

# Intracranial Hemorrhage Localization in a Population of Patients using Registration-based Techniques in CT Imaging

John Muschelli<sup>1,\*</sup>, Natalie L. Ullman<sup>2</sup>, Elizabeth M. Sweeney<sup>3</sup>, Ani Eloyan<sup>4</sup>, Daniel F. Hanley<sup>5</sup>,  
Ciprian M. Crainiceanu<sup>6</sup>

**1 John Muschelli Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA**

**2 Natalie L. Ullman Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA**

**3 Elizabeth M. Sweeney, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA**

**4 Ani Eloyan, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA**

**5 Daniel F. Hanley Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA**

**6 Ciprian M. Crainiceanu Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA**

**\* E-mail: Corresponding jmusche1@jhu.edu**

## Abstract

NA

**Keywords:** Intracranial Hemorrhage; CT Imaging Analysis; 3D Histograms;

## 1 Introduction

Intracranial hemorrhage (ICH) is a neurological condition that results from a blood vessel rupturing into tissues and possibly the ventricles of the brain. Bleeding may cause distension of the brain structures and increase in potentially lethal intracranial pressure (ICP). ICH is a serious condition; it accounts for approximately 10-15% of all strokes, corresponding to an estimated 79,500 annual cases [15] and approximately 30,000 deaths [34] in the US and approximately 5 million cases worldwide [20]. In addition to the increased likelihood of death, ICH has debilitating health effects on survivors who do not have full functional recovery after stroke.

The use of computed tomography (CT) scans allows clinicians and researchers to qualitatively and quantitatively describe the characteristics of a hemorrhage to guide interventions and treatments. CT scanning is widely available and is the most commonly used diagnostic tool in patients with ICH [39]. While much can be understood from careful CT image inspection, detailed quantification of information at the patient and population level is an open problem. Indeed, it is hard, for example, to quantify from visual inspection alone the percent of a particular anatomic ROI engaged by the stroke, compare locations of strokes across patients, and study the association between specific locations and adverse health outcomes.

To address these problems, we will 1) develop an analytic pipeline based on available neuroimaging tools to register patient CT scans to a CT template; 2) create a 3-dimensional (3D) density map of hemorrhages occurring in a population of patients who survived initial ICH and were enrolled in the MISTIE and ICES clinical trials; 3) provide a detailed quantification of hemorrhage engagement of individual neuroanatomic regions within the brain; 4) determine if differences in location relate to two common disability scores: the National Institutes of Health Stroke Scale (NIHSS) [6] and the modified Glasgow Coma Scale (GCS) [45, 46]; and 5) generate a stroke region of engagement that is likely to be associated with adverse health effects and test its predictive performance using within-sample validation.

It is a matter of debate whether the location of the ICH is relevant to functional outcome in patients who survived to make it to the hospital [7]. For example, Hemphill et al. [17] found that intratentorial ICH location, intraventricular involvement, and ICH volume, which may represent a number of anatomic locations, all predict 30-day mortality. Conversely, Chuang et al. [9], and Diringer, Edwards, and Zazulia [12] found that the site of ICH was associated with 30-day mortality, but was not statistically associated after adjusting for other covariates. Following the result that hydrocephalus was an independent predictor of 30-day mortality [12], Phan et al. [30] found that hydrocephalus was an independent prognostic indicator only for the group with hemorrhages in the putamen and not for those with hemorrhages in the caudate or thalamus. In contrast, other studies have not observed location of the ICH predicting prognosis [5, 8, 11, 16, 22, 23, 31, 33, 36, 40].

Other trials have estimated the effects of surgery in a sub-population of ICH patients on health outcomes. Phase I of the Surgical Trial in Intracerebral Haemorrhage (STICH I) found treatment effects of hemorrhage reduction through early surgery in a subset of patients who had bleeds in the proximity of the brain lobes (1 cm from the cortical surface of the brain) [24]. However, when the effect was re-examined in the second phase of the trial (STICH II), which included only patients with superficial lobar hematomas, the treatment effect was not found to be statistically significant [25]. Although STICH II did not identify a statistically significant treatment effect, this may be due to sub-optimal definition of localization. In the “Guidelines for the Management of Spontaneous Intracerebral Hemorrhage” Morgenstern et al. [26] noted that: “other randomized trials have had too few patients to determine outcomes in subgroups by location, randomized only patients with deep ICH, or did not report these results”. Moreover, the localization of ICH and its evolution before and after treatment administration may actually be a crucial baseline or longitudinal characteristic of the stroke that could predict whether treatments are effective, may be a confounding factor of treatment and outcomes, and may relate to ICH recurrence [13].

However, even the coarse classification of hemorrhage location is complicated, which may lead to labeling confusion even among the best trained neuroimage scientists. Indeed, a hemorrhage may extend into multiple brain areas, may affect different tissue classes, and may break through the ventricular wall causing the hemorrhage to spread to other areas of the brain. Such problems would challenge even the best scientists, as it becomes time-onerous to do anything more than calling a single region as the primary affected anatomic region (e.g. caudate or putamen) or describing the edge of a hemorrhage by an average or minimum distance to a given landmark [49]. Also, increased long- or short-term disability that corresponds to a significant difference in hemorrhage volume may be modulated by hemorrhage location. Alternatively, hemorrhage increases after a stabilization period may result in no changes in functional indices depending on the brain regions involved. Hence, there is a need for detailed information of brain hemorrhage location and its association with the stroke severity, while controlling for ICH volume. Such detailed localization information can be obtained by registering scans to a common template and by using the detailed anatomical information in the template to provide refined anatomical localization information.

Previous CT neuroimaging studies have not used registration-to-template-based methods for detailed location quantification, most likely due to a lack of a CT template brain. Registration to template space is a crucial first step for the type of analysis we are describing here; all patient scans are located in the same stereotaxic space so information may be combined across scans. Recently, Rorden et al. [38] published the first CT template of healthy adults in its MNI (Montreal Neurological Institute) space representation. Along with this template, Rorden et al. [38] released software for CT image registration to the template. This software provides an important analytic step forward compared to what was previously available. Indeed, previous studies have used average CT images as a template, but the patient population consisted of people affected by a specific neurological condition (cerebrovascular infarcts) and were not in MNI space [19]. In addition, many other studies have co-registered CT images to concurrent (or temporally proximal) MRI images from the same patient [4, 10, 14, 29]. This registration process is useful when structural MRIs are concurrently acquired with CT images. In the clinical trials we investigate here the

CT images have always been acquired according to the trial protocol, whereas MRI sequences were only acquired according to the local practices, which varied quite a bit by modality and frequency. Therefore, a CT-only pipeline provides the necessary flexibility to conduct the type of analyses we address here. Other methods have used CT image registration directly to an MRI template, but using a CT template provides a simpler registration as the reference and template images are the same modality [21, 32, 43].

Most importantly, this type of detailed mapping of strokes has never been done in the context of the highly visible MISTIE and ICES stroke clinical trials. MISTIE is the Minimally Invasive Surgery plus recombinant-tissue plasminogen activator (rtPA) for Intracerebral Evacuation trial and ICES is the Intraoperative CT-Guided Endoscopic Surgery for ICH trial. MISTIE was a phase II clinical trial with the goal of determining the safety, efficacy, and long-term effects of treatment of ICH using thrombolytics. ICES was a simultaneous phase I trial with the same inclusion/exclusion criteria as MISTIE and the goal to determine the effect of endoscopic surgery as opposed to thrombolytic therapy for hematoma removal. We are not focusing on treatment effects; instead we are mapping stroke locations and investigate the associations between location and stroke scores before treatment was administered. For this purpose we will use only scans, demographic and functional measures collected prior to randomization.

## 2 Materials and Methods

All statistical analysis was done in the R statistical programming language (<http://cran.r-project.org/>). We used the `oro.nifti` package [47] for reading and visualizing brain images and the `ggplot2` package [48] to visualize histograms.

### 2.1 Subjects and Demographics

The study consists of 111 patients from MISTIE ( $N = 94$ ) and ICES ( $N = 17$ ) recruited from 26 centers. Patients were eligible if: they were aged 18 to 80 years old with a spontaneous ICH of greater than 20cc that did not require an extraventricular drain (EVD), were non-pregnant, had a historical Rankin Scale [35, 44] score of 1 and a diagnosis that excluded other conditions such as ruptured aneurysms, arteriovenous malformations, or other neurovascular anomaly (for full criteria, see Mould et al. [27]). CT and clinical data were previously collected as part of the Johns Hopkins Medicine IRB-approved ICES and MISTIE research studies with written consent provided by participants. Each site adhered to local ethical procedures and consented to participation in the study. All clinical data were deidentified by using a unique patient identifier and anonymizing CT images by removing all patient-identifying information.

The NIHSS and GCS scores were recorded at enrollment into the study; 3 patients did not have a recorded NIHSS score, 1 patient did not have a recorded GCS score, and these patients were excluded from the analyses associating NIHSS and GCS scores and location, respectively. All patients were included in the construction of the 3D histogram image. Histograms of NIHSS score, GCS score, age, and baseline ICH volume are shown in Figure 1, and the descriptive statistics of age, sex, NIHSS and GCS scores are shown in Table 1. These measures reflect different aspects of a patient’s health: the NIHSS score reflects a stroke impact (higher is worse), while the GCS score reflects a patient’s level of consciousness (higher is better). The NIHSS scores indicate large heterogeneity of disease burden, with a range between 2 (no stroke symptoms) to 41 (severe stroke). This heterogeneity is similarly reflected in the GCS score distribution with a range of 3 (deep unconsciousness) to 15 (fully awake person). The age of patients is between 33 and 80 years, with a mean of 60.8 years. Women comprise 31.5% of the dataset.

[Figure 1 here.]

[Table 1 here.]

## 2.2 Imaging Data

CT images were acquired for all patients, with site-specific scanning parameters. Scans were acquired using GE ( $N = 46$ ), Siemens ( $N = 37$ ), Philips ( $N = 20$ ), and Toshiba ( $N = 8$ ) scanners. Gantry tilt was observed in 87 scans. Slice thickness of the image varied within the scan reconstruction for 14 scans, which we will refer to as variable slice thickness. For example, if a scan was reconstructed with 10 millimeter (mm) slices at the top and bottom of the brain, which are not of interest, but with 5mm slices in the middle where the hematoma is seen (see Figure 3a). Therefore, the scans analyzed had different voxel dimensions and image resolution prior to registration to the template. Different reconstructions of CT images are not available via the data-acquiring center, and represent how scans are presented for evaluation in many diagnostic cases.

## 2.3 Hemorrhage Segmentation and Location Identification

To register the CT scan to the CT template, the hemorrhage was excluded from the algorithm, i.e. “masked out”, using manual ICH segmentations. ICH was manually segmented on CT scans using the OsiriX imaging software by expert readers (OsiriX v. 4.1, Pixmeo; Geneva, Switzerland). Readers employed a semiautomated threshold-based approach using a Hounsfield unit (HU) range of 40 to 80 to select potential regions of ICH [3, 41]; these regions were then further quality controlled and refined by readers using direct inspection of images. Binary image masks were created for the hemorrhage region of interest (ROI) by setting voxel intensity to 1 if the voxel was classified as hemorrhage, and 0 otherwise. Readers also identified the region most engaged by the ICH before creating the binary image of the hemorrhage.

## 2.4 Image Processing: Brain Extraction, Reorientation, Registration

CT brain images and the binary mask were exported from OsiriX to DICOM (Digital Imaging and Communications in Medicine) format. Images with gantry tilt were corrected using a customized MATLAB (The Mathworks, Natick, Massachusetts, USA) user-written script (<http://bit.ly/1ltIM8c>). Images were converted to the Neuroimaging Informatics Technology Initiative (NIfTI) data format using `dcm2nii` (2009 version, provided with MRICro [37]). Images were constrained to values  $-1024$  and  $3071$  HU to remove potential image rescaling errors and artifacts. No interpolation was done for images with a variable slice thickness. Thickness was determined from the first slice converted and was assumed homogeneous throughout.

Brains were extracted to remove skull, eyes, facial and nasal features, extracranial skin, and more importantly non-human elements of the image captured by the CT scanner, such as the gantry, pillows, or medical devices. Using a variant of a previously published brain extraction protocol [38], removal of these elements was performed using the brain extraction tool (BET) [42], a function of the FSL [18] neuroimaging software (v5.0.4). Images were thresholded to a brain tissue range (0-100 HU) and BET was applied, using a fractional intensity threshold of 0.01. An image with its brain-extracted counterpart can be seen in Figure 2.

In order to use statistical parametric mapping (version 8, SPM8, Wellcome Trust Centre for Neuroimaging, London, United Kingdom) spatial registration algorithm, images must be centered at the line between anterior commissure (AC) and posterior commissure (PC). The AC-PC line was automatically estimated from the brain-extracted image using a custom MATLAB script (<http://bit.ly/1gUqMDw>), which provides the transformation matrix. The non-brain-extracted image and binary mask were then re-centered using this transformation matrix. The binary hemorrhage mask was smoothed using a Gaussian smoother with a full width at half maximum (FWHM) of 3mm. The brain image was masked by the inverse of the smoothed hemorrhage mask before registration.

The image was then spatially registered to the CT template using the Clinical toolbox [38], which employs SPM8’s unified normalization-segmentation routine [2] in MATLAB. The smoothed hemorrhage mask was transformed into the template space using the estimated registration transformation. The registered smoothed hemorrhage mask was then thresholded using the following rule:

$$v_{ij} \frac{\min(I_i) + \max(I_i)}{2}$$

where  $v_{ij}$  denotes voxel  $j$  for smoothed mask for person  $i$ , denoted  $I_i$ . This corresponds to voxels in the template space being classified as ICH if the transformed voxel on the transformed smoothed mask was larger than the midrange of the transformed smoothed mask.

## 2.5 Assessing Registration

Determining the quality of registration was done qualitatively by visual inspection by expert CT readers. We acknowledge that visual inspection is subjective, but currently no automatic computer-based gold standard exists for assessing registration results and analysis of registration quality is not commonly done with large deformations of tissue as in patients with ICH.

No scans were excluded due to inadequate registration.

## 2.6 Histograms of ICH in the Brain

To visualize and describe the localization of lesions, we first combined information from registered masks of 111 patients. Using the registered ICH masks we obtained the 3-dimensional (3D) histogram of the ICH localization for the study population. More precisely, for every voxel location in the template space, we calculated the proportion of patients who have an ICH at that particular voxel. In addition to displaying results at pre-selected slices of the 3D volume, we used the R package `brainR` to create an 3D interactive map of the 3D histogram. The interactive map is located at [http://muschellij2.github.io/CT\\_Pipeline/index.html](http://muschellij2.github.io/CT_Pipeline/index.html).

## 2.7 Prediction of Functional Score Based on Hemorrhage Location

There is large scientific interest in finding and quantifying the effect size of possible associations between ICH location and functional scores. We evaluated the association between hemorrhage location and stroke severity as measured by the total NIHSS score and GCS score using a series of models that account for increasing levels of confounder adjustment. In the population, patients had an ICH at a total of 1045174 voxels; a voxel in the template was counted if it was part of the ICH of any patient. We limited our analysis to voxels in the template space where at least 10 patients exhibit lesions to reduce the problems with low prevalence regression in moderate sample sizes. This reduced the number of voxels by roughly an order of magnitude to 166202. At each one of these voxels, we ran a linear regression model:

$$Y_i = \beta_0 + \beta_1(v)ICH_i(v) + X_i^t\gamma + \epsilon_{iv}, \quad (1)$$

where  $Y_i$  was either the NIHSS or GCS score for patient  $i = 1, \dots, 111$ ,  $ICH_i(v)$  is a binary indicator where  $ICH_i(v) = 1$  if patient  $i$ ’s ICH mask has a 1 at voxel  $v$  of the template space and  $ICH_i(v) = 0$  otherwise,  $X_i^t$  is a vector of patient-specific confounders with effects  $\gamma$  and  $\epsilon_{iv}$  are treated as independent homoscedastic errors. The vector of confounders,  $X_i$ , either was excluded for “unadjusted” voxel-level models, or contains a combination of 3 patient-specific confounders, age, sex, and total baseline ICH volume (TICHVol) in an “adjusted” model. We used the unadjusted Wilcoxon rank-sum test on  $Y$  at every voxel to confirm that results are robust to the choice of test statistic.

P-values of each voxel-wise regression model of the functional scores on ICH presence were calculated, testing the null hypothesis  $H_{0,v} : \beta_1(v) = 0$ —or in the case of the Wilcoxon rank-sum test:  $H_{0,v} :$

$Score\{ICH(v) = 1\} = Score\{ICH(v) = 0\}$  where  $Score$  denotes the distribution of functional scores of patients that either did or did not have ICH at the given voxel  $v$ . Figures 5 and 7 display the pointwise p-values. The p-value images displayed were not corrected as the goal is to see potential “hot spots” of ICH that relate to functional outcome.

We corrected the p-values, within each analysis of functional scores, using a Bonferroni correction with  $\alpha = 0.05$ . Some of the voxels were identical for presence for ICH over the population, e.g. voxels 1 and 2 each had the same people having ICH and same people not having ICH at that location. We excluded these from the correction, using only then number of unique tests ( $N = 48775$ ).

### 2.7.1 Region of Interest Generation and Analysis

In order to create a patient-level covariate that summarizes ICH location information, we created a sequence of nested regions of interest (ROIs) by selecting voxels based on the smallest p-values obtained from the undadjusted linear model, i.e. from a model where only the voxel-level ICH indicator was used, but none of the other covariates. We have obtained 6 different ROIs, 3 based on the smallest 1000, 2000, and 3000 lowest p-values and three based on all p-values below a particular threshold. We have used three thresholds equal to 0.05, 0.01, and 0.001, resulting in ROI with 2422, 19047, and 47736 voxels respectively.

We thresholded the image based on a range of p-values (0.05, 0.01, 0.001) and by taking a fixed number of voxels, ranked lowest to highest on p-values (1000, 2000, 3000 voxels). For each ROI, we have introduced a variable called “coverage” that represents the percentage of the voxels that were classified as hemorrhage in the subject-specific image:

$$\text{Coverage} = \frac{\# \text{ Voxels classified ICH in ROI}}{\# \text{ Voxels in ROI}} \quad (2)$$

For example, if a patient’s hemorrhage covers the entire ROI, the coverage is 100%, whereas if there is no overlap between the patient’s scan and ROI then coverage is 0%. This subject-specific covariate is then used as a predictor of the severity score in an adjusted model as

$$Y_i = \beta_0 + \beta_1 \text{Coverage}_i + \gamma_1 \text{Age}_i + \gamma_2 \text{Gender}_i + \gamma_3 \text{TICHVol}_i + \epsilon_i$$

We have further investigated whether the new variable coverage was an independent predictor of the severity score compared to accounting for the expert-specified ICH location classification. Specifically, we have used a categorical indicator of the expert-specified ICH location classification, with categories: Thalamus ( $N = 4$ ), Globus Pallidus ( $N = 6$ ), Lobar ( $N = 33$ ), and Putamen ( $N = 68$ ). Prediction performance and model fit were assessed using a series of model-fit measures:  $R^2$ , adjusted  $R^2$ , Akaike information criterion (AIC) [1], and root mean squared error (RMSE).

## 2.8 ICH Localization and Engagement

As each patient’s ICH mask is in MNI space, we can use previously segmented and labeled brain atlases to label spatial ICH engagement. We used the “Eve” atlas [28] because it segments and labels both gray matter (GM) and white matter (WM) regions. Ventricular regions are not explicitly segmented and are categorized as background. Thus, our labels are slightly modified by categorizing any region within the brain mask labeled as background in the Eve atlas as cerebrospinal fluid (CSF), though this may encompass a small number of voxels that are not CSF.

Registration to a carefully labeled template allowed the detailed localization of the ICH at the subject level. In particular, for each subject we obtain and report: 1) the percent of the ICH engaged by each labeled region of the EVE atlas; 2) the percent of each labeled region of the EVE atlas engaged by the ICH, and 3) the total volume of each region engaged by the ICH. These summaries of ICH engagement

provide a much finer scale description than what can currently be done by expert human readers. Once summaries are available, further simplifications are easy to obtain. Indeed, it is easy to provide the labeled region with the highest percentage of ICH engagement or the region containing the largest percentage of the ICH. We did not directly compare reader-classified regions and the most-engaged region, as the template did not directly segment into the same categories used by readers.

These results were further summarized at the population level by calculating the percentage of engagement of the 3D population histogram image, weighted by the proportion of patients with ICH at that location. More precisely, for every voxel in a labeled region we have calculated the proportion of subjects who have an ICH at that voxel. These proportions were then summed over the entire labeled region and this sum was divided by the total sum of the 3D population histogram image. We also calculated the percent engagement of regions for the best-performing ROIs for the NIHSS and GCS score analyses to characterize potential locations relating to these functional outcomes.

### 3 Results

#### 3.1 Image Registration

To illustrate registration results, the registered images for 2 patients are shown in Figure 3 with overlays (in pink) of the manually segmented blood in native space (panels a and d), the registered image and mask in template space (panels b and e), and the overlaid mask onto the template (panels c and f). Slices are shown at the centroid (average location) of the hemorrhage mask. We chose one patient with variable slice thickness with a large hemorrhage (Figure 3a, b, and c) and one patient with a small hemorrhage (Figure 3d, e, and f). Images indicate that gross brain features remain relatively unchanged, but large deformations of tissue, mainly due to ICH, seem well preserved by registration. Also, the registered brain in the patient with variable slice thickness (image looks distorted in native space) is the same as the template, which indicates that the non-linear registration seems to reasonably account for variable slice thickness, most likely by non-uniform scaling.

[Figure 3 here.]

#### 3.2 Histograms of ICH in the Brain

The histogram of the prevalence of hemorrhages over voxels (Figure 4a) shows the majority of voxels have a low prevalence; the median proportion of ICH at a given voxel is 3% ( $N = 3$ ), though some voxels ( $V = 5685$ ) have a high prevalence ( $> 40\%$  of the sample).

The resulting 3D histogram of the density of hemorrhages is shown in Figure 4b. In this image, voxels colored in shades of blue correspond to a lower proportion of patients with hemorrhage, whereas those colored in shades of red correspond to a higher proportion. This histogram indicates that in this cohort lesions are distributed medially in the brain in this cohort, with a lower concentration close to the cortical surface and higher on the left side of the brain. Combining areas of engagement regardless of the side of the brain may be worthwhile, though combination of these areas may not be straightforward for those by ICH that cross the mid-sagittal plane. Figure 4b indicates that the prevalence of strokes in the extreme anterior and posterior areas of the brain are very low. Nor do we see a high prevalence of bleeds in inferior areas of the brain, which is likely due to brain stems bleeds being excluded using the trial inclusion criteria.

These observations of ICH being centralized in the middle portion of the brain may also be an artifact of the inclusion criteria and may not generalize to other populations, but we note that no spatial location was preferred for inclusion. Regardless, creating these population images actually allows us to test if this location density does generalize to other populations given their density map.

[Figure 4 here.]

### 3.3 Prediction of Functional Score Based on Hemorrhage Location

We ran several versions of model (1) starting with the simplest case and then building the model by incorporating different choices of potential confounders to investigate their effects on the statistically significant association between ICH locations and functional score. More precisely, the five linear models are:

$$\begin{aligned}\mathcal{M}_1 : & Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \epsilon_{iv}, \\ \mathcal{M}_2 : & Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \gamma_1(v)Age_i + \epsilon_{iv}, \\ \mathcal{M}_3 : & Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \gamma_2(v)Gender_i + \epsilon_{iv}, \\ \mathcal{M}_4 : & Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \gamma_3(v)TICHVol_i + \epsilon_{iv}, \\ \mathcal{M}_5 : & Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \gamma_1(v)Age_i + \gamma_2(v)Gender_i + \gamma_3(v)TICHVol_i + \epsilon_{iv},\end{aligned}$$

where  $\mathcal{M}$  denotes a model. Also, we'll denote  $\mathcal{M}_6$  as the voxel-wise Wilcoxon rank-sum test on functional score.

No voxels were significantly related to NIHSS score or GCS score in any model using the Bonferroni correction. We present maps of p-values for the models with NIHSS score as the outcome to potential areas of interest, as the Bonferroni correction is conservative and information can be gained at looking at the location of high and low p-values.

The p-values for each model, with NIHSS score as the outcome, are presented in Figure 5, overlaid on a MRI T1 image for spatial localization of structures. The p-values are colored with cooler (blue) colors representing higher p-values and hotter (red) colors representing lower p-values. Colors correspond to p-values .1 – 1 (dark blue), .1 – .05, (light blue), .05 – .01 (green), .01 – .001 (yellow), .001 – .0001 (orange) < .0001 (red). We see that there may be some relationship to medial hemorrhages with NIHSS score seen in the unadjusted model  $\mathcal{M}_1$  (Figure 5a) and the Wilcoxon rank-sum test (Figure 5f) and observe this when modeling the GCS score (Supplemental Figure 7). The p-values are higher in the models adjusted for age (Figure 5b), baseline ICH volume (Figure 5d), and both age and baseline ICH volume (Figure 5e).

[Figure 5 here.]

#### 3.3.1 Region of Interest Analysis

Figures 6a and 6b illustrate the voxels (in red) that were selected as the ROI based on p-value or ranking of p-value for NIHSS and GCS score as the functional outcome, respectively. We observe that the ROIs appears larger on the left side compared to the right and many are near the ventricles, which may indicate ICH at these regions is important to functional score. We note that the ROI has spatial contiguity on either side of the brain, which is expected as the presence of ICH at a given voxel is highly correlated to that voxels neighbors. The number of voxels ( $V$ ) for each p-value threshold (not displayed) were: .05 ( $V = 47736$ ), .01 ( $V = 19047$ ), .001 ( $V = 2422$ ), which can correspond to rather large regions As the resolution of the template is  $1\text{mm}^3$ , this corresponds to the ROI for a 0.05 p-value threshold measuring 48ml.

The corresponding number of voxels for p-value thresholds, using GCS as the functional outcome are: .05 ( $V = 52368$ ), .01 ( $V = 22858$ ), .001 ( $V = 4669$ ), which again correspond to large ROIs for higher p-value thresholds.

Panels a and b of figure 6 show the unadjusted relationship of the percent ROI coverage with the functional score. We observe that the relationships are as expected, the larger the ROI coverage the higher (more severe stroke) NIHSS score and the lower (deeper unconsciousness) the GCS score.

[Figure 6 here.]

We present how these regions perform against using a location classification of ICH determined by readers. We summarized the model-fit measures for in models for the NIHSS score (Table 2) and GCS score (Table 3). We have subtracted the AIC scores by the minimum value as these measures are unit-less; the model with a value of 0 has the best (lowest) value for AIC.



We see that for NIHSS score, the ROI constructed using a p-value threshold of 0.0100 corresponded to the best model based on all our model-fit measures. For the GCS score, the best model was one using an ROI using the voxels with lowest 1000 p-values. With respect to comparing model-fit measures within models using the ROI coverage, all models due comparably well, defined as less than 5 unit difference in AIC, except for the ROI coverage model for ROI based on a p-value threshold of .05 modeling GCS score.

Though large differences in model-fit measures using the ROI coverage are not observed, there is significant gain compared to using the reader classification with respect to  $R^2$ . For the NIHSS models, the  $R^2$  was 0.178 for the reader-classified location model compared to 0.282 for the best ROI coverage model, corresponding to a 58% relative increase. The result is similar when modeling GCS score: reader-classified location model  $R^2 = 0.12$  vs best ROI coverage model  $R^2 = 0.243$ , corresponding to a 102% relative increase. Even though  $R^2$  is relatively low for these models in absolute terms, the main goal of this ROI analysis was not prediction or model fit but rather comparison. We see that if these are potential goals for researchers, then using a CT-based measure can be potentially more informative than a reader-based categorical location predictor.

We observe larger gains for the ROI coverage model using adjusted  $R^2$ , which penalizes the reader-classified locations more than the ROI coverage model due to the additional predictors. Similarly, AIC would not choose the reader-classified location model over any model using ROI coverage. Using the RMSE as a criteria for model selection, the ROI coverage model outperform the reader-classified location model.

[Tables 2 and 3 here]

Table 4 represents the beta coefficients from the models using reader-assessed location and the ROI coverage model with the best model-fit measures. The models for ROI coverage represent the best model based on the model-fit measures. We see that after adjusting for age, sex, and total baseline ICH volume, increasing 10% coverage is expected to increase NIHSS score by 2.0 (95% CI: 1.1, 2.8) points. We see that all locations, compared to lobar hemorrhages have higher estimated NIHSS scores, but putaminal hemorrhages were significantly higher by 4.2 (95% CI: 0.2, 8.2) points. Similarly for GCS score, adjusting for other covariates, increasing 10% coverage is expected to decrease GCS score by 0.4 (95% CI: 0.6, 0.2) points. All locations compared to lobar hemorrhages have lower estimated GCS scores, but none were statistically lower.

[Table 4 here]

### 3.3.2 ICH Localization and Engagement

Table 5 represents the 10 most-engaged regions for the population 3D histogram as well as the ROIs for the GCS and NIHSS score ROI analyses.

[Table 5 here]

## 4 Discussion

### 4.1 Semi-Automated Localization of ICH

We have characterized the localization of ICH in a population of stroke patients within this processing pipeline. We would like to stress that this is, as far as we know, the first 3D histogram of a population with ICH. We found that more patients had ICH engagement on the left side of the brain compared to right and both left and right sided ICHs are located towards the middle of the brain. The pipeline described is semi-automated allowing for more reproducible and objective analyses; the only non-automated steps in our pipeline are the ICH segmentation and export from OsiriX.

## 4.2 Voxel-wise Analysis of Functional Scores

Voxel-wise hypothesis tests were calculated using linear models or Wilcoxon rank-sum tests for NIHSS and GCS scores. The resulting p-values from these tests indicate that ICH at locations near the ventricles, particularly on the left side, may be related to functional scores. Although the voxel-wise regressions did not have any statistically significant results, this result is likely due to the strict threshold imposed by the Bonferroni correction. As each voxel is not independent of the other voxels as ICH is commonly a contiguous region of hemorrhage, this correction is inappropriate. Other methods that attempt to account of the smoothness properties, such as random field theory, may be more powerful. The voxel-wise analysis does provide a proof of concept and a screening procedure to generate a ROI for patient-level analysis.

## 4.3 Region of Interest Analysis

The region of interest analysis provides some evidence that using imaging-based location measures can result in better performance when estimating functional scores. Although some measures were still relatively low with respect to model fit in absolute terms, such as an  $R^2$  of 0.24, these measures showed demonstrable gains over using the reader-based categories commonly used in analysis.

This analysis focused on NIHSS and GCS scores as these were available at enrollment, but the process may be applied to long-term functional scores. As these populations had different interventions performed, we aimed to analyze outcomes that were prior to any separation of the groups induced by the trial.

## 4.4 Limitations

Although this set of patients represent a large proportion (79%) of the entire population ( $N = 141$ ) of the MISTIE and ICES trials, the sample size is relatively small. This size was due to our selection process for scans and the number of patients available. The framework does not impose a restriction on the number of patients to be processed; it is extendable to much larger (hundreds or thousands) populations.

The current description of hemorrhages and all subsequent analyses are reliant on one specific registration technique. If the registration is insufficient or poor, each patient's voxels are not spatially in the same location and aggregating over scans would not be appropriate or valid. However, at this time the current software is the sole CT-only registration software and CT template published and available. Though we have seen results indicating sufficient registration to the CT template, the template was made from healthy adults. Thus, brains with large deformations are registered to the template that was constructed from typical brains. In future studies, we may want to create a study-specific template for population-level analysis. However, in that case, comparisons between analyses may be impossible as templates would change from study to study. Also, assessments of registration by visual inspection may be subjective.

The results from the ROI-based analysis are flawed in that the ROI has information based on the functional score by how they were generated. Cross-validation can be used to split the data to a training group used to generate the ROI and a testing group to estimate model performance. We plan to validate these ROIs on the subset of patients in the MISTIE and ICES trials not analyzed here.

## 4.5 Extensions of Analysis and Future Work

Although this analysis focused on patients from a single trial cohort, these tools can be applied to large-scale epidemiologic studies to better understand the ICH distribution in specific patient populations of interest and help predict patient outcomes. Moreover, this framework allows for the direct comparison of population-level ICH incidence between two populations. As each voxel represents a proportion, voxel-wise 2-sample proportion tests can be done.

Future research will focus on different registration techniques approaches and the variability of results conditional on the registration technique chosen.

## 5 Funding and Disclosures

The project described was supported by the National Institutes of Health (NIH) grant RO1EB012547 from the National Institute of Biomedical Imaging And Bioengineering, training grant T32AG000247 from the National Institute on Aging, NIH grants RO1NS060910 and RO1NS085211 from the National Institute of Neurological Disorders and Stroke (NINDS), and by NIH grant RO1MH095836 from the National Institute of Mental Health.

Dr. Daniel F. Hanley was awarded significant research support of grants number R01NS046309 and 5U01NS062851 from NINDS. Johns Hopkins University holds a use patent for intraventricular tissue plasminogen activator.

## References

- [1] Hirotugu Akaike. “Information theory and an extension of the maximum likelihood principle”. In: *Second international symposium on information theory*. Akademina Kiado, 1973, 267281.
- [2] John Ashburner and Karl J. Friston. “Unified segmentation”. In: *NeuroImage* 26.3 (July 1, 2005), pp. 839–851.
- [3] M Bergstrm et al. “Variation with time of the attenuation values of intracranial hematomas”. In: *Journal of computer assisted tomography* 1.1 (1977). PMID: 615894, pp. 57–63.
- [4] M. Bhattacharya and D. D. Majumder. “Registration of CT and MR images of Alzheimers patient: a shape theoretic approach”. In: *Pattern Recognition Letters* 21.6 (June 2000), pp. 531–548.
- [5] J. P. Broderick et al. “Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality.” In: *Stroke* 24.7 (July 1, 1993). PMID: 8322400, pp. 987–993.
- [6] T. Brott et al. “Measurements of acute cerebral infarction: a clinical examination scale.” In: *Stroke* 20.7 (July 1, 1989). PMID: 2749846, pp. 864–870.
- [7] J. Ricardo Carhuapoma, Stephan A. Mayer, and Daniel F. Hanley. *Intracerebral Hemorrhage*. Cambridge University Press, Nov. 12, 2009. 277 pp.
- [8] Raymond Tak Fai Cheung and Liang-Yu Zou. “Use of the Original, Modified, or New Intracerebral Hemorrhage Score to Predict Mortality and Morbidity After Intracerebral Hemorrhage”. In: *Stroke* 34.7 (July 1, 2003). PMID: 12805488, pp. 1717–1722.
- [9] Ya-Ching Chuang et al. “Risk stratification for predicting 30-day mortality of intracerebral hemorrhage”. In: *International Journal for Quality in Health Care* 21.6 (Dec. 1, 2009). PMID: 19828550, pp. 441–447.
- [10] Arpita Das and Mahua Bhattacharya. “Affine-based registration of CT and MR modality images of human brain using multiresolution approaches: comparative study on genetic algorithm and particle swarm optimization”. In: *Neural Computing and Applications* 20.2 (Mar. 1, 2011), pp. 223–237.
- [11] P. Daverat et al. “Death and functional outcome after spontaneous intracerebral hemorrhage. A prospective study of 166 cases using multivariate analysis.” In: *Stroke* 22.1 (1991). PMID: 1987664, pp. 1–6.
- [12] Michael N. Diringer, Dorothy F. Edwards, and Allyson R. Zazulia. “Hydrocephalus: A Previously Unrecognized Predictor of Poor Outcome From Supratentorial Intracerebral Hemorrhage”. In: *Stroke* 29.7 (July 1, 1998). PMID: 9660386, pp. 1352–1357.
- [13] Emilie FitzMaurice et al. “Effect of Statins on Intracerebral Hemorrhage Outcome and Recurrence”. In: *Stroke* 39.7 (July 1, 2008). PMID: 18436876, pp. 2151–2154.
- [14] J.M. Fitzpatrick et al. “Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain”. In: *IEEE Transactions on Medical Imaging* 17.4 (Aug. 1998), pp. 571–585.
- [15] Alan S. Go et al. “Heart Disease and Stroke Statistics 2013 Update A Report From the American Heart Association”. In: *Circulation* 127.1 (2013). PMID: 23239837, e6–e245.
- [16] Chen Hallevy et al. “Spontaneous supratentorial intracerebral hemorrhage: Criteria for short-term functional outcome prediction”. In: *Journal of Neurology* 249.12 (Dec. 2002), pp. 1704–1709.
- [17] J. Claude Hemphill et al. “The ICH Score A Simple, Reliable Grading Scale for Intracerebral Hemorrhage”. In: *Stroke* 32.4 (Apr. 1, 2001). PMID: 11283388, pp. 891–897.
- [18] Mark Jenkinson et al. “FSL”. In: *NeuroImage* 62.2 (Aug. 15, 2012), pp. 782–790.
- [19] Cynthia Jongen et al. “Construction and evaluation of an average CT brain image for inter-subject registration”. In: *Computers in Biology and Medicine* 34.8 (Dec. 2004), pp. 647–662.

- [20] Rita V. Krishnamurthi et al. "The Global Burden of Hemorrhagic Stroke: A Summary of Findings From the GBD 2010 Study". In: *Global Heart*. The Global Burden of Cardiovascular Diseases 9.1 (Mar. 2014), pp. 101–106.
- [21] Yonghong Li et al. "Registration of head CT images with subarachnoid hemorrhage". In: *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*. Vol. 2. IEEE, 2010, 498502.
- [22] D. R. Lisk et al. "Early presentation of hemispheric intracerebral hemorrhage Prediction of outcome and guidelines for treatment allocation". In: *Neurology* 44.1 (1994). PMID: 8290048, pp. 133–133.
- [23] G. Mas et al. "Immediate prognosis of primary intracerebral hemorrhage using an easy model for the prediction of survival". In: *Acta Neurologica Scandinavica* 91.4 (Apr. 1, 1995), pp. 306–309.
- [24] A David Mendelow et al. "Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial". In: *The Lancet* 365.9457 (2005), pp. 387–397.
- [25] A David Mendelow et al. "Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial". In: *The Lancet* 382.9890 (Aug. 9, 2013), pp. 397–408.
- [26] Lewis B. Morgenstern et al. "Guidelines for the Management of Spontaneous Intracerebral Hemorrhage A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association". In: *Stroke* 41.9 (Sept. 1, 2010). PMID: 20651276, pp. 2108–2129.
- [27] W. Andrew Mould et al. "Minimally Invasive Surgery Plus Recombinant Tissue-type Plasminogen Activator for Intracerebral Hemorrhage Evacuation Decreases Perihematomal Edema". In: *Stroke* 44.3 (Mar. 1, 2013). PMID: 23391763, pp. 627–634.
- [28] Kenichi Oishi et al. "Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter". In: *NeuroImage* 43.3 (Nov. 15, 2008), pp. 447–457.
- [29] C A Pelizzari et al. "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain". In: *Journal of computer assisted tomography* 13.1 (Feb. 1989). PMID: 2492038, pp. 20–26.
- [30] Thanh G. Phan et al. "Hydrocephalus Is a Determinant of Early Mortality in Putaminal Hemorrhage". In: *Stroke* 31.9 (Sept. 1, 2000). PMID: 10978045, pp. 2157–2162.
- [31] R. K. Portenoy et al. "Intracerebral haemorrhage: a model for the prediction of outcome." In: *Journal of Neurology, Neurosurgery & Psychiatry* 50.8 (Aug. 1, 1987). PMID: 3655832, pp. 976–979.
- [32] Juan Pablo Princich et al. "Rapid and efficient localization of depth electrodes and cortical labeling using free and open source medical software in epilepsy surgery candidates". In: *Frontiers in Neuroscience* 7 (Dec. 31, 2013). PMID: 24427112 PMID: PMC3876273.
- [33] Adnan I. Qureshi et al. "Predictors of Early Deterioration and Mortality in Black Americans With Spontaneous Intracerebral Hemorrhage". In: *Stroke* 26.10 (Oct. 1, 1995). PMID: 7570722, pp. 1764–1767.
- [34] Adnan I. Qureshi et al. "Spontaneous Intracerebral Hemorrhage". In: *New England Journal of Medicine* 344.19 (2001). PMID: 11346811, pp. 1450–1460.
- [35] John Rankin. "Cerebral vascular accidents in patients over the age of 60. II. Prognosis." In: *Scottish medical journal* 2.5 (1957), 200215.
- [36] Ashfaq A. Razzaq and Rafat Hussain. "Determinants of 30-day mortality of spontaneous intracerebral hemorrhage in Pakistan". In: *Surgical Neurology* 50.4 (Oct. 1998), pp. 336–343.

- [37] Chris Rorden and Matthew Brett. “Stereotaxic Display of Brain Lesions”. In: *Behavioural Neurology* 12.4 (2000), pp. 191–200.
- [38] Christopher Rorden et al. “Age-specific CT and MRI templates for spatial normalization”. In: *NeuroImage* 61.4 (July 16, 2012), pp. 957–965.
- [39] Ramandeep Sahni and Jesse Weinberger. “Management of intracerebral hemorrhage”. In: *Vascular Health and Risk Management* 3.5 (Oct. 2007). PMID: 18078021 PMCID: PMC2291314, pp. 701–709.
- [40] J Senant et al. “[A multi-factorial approach in the vital prognosis of spontaneous intracerebral hematoma]”. In: *Revue neurologique* 144.4 (1988). PMID: 3047834, pp. 279–283.
- [41] Eric E. Smith, Jonathan Rosand, and Steven M. Greenberg. “Imaging of Hemorrhagic Stroke”. In: *Magnetic Resonance Imaging Clinics of North America* 14.2 (May 2006), pp. 127–140.
- [42] Stephen M. Smith. “Fast robust automated brain extraction”. In: *Human Brain Mapping* 17.3 (2002), 143155.
- [43] Jeffrey Solomon et al. “User-friendly software for the analysis of brain lesions (ABLe)”. In: *Computer Methods and Programs in Biomedicine* 86.3 (June 2007), pp. 245–254.
- [44] J. C. van Swieten et al. “Interobserver agreement for the assessment of handicap in stroke patients.” In: *Stroke* 19.5 (May 1, 1988). PMID: 3363593, pp. 604–607.
- [45] G. Teasdale and B. Jennett. “Assessment and prognosis of coma after head injury”. In: *Acta Neurochirurgica* 34.1 (Mar. 1, 1976), pp. 45–55.
- [46] Graham Teasdale and Bryan Jennett. “ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS: A Practical Scale”. In: *The Lancet*. Originally published as Volume 2, Issue 7872 304.7872 (July 13, 1974), pp. 81–84.
- [47] Brandon Whitcer, Volker J. Schmid, and Andrew Thornton. “Working with the DICOM and NIFTI Data Standards in R”. In: *Journal of Statistical Software* 44.6 (2011), 128.
- [48] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2009.
- [49] Wendy C. Ziai et al. “A multicenter, randomized, double-blinded, placebo-controlled phase III study of Clot Lysis Evaluation of Accelerated Resolution of Intraventricular Hemorrhage (CLEAR III)”. In: *International Journal of Stroke* (Sept. 2013), n/a–n/a.

## 6 Tables

Variable (N = 111)	N (%) or Mean (SD)
Age in Years: Mean (SD)	60.8 (11.2)
Gender: Female	35 (31.5%)
NIHSS Score: Mean (SD)	22.1 (8.7)
GCS Score: Mean (SD)	10.0 (3.0)
Race	
Caucasian not Hispanic	59 (53.2%)
African American not Hispanic	35 (31.5%)
Hispanic	12 (10.8%)
Asian or Pacific Islander	5 (4.5%)
Reader-Classified ICH Location	
Putamen	68 (61.3%)
Lobar	33 (29.7%)
Globus Palidus	6 (5.4%)
Thalamus	4 (3.6%)

**Table 1.** Descriptive statistics of the demographic information on the patients.

Number of Voxels	P-value	Adjusted $R^2$	$R^2$	AIC	RMSE
Location Model		0.129	0.178	18.60	8.116
1000	0.0005	0.236	0.265	2.47	7.598
2000	0.0009	0.234	0.263	2.81	7.610
2422	0.0010	0.247	0.275	0.98	7.545
3000	0.0013	0.244	0.272	1.46	7.562
19047	0.0100	0.254	0.282	0.00	7.511
47736	0.0500	0.248	0.276	0.77	7.538

**Table 2.** Table of model-fit measures for NIHSS score: reader-based location model vs. CT voxel-based ROI coverage models. We see that the model using a p-value threshold of 0.0100 corresponds to the best model on all model-fit measures: adjusted and unadjusted  $R^2$  are the highest; AIC, and RMSE are the smallest. The models using ROI coverage have comparable fit measures, but each ROI coverage model outperforms the reader-based location model, notably in the  $R^2$  measures. Thus, we infer that using CT-based regions of interest may help in prediction of NIHSS score.

Number of Voxels	P-value	Adjusted $R^2$	$R^2$	AIC	RMSE
Location Model		0.069	0.120	20.51	2.914
1000	0.0002	0.214	0.243	0.00	2.677
2000	0.0004	0.213	0.242	0.09	2.678
3000	0.0006	0.212	0.241	0.25	2.680
4669	0.0010	0.212	0.241	0.23	2.680
22858	0.0100	0.191	0.221	3.17	2.716
52368	0.0500	0.166	0.197	6.44	2.757

**Table 3.** Table of model-fit measures for GCS score: reader-based location model vs. CT voxel-based ROI coverage models. We see that the model using a using the voxels with lowest 1000 p-values corresponds to the best model on all model-fit measures: adjusted and unadjusted  $R^2$  are the highest; AIC, and RMSE are the smallest. The models using ROI coverage have comparable fit measures, but each ROI coverage model outperforms the reader-based location model, notably in the  $R^2$  measures. Thus, we infer that using CT-based regions of interest may help in prediction of NIHSS score.

	NIHSS Score		GCS Score	
	ROI Coverage	Reader-Location	ROI Coverage	Reader-Location
Age	-0.04 (-0.2, 0.1)	-0.1 (-0.2, 0.1)	0.02 (-0.03, 0.1)	0.02 (-0.03, 0.1)
Sex: Male vs. Female	-0.7 (-3.8, 2.4)	-1.7 (-5.0, 1.7)	0.03 (-1.1, 1.1)	0.1 (-1.1, 1.3)
TICHVol per 10 cc	0.8 (-0.003, 1.5)	1.6 (0.8, 2.4)	-0.2 (-0.5, 0.02)	-0.5 (-0.7, -0.2)
ROI Coverage per 10% Reader-Based Location	2.0 (1.1, 2.8)		-0.4 (-0.6, -0.2)	
Globus Pallidus		4.5 (-2.9, 11.9)		-1.8 (-4.5, 0.8)
Putamen		4.2 (0.3, 8.2)		-1.2 (-2.6, 0.2)
Thalamus		4.8 (-4.1, 13.6)		-1.0 (-4.2, 2.1)
Constant	18.8 (9.1, 28.4)	19.6 (7.5, 31.7)	10.6 (7.4, 13.7)	11.2 (6.9, 15.5)

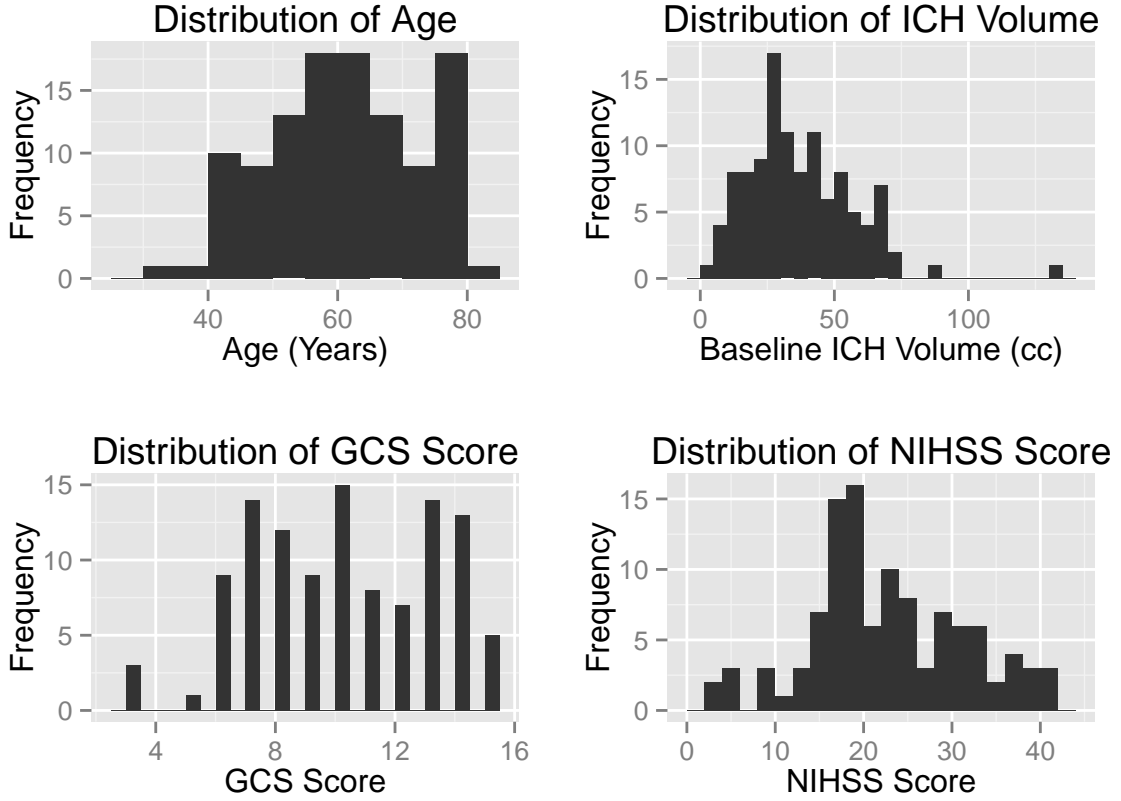
**Table 4.** Regression Models for ROI-Based Analysis. The models for ROI coverage represent the best model based on the model-fit measures. We see that after adjusting for age, sex, and total baseline ICH volume, increasing 10% coverage is expected to increase NIHSS score by 2.0 (95% CI: 1.1, 2.8) points. We see that all locations, compared to lobar hemorrhages have higher estimated NIHSS scores, but putaminal hemorrhages were significantly higher by 4.2 (95% CI: 0.2, 8.2) points. Adjusting for other covariates, increasing 10% coverage is expected to decrease GCS score by 0.4 (95% CI: 0.6, 0.2) points. We see that all locations, compared to lobar hemorrhages have lower estimated GCS scores, but none were statistically different.



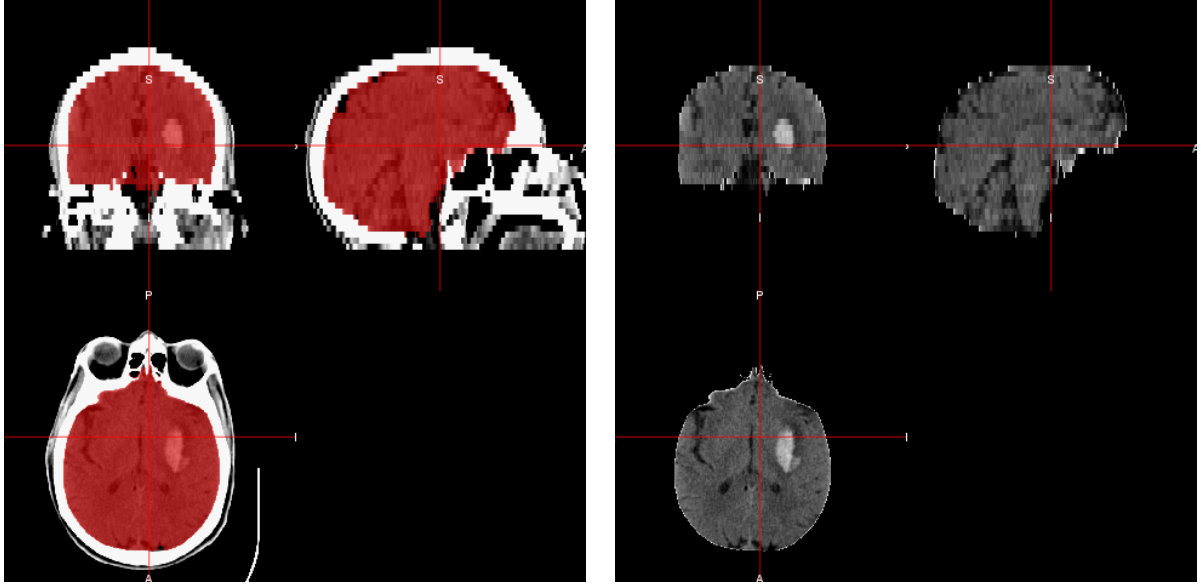
Area	Population Prevalence	NIHSS ROI	GCS ROI
CSF	7.9	10.0	4.2
Insular	4.7		
Superior temporal gyrus	3.8		
Putamen left	3.0		
Insular right	2.9		
External capsule left	2.3		
Superior corona radiata left	1.9	11.8	27.9
Superior temporal wm left	1.9		
Superior corona radiata right	1.8		
Putamen right	1.8		
Posterior limb of internal capsule left		10.1	3.9
Thalamus left		7.6	33.9
Caudate nucleus left		5.4	9.6
Superior longitudinal fasciculus left		4.9	5.9
Globus pallidus left		3.7	
Anterior limb of internal capsule left		3.6	
Outside brain mask		3.5	
Anterior limb of internal capsule right		3.0	
Postcentral wm left			6.7
Posterior corona radiata left			3.1
Precentral wm left			1.3
Supramarginal wm left			1.1

**Table 5.** Distribution of the top 10 areas of engagement for population 3D histogram, the NIHSS ROI was based on a p-value threshold of 0.01, the GCS ROI was based on voxels with 1000 smallest p-values. Each value represents the percentage of the ROI engaged in this area. The population-level areas are percentages weighted by proportion. Each distribution of areas is based on the Eve atlas. We see that the population is engaged in areas of the CSF, such as the ventricles, and the insular and putaminal regions. The ROI based on the NIHSS analysis engages primarily areas of the internal capsule and ventricular regions. The ROI based on the GCS analysis engages primarily the left thalamus and superior corona radiata. The Eve atlas can be used to calculate area engagement on a per-scan level as well.

## 7 Figure Legends



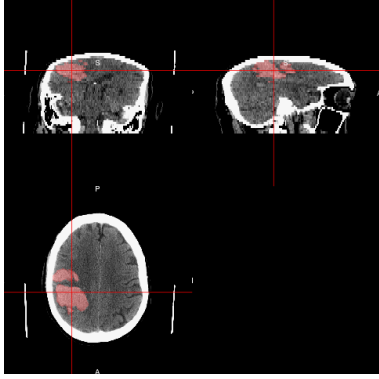
**Figure 1.** Histograms of GCS, age (in years), baseline ICH volume (in cc), and NIHSS scores for the patients in the study. We observe that our population is predominantly over 40 and that are on average 60.8 years of age. The distribution of baseline ICH has a median of 33cc, with some patients having very large ICH volumes. GCS scores indicate that a few patients ( $N = NA$ ) with a GCS of 3 indicating a deep unconsciousness, yet the mean GCS score is 10. The distribution of NIHSS scores shows a similar trend with a mean of 22.1.



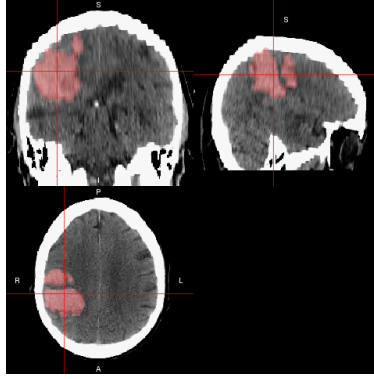
**(a) Original Image** (in native space) with overlaid brain-extraction mask colored in red. We observe that the mask does not appreciably drop out any areas of the brain nor does it add areas of the skull, eyes, or nose.

**(b) Brain-extracted image** used for re-centering to AC/PC line for SPM8 registration techniques.

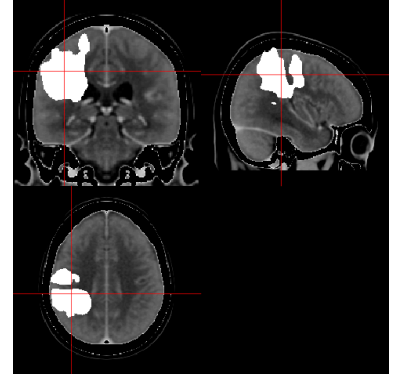
**Figure 2.** Image (a), displays the original image with the brain-extracted mask in red created from BET. Image (b) is the brain extracted image with the mask applied. We observe that the brain extraction appears to extract only the brain and not extra features. The image in (b) can be used for more quantitative patient-level measurements such as intracranial volume, mean Hounsfield units of areas, intensity-based normalization, and can restrict analyses to only brain tissue and not extracranial areas.



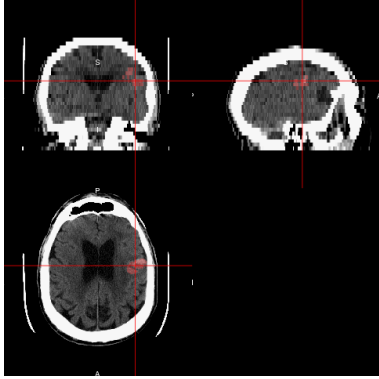
(a) **Native space** where the image looks “scrunched” vertically since the slice thicknesses are smaller for the middle and above part of the brain.



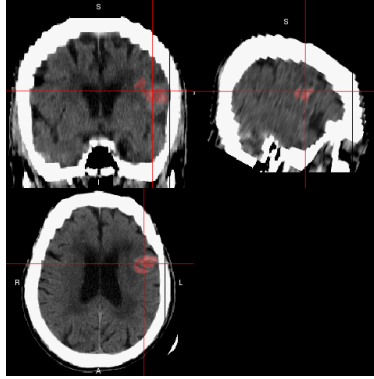
(b) **Registered image** with in template space with hemorrhage mask highlighted in pink. We observe that the brain has been stretched to match the dimensions of the template, yet structures on the patient image look to be in the same areas as that of the template in panel (c).



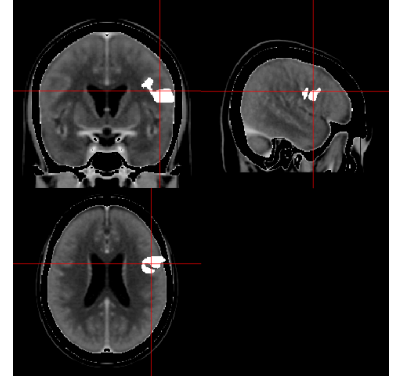
(c) **Template image** with hemorrhage mask overlaid in white.



(d) **Native space** where image acquired with homogeneous slice thickness and a small ICH

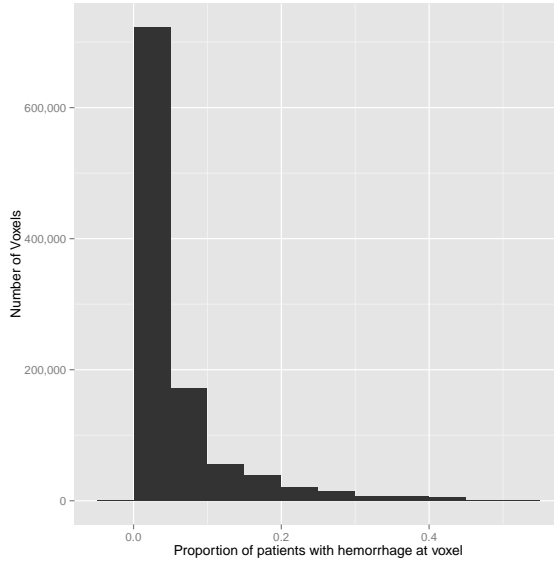


(e) **Registered image** in template space with hemorrhage mask highlighted in pink. We observe the skull (in white) may be scaled during registration, but ventricles in the middle of the brain look similar to those on the template in panel (f).

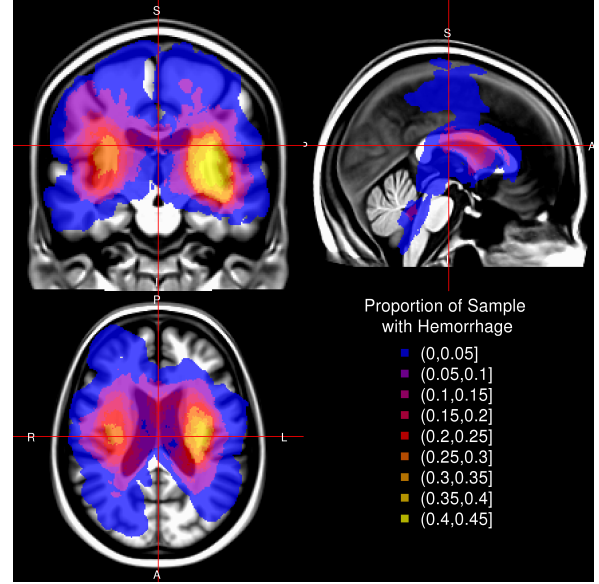


(f) **Template brain** with hemorrhage mask overlaid

**Figure 3.** Native (left, panels (a) and (d)), template-registered (middle, panels (b) and (e)) images from two patients with template brain for comparison (right, panels (c) and (f)). In image (a) shows an image acquired with variable slice thickness, which can be seen as the top of the brain has smaller voxel sizes (thinner slices) than the bottom of the brain. We see in (b), the template-registered image, that the transformation has scaled the brain to the same size as the template. We see that similar structures, such as the lateral ventricles, are observed on the same axial slice on the patient-level scan compared to the template image, which indicates adequate registration. We observe a patient with a small ICH in panel (d), where we see the registration preserves the hemorrhage with respect to surrounding tissue, but may scale areas, such as the skull around the brain, too much.

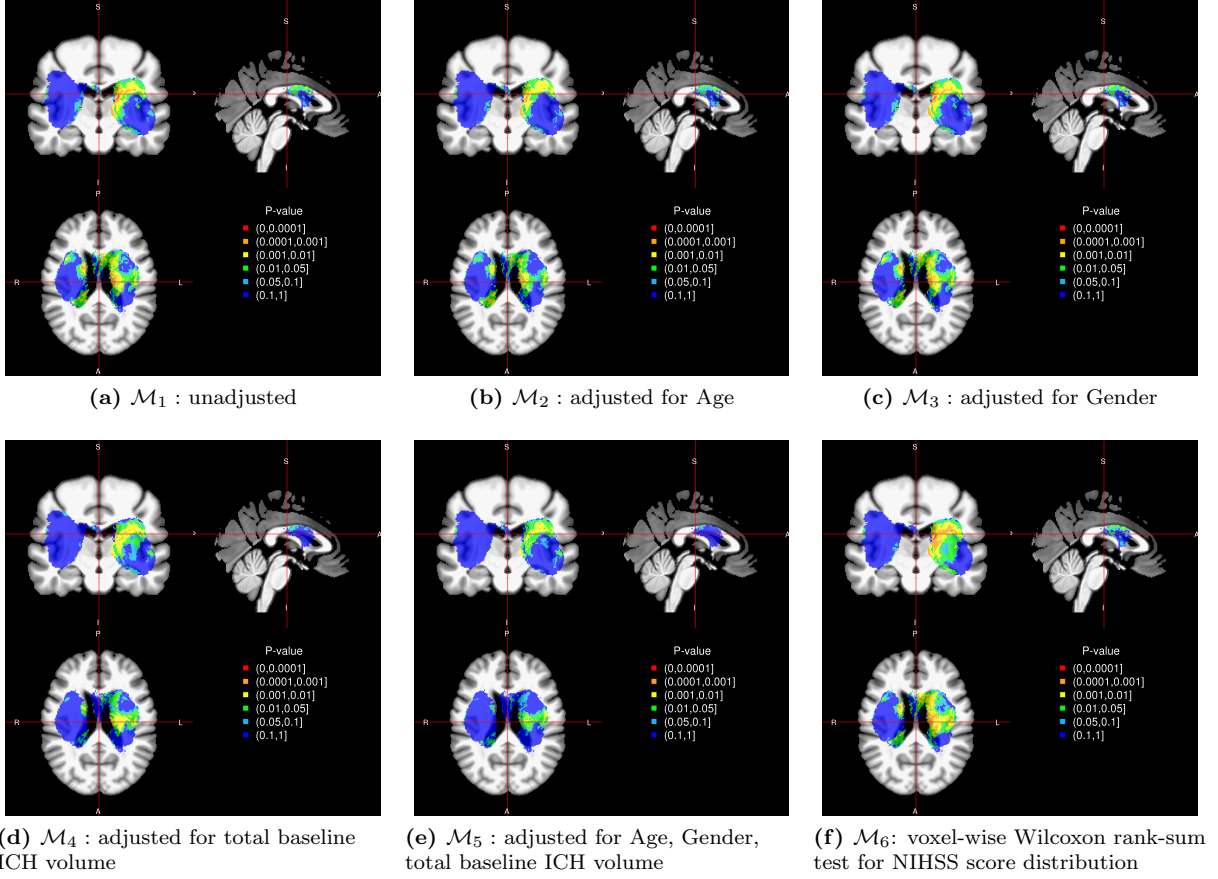


(a) **Distribution** of proportion of patients with hemorrhage. All voxels with 0 proportion are excluded.

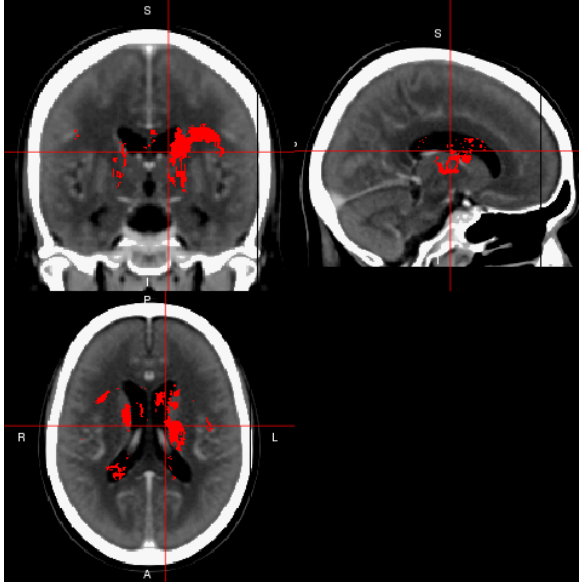


(b) **Template brain (MRI T1)** with proportion of hemorrhage overlaid

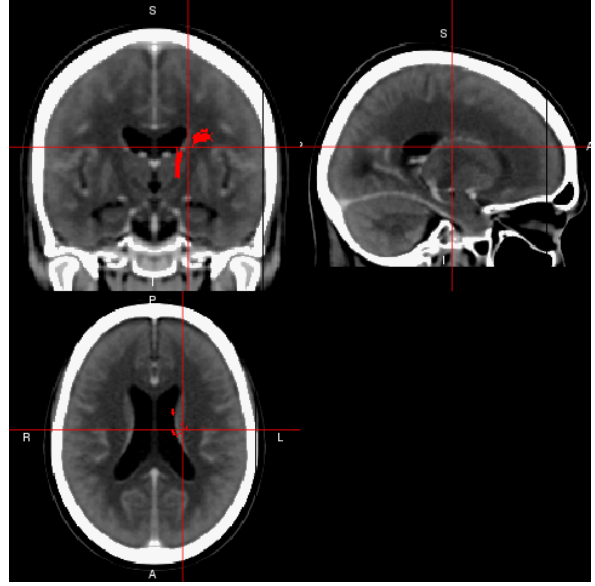
**Figure 4.** The histogram of proportion of patients for each voxel that had hemorrhage is presented in Figure (a). Voxels with 0 proportion are excluded. We observe the majority of voxels have a low prevalence, with a median of 3%, but some voxels ( $V = 5685$ ) have a high prevalence of over 40% in this sample. In Figure (b), we present these proportions in a 3D histogram image (radiological convention - right side of image is left side of brain). Overlaid on the image is a MRI T1 template for spatial orientation and depictions of brain structures. Brighter values denote a higher percentage of patients that have a hemorrhage at that specific voxel. We see some bi-laterality of the image, but more hemorrhages in the left side of the brain compared to the right. ICH is also somewhat localized in the middle of the brain, with few extensions in the far anterior and posterior areas of the brain. The interactive version of this figure is located at [http://muschellij2.github.io/CT\\_Pipeline/index.html](http://muschellij2.github.io/CT_Pipeline/index.html).



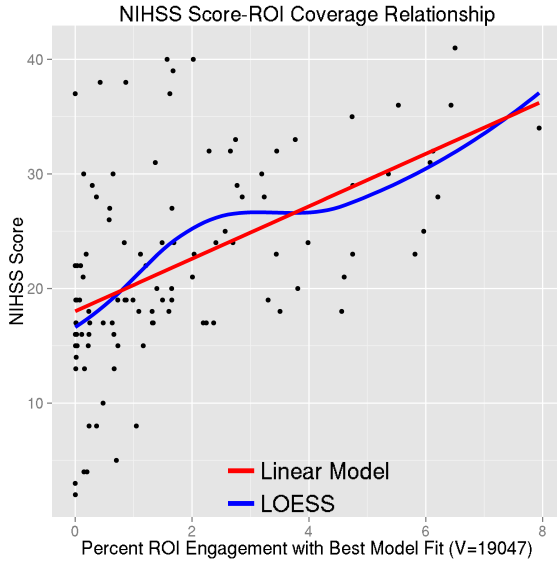
**Figure 5.** P-value maps for the 6 models, with NIHSS score as functional outcome. Colors correspond to p-values  $.1 - 1$  (dark blue),  $.1 - .05$ , (light blue),  $.05 - .01$  (green),  $.01 - .001$  (yellow),  $.001 - .0001$  (orange)  $< .0001$  (red). We see that in the unadjusted linear model (panel (a)) and after adjusting for gender (panel (c)) there are voxels with the smallest p-values seem to be near the ventricles. In models adjusting for age (panel (b)), or total baseline ICH volume (panel (d)), or both (panel (e)), the p-value relating to NIHSS scores appear higher (more bluish). Also, we see the Wilcoxon rank-sum test have lower p-values for the same area compared to the unadjusted linear model (panel (f)). Slices are presented at the middle of the template.



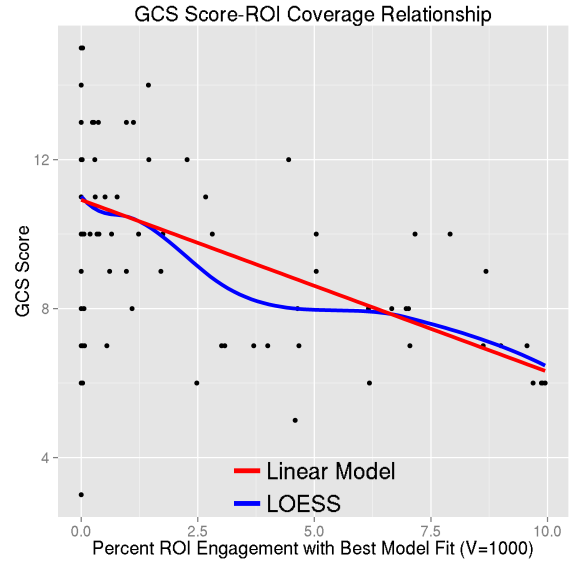
(a) ROI for NIHSS score; voxels less than 0.0100



(b) ROI for GCS score; voxels with the lowest 1000 p-values



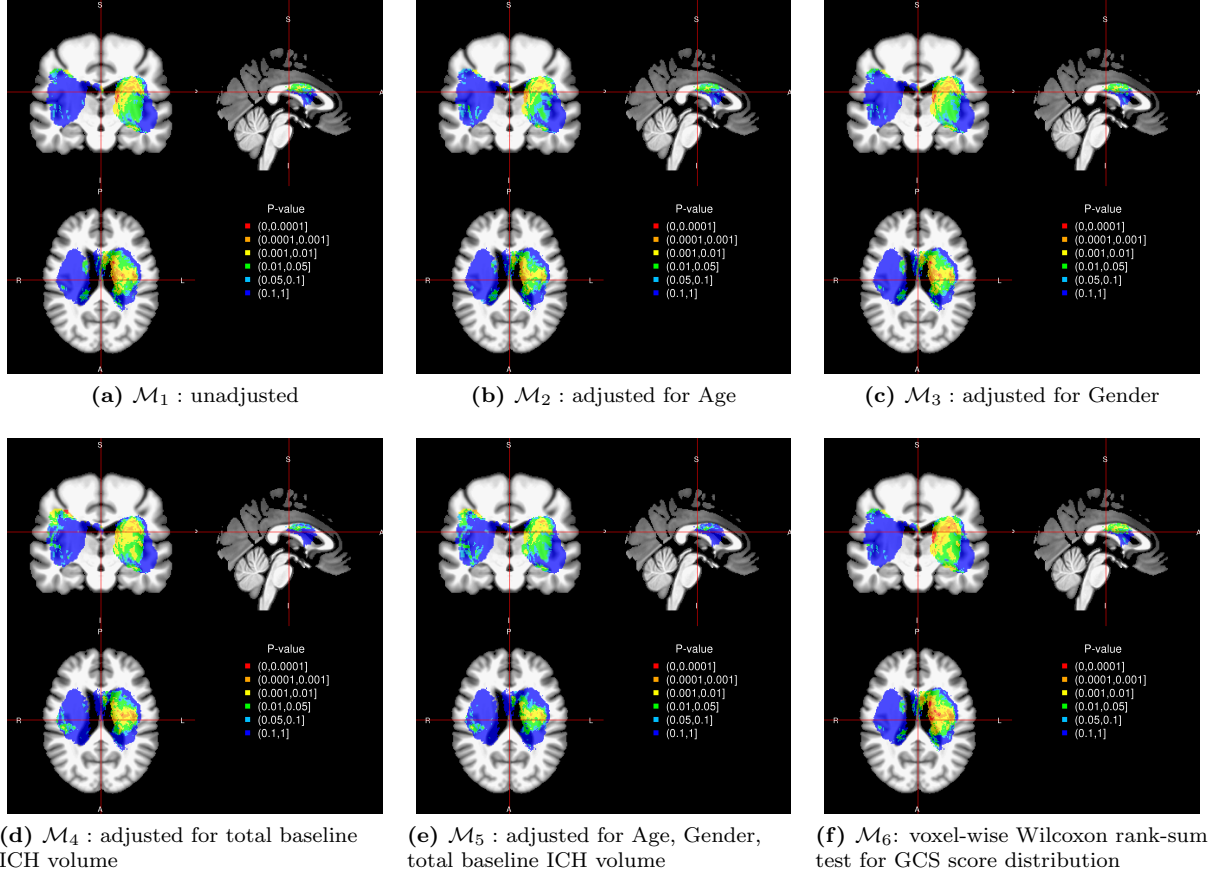
(c) Relationship of NIHSS score and ROI coverage for ROI in (a)



(d) Relationship of GCS score and ROI coverage for ROI in (b)

**Figure 6.** Region of Interest (ROI) Analysis. Panels (a) and (b) correspond to the ROI for the top-performing model for NIHSS and GCS scores, respectively. The ROI in panel (a) represents a p-value threshold of 0.0100 for the voxel-wise ICH on NIHSS score regressions, which includes 19047 voxels. The ROI in panel (b) represents 1000 with the lowest p-values for the voxel-wise ICH on GCS score regressions, corresponding to a p-value threshold of 0.0002. Panels (c) and (d) display the relationship of the ROI coverage to the functional score. The red line represents a linear fit (without adjustment of other covariates) of the ROI coverage and functional score and the blue line represents a LOESS fit. We see that the larger the ROI coverage the higher (more severe stroke) NIHSS score and the lower (deeper unconsciousness) the GCS score.

## 8 Supplemental Material



**Figure 7.** P-value maps for the 6 models, with GCS score as functional outcome. Colors correspond to p-values .1 – 1 (dark blue), .1 – .05, (light blue), .05 – .01 (green), .01 – .001 (yellow), .001 – .0001 (orange) < .0001 (red). We see that in the unadjusted linear model (panel (a)) and after adjusting for gender (panel (c)) there are voxels with the smallest p-values seem to be near the ventricles. In models adjusting for age (panel (b)), or total baseline ICH volume (panel (d)), or both (panel (e)), the p-value relating to GCS scores appear higher (more bluish). Also, we see the Wilcoxon rank-sum test have lower p-values for the same area compared to the unadjusted linear model (panel (f)). Slices are presented at the middle of the template.