Important Considerations in Lesion-Symptom Mapping: Illustrations From Studies of Word Comprehension

Hinna Shahid,[1] Rajani Sebastian,[1] Tatiana T. Schnur,[5] Taylor Hanayik,[6] Amy Wright,[1] Donna C. Tippett,[1,2,3] Julius Fridriksson,[6] Chris Rorden,[6†] and Argye E. Hillis [1,2,4†]*

[1]*Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287*
[2]*Department of Physical Medicine & Rehabilitation, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287*
[3]*Department of Otolaryngology & Head & Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287*
[4]*Department of Cognitive Science, Johns Hopkins University, Baltimore, Maryland 21218*
[5]*Department of Neurosurgery, Baylor College of Medicine, Houston, Texas 77030*
[6]*University of South Carolina, Columbia, South Carolina 29208*

◆───────────────────────────◆

Abstract: Lesion-symptom mapping is an important method of identifying networks of brain regions critical for functions. However, results might be influenced substantially by the imaging modality and timing of assessment. We tested the hypothesis that brain regions found to be associated with acute language deficits depend on (1) timing of behavioral measurement, (2) imaging sequences utilized to define the "lesion" (structural abnormality only or structural plus perfusion abnormality), and (3) power of the study. We studied 191 individuals with acute left hemisphere stroke with MRI and language testing to identify areas critical for spoken word comprehension. We use the data from this study to examine the potential impact of these three variables on lesion-symptom mapping. We found that only the combination of structural and perfusion imaging within 48 h of onset identified areas where more abnormal voxels was associated with more severe acute deficits, after controlling for lesion volume and multiple comparisons. The critical area identified with this methodology was the left posterior superior temporal gyrus, consistent with other methods that have identified an important role of this area in spoken word comprehension. Results have implications for interpretation of other lesion-symptom mapping studies, as well as for understanding areas critical for auditory word comprehension in the healthy brain. We propose that lesion-symptom mapping at the acute stage of stroke addresses a different sort of question about brain–behavior relationships than lesion-symptom

mapping at the chronic stage, but that timing of behavioral measurement and imaging modalities should be considered in either case. *Hum Brain Mapp 38:2990–3000, 2017.* © **2017 Wiley Periodicals, Inc.**

Key words: ischemic stroke; brain mapping; lesion studies; aphasia; language comprehension

INTRODUCTION

Current approaches to testing hypotheses about brain-behavior mapping using data from neurologically impaired individuals focus on establishing statistical associations between damaged regions and level of performance on a task or function. Lesioned regions (defined by any parcellation map) may be analyzed as a dichotomous or as a continuous variable [Crinion et al., 2013]. Performance may also be analyzed as a dichotomous or as a continuous variable [Bates et al., 2003; Megalooikonomou et al., 1999]. Results obtained (i.e., which areas are most critical for a particular task or function) are influenced by a number of variables. Variables typically taken into account include the statistical tests selected, variables that are controlled (e.g., total volume of lesion), corrections that are made (e.g., corrections for multiple comparisons), quality of images and normalization, and so on [Rorden et al., 2009]. There are other variables that potentially influence results which are seldom considered in lesion-deficit mapping studies, and these variables will be the focus of this article. One variable is the timing of behavioral testing. In many studies, imaging and/or behavioral testing are obtained at least 6 months after brain damage, when test performance and the lesion are considered "stable." However, it is rarely confirmed that performance is stable. Many people, especially those receiving rehabilitation, show considerable recovery [Baker et al., 2010; Crinion and Leff, 2007; Meinzer et al., 2008] or decline [Levine et al., 2015] many months after stroke. Therefore, different results might be obtained if such patients were tested at 6 months versus longer poststroke. Yet, many studies involve patients with wide ranges of times postonset [Buchsbaum et al., 2011; Dronkers et al., 2004; Schwartz et al., 2009].

Occasionally, patients are studied at a homogeneous time after stroke onset, usually when they are first admitted to the hospital [Cloutman et al., 2009; Croquelois et al., 2003; Medina et al., 2009; Philipose et al., 2007]. Behavior typically changes rapidly early after stroke, with greatest improvement (or deterioration) occurring across days rather than months, so it is particularly important to study acute patients within a narrow time period. A few studies have shown that changes in perfusion in the first week or so after stroke may account for some of the rapid changes in language performance [Croquelois et al., 2003; Hillis et al., 2001a]. Functional neuroimaging studies show that areas that are activated during performance change from the acute period to later periods, providing evidence of reorganization of structure/function relationships underlying recovery of language [Jarso et al., 2013; Saur et al., 2006; Sebastian et al., 2016] and motor function [Askim et al., 2009; Johansen-Berg et al., 2002; Schaechter et al., 2002]. An advantage in studying acute patients is that there has been little time for reorganization of structure–function relationships, so there may be less variability in lesion-deficit associations [Ochfeld et al., 2010].

Another important consideration, particularly in the acute stage, is that the deficits observed may be due to hypoperfused, dysfunctional tissue surrounding the infarct, as well as the infarct [Hillis et al., 2005; Motta et al., 2015; Olsen et al., 1986]. For example, Figure 1 shows the scans of a patient with severe aphasia and right-sided weakness, despite minimal infarct (left), which can be accounted for by the extensive area of hypoperfusion seen on perfusion-weighted imaging (PWI, right). Therefore, whether one uses only structural imaging (e.g., diffusion weighted imaging, DWI, which is most sensitive to acute infarct) or also perfusion imaging (which reveals the area of hypoperfused tissue) to define the "lesion" may influence results.

Finally, results are substantially influenced by the power of the study—the effect size, as well as the number of patients (with more observations aiding our ability to detect effects) and distribution of infarcts of the patients who are studied (cf.[Kimberg et al., 2007; Rudrauf et al., 2008]). The last effect may seem unintuitive, but can be derived from first principles: we have no power to detect effects in areas that are never (or always) injured in our population. On the other hand, we can detect effects with a relatively small sample group of individuals if the critical area is damaged in half the population. In sum, if no or relatively few patients have damage to a given area, it is impossible to determine the probability that damage to that area would affect performance on the studied task. Unfortunately, prior attempts to model power with continuous behavioral measures and lesion-symptom mapping (e.g., Gläscher et al., 2009) have focused exclusively on the incidence of injury without regard to the behavioral effect size. In the terminology of Faul et al. [2007], this prior approach models the group size allocation ratio (e.g., incidence of a region being injured) but does not model spatial variations in the effect size. The premise of lesion-symptom mapping presumes that the effect size varies, with some areas being involved with a task while other brain regions are not required. Here we use our observed pattern of both lesion incidence as well as behavioral deficits to provide a more refined power analysis.
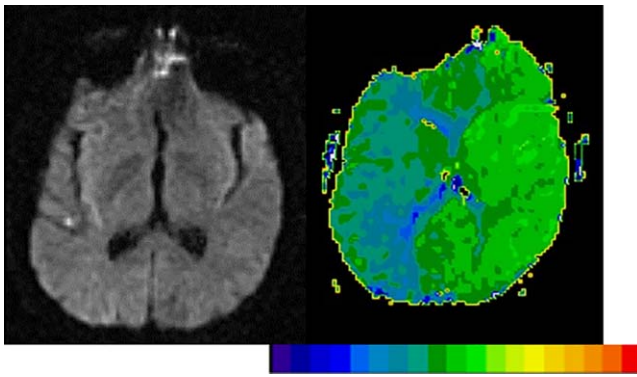
Figure 1.
DWI (left) and PWI (right) in a patient with severe deficits in all language tasks and right hemiplegia at Day 1. For the PWI, we computed time-to-peak (TTP), with bright green showing normal, fast blood flow while dark green and blue showing abnormally slow blood flow. Each color difference represents 2 s delay in TTP arrival of contrast. His deficits cannot be accounted for by the minimal infarct on DWI, but can be accounted for by the severe hypoperfusion of the whole left hemisphere. Scans in this and all figures are displayed in neurological convention (left hemisphere on left side). [Color figure can be viewed at wileyonlinelibrary.com]

In this article, we examine the potential impact of these variables for testing hypotheses about brain-behavior mapping, illustrated with lesion-deficit mapping studies of auditory word comprehension. We test three hypotheses concerning whether statistically significant associations identified between brain regions and acute deficits in spoken word comprehension depend on (1) the imaging sequences (DWI only or DWI plus PWI) used to define the "lesion," (2) the timing of measurement of word comprehension, and (3) the power of the study. Results have implications for interpreting other lesion-symptom mapping studies, as well as for understanding neural regions normally critical for auditory word comprehension.

MATERIALS AND METHODS

We report a retrospective analysis of prospectively collected data. The data were collected from 191 patients enrolled in a study of language processing in adult patients with acute ischemic left hemisphere stroke. Patients were enrolled at Johns Hopkins Medicine (Johns Hopkins Hospital or Johns Hopkins Bayview Medical Center). Mean age was 60.2 (SD 14.7) years. Mean education was 12.3 (SD 3.3) years; 49% were female. Additional inclusion criteria were right-handed, native English speaker, able to provide informed consent, or indicate another to provide informed consent. Exclusion criteria were previous neurological disease affecting the central nervous system, uncorrected hearing or visual loss, impaired level of consciousness, or ongoing sedation. All patients were enrolled within 48 h of symptom onset and had language testing within 24 h of MRI scan.

Testing of Auditory Word Comprehension

The test of auditory word comprehension was identical across the two sites, and consisted of a spoken word/picture verification test in which each of the 17 items was presented once with the target (e.g., goat/goat), once with a semantic foil (e.g. goat/cow) and once with a phonological foil (e.g., goat/coat) [Breese and Hillis, 2004]. The patient was asked, "Is this a ____" and had to respond yes or no verbally, or by pointing to the word "yes" or "no," or by nodding/shaking the head, whichever was the most reliable response of affirmation or negation. That is, because some patients could not speak or perseverated on "yes" or "no" in speaking, we determined their most reliable response to simple yes/no questions (e.g., "Is your name John?") prior to administration of the word/picture verification test. To obtain credit for understanding the item, the patient had to accept the correct response and reject both foils. A subset of patients ($n = 73$) were tested twice: once within 48 h of onset (Time 1) and once at 3–5 days post onset (Time 2). Patients were not tested at time 2 (or we did not include their data) if they (1) had a recurrent stroke or extension of the stroke or area of hypoperfusion; (2) were discharged from the hospital; or (3) declined to be tested.

Imaging

Sequences included the following: T2 and FLAIR to exclude old lesions, DWI and Apparent Diffusion Coefficient (ADC) to identify the acute structural lesion (area of dense ischemia/infarct), and PWI. Single-shot echo planar images with full-brain coverage were obtained in transaxial slices parallel to the AC–PC line. DWI trace images were acquired using a multislice, isotropic, single-shot EPI sequence, with $b_{max} = 1,000$ s/mm^2, 230×230 mm field of view, 192×192 matrix, TR/TE of 10,000/120 ms, and slice thickness of 5 mm. DWI b0 images were also acquired using the same parameters for normalization (see below). For PWI, single-shot gradient echo EPI perfusion images (TR/TE of 2,000/60 ms, 5 mm slices) were obtained with 20 cc GdDTPA (Gadolinium) bolus power injected at 5 cc/s at the start of the scan. Although several hemodynamic maps were constructed from the PWI, for the purposes of this study, we analyzed the time to peak (TTP) maps. Areas of significant hypoperfusion corresponding to dysfunction were defined as areas with >4 s delay in TTP, as this corresponds to areas of dysfunction defined by PET [Zaro-Weber et al., 2010] and as defined by impairment in function [Hillis et al., 2001b; Motta et al., 2015]. All patients also had either acute CT or SWI to exclude hemorrhage.
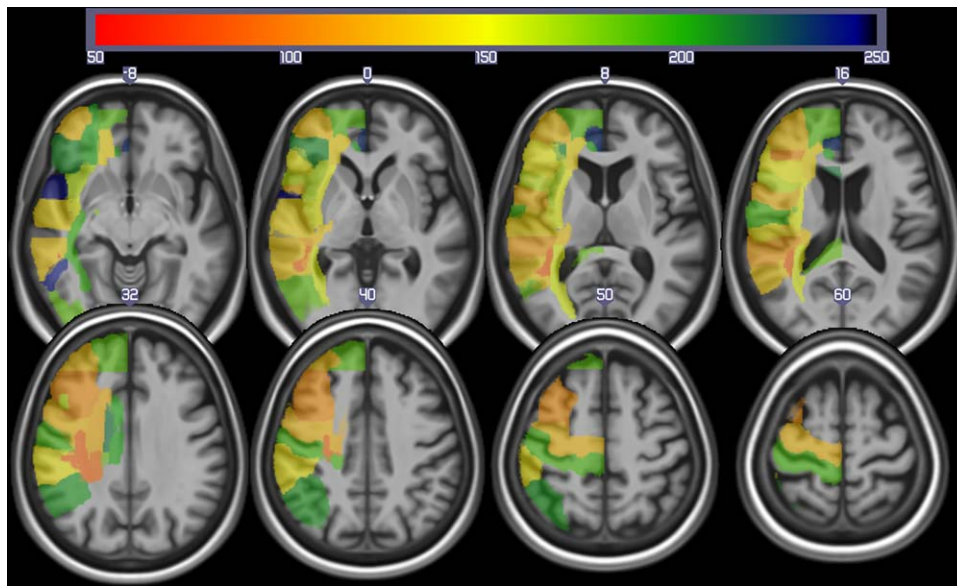
Figure 2.

Power map for identifying brain regions (from the JHU atlas) involved with auditory comprehension using lesions (as identified from the DWI scans). The colors reveal the sample size required to exceed $Z = 3.5$ in 60% of the replication studies. This map reveals maximal power in the center of the middle cerebral artery territory. [Color figure can be viewed at wileyonlinelibrary.com]

Image Analysis

Regions of infarct were drawn on DWI trace images and regions of hypoperfusion were identified on PWI images (TTP maps, co-registered to DWI) using MRIcron (www.mccauslandcenter.sc.edu), by a physician or trained technician who had no knowledge of the behavioral data. A single area of "dysfunctional tissue"—areas of abnormality on DWI and/or TTP—was drawn for each patient. It was necessary to include all areas that were either infarcted (on DWI) and/or hypoperfused (on PWI) in the area of dysfunctional tissue because there is often "luxury perfusion" of the infarct. That is, there is often normal perfusion where there is infarction (as well as hypoperfusion where there is no infarct) in acute stroke. All images were normalized to standard space using SPM12. First, the normalization transforms were computed for the DWI $B = 0$ image to a template based on age-matched controls [Rorden et al., 2012]. Then, the normalization parameters were applied to the DWI Trace-based lesion and TTP images. This method takes advantage of the fact that abnormalities visible on DWI Trace images and TTP maps are typically not apparent on DWI $B = 0$ images, allowing a normalization that is not disrupted by the abnormal signal caused by the stroke [Mah et al., 2014]. Proportions of damaged and hypoperfused tissue in different anatomical regions included in the JHU-MNI atlas (cmrm.med.jhmi.edu) were calculated for each patient. That is, we identified the percentage of voxels in each anatomical parcel that were affected by stroke. This proportion was entered into our

analyses. Associations between behavior and tissue dysfunction (i.e., the association between severity of deficit and percentage of voxels that were damaged in each parcel) were identified using GLM (pooled-variance t test, linear regression), and each analysis only included regions where at least 10 patients had damage/tissue dysfunction. All t scores were transformed to Z scores to aid interpretation (using SPM12's spm_t2z function). The benefit of this is that one does not need to know the degrees of freedom to interpret Z scores. Resulting statistical maps were thresholded to correct for multiple comparisons using 5,000 permutations (one-tailed $P < 0.05$, as we predicted injury would be associated with impairment) using our open source NiiStat (https://github.com/neurolabusc/NiiStat/blob/master/nii_stat_core.m) routines and the methods described by Rorden et al. (2007).

To compute our power analysis, we used the observed correlation (r) between proportion of injury in each region of interest to the initial comprehension test. Combining this with the given statistical threshold (here we set the threshold of Z to 3.5, inferred from the $P < 0.05$ controlling for multiple comparisons using permutation thresholding), we can estimate the power for every region of interest, which is the number of participants that would be required to replicate our findings in a given proportion of studies. Here, computed power for proportion of studies set to 0.6, so that in cases where there is a real effect of the observed magnitude one would detect it 60% of the time in each region. One wrinkle for lesion symptom mapping studies is that we need to control for multiple comparisons
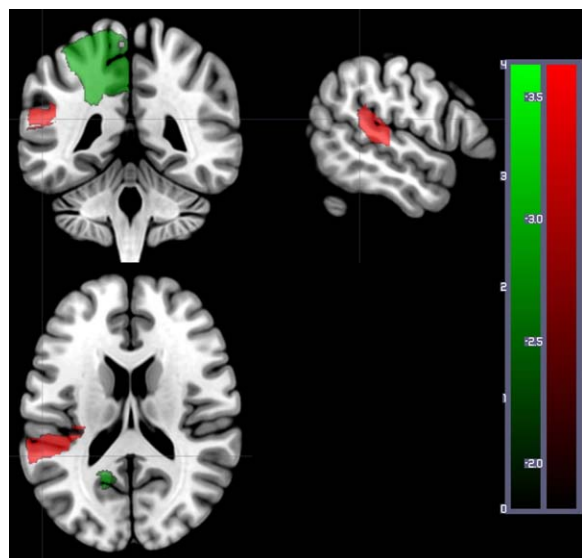
Figure 3.
Areas where dysfunctional tissue (abnormal on DWI and/or PWI) was significantly associated with percent correct on auditory word comprehension at Time 1 in 169 patients. Areas with significant negative Z scores (indicating more voxels with dysfunction predicts lower accuracy) are shown in red: in this analysis, only left posterior superior temporal gyrus survived threshold of 5,000 permutations, $P < 0.05$. Areas with positive Z scores (indicating more voxels with tissue dysfunctional in the region are associated with higher accuracy) are shown in green: left superior frontal gyrus, left superior parietal gyrus, and left precuneus. The scales show the Z scores (red for negative Z scores; green for positive Z scores). [Color figure can be viewed at wileyonlinelibrary.com]

across thousands of voxels or hundreds of regions of interest (which makes our alpha more extreme) and the fact that this threshold will actually be modulated for our sample size. This latter fact would be true even if Bonferroni thresholding were to be applied, as we will tend to analyze more voxels with larger studies (e.g., with small studies, we will have no observations of damage to many voxels, while with larger samples we will have observations of rarely injured regions). However, this effect will become exacerbated with permutation thresholding, as in small studies the number of unique lesion patterns is small. This latter effect largely explains why permutation thresholding provides much more power than Bonferroni thresholding, in particular for small studies (Rorden et al., 2009). However, it does pose a challenge to traditional power analyses, as alpha is assumed constant across group sizes. For simplicity, we chose a constant alpha of $Z = 3.5$, corresponding to permutation thresholds for our effects with about 191 participants. Consequently, the results of our power analyses should be considered a little conservative for voxels where power is reported below this number and a little liberal for voxels beyond this sample. We do

feel that this approach provides a traditional, principled, and justifiable solution to this problem. We appreciate there are other methods for estimating this, for example, using Monte Carlo simulations. However, on inspection, we felt that these approaches would have their own biases with respect to lesion coverage and sensitivity.

RESULTS

In 191 patients, the range of percent correct on the word-picture verification test was 6–100 (mean: 77.0; SD: 29.3). The effective power map for the whole group of 191 patients is shown in Figure 2. This figure shows the best statistical power to detect lesion-deficit correlations (i.e., the area where the greatest number of patients had damage) was in the left MCA territory, although there were patients with infarcts in other parts of the left hemisphere.

The Influence of Neuroimaging Sequence Used in Analysis

For this analysis, we included a subset of 169 patients who had both DWI and PWI scan at Day 1. Some patients did not have PWI because of (1) lack of IV access, (2) elevated glomerular filtration rate, which increases the risk of a rare dermatological adverse effect of Gadolinium contrast; (3) failure of the power injection or its timing; or (4)
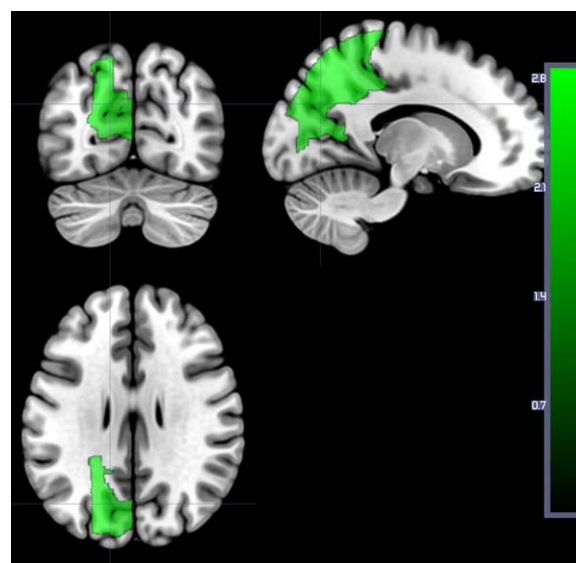

Figure 4.
Areas where more voxels with structural lesion (on DWI alone) was associated with higher percent correct on auditory word comprehension at Time 1 in the same 169 patients included in Figure 3. Regions include left superior parietal gyrus ($Z = 2.88$), left superior occipital gyrus ($Z = 3.32$), and left cuneus ($Z = 2.54$). [Color figure can be viewed at wileyonlinelibrary.com]
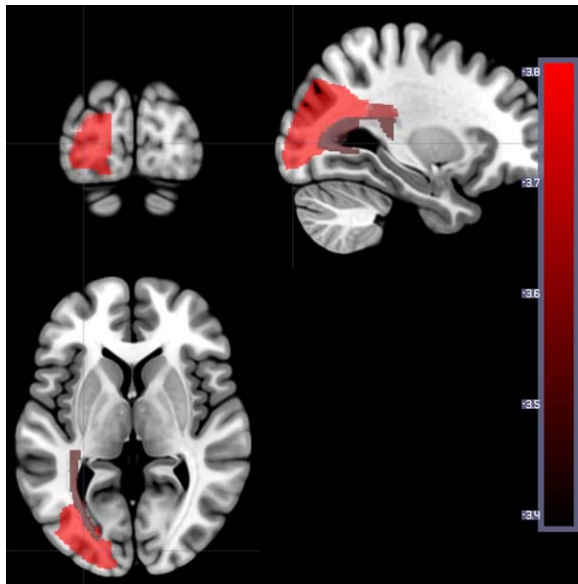
Figure 5.

Areas where the number of voxels with lesion (on DWI alone) is associated with lower percent correct on auditory word comprehension at Time 2. Areas included left middle frontal gyrus ($-Z = 3.37$), left middle occipital gyrus ($-Z = 3.82$), left posterior corona radiata ($-Z = 3.65$), left posterior thalamic radiation ($-Z = 3.50$), and left occipital lateral to ventricle ($-Z = 3.37$). [Color figure can be viewed at wileyonlinelibrary.com]

other reasons. The range of accuracy in 169 patients on the word-picture verification test was 6–100% (mean: 77.9; SD: 29.3).

When we analyzed the relationship between word comprehension performance and total dysfunctional tissue (abnormal on DWI and/or PWI), four regions survived permutation threshold after correction of lesion volume. Of the four regions, there was only one region where tissue dysfunction was associated with lower performance: left posterior superior temporal gyrus (pSTG; $-Z = 3.66$; Fig. 3). For the other three regions (left superior frontal gyrus, left superior parietal gyrus, and left precuneus), tissue abnormality was associated with *higher* accuracy (Z: 2.56 to 2.77; Fig. 3). Note while our one-tailed analysis is designed to identify regions where lesions are associated with deficits, we logically expect both negative and positive Z scores, because patients were selected on the basis of having lesions, not on the basis of having deficits. Specifically, since brain injury is determined by the vasculature involved, a brain territory not involved with the task often appears to show a paradoxical reverse correlation: where brain injury at this location predicts the absence of impairment. Obviously, injury to these regions does not imply better performance than healthy controls; it does suggest that injury to these regions leads to less impairment than damage to other territories. The results confirm that some lesions are associated with impaired auditory word comprehension (pSTG), while others are associated with relatively spared auditory word comprehension.
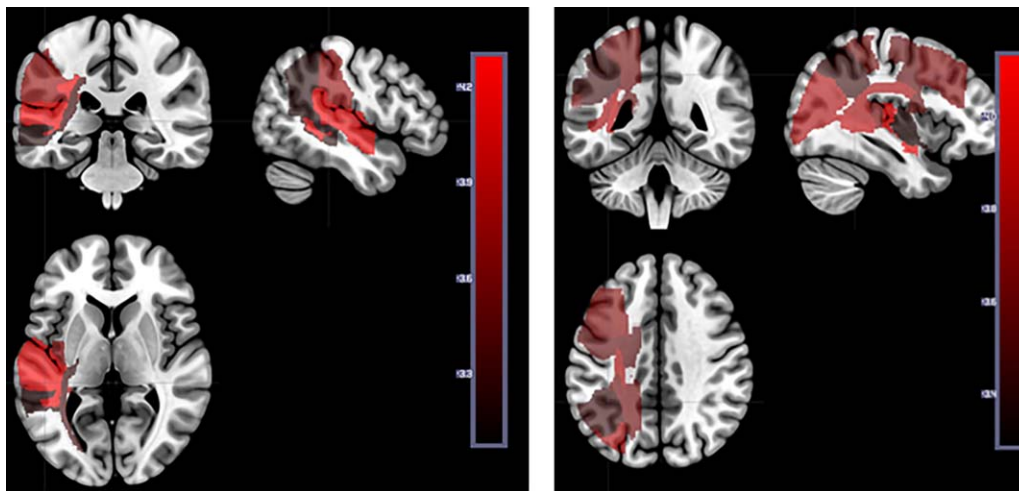

Figure 6.

Left panel: Areas where the number of voxels with tissue dysfunction (defined by DWI/PWI abnormality at *Time 1*) is significantly associated with lower accuracy on auditory word comprehension (defined by performance at Time 1). See Table II for Z scores for each region. Right panel: Areas where the number of voxels with tissue dysfunction (defined by DWI/PWI abnormality at *Time 2*) is significantly associated with lower accuracy on auditory word comprehension (defined by performance at Time 2). See Tables I and II for Z scores for each parcel. The scale shows the Z scores. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Areas where voxels with tissue dysfunction (defined by DWI/PWI abnormality at Time 1) are significantly associated with lower accuracy on auditory word comprehension (defined by performance at Time 1)

Auditory comprehension Time 1	
12 regions survived threshold (5,000 permutations; $P < 0.05$)	Z score
Superior longitudinal fasciculus left	−4.31
Posterior superior temporal gyrus left	−4.31
Superior temporal gyrus left	−4.24
Tapatum left	−4.16
Supramarginal gyrus left	−3.77
Posterior insula left	−3.69
Lateral ventricle_atrium left	−3.54
Posterior corona radiata left	−3.38
Posterior thalamic radiation (include optic radiation) left	−3.36
Retrolenticular part of internal capsule left	−3.31
Angular gyrus left	−3.22
Posterior middle temporal gyrus left	−3.21

When we analyzed only DWI (structural lesion) for the same 169 patients, three regions survived threshold, and these were all associated with *higher* accuracy in word comprehension: left superior parietal gyrus, left superior occipital gyrus, and left cuneus (Z: 2.54 to 3.32; Fig. 4). No areas were associated with lower accuracy in word comprehension using DWI alone. These results confirm that auditory word comprehension scores in the first 48 h after stroke are more strongly associated with the entire area of dysfunctional tissue (infarcted or significantly hypoperfused) than with the infarct itself.

The Influence of Time at Which Performance is Evaluated

These analyses included the subset of 73 patients who had language testing at Time 1 and Time 2 (3 to 5 days after onset of stroke). Initially, we analyzed the results as in previous analyses, but no areas survived threshold after correction for lesion volume. Therefore, to illustrate the effect of time poststroke on lesion-deficit correlations, we ran the analyses without correction for lesion volume (but still correcting for multiple comparisons). Using the Time 1 DWI scans (infarct), zero regions survived permutation threshold with language testing at time 1, and five regions survived threshold with language testing at time 2. Figure 5 shows the areas where more voxels with structural lesion (defined by DWI at Time 1) are associated with lower accuracy on auditory word comprehension (defined by performance at Time 2). Areas include left middle frontal gyrus, left middle occipital gyrus, left posterior corona radiata, left posterior thalamic radiation, and left occipital lateral to ventricle. As we did not correct for lesion

volume, this analysis may reveal areas where extent of structural lesion is associated with lower performance.

We also evaluated the lesion defined by total area of dysfunction brain tissue (abnormal on DWI and PWI). For this analysis, 63 patients had both DWI and PWI scans at Time 1 and language testing at Time 1 and Time 2. Figure 6 (left) and Table I show the areas where more voxels with tissue dysfunction (defined by DWI/PWI abnormality at Time 1) is associated with lower accuracy rate on auditory word comprehension (defined by performance at Time 1), again not correcting for lesion volume due to relatively small number of patients. Twelve regions survived threshold at Time 1. Figure 6 (right) and Table II show the areas where more voxels with tissue dysfunction (defined by DWI/PWI abnormality at Time 1) is associated with lower accuracy on auditory word comprehension defined by performance at Time 2, just 3 to 5 days later. Twenty regions survived threshold using Time 2 language testing. Again, this analysis may reveal areas where extent or volume of tissue dysfunction is associated with lower performance.

The Influence of Number of Patients and Distributions of Their Lesions

Here we compare results for the VSLM using structural imaging (infarct on DWI) in all 191 patients at Day 1 to

TABLE II. Areas where voxels with tissue dysfunction (defined by DWI/PWI abnormality at Time 1) are associated with lower accuracy on auditory word comprehension (defined by performance at Time 2)

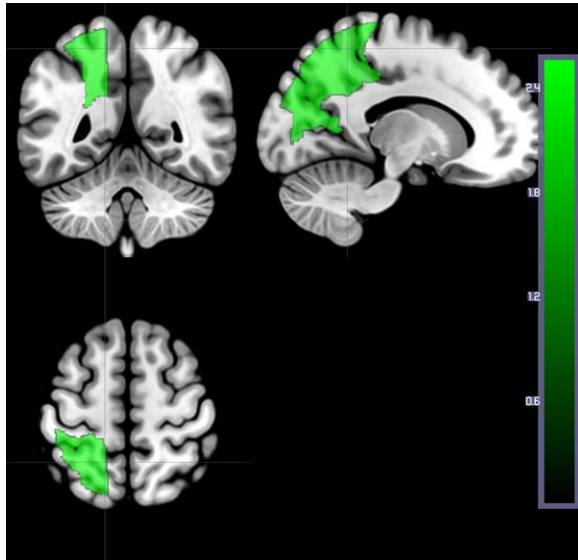| Auditory comprehension Time 2	
20 regions survive threshold (5,000 permutations; $P < 0.05$)	Z score
Superior temporal gyrus left	−4.15
Posterior thalamic radiation (include optic radiation) left	−3.92
Middle occipital gyrus left	−3.84
Superior longitudinal fasciculus left	−3.83
Anterior corona radiata left	−3.75
Lateral ventricle_atrium left	−3.71
Middle frontal gyrus (posterior segment) left	−3.70
Posterior superior temporal gyrus left	−3.67
Superior parietal gyrus left	−3.64
Tapatum left	−3.64
Splenium of corpus callosum left	−3.60
Precentral gyrus left	−3.59
Superior corona radiata left	−3.55
Angular gyrus left	−3.45
Posterior insula left	−3.38
Anterior limb of internal capsule left	−3.36
Lateral ventricle_occipital left	−3.30
Postcentral gyrus left	−3.28
Posterior middle temporal gyrus left	−3.22
Caudate nucleus left	−3.18

Figure 7.

Areas where the number of voxels with structural lesion (on DWI) was associated with higher accuracy on auditory word comprehension at Time 1 after correcting for lesion volume and multiple comparison. Regions include left superior parietal gyrus, left superior occipital gyrus, and left cuneus. [Color figure can be viewed at wileyonlinelibrary.com]

the same analysis carried out in a smaller number of patients (the 73 who had testing at Time 2). Figure 7 shows the area where the number of voxels with structural lesion (on DWI) was significantly associated with *higher* accuracy on auditory word comprehension at Time 1 after correcting for lesion volume and multiple comparisons. Regions include left superior parietal gyrus, left superior occipital gyrus, and left cuneus. There were no areas where damage was significantly associated with lower accuracy. In contrast, as noted above, when we ran the same analysis on the 73 patients with testing at Times 1 and 2, no areas survived threshold after controlling for lesion volume and multiple comparisons. Because there could be differences between patients who had Time 1 and Time 2 testing and those who had only Time 1 testing, we also ran the same analysis with a random subset of 100 patients from the total set of 191 patients who had only time 1 testing. Again, we found that no areas survived threshold after permutation correction for multiple comparisons and after controlling for lesion volume. In contrast, identical areas were identified when we ran the same analysis (using DWI scans only) on the subset of 169 patients who had both DWI and PWI at Time 1 (Fig. 4). That is, there was no difference in results when 191 patients were included versus when 169 patients were included, but there was a large difference in results when only 73 patients or a subset of 100 from the 191 were included.

DISCUSSION

Here we confirmed that lesion-deficit mapping is influenced by how "lesion" is defined (on which imaging sequence/s), time of testing, and number of patients. Most lesion-symptom mapping studies map level of performance of patients (measured at a variety of times postonset of stroke) to the structural lesion. We showed that, at least in the acute stage, different results are obtained if one uses structural imaging alone (DWI, which is sensitive to dense ischemia/infarct in acute stroke), versus if one also uses sequences that show the entire area of dysfunctional tissue (hypoperfused and/or infarcted). In our study, pSTG, which is typically considered important in word comprehension [Hart and Gordon, 1990; Naeser et al., 1987; Selnes et al., 1983] was associated with auditory word comprehension performance only when we evaluated the entire area of dysfunctional tissue (using both DWI and PWI). In contrast, when we evaluated only DWI in the same patients, no areas were identified where ischemia was significantly associated with impaired performance, although three areas were identified where ischemia was associated with better performance on auditory word comprehension—all areas in different vascular territories than pSTG. This result can be understood as showing that when the lesion is somewhere other than in the distribution of left inferior division left MCA (which supplies pSTG), auditory word comprehension tends to be preserved. We used dynamic-contrast (bolus-tracking) PWI, but similar results have been shown in small studies of acute aphasia using CT perfusion [Croquelois et al., 2003] and studies of chronic aphasia using arterial spin labeling perfusion MRI [Love et al., 2002; Richardson et al., 2011].

We also illustrated that results obtained may be influenced by the time post-onset at which behavior is measured, even if the lesion has not changed and performance is measured just a few days later. For the subset of patients included in the study, the lesion did not change from time 1 to time 2 (as confirmed by Time 2 imaging), but performance changed, yielding different results. Although these results were obtained in the acute stage of stroke, similar (although perhaps less dramatic) changes in level of performance can be seen after 6 months post-stroke, particularly when patients are involved in rehabilitation. Different results will be obtained before and after recovery or rehabilitation.

Finally, we show that the number of patients included in the study can substantially influence the number of areas where damage is significantly associated with the deficit. While this conclusion seems trivially obvious, many studies do not adequately consider the power required to draw valid conclusions. It is not only the total number of participants that is important but also the distribution of their lesions. One can only evaluate the role of areas where a sufficient number of people have, and do not have, a lesion. Studies that include a large number of

patients with and without damage to the regions of interest will provide the most reliable results.

Our penultimate comparison (showing the influence of time) also shows, indirectly, the importance of controlling for lesion volume. In this analysis of 73 patients, no areas survived after controlling for lesion volume, but 12 to 20 regions (depending on the timing of behavioral testing) survived when we corrected for multiple comparisons, but not for lesion volume. Some (or all) of the areas in Tables I and II may simply be associated with larger lesions—and larger lesions are associated with more impaired auditory word comprehension. For example, insular lesions are particularly associated with larger lesions because the insula is infarcted in the majority of stroke due to occlusion of the left MCA or internal carotid artery [Kodumuri et al., 2016].

Many other variables can influence the results of lesion-deficit association studies, such as the atlas or parcellation map used to evaluate the association, the size of the regions or voxel clusters examined, and the methods used for analysis. These variables include registration method, statistical tests, method of correction for multiple comparisons, and the variables that are controlled (including lesion volume). A range of performance on the experimental task is also essential if performance is treated as a continuous measure. Otherwise, tests that evaluate dichotomous associations are more appropriate (see Crinion et al. [2013]; Rorden et al. [2007]; Rorden et al. [2009] for discussion of these issues).

We have illustrated that it is important to consider not only the imaging sequences used to identify the "lesion," but also the timing of behavioral testing with respect to stroke onset when carrying out VSLM or other lesion-symptom mapping. We used only one task (word-picture verification), but the principles should apply to any behavior that might change over time after stroke. While most patients improve over time, 20 to 25% actually show continued decline in function after stroke [Dhamoon, 2016; Dhamoon et al., 2009, 2010; Levine et al., 2015]. Demographic variables that influence recovery [Hope et al., 2013] or decline [Dhamoon et al., 2012] after stroke, and their interaction with time, might also influence results of lesion-symptom mapping.

While lesion studies are invaluable to complement functional imaging studies, results need to be interpreted in the context of the sequences used to identify the "lesion," the time at which behavior is measured after the lesion, the power to detect the role of regions of interest in the behavior, and the demographic and imaging variables that are controlled. A recent study indicates that for mapping language, resting-state connectivity between homologous regions can also complement information from the lesion itself in predicting task performance, although the same was not found for mapping motor function [Siegel et al., 2016].

Limitations of our study include the fact that we did not evaluate the contribution of structural or resting-state connectivity (as relatively few of our patients in this study had resting-state data). Future studies will test the hypothesis that damage to connections to the pSTG produce similar deficits as dysfunction of pSTG itself. We will also evaluate areas where damage is associated with deficits in auditory word comprehension at 6 and 12 months post-stroke in the same patients.

We used performance on auditory word comprehension merely as the task to illustrate the influence of variables such as time and imaging sequence. Nevertheless, results also reveal areas critical for word comprehension. Only using the total area of dysfunctional tissue allowed us to identify voxels associated with more impaired performance on word comprehension, after correcting for lesion volume. All the analyses that revealed areas where the number of affected voxels was associated with lower accuracy at Time 1 or Time 2 showed the left STG or pSTG to be among the most strongly (or only) associated region. Left pSTG has been identified as critical for word comprehension in both acute and chronic stroke (Hart and Gordon, 1990; Hillis et al., 2001a,b; Naeser et al., 1987; Selnes et al., 1983). It seems plausible that DWI plus PWI abnormality in the acute stage reveals the area responsible for the acute deficit (indicating the area is normally associated with the auditory word comprehension), while structural imaging in the chronic stage reveals the areas that, when damaged, are associated with failure to recover the function. However, pSTG seems to be both: the area where acute dysfunction causes impaired word comprehension and the area where structural damage impedes recovery of word comprehension. Therefore, while we have shown the influence of imaging and time variables, at least some areas may be robustly associated with a deficit, independently of these variables.

ACKNOWLEDGMENTS

REFERENCES

Askim T, Indredavik B, Vangberg T, Haberg A (2009): Motor network changes associated with successful motor skill relearning after acute ischemic stroke: A longitudinal functional magnetic resonance imaging study. Neurorehabil Neural Repair 23: 295–304.

Baker JM, Rorden C, Fridriksson J (2010): Using transcranial direct-current stimulation to treat stroke patients with aphasia. Stroke 41:1229–1236.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003): Voxel-based lesion–symptom mapping. Nat Neurosci 6:448–450.

Breese EL, Hillis AE (2004): Auditory comprehension: Is multiple choice really good enough? Brain Lang 89:3–8.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G (2011): Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fMRI data. Brain Lang 119: 119–128.

Cloutman L, Gingis L, Newhart M, Davis C, Heidler-Gary J, Crinion J, Hillis AE (2009): A neural network critical for spelling. Ann Neurol 66:249–253.

Crinion J, Holland AL, Copland DA, Thompson CK, Hillis AE (2013): Neuroimaging in aphasia treatment research: Quantifying brain lesions after stroke. NeuroImage 73:208–214.

Crinion JT, Leff AP (2007): Recovery and treatment of aphasia after stroke: Functional imaging studies. Curr Opin Neurol 20: 667–673.

Croquelois A, Wintermark M, Reichhart M, Meuli R, Bogousslavsky J (2003): Aphasia in hyperacute stroke: Language follows brain penumbra dynamics. Ann Neurol 54:321–329.

Dhamoon MS (2016): The trajectory of functional status before and after vascular events, Columbia University Academic Commons.

Dhamoon MS, Moon YP, Paik MC, Boden-Albala B, Rundek T, Sacco RL, Elkind MS (2009): Long-term functional recovery after first ischemic stroke: The Northern Manhattan Study. Stroke 40:2805–2811.

Dhamoon MS, Moon YP, Paik MC, Boden-Albala B, Rundek T, Sacco RL, Elkind MS (2010): Quality of life declines after first ischemic stroke. The Northern Manhattan Study. Neurology 75:328–334.

Dhamoon MS, Moon YP, Paik MC, Sacco RL, Elkind MS (2012): Trajectory of functional decline before and after ischemic stroke: The Northern Manhattan Study. Stroke 43:2180–2184.

Dronkers NF, Wilkins DP, Van Valin RD, Redfern BB, Jaeger JJ (2004): Lesion analysis of the brain areas involved in language comprehension. Cognition 92:145–177.

Faul F, Erdfelder E, Lang A-G, Buchner A (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Met 39: 175–191.

Gläscher J, Tranel D, Paul LK, Rudrauf D, Rorden C, Hornaday A, Grabowski T, Damasio H, Adolphs R (2009): Lesion mapping of cognitive abilities linked to intelligence. Neuron 61:681–691.

Hart J, Gordon B (1990): Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. Ann Neurol 27:226–231.

Hillis AE, Kane A, Tuffiash E, Ulatowski JA, Barker PB, Beauchamp NJ, Wityk RJ (2001a): Reperfusion of specific brain regions by raising blood pressure restores selective language functions in subacute stroke. Brain Lang 79:495–510.

Hillis AE, Wityk RJ, Tuffiash E, Beauchamp NJ, Jacobs MA, Barker PB, Selnes OA (2001b): Hypoperfusion of Wernicke's area predicts severity of semantic deficit in acute stroke. Ann Neurol 50:561–566.

Hillis AE, Newhart M, Heidler J, Barker PB, Herskovits EH, Degaonkar M (2005): Anatomy of spatial attention: Insights from perfusion imaging and hemispatial neglect in acute stroke. J Neurosci 25:3161–3167.

Hope TM, Seghier ML, Leff AP, Price CJ (2013): Predicting outcome and recovery after stroke with lesions extracted from MRI images. NeuroImage Clin 2:424–433.

Jarso S, Li M, Faria A, Davis C, Leigh R, Sebastian R, Tsapkini K, Mori S, Hillis AE (2013): Distinct mechanisms and timing of language recovery after stroke. Cogn Neuropsychol 30:454–475.

Johansen-Berg H, Dawes H, Guy C, Smith SM, Wade DT, Matthews PM (2002): Correlation between motor improvements and altered fMRI activity after rehabilitative therapy. Brain 125:2731–2742.

Kimberg DY, Coslett HB, Schwartz MF (2007): Power in voxel-based lesion-symptom mapping. J Cogn Neurosci 19: 1067–1080.

Kodumuri N, Sebastian R, Davis C, Posner J, Kim EH, Tippett DC, Wright A, Hillis AE (2016): The association of insular stroke with lesion volume. NeuroImage Clin 11:41–45.

Levine DA, Galecki AT, Langa KM, Unverzagt FW, Kabeto MU, Giordani B, Wadley VG (2015): Trajectory of cognitive decline after incident stroke. JAMA 314:41–51.

Love T, Swinney D, Wong E, Buxton R (2002): Perfusion imaging and stroke: A more sensitive measure of the brain bases of cognitive deficits. Aphasiology 16:873–883.

Mah YH, Husain M, Rees G, Nachev P (2014): Human brain lesion-deficit inference remapped. Brain 137:2522–2531.

Medina J, Kannan V, Pawlak MA, Kleinman JT, Newhart M, Davis C, Heidler-Gary JE, Herskovits EH, Hillis AE (2009): Neural substrates of visuospatial processing in distinct reference frames: Evidence from unilateral spatial neglect. J Cogn Neurosci 21:2073–2084.

Megalooikonomou V, Davatzikos C, Herskovits EH (1999): Mining lesion-deficit associations in a brain image database. 347–351.

Meinzer M, Flaisch T, Breitenstein C, Wienbruch C, Elbert T, Rockstroh B (2008): Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. NeuroImage 39:2038–2046.

Motta M, Ramadan A, Hillis AE, Gottesman RF, Leigh R (2015): Diffusion–perfusion mismatch: An opportunity for improvement in cortical function. Front Neurol 5:280.

Naeser MA, Helm-Estabrooks N, Haas G, Auerbach S, Srinivasan M (1987): Relationship between lesion extent in 'Wernicke's area' on computed tomographic scan and predicting recovery of comprehension in Wernicke's aphasia. Arch Neurol 44:73–82.

Ochfeld E, Newhart M, Molitoris J, Leigh R, Cloutman L, Davis C, Crinion J, Hillis AE (2010): Ischemia in broca area is associated with broca aphasia more reliably in acute than in chronic stroke. Stroke 41:325–330.

Olsen TS, Bruhn P Oberg RG (1986): Cortical hypoperfusion as a possible cause of 'subcortical aphasia'. Brain 109: 393–410.

Philipose LE, Gottesman RF, Newhart M, Kleinman JT, Herskovits EH, Pawlak MA, Marsh EB, Davis C, Heidler -Gary J, Hillis AE (2007): Neural regions essential for reading and spelling of words and pseudowords. Ann Neurol 62:481–492.

Richardson JD, Baker JM, Morgan PS, Rorden C, Bonilha L, Fridriksson J (2011): Cerebral perfusion in chronic stroke: Implications for lesion-symptom mapping and functional MRI. Behav Neurol 24:117–122.

Rorden C, Fridriksson J, Karnath H (2009): An evaluation of traditional and novel tools for lesion behavior mapping. NeuroImage 44:1355–1362.

Rorden C, Karnath H, Bonilha L (2007): Improving lesion-symptom mapping. J Cogn Neurosci 19:1081–1088.

Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath H (2012): Age-specific CT and MRI templates for spatial normalization. NeuroImage 61:957–965.

Rudrauf D, Mehta S, Bruss J, Tranel D, Damasio H, Grabowski TJ (2008): Thresholding lesion overlap difference maps: Application to category-related naming and recognition deficits. NeuroImage 41:970–984.

Saur D, Lange R, Baumgaertner A, Schraknepper V, Willmes K, Rijntjes M, Weiller C (2006): Dynamics of language reorganization after stroke. Brain 129:1371–1384.

Schaechter JD, Kraft E, Hilliard TS, Dijkhuizen RM, Benner T, Finklestein SP, Rosen BR, Cramer SC (2002): Motor recovery and cortical reorganization after constraint-induced movement therapy in stroke patients: A preliminary study. Neurorehabil Neural Repair 16:326–338.

Schwartz MF, Kimberg DY, Walker GM, Faseyitan O, Brecher A, Dell GS, Coslett HB (2009): Anterior temporal involvement in semantic word retrieval: Voxel-based lesion-symptom mapping evidence from aphasia. Brain 132: 3411–3427.

Sebastian R, Long C, Purcell JJ, Faria AV, Lindquist M, Jarso S, Race D, Davis C, Posner J, Wright A, Hillis AE (2016): Imaging network level language recovery after left PCA stroke. Rest Neurol Neurosci 34:473–489.

Selnes OA, Knopman DS, Niccum N, Rubens AB, Larson D (1983): Computed tomographic scan correlates of auditory comprehension deficits in aphasia: A prospective recovery study. Ann Neurol 13:558–566.

Siegel JS, Ramsey LE, Snyder AZ, Metcalf NV, Chacko RV, Weinberger K, Baldassarre A, Hacker CD, Shulman GL, Corbetta M (2016): Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. Proc Natl Acad Sci USA 113:E4367–E4376.

Zaro-Weber O, Moeller-Hartmann W, Heiss WD, Sobesky J (2010): Maps of time to maximum and time to peak for mismatch definition in clinical stroke studies validated with positron emission tomography. Stroke 41:2817–2821.