Published in final edited form as:

Neuroimage. 2010 January 1; 49(1): 597–602. doi:10.1016/j.neuroimage.2009.07.061.

Voxel-based Bayesian lesion-symptom mapping

Rong Chen^{*} and Edward H Herskovits[†]

^{*} Department of Radiology, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104. Phone: 215-662-7797. E-mail: rong.chen@uphs.upenn.edu

[†] Department of Radiology, University of Pennsylvania, 3400 Spruce Street, Philadelphia, PA 19104. Phone: 215-615-3705. E-mail: ehh@ieee.org

Abstract

Most existing voxel-based lesion-symptom mapping methods are based on the same statistical foundation: null-hypothesis significance testing (NHST). The two major limitations of these methods are the inability to infer that there is no difference in lesion proportions, and a requirement for multiple-comparison correction. We propose a Bayesian approach that directly models the posterior distribution of lesion proportion difference, and makes decisions based on inference on this posterior distribution. Compared to NHST-based approaches, our Bayesian approach yields inference results with clearer semantics, and does not require multiple-comparison correction. We evaluated our Bayesian method using simulated data, and data from a study of acute ischemic left-hemispheric stroke. Results of both experiments indicate that the Bayesian approach is sensitive in detecting regions that characterize group differences.

1 Introduction

Studying brain-injured subjects (lesion-based analysis) is an important approach for investigating structure-function relationships using neuroimaging techniques. In particular, voxel-based lesion-symptom mapping has been widely used to detect structure-function associations in neuroimaging studies. This approach involves mapping the relationship between tissue damage and a clinical variable on a voxel-by-voxel basis. Let C denote the clinical variable of interest, which can represent the outcome of a neuropsychological test, or a demographic variable such as sex; subjects are grouped by C . Lesion-symptom mapping focuses on detecting differences in the spatial distributions of lesions across experimental groups.

Before analyzing a lesion-based study, we usually delineate lesions for each subject, to produce a binary lesion map. We accomplish this task either by manually tracing lesions, or by using an automated approach. We must then register subjects' lesion maps to a common stereotaxic space. The clinical variable C can be binary (e.g., presence or absence of a deficit), or continuous (e.g., a score representing task performance). Since C is most often used as a group-membership variable, we herein consider C to be binary.

Methods for voxel-based lesion-symptom mapping include MAP3 (Damasio et al. (1996)), VLSM (Bates et al. (2003)), χ^2 computation (Karnath et al. (2004)), logistic regression (Karnath et al. (2004)), correlation (Tyler et al. (2005)), resampling approach (Kimberg et al. (2007)),

and Liebermeister and Brunner-Munzel tests (Rorden et al. (2007)). Despite the differences in model assumption and formulation, these methods fall under the same statistical framework: null hypothesis significance testing (NHST). NHST is employed to choose between two competing statistical hypotheses: H_0 (the null hypothesis) and H_a (the alternative hypothesis). In NHST, a p -value is computed based on the probability distribution of the chosen statistic, under the null hypothesis. This p -value represents the probability of obtaining the statistic if H_0 were true. A small p -value leads us to reject H_0 . For example, let η_A and η_B denote the proportions of tissue damage for subjects with and without a particular clinical deficit, respectively. We can compute voxel-wise Fisher-exact statistics for $H_0 : \eta_A = \eta_B$ versus $H_a : \eta_A \neq \eta_B$. A small p -value would indicate a significant difference in lesion proportions across groups at that voxel.

In contrast, a Bayesian approach to the same analysis directly models the posterior distribution of lesion-proportion difference, $\eta_A - \eta_B$, and performs inference based on this posterior distribution. A Bayesian approach has the potential to address the two major limitations of NHST-based approaches. First, in the NHST framework, we cannot reject H_a ; that is, a statistic below our significance threshold does not allow us to infer that there is no difference between groups. This explanation of p -value based inference is not intuitive, and may cause misinterpretation of results (Krueger (2001)). In contrast to a NHST-based approach, a Bayesian approach directly models this difference, and therefore has a clear interpretation.

Second, in voxel-based lesion-symptom mapping, we perform hypothesis testing for each voxel, leading to the multiple-comparison problem: in order to preserve the overall type I error rate, the p -value of each individual test must be adjusted, to prevent false-positive results. In most cases, there is no normative way to determine how to adjust p -value thresholds, which greatly complicates the interpretation of NHST-based approaches (Kimberg et al. (2007); Rorden et al. (2007)). In contrast to a NHST-based approach, a Bayesian approach does not incur the multiple-comparison problem. For each voxel, we perform inference based on the posterior distribution of lesion-proportion differences. This obviates multiple-comparison procedures.

The proposed voxel-based Bayesian lesion-symptom mapping (vBLSM) algorithm is founded on an analytical formulation of the posterior distribution of the lesion proportion difference $\eta_A - \eta_B$; we use this distribution to calculate the probability that a proportion difference is greater than a threshold. Voxels with high probability of manifesting proportion differences are voxels that are associated with the clinical variable C . In comparison to NHST-based approaches, our Bayesian approach has more clearly interpretable inference results, and avoids the multiple-comparison problem.

The remainder of this paper is organized as follows. The Methods section provides details about computing the posterior distribution of lesion-proportion differences. In the Experimental Results section we evaluate the accuracy of our approach on simulated and real data. In section Conclusion and Discussions, we summarize and discuss our approach and results.

2 Methods

2.1 Background

The Bayesian analysis that we have presented employs Bayes' theorem to reformulate the hypothesis regarding differences among experimental groups in lesion-proportions; our approach is in keeping with that espoused by Gelman (Gelman et al. (1995)). Let C and V denote the clinical variable and lesion load at a voxel, respectively. Bayes theorem is as follows,

$$Pr(C|V) = \frac{Pr(C)Pr(V|C)}{Pr(V)} = \frac{Pr(C)Pr(V|C)}{\sum_C Pr(C)Pr(V|C)},$$

where $Pr(C|V)$ is referred to as the posterior distribution of C , $Pr(C)$ is the prior distribution of C , and $Pr(V|C)$ is the likelihood term. In Bayesian framework, the probability is a measure of uncertainty, while the frequency argument about probability is that it is the relative frequency obtained in a long sequences of experiments. The prior distribution $Pr(C)$ is our initial belief about C ; and Bayes theorem provides an approach to update our belief about C based on observations of V . This updated belief is the posterior distribution of C . Inference about C is based on the posterior distribution.

For instance, let C denote a normal/impaired cognitive function. Our prior of C is $Pr(C = \text{impaired}) = 0.5$ and $Pr(C = \text{normal}) = 0.5$. The likelihood function $Pr(V|C)$ is $Pr(V = \text{lesioned}|C = \text{impaired}) = 0.8$ and $Pr(V = \text{lesioned}|C = \text{normal}) = 0.3$. If we observe that the voxel is lesioned, then based on Bayes theorem, the posterior probability of $C = \text{impaired}$ is

$$Pr(C = \text{impaired}|V = \text{lesioned}) = \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.3 \times 0.5} = 0.72.$$

That is, our belief that the subject will have impaired cognitive function increases from 0.50 (prior) to 0.72 (posterior) if we observe that V is lesioned.

Note that when the sample size is large in a lesion-deficit study, for the task of detecting linear association between V and C , frequentist based methods and Bayesian methods will converge to the same conclusion, if we use a noninformative prior in Bayesian modeling (Gelman et al. (1995)).

2.2 Voxel-based Bayesian lesion-symptom mapping

Our algorithm is voxel-based; that is, we perform Bayesian analysis voxel-by-voxel. To simplify the notation, we do not include the voxel index. In this paper, we focus on discrete lesion maps, in which each voxel V can assume either state 0 (normal) or state 1 (lesioned), and a categorical clinical variable C , which divides subjects into two groups. Similar conditions commonly apply in lesion-deficit studies.

The lesion proportion η is the probability, across subjects, that V is lesioned. For V , let η_A and η_B denote the lesion proportions for groups A and B, respectively. In lesion-symptom mapping, we are interested in whether η_A and η_B are significantly different. In our Bayesian approach, the quantity of interest is $\eta_A - \eta_B$; we compute $d(\delta) = Pr((\eta_A - \eta_B) \geq \delta)$ to measure the difference between η_A and η_B , where δ is a threshold. For most applications, we set $\delta = 0$.

Since η_A and η_B are independent, the posterior probability distribution for the difference $\eta_A - \eta_B$ is

$$f_{\eta_A - \eta_B}(\tau) = \int_{-\infty}^{\infty} f_{\eta_A}^*(\theta) f_{\eta_B}^*(\theta - \tau) d\theta, \quad (1)$$

where $f_{\eta_A}^*$ and $f_{\eta_B}^*$ are the posterior distributions of proportions for groups A and B, respectively.

A widely used prior distribution for η_A and η_B is the Beta distribution. The Beta distribution with parameters α and β , denoted by $B(\alpha, \beta)$, has the form

$$f(\eta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \eta^{\alpha-1} (1-\eta)^{\beta-1}, \quad (2)$$

where Γ is the Gamma function. Under binomial sampling, the likelihood of η is

$$L(\eta|n_0, n_1) = \binom{n_0+n_1}{n_0} \eta^{n_0} (1-\eta)^{n_1}, \quad (3)$$

where n_0 and n_1 are the numbers of subjects for whom voxel V assumes state 0 and 1, respectively. Given the Beta prior and binomial sampling, the posterior distribution of η , $f^*(\eta)$, follows the Beta distribution $B(n_0 + \alpha, n_1 + \beta)$.

Therefore, $d(\delta)$ can be formulated as

$$\begin{aligned} Pr((\eta_A - \eta_B) \geq \delta) &= \int_{\delta}^{\infty} \int_{-\infty}^{\infty} f_{\eta_A}^*(x) f_{\eta_B}^*(x - \tau) dx d\tau \\ &= \int_{-\infty}^{\infty} f_{\eta_A}^*(x) \int_{-\infty}^{-\delta+x} f_{\eta_B}^*(\tau) d\tau dx \\ &= \int_{-\infty}^{\infty} f_{\eta_A}^*(x) F_{\eta_B}^*(x - \delta) dx. \end{aligned} \quad (4)$$

Since the posterior distributions of η_A and η_B are Beta distributions, $Pr((\eta_A - \eta_B) \geq \delta)$ is the integral over x of $betapdf(x, a_A, b_A) \times betacdf(x - \delta, a_B, b_B)$, where

$$a_A = n_0^A + \alpha^A, b_A = n_1^A + \beta^A, a_B = n_0^B + \alpha^B, b_B = n_1^B + \beta^B.$$

There are different approaches to constructing the prior distribution for η . First, if we don't have prior knowledge of η , we can set the prior to a uniform probability distribution on the unit interval; that is, we can set the prior distribution for η to $B(1, 1)$. This type of prior is referred to as a noninformative prior.

Second, in some applications, there are existing data we can utilize. For example, consider the case in which, prior to executing a large-scale study, a pilot study is performed. We could use the empirical proportions in the pilot data as the prior distribution. This approach is called the training-data approach.

A third source of prior information about η is clinicians, who could estimate η based on their experience.

For any voxel, if the probability of $\eta_A - \eta_B \geq \delta$ is greater than a threshold th , we assert that there is a difference in lesion proportions at that voxel location. In most applications, we can set $\delta = 0$. th is the significance level, in $[0, 1]$. Larger values for th will cause vBLSM to be more conservative.

We have developed a C++ program that calculates $Pr((\eta_A - \eta_B) \geq \delta)$. We use the libRmath library, provided by the R project (<http://www.r-project.org/>), to calculate *betapdf* and *betacdf*. We then use Equation (4) to calculate $Pr((\eta_A - \eta_B) \geq \delta)$, using a numerical-integration

method based on the extended-trapezoidal rule (Press et al. (1992)). This application, including source code and test data, is available from the authors.

3 Experimental results

We evaluated vBLSM on both simulated and real data. The simulated data provide ground truth for validation. Data from a study of acute ischemic left-hemispheric stroke allow us to determine whether our approach yields results that are consonant with functional neuroanatomy.

3.1 Simulated data

In this experiment, we generated simulated lesion maps for 40 subjects. We divided these subjects into two equal groups: subjects in group A had impaired cognitive function, and subjects in group B had normal cognitive function.

We constructed simulated lesion maps based on the CH2 or AAL template (Holmes et al. (1998)). First, we chose a region of interest. In this experiment, we chose Brodmann area (BA) 48 in the CH2 atlas as the region of interest. Second, we introduced lesions; that is, voxels in BA 48 would be abnormal with probability λ , and voxels outside of BA 48 would be abnormal with probability ν . λ and ν represent the signal and noise levels, respectively. In particular, ν represents a combination of a voxel's being lesioned outside the region-of-interest, background noise, and image-processing error. Third, to make the simulation more realistic, we applied an in-plane median filter to each subject's lesion map to increase spatial smoothness. Let D_i denote the smoothed lesion map for subject i . $\mathbf{D} = \{D_i\}$ constituted the input to our lesion-symptom mapping software.

In our first experiment, for subjects in group B, each brain voxel had a probability of 0.2 of assuming value '1' (being lesioned). For subjects in group A, a voxel inside the brain but not in BA 48 had a probability of 0.2 of being lesioned, and a voxel in BA 48 had a probability of 0.4 of being lesioned. We refer to this experiment as simulation study I.

We analyzed $\mathbf{D}$ using vBLSM, calculating $d(\delta) = Pr((\eta_A - \eta_B) \geq \delta)$, with a noninformative prior. We considered voxels with $d(\delta) \geq th$ as demonstrating a significant difference in lesion proportions.

We then compared our analysis with analyses based on NHST. The purpose of comparison is to compare vBLSM with the state-of-the-art NHST-based approaches. The Lieberman test with different correction methods (Rorden et al. (2007)) is a recently proposed method to solve the lesion-symptom mapping problem. This method is more sensitive than the χ^2 test. We analyzed $\mathbf{D}$ using the Lieberman test implemented in MricroN (<http://www.sph.sc.edu/comd/rorden/mricron/>), with a significance level of 0.05. We thresholded the resulting statistical map using three multiple-comparison correction methods: Bonferroni correction, false-discovery rate (FDR), and familywise error rate (FWE) with 1000 permutations. Details of the Lieberman test and these multiple-comparison correction methods are in Rorden et al. (2007).

Figure 1 shows the results of the simulated-data experiment I. Figure 1(a) shows BA 48 in the CH2 atlas. Figure 1(b) shows the lesion-density map for all subjects. For a given voxel, lesion density is defined as the fraction of subjects with a lesion overlapping that voxel, and therefore assumes a value in the range [0, 1]. In Figure 1, voxels demonstrating significant lesion-proportion differences are colored red, and are overlaid on the CH2 template. Let $\mathbf{A}$ denote the set of voxels demonstrating significant lesion-proportion differences, $\mathbf{B}$ denote the voxels in BA 48 in the CH2 template, and $\mathbf{V}$ denote all voxels inside the brain in the CH2 template. We

can evaluate a lesion-symptom mapping algorithm using the true-positive rate (TPR) and false-positive rate (FPR), defined as

$$TPR = \frac{\Omega(\mathbf{A} \cap \mathbf{B})}{\Omega(\mathbf{B})}$$

and

$$FPR = \frac{\Omega(\mathbf{A} \setminus \mathbf{B})}{\Omega(\mathbf{V} \setminus \mathbf{B})},$$

where Ω is an operator that calculates region volume.

Table 1 lists the TPR and FPR for the Lieberman test with Bonferroni, FDR, and permutation-based FWE correction, and for vBLSM with different δ and th settings. As expected, we found that the Lieberman test with Bonferroni correction is extremely conservative: it detected only one voxel, with $TPR=0.000013$ and $FPR=0.0$. For the Lieberman test with FDR correction, $TPR=0.59$ and $FPR=0.000087$. For the Lieberman test with permutation-based FWE correction, $TPR=0.16$ and $FPR=0.000001$. As expected, TPR and FPR of vBLSM vary across different δ and th values.

The Lieberman test with Bonferroni correction is extremely conservative, and provides little information about regions characterizing group differences. The Lieberman test with permutation FWE correction is still conservative, with $TPR=0.16$. Both the Lieberman test with FDR correction and vBLSM with ($\delta = 0$, $th = 0.95$) have high sensitivity. Note that $\delta = 0$ and $th = 0.95$ are the values that most researchers would use when analyzing their data with vBLSM. vBLSM with $\delta = 0$ and $th = 0.95$ ($TPR=0.63$) is slightly more sensitive than the Lieberman test with FDR ($TPR=0.59$), at the cost of slightly increased FPR. The set $\mathbf{A}$ (voxels demonstrating significant lesion-proportion differences) generated by vBLSM correctly identified 50289 of the 79377 voxels in BA 48, and included an extra 1312 voxels. The set $\mathbf{A}$ generated by the Lieberman test with FDR correctly identified 46669 voxels in BA 48, and included an extra 611 voxels. vBLSM correctly included 3620 more voxels in BA 48, at the cost of adding 701 more extra voxels. In this case, vBLSM is more accurate than the Lieberman test with FDR.

We analyzed another simulated-data set, to determine our algorithm's performance when there is no effect (i.e., the lesion proportion of BA 48 in group A is same as that of group B). In this experiment, for subjects in groups A and B, each brain voxel in BA 48 had a probability of 0.25 of assuming state '1', and voxels outside of BA 48 had a probability of 0.20 of assuming state '1'. That is, there is no effect across groups. We refer to this experiment as simulation study II.

In this case, since there is no region manifesting an effect, TPR is not defined. The only meaningful metric is FPR, which is defined as the number of voxels incorrectly labeled as demonstrating group differences, divided by the total number of voxels in the brain template. Table 2 lists FPRs for the Lieberman test with Bonferroni, FDR, and permutation-based FWE correction, and for vBLSM with different δ and th settings.

3.2 Acute ischemic left-hemispheric stroke

We analyzed data collected from 33 subjects with acute ischemic left-hemispheric stroke. These subjects were classified with respect to word-reading impairment: there were 18 subjects with impaired word reading, and 15 subjects with normal word reading.

All subjects underwent diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI) as part of the Johns Hopkins Acute Stroke Protocol, within 24 h of symptom onset. Combining DWI and PWI images, we identified voxels with tissue dysfunction due to either hypoperfusion or abnormally restricted diffusion. For DWI, two experts, who were blinded to the results of the reading-test results, manually delineated regions of abnormally restricted diffusion. For PWI, we compared each voxel to its contralateral (i.e., mirror-image) counterpart, and labeled voxels with time-to-peak differences of 4 sec or greater as regions of hypoperfusion. We refer to the resulting regions of hypoperfusion in PWI, and regions of acute ischemia in DWI, as the PWI and DWI lesion maps, respectively.

For each subject, we registered the subject's DWI and PWI images to the Montreal Neurological Institute's Jakob template, using a mutual-information maximization algorithm (Jenkinson and Smith (2001)). We then applied this transformation to that subject's PWI and DWI lesion maps. For a voxel in the MNI Jakob space, if it was infarcted in the normalized DWI image, or hypoperfused in the normalized PWI image, we labeled this voxel as abnormal. Details about data acquisition and image processing for these data are presented in Chen et al. (2008).

Figure 2 shows the results of analyses of these data. Figures 2(a) and (b) show the lesion-density maps for subjects without and with impairment of word reading, respectively. Figure 2(c) shows voxel-wise Lieberman statistics, which were in the range $[-2.9288, 2.7254]$. The cut-off thresholds for voxels demonstrating significant differences under significance level 0.05, for Bonferroni correction, FDR correction, and permutation FWE correction, were 5.094, 9.200, and 3.289, respectively. The set $\mathbf{A}$ was empty for the Lieberman test with Bonferroni, FDR, and permutation FWE corrections. Figure 2(d) depicts voxels with $d(0) \geq 0.95$ (colored red) based on analysis with vBLSM. These voxels were primarily in BA 37.

4 Conclusion and Discussions

We have described and implemented a Bayesian voxel-based lesion-symptom mapping algorithm, which represents an alternative to NHST-based approaches. The advantages of vBLSM over NHST-based approaches are: 1) vBLSM explicitly models lesion-proportion differences, which are of great clinical importance, whereas significance as reflected by a p -value may not have as much practical importance, particularly when the corresponding statistics fall above the significance threshold. 2) vBLSM obviates multiple-comparison correction. We have provided an analytical foundation for the calculation of $Pr((\eta_A - \eta_B) \geq \delta)$, the probability that a proportion difference is greater than a threshold. This Bayesian-inference formulation is exact, that is, it does not involve approximation procedures. Because we can compute this distribution exactly, vBLSM is much more computationally efficient than it would be if it were based on approximation techniques.

The results of simulation study I demonstrate that vBLSM is comparable to or more accurate than NHST-based approaches for the delineation of regions that characterize group differences. The results of simulation study I also demonstrate that vBLSM can balance TPR and FPR . The performance of vBLSM depends on the settings of δ and th . In general, vBLSM does not manifest the extreme conservatism associated with the Lieberman test with Bonferroni correction.

For simulation study II, in which there was no difference among groups, we could not define *TPR*; therefore, we used only *FPR* for evaluation. As we expected, conservative procedures achieved extremely low *FPR* (Table 2). vBLSM, with the typical setting ($\delta = 0$, $th = 0.95$) balancing *TPR* and *FPR*, tends to have higher *FPR* than more conservative procedures when there is no difference among groups. However, even in this case, the *FPR* for vBLSM with ($\delta = 0$, $th = 0.95$) is still very low.

In the study of acute ischemic left-hemispheric stroke, NHST-based approaches (Liebermeister test with Bonferroni, FDR, or permutation FWE correction) did not detect any voxels demonstrating a significant lesion-proportion difference. This result can be explained by the small effect size. For a test of two proportions, power is determined by sample size and the difference between the two proportions. In this experiment, the number of subjects with lesions at a given voxel was small, for both the impaired word-reading and the normal word-reading groups. Therefore, the difference between the two lesion proportions was small, for any voxel. If we had had a larger sample size, NHST-based approaches would probably be able to detect this difference. However, with a sample size of 33 and small lesion-proportion differences, NHST-based approaches have insufficient power to detect any voxels. In contrast, vBLSM detects voxels characterizing group differences: lesions in BA 37 are associated with impaired word reading. This area is known to be important for prelexical processing of letter strings (Hillis et al. (2005)), as evidenced by activation in response to written words as assessed by functional MR (Puce et al. (1996)).

In lesion-deficit studies, the amount of lesion overlap and the spatial distribution of lesions are highly variable. In the case that the number of subjects lesioned at a given voxel is small and the sample size is small, such as our study of acute ischemic left-hemispheric stroke, the capability of our Bayesian method to detect small changes in lesion proportions is crucial to identifying brain regions that are associated with impaired neuro-cognitive functioning.

Another benefit of vBLSM is that, in addition to inference regarding proportion differences among groups, we can use Equation (4) for other types of inference. For example, we may be interested in the probability that the lesion-proportion difference between groups is in an interval $[t_1, t_2]$. Since $Pr(t_1 \leq (\eta_A - \eta_B) \leq t_2) = Pr((\eta_A - \eta_B) = t_1) - Pr((\eta_A - \eta_B) \geq t_2)$, we can use Equation (4) to calculate this interval.

Functional brain imaging and lesion-deficit analysis are complementary methods for investigating structure function relationships. NHST-based methods are widely used to detect group differences in functional-imaging studies (Friston et al. (1995)), and Bayesian approaches have been proposed to address the limitations of NHST-based approaches (Friston et al. (2002a,b)). Our method is related to these Bayesian methods for functional imaging studies, yet differs from them in that the latter are based on the multivariate Gaussian distribution, whereas our method is based on the Beta distribution.

We previously described a Bayesian lesion-deficit analysis method, called voxel-wise Bayesian Lesion Deficit Analysis (vBLDA), in Chen et al. (2008). The differences between vBLSM and vBLDA are as follows. First, vBLSM falls under the lesion-symptom framework and is mass-univariate, whereas vBLDA is a multivariate analysis method. vBLDA centers on representing linear or nonlinear probabilistic associations among brain regions and C , while vBLSM centers on the associations between an individual voxel and C . In other words, we designed vBLSM to evaluate functional segregation (Friston (2005)), whereas vBLDA evaluates functional integration (Friston (2005)). Second, vBLDA incorporates spatial information in the form of Markov random fields to model spatial correlations among voxels, whereas the current version of vBLSM does not incorporate spatial information.

Assessing row-column associations in a two-way contingency table is a well-established topic in Bayesian analysis (Altham (1969); Gelman et al. (1995); Fredette and Angers (2002); Congdon (2005)). The quantity of interest used in this paper is the proportion difference, $\eta_A - \eta_B$. We use this quantity because of its clinical relevance and clear interpretation. Inference about whether C and V are associated is based on $Pr((\eta_A - \eta_B) \geq \delta)$. Other quantities, such as the log-odds ratio or Bayes factor, could also be used to detect an association between V and C . The log-odds ratio is defined as $\log(\frac{\eta_A/(1-\eta_A)}{\eta_B/(1-\eta_B)})$. The performance of approximation methods for the posterior distribution of log-odds ratio is presented in Fredette and Angers (2002).

One of limitations of the proposed approach is that it is designed to handle binary lesion maps (voxels are lesioned or normal) and a binary group-membership variable. For studies involving continuous lesion maps, voxel V assumes a continuous value. In this case, we can use a Bayesian inference procedure based on the Gaussian distribution (Neumann and Lohmann (2003)) to identify voxels characterizing group differences. This procedure assumes that η , the effect of interest, is Gaussian: $p(\eta_A) \sim N(\mu_A, \sigma_A^2)$ and $p(\eta_B) \sim N(\mu_B, \sigma_B^2)$. Let $\mathbf{D} = \{\hat{\eta}^1, \dots, \hat{\eta}^n\}$, represent the effects of interest for all subjects. The posterior distribution $d = \eta_A - \eta_B$ has the analytical form $p(d|\mathbf{D}) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{d-\mu}{\sigma})^2]$, where $\mu = \mu_A - \mu_B$ and $\sigma = (\sigma_A^2 + \sigma_B^2)^{1/2}$ and $\mu_A, \mu_B, \sigma_A, \sigma_B$ are estimated from $\mathbf{D}$. We can make inference based on the posterior distribution $d = \eta_A - \eta_B$.

The proposed Bayesian approach can analyze only one clinical variable. Some applications require the analysis of several categorical clinical variables; one variable is the outcome variable (denoted by C), and the other variables are covariates. The task is to detect differences in the spatial distributions of lesions across C , adjusting for the covariates. We plan to develop a method based on latent-variable modeling (Bartholomew and Knott (1999)) to support such analyses. It will be a two-stage algorithm. In the first stage of this analysis, for each voxel V , we stratify C by levels of other covariates, and generate a series of contingency tables. For example, if C represents a subject's having impaired or normal cognitive function, and there is a covariate Sex , we would generate two contingency tables between V and C , for $Sex = female$ and $Sex = male$, respectively. For contingency table k , we calculate $d_k(\delta)$ using the Bayesian method proposed in this paper. In the second stage, each voxel V is associated with a binary latent variable L that represents whether or not the lesion proportions are significantly different across groups defined by C . Assuming that $\{d_k(\delta)\}$ are conditionally independent given L (the standard conditional-independence assumption in latent-variable modeling), and deriving a formula to link $d_k(\delta)$ and L , we obtain the posterior distribution of L . Inference regarding lesion-proportion differences is based on the posterior distribution of L .

Our future work includes a series of validation experiments, based on simulated data, to assess how TPR and FPR vary with factors such sample size and the voxel-wise probability of being lesioned. Another direction of future research is an extension of our voxel-based Bayesian lesion-symptom mapping algorithm by incorporating spatial information. Our current version analyzes a voxel independently of its neighbors, which greatly simplifies inference. However, a more realistic spatial model for imaging data should include spatial interactions among neighboring voxels. That is, we expect voxels characterizing group differences to be grouped into regions, rather than randomly distributed in space. We will incorporate Markov random field modeling (Geman and Geman (1984)) into our Bayesian analysis, which should further improve the sensitivity and specificity of this approach.

Acknowledgments

Drs. Chen and Herskovits are supported by National Institutes of Health grant R01 AG13743, which is funded by the National Institute of Aging, the National Institute of Mental Health, and the National Cancer Institute. This work is also supported by National Institutes of Health grant R01 DC05375.

References

- Altham P. Exact Bayesian analysis of a 2x2 contingency table, and fishers exact significant test. *Journal of the Royal Statistical Society Series B* 1969;31 (2):261–269.
- Bartholomew, D.; Knott, M. Latent variable models and factor analysis. Vol. 2. Arnold; London: 1999.
- Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. Voxel-based lesion-symptom mapping. *Nature Neuroscience* 2003;6 (5):448–450.
- Chen R, Hillis AE, Pawlak M, Herskovits EH. Voxelwise bayesian lesion deficit analysis. *NeuroImage* 2008;40 (4):1633–1642. [PubMed: 18328733]
- Congdon, P. Bayesian models for categorical data. Wiley; 2005.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR. A neural basis for lexical retrieval. *Nature* 1996;380:499–505. [PubMed: 8606767]
- Fredette M, Angers J. A new approximation of the posterior distribution of the log-odds ratio. *Statistica Neerlandica* 2002;56 (3):314–329.
- Friston KJ. Models of brain function in neuroimaging. *Annu Rev Psychol* 2005;56:57–87. [PubMed: 15709929]
- Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications. *NeuroImage* 2002a;16 (2):484–512. [PubMed: 12030833]
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 1995;2:189–210.
- Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* 2002b;16 (2):465–483. [PubMed: 12030832]
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian Data Analysis. Chapman Hall/CRC; Boca Raton: 1995.
- Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans PAMI* 1984;6:721–741.
- Hillis AE, Newhart M, Heidler J, Barker P, Herskovits EH, Degaonkar M. The roles of the "visual word form area" in reading. *NeuroImage* 2005;24 (2):548–549. [PubMed: 15627597]
- Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography* 1998;22:324–333. [PubMed: 9530404]
- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 2001;5(2):143–156. [PubMed: 11516708]
- Karnath HO, Berger MF, Kuker W, Rorden C. The anatomy of spatial neglect based on voxelwise statistical analysis: a study of 140 patients. *Cereb Cortex* 2004;14(10):1164–1172. [PubMed: 15142954]
- Kimberg DY, Coslett HB, Schwartz MF. Power in voxel-based lesion-symptom mapping. *Journal of Cognitive Neuroscience* 2007;19:1067–1080. [PubMed: 17583984]
- Krueger J. Null hypothesis significance testing. On the survival of a flawed method. *Am Psychol* 2001;56 (1):16–26. [PubMed: 11242984]
- Neumann J, Lohmann G. Bayesian second-level analysis of functional magnetic resonance images. *NeuroImage* 2003;20 (2):1346–55. [PubMed: 14568503]
- Press, WH.; Flannery, BP.; Teukolsky, SA.; Vetterling, WT. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press; 1992.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 1996;16 (16):5205–5215. [PubMed: 8756449]
- Rorden C, Karnath H, Bonilha L. Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience* 2007;19 (7):1081–1088. [PubMed: 17583985]
- Tyler KL, Marslen-Wilson W, Stamatakis EA. Dissociating neuro-cognitive component processes: voxel-based correlational methodology. *Neuropsychologia* 2005;43:771–778. [PubMed: 15721189]

Figure 1.

Results for simulation study I, overlaid on the CH2 template (gray-scale). a) Brain region BA 48 (red) from the CH2 atlas. b) Lesion density. Voxels demonstrating significant difference between groups (red), at significance level 0.05, based on the Lieberman test with c) Bonferroni correction, d) FDR correction, and e) permutation FWE correction. f)-k) Results for vBLSM. f) Posterior of $d(0)$. g) Voxels with $d(0) \geq 0.99$ (red). h) $d(0) \geq 0.95$. i) Posterior of $d(0.05)$. j) $d(0.05) \geq 0.99$. k) $d(0.05) \geq 0.95$.

Figure 2.

Results of acute ischemic left-hemispheric stroke experiment. a) Lesion-density map for subjects with no word-reading deficit. b) Lesion-density map for subjects with word-reading deficit. c) Lieberman test statistics. d) Voxels with $d(0) \geq 0.95$ (red) based on vBLSM, overlaid on the MNI Jakob template (gray-scale).

Table 1

TPR and FPR for the simulated-data experiment I.

Method		TPR	FPR
Liebermeister Test	Bonferroni	0.000013	0.0
	FDR	0.59	0.000087
	Permutation	0.16	0.000001
vBLSM	$d(0) \geq 0.95$	0.63	0.000190
	$d(0) \geq 0.99$	0.28	0.000018
	$d(0.05) \geq 0.95$	0.43	0.000054
	$d(0.05) \geq 0.99$	0.16	0.000001

Table 2

TPR and FPR for simulation study II.

Method		FPR
Liebermeister Test	Bonferroni	0.00
	FDR	0.00
	Permutation	0.000001
vBLSM	$d(0) \geq 0.95$	0.000022
	$d(0) \geq 0.99$	0.000001
	$d(0.05) \geq 0.95$	0.000013
	$d(0.05) \geq 0.99$	0.00