NeuroImage xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Review

Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis

Marcus Meinzer [a,b,*], Pélagie M. Beeson [c], Stefano Cappa [d], Jenny Crinion [e], Swathi Kiran [f], Dorothee Saur [g], Todd Parrish [h], Bruce Crosson [i,1], Cynthia K. Thompson [j,1]

[a] Charite Universitätsmedizin, Department of Neurology, Center for Stroke Research Berlin & Cluster of Excellence NeuroCure, Berlin, Germany
[b] University of Florida, Department of Clinical and Health Psychology, Gainesville, Florida, USA
[c] University of Arizona, Department of Speech, Language and Hearing Sciences, Tucson, Arizona, USA
[d] Vita-Salute San Raffaele University, Department of Clinical Neurosciences, San Raffaele Turro, Milano, Italy
[e] University College London, Institute of Cognitive Neuroscience, London, UK
[f] Boston University, Sargent College of Health and Rehabilitation Sciences, Boston, Massachusetts, USA
[g] University of Leipzig, Department of Neurology, Language and Aphasia Laboratory, Leipzig, Germany
[h] Northwestern University, Department of Radiology, Chicago, Illinois, USA
[i] Department of Veterans Affairs Brain Rehabilitation Research Center of Excellence and University of Florida Department of Clinical & Health Psychology, Gainesville, Florida, USA
[j] Northwestern University, Departments of Communication Sciences and Disorders and Neurology, Evanston, Illinois, USA

ARTICLE INFO

Article history:
Accepted 19 February 2012
Available online xxxx

Keywords:
Structural anatomy
Language recovery
Stroke
Aphasia
Brain lesions

ABSTRACT

Functional magnetic resonance imaging is the most widely used imaging technique to study treatment-induced recovery in post-stroke aphasia. The longitudinal design of such studies adds to the challenges researchers face when studying patient populations with brain damage in cross-sectional settings. The present review focuses on issues specifically relevant to neuroimaging data analysis in aphasia treatment research identified in discussions among international researchers at the Neuroimaging in Aphasia Treatment Research Workshop held at Northwestern University (Evanston, Illinois, USA). In particular, we aim to provide the reader with a critical review of unique problems related to the pre-processing, statistical modeling and interpretation of such data sets. Despite the fact that data analysis procedures critically depend on specific design features of a given study, we aim to discuss and communicate a basic set of practical guidelines that should be applicable to a wide range of studies and useful as a reference for researchers pursuing this line of research.

Contents

* Corresponding author at: Charite Universitätsmedizin Berlin, CCM, Neurologie, Chariteplatz 1, 10117 Berlin, Germany.
E-mail address: marcus.meinzer@charite.de (M. Meinzer).
[1] Shared senior authorship.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.02.058

Introduction

Functional magnetic resonance imaging (fMRI) is one of the most widely used imaging techniques to study intact and impaired cognitive functions (Crosson et al., 2010). In particular, an increasing number of studies have used fMRI to assess functional brain activity changes in response to treatment of acquired aphasia (for recent reviews see Meinzer et al., 2011; Thompson and den Ouden, 2008; Meinzer and Breitenstein, 2008; Crinion and Leff, 2007). Although aphasia can be caused by various types of brain damage, it is most frequently observed after middle cerebral artery stroke in the left hemisphere (Nicholas, 2005). The analysis and interpretation of functional imaging data in brain damaged populations poses several challenges to the researcher, relating to a number of factors, including the presence of a lesion and possible hemodynamic changes due to vascular pathophysiology.

Compared to cross-sectional studies (e.g., examining individuals with aphasia at different stages of recovery), neuroimaging studies of treatment effects present additional challenges. They aim to assess treatment-induced plasticity of neural functions in a longitudinal design, typically involving repeated assessments in the same individuals (e.g., before and after treatment). Whereas some aspects of data processing are essentially the same as neuroimaging studies of healthy participants, several important differences with regard to the data processing and statistical analyses need to be considered when pursuing aphasia treatment research. Previous reviews have discussed general aspects of functional imaging in brain damaged populations (e.g., Crosson et al., 2010; Price et al., 2006) or have focused on specific language domains in aphasia (e.g., Crosson et al., 2007 for a review of imaging language production mechanisms). The present paper focuses on issues specifically relevant to assessing treatment-induced plasticity and conveys the consensus regarding critical aspects of data processing that was reached during the Neuroimaging in Aphasia Treatment Research Workshop, held at Northwestern University.

In this paper, we aim to provide the reader with a review of critical issues regarding data analysis in functional neuroimaging of aphasia treatment and provide guidelines regarding how to deal with these issues. We acknowledge that data analysis procedures depend on the goals and specific design features of a given study (e.g., experimental design, type of treatment, language modality assessed), so the recommendations are intended to have broad application. Although data analysis also includes the reporting of these procedures, this will not be the main focus of this manuscript, as general guidelines for describing methodological aspects of fMRI studies have been elaborated elsewhere (see, for example, Poldrack et al., 2008). However, because data analysis in brain-damaged individuals might differ substantially compared to that in healthy participants, we make recommendations for reporting specific procedures where necessary.

In summary, we discuss a basic set of practical suggestions for analyzing datasets collected to assess treatment-induced plasticity in aphasia patients. These guidelines are intended to provide a reference for researchers pursuing this line of research.

Processing of MRI data sets

Functional MRI datasets require several pre-processing steps that are implemented in similar ways in available and commonly used data analysis packages (e.g., Statistical Parametric Mapping, SPM, http://www.fil.ion.ucl.ac.uk/spm; Analysis of Functional Neuroimages, AFNI, Cox, 1996; Brain Voyager©). These include (1) realignment of the images of an fMRI time series to compensate for head movement during scanning and correction for slice timing differences, (2) co-registration of the functional data to a high-resolution structural image, (3) spatial normalization to account for inter-individual variations in brain size and anatomy, and (4) spatial smoothing of the data to increase statistical power for group analysis. Following pre-processing, a statistical model is designed to estimate neural activity in a single patient or within and between groups of study participants.

Realignment, slice timing correction, and co-registration procedures are relatively unaffected by the presence of brain damage. Thus, early pre-processing of lesioned and normal brains is essentially the same. However, spatial normalization, smoothing and, most importantly, aspects of statistical modeling of the data (model-specification, statistical inferences and interpretation) vary depending on whether healthy or aphasia participants' data are analyzed. Moreover, there are several important differences regarding the analysis of cross-sectional or longitudinal data in aphasia research. Thus, we will discuss them in more detail below.

Pre-processing fMRI data

Spatial normalization (between participant and/or session realignment)

There are basically two different ways to proceed after image realignment and co-registration. One is to statistically analyze the respective dataset in native space, which assures a valid relationship between an individual participant's anatomy and his/her activation. This approach, however, is limited to extraction of data from individual participants and cannot be used when the objective is to generate a group image reflecting statistical analysis of mean differences between groups or across sessions. An alternative way to proceed is to spatially normalize the functional imaging data. This is a necessary step for studies that rely on voxelwise comparisons, such as contrasts between healthy and brain- damaged groups, comparison of activity patterns of individual participants to a reference population, or correlation of treatment outcome with changes in activity patterns. Moreover, it is necessary to report activity patterns in standard coordinates (e.g., Talairach or Montreal Neurological Institute space (Mazziotta et al., 1995; Talairach and Tournoux, 1988)), which facilitates comparison of signal location with other published studies. The quality of the normalization has been shown to affect activity patterns in group studies of healthy subjects; i.e., inaccurate normalization leads to reduced sensitivity to detect functional activity (Ardekani et al., 2004). This is further complicated in brain-damaged participant groups where automated warping algorithms, as implemented in standard neuroimaging analysis platforms, may produce inappropriate solutions because of the presence of lesioned tissue, leading to inaccurately localized activation (Price et al., 2006). Similarly, misalignment of images in aphasic groups may result in falsely detected activity compared to a control group. Thus, precise and valid normalization is critical to understand the neural substrates of treatment-induced recovery.

Before proceeding further with the discussion of spatial normalization, caveats regarding such procedures in individuals with stroke

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

must be mentioned. First, group studies that analyze mean changes from one session to the next can obscure perilesional activity, especially when study samples have diverse lesion patterns (Crosson et al., 2007). Because the contribution of perilesional activity to recovery of function as a result of treatment frequently is an important issue, this kind of session-to-session comparison is problematic in studies of individuals with diverse lesion patterns, unless perilesional activity is not of interest. There are valid uses of group images in aphasia treatment research, and ways of addressing this problem are discussed later in this paper. Second, we use SPM as an example of how problems in registration of images for stroke patients have been addressed. This is not meant to imply that other programs have not addressed these problems or that different, equally effective solutions to those available in SPM are not possible.

Methodological advances in neuroimaging (Bandettini, 2009) have increased our ability to combine brain images from different brain-damaged participants into a common anatomical space and to analyze thousands of regions simultaneously (Godefroy et al., 1998; Rorden et al., 2007). There are many different methods of spatial normalization (also referred to as registration), some automated and some manual, with global or local warping to a given atlas (Crinion et al., 2007; Godefroy et al., 1998; Rorden and Brett, 2000; Seghier et al, 2008). All have strengths and weaknesses that should be recognized. Ideally, if it is decided that scans will be normalized, the deformation error should be quantified, for example, using forward and backward registration between each participant's scan and atlas or template space. Some methods may benefit from normalizing (registering) scans to an age-appropriate atlas to further minimize potential image registration error.

Automated normalization algorithms often use differences in intensity values between a given image in native space and a template to calculate the spatial transformation parameters that minimize the mismatch between the two images (Friston et al., 1995). This usually involves both linear (affine) and nonlinear distortions of the original image. Linear transformations apply uniform warps across the entire image to match the overall shape and orientation of the template. However, linear algorithms restrict the fitting of local anatomy (e.g., sulcal structure and size). Therefore, subsequent non-linear transformations that are concerned with local shape are required. Problems with automated normalization procedures arise when there are areas of large signal change, such as those reflecting the presence of a structural lesion. Although affine transformations are relatively robust to lesion effects, the quality of non-linear transformations are disproportionately affected. That is, in an attempt to reduce image mismatch introduced by the structural damage, the algorithm may over-fit the original image, distorting intact tissue and reducing the size of the lesion in the normalized images (Brett et al., 2001).

Hence, affine-only solutions cannot be recommended for lesioned brains, as they compromise the fitting of local anatomy, which is especially important in individuals with stroke-induced lesions that produce enlarged ventricles or local atrophy. Thus, there have been attempts to restrict the normalization to undamaged parts of the brain by masking the lesion (e.g., cost-function masking; Brett et al., 2001), thus minimizing the impact of the lesion on the non-linear component of the normalization of the remaining image. Masking the lesioned area does not mean these areas are not normalized, but rather that there is a continuation of the normalization solution for the remaining brain to the lesioned area. With regard to structural images, cost-function masking has been shown to be superior to affine-only solutions (Brett et al., 2001) and is considered the gold standard. The major limitation for aphasia studies is that cost-function masking is limited to individuals with unilateral pathology, because normalization of the area under the mask largely depends on intact homologous areas. In addition, cost function masking might be compromised by a lack of symmetry between brain structures (Binder et al., 1996) and masking of lesioned brain areas involves an operator dependent and laborious manual definition of the lesion boundary. We note, however, that this latter problem might be accounted for by applying more recent automated lesion identification procedures (see, for example, Seghier et al., 2008).

With regard to functional imaging, normalization of a high resolution anatomical image provides transformation parameters that are applied to the co-registered functional images. However, the impact of different normalization procedures on functional activity has only been formally assessed for one of the major functional imaging analysis platforms so far (SPM5). Crinion et al. (2007) compared the performance of three different normalization procedures (affine only, standard SPM normalization, and unified normalization as implemented in SPM5 and above) with and without cost-function masking. Compared to previous SPM normalization procedures (see Crinion et al., 2007 and Ashburner and Friston, 2005 for a comprehensive description of both procedures), unified normalization comprises segmentation (i.e., tissue classification as grey and white matter and cerebrospinal fluid), bias correction (modeling of tissue non-homogeneities, which in turn allows modeling of healthy and lesioned tissue separately within one tissue class), and spatial normalization in a single iterative model. In particular, the bias correction may act like an implicit cost-function mask (e.g., the effects of lesioned white matter should not affect the normalization of intact white matter that is modeled separately; see Ashburner and Friston, 2005 for details).

Performance of different normalization procedures and their impact on functional data were assessed in three experiments establishing the anatomical validity of each respective normalization procedure using anatomical landmarks (i.e., co-localization of anatomical landmarks across images) in intact brains and intact brains with simulated lesions. In addition, the impact on functional activity was assessed by using a previously published dataset of stroke patients obtained during an auditory speech comprehension paradigm (Crinion and Price, 2005). The main results of the study were that unified models (1) produced the best results in terms of anatomical co-localization and (2) resulted in greater sensitivity for functional activity. While cost-function masking improved the quality of the standard normalization, it did not further improve the quality of the unified solution.

In a subsequent study, Andersen et al. (2010) found that cost-function-masking used with the unified solution produced greater normalization accuracy of high-resolution structural scans in chronic stroke patients with relatively large lesions and secondary changes in brain morphology (e.g., dilation of ventricles). Moreover, no difference in normalization accuracy was found between different types of masks (precise, roughly-outlined, smoothed, or unsmoothed), indicating that even a time-efficient rough outlining of the lesioned area appears to improve normalization quality significantly compared to unified normalization without masking in such patients. However, it is worth noting that the localization errors introduced by both methods (i.e., unified normalization with or without cost-function masking) were significantly smaller than the typical smoothing kernels (6–8 mm) used in fMRI studies. Thus, the impact of these errors is not large enough to significantly affect group fMRI studies in patients with brain damage.

Taken together, these findings suggest that, when using SPM5 or later versions, unified normalization is recommended over other approaches when morphological changes are restricted to the region of primary pathology. When this normalization approach fails, additional cost-function-masking is advised, especially in patients with additional secondary changes as a consequence of large lesions. In the context of other imaging analysis platforms where the impact of different manual or automated normalization procedures on functional activity patterns has not been formally assessed, cost-function masking may be considered an appropriate solution to minimize inaccurate normalizations due to lesion effects. Moreover, quality assurance procedures should be established and reported (e.g., comparing the results against the image in native space to detect distortions of the

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

normalized image). This is particularly indicated in longitudinal group studies on aphasia rehabilitation that require not only good within-subject session-to-session registration of the images, but also good between-subject registration to enhance the sensitivity of the analysis.

Detrending for signal drift and spatial smooting

Detrending fMRI series for linear or low frequency drifts in signal baseline and spatial smoothing of image data are two procedures that can be performed prior to data analysis. Regarding the former, linear and low frequency drifts in baseline signal levels are common in fMRI data, and they reduce sensitivity to task-related activation. Detrending algorithms to remove the signal drift from the data vary in effectiveness. Tanabe et al. (2002) found that linear and quadratic detrending increased sensitivity to task-related activation in individual fMRI images, whereas cubic detrending actually decreased sensitivity to activation changes. Although a spline detrending method was superior to linear or quadratic detrending, an automatic method selecting the optimal detrending method on a voxel-by-voxel basis was superior to all other methods, increasing sensitivity to activity changes by 150%. These results indicate that even within a single data set, different kinds of baseline signal drift can be present and further suggest that a method that can select the best detrending method on a voxel-by-voxel basis is preferred.

The purpose of spatial smoothing is to increase signal to noise in the time series data by reducing random noise, to account for inter-subject variability in functional and structural anatomy by blurring the spatial details of the functional maps (i.e., increase statistical power in group studies), to allow for parametric statistical testing, and to assure that the data conform to the lattice assumption of Gaussian random field theory (Turner et al., 1998). Spatial smoothing is usually achieved by convolving the data with a Gaussian kernel of a given size that is determined by the voxel size and the size of the anticipated signal change (Hopfinger et al., 2000; Price et al., 2006). Drawbacks of spatial smoothing are that spatial resolution decreases, and blurring or shifting of activation may result in merging of adjacent clusters of activation. Hence, accuracy of localization may be compromised, which is most critical for single subject studies. Typical smoothing kernels in fMRI group studies involving healthy participants are usually two times or more the re-sampled voxel size. Choosing an optimal smoothing filter is not trivial as it may significantly affect the results of a given study. However, the impact on functional activity in brain-damaged populations has not been thoroughly evaluated. Although the degree of spatial smoothing clearly depends on study design and should be determined empirically, studies in healthy individuals can be used as a starting point to determine the optimal filter width for studies with neurologically impaired individuals (e.g., Hopfinger et al., 2000; Mikl et al., 2008).

The type of study and the number of participants are critical to determining the degree of spatial smoothing. In single participant studies precise and valid localization of focal activation is crucial. Studies in healthy participants have shown that larger smoothing filters (e.g., >10 mm) may induce shifting of local maxima up to 12 millimeters (e.g., Geissler et al., 2005; Mikl et al., 2008). Therefore, no smoothing or only low spatial kernels (e.g., not larger as twice the largest acquired voxel dimensions) should be used in studies examining activation in individual participants to assure accurate localization. This is particularly critical when peak activity is located in sulcal walls, and even minor shifts of activity may result in gross mislocalization relative to the cortical surface (e.g., on the opposite bank of the sulcus) (Brett et al., 2002). This latter problem might be of lesser concern when using cortical surface mapping techniques (e.g., Van Essen, 2004), but this feature is currently not implemented in most imaging analysis platforms (SPM, AFNI). For studies using group voxel-by-voxel images, between-subject variability in anatomy, functional activity and registration quality need to be taken into account, which may require increasing filter widths. In this context, extensive spatial smoothing may be indicated for individuals with brain damage compared to healthy individuals due to greater between-subject dispersion in the location of local structures (c.f., Price et al., 2006), increased between-subject registration errors, or overlapping lesion borders that occur even in highly homogeneous samples. Group size may also be an important factor. As a general rule, in healthy participants it has been suggested that with larger numbers (16+) a similar smoothing factor can be chosen compared to single subject studies; however, the smoothing kernal should be larger when examining data from smaller sample sizes in order to account for outlier effects in the spatial localization of activation patterns (Mikl et al., 2008). Individual factors like quality of the data (e.g., signal-to-noise ratio and quality of inter-subject registration) also need to be taken in account. The degree of smoothing may also depend on the anatomical region-of-interest (e.g., cortical vs. subcortical) and the correction level used. This was demonstrated in healthy participants by Hopfinger et al. (2000), who showed that smaller smoothing filters increased sensitivity in cortical regions, whereas larger filters increased sensitivity in subcortical areas.

In summary, studies that use group voxel-by-voxel analyses in healthy participants have suggested that the optimal degree of spatial smoothing is critical to the outcome of the study and that this depends on several factors. Some of these factors are influenced by the aims and design of the study (e.g., single-subject vs. group designs, anatomical region of interest), which need to be determined empirically (e.g., anticipated signal change and signal to noise). Quality assurance procedures (e.g., intersubject registration quality and variability of functional activity peaks, and peak activity of unsmoothed activity in native space) may help to assure that the optimal extent of spatial smoothing is chosen.

Specific problems related to motion artifacts during overt speech

The issue of motion artifacts when assessing overt speech production has been addressed in a recent review by Crosson et al. (2007). We discuss this issue briefly here because word-retrieval impairments are one of the most frequent symptoms of aphasia (Kohn and Goodglass, 1985) and most fMRI studies to date that have examined treatment-induced recovery of language functions, have used overt naming or other language production paradigm to evaluate treatment effects (e.g., picture naming, category generation; see Meinzer and Breitenstein, 2008, and Thompson and den Ouden, 2008, for review). Further, although covert paradigms have been shown to reliably elicit activation in language related brain areas in healthy participant (see Kielar et al., 2011), the lack of behavioral control limits their usefulness in aphasia treatment studies. That is, response accuracy and reaction time often are important to fully characterize aphasia recovery, and these data are not available using covert neuroimaging tasks.

There are several ways to deal with motion artifacts (which are predominantly false positive activity) during overt generation. At the design level, for example, motion-related artifacts can be avoided by using blocked designs and dropping images confounded by evidence of motion (e.g., Martin et al., 2005) or by using sparse acquisition paradigms that acquire the BOLD response after overt articulation (Fridriksson, 2010; Meinzer et al., 2008). However, these strategies are associated with a loss of information and reduced flexibility compared to event-related paradigms. Moreover, optimizing of presentation parameters has been shown to reduce motion-related artifacts when using ideal waveforms for analysis. These designs exploit the different temporal properties of (rapid) motion induced signal changes compared to more slowly evolving changes of the task-related hemodynamic signal (e.g., Birn et al., 2004). However, even when using such designs, overt articulation may still result in false positive activity and standard detrending algorithms that aim to remove motion-related signal from the time series non-selectively across all voxels, may result in reduced sensitivity (see Crosson et al., 2007 for details). More recently developed detrending algorithms consider the latter weakness by

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

selectively removing motion related signal changes from the images, which results in improved sensitivity and specificity (Gopinath et al, 2009).

Statistical model specification

The next data analysis step involves setting up a statistical model to estimate task-related activation. Most fMRI data are analyzed in the context of the General Linear Model (GLM). The first step here is setting up a design matrix consisting of factors that potentially contribute to the actual fMRI signal (experimental conditions and non-experimental sources of variability like head movements). The structure of the design matrix and factors included depend on the hypotheses to be tested. Extracting the respective task-related signal can be accomplished in different ways (see below) and treatment-induced plasticity in individuals with aphasia can be assessed in individual participants (first-level analysis) or in groups (second-level analysis). Both approaches pose several challenges with regard to statistical model specification and will be addressed in the following. For discussion of more general issues of design and statistical analysis of fMRI data in individuals with brain damage see Price et al. (2006).

Modeling the hemodynamic response

Extraction of task-related signal at the individual subject level can be accomplished with either constrained or unconstrained methodologies. Constrained methods rely on a predetermined model of the hemodynamic response form, whereas unconstrained methods make no assumptions about the shape of the response. Each methodology has its strengths and weaknesses. As examples of constrained and unconstrained methodologies, we discuss below use of a standard model of the hemodynamic response function (HRF) and deconvolution of the HRF with unconstrained modeling, respectively.

Many studies with healthy participants model blood oxygen-level dependent (BOLD) activity by using a standard HRF. To account for individual differences in shape or timing of the HRF, the design model can include additional factors (e.g., temporal or spatial derivatives) or the HRF can be modeled by using other types of basis functions (e.g., the gamma function). A different approach that uses participant-specific HRFs may substantially improve the model fit (Aguirre et al., 1998). Modeling the HRF poses a challenge in people who have suffered a stroke because neurovascular reactivity in perilesional, or even distal, brain areas may be compromised due to microvascular impairment. In fact, despite intact neural functioning in some brain-damaged individuals, no positive BOLD signals, reduced positive BOLD signals, or even negative signals may be observed (e.g., Bonakdarpour et al., 2007; Fridriksson et al., 2006a, 2006b; Murata et al., 2006; Rossini et al., 2004; Röther et al., 2002). It has also been shown that the shape or the timing of the HRF can be compromised, even in individual with chronic stroke-induced aphasia (Bonakdarpour et al., 2007; Peck et al., 2004). For example, Peck et al. (2004) investigated the temporal characteristics of the BOLD response in three individuals with chronic Broca's aphasia during a category generation task. Functional MRI revealed prolonged HRFs and longer time to peak (TTP) in right hemisphere regions of interest (ROIs) in two of the patients with impaired behavioral functioning when compared to healthy controls. Bonakdarpour et al. (2007) found similar abnormal HRFs in three of five chronic aphasic individuals and showed that, when adjusted for their true HRF, patients with delayed TTP showed activation (particularly in perilesional tissue) which was not apparent when a canonical HRF was used. Importantly, in a study, which examined activation associated with treatment-induced language recovery in aphasic individuals, Thompson et al. (2010a, 2010b) found that regions of the brain in which upregulation of neural activity was found correlated with the HRF TTP. That is, regions of the brain with more normal (i.e., faster TTP) were more likely to demonstrate treatment-induced recovery of language processing.

Several strategies are available to address the potential problems related to modeling the hemodynamic response function in stroke patients. First, some baseline individual's cerebral ischemic condition can be assessed in addition to BOLD imaging, as even in chronic stroke patients misery perfusion has been shown to be related to poor BOLD signal (Murata et al., 2006). An alternative strategy, recently described by van Oers et al. (2010), is to assess intact hemodynamic responsiveness by using a breath-hold paradigm. However, this strategy may be contraindicated in individuals with stroke because it could result in ischemia in areas with reduced hemodynamic reserve (Hillis A.E., personal communication). Second, with regard to the altered shape and timing of the HRF several different alternatives are conceivable. As in research with healthy individuals, several studies have successfully used standard canonical HRFs or other types of basis functions (e.g., gamma function) with the first (temporal) derivative to account for increased variability in stroke patients (e.g. Crinion and Price, 2005). A third strategy is to collect information about study participant's hemodynamic parameters using an event-related design and a long inter-stimulus interval. This method allows a true HRF to be identified for each participant, which can then be used to optimize modeling of each participant's fMRI data on an individual basis and, thereby, improve BOLD signal detection (Bonakdarpour et al., 2007; Thompson et al., 2010a). Also, TTP measures may be of interest to characterize treatment-induced improvement or rehabilitation status. For example, in the aforementioned study by Peck et al. (2004) TTP was delayed prior to a language intervention and decreased (i.e., became similar to that of a control group) after treatment. Fourth, successful detection of BOLD activity using model-driven frameworks such as the GLM is only optimal if the underlying modeling assumptions are correct (e.g., regarding the timing and shape of the HRF, noise characteristics, etc.), which may be more difficult to achieve in some brain-damaged individuals, such as those with known perfusion deficits due to carotid stenosis. Hence, the weakness of constrained modeling of hemodynamic responses is that it may miss important characteristics of hemodynamic responses if they are not anticipated. Research with stroke individuals is particularly vulnerable to this problem where the shape and timing of hemodynamic responses are known to be variable.

On the other hand, unconstrained data-driven techniques may offer an alternative means to assess relevant fluctuations in the measured signal. For example, there are different techniques that can be used to deconvolve HRFs, some of which are entirely data driven with respect to HRF shape, such as the earliest form of deconvolution implemented in AFNI (Cox, 2009). The advantage of this technique is that it can accommodate changes in hemodynamic response shape from voxel to voxel even within individual participants (See Glover, 1999 and Serences, 2004 for more details about deconvolution techniques). The disadvantage to this approach is that it is very sensitive to noise such as that generated during overt speech during scanning. In the latter instance it is necessary to have a technique for minimizing noise in the data (e.g., Gopinath et al., 2009), and such techniques are not perfect. Averaging of raw signals from response epochs timed to experimental manipulations is also assumption free, but is subject to distortions, for instance when there are sequential dependencies of HRFs (Serences, 2004). Independent component analysis (ICA, McKeown and Sejnowski, 1998) is another assumption-free method. ICA separates the signal into maximally independent spatiotemporal components and does not impose any constraints on the HRF and thus, the results are data driven. However, it is critical to have a method to determine which components are signal and which represent noise.

Temporal characteristics of the responses

Regardless of the language task or paradigm used (e.g., picture naming, sentence-picture matching, lexical decision by button

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

press), response latency often is delayed in aphasic individuals and must, therefore, be considered when analyzing fMRI data in these individuals (see Crosson et al., 2007 for a more detailed discussion of this issue and examples). Timing the analysis to the presentation of the stimulus (stimulus-locked analysis) or to the response itself (response-locked analysis) can make a difference in the location of significant brain activity. Processes more closely linked to the presentation of the stimulus (e.g., perceiving or comprehending the stimulus) would be favored in a stimulus-locked analysis. Whereas, processes more closely linked to the response (e.g., articulation of a given word or sentence) would be favored by a response-locked analysis. This consideration becomes more important with longer latency periods between the stimulus and the response, because the activation patterns become more variable. Thus, choosing the stimulus-based, response-based, or a combination of both analyses depends on which cognitive processes are targeted. Moreover, given the highly variable performance often seen in individuals with aphasia, it may not always be clear which type of analysis to chose, because the cognitive processes may be variably linked to the stimulus or the response onset (e.g., word-retrieval processes).

There is no easy solution to this problem and although there is a clear distinction between stimulus onset and response onset in some tasks (lexical decision, overt word production), others only allow a stimulus locked analysis (e.g., covert reading). A recent example of how to deal with such issues has been introduced by Crosson et al. (2009), who analyzed fMRI data obtained during an overt category generation task in five individuals with chronic aphasia and combined both types of analyses (stimulus- and response-locked) to account for delayed and variable responses in the participants. Because this type of analysis allows a better model fit for each voxel of the image, it may also account for different time-courses across different brain areas.

Statistical inferences and interpretation

Comparing activation across scan sessions

Studying the natural history of recovery or treatment-induced plasticity in aphasia requires a longitudinal experimental design, which involves subjecting study participants to two or more neuroimaging sessions and comparing activation across sessions. Differences between assessments may stem from different sources (e.g., test-retest effects due to repeated task exposure, scanner related changes, plasticity related changes). Thus, special considerations relevant to the replication or reliability of activation is an issue of concern. Several suggestions for dealing with this issue are discussed in Rapp et al. (this volume), one of which is to conduct repeated baseline scans to ascertain any variability in activation (e.g., see Fridriksson et al., 2007). Another is to include tasks that reflect both impaired and unimpaired functions (Leger et al., 2002) or items that participants can and cannot respond to correctly (Menke et al., 2009). In any case, the issue is how best to compare activation changes between sessions. Several different measures can be obtained, including voxel-counting approaches and measures of activation magnitude such as percent signal change. Given that even in healthy participants the probability for single voxels to be consistently activated across scanning sessions is relatively low (see Meltzer et al., 2009), examining for activation in larger regions of interest (ROIs) can yield much better repeatability across sessions (e.g., Machielsen et al., 2000; Maldjian et al., 2000; Swallow et al., 2003; Wei et al., 2004). Moreover, it has been suggested that magnitude of signal change measures are much more consistent across repeated sessions than simple voxel counting approaches (for examples of such approaches see Voyvodic, 2006; Friedman et al., 2008; Kimberley et al., 2008; Meltzer et al., 2009; Voyvodic et al., 2009).

Choice of responses for modeling

The compromised language abilities of individuals with aphasia create challenges for neuroimaging studies of language recovery because aphasic individuals may have difficulty performing selected neuroimaging tasks, particularly prior to treatment. Although it is possible to design tasks that can be performed with high accuracy (see Rapp et al. for discussion of issues related to selection of tasks for neuroimaging studies of aphasic individuals), error responses are common. Because studies have demonstrated that correct and error responses may differ with regard to their neural signatures (e.g., Fridriksson et al., 2009; Meinzer et al., 2006; Postman-Caucheteux et al., 2010), this issue is not trivial. Thus, the question arises: which types of responses should be included in the analysis when assessing treatment-induced recovery? Typically, studies that assess the impact of treatment on brain functions imply that (a) a given language function is impaired prior to treatment and (b) treatment results in improvement in that language function, which is reflected in participants' performance ability and, in turn, changes in neural activation patterns seen from pre- to post-treatment. Therefore, inclusion of only correct responses in pre-post comparisons may prevent detection of meaningful changes. This strategy also putatively would require an analysis with differences in the number of responses between scans, which also could compromise the results. Analyzing both correct and incorrect responses, however, may also lead to spurious findings because error responses likely reflect increased processing demands (for a comprehensive review see Price et al., 2006) and also influence the timing of the HRF (Peck et al., 2004). However, it can be argued that regardless of whether responses are correct or incorrect, participants use whatever processing resources are available to them when performing a given linguistic tasks. Thus, changes in language ability will be reflected by brain activation changes from pre-treatment, for example when inefficient and incorrect linguistic processing is prevalent, to post-treatment when access to more normal linguistic processing routines becomes possible (see Thompson et al., 2010b, who take this position when examining treatment-induced recovery of complex sentence processing in aphasia). Further, for some paradigms and tasks it is difficult to quantify correct vs. erroneous responses, for example, for complex paradigms such as story comprehension when task performance may not reflect linguistic processing routines or abilities. Thus, decisions regarding which types of responses to analyze and how, or whether, different types of responses should be grouped into the same analysis depend mainly on the goals of the study and the experimental paradigm employed.

Statistical comparisons between sessions can be made by directly comparing different sessions in the same statistical model. Meltzer et al. (2009) have argued that this procedure assures that changes in the amount of noise do not produce misleading "changes" in activation. Data also can be extracted from two sessions separately for subsequent comparison, and then some procedure should be used to ensure that the detection sensitivity for activation is equivalent across sessions (Parrish et al., 2000). For example, Gopinath et al. (2009) developed a technique to compensate for differences in detection sensitivity of BOLD measures across sessions. In short, the technique starts with the residuals of the regression of a deconvolved hemodynamic response series against the acquired time series, and uses this as the starting point for a mixed auto-regressive plus white noise model to estimate noise structure for two sessions on a voxel-by-voxel basis. Once a time series representing the noise structure of both sessions is modeled, then detection sensitivity can be estimated by adding simulated hemodynamic responses of known amplitude to the estimated noise time series at appropriate points and equating detection sensitivity between sessions (see Gopinath et al., 2005 or Crosson et al., 2007 for details).

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

The issue of single subject vs. group studies

The question of whether to analyze neural activation changes from pre- to post-treatment in individual study participants or to analyze the data for groups of participants is another important issue. What is quite clear is that individual participants in a given experiment will differ with regard to the precise location and extent of their lesions, a situation that directly affects the neural tissue available to support recovery. That is, activation patterns will necessarily differ across participants. Nonetheless, it is important to integrate and generalize findings across individuals with the same language deficits or other variables, such as age, handedness, gender, or time post stroke. However, we see no reason to group individuals by classic aphasic syndromes, which are heterogeneous both functionally and neurologically.

Single case studies

Analysis of individual cases of aphasia can be advantageous compared to group studies in aphasia research. Data from individual participants can be analyzed in native space, which avoids localization errors introduced by inter-subject registration procedures and assures anatomical correspondence between individual participant's anatomy and his/her activation. This procedure also enhances the ability to visualize perilesional activation, which might not be detected even in highly homogeneous groups of aphasic participants (see below for a discussion of this problem with regard to treatment studies). On the other hand, there are important drawbacks to the individual participant approach. Namely, there is an inherent lack of power to detect activation changes, which can only be resolved at the design level (e.g., by increasing the number of trials). Further, the results from individual participants cannot easily be generalized to other individuals with aphasia. The latter problem, however, can be resolved by analyzing series of aphasic cases. The results can then be interpreted with regard to commonalities and differences between activation patterns found across participants, in the context of the other information that is available. For example, this may allow investigators to assess activation patterns associated with treatment outcome (see, Crosson et al., 2009), or compare activation patterns for participants who respond well vs. poorly to treatment.

A critical aspect associated with individual participant or case series designs is that such studies require an appropriate and clearly stated a priori hypothesis regarding the anticipated mechanism of treatment, or brain activation changes that can be tested and potentially rejected. Even well designed case studies that only include a posteriori explanations of change in activation patterns are simply descriptive and do not provide information about the mechanisms of change. A good example of an a priori hypothesis that can be tested is discussed by Crosson et al. (2005). Here, the authors engaged participants in a specific intervention designed to shift activity from the left to the right frontal lobe.

Group studies

The types of analyses that are feasible for evaluating treatment-induced changes in activation in longitudinal designs are quite different from longitudinal studies with groups of healthy participants, for example, when evaluating the reliability or repeatability of activation over time. In non-brain-damaged participants, task-related activation changes can be assessed at the first level and then these images can be entered into the second level for a group analysis assuming a similar expression of potential effects across the group and time. Conversely, within groups of aphasic participants, individual differences in lesion patterns, functional reorganization following stroke, and the resultant language profiles may be associated with highly variable patterns of behavioral improvement or changes in functional activation. Even in highly homogeneous samples these variables may compromise detection of perilesional activation changes when using simple pre-post comparisons and entering them into a group analysis.

It is clear that alternative strategies for analyzing data from aphasic individuals at the group level are necessary. One option is to use a region-of-interest approach in which specific brain areas are chosen depending on the aims of the study. Then, activation changes in these pre-defined ROIs can be correlated with performance gains following treatment. Examples of this approach can be found in the studies by Richter et al. (2008) and Meinzer et al. (2008). Another approach that does not require a priori selection of ROIs has been used in three previous studies by Raboyeau et al. (2008), Menke et al. (2009) and Fridriksson (2010). In these studies pre-post activation patterns were compared directly on an individual level and the resulting images were entered into a whole brain regression analysis. Here, a behavioral regressor (e.g., performance gains after treatment) was used to predict which activated brain areas were associated with superior behavioral improvement in a group of aphasic participants. This statistically powerful approach does not suffer from many of the problems associated with averaging across a heterogeneous group, but it must, nonetheless, be approached with caution as issues of heterogeneity of language-lesion and language-deficit relations are still highly relevant for interpretation of the results. In addition, when using correlational approaches, small sample sizes (as is the case in most aphasia treatment studies) are prone to outlier effects. Thus, researchers are advised to closely inspect (and report) their data and deal with outliers in appropriate ways. With regard to homogeneity of the aphasic group, it has been noted in previous reviews that inhomogeneous samples in cross-sectional designs may reduce detection power and false positives (see Crosson et al., 2007, and Price et al., 2006). On the other hand, in the context of treatment studies, highly homogeneous samples may reduce variability of performance improvements and functional activation changes, which in turn reduce statistical power for correlational methods and may prevent identification of predictors associated with treatment success. Importantly, however, imaging of treatment-induced changes in neural activation in groups of individuals with aphasia can provide information about predictors of treatment success when activation changes over time are correlated with a given indicator of behavioral performance improvement (i.e., which functional activation changes produce the best outcome). This in turn may provide information about which patients are best suited for a particular treatment approach. Moreover, the results can then be generalized, at least with regard to the same treatment paradigm and similar patient populations. With regard to clinical rehabilitation, this is very important, as it may eventually guide the assignment of individual patients to specific treatment approaches.

Comparison with a group of healthy participants

Although longitudinal assessments of activation patterns in aphasic groups at a given recovery stage greatly benefit from a healthy control group, the assessment of treatment effects over time does not necessarily require a healthy control group. On the other hand, in some instances the inclusion of a healthy group of participants may have some advantages. First, instead of assuming that a given paradigm elicits activity in a given number of regions, the validity of the paradigm can be verified by including healthy participants. Second repeated assessment of healthy participants can provide a measure of reliability. Third, comparison of changed activation patterns in aphasic individuals to those of a non-brain-damaged control group allows assessing whether changes occur within or outside of the "normal" language network (e.g., Menke et al., 2009; see Warren et al., 2009 for a recent example of altered temporal lobe functional connectivity in aphasia). Finally, an interesting approach used by Raboyeau et al. (2008) compared the neural signatures of language re-training in

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

chronic anomia to activation changes in healthy controls who were trained to name objects in a previously learned foreign language with low current proficiency. Although there are critical differences between the two groups, the comparison allows examination of mechanisms involved in language reacquisition after stroke and language learning in healthy participants.

When choosing to obtain data from healthy control subjects, however, the participants should be closely matched to the aphasic group with regard to demographic and socio-economic variables. This is an important pre-requisite to allow for valid interpretation of the results. For example, with regard to language tasks, several studies have shown that the neural representation of language production or comprehension mechanisms differ in healthy young vs. older participants (e.g., Wingfield and Grossman, 2006; Fridriksson et al., 2006a, 2006b; Wierenga et al., 2008; Meinzer et al., 2009; Meinzer et al., 2012). Therefore, comparing older participants with stroke-induced aphasia who may evince fundamental changes in brain morphology and function due to age or other variables to a (younger) control group is contraindicated.

Functional network analyses

It is clear that cerebrovascular stroke results in local cortical dysfunction, as well as impaired functioning in remote areas and potentially a compensatory up-regulation of other areas (e.g., Warren et al., 2009). Recent developments in data analysis allow the investigation not only of functional segregation of brain areas related to a specific task, but also assessment of functional integration among different regions. This dynamic network approach has potentially interesting applications to the investigation of longitudinal changes in brain connectivity associated with treatment-induced behavioral changes in aphasia (see Price et al., 2006 for a review of common techniques). Integration within a distributed system is usually understood in terms of effective connectivity, which refers to the influence that one neuronal system exerts over another, either at a synaptic (i.e. synaptic efficacy) or population level (Friston, 2002). Effective connectivity may be measured, for example, using structural equation modeling (SEM) of fMRI data over time. SEM of fMRI time series estimates the effects (in terms of modulation of connection strengths) of experimental manipulation on connectivity among brain regions within specified constraints, based largely on consideration of anatomical connectivity of the brain (Büchel and Friston, 1997, 2000). This approach has been applied to the investigation of training and generalization effects in anomia rehabilitation (Vitali et al., 2009).

Changes in the coupling between different regions can also be investigated using dynamic causal modeling. In DCM, the brain is treated as a dynamic input–state–output system. A given experiment is considered as a designed perturbation of neuronal dynamics that is propagated throughout a network of interconnected anatomical nodes. The coupling between regions is estimated using a series of inputs (i.e., stimulus functions) and the changes in regionally-specific hemodynamic responses are measured (Friston et al., 2003). This approach has successfully been applied to language network changes in primary progressive aphasia (Sonty et al., 2007) and more recently, to the assessment of longitudinal changes associated with anomia treatment (Abutalebi et al., 2009). Variants of Granger Causality Modeling (GCM) are more assumption free than SEM and DCM and are being developed to address functional connectivity in fMRI data (e.g., Zhou et al, 2009). All of these forms of analysis yield insights regarding how areas of brain activity integrate into dynamic systems to perform various tasks that are not available from the simple observation of activity changes in various brain regions.

Importantly, network analyses can be accomplished even in single case studies. Simple pre-post comparisons in single participants (e.g., t-tests) can provide statistical tests of activation changes in various regions of the brain. However, complex dynamic changes at the system level can also be assessed by examining connectivity of brain areas supporting language recovery and changes in response to treatment and underlying driving forces, such as increased compensatory input from non-domain specific areas. The feasibility of a dynamic network approach to examine the effects of aphasia rehabilitation has recently been demonstrated in two case reports (see Abutalebi et al., 2009; Vitali et al., 2009). Moreover, hypothesis-driven modeling of network dynamics in case studies could be guided by obtaining additional information about structural connectivity, by evaluating the integrity of white matter tracts prior to and following treatment (see, for example, papers by Schlaug et al., 2009, and Gauthier et al., 2008). To date, network analyses have not been performed at the group level but such may be possible in future studies (see Warren et al., 2009 for a cross-sectional example in aphasia research).

Summary and conclusions

Neuroimaging in aphasia treatment research has the potential to provide insight into the neuroplastic capacities of the adult human brain and the mechanism of language recovery after brain damage. Moreover, understanding the neural substrates of treatment effects may prompt changes to existing approaches and/or the development of new treatment paradigms that may contribute to the efficacy of rehabilitation efforts. However, neuroimaging of aphasia treatment poses several challenges to researchers that have not been addressed in the past.

The present paper conveys the basic agreement among researchers about critical issues with respect to fMRI data processing in aphasia treatment research that was reached during a consensus conference in Fall 2009 at Northwestern University, Chicago. We reviewed critical issues specifically related to data analysis, including aspects of the pre-processing, the statistical modeling, and the interpretation of such data sets. Moreover, we aimed to provide the reader with a set of general practical guidelines and references to facilitate choosing adequate data analysis strategies.

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 Ansaldo, Roelien Bastiaanse, Pelagie Beeson, Stefano Cappa, David Caplin, Leora Cherney, David Copland, Jenny Crinion, Bruce Crosson, Dirk den Ouden, Susan Edwards, Evy Fitch-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, Steve Small, Benda Rapp, Dorothee Saur, Marion Smits, Yasmeen Faroqi-Shah, Cynthia Thompson, Eisha Wali.

This work was also supported in part by the Bundesministerium für Bildung und Forschung, BMBF: 01EO0801 (MM); the National Institutes of Health, Institute on Deafness and other Communication Disorders: DC007646 and DC008286 (PB); R01DC007387 and R21DC009247 (BC); R01DC01948 and R01DC007213 (CT); and the Department of Veterans Affairs Rehabilitation Research & Development Service, Grant B6364L Senior Research Career Scientist Award (BC).

References

Abutalebi, J., Rosa, P.A., Tettamant, M., Green, D.W., Cappa, S.F., 2009. Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. Brain Lang. 109, 141–156.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. Magn. Reson. Med. 39, 500–505.

Andersen, S.M., Rapcsak, S.Z., Beeson, P.M., 2010. Cost function masking during normalization of brains with focal lesions: still a necessity? Neuroimage 53, 78–84.
Ardekani, B., Bachman, A., Striother, S.C., Fujibayashi, Y., Yonekura, Y., 2004. Impact of inter-subject image registration on group analysis of fMRI data. Int. Congr. Ser. 1265, 209–217.
Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.
Bandettini, P.A., 2009. What's new in neuroimaging methods? Ann. N. Y. Acad. Sci. 1156, 260–293.
Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119, 1239–1247.
Birn, R.M., Cox, R.W., Bandettini, P.A., 2004. Experimental designs and processing strategies for fMRI studies involving overt verbal responses. Neuroimage 23, 1046–1058.
Bonakdarpour, J., Parrish, T.B., Thompson, C.K., 2007. Hemodynamic response function in patients with stroke-induced aphasia: implications for fMRI data analysis. Neuroimage 36, 322–331.
Brett, M., Leff, A.P., Rorden, C., Ashburner, J., 2001. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 14, 486–500.
Brett, M., Johnsrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3, 243–249.
Büchel, C., Friston, K., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.
Büchel, C., Friston, K., 2000. Assessing interactions among neuronal systems using functional neuroimaging. Neural Netw. 13, 871–882.
Cox, R.C., 1996. AFNI: software for the analysis and visualisation of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Cox, R.C., 2009. The AFNI and NIfTI Server at the NIMH. http://afni.nimh.nih.gov/2009.
Crinion, J.T., Leff, A.P., 2007. Recovery and treatment of aphasia after stroke: functional imaging studies. Curr. Opin. Neurol. 20, 667–673.
Crinion, J., Price, C.J., 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. Brain 128, 2858–2871.
Crinion, J., Ashburner, J., Leff, A., Brett, M., Price, C., Friston, K., 2007. Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. Neuroimage 37, 866–875.
Crosson, B., Moore, A.B., Gopinath, K., White, K.D., Wierenga, C.E., Gaiefsky, M.E., et al., 2005. Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. J. Cogn. Neurosci. 17, 392–406.
Crosson, B., McGregor, K., Gopinath, K.S., Conway, T.W., Benjamin, M., Chang, Y.L., et al., 2007. Functional MRI of language in aphasia: a review of the literature and the methodological challenges. Neuropsychol. Rev. 17, 157–177.
Crosson, B., Moore, A.B., McGregor, K.M., Chang, Y.L., Benjamin, M., Gopinath, K., et al., 2009. Regional changes in word-production laterality after a naming treatment designed to produce a rightward shift in frontal activity. Brain Lang. 111, 73–85.
Crosson, B., Ford, A., McGregor, K.M., Meinzer, M., Cheshkov, S., Li, X., et al., 2010. Functional imaging and related techniques: an introduction for rehabilitation researchers. J. Rehabil. Res. Dev. 47, vii–xxxiv.
Fridriksson, J., 2010. Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. J. Neurosci. 30, 11558–11564.
Fridriksson, J., Morrow, K.L., Moser, D., Baylis, G.C., 2006a. Age-related variability in cortical activity during language processing. J. Speech Lang. Hear. Res. 49, 690–697.
Fridriksson, J., Rorden, C., Morgan, P.S., Morrow, L., Baylis, G.C., 2006b. Measuring the hemodynamic response in the case of delayed perfusion. Neurocase 12, 146–150.
Fridriksson, J., Moser, D., Bonilha, L., Morrow-Odom, K.L., Shaw, H., Fridriksson, A., Baylis, G.C., Rorden, C., 2007. Neural correlates of phonological and semantic-based anomia treatment in aphasia. Neuropsychologia 45, 1812–1822.
Fridriksson, J., Baker, J.M., Moser, D., 2009. Cortical mapping of naming errors in aphasia. Hum. Brain Mapp. 30, 2487–2498.
Friedman, L., Stern, H., Brown, G.G., Mathalon, D.H., Turner, J., Glover, G.H., et al., 2008. Test-retest and between-site reliability in a multicenter fMRI study. Hum. Brain Mapp. 29, 958–972.
Friston, K.J., 2002. Functional integration and inference in the brain. Prog. Neurobiol. 68, 113–143.
Friston, K., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S., 1995. Spatial registration and normalization of images. Hum. Brain Mapp. 2, 165–189.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.
Gauthier, L.V., Taub, E., Perkins, C., Ortmann, M., Mark, V.W., Uswatte, G., 2008. Remodeling the brain: plastic structural brain changes produced by different motor therapies after stroke. Stroke 39, 1520–1525.
Geissler, A., Lanzenberger, R., Barth, M., Tahamtan, A.R., Milakara, D., Gartus, A., et al., 2005. Influence of fMRI smoothing procedures on replicability of fine scale motor localization. Neuroimage 24, 323–331.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416–429.
Godefroy, O., Duhamel, A., Leclerc, X., Saint Michel, T., Henon, H., Leys, D., 1998. Brain-behaviour relationships. Some models and related statistical procedures for the study of brain-damaged patients. Brain 121, 1545–1556.
Gopinath, K., Crosson, B., Peck, K.K., Moore, A.B., White, K.D., Briggs, R.W., 2005. Detection power adjustment method for improved comparisons between multiple-session individual-subject fMRI scans. Proc. Int. Soc. Magn. Res. Med. 13, 697.
Gopinath, K., Crosson, B., McGregor, K., Peck, K., Chang, Y.L., Moore, A., et al., 2009. Selective detrending method for reducing task-correlated motion artifact during speech in event-related FMRI. Hum. Brain Mapp. 30, 1105–1119.
Hopfinger, J.B., Buchel, C., Holmes, A.P., Friston, K.J., 2000. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage 11, 326–333.
Kielar, A., Milman, L., Bonakdarpour, B., Thompson, C.K., 2011. Neural correlates of covert and overt production of tense and agreement morphology: evidence from fMRI. J. Neurolinguist. 24, 183–201.
Kimberley, T.J., Birkholz, D.D., Hancock, R.A., VonBank, S.M., Werth, T.N., 2008. Reliability of fMRI during a continuous motor task: assessment of analysis techniques. J. Neuroimaging 18, 18–27.
Kohn, S.E., Goodglass, H., 1985. Picture-naming in aphasia. Brain Lang. 24, 266–283.
Leger, A., Demonet, J.F., Ruff, S., Aithamon, B., Touyeras, B., Puel, M., et al., 2002. Neural substrates of spoken language rehabilitation in an aphasic patient: an fMRI study. Neuroimage 17, 174–183.
Machielsen, W.C.M., Rombots, S.A.R.B., Barkhof, F., Scheltens, P., Witter, M.P., 2000. FMRI of visual encoding: Reproducibility of activation. Hum. Brain Mapp. 9, 156–164.
Maldjian, J.A., Laurienti, P.J., Driskill, L., Burdette, J.H., 2000. Multiple reproducibility indices for evaluation of cognitive functional MR imaging paradigms. Am. J. Neuroradiol. 23, 1030–1037.
Martin, P.I., Naeser, M.A., Doron, K.W., Bogdan, A., Baker, E.H., Kurland, J., et al., 2005. Overt naming in aphasia studied with a functional MRI hemodynamic delay design. Neuroimage 28, 194–204.
Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage 2, 89–101.
McKeown, M.J., Sejnowski, T.J., 1998. Independent component analysis of fMRI data: examining the assumptions. Hum. Brain Mapp. 6, 368–372.
Meinzer, M., Breitenstein, C., 2008. Functional imaging studies of treatment-induced recovery in chronic aphasia. Aphasiology 22, 1251–1268.
Meinzer, M., Flaisch, T., Obleser, J., Assadollahi, R., Djundja, D., Barthel, G., et al., 2006. Brain regions essential for improved lexical access in an aged aphasic patient: A case report. BMC Neurol. 6, 28.
Meinzer, M., Flaisch, T., Breitenstein, C., Wienbruch, C., Elbert, T., Rockstroh, B., 2008. Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. Neuroimage 39, 2038–2046.
Meinzer, M., Flaisch, T., Wilser, L., Eulitz, C., Rockstroh, B., Conway, T., et al., 2009. Neural signatures of semantic and phonemic fluency in young and old adults. J. Cogn. Neurosci. 21, 2007–2018.
Meinzer, M., Harnish, S., Conway, T., Crosson, B., 2011. Recent developments in functional and structural imaging of aphasia recovery after stroke. Aphasiology 25, 271–290.
Meinzer, M., Seeds, L., Flaisch, T., Harnish, S., Cohen, M.L., McGregor, K., Conway, T., Benjamin, M., Crosson, B., 2012. Impact of changed positive and negative task-related brain activity on word-retrieval in aging. Neurobiol. Aging 33, 656–669.
Meltzer, J.A., Postman-Caucheteux, W.A., McArdle, J.J., Braun, A.R., 2009. Strategies for longitudinal neuroimaging studies of overt language production. Neuroimage 47, 745–755.
Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgartner, A., Schiffbauer, H., et al., 2009. Imaging short- and long-term training success in chronic aphasia. BMC Neurosci. 10, 118.
Mikl, M., Marecek, R., Hlustik, P., Pavlicova, M., Drastich, A., Chlebus, P., et al., 2008. Effects of spatial smoothing on fMRI group inferences. Magn. Reson. Imaging 26, 490–503.
Murata, Y., Sakatani, K., Hoshino, T., Fujiwara, N., Kano, T., Nakamura, S., et al., 2006. Effects of cerebral ischemia on evoked cerebral blood oxygenation responses and BOLD contrast functional MRI in stroke patients. Stroke 37, 2514–2520.
Nicholas, M., 2005. Aphasia and dysarthria after stroke. In: Barnes, M., Dobkin, B.H., Bogousslavsky, J. (Eds.), Recovery after stroke. Cambridge University Press, Cambridge, pp. 474–502.
Parrish, T.B., Gitelman, D.R., LaBar, K.S., Mesulam, M.M., 2000. Impact of signal-to-noise on functional MR. Magn. Reson. Med. 44, 925–932.
Peck, K.K., Moore, A.B., Crosson, B.A., Gaiefsky, M., Gopinath, K.S., White, K., et al., 2004. Functional magnetic resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. Stroke 35, 554–559.
Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E., 2008. Guidelines for reporting an fMRI study. Neuroimage 40, 409–414.
Postman-Caucheteux, W.A., Birn, R.M., Pursley, R.H., Butman, J.A., Solomon, J.M., Picchioni, D., et al., 2010. Single-trial fMRI Shows Contralesional Activity Linked to Overt Naming Errors in Chronic Aphasic Patients. J. Cogn. Neurosci. 22, 1299–1318.
Price, C.J., Crinion, J., Friston, K.J., 2006. Design and analysis of fMRI studies with neurologically impaired patients. J. Magn. Reson. Imaging 23, 816–826.
Raboyeau, G., De Boissezon, X., Marie, N., Balduyck, S., Puel, M., Bezy, C., et al., 2008. Right hemisphere activation in recovery from aphasia: lesion effect or function recruitment? Neurology 70, 290–298.
Richter, M., Miltner, W.H., Straube, T., 2008. Association between therapy outcome and right-hemispheric activation in chronic aphasia. Brain 131, 1391–1401.
Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. Behav. Neurol. 12, 191–200.
Rorden, C., Karnath, H.O., Bonilha, L., 2007. Improving lesion-symptom mapping. J. Cogn. Neurosci. 19, 1081–1088.
Rossini, P.M., Altamura, C., Ferretti, A., Vernieri, F., Zappasodi, F., Caulo, M., et al., 2004. Does cerebrovascular disease affect the coupling between neuronal activity and local haemodynamics? Brain 127, 99–110.
Röther, J., Knab, R., Hamzei, F., Fiehler, J., Reichenbach, J.R., Büchel, C., Weiller, C., 2002. Negative dip in BOLD fMRI is caused by blood flow–oxygen consumption uncoupling in humans. Neuroimage 15, 98–102.
Schlaug, G., Marchina, S., Norton, A., 2009. Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. Ann. N. Y. Acad. Sci. 1169, 385–394.

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058

Seghier, M.L., Ramlackhansingh, A., Crinion, J., Leff, A.P., Price, C.J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. Neuroimage 41, 1253–1266.
Serences, J.T., 2004. A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. Neuroimage 21, 1690–1700.
Sonty, S.P., Mesulam, M.M., Weintraub, S., Johnson, N.A., Parrish, T.B., Gitelman, D.R., 2007. Altered effective connectivity within the language network in primary progressive aphasia. J. Neurosci. 27, 1334–1345.
Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zacks, J.M., 2003. Reliability of functional localization using fMRI. Neuroimage 20, 1561–1577.
Talairach, J.P., Tournoux, P., 1988. Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system – an approach to cerebral imaging. Thieme Medical Publishers, New York, NY.
Tanabe, J., Miller, D., Tregellas, J., Freedman, R., Meyer, F.G., 2002. Comparison of detrending methods for optimal fMRI preprocessing. Neuroimage 15, 902–907.
Thompson, C.K., den Ouden, D.B., 2008. Neuroimaging and recovery of language in aphasia. Curr. Neurol. Neurosci. Rep. 8, 475–483.
Thompson, C.K., Bonakdarpour, B., Fix, S.F., 2010a. Neural mechanisms of verb argument structure processing in agrammatic aphasic and healthy age-matched listeners. J. Cogn. Neurosci. 22, 1993–2011.
Thompson, C.K., den Ouden, D.B., Bonakdarpour, B., Garibaldi, K., Parrish, T.B., 2010b. Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. Neuropsychologia 48, 3211–3227.
Turner, R., Howseman, A., Rees, G.E., Josephs, O., Friston, K., 1998. Functional magnetic resonance imaging of the human brain: data acquisition and analysis. Exp. Brain Res. 123, 5–12.
Van Essen, D.C., 2004. Surface-based approaches to spatial localization and registration in primate cerebral cortex. Neuroimage 23, S97–S107.
van Oers, C.A., Vink, M., van Zandvoort, M.J., van der Worp, H.B., de Haan, E.H., Kappelle, L.J., et al., 2010. Contribution of the left and right inferior frontal gyrus in recovery from aphasia. A functional MRI study in stroke patients with preserved hemodynamic responsiveness. Neuroimage 49, 885–893.
Vitali, P., Tettamanti, M., Abutalebi, J., Ansaldo, A.I., Perani, D., Cappa, S.F., Joanette, Y., 2009. Generalization of the effects of phonological training for anomia using structural equation modelling: A multiple single-case study. Neurocase 4, 1–13.
Voyvodic, J.T., 2006. Activation mapping as a percentage of local excitation: fMRI stability within scans, between scans and across field strengths. Magn. Reson. Imaging 24, 1249–1261.
Voyvodic, J.T., Petrella, J.R., Friedman, A.H., 2009. fMRI activation mapping as a percentage of local excitation: consistent presurgical motor maps without threshold adjustment. J. Magn. Reson. Imaging 29 (4), 751–759.
Warren, J.E., Crinion, J.T., Lambon Ralph, M.A., Wise, R.J., 2009. Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. Brain 132, 3428–3442.
Wei, X., Yoo, S.S., Dickey, C.C., Zou, K.H., Guttmann, C.R.G., Panych, L.P., 2004. Functional MRI of auditory verbal working memory: Long-term reproducibility analysis. Neuroimage 21, 1000–1008.
Wierenga, C.E., Benjamin, M., Gopinath, K., Perlstein, W.M., Leonard, C.M., Rothi, L.J., Conway, T., Cato, M.A., Briggs, R., Crosson, B., 2008. Age-related changes in word retrieval: role of bilateral frontal and subcortical networks. Neurobiol. Aging 29, 436–451.
Wingfield, A., Grossman, M., 2006. Language and the aging brain: patterns of neural compensation revealed by functional brain imaging. J. Neurophysiol. 96, 2830–2839.
Zhou, Z., Chen, Y., Ding, M., Wright, P., Lu, Z., Liu, Y., 2009. Analyzing brain networks with PCA and conditional Granger causality. Hum. Brain Mapp. 30, 2197–2206.

Please cite this article as: Meinzer, M., et al., Neuroimaging in aphasia treatment research: Consensus and practical guidelines for data analysis, NeuroImage (2012), doi:10.1016/j.neuroimage.2012.02.058