

Scan Stratified Case-Control Sampling for Modeling Blood-Brain Barrier Integrity in Multiple Sclerosis

GM. Pomann,^{a*} E.M. Sweeney,^{bc} D.S. Reich,^c AM. Staicu^{a**} and R.T. Shinohara^{d**}

Multiple sclerosis (MS) is an immune-mediated neurological disease that causes morbidity and disability. In patients with MS, the accumulation of lesions in the white matter of the brain is associated with disease progression and worse clinical outcomes. Breakdown of the blood-brain barrier in newer lesions is indicative of more active disease-related processes and is a primary outcome considered in clinical trials of treatments for MS. Such abnormalities in active MS lesions are evaluated *in vivo* using contrast-enhanced structural magnetic resonance imaging (MRI), during which patients receive an intravenous infusion of a costly magnetic contrast agent. In some instances, the contrast agents can have toxic effects. Recently, local image regression techniques have been shown to have modest performance for assessing the integrity of the blood-brain barrier based on imaging without contrast agents. These models have centered on the problem of cross-sectional classification in which patients are imaged at a single study visit and pre-contrast images are used to predict post-contrast imaging. In this paper, we extend these methods to incorporate historical imaging information, and we find the proposed model to exhibit improved performance. We further develop scan-stratified case-control sampling techniques that reduce the computational burden of local image regression models, while respecting the low proportion of the brain that exhibits abnormal vascular permeability. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: Case-Control sampling; Logistic regression; Magnetic resonance imaging; Multiple sclerosis

1. Introduction

Multiple sclerosis (MS) is an immune-mediated neurological disease that causes morbidity and disability. In patients with relapsing-remitting MS, the most common first stage of the disease [1, Chapter 15], the accumulation of lesions in the

†** These authors contributed equally.

^aDepartment of Statistics, North Carolina State University, Raleigh, NC.

^bDepartment of Biostatistics, Johns Hopkins School of Public Health, Baltimore, MD.

^cTranslational Neuroradiology Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD.

^dDepartment of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

* Correspondence to: Department of Statistics, North Carolina State University, Raleigh, NC. E-mail: gpomann@gmail.com

white matter of the brain is associated with disease progression and worse clinical outcomes. Breakdown of the blood-brain barrier (BBB) in newer lesions is indicative of more active disease processes. Such abnormalities in active MS lesions are evaluated *in vivo* using contrast-enhanced structural magnetic resonance imaging (MRI). Lesions exhibiting these abnormalities are referred to as enhancing due to their hyperintense presentation on T1-weighted (T1w) images after intravenous administration of a contrast agent. The number and volume of such enhancing lesions are important for the clinical management of patients with MS and are primary outcomes of clinical trials of treatments for MS [2, 3].

The standard procedure for assessing changes in the BBB involves the comparison of pre-contrast MRIs with MRIs acquired after the intravenous infusion of gadolinium, a magnetic contrast agent. However, contrast-enhanced imaging can cost 38% more than taking an MRI without contrast [4]. Gadolinium-based contrast agents have, in rare cases, been associated with kidney problems as well as allergic reactions [5, 6, 7]. Thus, methodology for assessing the integrity of the BBB based on contrast-free imaging potentially has broad clinical implications.

Our contribution is twofold. First, using only pre-contrast images and historical information, we model the probability that a given voxel in an MRI would enhance if the patient had been given a contrast agent. The model is used to assess disease activity through BBB integrity using MRI without contrast. Our methodological developments are motivated by longitudinal data from an ongoing observational study at the National Institute of Neurological Disorders and Stroke (NINDS). In this work, we study a subset of high-resolution structural MRIs acquired in hundreds of patients with MS on a monthly basis with no specified end date. The subset of patients was chosen to be an observational cohort-study of the subjects who had multiple scans over a one year time period with a research-quality contrast-enhanced MRI protocol. Each image has over one million measurements within the brain; this data set is growing at a rapid pace and computationally scalable methods for fitting statistical models are crucial. Our second contribution is the proposal of a novel sub-sampling technique that greatly reduces the computational burden of the estimation procedure. The proposed model, which utilizes historical information, shows superior prediction performance in terms of the receiver operating characteristic (ROC) curve when compared to the model that does not use this information. Furthermore, using a sample of approximately 750 voxels from each image yields a comparable ROC curve to the case when the proposed model is fit on the full data set that consists of one million voxels per image.

The ground truth or 'gold standard' for identifying enhancing white matter lesions involves the manual comparison of pre- and post-contrast T1-weighted imaging conducted by an expert neuroradiologist [8, 9, 10, 11]. However, there are image features visible on high-resolution scans acquired on 7 tesla scanners that correlate with BBB disruption [12, 13]. Furthermore, [14] showed that in some cases the adjudication of BBB integrity does not require post-contrast imaging. They used a voxel-level logistic regression model for predicting enhancement using pre-contrast voxel intensities in T1w and T2-weighted (T2w) images along with the interaction between the intensities. To reduce the number of false positives, the model was fit on a subset of the voxels, defined by thresholding the T2w fluid attenuated inversion recovery (FLAIR) images within each scan, as FLAIR images are known to have high values for lesion voxels. Taking only regions with FLAIR intensity in the top 1% provides a good set of voxels that is likely to include most enhancing lesions. Sensitivity analysis showed that the method was robust to changes in this threshold value [14].

While thresholding on the FLAIR images does significantly reduce the sample size of the data, one criticism of this method is that all of the enhancing voxels may not be included in this sub-sample. However, in the case of MRI data of patients with MS, enhancement is rare; there are many fewer voxels that enhance compared to those that do not. It is important for our model that these rare events are all contained in the sub-sample. One can think of the data set as case-control data such that enhancing voxels are labeled as cases and non-enhancing voxels are labeled as controls. Using this setting we propose a novel sub-sampling method that takes advantage of well established theory from the case-control study literature [15, 16].

It is known that voxels belonging to new MS lesions are more likely to enhance than voxels in older lesions [17]. Subtraction-Based Logistic Inference for Modeling and Estimation (SuBLIME) was recently proposed to estimate the probability that a given voxel is part of a new or enlarging lesion between two study visits [18]. The model relies on the

historical study of patients, as the differences between images taken from the same patient at different visits are used as covariates. Then, for any given voxel in an image, one can obtain a corresponding probability that it belongs to an incident or enlarging lesion. Therefore, to predict enhancement in a contrast-free MRI study, we propose using SuBLIME voxel probabilities along with the T1w and T2w voxel intensities in a logistic regression model. This allows us to model the probability that voxels will enhance using only pre-contrast images and temporal information about lesion incidence and enlargement.

Section 3 presents the proposed model and compares the performance to that of [14]. Section 4 then proposes how the MR images acquired as part of the NINDS study can be analyzed in a computationally scalable fashion as case-control data where the cases (enhancing voxels) are rare events. Results from the proposed methods are evaluated using cross-validation on 15 patients who received structural MR scans at between two through nine visits. In Section 5, the results of the new model and sampling scheme are compared to those of [14], and we find that the proposed methodology improves prediction and reduces computational burden by reducing the number of voxels the model fits on.

2. NINDS Data

The MRI modalities of interest in this paper consist of FLAIR, pre-contrast T1w, and pre-contrast T2w images. For all voxels in the brain, the imaging modalities are normalized by subtracting their mean and dividing by their standard deviation [19, 20, 21]. All images were preprocessed using the procedure outlined in the Experimental Methods section of [14] which is outlined in the the supplementary material of this paper.

For the purpose of modeling, we also obtain SuBLIME maps that provide the probability that each voxel in the brain of a patient is part of a new or enlarging lesion since the patient's last study visit [18]. The probabilities are estimated with logistic regression in which the covariates are T1w, T2w, proton density-weighted (PDw), and FLAIR normalized intensities along with the lag time between consecutive visits, the relative change in these imaging modalities, and their interactions with time. SuBLIME maps have not previously been used as historical covariates in the context of image analysis. More details about estimation of these maps are provided in the supplementary material of this paper.

Additionally, we obtain binary radiologist gold standard manual segmentation masks (RGS), indicating where enhancing lesions exist, and brain masks, indicating where cerebral tissue lies in each image. In order to build the proposed predictive model, we use these RGS masks as the response variable. The masks are created by a neuroradiologist (D.S.R.) with over 10 years of clinical experience with MR imaging in MS. To define these masks, the neuroradiologist uses all available post-contrast imaging modalities including, but not limited to, the post-contrast T1w images. We consider this RGS mask to be our ground truth as this is the current standard in practice for identification of enhancing voxels. The details of the pre-processing of the images in this study are described in the supplementary material. For the analysis presented in this paper we use 77 brain MRIs from 15 individuals with MS, obtained under an institutional review board-approved protocol. Each of these patients was scanned between two and nine times. Informed consent was obtained from all participants.

3. Logistic Enhancement Models

We present a model for predicting voxel enhancement using pre-contrast MRIs by advancing the work of [14] and [18]. In [14] it was shown that voxel enhancement can be predicted using a cross-sectional logistic model that is fit using the pre-contrast T1w and T2w voxel intensities along with the interaction under the assumptions of independence between scans and voxels. We define enhancement using a binary variable, $E_i(v, t_{ij})$, which equals one if voxel v in the brain of patient i enhances at time t_{ij} and zero otherwise. The model presented in [14] assumes that enhancement across images

for patient i at time t_{ij} are independent so that,

$$\text{logit}\{\text{Pr}E_i(v, t_{ij})\} = \beta_0 + \beta_1 M_{i,1}(v, t_{ij}) + \beta_2 M_{i,2}(v, t_{ij}) + \beta_{12} M_{i,1}(v, t_{ij}) M_{i,2}(v, t_{ij}) \quad (1)$$

where intensity on imaging modality k in voxel v for subject i at time t_{ij} is denoted as $M_{i,k}(v, t_{ij})$ for $i = 1 : n$, $j = 1 : m_{i,k}$, and $m_{i,k}$ is the number of scans using modality k for subject i . Here, $M_{i,1}$ and $M_{i,2}$ refer to the T1w and T2w images for patient i respectively.

In Model 1, the fact that new or enlarging MS lesions are more likely to enhance [17] is not taken into account. It is our assertion that one may include a covariate that describes lesion age to improve predictive ability. This information can be obtained by having a radiologist manually indicate where new or enlarging lesions exist in an image. However, manual identification is extremely time consuming and prone to errors even for the most expert neuroradiologists in this field. Additionally, the process is subject to both intra-observer and inter-observer variability [10]. Therefore, we propose the use of the SuBLIME method to estimate the probability that a voxel belongs to a new or enlarging lesion [18].

One of the main contributions of this paper is the development of a predictive model that uses information about temporal lesion behavior to predict the probability of a voxel enhancing. The proposed modeling procedure is a two-step process. First, for each voxel used for fitting the model, we acquire historical information about whether the voxel belongs to a new or enlarging lesion. This can either be obtained from a mask developed by a radiologist or from the SuBLIME method [18], which estimates the corresponding probabilities. In this paper, we employ the latter technique. For this, we define the indicator of incidence as a binary variable, $W_i(v, t_{ij})$, which is equal to one if subject i has new lesion incidence in voxel v at time t_{ij} and zero otherwise. The SuBLIME probability maps, $SP_i(v, t_{ij}) := \text{Pr}[W_i(v, t_{ij}) = 1]$, are the probabilities that voxel v is with a new or enlarging lesion at time t_{ij} . As described in the previous section, these maps are obtained using longitudinal study information. Details about the estimation of this covariate using SuBLIME [18] are presented in the supplementary material for this paper.

The second step of our fitting procedure consists of modeling the probability that a voxel enhances by accounting for T1w, T2w pre-contrast images, their interaction, and the covariate describing the new or enlarging lesion incidence. Specifically we assume the following logistic model,

$$\begin{aligned} \text{logit}\{\text{Pr}E_i(v, t_{ij})\} = & \beta_0 + \beta_1 M_{i,1}(v, t_{ij}) + \beta_2 M_{i,2}(v, t_{ij}) + \beta_3 SP_i(v, t_{ij}) \\ & + \beta_{12} M_{i,1}(v, t_{ij}) M_{i,2}(v, t_{ij}). \end{aligned} \quad (2)$$

Note that one implicit assumption of this model is that the covariates are observed without error. Efforts are made in preprocessing and estimation to reduce the error as much as possible. However, any error that does remain may propagate through the model and impact prediction. Nevertheless, we achieve high predictive accuracy as presented in Section 5. We compare the predictions of Model (1) and Model (2) by analyzing the classification performance using the empirical receiver operating characteristic (ROC) curve. We consider both the full and partial area under the curve (AUC) of the ROC curve, obtained through cross-validation. We also investigate the impact of using different sub-sampling techniques that aim to reduce the computational expense and improve prediction. In the next section, we present a modified case-control sampling technique, which we show to further improve the predictions.

4. Sampling on 3D Magnetic Resonance Images

Our data consist of images that each contain over one million voxels, therefore the data set analyzed in this paper contains over 77 million voxels. Furthermore, in this data set, the average proportion of voxels that are known to enhance over all RGS masks is 0.004, indicating that enhancement is an extremely rare event in these data. Developing a computationally inexpensive fitting procedure would thus be very beneficial for researchers. However, it is also important that this procedure takes into account the rarity of enhancement. In this section we propose a procedure that not only accounts for this, but reduces the computational time by two orders of magnitude and maintains high predictive power. The sub-sampling method will facilitate large scale analysis of the proposed predictive model in future studies.

We begin by considering the population of interest to be all voxels in all brains of people with MS. The full data set under study consists of the voxels in the brain scans of the subjects in the observational study. Using case-control terminology, we refer to the voxels that are identified (by the RGS) as enhancing and non-enhancing by cases and controls respectively.

Sub-sampling without replacement from the total number of voxels would often result in a sample that only contains controls. Therefore, in order to ensure full representation of the cases and maintain some of the spatial dependence in the sampling scheme, we propose a scan-stratified case-control (SSCC) sampling approach. Let $RGS_i(v, t_{ij})$ be the binary RGS mask, which equals one if voxel v in the brain image of patient i taken at time t_{ij} is defined as enhancing and zero otherwise. Denote the mask over all voxels v in the brain image of patient i taken at time t_{ij} as RGS_{ij} . Furthermore, let l_{ij} be the number enhancing voxels in RGS_{ij} and \bar{n}_e be the average number of enhancing voxels over all masks. Then for each image taken of patient i at time t_{ij} , the sub-sample of cases consists of all enhancing voxels as defined by the RGS_{ij} . The control sample is constructed by taking a simple random sample from all non-enhancing voxels in each scan with the following rules:

- If $l_{ij} > 0$ (the RGS contains enhancing voxels): take a simple random sample of size five times l_{ij} from the non-enhancing voxels identified by RGS_{ij} .
- If $l_{ij} = 0$ (the RGS does not contain any enhancing voxels): take a simple random sample of size five times \bar{n}_e from the non-enhancing voxels identified by RGS_{ij} .

The constant five is recommended in [16] for the case of fitting a logistic model to a case control sample for rare event data. Let $SSCC_{ij}$ be the set of all voxels in the case-control sub-sample for patient i at time t_{ij} (determined using RGS_{ij}). Then the final case-control sub-sample of voxels used to fit the model is $SSCC = \bigcup_{ij} SSCC_{ij}$.

The SSCC sampling method ensures that all of the cases are included in the sub-sample used to fit logistic models (1) and (2). This is desirable because including enough cases in the sample can reduce the variation of logistic regression coefficient estimates [22, 23, 24, 16]. However, if only voxels $v \in SSCC$ are used to fit the proposed models, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_{12}$ will be consistent, but $\hat{\beta}_0$ may be biased [15, 16]. Since our goal is to maintain high predictive power when only using $\{v : v \in SSCC\}$ to fit models (1) and (2) it is important to address this problem.

The bias in this estimate can be easily removed by using the correction presented in [16]. In their paper, they focus on model estimation when fitting logistic models with a case-control sample for rare event data. However, they are motivated by the difficulty of obtaining large amounts of data. While it is not difficult for us to obtain a large amount of control data, it does greatly increase the computational expense of running the model with a large number of scans. Therefore, the goal is the same in both scenarios; to use a case-control sample to fit a logistic model without loss of prediction power.

Denote the intercept estimate from fitting a logistic model using only $\{v : v \in SSCC\}$ as $\hat{\beta}_0^{SSCC}$, where the superscript specifies the SSCC sub-sample used for fitting the model. Let τ be the population proportion of cases and \bar{y} be the proportion of cases in the sample. Then the bias-corrected coefficient estimate derived in [16] is, $\hat{\beta}_0 = \hat{\beta}_0^{SSCC} - \ln[(1 - \hat{\tau}/\tau)(\bar{y}/1 - \bar{y})]$. We obtain $\hat{\tau}$ by computing the proportion of enhancing voxels in $\bigcup_{ij} RGS_{ij}$. The number of enhancing voxels in $SSCC$ is computed to obtain \bar{y} . This proposed sampling technique and estimation procedure requires a working independence assumption of the case-control samples. While this may be incorrect, from Table 1 we find that using the SSCC sampling scheme does just as well as using the entire data set (All Voxels) in terms of pAUC. Using this scheme we maintain prediction accuracy while improving computational efficiency.

5. Enhancement Prediction Results

The full data set analyzed in this paper consists of 77 million voxels. Using the FLAIR sub-sampling method presented by [14] reduces the sample size to under 5.6 million voxels. The proposed SSCC sub-sampling method further reduces the sample size to 58 thousand voxels. In this section we compare the predictive performance of both models (1) and (2) when fitting on the full data set, the FLAIR sample, and the SSCC sample. The estimated ROC curves are used to determine

which of the six (2×3) possible model-sampling technique combinations achieves the best classification in terms of the false positive rate (FPR) and true positive rate (TPR) trade-off.

The ROC curves are estimated using cross-validation by resampling subjects. To accomplish this, we split the 15 subjects into two sets. All images from eight randomly selected subjects are placed in a training set and all images from the remaining seven subjects are placed in the validation set. The estimated ROC curve for each model-sample combination is taken to be the average ROC calculated on the validation set across 100 bootstrap replications. We conducted a sensitivity analysis (omitted) to investigate the stability of the average ROC curves, and we found 100 replications sufficient. To ensure proper comparison of the prediction performance, we fit the models using each of the three sets of data (all voxels, 1% FLAIR, SSCC), but validate using all brain voxels.

Due to the rarity of enhancement, analyzing FPRs larger than the actual event rate is not of interest. If the event rate is low, as in this data set, then investigating the ROC curves, for FPRs much greater than the event rate, does not give useful information about the predictive performance of the method in practice. To address this, we truncate the ROC curve at a fixed FPR, a common method of evaluating classification methods for rare event data [25]. To find a reasonable cutoff for such a partial ROC (pROC) analysis, we first note that in our data set the average rate of enhancing voxels in the entire brain over all scans is 4 per 1000. Since the data consist of 3D images, we consider each voxel in the context of a $3 \times 3 \times 3$ cube surrounding it; by multiplying .004 with $3^3 = 27$, we thus deem FPRs above 0.108 irrelevant.

Figure 1 provides plots of the pROC curves comparing models (1) and (2) when fit using each of the three data subsets (full data, 1% FLAIR, and SSCC). The partial AUC (pAUC) is a useful summary of the pROC reported in Table 1. This value is computed by taking the area under the partial curve and scaling it to lie between zero and one for the corresponding cutoff value. For reference, the full AUC results are included in the supplementary material. Note that model (2) outperforms model (1) when each is fit on all three data sets.

We observe that SSCC and full-brain sampling perform comparably for both methods, and both outperform FLAIR sampling. This can also be observed in Figure 2 which overlays the pROC curves corresponding to the three data sets for each model separately. From this pROC analysis, we conclude that i) using the SuBLIME covariate yields an improvement in the predictive accuracy of the model; and ii) fitting the models using the SSCC sampling method maintains predictive power, while vastly improving computational efficiency. Using a case-control sample we can find estimates for model (2) in 1.6 seconds using SSCC compared to 286.8 seconds using all voxels in the full data set.

An example of the improvement in classification of model (2) is presented in Figure 3 which displays an axial slice from one of patient's images used in this analysis. Figures 3A and 3B display the pre-contrast T1w and T2w images respectively. Figures 3C and 3D display the post-contrast FLAIR and T1w images respectively. The RGS which displays the location of the lesion defined by the neuroradiologist is shown in Figure 3H. Additionally, we present the resulting probability maps from models (1) and (2). The lesion (red) in the upper right portion of the images is detected in Figure 3F but not in Figure 3E, which display the predicted probabilities from model (2) and (1) respectively. However, it is identified as a newly enhancing lesion by the SuBLIME model which is displayed in Figure 3D. Note that the model (1) probability map (Figure 3E) also shows the ventricular CSF as enhancing lesion, indicating extrapolation errors outside of the FLAIR-hyperintense voxels, whereas model (2) probability map (Figure 3F) does not exhibit these distracting artifacts. [14] did not observe this artifact as they only predicted on the set of voxels corresponding to the top 1% of FLAIR histogram. We see this in our analysis because we predict on all voxels containing brain matter. Reducing this type of artifact in the predictions further motivates the use of the SSCC sampling method along with model (2). The coefficient estimates for the model-sample combinations are provided in Tables 2 and 3.

The proposed methodology does require a working independence assumption over voxels and scans. We find that with this assumption the method provides very good performance in terms of pAUC. To investigate if prediction improves if we account for some of the inherent dependence in the data we also consider a compound symmetric structure, by assuming constant dependence over voxels, but ignoring the dependence over time. This did not improve the predictions, but the results are included for reference in Section 4 of the Supplementary Material.

6. Discussion

We propose a new model that uses a historical covariate to predict lesion enhancement using contrast-free imaging. The gain in predictive performance that we observed is unprecedented, despite the simplicity and strong parametric assumptions of the model. Since using the historical covariate yields such an improvement, future work will focus on longitudinally modeling enhancement probabilities. Requiring estimation of such a historical covariate imposes that prior scans must have been collected. Therefore, the proposed model is not applicable for patients who are only scanned at one visit. This would not be a limitation in two-arm placebo-controlled trials where only post-randomization scans are relevant, as long as there is a baseline scan.

The proposed SSCC sub-sampling method relies on the estimation of the population parameter τ . However, this estimation will change depending on the application. This provides a natural way to calibrate the classifier for a different population of interest. The results show that by using the SSCC sub-sampling method, we greatly reduce the computational burden of fitting the model on the high-dimensional data set. We are able to fit the models 180 times faster than if we used the full data set. When using only this sub-sample to fit the proposed model, predictive performance comparable to using the full data set is achieved. This development allows for fitting next-generation local image regression models, which have recently shown promise in a variety of image analysis problems [18, 20, 14], on hundreds or thousands of subjects observed at many visits.

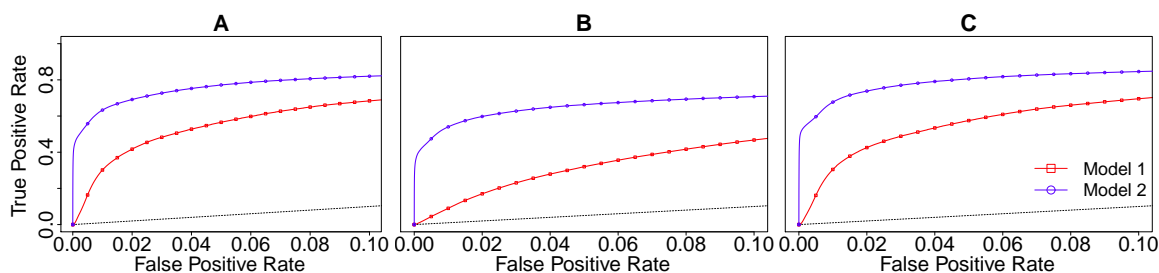


Figure 1. Partial ROC curves comparing models (1) and (2) (blue line (circles) and solid red line (squares) respectively) for; A-All voxels; B-Top 1% FLAIR sub-sampling; and C-SSCC sub-sampling. The dashed line in all curves corresponds to having TPR=FPR.

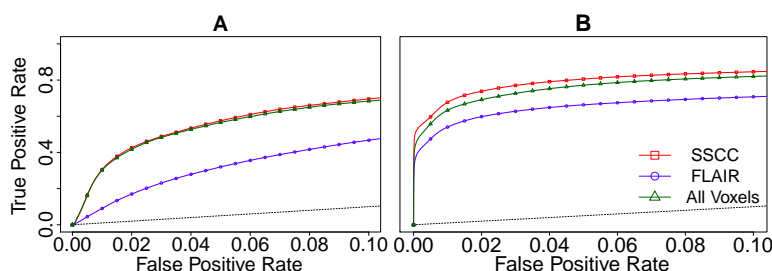


Figure 2. Partial ROCs curves from using the three data sets under study to fit model (1) (A) and model (2) (B).

Table 1. Partial AUC estimates and corresponding 95% confidence intervals (CI) for curves in Figures 1 and 2.

	Model 1		Model 2	
	pAUC	95% CI	pAUC	95% CI
All Voxels	0.52	(0.49,0.55)	0.74	(0.72,0.76)
1% FLAIR	0.29	(0.27,0.32)	0.64	(0.61,0.67)
SSCC	0.53	(0.50,0.56)	0.78	(0.76,0.80)

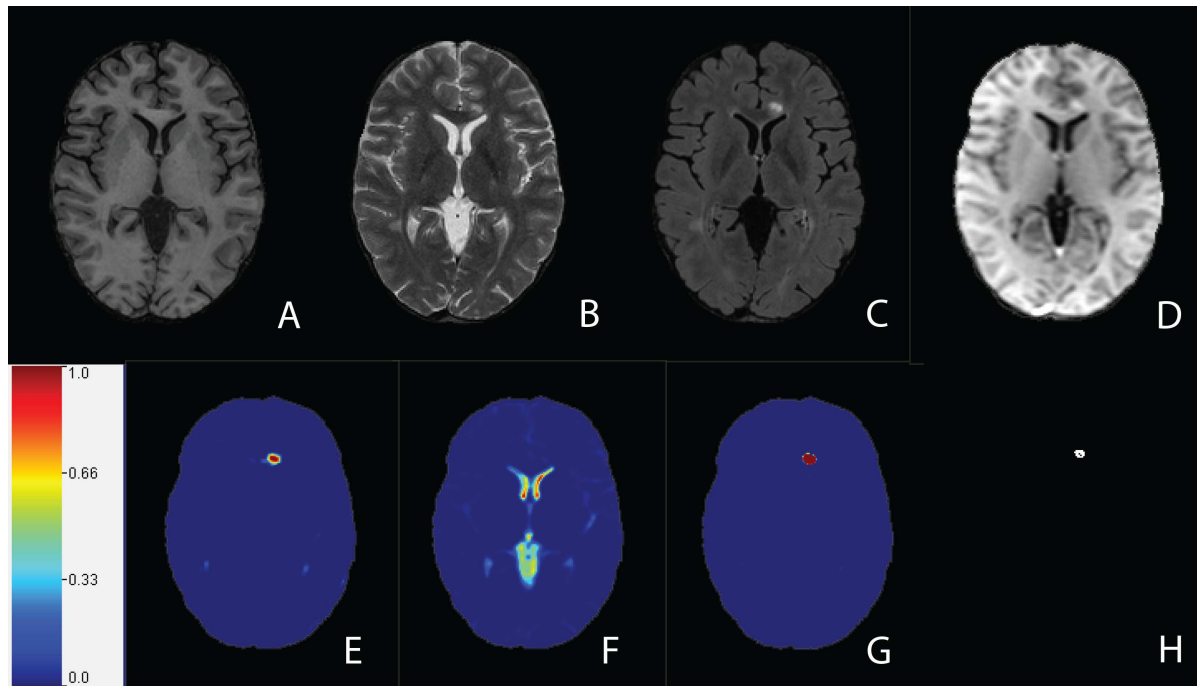


Figure 3. (A) T1w pre-contrast; (B) T2w pre-contrast; (C) FLAIR post-contrast; (D) T1w post-contrast; (E) SuBLIME Map ; (F) model 1 prediction map using FLAIR thresholding (top 1%) to fit and predicting on the full brain, as per Shinohara et al.(2012) ; (G) model 2 prediction map with case control sampling for fitting and predicting on the full data set; (H) RGS mask.

Table 2. Coefficient estimates when model 1 is fit using all voxels in the corresponding sets.

All Voxels		
	Estimate	95% CI
β_0	-10.18	(-10.21 , -10.14)
β_1	1.08	(1.03 , 1.13)
β_2	1.98	(1.96 , 2.00)
β_{12}	0.88	(0.85 , 0.90)
FLAIR 1%		
	Estimate	95% CI
β_0	-7.81	(-7.85 , -7.77)
β_1	0.01	(-0.06 , 0.09)
β_2	2.00	(1.98 , 2.04)
β_{12}	0.52	(0.48 , 0.56)
SSCC		
	Estimate	95% CI
β_0	-10.79	(-10.85 , -10.74)
β_1	2.02	(1.93 , 2.11)
β_2	2.88	(2.82 , 2.94)
β_{12}	1.32	(1.27 , 1.37)

Table 3. Coefficient estimates when model 2 is fit using all voxels in the corresponding sets.

All Voxels		
	Estimate	95% CI
β_0	-10.23	(-10.27 , -10.20)
β_1	0.814	(0.76 , 0.87)
β_2	1.53	(1.50 , 1.56)
β_3	10.03	(9.94 , 10.12)
β_{12}	0.95	(0.92 , 0.98)
FLAIR 1%		
	Estimate	95% CI
β_0	-7.65	(-7.69 , -7.60)
β_1	-0.51	(-0.59 , -0.43)
β_2	1.40	(1.37 , 1.44)
β_3	6.74	(6.64 , 6.83)
β_{12}	0.97	(0.92 , 1.02)
SSCC		
	Estimate	95% CI
β_0	-11.42	(-11.49 , -11.35)
β_1	1.94	(1.83 , 2.06)
β_2	2.53	(2.46 , 2.60)
β_3	37.09	(34.93 , 39.27)
β_{12}	1.09	(1.03 , 1.15)

Acknowledgements

Pomann's research is supported by the National Science Foundation under Grant No. DGE-0946818. Pomann, Shinohara, Staicu, and Sweeney are partially funded by the NIH grant RO1 NS085211 from the National Institute of Neurological Disorders and Stroke. The study was supported in part by the Intramural Research Program of NINDS. This work represents the opinions of the researchers and not necessarily that of the granting organizations.

References

- Goodin DS (ed.). *Handbook of Clinical Neurology*, vol. 122. Elsevier B.V., 2014.
- McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, *et al.*. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology* 2001; **50**(1):121–127.
- Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, Lublin FD, Metz LM, McFarland HF, O'Connor PW, *et al.*. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "mcdonald criteria". *Annals of Neurology* 2005; **58**(6):840–846.
- Medicare. Physician fee schedule. centers for medicare and medicaid services. <http://www.cms.gov/apps/physician-fee-schedule/license-agreement.aspx> 2014. URL <http://www.cms.gov/apps/physician-fee-schedule/license-agreement.aspx>
- Moreno-Romero J, Segura S, Mascaró J, Cowper S, Julia M, Poch E, Botey A, Herrero C. Nephrogenic systemic fibrosis: a case series suggesting gadolinium as a possible aetiological factor. *British Journal of Dermatology* 2007; **157**(4):783–787.
- Kay J. Nephrogenic systemic fibrosis: a gadolinium-associated fibrosing disorder in patients with renal dysfunction. *Annals of the rheumatic diseases* 2008; **67**(Suppl 3):iii66–iii69.
- Perazella MA. Current status of gadolinium toxicity in patients with kidney disease. *Clinical Journal of the American Society of Nephrology* 2009; **4**(2):461–469.
- Simon J, Li D, Traboulsee A, Coyle P, Arnold D, Barkhof F, Frank J, Grossman R, Paty D, Radue E, *et al.*. Standardized mr imaging protocol for multiple sclerosis: Consortium of ms centers consensus guidelines. *American Journal of Neuroradiology* 2006; **27**(2):455–461.
- Garcia-Lorenzo D, Prima S, Collins DL, Arnold DL, Morrissey SP, Barillot C, *et al.*. Combining robust expectation maximization and mean shift algorithms for multiple sclerosis brain segmentation. *MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (validation and*

- methodological issues*(MIAMS'2008), 2008; 82–91.
10. Llado X, Ganiler O, Oliver A, Marti R, Freixenet J, Valls L, Vilanova JC, Ramio-Torrenta L, Rovira A. Automated detection of multiple sclerosis lesions in serial brain mri. *Neuroradiology* 2012; **54**(8):787–807.
 11. Llado X, Oliver A, Cabezas M, Freixenet J, Vilanova JC, Quiles A, Valls L, Ramio-Torrenta L, Rovira A. Segmentation of multiple sclerosis lesions in brain mri: a review of automated approaches. *Information Sciences* 2012; **186**(1):164–185.
 12. Gaitan M, P S, and Reich DS IS. Initial investigation of the blood-brain barrier in ms lesions at 7 tesla. *Multiple Sclerosis Journal* 2012; **19**:1068–1073.
 13. Absinta M, Sati P, Gaitan MI, Maggi P, Cortese ICM, Filippi M, Reich DS. Seven-tesla phase imaging of acute multiple sclerosis lesions: A new window into the inflammatory process. *Annals of Neurology* 2013; **74**(5):669–678, doi:10.1002/ana.23959. URL <http://dx.doi.org/10.1002/ana.23959>.
 14. Shinohara R, Goldsmith J, Mateen F, Crainiceanu C, Reich D. Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents. *American Journal of Neuroradiology* 2012; **33**(8):1586–1590, doi:10.3174/ajnr.A2997. URL <http://www.ajnr.org/content/early/2012/03/22/ajnr.A2997.abstract>.
 15. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika* 1979; **66**(3):403–411, doi:10.1093/biomet/66.3.403. URL <http://biomet.oxfordjournals.org/content/66/3/403.abstract>.
 16. King G, Zeng L. Logistic regression in rare events data. *Political Analysis* 2001; **9**(2):137–163. URL <http://pan.oxfordjournals.org/content/9/2/137.abstract>.
 17. Gaitan MI, Shea CD, Evangelou IE, Stone RD, Fenton KM, Bielekova B, Massacesi L, Reich DS. Evolution of the blood–brain barrier in newly forming multiple sclerosis lesions. *Annals of neurology* 2011; **70**(1):22–29.
 18. Sweeney EM, TShinohara R, Shea CD, Reich DS, Crainiceanu CM. Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal mri. *American Journal of Neuroradiology* 2013; **34**(1):68–73.
 19. Shinohara RT, Crainiceanu CM, Caffo BS, Gaitán MI, Reich DS. Population-wide principal component-based quantification of blood–brain-barrier dynamics in multiple sclerosis. *NeuroImage* 2011; **57**(4):1430–1446.
 20. Sweeney EM, Shinohara RT, Shiee N, Mateen FJ, Chudgar AA, Cuzzocreo JL, Calabresi PA, Pham DL, Reich DS, Crainiceanu CM. Oasis is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in mri. *NeuroImage: clinical* 2013; **2**:402–413.
 21. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, Jarso S, Pham DL, Reich DS, Crainiceanu CM, *et al.*. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical* 2014; **6**:9–19.
 22. Imbens GW. An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica: Journal of the Econometric Society* 1992; **60**:1187–1214.
 23. Cosslett SR. Maximum likelihood estimator for choice-based samples. *Econometrica: Journal of the Econometric Society* 1981; **49**:1289–1316.
 24. Imbens GW, Lancaster T. Efficient estimation and stratified sampling. *Journal of Econometrics* 1996; **74**(2):289–318.
 25. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Statistics in Medicine* 2013; **32**(9):1467–1482.