2018

An Investigation Of Brain Normalization And Lesion Compensation Techniques Applied To Stroke

Taylor Hanayik
University of South Carolina - Columbia

An Investigation of Brain Normalization and Lesion Compensation Techniques Applied to Stroke

by

Taylor Hanayik

Bachelor of Arts
University of South Carolina, 2013

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Experimental Psychology

College of Arts and Sciences

University of South Carolina

2018

Accepted by:

Chris Rorden, Major Professor

Scott Decker, Committee Member

Dirk den Ouden, Committee Member

John Richards, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

DEDICATION

For my wife Kaitlin, my family, and my friends who have been on this journey with me. Kaitlin, a special dedication is warranted for your understanding, support, and love throughout my graduate school years. You have had to endure all the time I have spent “away” from you, even though we are in the same room most of the time. I’m most excited about what is in store for us once the next chapter of our life together starts. To Professor Rorden, you have been a fantastic mentor over the years. I have learned, matured, and adapted based on your guidance. I appreciate your knowledge, pursuit of truth, and generosity. I can only hope that my future students view me, as I do you. Thank you for all you have taught and provided.

ACKNOWLEDGEMENTS

I thank my committee members for their time, and guidance throughout the course of my dissertation project. Special thanks are devoted to Professor John Richards who dedicated his vast computing resources to my cause. I also thank the staff of the Research Cyber Infrastructure team at the university who have provided great advice over the past few months. Thank you to my fellow lab members Roger, Grigori, and Makayla for always offering a helping hand and being so supportive. I also thank my family for their support throughout my studies. To my uncle Michael, I owe you more thanks than can be expressed for encouraging my scientific curiosity from an early age. To my dad, thank you for supporting me always, no matter what. To my mom, thank you for all the guidance and support you have provide over the years. To Professor Fridriksson and the Aphasia Lab, thank you for your support, scientific encouragement, and faith in me over the past years. I will be sad to move on from the Rorden/Fridriksson lab family.

ABSTRACT

One of the most ubiquitous steps in neuroimaging is the normalization of brain images. The process of normalization attempts to match any given brain to a standardized template image (e.g. the MNI 152 image). However, clinical images such as those from stroke participants present many challenges when we attempt to warp them to the space of template images, which are typically representative of neurologically healthy individuals. Many software packages exist to facilitate normalization of brain images, but most have limited options available to compensate for brain injury, which is often disruptive to these algorithms. Of the injury compensation methods that do exist, they are varied across software packages. The current study aimed to assess the contemporary methods available in state of the software commonly used across the field. Specifically, we assessed SPM12's new tissue filling procedure on masked clinical images, and LINDA, a fully automated lesion segmentation algorithm combined with ANTs normalization. Across normalization methods, we compared each software package's default injury compensation strategy to the nonstandard enantiomorphic lesion healing procedure. We created an artificial dataset of more than 10,000 images representing stroke related injury, and assessed each normalization method (SPM's unified segmentation, DARTEL, ANTs) on multiple performance metrics. Overall, we found that the optimal injury compensation strategy for clinical images varied by the normalization method used, and the metric it was evaluated on. Finally, we present evidence of each

normalization method and brain injury compensation technique's effect on predicting behavior deficits from brain injury using support vector regression. Our results show that prediction accuracy (and error) can be affected by the normalization technique used.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTON

1.1 OVERVIEW

All group neuroimaging analyses rely on the "normalization" of brain images to a common space in order to make statistical inferences. These normalization routines attempt to match the variations in the size, shape and cortical folding pattern across individuals. These automated normalization routines can be disrupted by the presence of brain injury (Brett et al., 2001). Successful normalization of brain injury is critical to guide optimal surgery (Bonilha et al., 2007) make inferences about critical brain regions, predict recovery and guide therapy. Several strategies have been proposed to allow the robust normalization of images from brain injured individuals. Here we compare several of these methods, including promising recent developments in order to identify the best method for this task. We provide several different performance metrics to help evaluate these different routines.

Normalization is useful in that it provides a standardized coordinate system to be used to compare neuroanatomy across individuals and research institutions (Grabner et al., 2006). Normalization techniques are also useful to determine which brain areas are typically activated by a cognitive task (in group statistics), which brain areas typically atrophy in dementia, and which brain areas are related to certain behavior deficits when

damaged. There are many different tools available to normalize neuroimaging data from individuals with brain injury. These different tools use different mathematical methods that tend to trade-off various strengths and limitations that influence the robustness of the algorithm (how often it fails catastrophically), and average accuracy. For example, methods that have more degrees of freedom can in theory provide a better fit from one brain image to the next, but can also be adversely influenced by small features that will yield poor results. A common theme of the popular normalization methods is the expectation to use an "average" brain as a standard template. Each individual's brain is warped to match the size, orientation, and shape of this reference image. Typically, a template is based on hundreds of individuals, but the appropriate template may vary from one use case to another (e.g. a child has a smaller brain, and tissue composition than an adult; Sanchez, Richards, & Almli, 2012), and from one software package to another. For example, both SPM (Statistical Parametric Mapping Software, Wellcome Department of Imaging Neuroscience, University College, London, UK, www.fil.ion.ucl.ac.uk/spm/), and FSL (FMRIB's Software Library; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) include common template images relevant to most researchers, but also include unique template images created independently from other institutions. Neuroimagers generally accept that these template images represent a sample of neurologically healthy individuals from the population with typical neuroanatomy. Some templates are exceptions to this rule, for example the average template constructed by Pustina and colleagues (2016), which consisted of 115 elderly control participants and 93 patients with various diseases such as Parkinson's and dementia (though no individual had focal damage).

Outside of research needs, generally healthy individuals are not likely to undergo magnetic resonance imaging (MRI) without a justified cause. It is far more likely that individuals undergo MRI scanning for clinical purposes either ordered by their physician or as part of the patient's participation in medical research. This dichotomy of needs between clinical and healthy neuroimaging represents a missing link between the efforts of most neuroimaging standardization practices (e.g. generating brain templates), and their application to clinical data from brain injured patients. Individuals with brain damage such as stroke, tumor, or severe atrophy are by definition not anatomically normal, and therefore precisely matching their anatomy to any standardized template image is more challenging compared to working with neurologically healthy individuals (Brett, Leff, Rorden, & Ashburner, 2001; Fiez, Damasio, & Grabowski, 2000).

Over the years, different groups have advocated a variety of methods for normalizing scans of injured brains. The present goal is to directly compare these methods to quantify their performance. Further, by understanding the novel contributions of different clinical normalization techniques, we can investigate a hybrid approaches that could outperform previous methods.

1.2 BASICS OF IMAGE REGISTRATION

Normalization is merely a specialized application of general image registration (Friston et al., 1995). Image registration (aka. "image matching", or "co-registration") is widely used in many computer vision applications (e.g. special effects, or animation; Le, Ma, & Deng, 2012), and is not limited to medical image analysis. Regardless of the field of application, image registration can be applied to variations of the same scene (e.g. correcting for camera shake between two successive frames of a movie) or to completely

different scenes (e.g. matching photographs from the faces of two different individuals). A crucial realization is that in the first case there is an objectively perfect solution (the two images of the same scene can in theory be perfectly aligned), while in the second case we can make the faces similar in alignment but there is no objective perfect solution (e.g. Marilyn Monroe's mole has no optimal size or position on another person's photo). In either case, many registration algorithms are designed to minimize a “cost function”, which measures the difference between images based on the mathematical algorithm used (e.g. sum of squared differences).

Brain normalization is special in the sense that an individual’s brain is matched to an averaged template image, rather than to an image from a separate participant. Similar to the face example above, there is no perfect solution to match an individual brain to a template. For example, a method that attempts to carefully align folds will necessarily cause distortions in the brain volume, whereas a method that attempts to match the size of the gray matter will yield poorer sulcal matching. Different methods choose different tradeoffs in making these matches. A nice analogy is warping the spherical shape of the earth onto a 2-dimensional rectangle (as most maps are displayed). The Mercator projection emphasizes preserving angles, while the Peters projection emphasizes preserving surface area. Neither transform is correct, they just use a different cost function to decide what features are important. In contrast, other forms of image registration used in neuroimaging align one image from an individual to a different image from the same individual (e.g. motion correction, eddy current correction, T1w to T2w co-registration). In such cases, there is an objectively perfect solution to align the same individual’s anatomy. In all of these cases, image registration techniques are used to put

one image (the moving image) in the space of another (the reference), and to match gross anatomical features from one image to the next. Normalization facilitates later image processing stages and image based group statistics (Brett et al., 2001). Image registration (including normalization) is a fundamental part of most neuroimaging workflows (Jenkinson & Smith, 2001), and is crucial to making generalized inferences related to a sample and its population, and comparisons between groups. To make the most accurate inferences about brain-behavior relationships in group clinical settings, it is only logical that we strive to use image registration methods with the best image matching results, and those that preserve important anatomical spatial relationships.

When an individual's brain image has been registered to a template (normalized), the newly created image of that individual now conforms to "template" or "normalized" space. Any template can be used in theory. For example, one could choose a template that most closely represents the population of interest (minimizing distortion), or one could choose a universal template that allows comparison to other neuroimaging centers and earlier work. Most research has been reported following the latter approach, using the contemporary MNI152 template (Grabner et al., 2006) that roughly matches the coordinate system used by Talairach & Tournoux, (1988). This template is widely used as a standardized coordinate system across neuroimaging studies. The benefit of a standardized coordinate system is the ability to generalize and compare experimental results across samples, across time periods, and across research labs that use varying participants, techniques, and machinery (Brett et al., 2001).

One of the seminal standardized templates was developed by Talairach and Tournoux (1988), based on the brain of a single individual mapped post-mortem and

meticulously dissected into slices and photographed. These photographs were used to make a detailed atlas of this individual's neuroanatomy that has since been used as a reference when viewing clinical neuroimaging data to identify normal versus abnormal anatomy, and in viewing neuroimaging data from healthy volunteers alike.

Although it was highly detailed, the Talairach neuroanatomical atlas was not representative of any particular population since it consisted of data from only one deceased individual, nor did the images have similar contrast to popular neuroimaging modalities. In contrast, the MNI152 template is an average *in vivo* representation generated from a large population of neurologically healthy individuals who participated in MRI scanning for research purposes. It is a standard template included with many neuroimaging tools. The MNI152 template was originally created by linearly transforming 152 individual brain images to a previous reference image known as MNI305 (Collins, Neelin, Peters, & Evans, 1994), and then simply averaging the resulting images that were in a unified image space. However, recent versions of this template have been constructed differently, and the preferred method now makes use of nonlinear image registration techniques which are better able to match the complex folding patterns of the human cortex (yielding a less blurry template). This leads us to the next important concept in image registration.

When neuroimagers choose an image registration method, there are typically two categories of techniques: linear, and nonlinear (Brett et al., 2001; Ripollés et al., 2012). Linear techniques have the advantage of robustness, simplicity, and speed given their small degrees of freedom (Jenkinson & Smith, 2001). Linear methods can also make use of a minimal set of transforms, or a full set. For example, a minimal set of transforms

would simply move a target image left, right, up, down, forward, and back (translation), in addition to moving about the pitch, roll, and yaw axes (rotation) in relation to the reference image. This limited set of linear transforms is widely known as “rigid body” transformation, and is notable in that it does not manipulate the overall shape or size of the brain image that is being warped to match a template image. This means that the size of the brain, and the anatomical spatial relationships remain unchanged. Rigid body transformations are ideally suited for motion correction in time series data such as fMRI where we are scanning a single individual across time with a single image modality. In this situation, we do not expect the individual’s brain to grow or shrink in any appreciable way. In contrast, a full set of linear transforms includes those of the rigid body method, in addition to a scaling, and a shearing transformation (often collectively referred to as affine transformation). Scaling manipulates the overall size of the brain image being warped, and shearing manipulates the shape (i.e. stretching the anatomy in a linear manner to better match the reference brain’s features). While scaling and shearing can change the overall size of the image, these are applied equally to the whole image. An area that is twice as big as another will remain twice as big afterwards (they are both scaled equally). Likewise, any three points that were co-linear prior to the linear transform remain so afterwards (Affine transformation - Encyclopedia of Mathematics). These properties are both the strength and weakness of this transform. It is unable to deform local features without deforming global features in the same manner. This constrains the method, which limits its ability to fit fine details (like complex cortical folding), but also tends to limit its ability to eliminate real details (Andersen, Rapcsak, & Beeson, 2010). Note that each of the four linear transforms (translation, rotation, scaling,

and shear) can be applied independently in each of the three dimensions of an image. Therefore, this warping can be concisely defined with 12 parameters (e.g. it has 12 degrees of freedom). In sum, the linear transforms provide a robust but relatively crude match between images. Because they are rapid and robust, these transforms are typically computed prior to more aggressive approaches. Therefore, we adopted the phrase "linear-only normalization" to refer to an algorithm that terminates after only computing these 12 parameters when registering one image to another.

The more advanced, and often more accurate (though fragile) image registration methods are nonlinear (Ashburner & Friston, 1999). These methods offer a reduced image registration error (aka "cost function") between a target image and its reference, and are able to more accurately match the intricacies of anatomical differences. This enhancement is in part due to their increased degrees of freedom, which range from thousands to millions (Klein et al., 2009; Ripollés et al., 2012). Although the advanced nonlinear methods generally outperform purely linear registration methods, the improved registration comes at the cost of increased computational complexity and time (Klein et al., 2009), as well as the danger of local over-fitting. However, the benefit of improved anatomical precision can often outweigh this increased complexity, especially in a clinical research setting. This improvement in image registration extends to the process of normalization to template images as well.

1.3 CHALLENGES PRESENTED BY BRAIN INJURY

In a clinical research setting, even in the acute phase of symptom onset, we may wish to better understand how the injured anatomy from a group of participants differs compared to a control sample. It is also important to assess the similarity of injury across

a group of clinical images to understand the relationship between injuries and deficits (Stamatakis & Tyler, 2005). Given that brain injury can often result in physical or cognitive deficits, neuroimagers use methods such as voxel based morphometry (VBM; Ashburner & Friston, 2000, 2001) to measure anatomical differences between participant groups or across timepoints. We may also use methods such as voxel-based lesion symptom mapping (VLSM; Bates et al., 2003; Rorden, Karnath, & Bonilha, 2007), which require overlapping lesion images in the same stereotaxic coordinate space to examine commonalities among damaged neuroanatomy, and related behavior deficits (Brett et al., 2001). In fields such as chronic stroke research, we must necessarily normalize brain images to a standard space to perform statistical comparisons on those images. Unfortunately, the efforts of most registration methods are developed and validated using brain imaging data from healthy volunteers with typical anatomy. This presents a problem when researchers wish to warp a clinical image (i.e. brain injury due to stroke) to a template. The clinical image may have many anatomical structures missing, or partially damaged, which will result in increased error when matching to a template (Brett et al., 2001; Crinion et al., 2007).

The effect of abnormal, or missing anatomy in the case of stroke can have a significant impact on a normalization method's cost function. A cost function between two brain images, such as the sum of squared differences, or correlation ratio (among others), will produce extreme values in areas of low signal due to missing tissue compared to the tissue signal in the template image. Normalization algorithms work by iteratively deforming the input image by a small amount and seeing if this improves or hurts the costs function of the image with respect to the template, in the case where the

change leads to an improvement the change is stored for the next iteration. Cost function algorithms will attempt to minimize the difference between two images across these iterations. In the case of missing tissue, these algorithms can be lured into making catastrophic deformations of the image (Brett et al., 2001). Linear only normalization can be robust to the presence of a brain lesion, or other abnormality, but such methods only result in crude matches in overall brain shape and size to the reference template. Linear only methods do not deliver the desired level of precision when we wish to make generalizable claims about brain-behavior relationships in clinical neuroimaging.

Linear only methods may lack precision in favor of robustness, but the more precise non-linear methods can be used in conjunction with an initial linear fit (Brian B. Avants, Tustison, & Song, 2009). In the presence of normal anatomy, this combination of linear and non-linear methods can lead to a greater match to the template brain image. However, when tissue is missing or abnormal, even the non-linear normalization methods can give distorted results (Brett et al., 2001; Kim, Patel, Avants, & Whyte, 2015). Often the nonlinear methods will shrink the brain lesion or abnormal tissue, which indeed does result in a better match to the template, but introduces the new issue of lesion erosion as well as distortions in neighboring tissue surrounding the lesion. However, when the brain injury is compensated for, the effects of lesion erosion are minimized, or alleviated (Brett et al., 2001; Ripollés et al., 2012).

1.4 CLINICAL NEUROIMAGING ANALYSIS TECHNIQUES

Many groups have developed techniques to overcome the challenges presented by brain injury (particularly stroke research). The methods include cost function masking (CFM) (Brett et al., 2001), enantiomorphic lesion filling (Nachev et al., 2008), variations

of unified segmentation normalization (Ashburner & Friston, 2005; Seghier, Ramlackhansingh, Crinion, Leff, & Price, 2008), using an age matched template for stroke datasets (Christopher Rorden, Bonilha, Fridriksson, Bender, & Karnath, 2012), and a fully automated method using machine learning combined with diffeomorphic non-linear normalization (Pustina et al., 2016). Note though, that unified segmentation normalization, and some other high DOF normalization methods can work when applied directly to images with lesions, but prior evidence shows they often benefit by including CFM (Andersen et al., 2010; Brett et al., 2001) or lesion filling (Nachev et al., 2008). Otherwise, lesion erosion and displacement can be quite severe depending on the method (Ripollés et al., 2012).

To this day, CFM (Brett et al., 2001) is a popular method still applied to clinical neuroimaging data, and is implemented in many software packages. CFM works by simply limiting the mathematical operations of minimizing the cost function to non-lesioned brain areas. To facilitate the masking procedure, researchers must supply an image mask that is typically created manually by hand using medical image viewing software. This mask is binary in most cases, having zeros at the location of the lesion, and ones elsewhere. Areas indexed by a zero will have no effect on the cost function calculations, and therefore not influence the image registration procedure. However, note that other abnormal anatomy such as enlarged ventricles due to stroke will affect the image registration procedure if not included in the image mask. This may be desirable, as the nonlinear routines will tend to shrink these ventricles toward a normal size, enlarging the lesion to better match its volume prior to necrosis. The CFM procedure results in decreased root mean squared (RMS) displacement values compared to unmasked

procedures, indicating that voxels within the brain images were less deformed overall, as well as near the masked part of the image when compared to unlesioned versions of the same images (Brett et al., 2001).

The RMS displacement measure is a standard metric used in the field. One popular technique is to create simulated lesions in healthy brains, and compare the normalization of a healthy brain to the same brain with a lesion inserted. In this case, displacement reveals how much the lesion perturbed the normalization parameters from the gold standard case of warping a completely healthy brain. The RMS displacement across the brain can be represented as a single number, with lower values indicating less overall deviation from a reference. This is explained in greater detail in the subsequent methods section.

Although CFM is still widely used in clinical neuroimaging, and a readily available method in most neuroimaging software packages, it does have limitations. Lesion size, or the size of the masked region can dramatically affect normalization results. A larger lesion mask necessarily means that less of the image is useable in cost function calculations, and therefore larger lesions are correlated with more error in normalization procedures (Brett et al., 2001; Crinion et al., 2007; Nachev et al., 2008). CFM traditionally requires the user to supply a manually traced lesion mask, making the normalization process time consuming, and less objective compared to fully automated methods used on brain imaging data from healthy individuals, which do not require additional user generated input. A notable advantage of CFM is that it is not limited in the region in which the mask is applied, meaning that any type of injury or abnormality

can be omitted from the template matching procedure. Other lesion compensation techniques can be limited to unilateral injury only.

To overcome the main limitation of omitting image data due to CFM, Nachev et al. (2008) devised a solution which relies on the brain's natural symmetry between the left and right hemispheres. Their solution, known as enantiomorphic normalization, also relies on a user supplied lesion mask, but differs from CFM in that the mask is used to "fill" the lesioned parts of the brain image from the undamaged hemisphere. Essentially, the lesion mask is mirrored to the undamaged hemisphere, and image intensities from the undamaged hemisphere are used to replace tissue in lesioned regions. By replacing the lesioned parts of the image with data from the opposite hemisphere of the same individual, the authors state quite frankly that this filled in estimated data is better than no data at all (as in CFM). This "healed" brain image is then run through any normalization algorithm just as a healthy brain image would be, and the lesion can be restored after normalization is complete. Using the whole-brain RMS (voxel displacement) measure of performance, the authors show that the enantiomorphic normalization method significantly outperforms unassisted normalization methods, as well as the CFM procedure on a dataset of stroke participants, and on a dataset where artificial lesions were injected into otherwise normal brain images. Although this mirror image method clearly outperformed CFM as measured by RMS values, it also has its limitations. Namely, it is subject to the same bias as CFM given that the lesion masks are traditionally generated manually by the user (however automated methods do exist, and are discussed later). Also, the mirroring process implies that regions in the undamaged hemisphere are representative of the missing tissue from the lesioned area. In general this

assumption will work in many cases, but there have been measured asymmetries between the left and right hemispheres (Toga & Thompson, 2003). From first principles though, filling in the lesioned area should benefit normalization to a template image more so than omitting entire regions using CFM. In addition, the enantiomorphic method is limited to cases where there is no bilateral damage in homologous brain regions, since the mirroring process will be ineffective if a user replaces damaged tissue with equally damaged tissue from the opposite hemisphere. This can be especially problematic in the presence of extremely enlarged ventricles. The method is thus generally limited to unilateral brain injury. A minor criticism of the enantiomorphic method as presented by Nachev et al. (2008) is that their comparison to CFM can be viewed as slightly unfair. In their comparison they enlarged each lesion mask by 10% prior to normalization, but kept the volume of the enantiomorphically replaced tissue much closer to the original lesion volume (with the exception of smoothing near the edges). The choice to dilate (enlarge) lesion masks by 10% ignores a previous report suggesting that minimal lesion dilation provides ideal results with CFM (Brett et al., 2001). Lastly, the effect of lesion size on the precision of the enantiomorphic normalization method is similar to the effect of CFM, in that as lesion size increases so does the RMS displacement. Figures 4B and 5B from Nachev et al. (2008) suggest a linear relationship in that as lesion size increases, so does the measured voxel displacement which could be explained by hemispheric asymmetry (and related mismatches) becoming more apparent as the replaced region increases in size. As a final note, while the enantiomorphic method seems intuitively to be useful, the authors did not provide software to emulate their method, and they examined its effectiveness using SPM2 which is no longer state-of-the-art. Here we revive this clever

method, provide an open source implementation, and examine its relevance with the current generation of normalization tools.

Both the CFM and enantiomorphic lesion filling techniques rely on supplying an additional mask image to existing normalization strategies. However, the unified segmentation normalization method introduced in SPM5 (Ashburner & Friston, 2005) is often used on clinical data as well. This method does not explicitly require a lesion mask to operate, but can be significantly enhanced by providing one when working with clinical data (Andersen et al., 2010; though see Crinion et al., 2007). Historically, normalization (aligning a brain to a template) and segmentation (classifying the proportion of gray matter, white matter and cerebral spinal fluid for each voxel) were seen as independent operations. In contrast, unified segmentation normalization (referred to as the unified method) creates a virtuous cycle: improving the accuracy of normalization provides a better estimate of where we expect to see different tissue. Likewise, improved segmentation allows us to better normalize one individual's gray matter to a gray matter template. Finally, estimating both normalization and segmentation can help provide better estimates of field inhomogeneity, which in turn benefit the estimation of the other parameters (Ashburner & Friston, 2005). An extension of the unified method is DARTEL, which (at least with healthy populations) can further refine the normalization of segmented images (Ashburner, 2007). Additionally, DARTEL was considered a top performing method in an assessment of automated normalization of brain injured images (Ripollés et al., 2012).

To explicitly test how well the unified method works with stroke imaging data, Crinion et al. (2007) tested it against the standard linear and non-linear normalization

(without segmentation) methods included in SPM5. On average, the unified normalization procedure significantly outperformed all other methods assessed, indicating that it is both effective, and can work in the absence of user generated lesion masks. Additionally, the authors explicitly compared the unified method both with and without CFM. There was no significant difference, indicating that unified segmentation normalization could perhaps be a well suited method for automatic normalization of stroke neuroimaging data without user defined masks. However, precisely delineated lesion masks are often necessary for lesion symptom mapping analyses common in stroke research, and this method does not provide the user with such as lesion map even though it is capable of normalizing clinical images. However, as noted by Andersen et al. (2010), the images (Crinion et al., 2007) may not accurately represent real pathological brains since lesions were synthesized from real patients but artificially placed into otherwise healthy brain images (same data as Brett et al., 2001). Unfortunately, the use of artificially lesioned brain images cannot be avoided in most studies that measure the performance of brain normalization methods on clinical data. One concern raised by Andersen et al. (2010) is that artificial lesion masks injected into healthy brain tissue may omit other pathological features resulting from stroke such as enlarged ventricles and widened sulci in perilesional areas. Considerations must be taken into account, such as matching the healthy imaging participants for age compared to their lesion "donors"; a point we address in the current experiment.

A novel aspect of the Crinion et al. (2007) paper is their performance metric that measured the average displacement of particular anatomical landmarks defined prior to normalization, rather than just relying on RMSD (which is measured across every voxel

in the brain). Choosing specific landmarks in each individual's image and then measuring their proximity after normalization (in template space) is far more informative than RMSD alone. RMSD can give the false impression of good normalization depending on the method used. Lower RMSD values when comparing the same brain with and without injury should indicate that the lesion had little effect and the result closely matched the gold standard, unlesioned brain. However, one could also observe low RMSD values in cases where the normalization found a poor solution in both images, or in cases where the normalization parameters are too constrained. In the current experiment, we extent the clever landmark displacement measure to the lesioned images, whereas Crinion et al. only measured it in their neurologically healthy control group.

Building on the promising results for automated tissue segmentation combined with image normalization, Seghier et al. (2008) modified the unified normalization procedure (Ashburner & Friston, 2005) by adding an additional tissue class that can be used to define abnormal tissue intensities (specifically tuned for stroke data). This tissue class is defined as the mean intensity of the known WM and CSF classes. In addition to this artificial tissue class, the segmentation algorithm is given additional constraints that control how tissue misclassifications are to be handled. For example, misclassified GM within the boundaries of the a priori WM tissue map from the template image will get reclassified as abnormal ("extra") if they have both lower than expected GM probability values, and low probability of being WM. By adding this extra tissue class for lesion identification, and modifying the segmentation computations the authors show moderate DICE coefficients (measure of similarity) between manually traced and automatically identified lesion maps. Importantly, this addition to SPM's unified normalization method

is able to output a lesion map, which can be useful for further VLSM analyses. While this modified segmentation procedure is promising, it also has some limitations. First, the tissue priors for GM, WM, and CSF are computed using subjects from the International Consortium for Brain Mapping (ICBM) which are younger than the typical stroke participant (Karnath, Berger, Küker, & Rorden, 2004). This could lead to age related tissue mismatch. Second, the average template images for the tissue priors were generated using linear only registration methods, which have been proven to be less precise in matching complex cortical folds (Klein et al., 2009). This could generate an oversimplified template for tissue priors. Lastly, the authors do not explicitly note that they achieved better normalization results after segmenting lesioned anatomical images using their enhanced method. One would assume that the results would be similar to previous tests of the unified segmentation normalization procedure, given that the actual normalization process was not modified in Seghier et al. (2008), but rather tissue segmentation was just enhanced. It is likely that most users simply use the automatically defined lesion tissue map as a mask to facilitate CFM normalization to a template as is traditionally done with these earlier versions of the SPM software.

In an attempt to definitively assess the efficacy of SPM's unified method on stroke neuroimaging data, Andersen et al. (2010) performed a similar analysis to Crinion et al. (2007) with the major difference of applying the method to real, unaltered data and simulated (artificially lesioned). The authors manually traced each participant's lesion both precisely (taking 1-8 hours per image), and roughly (taking 5-30 minutes per image). Using only the unified normalization method as implemented in SPM5, they compared voxel displacement results (RMSD) across conditions where the precise and

rough lesion masks were used in addition to a no mask condition (no CFM). They found that with the no mask condition where cost function masking was omitted, the RMSD was significantly higher compared to all conditions where CFM was used. This indicates that applying CFM to the unified method indeed reduces the normalization error, contrasting the results of Crinion et al., (2007). They did not see any significant differences between precise lesion masks and roughly sketched lesion masks, indicating that the precision of manual lesion tracing has little to no effect on the unified normalization procedure as measured by RMS displacement from a reference image. However, in practice it is perhaps wasteful to not make a precise lesion mask. It would be expected that rough lesion drawings would not have much specificity in a VLSM analysis. VLSM analyses are often of great interest to stroke researchers who use neuroimaging methods to infer brain-behavior relationships across a group of brain injured participants. Together, these results suggest that although a virtuous cycle of segmentation and normalization can enhance normalizing clinical neuroimaging data, CFM was still required for best performance in these earlier versions of SPM.

In a dual effort to further improve the results obtained using unified segmentation normalization with stroke datasets, and to provide an appropriately aged template image, Rorden et al., (2012) created the Clinical Toolbox. This toolbox is distributed as an add-on for the SPM neuroimaging software. The Clinical Toolbox includes an alternative template image to be used in the unified normalization process that was generated using an older sample of participants with a mean age of 72.9 years old. Participants in stroke studies are typically 60 years old or greater (Karnath et al., 2004; Karnath, Rennig, Johannsen, & Rorden, 2011), but are routinely normalized to the MNI152 template

generated from participants with an average age of 25 years old (SPM source documentation). This age related difference results in anatomical discrepancies present in older individuals but absent in younger individuals such as general widespread cortical atrophy, widening sulci, and larger ventricles (Rorden et al., 2012; Salthouse, 2011). By using a standardized template that accounts for general age related effects, perhaps a more accurate registration between a stoke participant and the age matched template can be computed. Indeed, Rorden et al. (2012) show that the aged template combined with unified normalization and CFM resulted in significantly reduced RMS displacement values compared to linear only normalization, and unified normalization using the MNI152 template without CFM. However, there was no statistical difference between the aged template combined with unified normalization with CFM, and the MNI152 template combined with unified normalization with CFM. In summary, within their sample there was no significant effect of template (older vs. younger) on normalization results when combined with CFM, but CFM resulted in less normalization error overall, further supporting Andersen et al. (2010). The average RMS (4.89 mm) for the aged template method was numerically lower than the RMS values computed using the standard MNI152 template (4.96 mm) indicating only a small performance boost, but perhaps far more validity given that similarly aged brains were compared. It is possible that the lack of statistical difference is due to the inherent smoothness of template images being generated from many individuals.

Again, the RMS performance metric may not accurately assess the quality of normalization across methods. It may be more applicable with a method, when testing variations of parameters. If a normalization does very little to change an image, then both

the reference deformation field and its comparison will be similarly displaced, resulting in lower RMS values overall. Therefore, a performance metric such as landmark displacement should be employed such as that used in Crinion et al. (2007).

Last, but certainly not least, is a non-SPM attempt to provide better normalization to stroke neuroimaging data. The method is known as LINDA (Lesion Identification with Neighborhood Data Analysis) and was developed and tested by Pustina et al. (2016). The authors created a new algorithm that automatically defines the features of lesioned tissue in a typical T1-weighted MRI scan and uses that lesion mask as an input to a high DOF normalization technique implemented in the Advanced Normalization Tools (ANTs) software package (Brian B. Avants et al., 2009). The method makes use of a trained Random Forrest classifier to predict lesioned tissue, which is then used as a mask when images are normalized to a template. In general, the motivation is the same as that of the enhanced unified segmentation method by Seghier et al. (2008) implemented in SPM: to segment injured tissue, and enhance normalization. LINDA uses six empirically defined features (selected as a subset of 12 original features) in the clinical image that are related to the image geometry, hemispheric asymmetry, and deviation from a template of control subjects. The LINDA method also makes use of an iterative register-predict-register cycle where both normalization to the standardized template and prediction of lesioned tissue increase in accuracy. The cycle is also carried out across image resolutions that increase in detail from earlier to later stages (e.g. from 6 mm voxel sizes to 2 mm). LINDA normalizes images to a template constructed from elderly individuals combined with patients that have diseases such as Parkinson's, mild cognitive impairment, and Alzheimer's (Pustina et al., 2016). Recall from Rorden et al. (2012) that the use of a well

matched template may result in reduced normalization error in some cases. LINDA differs from previously mentioned methods in that it utilizes the ANTs high DOF SyN (B. B. Avants, Epstein, Grossman, & Gee, 2008) registration method which has been demonstrated to be "best in class" on many datasets consisting of neurologically healthy individuals (Klein et al., 2009). LINDA is perhaps the most complex procedure to facilitate normalization of lesioned images, and automatically define the lesion without user input. However, it still shares some similarities with previously mentioned techniques, and it is not without its own limitations. To facilitate the normalization procedure using the ANTs algorithms, CFM masking is still a necessity, however constrained cost function masking is used (CCFM), which infers the deformations needed in the lesioned area based on surrounding tissue (Kim, Avants, Patel, & Whyte, 2007) rather than just omitting the lesioned area. LINDA also has similar limitations to the enantiomorphic normalization method. Namely, most use cases will be restricted to unilateral lesions. Furthermore, LINDA makes use of a pretrained classifier when predicting new data that the model has not been exposed to. The pretrained classifier is limited to left hemisphere injury, but the authors do state that right hemisphere injured images can be successfully normalized (with predicted lesions) if the image is first flipped across the axial midline (but then subsequently flipped to its original state after processing). The authors do not compare other normalization techniques against LINDA, but they do explicitly compare the method used by Seghier et al. (2008) to segment lesions, which in turn can affect normalization when that lesion map is used as a cost function mask. LINDA results in an increased match to manually traced lesions (DICE 0.696) compared to Seghier et al.'s (2008) enhanced unified segmentation technique

(DICE 0.44). This indicates a significant improvement in automated lesion segmentation (and has a smaller failure rate as measured by the inability to detect a lesion or poor normalization). Lastly, VLSM results from lesions identified using LINDA had similar statistical maps compared to hand drawn lesions (DICE of 0.6), but differed as a function of the behavior being assessed. For example, the peak statistical scores from each analysis (manual vs. automated lesion tracing) for the behavioral measure of auditory comprehension differed by 64 mm. Statistical peaks related to other behavior analyses differed as little as 5 mm. From this data alone, it is difficult to discern if some extreme spatial differences such as 64 mm are related to variance in lesions masks, or any number of interactions between lesion size, position, and behavioral variability as a factor of "mask type" (manually traced vs. automatically identified).

1.5 THE CURRENT APPROACH

To date, there is no universally accepted method to normalize clinical neuroimaging data in the stroke research community. However, the view expressed in the current article is that there may never be a method that performs well across the board of all performance metrics. Here we test multiple lesion compensation techniques within each method to provide evidence of each technique's effect on the normalization procedure. We used a variety of performance metrics across all normalized images. Each method has its limitations, and corner cases where it may fail altogether (e.g. incorrectly set origin coordinates), but there are solutions to the clinical image normalization problem such as those mentioned above. Practitioners in the field have no updated comparison of normalization methods that directly measures the performance of each contemporary method on the same data. For example, no previous comparison study has

included LINDA since its publication in 2016. Additionally, no comparison studies on stroke image normalization have included the newly updated SPM12 software. In the most recent study (Ripollés et al., 2012) similar in design to our current experiment, SPM8 was used. Since that time, many improvements have been released in SPM12 that dramatically change normalization results from the unified segmentation method. Notably, the most recent version of SPM12 now performs an implicit lesion filling step on segmented images (SPM12 change log, Oct. 3, 2017), with the filled in regions derived from template tissue probability values. This is quite a radical change to the traditional CFM procedure. Now, zeroed out regions are implicitly filled in. This recent version of SPM12 directly rivals the enantiomorphic lesion filling method.

Comparisons of multiple methods to accomplish the same goal are crucial to establish evidence for "best practices" within a field. There is a clear trend in neuroimaging (including clinical data) to develop and validate fully automated image processing pipelines. Very few methods exist that fulfill this desire, and the ones that do must be put to the test when considered for inclusion in future investigations. For this reason, we have specifically included the LINDA method in our comparisons. Many projects that make use of hundreds or thousands of datasets would benefit greatly from using state-of-the-art processing methods to normalize neuroimaging data from brain injured participants. In some cases, using the "best in class" methods can even enhance anatomical specificity when answering neuropsychological questions (Crinion et al., 2007; Pustina et al., 2016). Additionally, we can combine multiple techniques that may result in more accurate normalizations, and in turn greater matching of anatomy among groups of images containing injured regions. One such novel combination is an

enantiomorphic version of LINDA which has been developed for the current experiment and tested for the first time in this series of comparisons. Our aim is to provide readers with empirical evidence of the effect each technique has on normalization of clinical images.

The current article will compare the effects of brain injury on normalization methods using a variety of free, and commonly available tools. Since the most recent comparison of this kind (Ripollés et al., 2012), many advances have been made to some software packages commonly used. Each software package has its own injury compensation technique. However, typically there is only one option available to users. Here, we test each software's normalization routines with their built-in lesion compensation strategy, in addition to a nonstandard lesion healing technique referred to as enantiomorphic normalization (Nachev, Coulthard, Jäger, Kennard, & Husain, 2008). That seminal work only described a proprietary implementation, and used a normalization method that is no longer considered state of the art (indeed it is referred to as "old Norm" in SPM12). Here we introduce an advanced open source implementation of this enantiomorphic technique that can be applied to any normalization tool. This allows us to test whether this novel healing method aids modern normalization. This work provides insight to which methods provide the most value to clinical neuroimagers.

CHAPTER 2

METHODS

2.1 PARTICIPANTS

Two sets of archived images from participants were included in this study. One group provided neuroimaging control data, and the other group of stroke survivors provided the lesion data. The control group consisted of 57 healthy participants (47 female; mean age 55 years old; range 40-69 years old) that participated in a separate neuroimaging study at the Medical University of South Carolina (MUSC). The control group were recruited based on criteria that they were at risk for stroke. The stroke group consisted of 177 participants (69 female; mean age 60 years old; range 29-83 years old; mean months post stroke 36.2). Image data from the stroke group was collected between 2006-2018 at the University of South Carolina (USC), and MUSC. All participants were informed of study procedures via approved Institutional Review Board documents and consented to participation.

2.2 MRI DATA ACQUISITION

Data for healthy control participants were acquired using a 3 Tesla Siemens Trio MRI system with a 12-channel head coil. Whole brain T1-weighted (T1w) 3-Dimenstional magnetization-prepared rapid gradient echo (MPRAGE) scans were

acquired for each individual with the following parameters: flip angle = 9°, TR = 2250 ms, TE = 4.18 ms, voxel resolution = 1.00 mm isotropic, field of view = 256x256 mm, slices = 176, GRAPPA = 2.

Data for stroke participants were acquired using a 3 Tesla Siemens Trio MRI system with a 12-channel head coil or a 3 Tesla Siemens PRISMA MRI system with a 20 channel head coil. On both systems, volumetric T1 and T2 weighted (T2w) MRI scans were obtained. The parameters for the Trio system are listed first. Settings for the T1w images were as follows: flip angle = 9°, TR = 2250 ms, TE = 4.15 ms, voxel resolution = 1.00 mm isotropic, field of view = 256x256 mm, slices = 192, GRAPPA = 2. T2w scanning parameters were: 3D SPACE, voxel resolution = 1 mm isotropic, field of view = 256 x 256 mm, 160 sagittal slices, variable flip angle, TR = 3200 ms, TE = 352 ms, with no slice acceleration. Slice center and angulation were similar to the T1 image sequence.

For the PRISMA system (stroke participants), the T1w images were acquired with the following parameters: flip angle = 9°, TR = 2250 ms, TE = 4.11 ms, voxel resolution = 1.00 mm isotropic, field of view = 256x256 mm, slices = 192, GRAPPA = 2. T2w scanning parameters were: 3D SPACE, voxel resolution = 1 mm isotropic, field of view = 256 x 256 mm, 160 sagittal slices, variable flip angle, TR = 3200 ms, TE = 567 ms, GRAPPA = 2. Slice center and angulation were similar to the T1 image sequence.

2.3 SOFTWARE IMPLEMENTATIONS

In order to accomplish the task of normalization, neuroimagers may choose to use a single software package, or even combine software packages and techniques across programming languages to suit their needs. There are many software packages and

techniques available, but the scope of this paper was limited to the selection of choices that have been routinely used in clinical research on brain damaged individuals with evidence of performance from previous comparisons. The complete listing of software used is: FLIRT, SPM12 old normalize with CFM, SPM12 old normalize with enantiomorphic lesion filling, SPM12's unified method with tissue map filling, SPM12's unified method with enantiomorphic lesion filling, SPM12's unified method with DARTEL combined with masked tissue map filling, SPM12's DARTEL unified method combined with enantiomorphic lesion filling, ANTs with LINDA and CCFM, and ANTs with LINDA and enantiomorphic lesion filling.

In order to generate the artificial lesions we used a different normalization method than any being subsequently evaluated. The intention was to avoid a specific bias toward one of the methods. Specifically, we used FLIRT (FMRIB's Linear Registration Tool) as the independent method (Jenkinson & Smith, 2001) to perform linear registration at the dataset creation stage. The use of FLIRT also avoids introducing unintended nonlinear deformations prior to the assessment of methods.

All image processing was carried out on a high performance computing cluster (HPC). This clustered computing system enabled highly parallel processing for all normalization methods. In total, 10 computed nodes with 28 cores each were used enabling access to 280 CPUs. In most cases 2 CPUs were assigned to a specific subject to be processed, meaning that if all available CPUs were utilized then 140 subjects from the artificial dataset could be processed at a time. The speed benefit of parallel computing is critical to this project. If the same dataset were analyzed in serial on a typical laptop with 2 CPUs the processing would take many months (nearly a year). Here, we were able to

process over 100,000 thousand images (~10,000 images x 10 normalization methods) in the timespan of one week. Additionally, all performance metrics were computed in parallel on the HPC. See Appendix A for an example a python script used to submit individual subject "jobs" to the HPC SLURM scheduler.

2.4 DATASET CREATION

Similar to other comparison studies (Andersen et al., 2010; Brett et al., 2001; Crinion et al., 2007; Ripollés et al., 2012), we assessed the performance of each normalization technique on brain images with artificially injected lesions. A dataset of artificially lesioned images affords some advantages. In particular, we can directly compare normalization performance with and without a lesion within the same individual. This is not possible in data from participants where the lesion has occurred naturally, and no pre-injury scan exists. Furthermore, we can generate datasets that are orders of magnitude larger, since we can apply artificial lesions from every stroke participant's image to every control participant's image. If all combinations are exhausted, datasets that began with hundreds of images will combine to create thousands of images. This can enhance our estimations of normalization performance metrics (additional variability is introduced by the unique combinations of injected injury to control participant brain images), although many images will end up containing similar features (e.g. the same lesion placed in all healthy control images will now make them uniquely related as well).

The artificially lesioned brain images were created using similar procedures to previous studies (Brett et al., 2001; Crinion et al., 2007). However, some novel methods were incorporated. Importantly, the particular application of methods used for dataset

creation were not included in the subsequent comparison of techniques. This ensures that no normalization method benefits more because of its similarity to the toolset used to create the images. All processing (except for manual lesion tracing) was automated with custom developed Python functions that interact with multiple additional programs. First, lesions masks were manually traced by experienced neuroimagers on each stroke participant's T2w image. Stroke related injury is often more apparent in these images. Next, each lesioned image, as well as the control image was reoriented to the stereotaxic coordinate system defined by the MNI152 template using FSL's "reorient2std" program. This reorientation does not perform any calculations or data interpolation, but merely ensures all images conform to a standardized data organization format and orientation. The T1w lesion image and T1w control image are then cropped in the z (head-foot) direction to remove extraneous non-brain tissues (e.g. neck) using FSL's "robustfov" program. Next, the lesioned T2w image is registered to the lesioned T1w image using FSL's FLIRT program with default parameters (Jenkinson & Smith, 2001). This registration is applied to the manually traced lesion mask generated from the T2w image and results in a mask that now conforms to the space defined by the T1w image. Due to interpolation effects, the lesion is then re-binarized using a threshold of 0.5. Next, the lesion mask is smoothed using a full-width half maximum (FWHM) of 3 mm. Smoothing generates a gradient from 1 to 0 only near the edges of the mask (feathering). The lesion mask is then mirrored to the undamaged hemisphere of the lesioned T1w image and the enantiomorphic lesion filling method (Nachev et al., 2008) is used to "heal" the lesioned T1w image (this procedure differs from the one used prior to normalization and is outlined below). This step is crucial, and strays from previous methods used to create

artificially lesioned brain images. The enantiomorphic healing is necessary to facilitate a decent brain extraction (e.g. skull stripping) of the lesioned T1w image. Brain extraction is carried out using FSL's "bet" (Smith, 2002) with a fractional intensity value of 0.4 which is more conservative than the default value, thus preserving more true brain signal. The same brain extraction is also carried out on the control T1w image. Then, the lesioned T1w image (brain only) is linearly (12 DOF) registered to the control T1w brain using trilinear interpolation with FLIRT. By removing non-brain tissue prior to this registration step, we can ensure a more accurate brain-to-brain registration since the cost function will be driven by intracranial tissue instead of non-brain tissue (e.g. skull). These registration parameters are then used to register the lesion mask to the control T1w image space as well. The brain-to-brain registration parameters are also used to then register the whole head lesioned T1w image to the space of the control T1w image. Next, both T1w images are matched for intensity using mean scaling so that when tissue from the lesioned imaged is placed into the control image the result will not be abnormally dark or bright. The lesion mask is then smoothed using a FWHM of 8 mm, which facilitates a gradual blending of tissues near the edges of the lesion when it is injected into the control image. This large smoothing value also allows us to include portions of enlarged ventricles when they are present, further making the artificially lesioned brain more representative of the effects related to a natural lesion. Lastly, tissue from the lesioned T1w image within the mask is placed into the control T1w image, and blended via the gradient present near the edges of the mask.

Together, all the artificial lesioning steps are performed using an exhaustive pairwise combination of control images and lesioned images to generate a dataset of

more than 10,000 artificially lesioned images. The resulting dataset contains whole head images (including skull) with real lesions "donated" into a cohort of age matched, neurologically healthy control participants. All images were processed identically, and with minimal interpolation. Therefore, there should be no systematic difference introduced in the artificial dataset prior to comparing each normalization method and lesion compensation technique.

2.5 SPM12 ENANTIOMORPHIC LESION HEALING PROCEDURE

Across all methods and participant images (both real and artificially lesioned) the enantiomorphic lesion filling process was identical. Each "healed" T1w image was created using the following steps in SPM12. First, voxels in the T1w anatomical scan within the lesion mask were set to a value of zero. Then the T1w image was segmented using SPM12's segmentation routines. The segmentation produced native space tissue maps for gray and white matter (GM, WM). Next, the GM and WM tissue images were summed, and regions within the native space lesion mask were given a value of one. This combined tissue map image, and the original T1w image were then left-right flipped and saved as new NIFTI files. The flipped images were then registered to the original image so that they were midline aligned using SPM12's Old Norm nonlinear registration procedure with trilinear interpolation (Ashburner & Friston, 1999). All other Old Norm settings remained at their default values. After the flipped images were registered to the original, the lesion mask was smoothed with a 4mm kernel and binarized using a 0.05 threshold (5%). The smoothed mask was added to the original resulting in a slightly dilated new mask image. This new image was smoothed one final time with an 8mm kernel allowing for the edges to be feathered, which enabled a gradual blend in the

voxelwise multiplication step of filling in the lesioned area with tissue from the flipped image (e.g. the opposite hemisphere). Source code is available at https://github.com/neurolabusc/nii_preprocess. Note that this open source solution extends the proprietary method described by Nachev. In particular, the human brain is not perfectly symmetrical, and some individuals have very asymmetric brains. Here we use a non-linear method that can match tissue even in these cases. See Figure 2.1 for an example applied to an asymmetric brain and Figure 2.2 for an example where we apply the method to one of the artificial images in our data set.

2.6 NORMALIZATION PROCEDURES OF LESIONED IMAGES

Normalization was performed using SPM12's Old Normalize procedure, SPM12's unified segmentation normalization, SPM12's DARTEL, and with ANTs SyN in combination with LINDA. All normalization methods were assessed using enantiomorphic lesion filling. Other lesion compensation techniques assessed included SPM12's tissue probability map lesion filling, or CFM with SPM's Old Normalize, or CCFM with the ANTs SyN method used by LINDA. In total, 10 normalization methods with unique lesion compensation strategies were used. Within all normalization methods, enantiomorphic lesion filling was compared to each particular method's implemented default alternative. See Table A for a summary of each software implementation and its main parameters.

Using SPM12's Old Normalize procedure, each lesioned T1w image was input to the algorithm along with the corresponding weighting image. The weighting image was constructed by taking the binary lesion mask and subtracting 1, then multiplying by -1. This converts all zeros to a value of 1 and all ones to a value of zero. The weighting

image is necessary for the traditional CFM technique implemented in SPM12's Old Normalize. Each image was aligned to the MNI152 T1w template distributed with SPM12. All other settings were kept at their default values. The transformation parameters from the normalization were saved, and subsequently applied to normalize the lesion mask as well. In the case of enantiomorphically healed images, no weighting image was supplied to the algorithm, but all other settings were the same as used with the CFM procedure. The normalization parameters were used to create a deformation field image for each technique, which saves the displacement needed to move each voxel from its native space image to the template is was warped to (in units of millimeters).

SPM12's unified segmentation normalization procedure was used in conjunction with the age appropriate template from the Clinical Toolbox (Rorden et al., 2012). Until recently (October 3rd, 2017) CFM was achieved in the unified procedure by zeroing out tissue in the anatomical scan (T1w) delineated by a lesion mask (SPM12 change log: https://www.fil.ion.ucl.ac.uk/spm/download/spm12_updates/README.txt). Once segmented, these resulting tissue maps would contain zeros at the location of the lesion. However, after this date, areas of the anatomical scan that have been zeroed out using a lesion mask are now filled in with the tissue values from the a priori tissue probability map (TPM) best representing the tissue of the missing anatomy. In the case of our experiment, these filled in values were taken from the tissue maps distributed with the Clinical Toolbox. Since this recent enhancement to SPM's segmentation procedure, we cannot necessarily refer to the method of accounting for the lesioned area as CFM. Instead, we will refer to it as the "TPM filled" method, which rivals the older enantiomorphic lesion filling method (See Figure 2.3 for an example). Unified

segmentation normalization was performed with the now standard TPM filled method, and the enantiomorphic method. We also analyzed the data with both high regularization and medium regularization (default). For high regularization, the nonlinear warping parameters were two orders of magnitude higher than the default values, which is in line with previous work (Crinion et al., 2007). All other settings were kept at their defaults. Each normalization produced a normalized T1w image and its corresponding normalized lesion mask. Additionally, each normalization produced a deformation field used in subsequent analyses.

SPM12's DARTEL (Ashburner, 2007) was used to normalize each image in the dataset as well. Each lesioned image was either "healed" prior to DARTEL operations using the enantiomorphic method described earlier, or filled in with TPM values resulting from the prior step of unified segmentation (the default method). DARTEL requires that images first be processed using the unified segmentation routines included in SPM. Here, we use the age appropriate tissue probability maps included in the Clinical Toolbox to facilitate segmentation (and TPM filling) in SPM prior to running the DARTEL procedure. A study specific template was first created in DARTEL for subsequent registrations. The study template was created from the same neurologically healthy subjects defined in Rorden et al. (2012), which were used to create the age appropriate template distributed in the Clinical Toolbox. Each subject's segmented images (WM and GM) are then registered to the study template using DARTEL's "warp to existing template" routines. We used the same registration parameters as Ripollés et al. (2012), who performed a parameter search to find optimal values for DARTEL since it had previously not been investigated. DARTEL also creates deformation fields that map the

world based transforms needed to move each voxel to its registered position in the template (measured in mm). Similar to the other methods listed, the deformation fields are used in subsequent analyses to measure performance across techniques.

The final normalization method assessed was SyN as part of ANTs in combination with LINDA. LINDA is the only method included that is advertised by its creators as a fully automated lesion segmentation and normalization procedure. First, LINDA defines parts of the lesioned image as lesion or not (no mask needed). Then, a register-predict-register cycle is carried out using higher resolution versions of the images in later cycles. LINDA makes use of the ANTs SyN normalization method, which shares some similarities to DARTEL. Since LINDA is a fully automated, and only requires that the user supply a T1w image, all processing steps are carried out without user interaction. This includes warping the identified lesion to template space. Here, we simply supply input images to LINDA's algorithms. Note that all lesioned images in our dataset conform to the requirements of LINDA. Namely that lesions are unilateral, limited to the left hemisphere, and are a result of stroke. This is relevant given that LINDA uses a pretrained lesion classifier based on left hemisphere stroke which is a core component of its register-predict-register cycle. In the case of enantiomorphically healed images, LINDA's lesion segmentation stages were skipped, and SyN normalization was used without CCFM.

2.7 PERFORMANCE METRICS

The performance of each normalization method is determined by multiple factors including the voxelwise root mean square displacement (RMSD), the displacement of particular anatomical locations, the normalized cross correlation (NCC) within the lesion

and whole brain, and the prediction accuracy when the normalized brain images are supplied to support vector regression (SVR) classification analyses. For each performance metric a repeated measures analysis of variance (ANOVA) was performed to compare overall main effects. Pairwise comparisons were also conducted between enantiomorphic lesion filling and the alternative injury compensation technique within each normalization method. Lastly, prior to the ANOVA and paired comparison procedures, the data from the artificial dataset were aggregated across "patients". This simply means that since each patient's injury was represented in each control subject's image, the aggregating procedure created an average "artificial patient". This reduced the complexity of the analyses since the artificial dataset inherently created data typically analyzed using a mixed design. The aggregation was chosen to reduce the variability induced within each representation of a patient's injury. Finally, the aggregated dataset more closely resembles a realistic dataset of a typical stroke study.

The RMSD metric is a single value that summarizes the average difference in voxel displacement from one normalization to the next. Here we use it to measure the amount that a lesion perturbs a normalization. Specifically, we compare the normalization of a healthy individual's brain to the normalization of the individual's brain with an artificial lesion. If the lesion has no influence, all voxels will be warped to precisely the same location, so the RMSD measure would be zero. On the other hand, if the presence of the lesion disrupts the process, the voxels will be warped to different locations. Voxel displacement is measured in world space (millimeters). The RMSD is therefore a value describing the Euclidean distance of a voxel's new position between images normalized with and without a lesion. Since this is a measure of distance

(displacement in mm) it is a metric considered separate from the cost function used to compute where that voxel was move to. In practice, RMSD measurements are impacted more by nonlinear regularization, rather than cost function used. To assess RMSD, each control participant's brain image is normalized with, and without lesions inserted. The difference in amount of displacement between the lesioned, and unlesioned versions of the voxelwise deformation fields measures the impact that the lesion had on the normalization algorithm. A perfect solution would result in a value of zero, meaning the lesion had no influence on a match to the template image (compared to the unlesioned reference). Therefore, lower RMSD values within a method are clearly desirable. However, comparisons across methods may be heavily misleading. Specifically, good normalization routine that matches local features should be expected to be more influenced by lesions than a coarse normalization that ignores these features. Therefore we compare within each method between the available lesion compensation strategies.

Similar to Crinion et al., (2007), the location of easily identifiable anatomical landmarks were recorded in each image from the neurologically healthy group in order to measure how well those particular points are aligned using each normalization method. We identified 18 landmarks across both cortical and subcortical anatomy. The complete list is: anterior commissure, left/right temporoparietal junction, posterior commissure, left/right anterior tip of the later ventricle, left/right posterior tip of the lateral ventricle, left/right calcarine fissure, left/right inferior frontal gyrus (pars obitalis), left/right central sulcus (superior portion near motor hand area), left/right anterior cerebellar lobe, and left/right inferior colliculus. First the origin of each neurologically healthy image was set to the anterior commissure using SPM12. Then, each easily identifiable landmark was

identified by a trained lab member, and verified for accuracy by a second member. Landmark identification was carried out using the MRIcron software (Chris Rorden & Brett, 2000). A three-dimensional coordinate point (x, y, z) corresponding to millimeters from the image origin is saved for each landmark. These millimeter coordinates are then converted to voxel indices and a 3x3x3 voxel cube is generated at each point and saved as a NIFTI image file with the same characteristics as the image the points were derived from. Each cube in this image was given a value ranging from 1..18 to indicate which landmark label it corresponded to, and all other voxels are given a value of zero. Since the landmarks were defined in each participant from the neurologically healthy group, we can measure their displacement regardless of the fact that lesions have been artificially injected. Computed normalization parameters from each normalization method (per participant) are applied to these saved landmark images using nearest neighbor interpolation which resulted in landmarks in the space of the template image. Once in template space, the average center of mass of each landmark cube was recorded as the average of each landmark's x, y, and z indices in the normalized image. This resulted in 18 sets of average coordinates (centroids). Similar to RMSD, normalization performance in the lesioned dataset was measured across methods as the average Euclidean distance (in mm) of each landmark from its associated centroid. Normalization methods that result in a smaller spread of these points (and therefor decreased distance from the centroid) are objectively better at aligning anatomy across individuals.

The next metric, NCC, was computed both across the whole image, and within the average, normalized lesion mask per participant and method. For each input image, the mean was subtracted from each voxel prior to computing the NCC coefficient. The NCC

coefficient measures the similarity between two images, or in our case also between masked parts of a images. It is important to note that NCC itself is a common cost function, and was indeed used in ANTs with LINDA in the current study. Therefore, this performance metric can be confounded when the method used to register images is related. However, when compared within a method, this confound is not present since we only measure the effect of lesion compensation strategy. When measuring performance higher NCC scores indicate increased similarity, with a perfect score being 1. Average lesion masks for each lesion were created by summing all normalized versions of the binary lesion masks for each image created with the same patient, then dividing by the total number of images. Once averaged, this final mask image was then re-binarized using a threshold of 0.5. We obtained one average lesion per patient, per method. This ensured that the same mask was used to compute the within mask NCC score for each particular lesion image (e.g. from all images of an artificial patient) per method (reduces variability in comparisons), but also preserved the unique characteristics (accuracy of registration) inherent to each normalization method (e.g. average lesion s from SPM Old Norm are not identical to average lesions from SPM unified segmentation). Finally, the NCC metric is computed on pairs of images, so all combinations were exhausted for each method in both control participant images and the artificially lesioned images.

The final metric of normalization performance is the predictive ability each normalized lesion set had when estimating behavioral scores from normalized lesion masks in real patient images. We used support vector regression (SVR) as implemented in LIBSVM (Chang & Lin, 2011; Smola & Schölkopf, 2004). We used a subset of 159 participants from the patient group for which we had recorded behavioral data. All

participants presented with language deficits as a result of left hemisphere stroke. Trained speech language pathologists assessed each patient using the Western Aphasia Battery (Shewan & Kertesz, 1980). Among other scores, this battery of tests produces a score referred to as the Aphasia Quotient (WAB AQ). The AQ score is a continuous value and indicates the severity of language impairment (in this case resulting from stroke). We predicted each participant's AQ score using an SVR procedure similar to (Yourganov, Fridriksson, Rorden, Gleichgerrcht, & Bonilha, 2016). Specifically, we used a linear SVR kernel on normalized patient lesion masks where predictors consisted of voxels containing either a 1 or 0 for lesioned or unlesioned tissue in the mask images. Each normalization method was assessed separately using the SVR procedure. Model accuracy was evaluated using a leave-one-participant-out cross-validation procedure. The procedure set aside one participant, and used the lesion masks from all remaining participants to estimate the model coefficients. Then, the left out participant's AQ score was predicted based on the trained model that consisted of all other participant's lesion mask images and their known AQ scores. AQ scores were scaled to the range of 0..1, and the left out participant's score (the test case) was scaled to the values of the training cases on every iteration. The leave one out prediction procedure was repeated iteratively, leaving out a new participant each time. SVR prediction accuracy was measured as the Pearson correlation coefficient between actual and predicted WAB AQ scores for each patient (Smola & Schölkopf, 2004; Yourganov et al., 2016). In addition to the Pearson correlation coefficient, the residual values resulting from the SVR procedure were compared in a repeated measures ANOVA. Since the residuals are a measure of error, normalization methods that result in lower residual values for behavior prediction would

be more desirable. Similar to Ashburner, (2007), the idea is that better image normalizations should provide better prediction accuracies in neuroimaging data.

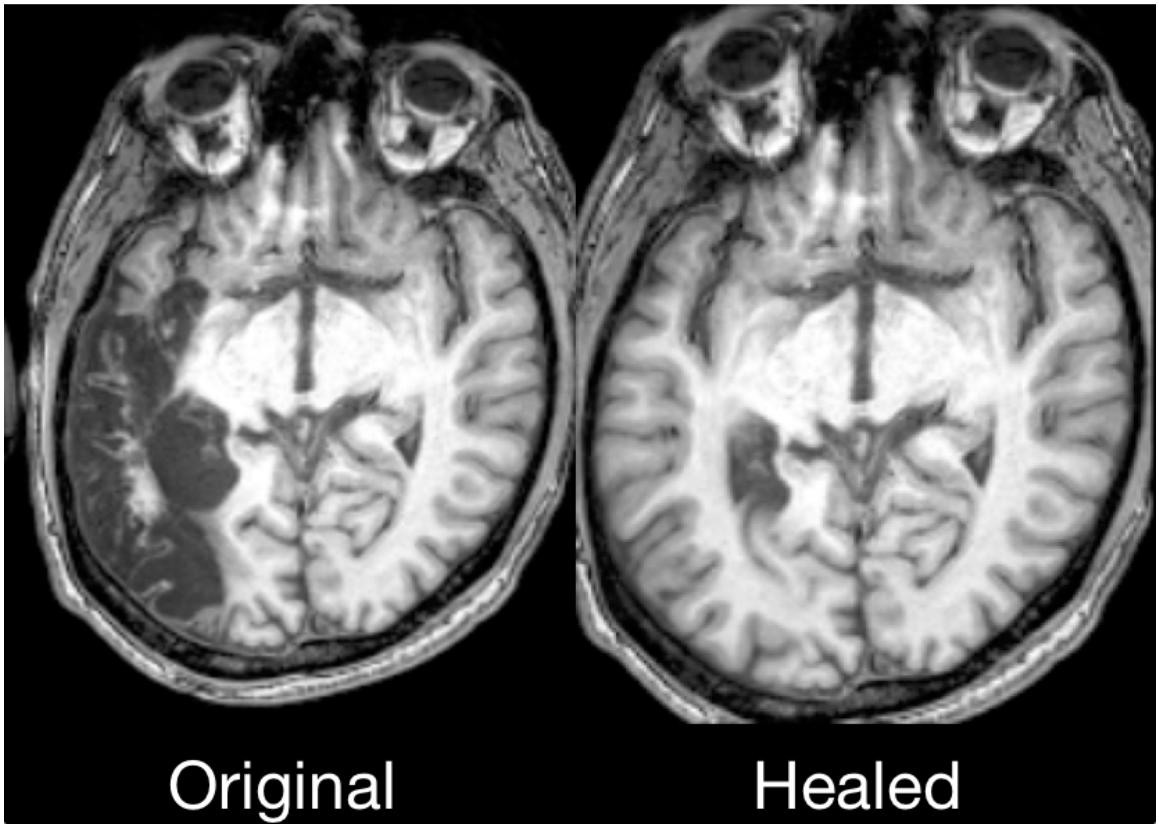

Figure 2.1: Enantiomorphic healing applied to an asymmetric brain. This is an improvement upon Nachev's (2007) method.

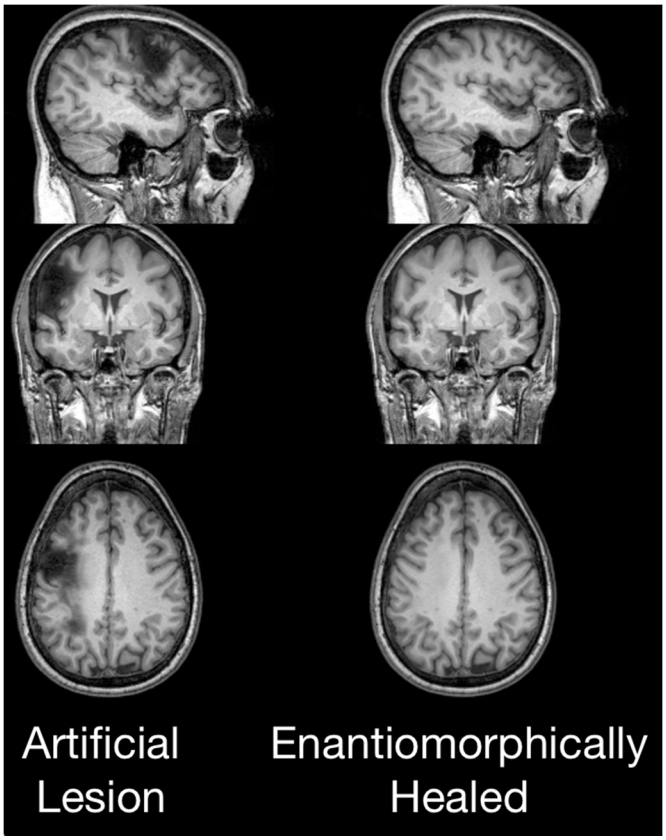

Figure 2.2: Example of the Enantiomorphic healing method applied to one of the artificial images in our data set

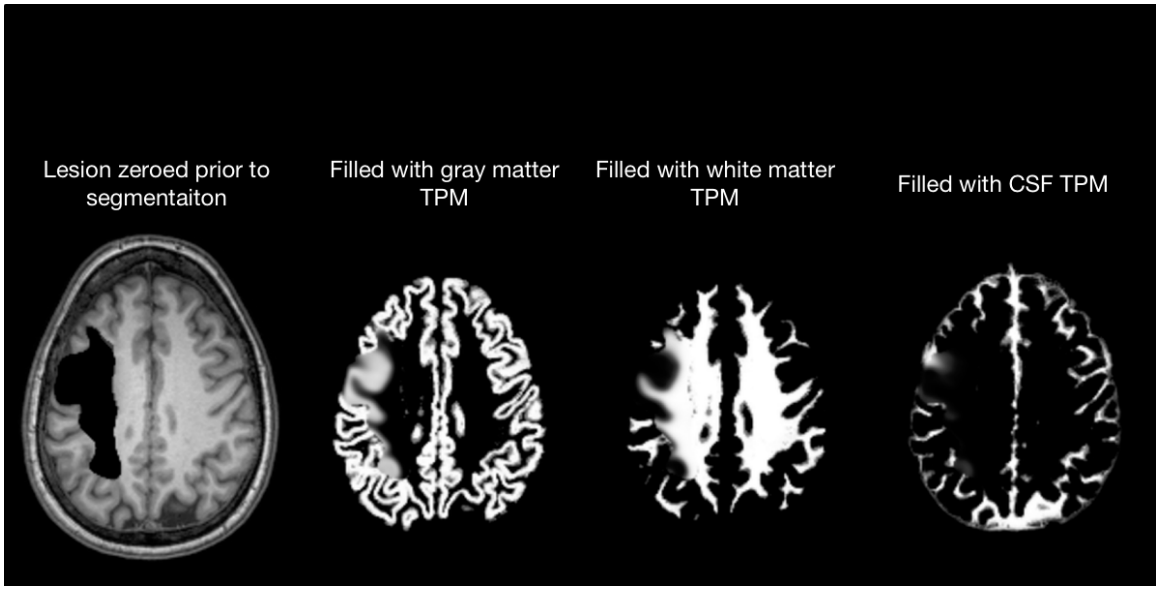

Figure 2.3: Example of SPM12's new tissue probability map filling method in segmented tissue.

CHAPTER 3

RESULTS

3.1 LANDMARK DISPLACEMENT

Repeated measures ANOVA analysis with Greenhouse–Geisser correction on landmark displacement scores showed a significant effect of normalization method F (3.96, 696.13) = 7276.99, $p < .001$, $\eta_p^2 = .967$. See Figure 3.1 for an illustration of each method's average landmark distance from the centroid (averaged across all landmarks). Overall, SPM12 DARTEL ENAT and ANTs LINDA CFM had the lowest average displacement from landmark centroids ($M = 4.86$, $SD = .11$; $M = 4.98$, $SD = .13$ respectively). SPM12 with high regularization (both with TPM lesion filling, and enantiomorphic lesion filling) resulted in the highest landmark displacement ($M = 6.14$, $SD = .05$; $M = 6.05$, $SD = .06$ respectively) indicating less precision in matching landmarks across images. Pairwise comparisons within each normalization method between enantiomorphic and the alternative lesion compensation technique showed significant differences in all pairs, but not always in the same direction. Within each method, enantiomorphic normalization outperformed the alternative (lower landmark displacement), with the exception of ANTs

LINDA with CCFM, which outperformed ANTs LINDA with ENAT. See Table 3.1 for the paired comparison statistics.

3.2 ROOT MEAN SQUARED DISPLACEMENT (RMSD)

Repeated measures ANOVA analysis with Greenhouse–Geisser correction on RMSD showed a significant effect of normalization method F (2.59, 454.96) = 870.58, p < .001, η_p^2 = .83. Figure 3.2 shows each method's average RMSD. The normalization methods with the lowest average RMSD values were SPM12 unified segmentation with high regularization using the TPM filling procedure (M = .286, SD = .17), along with both versions of SPM12 Old Norm which used ENAT and CFM methods to compensate for the lesion (M = .3, SD = .31; M = .34, SD = .22 respectively). The methods with the highest RMSD values were ANTs LINDA using CCFM (M = 1.41, SD = .11) and SPM12 DARTEL using the TPM filling procedure (M = 1.08, SD = .48). Pairwise comparisons within each method between enantiomorphic and the alternative showed significant differences in all pairs. Specifically, SPM12's Old Norm using ENAT lesion filling resulted in significantly lower RMSD values compared to traditional CFM (replicating Nachev et al., 2008). However, using SPM12's most recent unified segmentation method (with TPM lesion filling) resulted in significantly lower RMSD values compared to unified segmentation with enantiomorphically healed lesions. Finally, enantiomorphic lesion healing resulted in significantly lower RMSD values for SPM12 DARTEL and ANTs with LINDA. The observed reduction of mean RMSD resulting from enantiomorphic lesion healing was most predominant in DARTEL and ANTs, which both make use of millions of degrees of freedom and diffeomorphic image registration. All comparison statistics are detailed in Table 3.2.

3.3 NORMALIZED CROSS CORRELATION (NCC)

The NCC score was computed across methods for the control participant images and the artificially lesioned images. NCC scores were computed both within the average lesion mask per lesion (excluding the rest of the image), and within the whole image (including the lesion). Each image in the set was compared to every other image in the set (control and artificial lesion images were separate sets). For the control participant images without using lesion masks (whole image NCC) a repeated measures ANOVA analysis with Greenhouse–Geisser correction was used. Overall, there was a significant effect of normalization method F (1.89, 3015.5) = 12397.4, $p < .001$, $\eta_p^2 = .89$. ANTs resulted in the highest NCC score within control participants (M = .97, SD = .01), followed by SPM DARTEL (M = .81, SD = .07). SPM12 unified with high regularization (and TPM lesion filling) obtained the lowest NCC score (M = .67, SD = .08). Post hoc comparisons were significant across all pairs. Post hoc comparisons are reported in this case because there are no lesion compensation methods to compare within the control participants only. See Table 3.3 for detailed pairwise statistics, and Figure 3.3 for the illustration of each method's performance.

Again, repeated measures ANOVA analysis with Greenhouse–Geisser correction was used to analyze NCC scores from control participants, but this time within the average lesion masks per participant. There was a significant effect of normalization method on NCC scores within the mask region F (1.76, 289.9) = 975.36, $p < 0.001$, $\eta_p^2 = .855$. Figure 3.4 shows each method's average NCC score for the masked region within control participants. The masked NCC scores were generally lower compared to unmasked scores, but overall, the pattern of results closely resembled that of the

unmasked control participant NCC scores. There was an exception to this pattern of results among ANTs LINDA both with CCFM and ENAT normalized lesion masks (both obtained lower scores within the lesioned region). SPM12 DARTEL with the unified TPM lesion filling had the highest average masked NCC score within the control participants (M = .85, SD = .03), followed by ANTs LINDA CCFM (M = .67, SD = .10). The methods with the lowest NCC scores (nearly identical) were SPM12 unified with high regularization, with both the unified TPM lesion filling, and enantiomorphic filling compensation methods (M = .52, SD = .07; M = .52, SD = .07 respectively). Pairwise comparisons within each normalization method between enantiomorphic and the alternative lesion compensation technique showed significant differences between SPM12 unified default regularization with TPM lesion filling and ENAT lesion filling ($t(165) = 11.36$, $p < .001$), SPM12 DARTEL with TPM lesion filling and ENAT lesion filling ($t(165) = 11.17$, $p < .001$), and finally ANTs LINDA CCFM and ANTs LINDA ENAT ($t(165) = 17.2$, $p < .001$). There was no significant difference among SPM12 Old Norm or SPM12 unified with high regularization between the lesion compensation methods used (See Table 3.4).

For the artificially lesioned dataset, NCC scores were also computed both within the lesion mask, and across the whole image (inclusive of the lesioned region). There was a significant effect of normalization method on NCC scores for the whole image comparison $F(2.02, 354.87) = 311097$, $p < 0.001$, $\eta_p^2 = 1$ (repeated measures ANOVA analysis with Greenhouse–Geisser correction). The results closely resemble those of the control participant whole image NCC analysis. Overall, ANTS LINDA and SPM DARTEL had the highest average NCC scores indicating a better average match among

all normalized images, while SPM12 unified with high regularization had the lowest average NCC score for the whole image comparison (see Figure 3.5). The pairwise comparison between lesion compensation techniques within each normalization method showed significant results, with the exception of SPM Old Norm, for which there was no significant difference between techniques. For the other methods, enantiomorphic lesion filling outperformed the alternative techniques. However, the effect size of the increased NCC score is small across all comparisons.

Within the artificially lesioned dataset a significant effect of normalization method was also found for the masked comparison of NCC scores $F(1.61, 265.77) = 278.5$, $p < 0.001$, $\eta_p^2 = .63$. Again, these results aligned closely with those of the masked control participant analysis. SPM12 DARTEL with TPM filling had the highest NCC score (M = .81, SD = .05), followed by SPM12 Old Norm with ENAT (M = .8, SD = .06) and SPM12 Old Norm with CFM (M = .79, SD = .06). The methods with the lowest average NCC scores were ANTS LINDA ENAT (M = .65, SD = .09), and ANTS LINDA with CCFM (M = .69, SD = .11). Pairwise comparisons between lesion compensation techniques within each method showed significant differences within all pairs, but the direction of difference varied. For both SPM12 DARTEL and ANTS LINDA, the available alternative techniques (TPM lesion filling, and CCFM respectively) had significantly higher NCC scores in the masked region. For all other normalization methods, the enantiomorphic lesion filling technique significantly outperformed the compared alternative. See Figure 3.6 and Table 3.5 for details of the pairwise comparisons.

3.4 SUPPORT VECTOR REGRESSION (SVR)

Whereas the previous metrics aimed to quantify the quality of normalization, the goal of the SVR analysis was to assess the "real world" effects that normalization has on predicting behavior outcome from stroke related brain injury. All normalization methods and their lesion compensation techniques tested could predict WAB AQ scores as measured by the significant Pearson correlation coefficients between actual and predicted scores (see Table 3.6 and 3.7). However, to compare the error associated with the predicted scores within each method we performed a repeated measures ANOVA with Greenhouse–Geisser correction on the residuals obtained from each SVR analysis. Lower residuals are indeed related to higher correlation, but simply knowing which method is most predictive does not reveal any information about how the lesion compensation technique has influenced the error associated with the predictions. The ANOVA showed a significant effect of normalization method $F(3.84, 607.3) = 5.55$, $p < 0.001$, $\eta_p^2 = .03$ (see Figure 3.7). SPM12 unified default regularization with ENAT and TPM lesion filling had the lowest residual average (M = .06, SD = .09; M = .06, SD = .09 respectively), while SPM12 DARTEL with ENAT (M = .09, SD = .12), and SPM12 Old Norm with CFM (M = .08, SD = .11) had the highest residual average. Pairwise comparisons between lesion compensation techniques within each normalization method showed a significant difference between SPM12 DARTEL TPM filling and SPM12 DARTEL with ENAT ($t(158) = -3.83$, $p < 0.001$, Cohens $d = .29$). All other comparisons between the lesion compensation techniques within each normalization method were not significant at an alpha level of .05 (Bonferroni corrected).

Table 3.1: Within Method Comparisons of Lesion Compensation on Landmark Displacement (mm)

Method	Mean Diff.	95% CI. Lower	Upper	*t*	*p*	*d*
Old Norm CFM vs. ENAT	.01	.00	.01	2.83	< .001	.33
Unified High TPM vs. ENAT	.09	.09	.1	21.23	< .001	1.63
Unified Default TPM vs. ENAT	.07	.06	.07	16.5	< .001	1.09
DARTEL TPM vs. ENAT	.18	.16	.19	17.8	< .001	1.48
ANTs LINDA CCFM vs. ENAT	-.10	-.12	-.09	-10.98	< .001	.74

Table 3.2: Within Method Comparisons of Lesion Compensation on RMSD (mm)

Method	Mean Diff.	95% CI. Lower	Upper	*t*	*p*	*d*
Old Norm CFM vs. ENAT	.04	.01	.07	2.88	0.0045	.15
Unified High TPM vs. ENAT	-.11	-.14	-.08	-6.81	< .001	.45
Unified Default TPM vs. ENAT	-.15	-.19	-.11	-7.22	< .001	.46
DARTEL TPM vs. ENAT	.51	.47	.55	27.55	< .001	1.23
ANTs LINDA CCFM vs. ENAT	.69	.66	.72	41.95	< .001	3.66

Table 3.3: Post Hoc Pairwise Comparisons Between Each Normalization Method and NCC Score in Control Participants

Method		Mean Diff.	95% CI. L	95% CI. U	*t*	*p*	*d*
Old Norm	Unified High	.12	.11	.12	82.19	< .001	1.65
	Unified Default	.07	.06	.07	42.42	< .001	0.96
	DARTEL	-.03	.03	.02	-26.55	< .001	0.49
	ANTs	-.19	.19	.19	-184.47	< .001	5.27
Unified High	Unified Default	-.05	.05	.05	-106.38	< .001	0.47
	DARTEL	-.14	.15	.14	-107.69	< .001	1.86
	ANTs	-.30	.31	.30	-154.64	< .001	5.26
Unified Default	DARTEL	-.10	.10	.09	-67.11	< .001	1.24
	ANTs	-.26	.26	.25	-126.21	< .001	4.06
DARTEL	ANTs	-.16	.17	.16	98.58	< .001	3.20

Table 3.4: Within Method Comparisons of NCC Score from Control Participants Within the Masked Region

Method	Mean Diff.	95% CI. Lower	95% CI. Upper	*t*	*p*	*d*
Old Norm CFM vs. ENAT	0	0	0	2.07	0.04	-
Unified High TPM vs. ENAT	0	0	0	-.34	.73	-
Unified Default TPM vs. ENAT	.01	.01	.01	11.36	< .001	1
DARTEL TPM vs. ENAT	.01	.01	.02	11.17	< .001	.5
ANTs LINDA CCFM vs. ENAT	.08	.07	.09	17.20	< .001	1.33

Table 3.5: Within Method Comparisons of NCC Score from Artificial Patients Within the Masked Region

Method	Mean Diff.	95% CI. Lower	Upper	*t*	*p*	*d*
Old Norm CFM vs. ENAT	-.01	-.01	0	-8.36	< .001	1
Unified High TPM vs. ENAT	-.02	-.02	-.02	-19.58	< .001	2
Unified Default TPM vs. ENAT	-.02	-.02	-.02	-14.09	< .001	1
DARTEL TPM vs. ENAT	.04	.04	.05	23.57	< .001	2
ANTs LINDA CCFM vs. ENAT	.04	.03	.05	11.91	< .001	1

Table 3.6: Within Method Comparisons of SVR Residuals

Method	Mean Diff.	95% CI. Lower	Upper	*t*	*p*	*d*
Old Norm CFM vs. ENAT	-01	0	.02	2.06	0.04	-
Unified High TPM vs. ENAT	0	0	0	.34	0.74	-
Unified Default TPM vs. ENAT	0	0	0	.85	0.4	-
DARTEL TPM vs. ENAT	-.02	-.03	-.01	-3.83	< .001	.29
ANTs LINDA CCFM vs. ENAT	0	-.01	.01	-.26	0.8	-

Table 3.7: Correlation Values Between Predicted and Actual WAB AQ Scores from the SVR Analysis

Method	*r*	*p*
Old Norm CFM	0.45	< .001
Old Norm ENAT	.65	< .001
Unified High	0.58	< .001
Unified High ENAT	.59	< .001
Unified Default	.59	< .001
Unified Default ENAT	.62	< .001
DARTEL	.55	< .001
DARTEL ENAT	.45	< .001
ANTs LINDA CCFM	.56	< .001
ANTs LINDA ENAT	.56	< .001

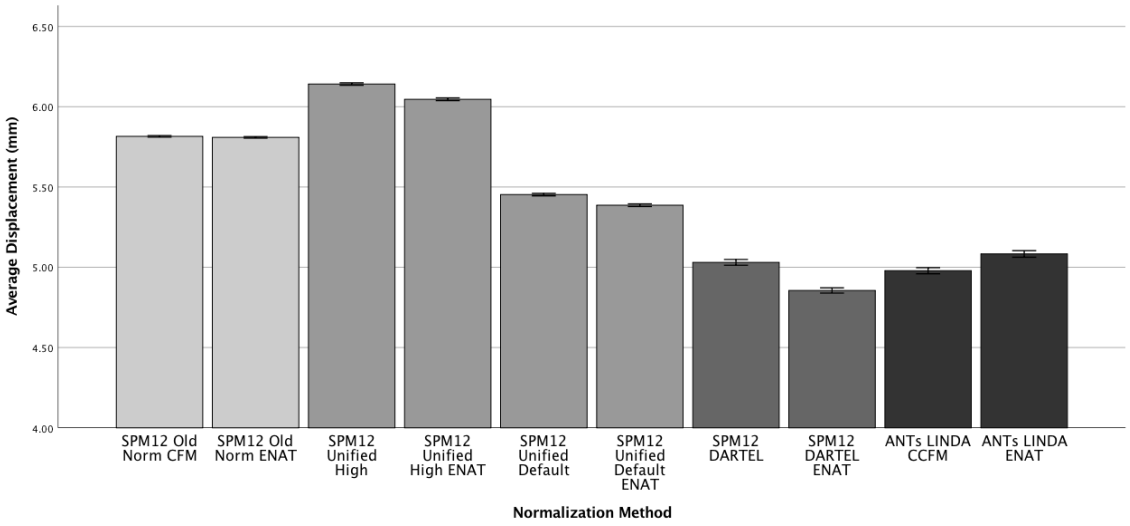

Figure 3.1: Average landmark displacement from the group centroids (averaged across all landmarks). Error bars represent 95% CI. Lower values indicate better normalization.

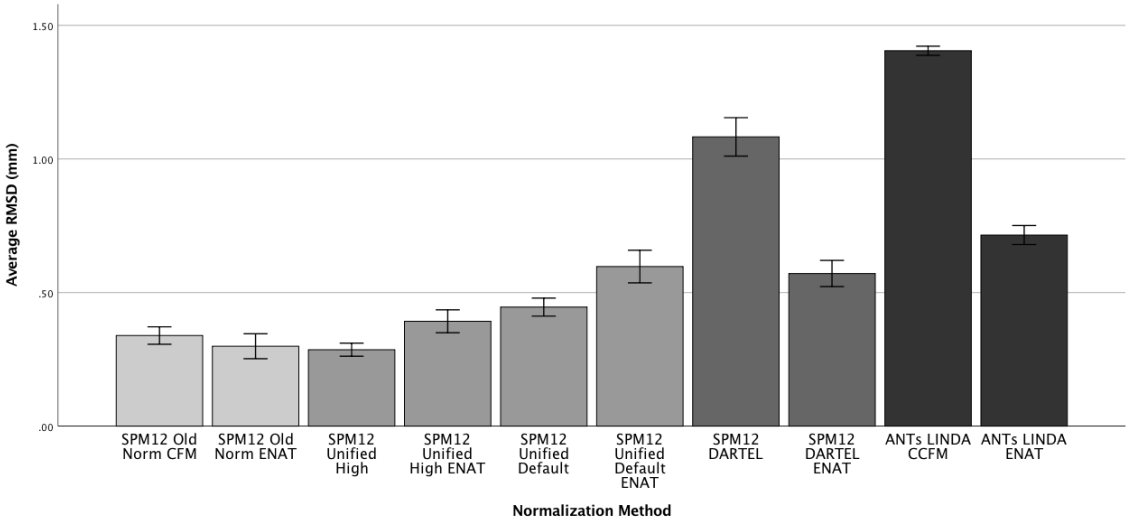

Figure 3.2: Average RMSD. Error bars represent 95% CI. Lower values indicate that the lesioned image was distorted similarly to the unlesioned reference image across all voxels.

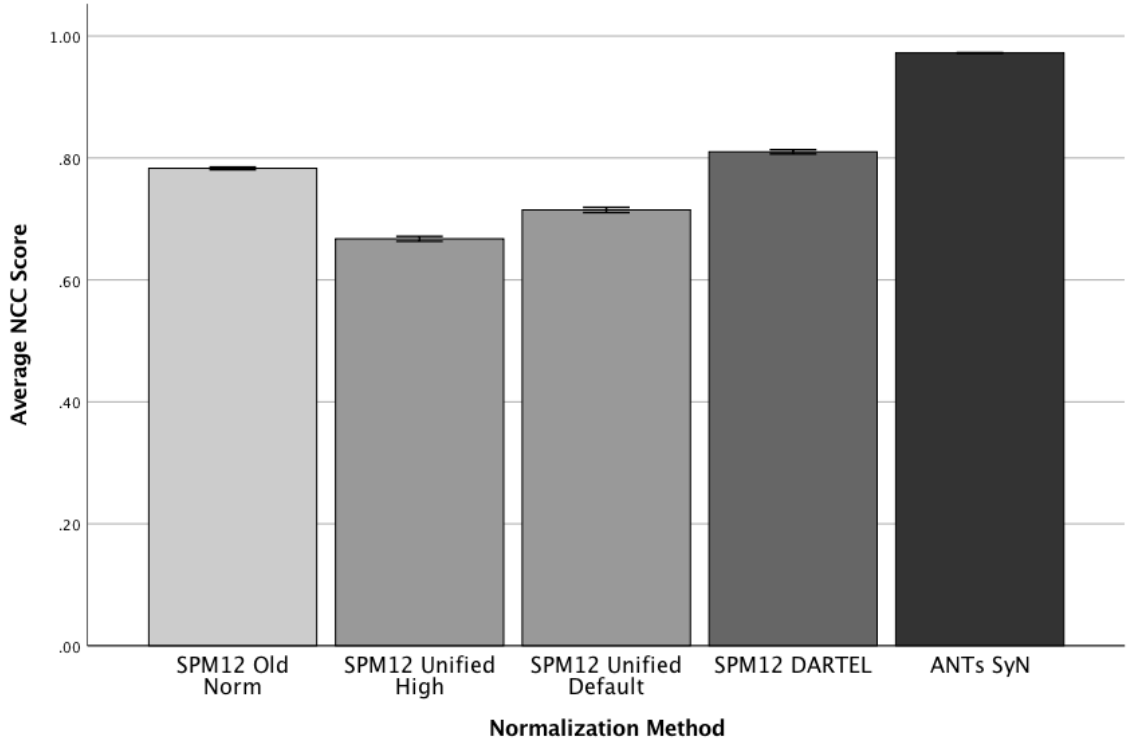

Figure 3.3: Average NCC scores measured across all combinations of normalized control participant images. Error bars represent 95% CI. Higher values indicate a better match among normalized images.

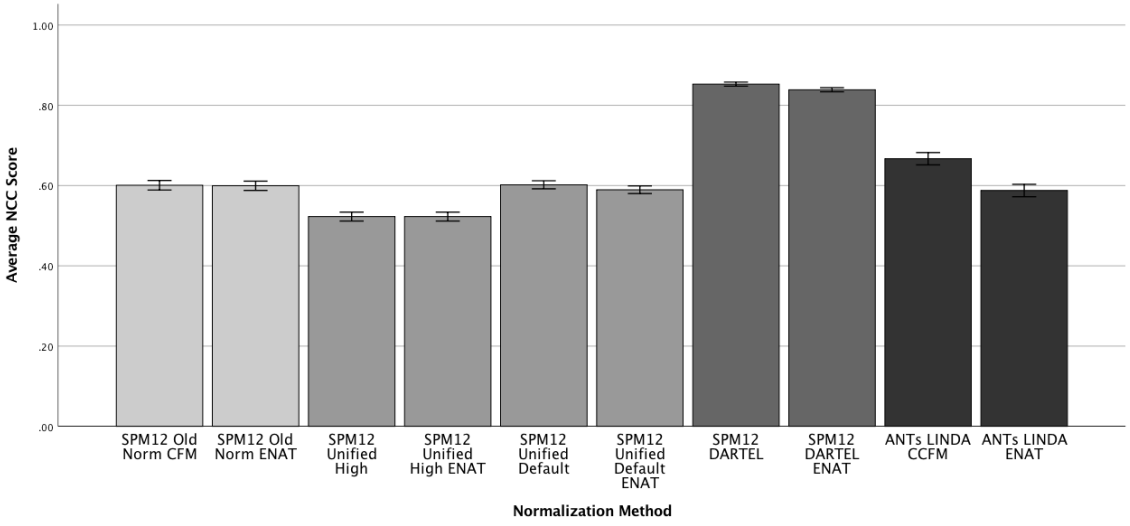

Figure 3.4: Average NCC score of control participant images within the masked region defined by the normalized masks per method. Error bars represent 95% CI. Higher values indicated a better match within the masked region.

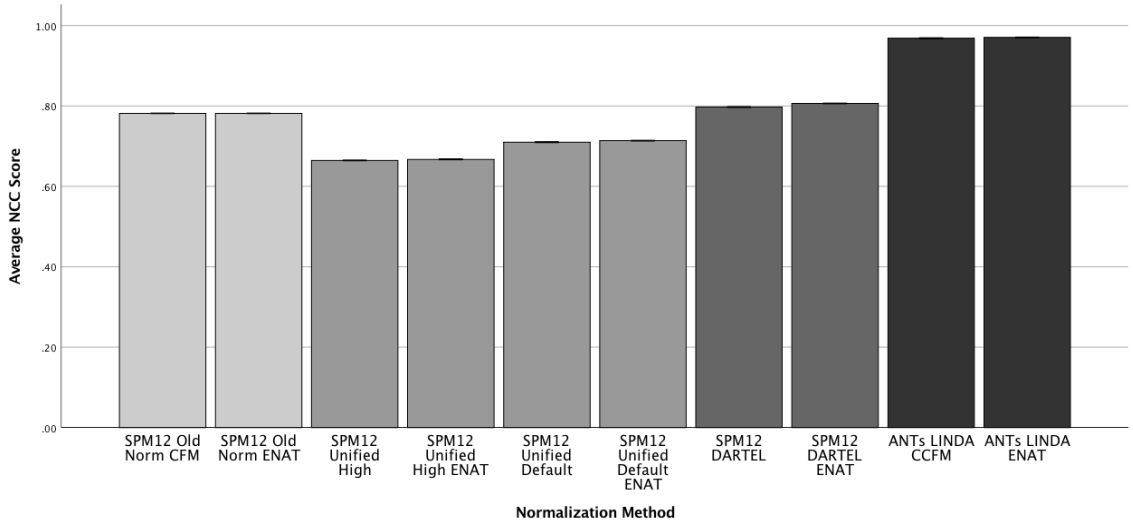

Figure 3.5: Average NCC score of artificial patients within the whole image. Error bars represent 95% CI. Higher values indicate a better match across all images per method.

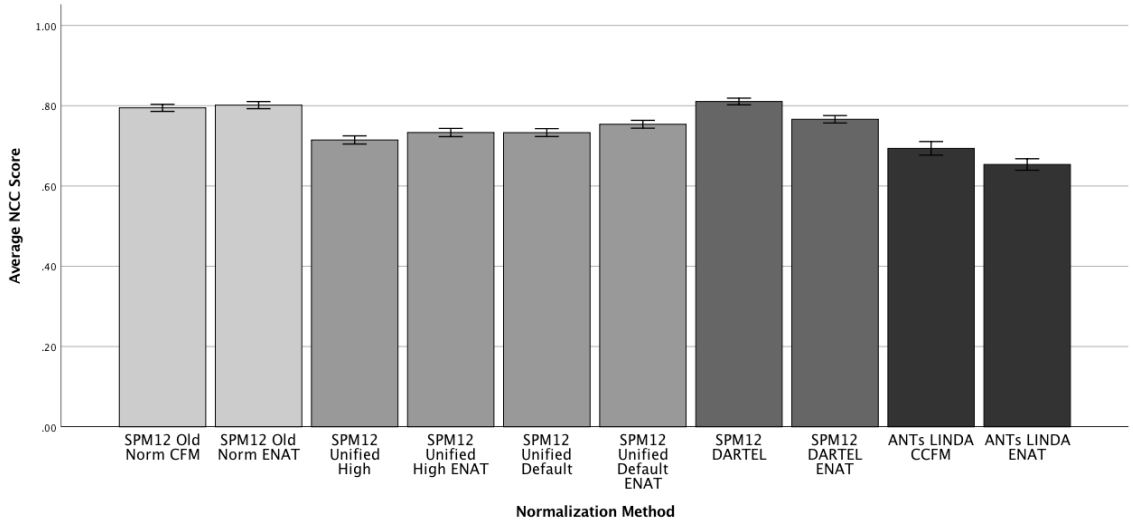

Figure 3.6: Average NCC score of artificial patients within the masked region defined by the average normalized masks per method. Error bars represent 95% CI. Higher values indicate a better match within the masked region.

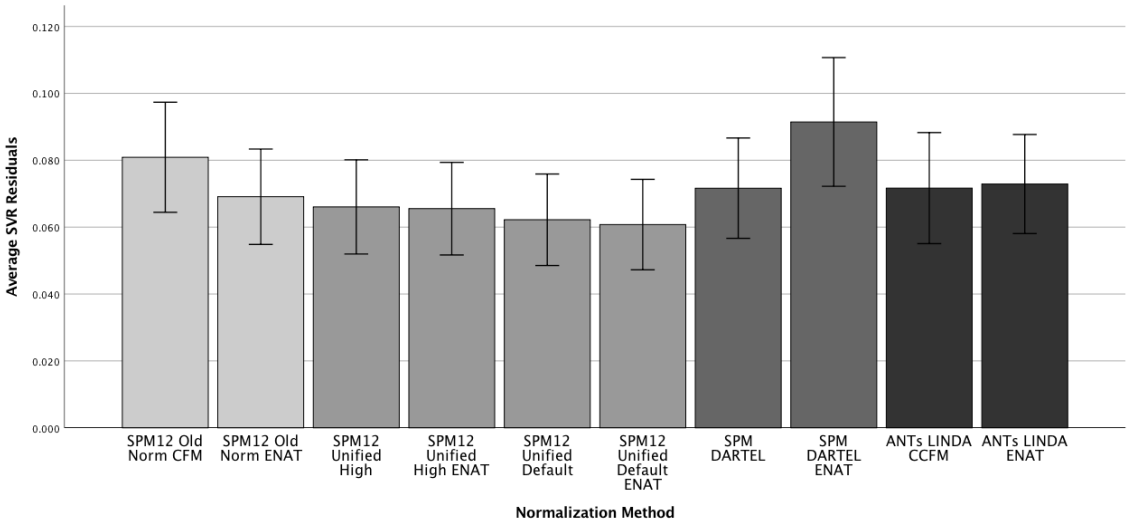

Figure 3.7: Average residual value from the SVR prediction. Error bars represent 95% CI. Lower residuals indicate less prediction error.

CHAPTER 4

DISCUSSION

Overall, we have shown that within popular normalization software, there is more than one way to compensate for brain injury when registering a clinical image to a template. Specifically, we tested each normalization method's default strategy to the nonstandard enantiomorphic technique. We found that the best performing technique differs across methods and the performance metric used. A major component of the current work is our use of multiple quantitative metrics for determining the quality of normalization. For example, if only assessing performance by lowest voxelwise RMSD value, our data show that the oldest methods (SPM12 Old Norm) have lower (i.e. better) scores, compared to ANTs and DARTEL, which is contradictory to the typical pattern of performance in healthy subjects (Klein et al., 2009). Within a method, the ideal lesion compensation strategy also differs. However, if performance is assessed by landmark displacement (a more intuitive measure), the lowest values (better) are produced by ANTs LINDA and SPM12's DARTEL, which are both methods with evidence showing their superiority in studies of healthy subjects (Klein et al., 2009; Ripollés et al., 2012). Also, the within method landmark comparisons show that DARTEL benefits more from the enantiomorphic technique, whereas ANTs benefits more by using its default technique (CCFM). Additionally, our NCC results generally support the pattern of performance seen in the landmark results. The interpretation of our SVR behavior

prediction is less clear though. Each method can predict behavior well, meaning that all methods are useful in these types of analyses. A more in depth assessment with parcellated atlas maps might reveal more dramatic differences compared to voxelwise methods used in the current study. Additionally, prediction accuracy (and error rate) could be highly dependent on the behavior being assessed and that behavior's suspected neuroanatomical substrates (Pustina et al., 2016).

Our comparison study is timely in that no previous study of this nature has included SPM12 and its new TPM filling feature, nor have any included LINDA (Pustina et al., 2016) in combination with ANTs. A noted limitation of ANTs in a previous comparison (Ripollés et al., 2012), was its lack of a method, or addon for automated lesion identification. This limitation has since been directly address by LINDA, which is deeply integrated with ANTs. Further, apart from its debut release, no other study has directly compared enantiomorphic lesion healing to the available options included in other software. Assessing the performance of the most recent version of SPM12 is particularly important because its new default behavior for masked portions of images (filling in with TPM values) changes the status quo for all users who use the software for clinical neuroimaging analysis that previously relied on its earlier techniques. The assessment of LINDA is also important given the evidence of its improved superiority, and reliability compared to its popular predecessor, the ALI toolbox (Seghier et al., 2008). As with previous studies that aim to compare normalization methods on clinical data (Andersen et al., 2010; Crinion et al., 2007), we used an artificially created dataset of brain images containing lesions resembling those of stroke patients. Although the term "artificial" accurately describes this dataset, the images closely resembled true clinical

images. This close resemblance to true stroke images is a result of the methods used to create the artificial lesions, and a result of injecting precisely delineated lesions from real patient data into a dataset of elderly participants. The use of control images from participants in a similar age range is not unique to our study. Andersen and colleagues (2010) also used a small sample of elderly control participants for their artificially created dataset, and Ripollés et al. (2012) also used some older participants to create their dataset, although the age range was quite wide (21-71 years old). Our control group is unique in that they are within a similar age range as the stroke participants, and were selectively recruited (in a previous study) for being at risk for stroke, but had not yet experienced one. Further, our dataset contained over 10,000 unique combinations of patient lesions injected into the control participants. An argument could be made that we have inflated the similarity of images in our dataset by injecting the same lesion from each stroke participant's image into each control participant. However, by creating this dataset we have also introduced additional variability compared to only using patient data. Every image still retains the unique characteristics of the control participant (e.g. their specific anatomy outside of the lesioned area). Unfortunately, there does not seem to be a clear alternative method to creating datasets such as these, and any limitations inherent to these methods have also affected similar studies in the same manner (Andersen et al., 2010; Brett et al., 2001; Crinion et al., 2007; Ripollés et al., 2012).

Previous studies have heavily relied on RMSD as the metric of normalization performance (Andersen et al., 2010; Brett et al., 2001). We also include this measure in our analysis. However, this metric may not be ideal for comparisons across normalization algorithms (which we do not perform). For example, ANTs LINDA uses its SyN image

registration algorithm which can have millions of degrees of freedom, compared to SPM's unified method which has about 1,000 degrees of freedom (Klein et al., 2009). Also, an approach such as the unified method makes many assumptions about the distribution of tissues within the brain (based on its a priori maps). Additionally, from first principles, SyN has more freedom to move voxels to fit the needs of the parameters defined by the cost function. Therefore, it should on average move voxels more in the lesioned image compared to the reference unlesioned image, which would result in higher RMSD values. The same argument applies to SPM's DARTEL (also millions of DOF) when compared to normalization methods with lower degrees of freedom. In our data, we see that the methods with the most constraint on nonlinear warping (SPM Old Norm, and SPM unified with high regularization) obtain the lowest RMDS scores. This could in part be due to their decreased freedom of parameters, and therefore less overall influence on image warping. In our experiment, we only statistically test the RMSD within each normalization method between the method's default lesion compensation technique, and the nonstandard enantiomorphic healing method. In these comparisons, we show that within SPM12's Old Norm, enantiomorphic normalization results in lower RMSD compared to traditional CFM, which directly replicates the original Nachev et al. (2008) finding. However, in both versions of SPM12's unified method (high and default regularization) the new standard SPM12 TPM filling technique results in lower RMSD values. The most contrasted RMSD results are in both SPM12's DARTEL and ANTs LINDA, which benefit more from enantiomorphic lesion healing compared to the alternatives of SPM12's unified TPM filling, and CCFM respectively. The benefit of enantiomorphic healing seen in both methods with millions of degrees of freedom could

be simply related to the mirrored tissue better fitting the needs of the cost function and thus resulting in less displacement in and around the lesioned area. In terms of RMSD, we recommend that users of ANTs and LINDA apply the enantiomorphic healing technique since their normalizations would benefit more. For users of SPM12's unified method, the now default TPM filling is perhaps better suited for lesioned images if their only concern is RMSD (assuming the lesioned region has been zeroed out first).

In contrast to RMSD, the measure of landmark displacement is arguably less ambiguous to interpret. The main goal of brain image normalization is to align similar anatomical features between and individual and a template (Brett et al., 2001; Crinion et al., 2007; Friston et al., 1995). This is especially challenging in data from stroke participants with focal lesions. Similar to Crinion et al. (2007), we identified a set of landmarks (individual 3D coordinates) in each healthy individual. However, the previous authors measured the root mean squared error (RMSE) of landmark displacement within their *healthy control* group, and not in the lesioned images. Here we extended the landmark technique to our artificially *lesioned* dataset. This allowed us to measure the impact of landmark displacement in a the artificial dataset, which aimed to mimic a real stroke dataset. Again, this dataset affords us some advantages. We identified and recorded the location of each landmark in the healthy images prior to lesion injection. We also created warped landmark images (landmarks were the identified coordinates with 3x3x3 voxel cubes placed inside these images). Rather than measure RMSE among landmarks, we measured the Euclidean distance of each participant's warped landmarks from their respective mean landmark locations across the group. We did not perform a specific test of the reliability of landmark labeling in the control participant images, but

since they were all done by a single individual, and then visually verified, any prior error associated with labeling was identical in all methods that were tested. Overall the pattern of results are on par with the ranked order of normalization methods in Klein et al. (2009), in that both SPM DARTEL and ANTS achieved the lowest landmark displacement scores (see Figure 3.1), followed by SPM12's unified method. Of the methods with the lowest (most desirable) mean distances, DARTEL with enantiomorphic lesion healing outperformed DARTEL with unified TPM filling, and ANTs LINDA with CCFM outperformed its enantiomorphic counterpart. Although our comparison included additional normalization methods, and we measured our landmark effects in lesioned images, we obtained results that agree with the pattern of results in Crinion et al. (2007). On one hand, these landmark results replicate their previous work, and on the other, our voxelwise RMSD result is at odds with the previous authors' pattern of results for RMSD. However, the previous study was conducted with SPM5, and our current work used SPM12, which contains many improvements beyond just the unified method's TPM filling procedure we have focused on. For users of each software package we recommend using the enantiomorphic procedure with DARTEL on lesioned images, and the default CCFM method if using ANTs. If using SPM12's unified methods, we recommend the enantiomorphic procedure, as it results in less deviation from each group landmark compared to the SPM12 TPM filling procedure. Within Old Norm the effect size between the two lesion compensation strategies is quite small. Both methods yield similar results.

Inspired by Ripollés et al. (2012), we chose to measure the normalized cross correlation (NCC) within our datasets in addition to the RMSD and landmark displacement. The NCC score summarizes the similarity (or dissimilarity) between pairs

of normalized images (Ripollés et al., 2012; Tahmasebi, Abolmaesumi, Zheng, Munhall, & Johnsrude, 2009). This analysis extends that of Ripollés et al. (2012) who only calculated NCC within their control participant group. Here we measure NCC in both our control group and our artificially lesioned group. Within our control group, ANTs significantly outperformed all other methods, indicating that its normalization results in images more similar to each other in template space. The high NCC scores indicate that the underlying anatomy is well aligned given that incorrectly aligned anatomy in normalized space would in theory attempt to correlate gray matter areas with white matter areas (an extreme example), resulting in reduced NCC values. However, we must note the confound present by using NCC as our similarity metric. It could be the case that the ANTs method resulted in the highest NCC scores within the normalized control participant images because NCC was used as the cost function to match to the template during the normalization process. However, as most versions of the SPM12 algorithms use the sum of squared differences (SSD) cost function, and the unified method incorporates the mutual information cost function (MI), using other similarity metrics in this manner also present potential confounds. ANTs SyN with NCC is built into the design of LINDA so we chose to use it across our analyses for consistency with the intended use of the tool. However, independent investigations could explore using ANTs and SyN with other cost functions, so that NCC would not be confounded with any tool across both SPM12 and ANTs. This confound is only present in our comparison of normalized control participants.

SPM12's DARTEL had high NCC scores (but less than ANTs), and was significantly higher than most other normalization methods. These results align with the

pattern seen in the landmark analysis where ANTs and DARTEL also resulted in the lowest deviation from the landmark means across participants in the created lesion dataset. Interestingly, SPM12's Old Norm method obtained significantly higher NCC scores in the control participants compared to both versions of SPM12's unified method we tested. This could be explained by the fact that Old Norm uses the SSD cost function, rather than the hybrid approach in the unified method. It may be the case that the hybrid approach is negatively impacted by the limits imposed by the a priori maps. If our results are limited to only SPM12's unified method, DARTEL, and ANTs then the pattern of NCC scores replicates those of the control group NCC scores in Ripollés et al. (2012).

The results of the lesion masked control participant analysis closely match the unmasked results with the major exception of ANTs with LINDA, which had lower NCC scores compared to the unmasked analysis. The masked comparison in control participants is mainly possible given our use of the artificially lesioned data, and its relationship to the control group. Masking and then computing NCC within the control participants gives us a general idea of how well the normalization methods have matched the tissue within the mask (how similar the signal is). In the masked analysis on controls both versions of ANTs with LINDA had lower scores than SPM12's DARTEL (see Figure 3.4). However, we only statistically compare NCC within a method rather than across all methods and lesion compensation techniques. The dramatic difference seen between masked and unmasked analyses could be inherent to the LINDA method itself since it automatically delineates the lesions for every image, whereas all other methods used the same hand drawn lesions. By providing a new prediction for every image, LINDA introduced additional variability (SD = .1, compared to DARTEL's SD = .03)

into the masked analysis that was absent from the hand drawn lesions. This could be an important consideration for future assessments of LINDA for fully automated analyses. This masked NCC score reduction could also be related to our dataset, rather than an inherent flaw in LINDA. Here, LINDA was forced to predict a new lesion for every image even when the same injury was present in the different control images. The hand draw lesions were not subject to this same iterative prediction variability. However, if a new lesion were to be hand drawn in each artificial image (such as LINDA needing to predict the lesion every time) we would expect that additional variability would be added to the other methods as well given the prior evidence that intra- and interrater reliability measurements of lesion identification is not perfect (Luby, Bykowski, Schellinger, Merino, & Warach, 2006). Importantly though, we only compare the NCC scores within each method. Among the comparisons, the enantiomorphic method generally resulted in lower NCC scores.

The whole image NCC scores for the artificial dataset closely resembled the pattern of NCC scores of the whole image analysis of the control group (see Figure 3.3). ANTs and SPM12's DARTEL both obtained the highest NCC scores in the artificially lesioned images, indicating their ability to generate images in template space with high similarity. Again, this result also supports the landmark displacement data. NCC scores within the masked region of the artificial dataset somewhat matched the pattern of results of the mask control NCC scores. Again, in the masked condition SPM12's DARTEL (with the default TPM filling procedure) obtained the highest NCC score, indicating a better match among normalized images, and it was significantly higher than its enantiomorphic counterpart. Similar to the masked control analysis, ANTS had lower

performance in the masked lesion data, likely due to the same limitations outlined previously. SPM12's Old Norm also obtained high NCC scores in the masked lesion data, likely due to the same logic described in relation to the masked control data. Within Old Norm, enantiomorphic outperformed traditional CFM, indicating a better match among images analyzed with that technique. Within ANTs LINDA, the default CCFM outperformed its enantiomorphic comparison, but the scores for both ANTs LINDA versions were lower than DARTEL and SPM12's Old Norm (see Figure 3.6). A future study of a less automated ANTs based normalization (no LINDA prediction) would likely show less variability in the masked NCC scores, and subsequently more closely resemble the unmasked results. The goal of including LINDA in this paper was to assess two versions of lesion compensation in normalization when those lesions were predicted using LINDA.

The above discussion relates to our use of three performance metrics seen in previous normalization comparisons in clinical data. We assessed the normalization of our artificially lesioned dataset using each of these metrics between multiple lesion compensation strategies. The current study is the first to comprehensively combine voxelwise RMSD, landmark displacement, and NCC scores into one unified assessment, where previously they have been reported separately, and not always on the same dataset (Andersen et al., 2010; Brett et al., 2001; Crinion et al., 2007; Nachev et al., 2008; Ripollés et al., 2012).

However, we performed an additional, novel analysis where the goal was to predict behavior from normalized lesion maps using support vector regression (Smola & Schölkopf, 2004; Yourganov et al., 2016). Crucially, the statistical procedure used for the

SVR analysis remained constant for each prediction. Instead, like the other analyses, we varied the way in which the data were normalized using the same methods performed on the artificially lesioned dataset. The SVR procedure was carried out on real patient data with associated WAB AQ scores (a measure of overall language impairment in the stroke participant group). Each normalization method could significantly predict WAB AQ scores as measured by the correlation of actual and predicted scores across participants (see Table 3.7). However, the method with the highest correlation coefficient and lowest residual average was SPM12's unified method (default regularization) with enantiomorphic lesion healing, followed closely by its unified TPM filling counterpart (see Figure 3.7). In the comparison of residuals within each method, the only significant difference was between SPM12's DARTEL with TPM filling, and its enantiomorphic counterpart, where the TPM technique resulted in lower residuals. Although each method could significantly predict WAB AQ scores (see Table 3.7), it is clear that the most desirable option is the one that results in the smallest error in prediction. The observed variability in correlation coefficients among the normalization methods was surprising. It is clear that normalization method has an overall effect on the associated error of predictions, but the low effect size implies that such differences are trivial. Of course, this claim is based on our data, and the specific prediction procedures used, but the effects of normalization can easily be extended to other behavior predictions. Additionally, differences between normalization methods as measured by residual error may become more apparent in region of interest (ROI) analyses which use parcellated, labeled atlases that represent percent of damage to regions, rather than using each voxel as a predictor. Although the downside of such an analysis could be less precision in anatomical

specificity of results, since many voxels are grouped into atlas regions. However, these analyses are quite common (Gleichgerrcht, Fridriksson, Rorden, & Bonilha, 2017; Shahid et al., 2018; Yourganov et al., 2016; Yourganov, Smith, Fridriksson, & Rorden, 2015), and would benefit from further testing of normalization methods and lesion compensation techniques. Furthermore, LINDA can produce a probabilistic map of its lesion prediction, which could serve as a more informative predictor in SVR, and similar analyses since each voxel identified as lesioned would have an associated probability. These probability values could implicitly map on to less necrotic tissue present at the boundaries of the focal injury.

In conclusion, we set out to test multiple methods of normalizing stroke neuroimaging data. Our purpose was not to definitively state which method was the gold standard, but rather illustrate that performance varies based on the method of assessment, and lesion compensation technique. We argue that good normalization is a tradeoff of sulcal alignment and preserving volume, this is analogous to representing a spherical surface as a 2D rectangular map. The Peters projection preserves volumes, the Mercator projection preserves directions. They represent tradeoffs. Likewise, the hunt for gold standard in normalization will depend on the question. For functional modules that are bound to sulci, more spatial distortion is desirable than functions that are more bound to volumetrics. Of the performance metrics used, both the landmark displacement and the SVR prediction have strong ties to the goals of most clinical imaging studies. One goal being to match a stroke individual's anatomy to the healthy template anatomy, and the other being the prediction of behavior from imaging data. Given our data, we suggest users of ANTs and LINDA adopt the default methods of the software, and recommend

users of SPM12 to perhaps adopt DARTEL with enantiomorphic lesion healing. However, DARTEL may benefit by fine tuning parameters across studies and datasets. All other methods assessed are also valid for answering clinical neuroimaging questions, but the more advanced provide increased performance on the metrics most useful to clinicians.

REFERENCES

Andersen, S. M., Rapcsak, S. Z., & Beeson, P. M. (2010). Cost function masking during normalization of brains with focal lesions: Still a necessity? *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2010.06.003

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*. https://doi.org/10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*. https://doi.org/10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2001). Why Voxel-based morphometry should be used. *NeuroImage*. https://doi.org/10.1006/nimg.2001.0961

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced Normalization Tools (ANTS). Insight Journal. https://doi.org/http://hdl.handle.net/10380/3113

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronker, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*. https://doi.org/10.1038/nn1050

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., … Muntner, P. (2017). Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*. https://doi.org/10.1161/CIR.0000000000000485

Bonilha, L., Yasuda, C. L., Rorden, C., Li, L. M., Tedeschi, H., De Oliveira, E., & Cendes, F. (2007). Does resection of the medial temporal lobe improve the outcome of temporal lobe epilepsy surgery? *Epilepsia*. https://doi.org/10.1111/j.1528-1167.2006.00958.x

Brett, M., Leff, A. P., Rorden, C., & Ashburner, J. (2001). Spatial normalization of brain images with focal lesions using cost function masking. *NeuroImage*, *14*(2), 486–500. https://doi.org/10.1006/nimg.2001.0845

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/1961189.1961199

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *Journal of Computer Assisted Tomography*. https://doi.org/10.1097/00004728-199403000-00005

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*(3), 866–875.

https://doi.org/10.1016/j.neuroimage.2007.04.065

Fiez, J. A., Damasio, H., & Grabowski, T. J. (2000). Lesion segmentation and manual warping to a reference brain: Intra- and interobserver reliability. *Human Brain Mapping*. https://doi.org/10.1002/(SICI)1097-0193(200004)9:4<192::AID-HBM2>3.0.CO;2-Y

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. □B, Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*. https://doi.org/10.1002/hbm.460030303

Gleichgerrcht, E., Fridriksson, J., Rorden, C., & Bonilha, L. (2017). Connectome-based lesion-symptom mapping (CLSM): A novel approach to map neurological function. *NeuroImage: Clinical*. https://doi.org/10.1016/j.nicl.2017.08.018

Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., & Collins, D. L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*. https://doi.org/10.1007/11866763_8

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Karnath, H. O., Berger, M. F., Küker, W., & Rorden, C. (2004). The anatomy of spatial neglect based on voxelwise statistical analysis: A study of 140 patients. *Cerebral*

Cortex. https://doi.org/10.1093/cercor/bhh076

Karnath, H. O., Rennig, J., Johannsen, L., & Rorden, C. (2011). The anatomy underlying acute versus chronic spatial neglect: A longitudinal study. *Brain*. https://doi.org/10.1093/brain/awq355

Kim, J., Avants, B., Patel, S., & Whyte, J. (2007). Spatial normalization of injured brains for neuroimaging research: An illustrative introduction of available options. *English*.

Kim, J., Patel, S., Avants, B., & Whyte, J. (2015). Spatial normalization of injured brains for neuroimaging research : An illustrative introduction of available options, (October).

Klein, A., , Jesper Andersson, Babak A. Ardekani, John Ashburner, B. A., Ming-Chang Chiang, Gary E. Christensen, D. L. C., Gee, J., Hellier, P., Joo Hyun Song, Mark Jenkinson, C. L., … Tom Vercauteren, Roger P. Woods, J. John Mann, and R. V. P. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*, *46*(3), 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037.Evaluation

Le, B. H., Ma, X., & Deng, Z. (2012). Live speech driven head-and-eye motion generators. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2012.74

Luby, M., Bykowski, J. L., Schellinger, P. D., Merino, J. G., & Warach, S. (2006). Intra- and interrater reliability of ischemic lesion volume measurements on diffusion-weighted, mean transit time and fluid-attenuated inversion recovery MRI. *Stroke*. https://doi.org/10.1161/01.STR.0000249416.77132.1a

Nachev, P., Coulthard, E., Jäger, H. R., Kennard, C., & Husain, M. (2008).

Enantiomorphic normalization of focally lesioned brains. *NeuroImage*, *39*(3), 1215–1226. https://doi.org/10.1016/j.neuroimage.2007.10.002

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., & Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Human Brain Mapping*. https://doi.org/10.1002/hbm.23110

Ripollés, P., Marco-Pallarés, J., de Diego-Balaguer, R., Miró, J., Falip, M., Juncadella, M., … Rodriguez-Fornells, A. (2012). Analysis of automated methods for spatial normalization of lesioned brains. *NeuroImage*, *60*(2), 1296–1306. https://doi.org/10.1016/j.neuroimage.2012.01.094

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*. https://doi.org/10.1155/2000/421719

Rorden, C., Karnath, H. O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2007.19.7.1081

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin*. https://doi.org/10.1037/a0023262

Sanchez, C. E., Richards, J. E., & Almli, C. R. (2012). Age-specific MRI templates for pediatric neuroimaging. *Developmental Neuropsychology*, *37*(5), 379–399. https://doi.org/10.1080/87565641.2012.688900

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*(4), 1253–1266. https://doi.org/10.1016/j.neuroimage.2008.03.028

Shahid, H., Sebastian, R., Tippett, D. C., Saxena, S., Wright, A., Hanayik, T., … Hillis, A. E. (2018). Regional Brain Dysfunction Associated with Semantic Errors in Comprehension. *Seminars in Speech and Language*, *39*(1). https://doi.org/10.1055/s-0037-1608858

Shewan, C. M., & Kertesz, A. (1980). Reliability and Validity Characteristics of the Western Aphasia Battery (WAB). *Journal of Speech and Hearing Disorders*. https://doi.org/10.1044/jshd.4503.308

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*. https://doi.org/10.1002/hbm.10062

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Stamatakis, E. A., & Tyler, L. K. (2005). Identifying lesions on structural brain images - Validation of the method and application to neuropsychological patients. *Brain and Language*. https://doi.org/10.1016/j.bandl.2004.12.010

Tahmasebi, A. M., Abolmaesumi, P., Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Reducing inter-subject anatomical variation: Effect of normalization method on sensitivity of functional magnetic resonance imaging data analysis in auditory cortex and the superior temporal region. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2009.05.047

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxis Atlas of the Human Brain*.

Direct.

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews Neuroscience*, *4*(1), 37–48. https://doi.org/10.1038/nrn1009

Yourganov, G., Fridriksson, J., Rorden, C., Gleichgerrcht, E., & Bonilha, L. (2016). Multivariate Connectome-Based Symptom Mapping in Post-Stroke Patients: Networks Supporting Language and Speech. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4396-15.2016

Yourganov, G., Smith, K. G., Fridriksson, J., & Rorden, C. (2015). Predicting aphasia type from brain damage measured with structural MRI. *Cortex*. https://doi.org/10.1016/j.cortex.2015.09.005

APPENDIX A: SOFTWARE DETAILS AND PROGRAM CODE

Table A.1: Software Summary

Method	Cost Function	Default Lesion Strategy	Degrees of Freedom
SPM12 Old Norm	SSD	CFM	~1000
SPM12 US High reg	Hybrid MI	TPM filling	~1000
SPM12 US Medium reg	Hybrid MI	TPM filling	~1000
SPM12 DARTEL	Hybris MI SSD	TPM filling	~6.4M
ANTs LINDA	CC	CCFM	~28M

Table A.2: Single Subject Run Rime, HPC Setup and Normalization Technique Details

Method	Run Time (min)	CPUs used	Regularization	Deformation
Old Norm CFM	~0.4	2	Bending energy, basis cutoff	Discrete cosine transforms
Old Norm ENAT	~2.18	2	Bending energy, basis cutoff	Discrete cosine transforms
Unified High	~2.17	2	Bending energy, basis cutoff	Discrete cosine transforms
Unified High ENAT	~4.14	2	Bending energy, basis cutoff	Discrete cosine transforms
Unified Default	~2.27	2	Bending energy, basis cutoff	Discrete cosine transforms
Unified Default ENAT	~3.87	2	Bending energy, basis cutoff	Discrete cosine transforms
DARTEL	~4.68	2	Linear-elasticity; MRes: full-multigrid (recursive)	Finite difference model of a velocity field (constant over time, diffeomorphic)
DARTEL ENAT	~5.95	2	Linear-elasticity; MRes: full-multigrid (recursive)	Finite difference model of a velocity field (constant over time, diffeomorphic)
ANTs LINDA CCFM	~28	8	MRes Gaussian smoothing of the velocity field; transformation symmetry	Bi-directional diffeomorphism
ANTs LINDA ENAT	~30	8	MRes Gaussian smoothing of the velocity field; transformation symmetry	Bi-directional diffeomorphism

a. The HPC utilized a SLURM scheduler for job (subject) submission

b. Regularization and deformation details are the same as Klein et al., (2009)

Code Sample 1: Python code sample illustrating how subjects were processed in parallel on the HPC. This example is specific to the SPM12 default unified TPM analysis, but is representative of all other subject processing scripts (python).

```
import os
import subprocess
import time
import shutil
import sys
import glob
from subprocess import call
import time

# set up some constants
STUDY        = "normalization_project"
STUDYLEG     = "artificial_lesion"
STUDYPART    = "spm_unified_tpm"
HOMEDIR      = os.environ['HOME']
STUDYDIR     = os.path.join(os.sep, "data", "userdata", "hanayik", STUDY,
       STUDYLEG, STUDYPART) #does not have trailing slash
SUBJDIRS     = []
CORESPERSUBJ = '2'
NODESTOUSE   = '1'
SLEEPTIME    = 1 # secs
SLEEPTIMEWAITING = 20 # secs
MAXQJOBS     = 200

print('HOMEDIR: ' + HOMEDIR)
print("STUDYDIR: " + STUDYDIR)
studyContents = glob.glob(os.path.join(STUDYDIR, 'NS*'))
studyContents.sort() # make list alphabetical
print("SUBJ DIRS: ")
for sc in studyContents: # loop through the list and find subj folders
    if os.path.isdir(sc): # if its a folder then append it to the subj array
        SUBJDIRS.append(sc)
        print(sc) # print it for visual confirmation

numjobs = len(SUBJDIRS)

if not os.path.exists(os.path.join(HOMEDIR,"logs",STUDYPART)):
    os.makedirs(os.path.join(HOMEDIR,"logs",STUDYPART))

if not os.path.exists(os.path.join(HOMEDIR,"logs",STUDYPART)):
    os.makedirs(os.path.join(HOMEDIR,"logs",STUDYPART))
```

```
# loop through subject folders and do processing
qjobs = 0
i = 0
# run until the entire list of jobs "subjects" has completed
while i < numjobs:
        # check how many jobs are running
    p = subprocess.Popen("squeue -u hanayik | wc -l", stdout=subprocess.PIPE,
        shell=True)
    (output, err) = p.communicate()
    if err is not None:
        time.sleep(1)
        continue
    # convert num jobs string to float for calculation
    qjobs = float(output) - 1 # minus one for header in output (counts as a line)

    # if space available, submit a new job ("subject")
    if qjobs <= MAXQJOBS:
        print("{} jobs in queue".format(qjobs))
        print("space available, submitting job now...")
        subj = SUBJDIRS[i]
        thisSubj = os.path.basename(subj)
        cmd = ["sbatch","-p","soph,jerlab","-N","1", "--job-name="+thisSubj,"-
        n",CORESPERSUBJ,"--
        output",os.path.join(HOMEDIR,"logs",STUDYPART,thisSubj+".out"),"--
        error",os.path.join(HOMEDIR,"logs",STUDYPART,thisSubj+".err"),"--
        wrap="+"python "+os.path.join(HOMEDIR,"run_sub_spm_unified_tpm.py")+"
        "+subj]
        print(str(cmd))
        e = call(cmd)
        if e != 0:
            print("Some sort of error submitting to sbatch, will try again in a sec...")
            time.sleep(1)
            continue
        time.sleep(SLEEPTIME)
        i += 1
    else:
        print("queue full, waiting on space...")
        time.sleep(SLEEPTIMEWAITING)
```

Code sample 2: Enantiomorphic lesion healing procedure (an enhanced, open source version recreated from the methods in Nachev et al., 2007) (matlab).

```
function et1 = lesion_heal(t1, les)
[p,n,x] = fileparts(t1);
et1 = fullfile(p, ['e' n x]);
if exist(et1, 'file')
   return
end
%replace tissue in location of lesion (les) with homologous healthy tissue
addpath('/home/hanayik/spm12/');
disp('added spm12 to path');
spm('defaults','fmri');
spm_jobman('initcfg');
spm_get_defaults('cmdline',true);
spmDir = spm('Dir');
if ~exist('t1','var'), t1 = spm_select(1,'image','Select T1 images'); end;
if ~exist('les','var'), les = spm_select(1,'image','Optional: select lesion map'); end;
%load data
hdr = spm_vol(t1);
img = spm_read_vols(hdr);
hdrLesion = spm_vol(les);
imgLesion = spm_read_vols(hdrLesion);
if ~isequal(size(img), size(imgLesion)), error('Dimensions do not match %s %s', les, t1);
      end;
%save with zeros in lesion
img(imgLesion ~= 0) = 0;
[p,n,x] = spm_fileparts(t1);
hdr.fname = fullfile(p, ['z',n,x]);
spm_write_vol(hdr,img);
%segment
[c1t1, c2t1] = newSegSub(hdr.fname);
t1LR = flipSub (t1);
st1 = combineSub(c1t1, c2t1, les);
st1LR = flipSub (st1);

%template = 'sc1zT1_M2127_LIME.nii';
%source = 'LRsc1zT1_M2127_LIME.nii';
%other = 'LRT1_M2127_LIME.nii';
rt1LR = coreg12EstWriteSub(st1, st1LR, t1LR);
et1 = insertSub(t1, rt1LR, les);

function namFilled = insertSub(nam, namLR, lesion)
%namLR donates voxels masked by lesion to image nam
if isempty(nam), namFilled =''; return; end;
hdrLesion = spm_vol(lesion);
```

```
imgLesion = spm_read_vols(hdrLesion);
rdata = +(imgLesion > (max(imgLesion(:))/2)); %binarize raw lesion data, + converts
      logical to double
spm_smooth(rdata,imgLesion,4); %blur data
rdata = +(imgLesion > 0.05); %dilate: more than 5%
spm_smooth(rdata,imgLesion,8); %blur data
%now use lesion map to blend flipped and original image
hdr = spm_vol(nam);
img = spm_read_vols(hdr);
hdr_flip = spm_vol(namLR);
imgFlip = spm_read_vols(hdr_flip);
if ~isequal(size(img), size(imgLesion)), error('Dimensions do not match %s %s', lesion,
      nam); end;
rdata = (img(:) .* (1.0-imgLesion(:)))+ (imgFlip(:) .* imgLesion(:));
rdata = reshape(rdata, size(img));
[pth, nam, ext] = spm_fileparts(hdr.fname);
hdr_flip.fname = fullfile(pth,['e' nam ext]);%image with lesion filled with intact
      hemisphere
spm_write_vol(hdr_flip,rdata);
namFilled = hdr_flip.fname;
%insertSub()

function resliced = coreg12EstWriteSub(template, source, other)
%coregister source to match template image, apply to lesion (use 12-dof normalization
      instead of 6 dof coregister)
if isempty(template) || isempty(source), return; end;
fprintf('Nonlinear Coregistration of %s to match %s\n',source,template);
matlabbatch{1}.spm.tools.oldnorm.estwrite.subj.source = {source};
matlabbatch{1}.spm.tools.oldnorm.estwrite.subj.wtsrc = [];
if ~exist('other','var') || isempty(other)
   matlabbatch{1}.spm.tools.oldnorm.estwrite.subj.resample = {[source]};
else
   matlabbatch{1}.spm.tools.oldnorm.estwrite.subj.resample = {[other]};
end
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.template = {template};
%n.b. masking tends to make problem worse
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.weight = [];
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.smosrc = 8;
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.smoref = 9;
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.regtype = 'mni';
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.cutoff = 25;
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.nits = 16;
matlabbatch{1}.spm.tools.oldnorm.estwrite.eoptions.reg = 1;
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.preserve = 0;
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.bb = [NaN NaN NaN; NaN NaN
      NaN];
```

```
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.vox = [NaN NaN NaN];
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.interp = 2;
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.wrap = [0 0 0];
matlabbatch{1}.spm.tools.oldnorm.estwrite.roptions.prefix = 'r';
spm_jobman('run',matlabbatch);
resliced = prefixSub('r',source);
if ~isempty(other), resliced = prefixSub('r',other); end;
%end coreg12EstWriteSub()

function nam = prefixSub (pre, nam)
[p, n, x] = spm_fileparts(nam);
nam = fullfile(p, [pre, n, x]);
%end prefixSub()

function sab = combineSub(a, b, c)
%sum two images, smooth with 4 voxel blur
fprintf('Combining of %s %s\n', a, b);
hdr = spm_vol(a);
img = spm_read_vols(hdr);
hdr = spm_vol(b);
img = img+spm_read_vols(hdr);
hdrc = spm_vol(c);
imgc = spm_read_vols(hdrc);
img(imgc > 0) = 1;
%we can smooth in normalization
%inimg = img + 0;
%spm_smooth(inimg,img,4); %blur data
[p,n,x] = spm_fileparts(a);
hdr.fname = fullfile(p, ['c',n,x]);
sab = hdr.fname;
spm_write_vol(hdr,img);
%end combineSmoothSub()

function namLR = flipSub (nam)
if isempty(nam), namLR = ''; return; end;
hdr = spm_vol(nam);
img = spm_read_vols(hdr);
[pth, nam, ext] = spm_fileparts(hdr.fname);
namLR = fullfile(pth, ['LR', nam, ext]);
hdr_flip = hdr;
hdr_flip.fname = namLR;
hdr_flip.mat = [-1 0 0 0; 0 1 0 0; 0 0 1 0; 0 0 0 1] * hdr_flip.mat;
spm_write_vol(hdr_flip,img);

function [c1t1, c2t1] = newSegSub(t1)
template = fullfile(spm('Dir'),'toolbox','Clinical','TPM4mm.nii');
```

```
if ~exist(template,'file')
   warning('Unable to find template named %s',template);
   template = fullfile(spm('Dir'),'tpm','TPM.nii');
   warning('falling back to template named %s',template);
end
[p,n,x] = spm_fileparts(t1);
c1t1 = fullfile(p,['c1',n,x]);
c2t1 = fullfile(p,['c2',n,x]);
fprintf('NewSegment of %s\n', t1);
matlabbatch{1}.spm.spatial.preproc.channel.vols = {t1};
matlabbatch{1}.spm.spatial.preproc.channel.biasreg = 0.001;
matlabbatch{1}.spm.spatial.preproc.channel.biasfwhm = 60;
matlabbatch{1}.spm.spatial.preproc.channel.write = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(1).tpm = {[template ',1']};
matlabbatch{1}.spm.spatial.preproc.tissue(1).ngaus = 1;
matlabbatch{1}.spm.spatial.preproc.tissue(1).native = [1 0];
matlabbatch{1}.spm.spatial.preproc.tissue(1).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(2).tpm = {[template ',2']};
matlabbatch{1}.spm.spatial.preproc.tissue(2).ngaus = 1;
matlabbatch{1}.spm.spatial.preproc.tissue(2).native = [1 0];
matlabbatch{1}.spm.spatial.preproc.tissue(2).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(3).tpm = {[template ',3']};
matlabbatch{1}.spm.spatial.preproc.tissue(3).ngaus = 2;
matlabbatch{1}.spm.spatial.preproc.tissue(3).native = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(3).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(4).tpm = {[template ',4']};
matlabbatch{1}.spm.spatial.preproc.tissue(4).ngaus = 3;
matlabbatch{1}.spm.spatial.preproc.tissue(4).native = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(4).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(5).tpm = {[template ',5']};
matlabbatch{1}.spm.spatial.preproc.tissue(5).ngaus = 4;
matlabbatch{1}.spm.spatial.preproc.tissue(5).native = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(5).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(6).tpm = {[template ',6']};
matlabbatch{1}.spm.spatial.preproc.tissue(6).ngaus = 2;
matlabbatch{1}.spm.spatial.preproc.tissue(6).native = [0 0];
matlabbatch{1}.spm.spatial.preproc.tissue(6).warped = [0 0];
matlabbatch{1}.spm.spatial.preproc.warp.mrf = 1;
matlabbatch{1}.spm.spatial.preproc.warp.cleanup = 1;
matlabbatch{1}.spm.spatial.preproc.warp.reg = [0 0.001 0.5 0.05 0.2];
matlabbatch{1}.spm.spatial.preproc.warp.affreg = 'mni';
matlabbatch{1}.spm.spatial.preproc.warp.fwhm = 0;
matlabbatch{1}.spm.spatial.preproc.warp.samp = 3;
matlabbatch{1}.spm.spatial.preproc.warp.write = [0 0];
spm_jobman('run',matlabbatch);
%end newSegSub()
```

Code sample 3: RMSD calculation (matlab)

```
function [rms_whole] = calc_def_rms(defA, defB)
addpath('/home/hanayik/spm12/');
addpath('/home/hanayik/dicm2nii/');
mfilepth = mfilename('fullpath');
spmDir = spm('Dir');
brainmask = fullfile(spmDir,'canonical','single_subj_T1_mask.nii');
if nargin < 1
   [ui_na, ui_pa, ~] = uigetfile('*.nii', 'Choose deformation Image');
   defA = fullfile(ui_pa, ui_na);
   [ui_nb, ui_pb, ~] = uigetfile('*.nii', 'Choose deformation Image');
   defB = fullfile(ui_pb, ui_nb);
end

disp(defA);
disp(defB);
lesmask = '';
les_pattern_id = '_M';
bmask_pattern = 'BrainMask.nii';
if strfind(defA, les_pattern_id)
   pth = fileparts(defA);
   lesmaskfile = dir(fullfile(pth, 'cLesion*.nii'));
   lesmaskfileALT = dir(fullfile(pth, 'Prediction3_native*.nii'));
   if ~isempty(lesmaskfile)
      lesmask = fullfile(pth,lesmaskfile(1).name);
   elseif ~isempty(lesmaskfileALT)
      lesmask = fullfile(pth,lesmaskfileALT(1).name);
   end
   %ANTS only
   bmaskfile = dir(fullfile(fileparts(defB), bmask_pattern)); % if lesion found here, use
      BrainMask from other folder (syn only vs. Linda)
   if ~isempty(bmaskfile)
      brainmask = fullfile(fileparts(defB), bmaskfile(1).name);
   end
elseif strfind(defB, les_pattern_id)
   pth = fileparts(defB);
   lesmaskfile = dir(fullfile(pth, 'cLesion*.nii'));
   lesmaskfileALT = dir(fullfile(pth, 'Prediction3_native*.nii'));
   if ~isempty(lesmaskfile)
      lesmask = fullfile(pth,lesmaskfile(1).name);
   elseif ~isempty(lesmaskfileALT)
      lesmask = fullfile(pth,lesmaskfileALT(1).name);
   end
   %ANTS only
```

```
    bmaskfile = dir(fullfile(fileparts(defA), bmask_pattern)); % if lesion found here, use
        BrainMask from other folder (syn only vs. Linda)
    if ~isempty(bmaskfile)
        brainmask = fullfile(fileparts(defA), bmaskfile(1).name);
    end
end
%brainmask = fullfile(ui_pa,'BrainMask.nii'); % for ANTS
niiA = nii_tool('load', defA);
niiB = nii_tool('load', defB);
niiM = nii_tool('load', brainmask);
imgM = double(niiM.img);

if ~isempty(lesmask)
    niiLesM = nii_tool('load', lesmask);
    imgLesM = double(niiLesM.img);
else
    %imgLesM = double(ones(size(niiM.img)));
end


sA = size(niiA.img);
sB = size(niiB.img);
pdA = round(niiA.hdr.pixdim(2:4),1);
pdB = round(niiB.hdr.pixdim(2:4),1);

if ~isequal(sA, sB)
    error('Images must be same dimensions');
end

if ~isequal(pdA, pdB)
    error('pixle dimensions do not match between images');
end

pixdim = round(niiA.hdr.pixdim(2:4),1);
%%%%%%%%%%%%%%%%%%%% whole brain (lesion + outside lesion)
imgAx = niiA.img(:,:,:,:,1);
imgAy = niiA.img(:,:,:,:,2);
imgAz = niiA.img(:,:,:,:,3);

imgBx = niiB.img(:,:,:,:,1);
imgBy = niiB.img(:,:,:,:,2);
imgBz = niiB.img(:,:,:,:,3);

imgAx = imgAx(imgM ~= 0);
imgAy = imgAy(imgM ~= 0);
imgAz = imgAz(imgM ~= 0);
```

```
imgBx = imgBx(imgM ~= 0);
imgBy = imgBy(imgM ~= 0);
imgBz = imgBz(imgM ~= 0);

dab = sqrt((imgAx-imgBx).^2 + (imgAy-imgBy).^2 + (imgAz-imgBz).^2);
rms_whole = sqrt (mean (dab.^2));
save(fullfile(pth,'rms_whole.mat'), 'rms_whole');
```

Code sample 4. Calculate landmark distance from average landmarks (matlab)

```
function get_warped_landmark_loc(healthyLm, lesionLm)

addpath('/home/hanayik/spm12/');
addpath('/home/hanayik/dicm2nii/');
maxNlandmarks = 18;
disp(healthyLm);
disp(lesionLm);

niiH = nii_tool('load', healthyLm);
niiL = nii_tool('load', lesionLm);

sH = size(niiH.img);
sL = size(niiL.img);
if ~isequal(sH, sL)
   error('Images must be same dimensions');
end

pdH = round(niiH.hdr.pixdim(2:4),1);
pdL = round(niiL.hdr.pixdim(2:4),1);
if ~isequal(pdH, pdL)
   error('pixle dimensions do not match between images');
end

%For calculating landmark center of mass position: [x,y,z] =
      ind2sub(size(nii.img),find(nii.img == 1));
% should be 18 landmarks per image
mxH = max(max(max(niiH.img)));
mxL = max(max(max(niiL.img)));
disp(mxH);
disp(mxL);
d = zeros(1,mxH);
les_lm_com = zeros(mxH, 3);
avgd = [];
if mxH < maxNlandmarks
   [p, nm, ex] = fileparts(lesionLm);
   % save to same folder as lesion landmark image
   lmdisplacementfile = fullfile(p, ['avg_lm_disp_' nm '.mat']);
   save(lmdisplacementfile, 'avgd');

   les_lm_com_file = fullfile(p, ['les_lm_com_' nm '.mat']);
   save(les_lm_com_file, 'les_lm_com');
   error('less than 18 landmarks found in image')
end
if mxL < maxNlandmarks
```

```
    [p, nm, ex] = fileparts(lesionLm);
    % save to same folder as lesion landmark image
    lmdisplacementfile = fullfile(p, ['avg_lm_disp_' nm '.mat']);
    save(lmdisplacementfile, 'avgd');

    les_lm_com_file = fullfile(p, ['les_lm_com_' nm '.mat']);
    save(les_lm_com_file, 'les_lm_com');
    error('less than 18 landmarks found in image')
end

for i = 1:mxH
    [xH,yH,zH] = ind2sub(size(niiH.img),find(niiH.img == i));
    [xL,yL,zL] = ind2sub(size(niiL.img),find(niiL.img == i));
    cxH = mean(xH); cyH = mean(yH); czH = mean(zH);
    cxL = mean(xL); cyL = mean(yL); czL = mean(zL);

    disp(sprintf('Control mean x,y,z (%i): %d, %d, %d',i, cxH, cyH, czH));
    disp(sprintf('Lesion mean x,y,z (%i): %d, %d, %d',i, cxL, cyL, czL));

    % calculate 3D euclidean distance
    d(1,i) = sqrt((cxH - cxL)^2 + (cyH - cyL)^2 + (czH - czL)^2);
    les_lm_com(i,:) = [cxL cyL czL];

end

avgd = mean(d); % * voxel dimensions to get mm
disp(sprintf('Average landmark displacement(mm): %d', avgd));
[p, nm, ex] = fileparts(lesionLm);

% save to same folder as lesion landmark image
lmdisplacementfile = fullfile(p, ['avg_lm_disp_' nm '.mat']);
save(lmdisplacementfile, 'avgd');

les_lm_com_file = fullfile(p, ['les_lm_com_' nm '.mat']);
save(les_lm_com_file, 'les_lm_com');
```

Code sample 5. NCC calculation code (matlab)

```
function [ncc, ncc_les] = computeWarpedLesionImgNCC_with_lesion(warpedListStr,
      maskfile, lestype, jobsavepth)
addpath('/home/hanayik/dicm2nii/');
warpedList = strsplit(warpedListStr,',')';
disp(warpedList);
disp(maskfile);
n = size(warpedList,1);
[pM, nM, eM] = fileparts(maskfile);
nccfile = fullfile(jobsavepth, ['ncc_' lestype '_' nM '.mat']);
if exist(nccfile,'file')
   return;
end
ncc = zeros(n,n)+NaN;
ncc_les = zeros(n,n)+NaN;
compname = cell(n,n);
poolobj = gcp('nocreate');
delete(poolobj);
parpool(24);
parfor i = 1:n
   for j = 1:n
      warpA = warpedList{i};
      warpB = warpedList{j};
      if strcmpi(warpA, warpB) % omit same file combos
         ncc(i,j) = NaN;
         ncc_les(i,j) = NaN;
         compname{i,j} = NaN;
         continue
      else
         disp(warpA);
         disp(warpB);
         niiA = nii_tool('load', warpA);
         niiB = nii_tool('load', warpB);
         niiA.img = double(niiA.img);
         niiB.img = double(niiB.img);
         niiMask = nii_tool('load', maskfile);
         niiMask.img = double(niiMask.img);
         if isequal(size(niiA.img), size(niiB.img), size(niiMask.img))
            disp('All image sizes are equal! yay!');
         else
            disp('Image sizes not equal!')
         end
         imgA = niiA.img(:);
         imgB = niiB.img(:);
         imgM = niiMask.img(:);
```

```
        imgA(isnan(imgA))=0;
        imgB(isnan(imgB))=0;
        imgM(isnan(imgM))=0;
        %%% this matches values returned by fslcc function
        imgAd = imgA-mean(imgA);
        imgBd = imgB-mean(imgB);
        ncc(i,j) = sum(imgAd.*imgBd)/sqrt(sum(imgAd.^2).*sum(imgBd.^2));

        Ales = imgA(imgM > 0);
        Bles = imgB(imgM > 0);
        Alesd = Ales-mean(Ales);
        Blesd = Bles-mean(Bles);

        ncc_les(i,j) = sum(Alesd.*Blesd)/sqrt(sum(Alesd.^2).*sum(Blesd.^2));

        [pA, nA, eA] = fileparts(warpA);
        [pB, nB, eB] = fileparts(warpB);

        if strfind(warpA, 'eSubject')
           disp('fixing ncc filename for ants enat results');
           sA = strsplit(warpA,'/');
           nA = ['e_' sA{end-2}];
        elseif strfind(warpA, 'Subject')
           disp('fixing ncc filename for ants cfm results');
           sA = strsplit(warpA,'/');
           nA = sA{end-2};
        end

        if strfind(warpB, 'eSubject')
           disp('fixing ncc filename for ants enat results');
           sB = strsplit(warpB,'/');
           nB = ['e_' sB{end-2}];
        elseif strfind(warpB, 'Subject')
           disp('fixing ncc filename for ants cfm results');
           sB = strsplit(warpB,'/');
           nB = sB{end-2};
        end

        compname{i,j} = [nM '_' nA '_to_' nB];
      end
   end
end
save(nccfile,'ncc','ncc_les','compname');
```