Investigating Lesion Size and Neuroimaging Methods with Magnetic Resonance Imaging in People With Aphasia

Meara Brown, Lisa Filion, Margaux Keith, Abbey Massar

Supervisors: Drs. Jacqueline Cummine and Esther Kim

Lesion size and tracing methods with MRI

ABSTRACT

The aim of this study was to investigate the effect of lesion size on performance on various cognitive and linguistic tasks in three individuals with left-hemisphere stroke. Findings from previous research studies have indicated mixed results, with some studies arguing that the location of the lesion has a greater impact on recovery after stroke. Researchers were also interested in determining which method of lesion mapping generated the most accurate measure of the individual's lesion size while reducing mapping duration. Subjects underwent an MRI scan, after which lesions were mapped onto brain images using the MRIcron program. Graphs were generated that plotted lesion size against participant performance on each task and data was analyzed through visual inspection and slope analysis of graphed lines. No consistent pattern of decline in task performance was observed with increasing lesion size across linguistic and cognitive tasks. In addition, it was determined that the manual method of mapping (MTM) was superior to the semi-automated technique (STM), as it provided greater flexibility and a higher level of accuracy, especially for larger lesions.

INTRODUCTION

Aphasia is defined as an acquired disturbance in expressive or receptive language function as a result of damage to brain areas responsible for language. The functional impact (e.g., severity of aphasia) of brain lesions can be evaluated through the calculation of lesion size, and a comparison of that lesion size with performance on behavioural tasks. A number of factors, including lesion size, lesion location and time post onset have been correlated with task performance (Marcotte, et al., 2012; Parkinson, Raymer, Chang, FitzGerald & Crosson, 2009;

Sebastian & Kiran 2012; Walker, Scwartz, et al., 2011), yet there is still some uncertainty about the relationship between lesion characteristics and behavioural performance. This study aims to evaluate the impact of lesion size on task performance, as well as the effectiveness of two different types of lesion mapping methods in accurately representing the areas encompassed by lesion.

Lesion Characteristics and Task Performance

While the significance of lesion size in predicting task performance is relatively consistent across research investigating acute stroke patients (Tsapkini, Frangakis & Hillis, 2011; Mazzoni et al., 1992), results from studies of chronic stroke patients are mixed. Chronic stroke generally refers to those patients who are at least one year post stroke and onwards. Many investigators have presented evidence suggesting no correlation between lesion size and task performance in chronic stroke patients (Marcotte et al., 2012; Walker et al., 2011). For example, Marcotte et al. (2012) investigated the cortical changes that occur after Semantic Feature Analysis (SFA) therapy. Results demonstrated that SFA therapy positively impacted clients; however, there was no correlation found with lesion size. Individuals with a lesion involving Broca's area demonstrated less improvement on naming tasks. Therefore, specific cortical areas impacted by a lesion seemed to be a better predictor of therapy outcome than size (Marcotte et al., 2012).

Recent work on the functional neuroanatomy of language has informed our knowledge of specific cortical areas that are involved in language tasks. For example semantic decision tasks such as determining if a sentence is plausible or implausible involves the activation of areas involved with semantic knowledge, speech perception, language comprehension and orthographic knowledge (if reading is involved). Such areas include the left anterior temporal lobe, prefrontal cortex and inferior parietal cortex (Walker et al., 2011; Hickok, 2009). A lesion involving any of these areas could directly influence semantic task performance. Other tasks commonly used to measure performance for patients with aphasia include picture naming, picture word matching and single-word lexical decision-making. Picture naming tasks typically activate extensive cortical networks (Sebastian et al., 2012). These include bilateral superior and middle temporal lobes, the left angular gyrus, left inferior frontal lobe and bilateral occipital lobes (Sebastian et al., 2012). A Picture Word Matching (PWM) task involves activation of areas associated with semantics as well as orthography. These include the posterior inferior occipito-temporal cortex for orthographic processing and the anterior temporal lobe, angular gyrus and anterior inferior frontal gyrus for semantic processing (Kim, 2012; Hickok, 2009). Lexical decision tasks, such as determining whether a written word is plausible or implausible, also involve the cortical areas involved with orthographic and semantic processing. It is important to understand the neuroanatomy of language in order to fully comprehend the relationship between lesion location and post-treatment performance.

Research investigating how lesion size and cortical area impact therapy outcome has led to some noteworthy findings. A study by Parkinson et al., (2009) examined the impact of lesion size and location in chronic aphasia on performance and improvement observed on object and action naming tasks. Baseline measures showed that larger anterior cortical lesions were correlated with higher pre-treatment naming. In addition post-treatment measures revealed that those with larger anterior cortical lesions demonstrated greater improvement in these tasks. Parkinson and colleagues acknowledged that these findings (i.e. larger lesions lead to better cognitive functioning) seem controversial and discussed a possible hypothesis for this correlation. Activity in the left frontal lobe may lead to interference with naming tasks. A large role for language centres of the brain is to suppress competing information when selecting the target word. Significant damage to the anterior regions of the brain may remove the competing 'noise' during naming tasks to allow for reorganization to occur (Parkinson et al., 2009).

The inconsistent findings have led some researchers to conclude that performance on tasks is influenced by several factors; with lesion location being the most influential (Marcotte et al., 2012; Sebastian et al., 2012; Walker et al., 2011). For example, a study by Sebastian et al., (2012) demonstrated that the greater the amount of peri-lesional area preserved, the more successful patients were in oral picture-naming and semantic judgement tasks. Identifying areas impacted by the lesion through an accurate and representative map of the damaged area is essential in determining task performance as well as appropriate therapy approaches to address the skill set of each individual.

Lesion Mapping

One potential contributing factor to the mixed results regarding lesion size/location and behavioural performance could be how the lesion is delineated. Brain morphometry has been used in research on many different populations for various reasons; in neuropathological populations it is used to pinpoint brain regions that are different from typical brains. Since Broca declared that the inferior frontal gyrus was used for speech production, the 'lesion method' of evaluating lesions in post-mortem brains, and later in MRI structural analyses, was seminal for understanding neurophysiology in humans (Rorden & Karnath, 2004). Manual tracing of lesions, done by an expert, is considered the gold-standard of lesion localization

(Wilke, de Hann, Juenger & Karnath, 2011). This method is used in experimentation, research, and medical diagnosis (Fillipi et al., 1995; Gonzalez-Fernandez et al., 2011; Tsapkini et al., 2011; Wicks et al., 1992; Wilke et al., 2011). The process of Manual Tracing Method (MTM) is outlined in Appendix A. It is precise and accurate, however the trade-off is time; an expert lesion mapper (characterized as someone with more than 20 hours of experience mapping lesions [Bogovic, et al., 2012]) takes 5 to 10 minutes per slice (Wilke et al., 2011). As such, creating a three-dimensional volume of interest (VOI) on a standard 144 slice MRI scan could take 12 to 24 hours to depending on the complexity of the lesions. Wilke et al. (2011) reported 4.8-9.6 hours per subject in their study of unilateral middle cerebral artery stroke. For white matter lesions, Filippi et al. (1995) indicate that manual delineation technique was on average 90 minutes for each brain. Ultimately, while expert manual tracing is and should stay the gold standard for lesion mapping, it is a time intensive process.

In order to increase the speed of lesion mapping, algorithms based on voxel intensity have been created (Lerch, et al., 2008). These programs can be semi-automatic (requiring some manipulation by an expert) or automatic (requiring no manipulation) (Wilke et al., 2011). The semi-automatic tracing (SAT) method is preferred over the automatic method in instances where the ventricles and lesion sites are in close proximity (such as the participants in the present paper). In such cases, the automatic method alone cannot differentiate between damaged brain tissue and the cerebral spinal fluid of the ventricles. The SAT does not suffer from such a limitation and was created to address the concerns that manual delineation was not suited to recognize changes in the brain that are less noticeable than a focal lesion. For example, areas of the brain far from the lesion can also be affected functionally even if it appears structurally the same (Wilke et al., 2011). SAT is a tool that can be used to recognize very small differences in voxels intensity in areas of the brain adjacent to the focal lesion, which can provide researchers with a more comprehensive understanding of the brain (Wilkes et al., 2011). There are several computer programs on which to run SAT lesion mapping, and some research has shown these programs to have an advantage over manual mapping in regards to precision and speed (Ashton et al., 2003).

Understandably, inter-rater reliability has been a concern in lesion mapping since the technique was introduced. Much of MRI research is conducted with multi-site studies and hundreds of participants (Filippi et al., 1995). Filippi et al. (1995) found that there was a high intra- and inter-rater agreement in SAT lesion measurements compared to MTM when evaluating white matter lesion volume in patients with multiple sclerosis. Additionally, there was significantly lower intra- and inter-rater variability of SAT compared to the MTM (Filippi et al., 1995).

Purpose of Study

This study was conducted with two research questions in mind. First, we were interested in determining the effect of lesion size on behavioural performance for various linguistic and cognitive tasks. Based on previous studies regarding lesion size, we posit that individuals with larger, more diffuse lesions will demonstrate reduced levels of performance for these tasks. The second question of interest was whether use of the semi-automatic method of lesion delineation would result in a more exact representation of participant lesions while reducing the required time demands. Based on research already performed on optimal tracing methods, we anticipated that the manual tracing method would be more exact but less efficient.

METHODS

Participants

Three participants with left-hemisphere stroke participated in this study. Participant 1 (TH) was a 67 year-old male, 4 years post-stroke at the time of testing. He was a monolingual English speaker, had 8 years of formal education and worked for a pipeline drilling company prior to his stroke. Participant 2 (RD) was also a 67 year-old male, 4 years post-stroke at the time of testing. His highest level of education completed was grade 8, and he was retired at the time of his stroke (previously worked as a janitor). RD was fluent in both English and French, but reported English was his primary language at the time of testing. Participant 3 (RJ) was a 53 year-old female, 5 years post-stroke at the time of testing. Her highest level of education was a two-year diploma from a technical institute. She worked as a human resources recruiter prior to her stroke. RJ spoke both English and German; however, she indicated that English was her primary language at the time of testing.

Behavioural Measures

Behavioural measures were taken from standardized tests, as well as four non-standardized linguistic tasks. Standardized test measures comprised subtests from two test batteries commonly used with individuals with aphasia. These included the Composite Auditory Comprehension score from the *Western Aphasia Battery-Revised (WAB-R;* Kertesz, 2007), as well as the Attention Composite score and the Memory Composite score from the *Cognitive Linguistic Quick Test (CLQT;* Helm-Estabrooks, 2001).

In addition to these measures, participant's accuracy scores from four non-standardized linguistic tasks were used. These consisted of a Plausible/Implausible Sentence Decision (SD) task (discriminating between 30 plausible and 30 implausible sentences), a Picture-Naming task, a Picture-Word Matching task, and a Word/Pseudoword Lexical Decision (LD) task. In the SD task, subjects listened to sentences through headphones, and indicated whether they thought the sentence was plausible or implausible by pressing a button. In the Picture-Naming task, participants were asked to name pictures as they appeared on a computer screen. Participant's responses were transcribed both online and after the task through a recording. In the Picture-Word Matching task, a picture appeared on one side of the computer screen, and a word appeared on the other side of the screen. Participants used a mouse to indicate whether the word presented was the label for the picture. The Word/Pseudoword Lexical Decision task was similar to the Sentence Decision task; participants heard words and used a mouse to indicate whether or not they thought the word was a real word.

Imaging

Subjects were scanned using a 1.5 Tesla Siemens Sonata MRI. High resolution anatomical images were acquired using an AXIAL T1 MPRAGE sequence with the following parameters: T_R = 2000ms, T_E = 4.38ms, slice thickness 1 x 1 x 1, number of slices = 144, base resolution = 256x256.

Lesion Mapping

MRI images were processed through *SPM8©* (FIL, 1994-2012) to create an image that could be used for creating lesion maps in the program *MRICron ©* (Rorden, 2009). All three participants' brains were mapped using two methods: Manual Tracing Method (MTM) and Semi-Automatic Tracing (SAT). Both types of mapping were performed by a group of four graduate students in the M.Sc program for Speech-Language Pathology at the University of Alberta.

MTM of the lesions was performed using *MRICron©.* The four graduate students ("mappers") manually delineated lesions on an axial view of the brain in *MRICron*© (consult Figure 1 below). This was performed using the *Draw* function to outline the area deemed to be lesion using a mouse, and the *Fill* function to incorporate all of the brain mass within the border created into a lesion map (see MRICron Index [Rorden, 2009] for information regarding these functions). This was done on a 'slice-by-slice' basis. In the initial stage in the process, each "mapper" mapped the same brain without consulting each other. After this, the four "mappers" met to compare maps. A standard protocol was devised based on these comparisons and was used to map all three subject brains. Consult Appendix A for the standard mapping protocol used with the manual method of mapping.

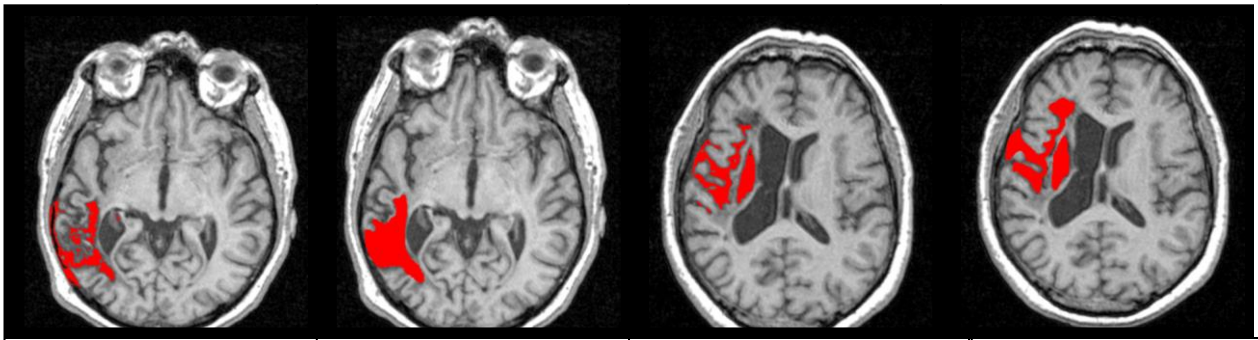
Each "mapper" used the standard protocol to create lesion maps for the three participants' brains. After all of the brains had been mapped individually, the four "mappers" met to compare lesion maps. One "Master Map" for each subject was chosen based on consensus. Four criteria were used to choose the most appropriate "Master Map": 1) the lesion delineation was consistent across slices within the same brain, 2) the map was consistent with the logic used for other brains, 3) the map adhered to the pre-established protocol, and 4) the selected map was representative of all four “mappers” delineations (i.e. was closest to “average” of all four maps). One “mapper’s” lesion maps appeared to be the most consistent across subjects. For this reason, this person’s map was used as the default master map in cases of uncertainty of which map to choose. It should be noted that one participant’s MRI images were used as a sample for all four “mappers” to devise a protocol. This should be taken into consideration when evaluating inter-rater reliability.

The SAT of lesions was performed with all four “mappers” consulting on one lesion map per slice on each of the subjects’ brains (see Figure 1 below). Again, a protocol was devised to ensure consistency across subject brains. Consult Appendix B for protocol used with the semi-automatic method of mapping.

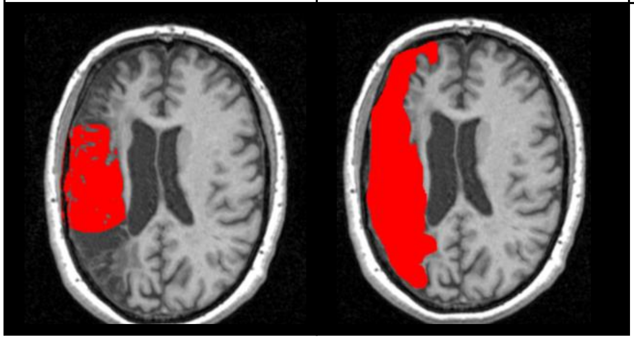
Following creation of each lesion map, lesion size in voxels was calculated using *Compute Statistics* function in *MRICron©*. These calculations were analyzed against the aforementioned behavioural measures, as well as both methods of mapping.

RESULTS

Slope analyses and visual analyses were completed for graphed data plotting lesion size against behavioural measures (subtests of WAB-R, CLQT, non-standardized tasks). Slopes were non-standardized and calculated by subtracting each plotted point on a line and dividing the resulting numbers (i.e. calculating the ratio of the rise divided by the run). Slope analysis was supplemented by a visual analysis, in which the researchers would visually examine the steepness of each line plotted on the graph relative to other lines to determine whether an effect existed. Slope and visual analyses were then compared to determine instances in which the two methods of analysis were in agreement. When this was not the case, greater weight was given to the slope analysis.



P01 SAT (Slice 54) | P01 MTM (Slice 54) | P02 SAT (Slice 54) | P02 MTM (Slice 54)



P03 SAT (Slice 80) | P03 MTM (Slice 80)

Figure 1. Lesion delineations performed by “mappers” for each participant using semi-automated and manual methods of mapping.

Effect of Lesion Size on Language Task Performance

For the Picture-Word Matching task, visual analysis for both the semi-automated and manual methods indicated no relationship between task performance and lesion size (measured by the number of voxels circumscribed within the lesion) for the graphed line between participant 1 and 2. Slope analysis further supported this finding, with the slope of the line measuring -1.08 for the semi-automated technique and -0.43 for the manual method of mapping (consult Figure 2 and 3 below for graphed data used during the analyses of each tasks). Visual observation of the line passing between participant 1 and 2 for the Picture Naming and Plausible/Implausible SD tasks showed a relationship between these tasks and lesion size for both manual and semi-automated methods. Once again this finding was further supported by slope analyses computed for the graphed results (SAT [picture naming] =12.97; MTM [picture naming] = 5.21; SAT [plaus vs. imp] = 8.11; MTM [plaus vs. imp] = 3.25). It should be noted that although the line between participant 1 and 2 showed an effect for these tasks, the value of each slope was positive, indicating an increase in task accuracy with increasing lesion size. Lastly, though a decrease in task accuracy was evident for the Word/Pseudowords task, visual and slope analysis did not determine it to be great enough to denote a relationship between lesion size and performance on this task.

Observation and slope analyses calculated for the line passing between participant 2 and 3 for the Picture Word Matching task revealed no relationship between lesion size and performance on this task (SAT= .03; MTM= .01). With regards to the Implausible/Plausible SD and Word/Pseudowords tasks, the lines for these tasks did not visually appear to be steep enough to indicate that a relationship exists (SAT [Plaus vs. Implaus] = -0.40; MTM [Plaus vs. Implaus] = -0.13; SAT [Pseudowords]= -0.4; MTM [Pseudowords] = -0.1). A relationship was believed to exist for the Picture Naming task, which appeared to visually demonstrate a notable decline in performance with increasing lesion size (SAT= -0.8; MTM= -0.26).

In sum, the only task that demonstrated a steady decline in task performance was the Word/Pseudowords task, although this decline was not great enough to indicate that a relationship exists. Other tasks either displayed no relationship between task performance and lesion size (i.e. the line was horizontal) or a varied pattern of performance. In-fact, in two cases participant 2 outperformed participant 1, resulting in a spike in the graph for the Plausible/ Implausible and Picture Naming tasks. Although declines in task performance for participant 3 were present, the only task in which these declines were determined to be noteworthy was the Picture Naming task.

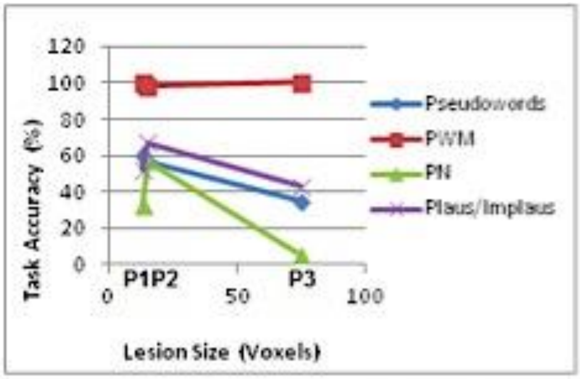


Figure 2. Effect of lesion size on various language tasks as measured using the semi-automated method of lesion mapping.

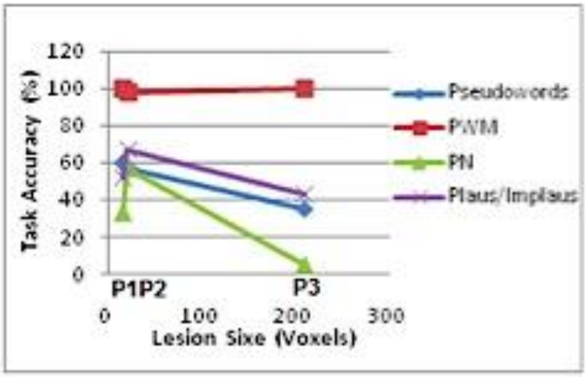


Figure 3. Effect of lesion size on various language tasks as measured using the manual method of lesion mapping.

Effect of Lesion Size on Standardized Measures of Auditory Comprehension

Visual and slope analyses were computed for the effect of lesion size on auditory comprehension scores from the *Western Aphasia Battery (WAB*), a standardized measure used to assess adults with aphasia. A steady decline in WAB auditory comprehension scores was observed as the participant's lesion size increased. Participant 2's score in this case was substantially lower than participant 1, resulting in a steep line between the two participants (SAT= 0.32; MTM=0.13). Participant 3 displayed a lower WAB score than participant 2 and therefore the line connecting these two participants had a negative slope. However, analysis of the slope of this line and a visual analysis did not denote a relationship between lesion size and WAB auditory comprehension scores (SAT= -0.04; MTM= -.01).

Overall, a steady decline in WAB auditory comprehension scores was observed as participant lesion size increased. However; a relationship was only found to exist for the line connecting participant 1 and 2 and not for participant 3, who has a significantly larger lesion (see Figures 4 and 5).

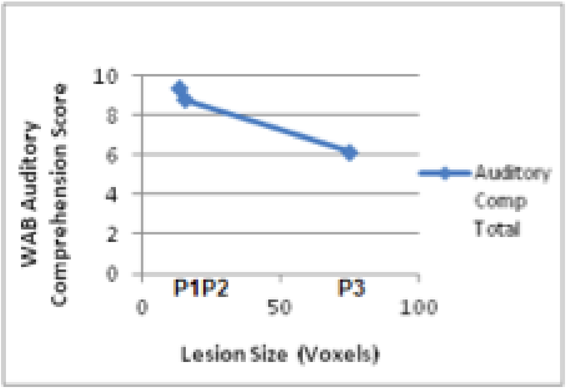


Figure 4. Effect of lesion size on auditory comprehension scores using the semi-automated method of lesion mapping.

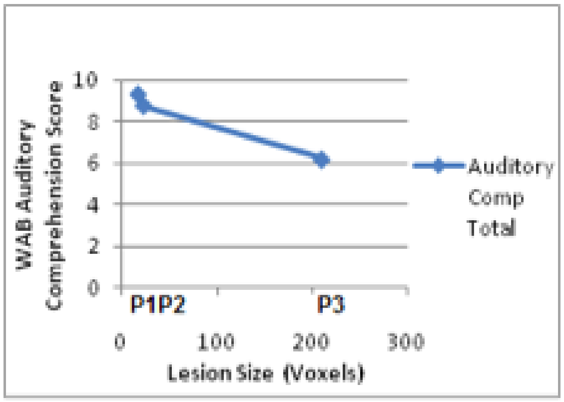


Figure 5. Effect of lesion size on auditory comprehension scores using the manual method of lesion mapping.

Effect of Lesion Size on Measures of the CLQT

The effect of lesion size on measures of attention and memory for *the Cognitive Linguistic Quick Test (CLQT),* a standardized measure assessing various cognitive domains, was determined through visual and slope analyses. Scores for attention revealed a decline with increasing lesion size (SAT= -9.73; MTM= -3.69). A notable decline was also observed when analyzing the line between participant 2 and 3 (SAT= -1.46; MTM= -0.46).

With regards to the memory component, the line connecting participant 1 and 2 exhibited a fairly steep decline (SAT= -9.19; MTM= -3.90). However, participant 3's performance on tests of memory surpassed that of participant 2 and was equivalent to participant 1's memory score (SAT= 0.28; MTM=0.09; see Figures 6 and 7).

In sum, a decline in attention scores was apparent as lesion size increased across participants. This decline indicates that a relationship may exist between participant attention scores and lesion size. In addition, memory scores were variable across participants, with participant 3 scoring better than participant 2 and equivalent to participant 1 who has the smallest lesion size. Thus, a relationship does not seem to exist between participant memory scores and lesion size.

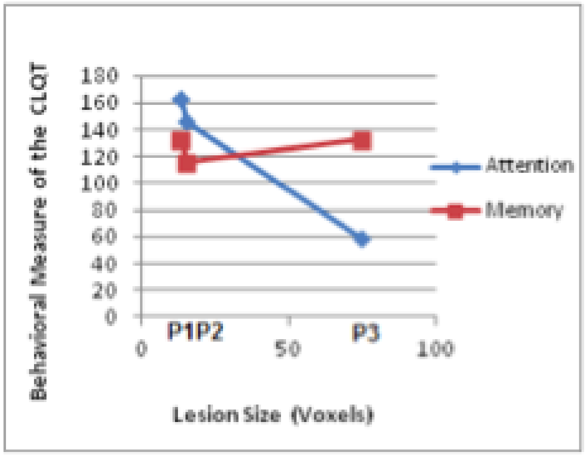


Figure 6. Effect of lesion size on attention and memory scores of the CLQT using the semi-automated method of lesion mapping.

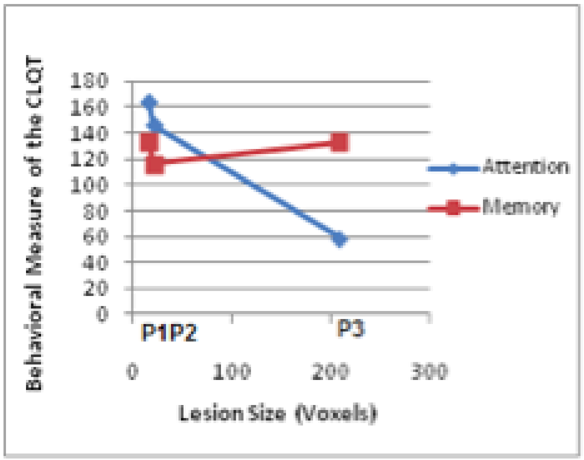


Figure 7. Effect of lesion size on attention and memory scores of the CLQT using the manual method of lesion mapping.

Comparison of Lesion Size for Manual vs. Automated Methods

When comparing the lesion size computed for each method for participant 1, the manual technique generated a lesion size that was 3.32 voxels greater than that obtained for the semi-automated method of lesion delineation (MTM= 16.56 mm^3; SAT= 13.24 mm^3). Participant 2 exhibited similar results, with the manual method generating a lesion size of 6.08 voxels greater than the semi-automated method (MTM=21.17 mm^3; SAT=15.09 mm^3). Participant 3 displayed markedly different voxel counts for the two methods, with a difference of 134.91 voxels between the two methods (MTM=209.6 mm^3; SAT=74.69 mm^3). In each case the semi-automated technique of lesion delineation had a tendency to calculate lesions as having fewer voxels whereas the manual method was more inclined to circumscribe a greater number of voxels during delineation (mean of differences= 48.10 mm^3). Though differences in lesion size did exist between the two methods, the values obtained were comparable for participant 1 and 2. Participant 3’s lesion size values were not comparable for the two methods.

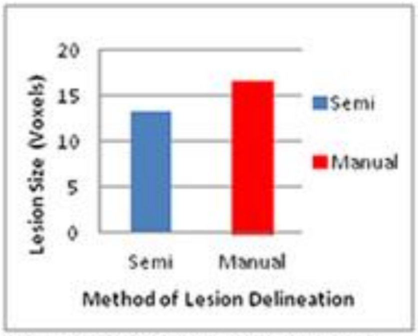


Figure 8. Comparison of lesion size (voxels) for manual vs. semi-automated methods of lesion mapping for participant 1.

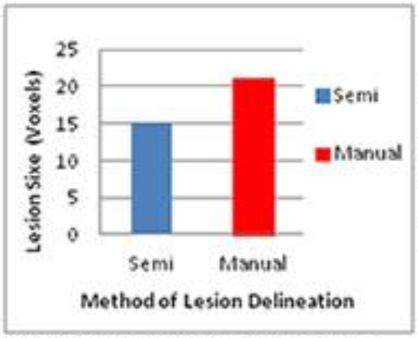


Figure 9. Comparison of lesion size (voxels) for manual vs. semi-automated methods of lesion mapping for participant 2.

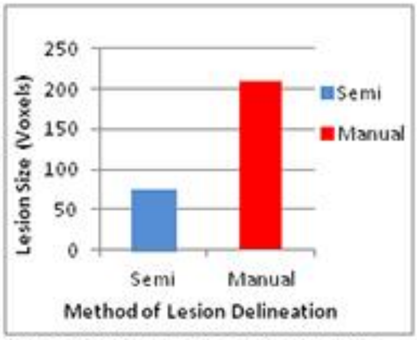


Figure 10. Comparison of lesion size (voxels) for manual vs. semi-automated methods of lesion mapping for participant 3.

DISCUSSION

Evaluation of the Effect of Lesion Size on Task Performance

Analysis of the data obtained did not reveal a consistent pattern of decline in performance with increasing lesion size for the linguistic and cognitive tasks measured in this study. Rather, some linguistic and cognitive tasks demonstrated superior performance as the lesion size increased (e.g. CLQT measure of memory). For tasks in which declines in performance were noted, only the decrease in performance for the Picture Naming task was found to be large enough to indicate that lesion size had an effect on performance.

The dramatic decline in accuracy for the picture naming task for participant 3 is not believed to be the result of the extensive lesion size impacting her ability to access semantic knowledge but rather her severe apraxia of speech. The reasons for this are twofold; first, she was successful in the Picture-Word Matching task, and second, she did not misname the items in the Picture Naming task. Using knowledge of the nature of the Picture-Word Matching task, what it tests, as well as componential analysis (Davis, 2007), we can infer that participant 3 has a fairly intact object recognition and semantic knowledge system. Componential analysis refers to the theory that semantic information is stored in a system of unique properties using the minimal number of features (Lewis-Beck, Bryman, and Liao, 2004). Therefore, the cognitive neuropsychology model of naming (Bruce & Howell, 1988) tells us that the breakdown must have occurred in the phonological output lexicon or the speech motor mechanisms. Our second reason supports the first if we consider it in the context of the cognitive neuropsychology model of naming. The low accuracy score participant 3 obtained for the Picture Naming task was not because she named 57 words incorrectly but instead due to the fact that she could only produce three of the words presented. Thus, poor performance on this task cannot be attributed to difficulties with picture naming caused by her extensive lesion size.

With regards to the declines in attention observed across participants, the reasoning behind this finding may result from an interaction between both lesion size and location. When examining brain images presented in Figure 1, it is apparent that participant 1's lesion is located in the posterior regions of the brain, whereas participant 2's lesion is more anterior and subsumes a portion of the frontal lobe. Participant 3 on the other hand, has a large area of the frontal lobe that has been damaged. According to Filley (2002), networks involved in attention are distributed through many cortical and subcortical structures; including a wide network of thamalic and bihemispheric frontal lobe structures. Thus, the large number of networks involved in attention that are located within the frontal lobes may explain why participant 2 and 3, whose lesions encapsulated regions of the frontal lobe, performed poorer than participant 1, whose lesion did not appear to impinge on frontal brain regions. Further, since attention is an intricate network of structures throughout the brain, it is unsurprising that a person with a large diffuse lesion, such as participant 3's lesion, would display increased difficulty on tasks that stress the attention system.

Overall, results of this study support the notion that lesion size and location work in combination to affect performance on measures of attention. So, what accounts for the variability in performance evident within several linguistic tasks and measures of memory for the CLQT? Previous research indicates that factors such as age and level of education can influence an individual's ability to perform various tasks after stroke. Krugar et al. (2003) investigated the effect of patient age in early stroke recovery and concluded that relative improvement decreased as age increased and that younger patients demonstrated a more complete recovery. In addition, studies examining the effect of education on aphasia after stroke indicate that individuals with a higher level of education may have greater synaptic function and be more resistant to the effects of stroke, resulting in a reduced level of aphasia severity after stroke (González-Fernández, 2011; Conner et al., 2001). Yet the results of this study do not appear to contradict or offer support for these findings. If age and level of education played a role, one would expect participant 2 to be younger and more educated than participant 1. However, participant 1 and 2 were both of the same age (67 years) and level of education (8 years). Also, one could argue that participant 3, who was younger and more educated, did relatively well compared to her counterparts with significantly smaller lesions.

In addition to age and level of education, the concept of cognitive reserve and compensation might also be an explanation for the variability in task performance observed across participants. Stern (2002) describes cognitive reserve and compensation as the ability to optimize or maximize normal performance in the event of brain injury by recruiting brain networks not typically used during normal processing. He stated that, “People with greater cognitive reserve can experience a larger lesion size before functional impairment is apparent” (Stern, 2002, p. 451). A comparison of the brain activation for various cognitive tasks, across individuals with varying lesions sizes would address this notion and is an avenue for future research.

Evaluation of Mapping Methods

It has been largely agreed upon that Manual Tracing Method (MTM) is the gold standard for lesion delineation under all circumstances (Wilke et. al, 2011, Lerch et al., 2008). However, it has also been widely argued that Semi-Automatic Tracing (SAT) is a viable option in lesion analysis (Wilke et. al, 2011, Lerch et al., 2008). The rationale for this statement is that SAT is significantly more efficient, with minimal disruptions to the integrity of the resulting calculation of lesion size in voxels. In this study, the authors compared these two methods to determine whether there were cases where SAT was inappropriate for accurate calculation of lesion volume. The analysis of the two different mapping methods in three cases enabled the authors to evaluate the accuracy of each method against characteristics of the specific lesion. This process supported the notion that MTM should remain the gold standard, however it also highlighted a case where MTM was the only way to accurately account for brain areas encompassed by lesion. This exploratory research may indicate a potential area for further study. The first evaluation of MTM vs. SAT was between the lesion volume calculated for the same three subjects through the differing methods of mapping. In this comparison, the SAT method of lesion delineation appears to be more conservative, generating smaller lesion sizes with fewer voxels for each participant, while the MTM had a tendency to delineate a greater number of voxels. However, these methods were comparable only for individuals whose lesions were contained within a localized region of the brain (ie. more focal damage). This can be seen in participant 1 and 2, whose lesion size calculations were only distinguished by a few voxels. Alternatively, it appears as though in cases where the lesion is relatively large and involves diffuse damage to brain regions, as is seen in the case of participant 3, SAT may not be appropriate. This is likely a result of the steps required in order to maintain a truly semi-automatic mapping protocol. The "mappers" in this study noted that, while using this protocol matched their perception of the brain regions encompassed by lesion in participants 1 and 2, these steps alone were not sufficient for defining the lesion in participant 3. SAT requires that a "seed" be planted within the lesion. The program then delineates the lesion based on a previously selected radius, determined by pre-established settings. This presented significant challenges when using the semi-automatic method to delineate participant 3's lesion. This participant's lesion encompassed a vast portion of the left hemisphere. "Mappers" found that planting a single seed was not sufficient to delineate the entire lesion, given large and diffuse nature of the damaged area. After using the established protocol for mapping in order to maintain comparable results for this study, "mappers" attempted to manipulate the variables of *MRICron©* to best represent the lesion. In order to accurately account for participant 3's lesion, it was necessary to plant multiple 'seeds', defeating the purpose of using this approach as a quick, automatic method of lesion delineation. An additional complication resulting from this process was that increasing the radius computed by the “seed” resulted in healthy tissue being incorporated into the resulting lesion map. Significant review and adjustment was necessary in order to create the most accurate (i.e. most representative of visual perception of lesion area) map using SAT.

These findings suggest that while semi-automatic lesion mapping is a valid option for many cases of lesion delineation, caution should be exercised when lesions have particularly large or diffuse characteristics. It appears as though one single protocol for both larger, diffuse lesions and smaller more localized lesions is not practical and would be difficult to standardize. The manual method of mapping appears to be better suited when circumscribing larger lesions that comprise diffuse areas of the brain, as it allows the expert to generate a more accurate approximation of the number of voxels impacted by the lesion and does not require that an alternate protocol be generated for special cases of lesions (i.e., lesions where the simple post-calculation adjustments that are typical of semi-automatic lesion mapping [e.g., removal of the ventricle] will not suffice to match expert visual judgement). Essentially, the results of this study show that the manual method of mapping is more flexible than the semi-automatic method of mapping, as no adjustments need to be made based on size or complexity of lesion.

While the research performed in this study is exploratory, the findings suggest that the method of mapping used in a given research study may have implications regarding its’ validity. It could be argued from these findings that MTM is essentially a more conservative test when calculating lesion size. This is to say that if MTM is significantly more representative of expert visual perception of lesion, that its’ use in research contexts suggests a more reliable measure than other methods (i.e. Semi-Automatic Tracing or Automatic lesion mapping). Further research may investigate the validity of the claims made in this study by comparing the effectiveness of the aforementioned tracing methods for lesions grouped by varying size and complexity. The authors suspect that if lesions were separated into large diffuse and small/localized lesions, the semi-automatic method may only prove to be sufficient for the latter group. Further research into this area may inspire a trend towards pre-selection of mapping methods used in a study based on a preliminary judgement of types of lesions involved (i.e. if localized lesions are the focus of a given study, semi-automatic lesion mapping could be used without retracting from the validity of the study).

Limitations

There were several limitations with this study that can be improved upon in future research. The sample size (N=3) severely limited the statistical analysis that could be performed on the data obtained. Many models in statistics require a sample size larger than three in order to be valid. Additionally, with the heterogeneous sample of people with brain injury due to stroke, three individuals are unlikely to be a good example of the wider population. On the other hand, Rorden, Karnath, and Bonilha (2007) comment that most lesion studies are based on smaller sample sizes. This is due to the nature of the lesion method research and the population accessed in hospitals and research facilities. This compilation of several smaller studies begins to show a general overview of brain structure and function. Future studies may consider conducting similar comparisons of lesion mapping methods with a greater number of participants in order to support the conclusions drawn in this paper. It would be informative to evaluate two pre-selected lesion groups (i.e. large/diffuse vs. small/focal) and compare the lesion size generated from semi-automatic and manual methods of lesion delineation.

Limitations with the behavioural tasks and measures must also be addressed. Performance on the Picture-Word Matching task demonstrated a ceiling effect, as almost all participants were able to perform at an extremely high level of accuracy (P1=100%; P2=98%; P3=100%). In addition, information regarding reaction time was not available for the cognitive-linguistic tasks. The authors believe that having this additional information may have given more insight into processing times and the length of time participants needed to formulate responses. Differences in reaction time during the Picture-Word matching task may have revealed variability among participants and thus, may have more accurately captured the behavioural limitations that occurred as a result of the lesion.

Researchers were unable to determine inter-rater reliability for the SAT approach. Semi-automatic delineation of lesions was performed with all four “mappers” consulting on one lesion map per slice on each of the subject’s brains. A research article by Fillippi and colleagues (1995) stated that one single mapper for a semi-automatic method is sufficient, especially when studies involve very small sample sizes. Inter-observer agreements for semi-automated approaches were relatively high largely due to agreement on choice for the threshold for lesion segmentation. Fillippi et al., (1995) also identified that when all observers work closely together there is a factor for inter-observer consistency, thus they were more likely to make similar choices when inconsistencies arose. The four “mappers” developed a mapping protocol together and were trained at the same institution. Therefore, according to Fillippi et al. (1995), similar results are likely to emerge. Having all “mappers” consult on one lesion map per slice during the SAT approach is still consistent with the manual protocol established, and did not appear to hinder production of a representative lesion map.

CONCLUSION

The current study addressed two questions; the effect of lesion size on task performance and whether the semi-automatic method of lesion delineation would produce a more representative lesion map while reducing time demands. First, no consistent pattern of decline on performance was observed with increasing lesion size across linguistic and cognitive tasks. Second, this study revealed that MTM should still remain the gold standard for lesion mapping methods. MTM was more flexible than SAT in that no adjustments need to be made based on size or complexity of lesion. Future research in methods of lesion mapping may consider pre-selection of mapping methods or lesion groups based on size and complexity.

Appendix A

Protocol Established for the Manual Method of Mapping

The devised protocol consisted of the following criteria:

- At each slice, the “mappers” outlined what they deemed to be lesion area based on darkness on the image, paired with general knowledge of brain anatomy
- In cases where a lesion was called into question (e.g. ambiguity as to whether an area contained distinctive sulcus patterns or was part of lesion area), the area was compared with the right (non-damaged) hemisphere for symmetry. If the area in question was structurally similar to that of the right hemisphere, it was not counted as part of the lesion.
- In order to complement the subjective measure of relative darkness, a reference image depicting relative density measures in the same brain was used. The image was obtained using *imagej* software (Rasband), and was visually compared with the MRICron image to aid in the judgement of lesion delineation.
- Some cases occurred where there appeared to be a portion of grey matter within the lesion. If this portion of grey matter was completely surrounded by lesion, the section was encapsulated within the delineation of the lesion.
- Slices above and below area in question were also used as judgement measures to determine whether the area should be included as lesion or healthy

Appendix B

Protocol Established for the Semi-Automated Method of Mapping

The devised protocol is as follows:

- Default settings in MRICron were used for automatic lesion mapping. This was primarily based on the premise set by Wilke et. al. (2011), where the authors noted that changing the pre-set parameters was beyond the scope of this type of research
- All four “Mappers” then evaluated each slice to determine if any portion of the ventricle was included. If ventricle was included, the draw function was used to remove that portion of the automatically generated map.

No other adjustments were made to the automatically generated map.

REFERENCES

Conner, L.T., Obler, L.K., Tocco, M., Fitzpatrick, P.M. & Albert, M.L. (2001). Effect of socioeconomic status on aphasia severity and recovery. *Brain and Language, 78 (2),* 254 257.

Davis, G., (2007). Aphasiology: disorders and clinical practice. Boston, MA: Pearson/Allyn and Bacon.

Filippi, M., Horsfield, M. A., Bressi, S., Martinelli, V., Baratti, C., Reganati, P., . . . Comi, G. (1995). Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis. A comparison of techniques. Brain, 118(6), 1593-1600.

Filley, C. M. (2002). The neuroanatomy of attention. Seminars in Speech and Language, 23(2), 89-98. doi: 10.1055/s-2002-24985.

Gonzalez-Fernandez, M., Davis, C., Molitoris, J., Newhart, M., Leigh, R., Hillis, A. (2011). Formal education, socio-economic status and the severity of stroke after aphasia. *Arch Phys Med Rehabil, 92*, 1809-1813.

Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test*. The Psychological Corporation.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews, 6*, 121- 143.

Kim, E. (2012). SPA 520: *The neural substrates of language, week 1 notes.* [PowerPoint Slides]. Retrieved from https://eclass.srv.ualberta.ca/my/

Kertesz, A. (2007). *Western Aphasia Battery-Revised*, New York: Grune & Stratton.

Kugler, C., Altenhoner, T., Lochner, P. & Ferbert, A. (2003). Does age influence early recovery in ischemic stroke? *Journal of Neurology, 250 (6),* 676-681.

Lerch, J. P., Pruessner, J., Zijdenbos, A. P., Collins, D. L., Teipel, S. J., Hampel, H., & Evans, A. C. (2008). Automated cortical thickness measurements from MRI can accurately separate alzheimer's patients from normal elderly controls. Neurobiology of Aging, 29(1), 23-30. doi: 10.1016/j.neurobiolaging.2006.09.013.

Lewis-Beck, M.S., Bryman, A., Liao, T.F. (2004). Componential Analysis in *The SAGE Encyclopedia of Social Science Research Methods*. Retrieved from http://srmo.sagepub.com/view/the sage-encyclopedia-of-social-science-research-methods/n145.xml doi: 10.4135/9781412950589.

Marcotte, K., Adrover-Roig, D., Damien, B., de Preaumont, M., Genereux, S., Hubert, M. & Asaldo, A. I. (2012). Therapy-induced neuroplasticity in chronic aphasia. *Neuropsychologia, 50*, 1776-1786.

Mazzoni, M., Vista, M., Paradossi, L., Avila, L., Bianchi, F., & Moretti, P. (1992). Spontaneous evolution of aphasia after ischemic stroke. *Aphasiology, 6,* (4), 387-396.

Parkinson, R. B., Raymer, A., Chang, Y., FitzGerald, D. B., & Crosson, B. (2009). Lesion characteristics related to treatment improvement in object and action naming for patients with chronic aphasia. *Brain & Language, 110*, 61-70.

Rasband, W.S. (1997) ImageJ [computer software]. Available from: http://imagej.nih.gov/ij/

Rorden, C. (2009). MRICron [computer software]. Available from: http://www.mccauslandcenter.sc.edu/mricro/mricron/install.html

Rorden, C. (2009). MRICron Index. Retrieved from: http://www.mccauslandcenter.sc.edu/mricro/mricron/

Rorden, C., & Karnath, H. (2004). Using human brain lesions to infer function: A relic from a past era in the fMRI age? Nature Reviews Neuroscience, 5(10), 813-819.

Rorden, C., Karnath, H., & Bonilha, L. (2007). Improving lesion-symptom mapping. Journal of Cognitive Neuroscience, 19(7), 1081-1088.

Sebastian, R., & Kiran, S. (2001). Task-modulated neural activation patterns in chronic stroke patients with aphasia. *Aphasiology, 25,* (8), 927-951.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. Journal of International Neuropsychological Society, 8(3), 448-460.

Tsapkini, K., Frangakis, C. E., & Hills, A. E. (2011). The function of the left anterior temporal pole: Evidence from acute stroke and infarct volume. *Brain, 134*, 3094-3105.

Walker, G. M., Schwartz, M. F., Kimberg, D. Y., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2011). Support for anterior temporal involvement in semantic error production in aphasia: New evidence from VLSM. *Brain & Language, 117*, 110-122.

Welcome Trust Center for Neuroimaging (1994-2012). SPM5 [computer software]. Available from http://www.fil.ion.ucl.ac.uk/spm/software/spm5/

Wicks, D., Tofts, P., Miller, D., Duboulay, G., Feinstein, A., Sacares, R., . . . Mcdonald, W. (1992). Volume measurement of multiple-sclerosis lesions with magnetic-resonance images – a preliminary-study. Neuroradiology, 34(6), 475-479.

Wilke, M., Kassubek, J., Ziyeh, S., Schulze-Bonhage, A., & Huppertz, H. J. (2003). Automated detection of gray matter malformations using optimized voxel-based morphometry: A systematic approach. NeuroImage, 20(1), 330-343. doi: 10.1016/S1053-8119(03)00296 9.