the session (e.g., participant very tired; alarm went off during PNT administration; etc.). Then, the assessor will check capture quality and file integrity and back up all recordings to data storage described in **Project 1** and **Project 2.**[47]

Following completion of assessment, the assessor will have a maximum of one week to complete scoring, which will include a combination of their online and offline scoring. For both **Projects 1** and **2**, the assessor will score the Apraxia of Speech Rating Scale,[48] Pyramids and Palm Trees Test,[49] and Western Aphasia Battery- Revised.[50] For **Project 1**, the assessor will also score the Kissing and Dancing Test,[51] Northwestern Assessment of Verbs and Sentences,[52] Philadelphia Repetition Test (PRT),[53] subtests of Psycholinguistic Assessments of Language Processing in Aphasia (PALPA),[54] matrix subtest of Wechsler Adult Intelligence Scale,[55] short-term memory tests,[56] and visual perception tests. For **Project 2**, the assessor will also score the Boston Naming Test (BNT),[47] Patient Health Questionnaire-9 (PHQ-9),[57] Raven's Coloured Progressive Matrices (RCPM),[58] NIH Stroke Scale[59] (with analysis of the Content Units, syllables/Content Unit in the “Cookie Theft” picture,[60-62] and Modified Rankin Scale.[63] Assessors will then immediately input the data into WebDCU™. As described in the Research Plans for **Project 1** and **Project 2**, the following data are automatically stored on the WebDCU™: for **Project 1** and **2**, audiovisual files for pre-treatment and all follow- up assessments of naming untrained pictures (Philadelphia Naming Test [PNT][64,65] Roach et al., 1996; Walker

& Schwartz, 2012) and discourse measures; for **Project 1**, output for Aphasia Communication Outcome Measure (ACOM; Doyle et al., 2013); and for **Project 2**, output for computerized treatment measures. Logging of these behavioral data, as well as scoring and analyses will be conducted by the Clinical Core staff through the WebDCU™ (the videocam in the computer sends videos and responses directly to the WebDCU)™. These steps will be detailed in the assessment procedure manual that will be created by the Clinical Core and reviewed thoroughly in training.

Rater – After a participant has completed all phases of the study, raters will obtain color-coded media files (with participant and session codes that do not reveal the treatment condition or order) for outcome measures (e.g. PNT, discourse).The order in which the blinded rater will score these assessments is determined by a random sequence generator, and will differ for each participant and for each repeated measure. After scoring, raters will immediately input the scored data into WebDCU™ as above.

Clinician – For Project 2, several additional data points need to be entered by the clinician following select sessions. These include participant code, heart rate, blood pressure, Wong-Baker FACES Pain Rating Scale (Wong, 1988), any symptoms noted by the participant, and evaluations of blinding. Date, time, and tDCS blinding is automatically recorded by the WebDCU™ . In addition, the Clinical Core will coordinate the videotaping of randomly selected treatment sessions (1-2 per week) to be used for monitoring of treatment fidelity, discussed in 3.15 and 3.16.

3.7 Data security and storage. Back-up audio and audiovisual files will be stored on equipment supplied by **Project 1** and **Project 2**. All behavioral data, both scored, and unscored, will be stored by the DCU, described in the Administrative Core.

Aim 2: To optimize and monitor assessment fidelity.

3.8 Rationale Poor adherence to assessment procedures threatens both statistical conclusion and internal validity. Specifically, the reliability of measure and the stability of the measurement instrument over time are important for arriving at the correct conclusions about relationships between variables and for ensuring that a false treatment effect is detected when in fact it could be due to change in the instrumentation.[18] While **Projects 1 and 2** rely upon reliable instruments and have no plans to change measurements mid-study, it must be remembered that the clinician is a part of the instrumentation. For most tests, the clinician, operating from an assessment manual, sets the pace, monitors the time limits, gives instructions and in some cases cues or feedback, and must make judgments when scoring performance. The instruments are therefore only as reliable and resistant to change over time as the person administering them.

If assessment instruments were more aligned in their procedures, perhaps this would be less of a concern. However, as is illustrated in Table 3, this is not the case, unsurprising as they originate from different authors, publishers, theoretical underpinnings, etc. Though not an exhaustive review of the variability of assessment administration and scoring procedures included in **Project 1** and **Project 2**, we highlight differences in instructions to the assessor regarding how the clinician is allowed to respond following item presentation, timing, and scoring (focusing on the language tests). As is evident, there are multiple opportunities for drift from assessment procedures and contamination of one assessment instrument by another. For some measures, feedback is not allowed (e.g., KDT) while for others it is prescribed (e.g., PNT). Time limits, if provided, differ between measures and often even between subtests of the same measure (e.g., NAVS). Regarding scoring, some measures score the final attempts (including self-corrections) (e.g., NAVS), others score only the first attempt (e.g., PNT), and some measures even give credit for no responses (e.g., PPTT). So far, this has only highlighted variability; when the complexity of some of the scoring systems is also considered, specifically for discourse and PNT, the need for a plan for monitoring assessment fidelity is obvious. Combining information regarding threats to validity,[18,22] we have identified the following threats that will fall under the purview of the Clinical Core: variability in clinician qualifications, drift, contamination, and clinician turnover. We will apply well- established treatment fidelity guidelines[19,22,27,66,67] to our assessment process and outline here the assessment fidelity plans that focus on assessor and rater training, assessment delivery, and scoring reliability.

Table 3. Sample instructions to clinician regarding assessment administration and scoring.

WAB-R, Naming	"If the patient does not respond or responds incorrectly, ask him or her to hold the object (tactile cue) and to tell you what it is. If the patient still does not respond or responds incorrectly, present the first phoneme of the word (phonemic cue), or, if it is a compound word, the first half of the word (semantic cue)." (cueing)
KDT	(Following two practice items with feedback), "No more feedback is provided to the participant. If the participant does not know what a picture is, do not tell them. If they are unsure of the answer, participants should be encouraged to guess." (feedback)
PNT	"After the subject has finished responding to an item, give him/her feedback: e.g., "good, that's a fish", or "actually, they're looking for garage here", or some variation thereof." (feedback)
NAVS, VCT/SCT	"Items/Sentences may be repeated once." (repetition)
WAB-R, AC	"Repeat the directions and the question if the patient gives an ambiguous or confabulatory response." (repetition)

Timing	
NAVS, VCT	Allow 5 seconds to respond for both first and (after prompt) second attempts.
NAVS, SPPT	Allow 15 seconds for both first and (after prompt) second attempts.
WAB-R, Naming	Allow 20 seconds maximum for each item. [including cues]
PNT	Allow a 30 second deadline to name each picture
Scoring and Attempts	
PALPA, Rhyme	"In a few instances, regional pronunciation differences may result in disagreement about whether stimulus and target rhyme. Check by testing appropriate control subjects and do not count any of these items in your analyses."
PNT	"On each trial, score the subject's first "Complete Attempt", defined as the first minimally CV (consonant-vowel) or VC response (schwa is not counted as a vowel) that meets one of the following criteria: 1) The attempt is not self-interrupted (cut-off) and has clear downward or upward/questioning intonation; may or may not be separated from a subsequent attempt by a noticeable pause; 2) The attempt is not self-interrupted (cut-off), is spoken with level intonation,...."
PPTT	"When a subject completely refuses to respond to a triad, it should be given a chance score – 0.5."
NAVS, VNT	"If the subject self-corrects within 10 seconds, score the final response produced."
NAVS, ASPT	"Self-corrections/reformulations should be put in parentheses. Score *only* the final attempt at a target utterance."
WAB-R, AC	"Score 3 points for each correct response and 0 points for each incorrect (ambiguous or confabulatory) response. If the patient self-corrects, score the last response he or she gives."
WAB-R, Naming	"Score 3 points if the object is named correctly or with a minor articulatory error (e.g., dysarthric slurring) and no cue is needed. Score 2 points if the object name is recognizable, but with a phonemic paraphasia (e.g., "fife" for "knife") and no cue is needed. If a tactile, phonemic, or semantic cue is needed... score as 1 point. Score an incorrect or no response after cueing as 0 points."

3.9 Assessor Training One suggestion for improving the quality of assessment measures, and thereby improving study validity and increasing power, is to conduct "better training of raters" (reference 18; p. 49). The Clinical Core will standardize assessor training for **Projects 1 and 2**, beginning with the development of an assessment manual containing the following sections: project explanation, relationship of assessment measures to project goals, administration procedures for each individual test, and integrated information across the tests highlighting "danger zones" for contamination. Though assessors will be speech-language pathologists who should already have a basic understanding of the project (through baseline and informed consent training, see sections 3.3 and 3.4), illustrating the relationships between the assessments the clinicians are giving and the specific project goals will emphasize the importance of the assessment process and adherence to prescribed procedures. This development of clinician "meta-competence" or "buy-in" is thought to be important in treatment fidelity[22,27] (Borrelli, 2011; Gearing et al., 2011), and is intuitive in assessment as well – understanding the "why" facilitates investment in the "how", or adherence to procedures. In addition, highlighting the opportunities most susceptible to drift or contamination serves to bring the information to the forefront so that it can be actively avoided. Finally, clinician attrition is likely to be observed less in clinicians who are invested in the project goals and feel as if their contribution is important.

The Clinical Core will facilitate the following assessment training procedures (Appendix 1) for both **Project 1** and **Project 2**: (1) independent reading of the assessment manual; (2) video observation of expert administration of each test, administered to persons with varying types and severity of aphasia; (3) small group

training session that will include manual review, highlighting of similarities and differences between administration procedures via discussion and video observation, and supervised role-play with feedback; (4) at project initiation, three supervised assessment sessions with expert feedback; and (5) throughout study, yearly "booster" small group training and 1-2 supervised assessment sessions with

expert feedback. These training procedures will guard against clinician-to-clinician variability, drift, and contamination.

3.10 Assessment Delivery. The Clinical Core will monitor adherence to assessment administration procedures throughout the entire study. A trained independent rater will watch videos of recorded sessions and rate the following for all assessments except the repeated measures (discourse, PNT): adherence to assessment administration procedures; frequency of additional, non-prescribed elements; frequency of omitted prescribed elements; frequency of cross-contamination; and assessor enactment and engagement. The blinded rater will conduct the implementation assessment for the repeated measures, as they will already be spending ample time observing the videos (as discourse transcription and PNT scoring often requires multiple viewings) and will also have expert knowledge of the assessment procedures and scoring. Each assessor will have 20% of their total assessment session time monitored quarterly for accurate implementation**.** Continuous monitoring the delivery of the assessment will further guard against drift and contamination and alert the Clinical Core to the need for action if fidelity is compromised.

3.11 Rater Training

Assessor – As described in 3.6, the assessor will administer and score several tests for each project. The assessor training described above (3.9) includes several opportunities for expert feedback. The feedback will address administration procedures for all assessments as well as scoring procedures and accuracy for all assessments except the repeated measures.

Rater – Blinded raters will be responsible for scoring of the repeated measures. There will be some overlap between training of the assessor and blinded raters. The Clinical Core will expect blinded raters to complete the following alongside assessors (Appendix 1): (1) independent reading of the assessment manual; (2) video observation of expert administration of the repeated measures administered to persons with varying types and severity of aphasia; and (3) attendance to small group training session that will include manual review, highlighting of similarities and differences between administration procedures via discussion and video observation, and supervised role-play with feedback.

Raters will then participate in specialized training sessions for the repeated measures that serve as the primary and secondary outcome measures for the proposed projects. For **Project 1**, this includes the PNT and the discourse production tasks. (The ACOM is also given more than once, but since it is computerized, the only rater responsibility will be checking the files and entry into DCU). For **Project 2**, this includes also the PNT and a discourse production task.

The specialized training for PNT will include the following phases, facilitated by the Clinical Core: (1) independent reading of material included in a blinded rater manual, which will contain *"The Philadelphia Naming Test: Scoring and Rationale"* [64], the *"Detailed Guide to PNT Scoring"* (Moss Rehabilitation Institute), and the *"Scoring Protocol"* (Moss Rehabilitation Institute); (2) co-scoring of two PNT training videos (one mild aphasia, one moderate-severe aphasia) during which the expert rater will narrate the decision-making process and discuss with the trainee any pertinent topics; (3) independent scoring of four PNT training videos, two with nonfluent aphasia (1 high, 1 low severity) and two with fluent aphasia (1 high, 1 low severity), to be checked by the expert rater; (4) feedback session reviewing independent scoring; (5) repeat of steps 2 and 3 if acceptable point-to-point agreement (> 95%) was not achieved; and (6) throughout the study, yearly "booster" small group training and scoring review. Raters will also be instructed about a scoring modification needed for Project 4, where the final self-corrected responses are also scored with PNT scoring conventions. This means that for the majority of aims, the "first complete attempt" is rated as indicated by PNT scoring guidelines and used for analyses, but for Project 4, it is necessary for the raters to also score self corrections on the PNT.

The specialized training for discourse assessment will include the following phases, facilitated by the Clinical Core: (1) direct instruction regarding use of Codes for the Human Analysis of Transcripts (CHAT) and Computerized Language Analysis (CLAN) tools,[68] including EVAL,[69] with relevant sections of corresponding manuals thoroughly reviewed; (2) independent reading of training documents available via AphasiaBank (*"Coding Cheat Sheet", "Error Coding",* and *"CLAN Glossary";* http://www.talkbank.org/AphasiaBank/) and two journal articles (*"Presence, completeness, and accuracy of main concepts in the connected speech of non- brain-damaged adults and adults with aphasia;"* [70] and *"Main concepts for three different discourse tasks in a large non-clinical sample;"* [71] (3) guided transcription of four training videos, two with nonfluent aphasia (1 high,

1 low severity) and two with fluent aphasia (1 high, 1 low severity) to be checked by the expert transcriptionist/rater; (4) feedback session reviewing transcription; (5) co-scoring of lexical and gist discourse measures for two of the four transcribed sessions, during which the expert rater will narrate the decision- making process and discuss with the trainee any pertinent topics; (6) independent scoring of the remaining two transcribed sessions; (7) feedback session reviewing independent scoring; (8) repeat of steps 5 and 6 if acceptable point-to-point agreement (list from literature) was not achieved; and (9) throughout the study, yearly "booster" small group training and transcription and scoring review.

These thorough training procedures for scoring of naming and discourse performance will improve reliability (discussed below) and guard against rater drift. The blinding of raters further improves study validity by reducing the potential for scoring bias. Raters will also be trained how to rate the assessor's adherence to administration procedures, rating the following (as described above): adherence to assessment administration procedures; frequency of additions, omissions, and cross-contamination; and, assessor enactment and engagement.

3.12 Reliability Point-by-point[72] intra- and inter- rater reliability will be conducted for 25% of PNT sessions and 25% of discourse sessions, equally distributed across raters. Intra-class correlation coefficients for 25% of all assessment measures will be calculated.

Aim 3: To optimize and monitor treatment fidelity.

3.13 Rationale As discussed in the Significance section of this Research Plan, treatment fidelity can determine the detection (and if present, magnitude) of treatment effects. Eights of the 13 aims of this P50 application rely on the assumption that treatment will be implemented as intended across clinicians, settings, participants, and for the entire duration of the study. Unreliable treatment implementation is a known threat to statistical conclusion validity;[18] and Gearing and colleagues[22] identified threats to treatment fidelity at every component (design, training, delivery, receipt), which we use here to outline the Clinical Core strategy for optimizing treatment fidelity. The strategies for Project 1 and Project 2 will be discussed separately.

3.14 Clinician Training

Project 1 – Standardized clinician training, relying upon a detailed manual of treatment procedures, is recommended if treatment fidelity is desired.[19] The Clinical Core will be responsible for the development of the treatment manual, which will include the following sections[22,66]: project explanation, relationship of treatment activities to project goals, treatment administration procedures, comparison of active ingredients of each treatment approach (including highlighting of "danger zones" for contamination as in the assessment manual), and troubleshooting tips. Similar to the assessment manual, the purpose of explaining the project and emphasizing the importance of treatment activities, and the differentiation thereof, promotes clinician buy-in and subsequent adherence to procedures.[22,27] Explicit instructions for treatment tasks are included and are likely to reduce the likelihood of cross-contamination.[66] The (non-computerized) tasks used in **Project 1** are no more complex than those practiced in everyday clinical practice, as indeed they arose from those practices; they are simply operationalized so that procedures would be standardized across persons and settings.

The Clinical Core will facilitate the following treatment training: (1) independent reading of the treatment manual; (2) video observation of treatment delivery (available from pilot study) to persons with varying types and severity of aphasia; (3) small group training session that will include treatment manual review, with highlighting of "danger zones", as well as supervised role-play with feedback; (4) at project initiation, three supervised sessions of each treatment arm with expert feedback; and (5) throughout the study, yearly "booster" small group training and 1-2 supervised treatment sessions (both treatment arms) with expert feedback. These training procedures will guard against clinician-to-clinician variability, drift, and contamination.

Project 2 – Standardized training for **Project 2** will also include a detailed manual of treatment session procedures, though the challenges for treatment fidelity are different from **Project 1** due to the computerized nature of the speech-language treatment, and the inclusion of brain stimulation. Because of steps taken during stimuli development, there is little threat of contamination. The threats to treatment fidelity that need to be addressed during training are related to the administration of brain stimulation. Perhaps the most complex element of treatment for **Project 2** is electrode placement for transcranial direct current stimulation (tDCS). With training and experience, this process becomes simple and routine. We have had no difficulty training clinicians to do this for the CATES trial (PI: Fridriksson) or for the studies at JHMI (Hillis). During the first two years of the project, Dr. Richardson will visit JHMI to ensure the electrode placement (particularly for participants randomized to fMRI-guided localization of electrode placement) is conducted in the same way as in the CATES trial. The greatest potential threat to validity concerns blinding, and automatic blinding of both the participant and the clinician the WebDCU^{TM} makes this threat negligible. The clinician enrolling the participant will not see the treatment assignment, only a numeric participant number generated by the DCU server, which is entered into a software package that controls whether tDCS or Sham is delivered to the participant. The in-house software/hardware setup allows for switching the tDCS on and off without any involvement from the participant or experimenter. The Clinical Core will develop the treatment manual for **Project 2**, which will include the following sections: project explanation, relationship of treatment activities to project goals, treatment administration procedures, and troubleshooting tips.

Training procedures facilitated by the Clinical Core for **Project 2**, most of which have been implemented with success for CATES, will include the following (Appendix 1): (1) independent reading of the treatment manual; (2) video observation of treatment delivery; (3) small group training sessions that will include treatment manual review, with a focus on the importance of blinding, as well as direct instruction and practice with electrode placement and device operations; (4) at project initiation, two supervised treatment sessions, with brain stimulation portions co-led with experts (Celnik, Richardson); (5) two supervised treatment sessions with expert feedback; and (6) throughout study, yearly booster training. These training procedures will guard against clinician-to-clinician variability, drift, and unblinding.

3.15 Treatment Delivery The Clinical Core will monitor adherence to treatment procedures throughout the entire study. Based upon recommendations by Gearing et al.[22], a trained independent rater will watch videos of recorded sessions and rate the following for both **Project 1** and **Project 2**: adherence to treatment procedures (Project 1, Appendix 2; Project 2, Appendix 1); frequency of additional, non-prescribed elements; frequency of omitted prescribed elements; frequency of cross-contamination; and clinician enactment and engagement. Each clinician will have 20% of their total treatment session time monitored quarterly for accurate implementation**.** Continuous monitoring of treatment delivery will guard against drift and contamination, and alert the Clinical Core to the need for action if fidelity is compromised.

For **Project 2**, the rater will also make a judgment about whether or not they thought the brain stimulation was sham or active. The rater will enter this information into the DCU, so that they can be alerted if the match of rater judgments to actual administration is above chance, which would indicate that there might be something observable in the treatment that could potentially unblind the participant that would need to be addressed immediately.

3.16 Treatment Receipt Another component of treatment fidelity involves monitoring whether or not participants demonstrate behaviors that indicate they have received the treatment, which can be assessed by rating within-session participant comprehension or utilization of treated behaviors.[22,66] The trained independent rater assessing treatment delivery (3.15) will also rate the following for **Project 1**: participant engagement, participant comprehension of treatment activities, and frequency of participant attempts to perform treatment activities. This information combined with the within-session data collected by the clinician regarding success, cueing needed, etc. will serve as an indicator of treatment receipt. For **Project 2**, the rater will also assess participant engagement, and this rating combined with data produced after each computerized session will indicate treatment receipt.

3.17 Clinical Core Management Plan – Overall management of the Clinical Core will be in the hand of Dr. Hillis. She will be assisted by Drs. Fridriksson (USC), *to be hired* Clinical Care Coordinator (USC), and Bonilha (MUSC). The Clinical Core Coordinator will manage and implement day-to-day operations, as described above, with *ad hoc* supervision by Hillis. Priorities will be determined by Hillis, Clinical Core Coordinator, and Fridriksson together in weekly videoconferences via www.GoToMeeting.com. Potential disputes will be resolved by this core leadership, using a conflict resolution model described by Shearouse,[73] as described in detail in the Administrative Core.

3.18 Potential Overlap - There is no budgetary overlap between the Research Subprojects and the Clinical Core. Dr. Hillis will devote a total of 35% effort to the C-STAR (25% to Project 2, 10% to the Clinical Core).