Published in final edited form as:
*Neuroimage*. 2013 August 1; 76: 428–435. doi:10.1016/j.neuroimage.2012.10.011.

# Neuroimaging in aphasia treatment research: Standards for establishing the effects of treatment

**Swathi Kiran**[1], **Ana Ansaldo**[2], **Roelien Bastiaanse**[3], **Leora R. Cherney**[4,5], **David Howard**[6], **Yasmeen Faroqi-Shah**[7], **Marcus Meinzer**[8], and **Cynthia K Thompson**[9,10]

[1]Boston University, Sargent College of Health and Rehabilitation Sciences, Boston, Massachusetts, USA [2]Centre de recherché de l'Institut universitaire de gériatrie de Montréal, Montréal, Québec, Canada [3]University of Groningen, Center for Language and Cognition Groningen (CLCG), Groningen, The Netherlands [4]Rehabilitation Institute of Chicago, Center for Aphasia Research Language and Treatment, Chicago, Illinois, USA [5]Northwestern University, Department of Physical Medicine & Rehabilitation, Feinberg School of Medicine, Chicago, Illinois, USA [6]Newcastle University, Centre for Research in Linguistics and Language Sciences, Newcastle, UK [7]University of Maryland, Department of Hearing and Speech sciences, College Park, Maryland, USA [8]Charite Universitätsmedizin, Department of Neurology, Center for Stroke Research Berlin & Cluster of Excellence NeuroCure, Berlin, Germany [9]Northwestern University, Department of Communication Sciences and Disorders, Evanston, Illinois, USA [10]Northwestern University, Department of Neurology and the Cognitive Neurology and Alzheimer's Disease Center, Feinberg School of Medicine, Chicago, Illinois, USA

## Abstract

The goal of this paper is to discuss experimental design options available for establishing the effects of treatment in studies that aim to examine the neural mechanisms associated with treatment-induced language recovery in aphasia, using functional magnetic resonance imaging (fMRI). We present both group and single-subject experimental or case-series design options for doing this and address advantages and disadvantages of each. We also discuss general components of and requirements for treatment research studies, including operational definitions of variables, criteria for defining behavioral change and treatment efficacy, and reliability of measurement. Important considerations that are unique to neuroimaging-based treatment research are addressed, pertaining to the relation between the selected treatment approach and anticipated changes in language processes/functions and how such changes are hypothesized to map onto the brain.

### Keywords

Aphasia; language recovery; fMRI; treatment efficacy

## 1.0. Introduction

The goal of this paper is to provide guidelines for designing and implementing treatment studies that aim to examine the neural mechanisms associated with language recovery in aphasia, using functional brain imaging. This research requires measurement of neural changes from pre to post intervention using functional magnetic resonance imaging (fMRI), PET (positron emission tomography) or other methods (e.g. ERPs). In addition, and the

Correspond with: Swathi Kiran, Boston University, 635 Commonwealth Ave. Boston, MA 02215, Phone: (617)-358-5478, Fax: (617)-353-5074, kirans@bu.edu.

focus of the present paper, careful measurement of language and/or cognitive changes from pre- to post-intervention and interpretation of the relationship between the two sets of changes (neural and behavioral) are required. As pointed out in recent reviews, there is variability in regions of the brain recruited by people with aphasia to support language recovery both within and across studies (see Crinion & Leff, 2007; Thompson & Den Ouden, 2008, Meinzer, 2011). Possible reasons for this may be related to the treatment provided and the experimental designs used to evaluate its efficacy. Although there have been recent methodological advances in the measurement of language behavior in individuals who have suffered a stroke using fMRI (Abutalebi, Rosa, Tettamanti, Green, & Cappa, 2009; Bonakdarpour, Parrish, & Thompson, 2007; Fridriksson, Morrow-Odom, Moser, Fridriksson & Baylis, 2006; Kurland et al., 2004; Peck et al., 2004; Rorden, Fridriksson, & Karnath, 2009, Marcotte & Ansaldo, 2010), few studies have systematically investigated the effects of rehabilitation on brain mechanisms recruited to support recovery. In this paper, we address a series of questions on the design of treatment studies when treatment effects are assessed both behaviorally and in terms of brain activations, presenting the consensus derived from discussions among experts in neuroimaging and aphasia at the Neuroimaging in Aphasia Treatment Research Workshop, held at Northwestern University in September, 2009. Because the nature of the experimental design, task manipulations and spatio-temporal manifestations of the data are different for fMRI studies and ERP studies, we limit our discussion to fMRI studies in this paper.

The first section of the paper considers different options for designing treatment experiments. Specifically, we discuss group versus single-subject experimental or case-series design options for establishing the effects of treatment and consider their advantages and disadvantages. We examine the general components of and requirements for treatment research studies, including the operational definition of variables, the criteria for defining behavioral change and treatment efficacy, and the reliability of measurement. We also point out unique considerations required in neuroimaging-based treatment research, concerning the relation between the treatment approach selected and the anticipated changes in language processes/functions and hypotheses about how changes in language function are expected to map onto changes in brain function. Other design considerations relevant to relating the effects of treatment directly to changes in brain function are covered in other papers in this series. For example, questions related to the reliability of activation patterns seen on repeated scans, fMRI task selection, and single-subject versus group approaches to analysis of the fMRI data are discussed in Rapp et al. (this volume) and Meinzer et al. (this volume).

## 2.0. Establishing the Effects of Treatment (Internal Validity)

The first essential requirement in designing a treatment study to evaluate treatment-induced neural plasticity is that the experiment uses a design that allows the researcher to establish that behavioral changes are a result of treatment (internal validity). There are several experimental approaches for accomplishing this – group approaches that compare the performance of experimental and control groups, and single-subject approaches that compare performance between experimental and control phases in the same participant. Both design types, if implemented properly, rule out the influence of extraneous variables, (e.g., environmental or participant factors), on the language behaviors or processes under study. The philosophy is the same for both: between-group experimental designs compare the performance of groups of individuals (experimental and control groups), whereas, single-subject experimental designs compare the performance of individual participants during experimental and control (baseline) phases. The idea is that similar extraneous variables are at play in both the experimental and control groups or conditions and that the influence of these variables on the behavior(s) under study can be ruled out by comparing patterns of performance between the two groups or conditions (see Thompson, 2006).

In studies examining the neural mechanisms of treatment-induced language recovery, the experimental treatment design employed is not only relevant to establishing the efficacy of treatment, but it also impacts analysis of the neuroimaging data. Between-group treatment design requires averaging the treatment effect in the experimental groups and comparing change over time between the treated and untreated groups. Thus, to estimate the effects of treatment on brain processing, a group approach to analysis of the fMRI data is required. However, the group approach may be confounded because it is possible (and likely) that not all participants in the experimental group will change to the same extent. As pointed out by Meinzer et al., (this volume), group analyses of aphasic neuroimaging data, in general, should be approached with caution because of individual differences in variables such as lesion site and extent, unless the goal of the study is to account for the effects of such variables on either treatment-induced behavioral performance or neural recruitment patterns, which requires large, rather than small sample sizes. Conversely, single-subject/case-series designs require measurement of language change throughout the treatment period, with no data averaging across study participants. The neuroimaging data derived from pre- and post-treatment scans of individual participants then can be examined and evaluated with regard to treatment improvement. However, there is an inherent lack of power to detect changes in activation over time when comparing changes in neural activation in individual study participants. It is therefore, important during the experimental design phase to include a sufficient number of experimental trials in the neuroimaging task. In addition, this practice has drawbacks with regard to external validity, or generalization to other individuals with aphasia. However, this latter problem can be addressed by replication of treatment across participants (see below for further discussion of single-subject experimental designs with regard to replication across study participants).

Independent of experimental design, it is important to conduct a power analysis and sample size estimation to justify the inclusion of a particular sample size and interpretation of a particular effect size. Particularly for neuroimaging treatment studies that are inherently clinical in nature, justifying the sample size and benchmarks for effect size can be very beneficial in evaluating what constitutes a clinically (or theoretically) important effect.

### 2.1. Establishing Experimental Control Between Groups

Between-group designs require at least two groups of participants, an experimental group that receives the (experimental) treatment, and a control group that either does not receive treatment or is provided with an alternative treatment or placebo. At the beginning of the study, both experimental and control participants are tested on one or all dependent measures, both behavioral and neuroimaging, and at the end of the study these measures are repeated. Performance on each measure is averaged across participants in each group at each time point and a treatment effect is established when the experimental group shows significantly greater pre- to post-treatment changes than the control group. One requirement of between groups experimental designs is that study participants be randomly selected from a population of people (e.g., those with aphasia or a particular aphasia profile). When random selection is not accomplished, an unwarranted extrapolation from the sample to the study population may occur, creating a problem of sample bias. Notably, when studying disorders such as aphasia, random selection of participants from the entire population of people with the disorder is not possible. Hence, researchers generally use “populations of convenience” from which to select their study participants (e.g., aphasic individuals in a particular geographic region). Although this practice itself does not preclude hypothesis testing and the use of parametric statistics, researchers rarely, if ever, randomly select study participants. Indeed, one of the advantages of well-designed group studies is that inferential statistics can be applied elegantly to estimate the generality of the findings to the population.

Importantly, however, this is not possible if the study sample is not randomly selected from the population.

Relatedly, once selected, potential participants meeting pre-specified inclusionary and exclusionary criteria must be randomly assigned to either the experimental or control group or a match pairs approach may be taken, where study participants are stratified based on their lesion patterns and/or other variables. Even though some studies examining the neurobiology of treatment-induced recovery from aphasia have studied relatively large groups of participants with aphasia (see, for example, Richter, Miltner, & Straube, 2008 (n = 16); Fridriksson 2010 (n=26)), few have employed random selection or assignment strategies to generate experimental and control groups (e.g., Cherney, Erickson, & Small, 2010). In fact, no studies to our knowledge have included a control group of aphasic individuals at all. Thus, even though studies report change in the treated groups' language behavior from pre- to post-treatment (and associated changes in fMRI activation) the extent to which behavioral or neural changes noted over the course of the study can be attributed to the treatment provided (rather than other uncontrolled variables including spontaneous change) is unclear.

As an alternative to using a control group of individuals with aphasia some studies have included a control group of non-brain-damaged participants, examining the neural correlates of their learning. For example, Raboyeau et al., (2008) trained non-brain damaged French-speaking individuals to produce novel words in either Spanish or English and compared pre- to post-treatment activation patterns associated with naming them. This approach provides information about learning in healthy adult brains, enhancing understanding of the neural mechanisms engaged for learning (or re-learning) in compromised brains. However, it does not provide experimental control, substantiating that learning (in either the aphasic or normal group) actually took place. For this, a control group of matched study participants who do not receive the experimental treatment is required.

One reason that control groups often are not included in studies of aphasia treatment concerns the issue of withholding treatment. Although, in theory, withholding an experimental treatment (particularly when the effect of the treatment is unknown) is probably not unethical, the idea of withholding treatment, even an inadequate one, is not popular among clinicians, individuals with aphasia, or their family members. Because of this, researchers are often faced with the less preferred option of including, as control participants, individuals with aphasia, who for reasons such as motivation, family support, transportation to the laboratory, and the like, are not able to participate in the experimental group (although this strategy has not been used to date in studies examining the neural mechanisms supporting language recovery). Both of these situations – that is, failure to include a control group, or including groups of individuals who are unable to participate in the study as controls – are problematic from a methodological point of view: as pointed out above, a control group is needed in order to generate internally valid data and the control participants must be selected from the same population as the experimental participants with an equal chance of being assigned to the experimental group. One possible group design strategy, which avoids withholding treatment, is to use a crossover design in which control participants are entered into treatment after it has been completed for the experimental group. In this design, participants are randomly assigned to a specific treatment sequence. Participants who initially do not receive intervention serve as no-treatment controls and are then entered into treatment during the second arm of the study. We are not aware, however, of any fMRI studies that have used this approach (but see, Fridriksson (2011) who used this design to evaluate the impact of tDCS).

Another requirement of group designs is that a pre-specified number of participants be included in order to insure sufficient statistical power to demonstrate a treatment effect. The number of participants also is relevant to studies concerned with examining changes in neural activity associated with treatment improvement in groups of participants. That is, the statistical power of the data is reduced when too few participants are included in the analysis. In addition, as pointed out by Meinzer et al. (this volume), one of the advantages of using a group approach for analyzing neuroimaging data is that this allows researchers to explore (e.g., correlate) the relation between behavioral and/or lesion variables and treatment outcome. Notably, however, most studies examining the functional reorganization of brain tissue associated with treatment-induced language recovery have included small numbers of participants. For example, of the 24 studies summarized by Thompson (this volume), only three included more than 10 participants, and 18 included three or fewer. Nevertheless, some group studies have performed regression or correlational analyses on the experimental group, even with data from few patients. For example,Menke et al. (2009) examined the relation between short -term training effects (i.e., percent accuracy) and BOLD signal change from pre- to post-treatment in eight anomic aphasic participants and found positive correlations between training success and signal changes in memory related structures, including the hippocampus.

In considering whether or not to utilize a between-groups design to examine the effects of treatment for aphasia, the issue of homogeneity is important to consider. It is well known that individuals with aphasia differ greatly with often varying language patterns and associated lesions, and even study participants carefully selected for their deficit patterns are seldom, if ever, homogeneous. They can, and do, differ markedly. Given this heterogeneity, it is often the case that treatment effects differ across individuals, and in turn, the neural recruitment patterns associated with behavioral change will likely differ across participant. Thus, averaging changes either behaviorally or neurologically across participants from pre- to post-treatment may be contraindicated and lead to inaccurate and / or misleading interpretation of the data. On the other hand, an advantage of group designs is that if participants are somewhat homogeneous (for example, grouped by similar lesions or similar behavior) the data can be potentially powerful for identifying predictors for treatment success (Menke, et al., 2009). Clearly, however, this practice has the potential to mask information about how certain individuals respond to treatment as well as the brain tissue recruited to support recovery. For example, treatment outcomes associated with right hemisphere and/or perilesional activation may be masked by individual variability (Crosson et al. 2007b). Therefore, it can be very difficult to draw any meaningful conclusions from a group of aphasic individuals.

**2.1.1. Summary—**One of the cornerstones of experimental treatment research is that proper controls be put in place such that the effects of the experimental treatment (either behavioral and neurological) can be established. The reader is referred to an analogous set of standards in physical rehabilitation studies for assistance in designing group studies (http://www.otseeker.com/PDF/PEDroScalePartitionedGuidelinesExplanations.pdf) Group experimental designs accomplish this by randomly selecting and assigning groups of study participants to either experimental or control groups. This practice allows the results of the study to be generalized to the population from which the participants were selected and, if enough participants are included, group studies have the advantage of allowing researchers to explore the relation between behavioral and neurological variables and recovery. Notably, no studies examining the neural mechanisms associated with treatment-induced recovery from aphasia have included a control group of aphasic individuals, perhaps because large numbers of study participants are required when using this approach and/or because such designs require withholding of treatment (or application of a placebo treatment) from the control group.

## 2.2. Using Single-Subject, Case-Series Strategies to Establish Experimental Control

As discussed above single-subject and/or case-series designs involve control and experimental phases, which are compared to one another for each participant in the study, such that the effects of the treatment can be determined. In this case experimental control, demonstrating that participants improve only when they are treated, is achieved by comparing the experimental phase with the baseline/control phase within each participant. Hence, no control groups are required. It is important to point out at the outset that these designs are not the same as case studies which document the effects of treatment without establishing experimental control. Technically, single-subject experimental designs also require study of more than one participant (and, therefore, are sometimes referred to as 'case series design') because replication of the treatment effect within and across study participants is required (McReynolds & Thompson, 1986). Therefore, they are not synonymous with N=1 studies.

The control (i.e, baseline) and experimental phases in single-subject, case-series designs are typically labeled A and B, respectively, and the behavior under study is continuously measured throughout these phases. This is accomplished by administration of the dependent measures regularly, using identical procedures throughout all phases of the experiment. As such, in these designs behavioral change is examined as it unfolds over the course of the study for individual participants, allowing close examination of behavioral variability as a function of the time series data. In turn, associated changes in neural activation associated with treatment can be examined, and where appropriate, common patterns/trends in activation across participants can be noted. In this regard, single-subject, case-series designs are unlike group designs, which require that the dependent measures be measured only twice – once prior to treatment and once following its completion – with averaging of group performance at the two test points.

**2.2.1. Types of designs—**There are several types of single-subject, case-series designs, which have been described extensively by others (see Kazdin, 1982; McReyolds & Kearns, 1983; McReynolds & Thompson, 1986 and many others). However, some designs are more appropriate than others for studying treatment induced recovery of language in aphasia. Below we discuss major single-subject experimental design types. The reader is also referred to Tate, McDonald, Perdices, Togher, Schultz, & Savageet (2008) for suggestions on designing and reporting single subjects/case series designs.

**2.2.1.1. The A-B-A-B design:** A common design is the A-B-A-B design. In this design, the behavior(s) under study are first measured in the baseline (A phase), then treatment is applied in a B phase, following baseline. The treatment is then withdrawn in a second A phase, and finally treatment is reapplied in a second B phase. In order to demonstrate experimental control using such a design (a) performance on the dependent variable must be stable during the first A phase, (b) a change in the dependent variable(s) must be seen when comparing performance in the first A phase with that in the first B phase, (c) the dependent variable(s) must reverse in the second A phase, that is, return to baseline levels, and (d) during the second B phase, the treatment effect must be re-established, that is, change in the dependent variable(s) is once again seen. This sequence of events allows within subject replication and when shown in several participants, across participant replication is established.

The methodological requirement that the dependent variable(s) return(s) to baseline levels in the second A phase presents a major problem for treatment research in aphasia, because the goal of such research is to improve language function. If the treatment is successful, a reversal is undesirable, and may not be possible. Although there are methods for "forcing a

reversal", such as training erroneous responding during the second A phase, this practice in aphasia treatment is not recommended. Furthermore, most aphasia treatment studies are geared toward showing longer-lasting effects. Demonstrating long term change confounds the reversal to baseline requirements in these ABAB design, and thus need to be implemented and interpreted with caution.

**2.2.1.2. The multiple baseline design:** A frequently used alternative to the A-B-A-B design is the multiple baseline design across behaviors, which does not require returns to baseline levels of responding to demonstrate internal validity. This design, in essence, is a series of A-B designs, with sequential iterations of treatment applied to sets of stimuli, with increasing baseline periods for each set. For example, baseline (A phase) data are collected on two or more sets of stimuli for each participant and following this, treatment is applied to the first set (in the B phase), while the A phase is continued for untrained sets. When a treatment effect is established for the first set, treatment is extended to the second set, and so on, until all have been treated. Experimental control is demonstrated when changes in the dependent variable(s) occur only when the B phase is in effect for each behavior; baseline performance of untreated behaviors remains stable, until treated. Replication of the treatment effect within participants is established by showing improved performance when treatment is applied to each set of stimuli. Across subject replication is established by entering more than one participant into the study.

Because the multiple baseline design requires sequential application of treatment to separate sets of stimuli, order effects must be ruled out. Thus, application of treatment to selected stimulus sets is counterbalanced across participants and the number of participants required for a particular study depends on the number of sets. Take for example, a study examining the effects of treatment on naming. The experimenter decides to study two sets of words, with each set tested in the baseline phase and subsequently sequentially trained. To rule out order effects, the order of training each word set is counterbalanced; hence a minimum of two participants is required with each receiving a different order (i.e., word set order : 1, 2; 2, 1). For full replication in such a study, four participants are required. Notably, as the number of behaviors selected for treatment increases, the number of participants required also increases. For example, for a study with three sets of words, six participants would be required for complete counterbalancing (i.e., word sets in the order 1,2,3; 1,3,2; 2,1,3; 2,3,1; 3,1,2; 3,2,1) and an additional six for full replication. The nature of single-subject experimental design, however, allows researchers some flexibility with regard to participant numbers. For example, if in a study that technically requires 12 participants (i.e., in an experiment using a multiple baseline design across behaviors, involving three behaviors), the first six participants all respond to treatment as expected (i.e., they show acquisition of trained word sets as each is trained and maintain baseline levels of performance on untrained sets), this would constitute 18 replications of the treatment effect (3 replications×6 participants), which is adequate for demonstrating the effects of treatment and for establishing internal validity (see Thompson & McReynolds, 1986; Connell & Thompson, 1986)

Another issue relevant to the multiple baseline design across behaviors is that the behaviors must be functionally independent and, at the same time, amenable to the treatment under investigation. This means, for example, that in a naming study using three sets of words, training one set would have no effect on the untreated sets. If the behaviors are not functionally independent, treatment of one set may influence the others, that is, generalization may occur across sets. Although such an effect is often desired, particularly in aphasia treatment research when one goal is to examine for (and promote) generalization to untrained language behavior, this situation is experimentally problematic, that is, experimental control is lost.

Rather than using a multiple baseline design across behaviors, which requires sequential training of selected sets of items following baselines that remain stable according to a pre-set criteria (e.g., no greater than 20% change across three sessions) some researchers select multiple sets of items, with the intent of leaving one set untrained, and expecting behavioral change on the trained, but not on the untrained (control) items. Changes in the experimental compared to control items is then compared statistically. This strategy provides internally valid data, if enough items are included in the trained and untrained lists. Note that the larger the stimulus set sizes, the greater the power for detecting changes as a function of treatment. However, this strategy carries a risk that generalization may occur from the trained to the untrained items. Although this often is a goal of aphasia intervention, its occurrence in this situation would result in a lack of experimental control, and hence a failed treatment study. It also is possible that untrained items may, for some reason, not be amenable to improvement under any circumstances. To avoid this potential issue, stimuli can be randomly assigned to trained and untrained sets such that the comparisons of change are meaningful. Another approach to circumvent the potential confound of generalization/ experimental control is to use a multiple baseline design across behaviors. Such a design requires checking for generalization to untrained sets throughout the course of treatment and applying the treatment to any untrained sets to which generalization does not occur.

An alternative to the multiple baseline design across behaviors is the multiple baseline design across participants. Rather than using behaviors or stimulus sets to demonstrate experimental control, study participants are employed for this. Specifically, treatment is applied to one (of several) study participants following baselines phases of varying length. The logic here is that treatment will improve language when and only when it is applied. Thus, if it is the treatment, and not extraneous variables, that are responsible for the behavioral effect, no change will be seen for any participant during the baseline phase, regardless of its length.Fridriksson et al. (2007), for example, employed this strategy by varying the order of treatment application across participants in their examination of the effects of phonological and semantic cueing in three aphasic individuals.

Combining the multiple baseline across behaviors and participants design is a particularly useful strategy for studying the effects of treatment for aphasia if one goal of the work is to examine for generalization from trained to untrained items. For example, Thompson, den Ouden, Bonakdarpour, Garibaldi, & Parrish (2010) used this experimental design to study changes in neural activation associated with treatment of sentence level deficits in six individuals with Broca's (and concomitant agrammatic) aphasia. Three different, but psycholinguistically related, sentence types were selected to comprise the multiple baseline across behaviors and participants were tested for their ability to comprehend and produce them in baseline phases of differing lengths. One sentence type at a time was then trained while generalization was examined to the untrained sentence types. Results showed successful generalization across sentences for all participants, as expected, precluding the necessity of training all sentence sets, but resulting in a lack of ability to show experimental control across behaviors (i.e., sentence sets). Rather, experimental control was demonstrated across participants, in that no behavioral change was noted for any sets of stimuli during the baseline phase for any participant. Changes in the dependent measures (i.e., sentence comprehension and production) only occurred when treatment was applied. This extra design component serves as an insurance policy; if generalization occurs, experimental control is maintained. There are other types of single subject controlled experimental designs that can be used to examine the effects of treatment, however, the multiple baseline strategy – either across behaviors or participants – is the most commonly used and is likely the best suited for most studies of aphasia.

The primary limitation of single subject experimental, case-series designs pertains to the putative lack of ability to generalize findings derived from such studies to a larger population. This idea is true if one relies on inferential statistics to estimate the generalizability of findings. Indeed, parametric statistics is inappropriate for use with data derived from single subject experimentation (i.e., comparing performance in baseline compared to treatment phases) for a number of reasons including serial dependency. In single subject, case series designs external validity is addressed through replication of treatment effects within and across participants, both within individual studies and across studies. The logic is simple: the greater the number of replications, the greater the generality of the effect. This is no different than logical, non-statistical generality statements derived from between-group studies in which random selection is not accomplished.

**2.2.2. Summary—**Single-subject experimental or case series designs are powerful alternatives to group experimental design strategies and have several advantages. Not only are fewer participants required, control groups of participants are not necessary because experimental control is demonstrated within participants rather than between participant groups. Additionally, and of particular importance for establishing the neural mechanisms of language recovery, these designs afford careful inspection of individual participant's learning patterns over time which can be captured as changes in BOLD signal as a function of treatment. We emphasize, however, that regardless of which design strategy is used, both single-subject and group approaches require that the proper experimental controls be put in place such that both behavioral (i.e., treatment-induced) and BOLD signal changes occurring from pre- to post-treatment can be directly attributed to the treatment provided. The precise design selected for this, of course, is at the discretion of the researcher and depends on the aims of the study.

## 3.0. Requirements of Experimental Treatment Research

Independent of the experimental design used to establish experimental control, there are a number of other important requirements and considerations for designing studies to examine the neurobiology of language recovery. Several of these are discussed in other papers in the series, pertaining to describing and quantifying participant criteria associated with brain lesions (see Crinion et al., this volume) as well as the disrupted language system (see Rapp et al., this volume). Here we address requirements for specification of and rationale for the treatment selected, including the dosage of treatment (i.e., the intensity and duration of treatment application), and the behavioral tasks included to evaluate the outcome of treatment. We also address the linking between treatment outcome variables and the neuroimaging tasks selected. We also briefly discuss reliability of measurement.

### 3.1. Treatment and Outcome Measures

All experimental studies require specification of the independent and dependent variables. In research examining the neural mechanisms of treatment, the former include primarily the treatment itself (although other independent variables such as lesion age or volume, the extent of hypoperfused tissue, etc. also may be included in such studies), whereas the latter refer to measures employed to examine for changes in behavior and brain processing.

Defining the treatment under investigation in neuroimaging studies is no different than in any other study examining treatment efficacy in that a detailed description of all aspects of the treatment, including the stimuli, response criteria, and any training procedures is required. This is important such that the treatment can be replicated in future studies and applied with precision clinically. The dependent measures also require precise description, detailing how the outcome of treatment is to be measured. A common practice in treatment research is to develop probe tasks explicitly designed to measure the language behavior

under study, as well as related behaviors, in conditions in which no feedback is provided. These tasks can include both on-line measures, such as reaction time, and/or off-line measures, depending on the goal of a particular experiment. It is the participant's responding to these probe tasks that serve as the primary dependent variable throughout the study.

For neuroimaging-based treatment research there are additional considerations pertaining to the independent and dependent variables. First, the relation between the selected treatment, and the impaired process(s) it is putatively addressing, and the behavioral outcome measure needs to be considered. That is, a clear explanation of how the treatment task addresses the impairment and how the dependent measure captures any change in the impaired process needs to be provided. For instance, in Marcotte & Ansaldo (2010) a semantic feature approach addressed severe anomia by boosting the targets' semantic representations to improve access to the impaired phonological word forms, a rationale based on Spreading activation Theory (Collins & Loftus, 1975); Also, Davis, Harrington, & Baynes (2006) sought to improve naming in a patient with Wernicke's aphasia with a word comprehension deficit, putatively related to difficulty selecting items from lexical-semantic competitors. To remediate this problem, Davis et al., used a semantic-feature treatment, which trained the aphasic individual to select target items based on their semantic attributes. The idea was that this treatment would influence the ability to inhibit competitors and, hence, improve word comprehension.

Second, in neuroimaging studies of aphasia treatment it is necessary to elucidate how changes in processes targeted in treatment will trigger changes in brain processing, measurable using fMRI. That is, the brain regions engaged to support that process in healthy individuals and regions expected to be engaged to support recovery need to be considered.Crosson et al. (2005), for example, trained participants with aphasia to name objects as they performed a complex left-handed movement task, with the idea that this pairing should facilitate engagement of a right medial frontal intention mechanism and, hence, result in an increase in right pre-SMA activation. In addition, they hypothesized that treatment would result in an increase in right lateral frontal activation, associated with improved naming. In Marcotte & Ansaldo (2010), the semantic feature approach used was expected to promote the development of a semantic strategy for word retrieval, which could rely upon preserved semantic processing areas in the left and the right hemispheres.

A third consideration is the fMRI task used to evaluate the effects of treatment. That is, the task(s) must be designed such that the neural mechanisms underlying the language process under study are elucidated. Therefore, it is important for researchers to integrate the fMRI tasks and the tasks used to evaluate the behavioral outcome of treatment. This allows the activation patterns noted during fMRI tasks to be linked with behavioral changes associated with treatment. For example, Fridriksson et al. (2007, 2010), who used phonological and semantic cuing strategies to improve naming in individuals with anomic aphasia, directly examined naming ability prior to and following treatment, utilizing picture naming as the primary outcome variable associated with treatment improvement and as the fMRI task (Fridriksson, et al., 2006; Fridriksson, et al., 2007). Marcotte & Ansaldo (2010) examined oral naming during fMRI scanning prior and after semantic feature therapy for anomia. The authors showed that plasticity operated differently in either case, despite the similarity of naming recovery profiles. In another study, Kiran, Sebastian, Chettiar, & Devous (2008) aimed to strengthen semantic representations in aphasic participants who presented with naming deficits resulting from an underlying semantic impairment. Hence, treatment focused on strengthening semantic representations through feature verification and the fMRI tasks included picture naming as well as a semantic feature verification task (Kiran, et al., 2008). By incorporating fMRI tasks that relate directly to the treatment tasks, interpretation of changes in patterns of activation as a function of treatment can be elucidated. When

different tasks are used to evaluate the behavioral effects of treatment and the neurological impact of treatment, it is difficult to link improvement in treatment to changes from pre- to post-post fMRI scans.

### 3.2. Treatment Dosage

The dosage of treatment, that is, treatment intensity and duration, is also important to consider. Indeed, the intensity of treatment for aphasia varies widely, with some treatments provided on a dense treatment schedule, for example, several hours a day. Meinzer and colleagues (2004, 2006) examined the effects of Constraint Induced Aphasia Therapy (CIAT), with treatment provided for three to four hours per day. Other treatments evaluated for their effects on brain function have been provided on less dense daily schedules (e.g., 15 minutes to 1 hour a day) (Leger et al. 2002; Raboyeau et al., 2008) or are provided for two to three days per week for one to two hours (Marcotte & Ansaldo, 2010; Thompson et al., 2010). Importantly, the effect of the intensity of aphasia treatment is still not clear, even though it may critically impact treatment efficacy, including how the brain recovers language. Therefore, we cannot make specific recommendations for treatment intensity here. Researchers, however, need to specify how frequently treatment is applied and, ideally, justify the choice of treatment dosage within the context of the presumed mechanisms targeted in treatment.

Another issue is the duration of treatment, which may also directly impact brain function. Some researchers provide treatment for a predetermined period of time, which varies across studies. For example, in studies by Fridriksson et al. (2006), Raboyeau et al., (2008), and Leger et al (2002) naming treatment was applied for two, four, and six weeks, respectively. This approach can be problematic because all participants may not respond equally well to treatment in the specified time period, and the neural recruitment patterns may vary because of variation in the degree to which the language behavior or process under treatment recovered. As a case in point,Vitali et al. (2007) showed differential learning (re-learning) patterns in their two participants with aphasia, with one participant reaching 50% naming accuracy on a set of trained items with four weeks of treatment and the other requiring eight weeks to achieve this level of performance. An alternative to setting an *a priori* treatment duration is to impose a behavioral criteria for termination of treatment. For instance,Thompson et al. (2010), interested in examining brain function associated with improved sentence comprehension and production in aphasia, provided treatment until their participants achieved an 80% accuracy level (with the idea that treatment would be terminated if this criterion were not met within 20 treatment sessions). Similarly, Marcotte & Ansaldo (2010) examined adaptive brain plasticity in two anomia cases, by describing the neural changes associated with a minimum of an 80% success rate following semantic feature therapy, delivered at a frequency of 3 weekly one-hour sessions, for a maximum of 3 weeks. In another study, Meinzer et al., (2006) terminated treatment after 10 consecutive treatment sessions and the patient was trained in the context of the constrained induced aphasia therapy protocol. Clearly, there may be reasons for selecting one approach versus the other (i.e., a predetermined temporal or behavioral criterion-based approach (or a combination of the two)). However, we emphasize here the need to provide a rationale for the approach taken and how this may influence reorganization of the language network. Furthermore, regardless of approach, in order to ascribe changes in neural activation resulting from treatment application, it is necessary to distinguish between responders and non-responders to treatment.

### 3.3. Reliability

One other important aspect of treatment research concerns the reliability of measurement of both the independent and dependent variables included in the experiment. Treatment

research largely involves observation of human behavior and inherent in human observation is human error, as well as observer bias. Although it is difficult, if not impossible, to overcome human error, estimates of reliability can be made using measures of inter-observer agreement. Such measures involve use of an independent observer who, together with the primary experimenter, scores important events in the study, including details pertaining to delivery of treatment (reliability on the independent variable) and responses made on the dependent measures (reliability on the dependent variable). With regard to the independent variable, the observer quantifies salient aspects of treatment, for example, the number of experimental trials delivered per treatment session, adherence to procedural detail within trials, and so on. Reliability on the dependent variable involves scoring of participant responses on the probe task(s) based on a pre-established criterion. When the independent observers agree to a high degree, it is unlikely that human error or observer bias is operating, adding an element of believability to the data. Lack of agreement between observers alerts the experimenter to problems with the experiment, for example, imprecise operational definitions of the study variables, which if discovered early in the course of an experiment can be modified.

Reliability of performance during scan tasks also is important. Whereas reaction times are automatically recorded in tasks that require a button press response, tasks that require production require that at least a subset of responses be coded by independent observers, with subsequent calculation of inter-observer agreement.

Finally, test-retest reliability is a critical issue in treatment studies, since most of the studies that have been published to date have not had a control group or a group of patients scanned multiple times. It is important to establish that imaging changes are actually attributable to the intervention and not due to scanning a single patient or a group of patients twice. Very few studies have been conducted in which imaging studies have been done multiple times over the course of the study (Kurland, Pulvermuller, Silva, Burke, & Andrianopoulos, 2012; Sarasso et al., 2010). In one study,Sarasso et al. (2010), conducted six fMRI scans during the course of an intervention study, three sessions were conducted prior to the start of therapy, the fourth and fifth three and six weeks after the initiation of therapy and the fifth week and the sixth fMRI session conducted nine months after therapy. This study does not specifically examine habituation effects on signal intensity changes but finds changes in functional connectivity only in the fMRI scans subsequent to treatment and not before treatment.

## 4.0. Conclusion

To conclude, it is clear that the initial wave of exploratory studies examining the neural mechanisms associated with treatment-induced language recovery have been completed. As in any science, much of this early work was not subjected to rigorous scientific scrutiny, because the novelty of the findings outweighed the methodological shortcomings of the research. Nevertheless, the general finding derived from these studies is that changes in language performance are associated with functional changes in the neural architecture of language processing. The next phase of neuroimaging-based treatment studies using fMRI need to be carefully designed and implemented such that any changes in neural activation following a period of treatment can be directly associated with the treatment provided and not to other uncontrolled variables. For instance, future studies will need to consider what change in BOLD signal as a function of treatment may indicate. It may be possible that an increase in task-dependent BOLD is a sign of increased neural processing (i.e., more effort requires more BOLD signal), while others see therapy (or time-related) behavioral improvements associated with a decrease in BOLD. Clearly, these kinds of BOLD effects may differ between the two hemispheres and will need to be considered in the fMRI analyses as well as interpreted within the context of treatment effects.

We point out here that this can be accomplished using either single-subject/case series or group experimental designs, but urge that researchers, when implementing these approaches, adhere to the methodological requirements inherent in each. We also point out that in order to fully understand the impact of treatment on brain function, special attention to issues related to the treatment selected, behavioral and neuroimaging outcome variables and reliability of measurement must be considered.

Given that considerable effort is currently focused on examining resting-state and functional connectivity changes in stroke patients with aphasia to better understand mechanisms of language recovery, it is expected that principles underlying accurate and systematic examination of the effects of treatment will be the same as what is discussed in this paper. Indeed, the ultimate goal of this work is to understand the optimal conditions for promoting language recovery in aphasia. What we learn will only be as robust as our science.

## Acknowledgments

The authors wish to thank the School of Communication and the Vice President for Research at Northwestern University for providing funds to support the Neuroimaging in Aphasia Treatment Research Workshop held at Northwestern University, July 2009. Participants of the workshop, who contributed to the ideas in this paper include: Ana Inés Ansaldo, Roelien Bastiaanse, Pelagie Beeson, Stefano Cappa, David Caplan, Leora Cherney, David Copland, Jenny Crinion, Bruce Crosson, Dirk-Bart den Ouden, Susan Edwards, Evy Visch-Brink, Julius Fridriksson, Argye Hillis, Audrey Holland, Chien-Ju Hsu, Aneta Kielar, Monique King, Swathi Kiran, Jim Kloet, Sladjana Lukic, Marcus Meinzer, Charis Price, Ellyn Riley, Steven Small, Brenda Rapp, Dorothee Saur, Marion Smits, Yasmeen Faroqi-Shah, Cynthia Thompson, Eisha Wali.

This work was also supported in part by the National Institutes of Health, Institute on Deafness and Other Communication Disorders R01DC01948 and R01DC007213 (CT) K18 DC011517-01 (SK); and R21DC9876 (LRC); the Bundesministerium für Bildung und Forschung, BMBF: 01EO0801 (MM) and The Heart and Stroke Foundation of Canada, and Fonds de la recherche en Santé du Québec (AIA).

## References

Abutalebi J, Rosa PA, Tettamanti M, Green DW, Cappa SF. Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. Brain Lang. 2009; 109(2–3):141–156. [PubMed: 19427522]

Barlow, DH.; Hersen, M. Single case experimental designs. New York: Pergamon Press; 1984.

Bonakdarpour B, Parrish TB, Thompson CK. Hemodynamic response function in patients with stroke-induced aphasia: implications for fMRI data analysis. Neuroimage. 2007; 36(2):322–331. [PubMed: 17467297]

Collins A, Loftus E. A spreading activation theory of semantic processing. Psychological Review. 1975; 82(6):407–428.

Crinion JT, Leff AP. Recovery and treatment of aphasia after stroke: functional imaging studies. Curr Opin Neurol. 2007; 20(6):667–673. [PubMed: 17992087]

Crosson B, McGregor K, Gopinath KS, Conway TW, Benjamin M, Chang YL, et al. Functional MRI of language in aphasia: a review of the literature and the methodological challenges. Neuropsychol Rev. 2007; 17(2):157–177. [PubMed: 17525865]

Crosson B, Moore AB, Gopinath K, White KD, Wierenga CE, Gaiefsky ME, et al. Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. Journal of Cognitive Neuroscience. 2005; 17(3):392–406. [PubMed: 15814000]

Davis CH, Harrington G, Baynes K. Intensive semantic intervention in fluent aphasia: A pilot study with fMRI. Aphasiology. 2006; 20(1):59–83.

Fridriksson J, Richardson JD, Baker JM, Rorden C. Transcranial direct current stimulation improves naming reaction time in fluent aphasia: a double-blind, sham-controlled study. Stroke; a journal of cerebral circulation. 2011; 42(3):819–821.

Fridriksson J. Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. J Neurosci. 2010; 30(35):11558–11564. [PubMed: 20810877]

Fridriksson J, Morrow-Odom L, Moser D, Fridriksson A, Baylis G. Neural recruitment associated with anomia treatment in aphasia. Neuroimage. 2006; 32(3):1403–1412. [PubMed: 16766207]

Fridriksson J, Moser D, Bonilha L, Morrow-Odom KL, Shaw H, Fridriksson A, et al. Neural correlates of phonological and semantic-based anomia treatment in aphasia. Neuropsychologia. 2007; 45(8): 1812–1822. [PubMed: 17292928]

Kazdin, A. Single Case Research Designs: Methods for clinical and applied settings. New York: Oxford University Press; 1982.

Kearns KP. Flexibility of single-subject experimental designs. Part II: Desng selection and arrangement of experimental phases. Journal of Speech and Hearing Disorders. 1986; 51:204–214. [PubMed: 3525987]

Kiran, S.; Sebastian, R.; Chettiar, P.; Devous, M. Neural correlates of lexical semantic recovery after treatment in aphasia. Paper presented at the Human Brain Mapping; 2008.

Kurland J, Naeser MA, Baker EH, Doron K, Martin PI, Seekins HE, et al. Test-retest reliability of fMRI during nonverbal semantic decisions in moderate-severe nonfluent aphasia patients. Behav Neurol. 2004; 15(3–4):87–97. [PubMed: 15706052]

Kurland J, Pulvermuller F, Silva N, Burke K, Andrianopoulos M. Constrained vs. Unconstrained Intensive Language Therapy in Two Individuals with Chronic, Moderate-to-Severe Aphasia and Apraxia of Speech: Behavioral and fMRI Outcomes. American Journal of Speech-Language Pathology. 2012 1058-0360_2012_1011-0113.

Leger A, Demonet JF, Ruff S, Aithamon B, Touyeras B, Puel M, et al. Neural substrates of spoken language rehabilitation in an aphasic patient: an fMRI study. Neuroimage. 2002; 17(1):174–183. [PubMed: 12482075]

Marcotte K, Ansaldo AI. The neural correlates of Semantic Feature Analysis in Broca's aphasia: Discordant patterns according to etiology. Seminars in Speech and Language. Special Issue on Treatment of Chronic Aphasia: International Perspectives. 2010; 31(No.1):52–63.

McReynolds, LV.; Kearns, KP. Single-subject experimental designs in communicative disorders. Austin, TX: Pro-Ed.; 1983.

McReynolds LV, Thompson CK. Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject design. Journal of Speech and Hearing Disorders. 1986; 51:194–203. [PubMed: 3525986]

Meinzer M, Harnish S, Conway T, Crosson B. Recent developments in functional and structural imaging of aphasia recovery after stroke'. Aphasiology. 2011; 25(3):271–290. [PubMed: 21532927]

Meinzer M, Elbert T, Wienbruch C, Djundja D, Barthel G, Rockstroh B. Intensive language training enhances brain plasticity in chronic aphasia. BMC Biol. 2004; 2:20. [PubMed: 15331014]

Meinzer M, Flaisch T, Obleser J, Assadollahi R, Djundja D, Barthel G, et al. Brain regions essential for improved lexical access in an aged aphasic patient: a case report. BMC Neurol. 2006; 6:28. [PubMed: 16916464]

Menke R, Meinzer M, Kugel H, Deppe M, Baumgartner A, Schiffbauer H, et al. Imaging short- and long-term training success in chronic aphasia. BMC Neurosci. 2009; 10:118. [PubMed: 19772660]

Peck KK, Moore AB, Crosson BA, Gaiefsky M, Gopinath KS, White K, et al. Functional magnetic resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. Stroke. 2004; 35(2):554–559. [PubMed: 14739418]

Raboyeau G, De Boissezon X, Marie N, Balduyck S, Puel M, Bezy C, et al. Right hemisphere activation in recovery from aphasia: lesion effect or function recruitment? Neurology. 2008; 70(4): 290–298. [PubMed: 18209203]

Richter M, Miltner WH, Straube T. Association between therapy outcome and right-hemispheric activation in chronic aphasia. Brain : a journal of neurology. 2008; 131(Pt 5):1391–1401. [PubMed: 18349055]

Rorden C, Fridriksson J, Karnath HO. An evaluation of traditional and novel tools for lesion behavior mapping. Neuroimage. 2009; 44(4):1355–1362. [PubMed: 18950719]

Sarasso S, Santhanam P, Määtta S, Poryazova R, Ferrarelli F, Tononi G, Small SL. Non-fluent aphasia and neural reorganization after speech therapy: insights from human sleep electrophysiology and

functional magnetic resonance imaging. Archives Italiennes de Biologie. 2010; 148(3):271–278. [PubMed: 21175013]

Tate R, McDonald S, Perdices M, Togher L, Schultz R, Savageet S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. Neuropsychological Rehabilitation. 2008; 18(4):385–401. [PubMed: 18576270]

Thompson CK. Single subject controlled experiments in aphasia: the science and the state of the science. J Commun Disord. 2006; 39(4):266–291. [PubMed: 16635494]

Thompson CK, den Ouden DB. Neuroimaging and recovery of language in aphasia. Curr Neurol Neurosci Rep. 2008; 8(6):475–483. [PubMed: 18957184]

Thompson CK, den Ouden DB, Bonakdarpour B, Garibaldi K, Parrish TB. Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. Neuropsychologia. 2010; 48(11):3211–3227. [PubMed: 20603138]

Vitali P, Abutalebi J, Tettamanti M, Danna M, Ansaldo AI, Perani D, et al. Training-induced brain remapping in chronic aphasia: a pilot study. Neurorehabil Neural Repair. 2007; 21(2):152–160. [PubMed: 17312090]