

All authors have read and approved the submitted manuscript, the manuscript has not been submitted elsewhere nor published elsewhere in whole or in part.

Suggested Reviewers:

1. Dr. Kennedy Lees
- 2.
- 3.

Quantitative Localization and Predictive Performance of Intracranial Hemorrhage

John Muschelli^{1,*}, ScM; Natalie L. Ullman², BS; Elizabeth M. Sweeney¹, ScM; Ani Eloyan¹, PhD; Neil Martin³, MD, FACS; Paul Vespa³, MD, FCCM, FAANS; Daniel F. Hanley², MD; Ciprian M. Crainiceanu¹, PhD

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

2 Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA

3 Department of Neurosurgery, David Geffen School of Medicine at UCLA, Los Angeles, California

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

Abstract

Background and Purpose

Current studies of association between intracranial hemorrhage (ICH) localization and stroke severity present conflicting evidence. Using advanced registration techniques of computed tomography (CT) scans, we provide detailed quantification of ICH location and its association with stroke outcomes.

Methods

We analyzed 111 scans from 111 patients enrolled in the MISTIE (Minimally Invasive Surgery plus recombinant-tissue plasminogen activator (rtPA) for Intracerebral Evacuation) trial. We registered CT scans to a CT template, estimated a 3-dimensional map of where ICH engagement occurs, and derived regions predictive of 2 severity scores: the National Institutes of Health Stroke Scale (NIHSS) and Glasgow Coma Scale (GCS) scores. We compared the prediction performance of these regions compared to using location information determined by human readers using the adjusted R^2 from linear regression models.

Results

In this sample, ICH is primarily located in deep brain nuclei and and lobar white matter, including the superior corona radiata and the thalamus, is related to severity scores. The prediction performance was twice as good for the image-based model compared to the reader-based model (adjusted R^2 : 0.254 vs. 0.129 for NIHSS, 0.214 vs. 0.069 for GCS).

Conclusions

Measures of ICH location and engagement using advanced CT imaging processing provide finer, more quantitative, and more predictive anatomic information than that provided by expert human readers. These approaches provide an analytic platform for hypothesis testing and discovery of new hypotheses.

Clinical Trial Registration Information

Clinical Trial Registration-URL: <http://www.clinicaltrials.gov>. Unique identifier: NCT00224770.

Keywords: intracranial Hemorrhage; CT imaging analysis; stroke; severity score prediction

1 Introduction

Intracranial hemorrhage (ICH) results from a blood vessel rupturing into brain tissues and possibly the ventricles. Bleeding causes distension of brain structures and increases the likelihood of intracranial pressure (ICP) elevation. ICH accounts for 10-15% of all strokes, corresponding to approximately 80,000 annual cases [1], 30,000 deaths in the US [2], and 5 million cases worldwide [3]. CT scanning is widely available and is the most commonly used diagnostic tool in patients with ICH [4]. Clinicians utilize CT to define the location of bleeding, clinically assess severity, and plan patient management.

Despite robust correlation of cerebral location and functional performance in normal humans, location of ICH surprisingly is not an important factor in predicting severity of injury or prognosis [5, 6, 7, 8, 9, 10]. Furthermore, recent clinical trials do not demonstrate an important role for location as a factor associated with beneficial clinical outcome [11, 12, 13, 14].

1.1 Problems with Visual Inspection

The classification of hemorrhage location is complicated for even the best-trained neuroimaging scientists. For example, a hemorrhage may extend into multiple brain areas, distend tissues altering anatomic relationships, and may break through the ventricular wall. Evaluating these anatomic possibilities challenges even the best clinicians, thus routine practice identifies a single location as the primary affected anatomic region (e.g. caudate, putamen, etc.) or describes the location of the edge of the hemorrhage in relation to a given landmark [15]. Outcome is strongly associated with hemorrhage volume; importantly, the modulation of the relationship between volume and location has not been studied. To investigate these anatomic issues, detailed localization information can be obtained by registering scans to a common template to provide refined anatomical localization information.

1.2 Previous CT Registration Work

Registration to template space is a crucial first step for any across-patient analysis; this allows each patient's scan to be located in the same stereotaxic space so information may be combined spatially across scans. Moreover, brain atlases with spatially defined anatomic structures are available in template space. Recently, Rorden et al. [16] released the first publicly available CT template of healthy adults in MNI (Montreal Neurological Institute) space. This image provides a simple registration where the reference and template images are in a single imaging modality as opposed to more complex referencing systems [17, 18, 19].

1.3 Hypothesis

We propose to use CT images from the MISTIE (Minimally Invasive Surgery plus recombinant-tissue plasminogen activator (rtPA) for Intracerebral Evacuation) trial to investigate the benefit or lack thereof in utilizing anatomic location as a biologically plausible predictor of ICH severity. Thus, we propose to test the hypothesis that routine clinical anatomic localization was no different than quantitative localization derived from registered-to-template images with atlas-based labeling for prediction of severity of injury.

2 Methods

To address this hypothesis, we will 1) create a 3-dimensional (3D) density map of hemorrhages occurring in a population of patients with ICH; 2) provide detailed quantification of hemorrhage engagement of individual neuroanatomic regions within the brain; 3) determine if differences in location relate to two common disability scores: the National Institutes of Health Stroke Scale (NIHSS) [20] and the modified

Glasgow Coma Scale (GCS) [21, 22]; and 4) generate a stroke region of engagement that is likely to be associated with stroke severity as well as test its predictive performance using within-sample validation.

All statistical analysis was done in the R statistical programming language (<http://cran.r-project.org/>).

2.1 Subjects and Demographics

The population studied consists of 111 patients from MISTIE recruited from 26 centers with lobar and deep ICHs ≥ 20 cc in volume [23]. CT and clinical data were collected as part of the Johns Hopkins Medicine IRB-approved MISTIE research studies with written consent from participants.

The NIHSS and GCS scores were recorded at enrollment; 3 patients did not have a recorded NIHSS score, 1 patient did not have a recorded GCS score. 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.

Descriptive demographic of age, sex, race, baseline ICH and IVH volume, and NIHSS and GCS scores are shown in Table 1. The NIHSS score reflects stroke-related impairment (higher is worse), while the GCS score reflects a patient’s level of consciousness (higher is better).

[Table 1 here.]

2.2 Imaging Data

Standard diagnostic CT images of the MISTIE trial subjects were registered to template space. The study protocol was executed with minor, but important, differences across the 26 sites. 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 for 14 scans, referred to as variable slice thickness. For example, a scan may have 10 millimeter (mm) slices at the top and bottom of the brain, where no hematoma is present, but with 5mm slices in the middle where the hematoma is seen (see Supplemental Figure IA). Therefore, the scans analyzed had different voxel (volume element) dimensions and image resolution prior to registration to the template. These conditions represent a pragmatic sample of stroke center diagnostic imaging.

2.3 Hemorrhage Segmentation and Location Identification

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 [24, 25]; these regions were then further quality controlled and refined by readers using direct inspection of images. Readers also identified the specific anatomic location most engaged by the ICH (Table 1).

2.4 Image Registration

The DICOM (Digital Imaging and Communications in Medicine) data was preprocessed to obtain a 3D brain image (see Supplemental Section 10.1.1 for details). The image was then spatially registered to the CT template using the Clinical toolbox [16], which employs the unified normalization-segmentation routine [26] from the statistical parametric mapping (version 8, SPM8, Wellcome Trust Centre for Neuroimaging, London, United Kingdom) software in MATLAB (The Mathworks, Natick, Massachusetts, USA). The binary hemorrhage mask was transformed into the template space.

No scans were excluded due to inadequate registration determined by visual inspection by expert CT readers

2.5 Histograms of ICH in the Brain

To visualize and describe the localization of ICH, we first combined information from registered masks of 111 patients. Using these ICH masks, we obtained the 3-dimensional (3D) histogram of ICH localization for the study population. More precisely, for every voxel in the template space, we calculated the proportion of patients who have an ICH at that particular voxel. We used the R package `brainR` to create an 3D interactive map of the 3D histogram [27]. The interactive map is located at http://muschellij2.github.io/CT_Pipeline/index.html.

2.6 Prediction of Severity Score Based on Hemorrhage Location

In the study population, 1,045,174 voxels had at least one patient with ICH. We limited our analysis to voxels in the template space where at least 10 patients exhibit ICH (166202 voxels) to optimize the models for a substantial proportion of the study population.

We tested the association between hemorrhage location and stroke severity as measured by the NIHSS score and GCS score running a series of models, accounting for confounders. At each voxel, we ran a linear regression model:

$$Y_i = \beta_0 + \beta_1(v)ICH_i(v) + \gamma X_i + \epsilon_i(v), \quad (1)$$

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 , and $ICH_i(v) = 0$ otherwise. X_i is a vector of patient-specific confounders with effects γ and $\epsilon_i(v)$ are assumed independent homoscedastic errors. The confounders, X_i , either were excluded for "unadjusted" voxel-level model, or contains a combination of 3 patient-specific confounders: age, sex, and total baseline ICH volume (TotalVol) in the "adjusted" models. 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 model were calculated, testing the null hypothesis $H_{0,v} : \beta_1(v) = 0$, or in the case of the Wilcoxon rank-sum test: $H_{0,v} : Severity\{ICH(v) = 1\} = Severity\{ICH(v) = 0\}$ where *Severity* denotes the distribution of severity scores of patients. Figure 2 displays the voxel-wise p-values from NIHSS score models (see Supplemental Figure II for GCS). The p-value images displayed were not corrected for multiple comparisons since the purpose is investigate regional brain anatomy where ICH engagement could relate to severity score.

We did investigate whether individual locations are predictive of severity score after accounting for multiple comparisons using a Bonferroni correction with a family wise error rate of $\alpha = .05$. At this stringent level no location was found to be significantly associated with the scores.

2.6.1 Highest Predictive Region Generation and Analysis

Although no voxel passed this strict correction, voxels with low p-values indicate candidate regions which may improve prediction of severity scores. In order to create a patient-level covariate that summarizes ICH location information, we created a sequence of nested regions of interest by selecting voxels based on the smallest p-values obtained from the unadjusted linear model, i.e. from a model where only the voxel-level ICH indicator was used, but no other covariates. We call these regions "highest predictive regions" (HPR) because they contain the locations of those voxels that are most predictive of the severity scores. We obtained 6 different HPR, 3 based on the smallest 1000, 2000, or 3000 lowest p-values and three based on p-values thresholds of .05, .01, and .001. For each HPR, we calculated the HPR "coverage" which represents the percentage of the voxels in the HPR that were classified as hemorrhage in the subject-specific image:

$$Coverage_i = \frac{\# \text{ Voxels classified ICH in HPR for scan } i}{\# \text{ Voxels in HPR}} \times 100\%$$

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

$$Y_i = \beta_0 + \beta_1 \text{Coverage}_i + \gamma_1 \text{Age}_i + \gamma_2 \text{Sex}_i + \gamma_3 \text{TotalVol}_i + \epsilon_i \quad (2)$$

We compared model (2) to one using a categorical indicator of the expert-specified ICH location, with categories: Thalamus ($N = 4$), Globus Pallidus ($N = 6$), Putamen ($N = 68$), and Lobar ($N = 33$). Prediction performance and model fit were assessed using R^2 , adjusted R^2 , Akaike information criterion (AIC) [28], and root mean squared error (RMSE).

2.7 ICH Localization and Engagement

Although prediction of severity score is of interest, standard practice conveys information based on known neuroanatomic regions. We automatically calculated spatial ICH engagement by neuroanatomic region using brain atlases with defined segmentations. We used the “Eve” atlas [29], which segments gray matter (GM) and white matter (WM) regions. Ventricular regions were not explicitly segmented; any region not classified as GM or WM were classified as cerebrospinal fluid (CSF).

From this atlas, we estimated for each patient scan: 1) the percent of the ICH engaged by region and 2) the percent of each region engaged by the ICH (see Supplemental Section 10.1.2 for further details). These summaries of ICH engagement provide a much finer description of location than what can currently be done by expert human readers. For example, instead of classifying a region as a putaminal bleed, we may indicate that an ICH engages 78% of the putamen. We did not directly compare reader-classified regions and the most-engaged region classified by the atlas, as the atlas labels do not directly map to the reader-classified categories.

We further summarized neuroanatomic engagement at the population level (see Supplemental Section 10.1.2 for details).

3 Results

3.1 Prevalence of ICH Engagement in the Brain

The non-spatial distribution of the prevalence of hemorrhages over all voxels (Figure 1A) shows the majority of voxels have a low prevalence of ICH engagement; the median number of patients with ICH at a given voxel is 3 (3%), though a small group of voxels ($V = 5685$) have a high prevalence of $> 40\%$ of the sample population. Figure 1B represents the 3D histogram of hemorrhage prevalence, where colors represent the percentage of patients with ICH engagement at that given area. This image indicates that ICH is distributed medially in the brain in this cohort, with a lower concentration at the cortical surface and higher on the left side of the brain. Figure 1B also indicates that the prevalence of strokes in the extreme anterior and posterior areas of the brain is very low. These observations may be a result of the inclusion criteria, but the inclusion criteria [23] did not specifically prefer some spatial locations over others.

[Figure 1 here.]

3.2 Prediction of Functional Score Based on Hemorrhage Location

To study the association between localization and stroke severity scores, we have fit model (1) using a sequence of potential confounding adjustments. More precisely, the five voxel-wise linear models fit were:

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

where \mathcal{M} denotes a model. For consistency of notation (as it is not an explicit model) we will refer to Wilcoxon rank-sum test as \mathcal{M}_6 .

No voxels were significantly related to NIHSS score or GCS score in any model after using the Bonferroni correction. The p-values from the models, with NIHSS score as the outcome, are presented in Figure 2, overlaid on a MRI T1 image for spatial localization of structures (\mathcal{M}_3 and \mathcal{M}_4 are not shown, and appear similar to Figure 2C).

The estimated association of location and NIHSS can easily be visualized by comparing the relative location of high and low p-values. Figure 2 indicates that the strongest associations are clustered together in the immediate vicinity of the medial plane both for the GCS (Supplemental Figure II) and NIHSS scores. The p-values are higher (more blue) in the models adjusted for age (not shown), baseline ICH volume (not shown), and both age and baseline ICH volume (Figure 2C) compared to the unadjusted model \mathcal{M}_1 .

[Figure 2 here.]

3.2.1 Highest Predictive Region Analysis

Inspired by the inspection of Figure 2, we investigated whether we can define areas of the brain that can improve prediction of stroke severity scores. To investigate this, we have obtained 6 different HPR, 3 based on the smallest 1000, 2000, and 3000 p-values and three based on all p-values below a particular threshold. We used three thresholds: .05, .01, and .001, corresponding to HPR with 47736, 19047, and 2422 voxels, respectively, for NIHSS and 52368, 22858, and 4669 voxels for GCS, respectively. For illustration, panel (A) in Figure 3 displays the region obtained by choosing all the p-values smaller than .01 in the unadjusted NIHSS regression on voxel location. Panel (B) in Figure 3 displays the region obtained by retaining only the smallest 1000 p-values for the unadjusted regression of GCS score on ICH location. Although we show three orthographic slices for each HPR, the entire regions are available in MNI coordinates.

While the voxel-wise p-values identify areas that are potentially highly associated with the outcome, we want to reduce these complex HPR to a simple subject-specific covariate, HPR coverage. Panel C of Figure 3 displays the NIHSS score (y-axis) as a function of the coverage of the HPR from Figure 3A. Similarly, Panel D of Figure 3 displays the GCS score as a function of the coverage of the HPR in Figure 3B of Figure 3. The blue line represents a non-parametric LOESS (local regression) fit to estimate the relationship between severity and coverage, the red line represents an unadjusted linear model fit. As expected, the larger the HPR coverage the higher (more severe stroke) the NIHSS score and the lower (deeper unconsciousness) the GCS score.

[Figure 3 here.]

Combining regions of engagement from the left and right sides of the brain may be worthwhile as regions will likely affect severity regardless of the hemisphere engaged. We did not combine these areas, as it may not be straightforward to combine ICH that crosses the mid-sagittal plane. We did attempt to “symmetrize” the HPR by including voxels on the contra-lateral side of the thresholded voxels and saw similar results (see Supplemental Figure III).

We further investigated how these regions predict severity compared to an ICH location classification provided by expert readers. Table 2 compares prediction performance based on the expert readers location (labeled “Location model”) and the same models using HPR coverage for NIHSS and GCS scores. Prediction performance is measured primarily using the adjusted R^2 , though conclusions were the same using R^2 , AIC, and RMSE. We concluded that all HPR coverage models strongly outperform

reader-classified location models. Indeed, for NIHSS, the adjusted R^2 almost doubled from 0.129 for the reader-classified location model compared to 0.254 for the best HPR coverage model. For GCS, the adjusted R^2 more than tripled from 0.069 for the reader-classified location model to 0.214 for the HPR coverage model. The other HPR coverage models provide relatively similar prediction performance. This indicates that the choice of one HPR versus another may need to be based on other criteria, such as total HPR volume, ICH population prevalence, and prior biological information. The coefficients for the best HPR and location models can be seen in Supplemental Table I.

[Table 2 here.]

3.2.2 ICH Localization and Engagement

Table 3 represents the 10 most-engaged regions for the population 3D histogram as well as the HPR for the GCS and NIHSS score analyses. The population ICH is engaged in areas of the insular, putamen, and primarily the CSF, such as the ventricles. The HPR based on the NIHSS analysis engages primarily areas of the areas of the internal capsule, thalamus, superior corona radiata, and CSF regions. The HPR based on the GCS analysis engages primarily the thalamus and superior corona radiata.

[Table 3 here.]

Though Table 3 represents the empirical regions most engaged, engagement by specified regions is commonly of interest. We calculated the engagement of the thalamus, putamen and globus pallidus by the population 3D histogram and the HPR for the GCS and NIHSS score analyses (Supplemental Table II). The population engagement represents the mean proportion of the population with ICH engagement for that brain region. The HPR engagement represent the percent of voxels in that brain region (e.g. putamen) that are in the HPR from NIHSS and GCS scores. On average, 23% of the putamen, 20% of the globus pallidus, and 8% of the thalamus are engaged with ICH from patients in this study. The HPR from the NIHSS analysis engages 40% of the globus pallidus, 6% of the putamen, and 9% of the thalamus. The HPR from the GCS analysis engages only 2% of the thalamus, but not the putamen or the globus pallidus; the GCS HPR again is only 1000 voxels.

4 Discussion

4.1 Semi-Automated Localization of ICH

We have characterized the localization of ICH in a population of stroke patients by a 3D histogram of a population with ICH from prospective clinical trials. We found, in this study, ICH was located towards the middle of the brain and more patients had ICH engagement on the left side of the brain. Overall, population maps such as these provide location information, allow researchers to sharpen their hypotheses about ICH location, and allow for such hypotheses to be tested. While this 3D histogram may not generalize to other populations, it will be crucial to obtain such maps for other studies and compare the degree of similarity to ours.

We can use this process to create a 3D histogram based on different studies or subgroups of this population. As the values in the 3D histogram represent proportions, these images can then be used to test differences between the groups using standard proportion tests. This process allows us to directly compare ICH location in 2 groups at a finer scale than currently available.

Moreover, 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 the export of ICH masks from OsiriX.

4.2 Voxel-wise Analysis of Severity Scores

Voxel-wise hypothesis tests were performed using linear models and Wilcoxon rank-sum tests for NIHSS and GCS scores. The resulting p-values from these tests indicate that ICH engagement near the lateral ventricles may be related to severity scores.

None of the voxels passed the stringent Bonferroni correction due to the large number of voxels being tested. Each voxel is not independent of the other voxels since ICH is commonly a contiguous region of hemorrhage; therefore, this correction is too conservative. Other methods that attempt to account for this dependence may be more powerful at voxel detection.

These maps allow us to explore the relationship of ICH location and severity scores, which can be used for hypothesis generation and testing. More importantly, these maps can be used as a potential diagnostic clinical marker or guide surgical interventions. In addition, we have shown the voxel-wise p-values provide a voxel screening procedure to generate a HPR for patient-level analysis.

4.3 Estimation of ICH Engagement with Neuroanatomic Regions

Although 3D spatial maps are worthwhile for viewing, many clinicians want description of location in terms of known neuroanatomic regions. We demonstrate how a segmented atlas (Eve) can be used to automatically describe ICH engagement by neuroanatomic regions at a patient or population level. These measures are more interpretable for clinical relevance and may translate to better determination of disability. Additionally, these measures can be used to objectively compare different groups or populations without the requiring expert readers to determine ICH engagement.

4.4 Highest Predictive Region Analysis

The HPR analysis rejects our hypothesis that routine clinical anatomic localization was no different than quantitative localization derived from registered-to-template images for prediction of severity of injury. Although prediction performance was quite low even for the best models, prediction performance doubled or tripled, depending on the severity score. 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 groups with different interventions, we aimed to analyze outcomes that were prior to any separation of the groups induced by the intervention. This procedure can be used for any patient outcome, such as the modified Rankin scale score [30, 31] at long-term followup visits.

4.5 Summary

The summary and visualization of ICH engagement presented provide a much finer description of location than currently possible by expert human readers. We also have shown that using image-based measurements can substantially better predict severity scores compared to using location determined by experts. This type of analysis opens a framework for rigorous testing of location and derivation of objective measures of ICH engagement.

4.6 Limitations

Although this set of patients represents a large proportion (79%) of the MISTIE and ICES trials, the sample size is relatively small. Also, the current description of hemorrhages and all subsequent analyses are reliant on one specific registration technique. It is probably desirable [32] to try multiple registration approaches and compare results across registrations.

One of the limitation with the HPR analysis is the severity score is used twice: once to find voxels that are most associated with severity and the second time regressing coverage on severity. Cross-validation can be used for internal validation: splitting the data into a training group used to generate the HPR and a testing group to estimate model performance on HPR coverage. Another approach is to externally validate by studying the performance of the HPR on a subset of the MISTIE and ICES patients not analyzed here. We will investigate external validation in our future studies.

5 Sources of Funding

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.

6 Disclosures

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] Alan S. Go et al. “Heart Disease and Stroke Statistics 2013 Update A Report From the American Heart Association”. In: *Circulation* 127.1 (2013), e6–e245.
- [2] Adnan I. Qureshi et al. “Spontaneous Intracerebral Hemorrhage”. In: *New England Journal of Medicine* 344.19 (2001), pp. 1450–1460.
- [3] 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.
- [4] Ramandeep Sahni and Jesse Weinberger. “Management of intracerebral hemorrhage”. In: *Vascular Health and Risk Management* 3.5 (Oct. 2007), pp. 701–709.
- [5] Stanley Tuhim et al. “Validation and comparison of models predicting survival following intracerebral hemorrhage”. In: *Critical care medicine* 23.5 (1995), 950954.
- [6] M. J. Ariesen et al. “Applicability and relevance of models that predict short term outcome after intracerebral haemorrhage”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 76.6 (2005), 839844.
- [7] Magdy Selim et al. “Predictors of Hemorrhagic Transformation After Intravenous Recombinant Tissue Plasminogen Activator Prognostic Value of the Initial Apparent Diffusion Coefficient and Diffusion-Weighted Lesion Volume”. In: *Stroke* 33.8 (Aug. 1, 2002), pp. 2047–2052.
- [8] J. Claude Hemphill et al. “The ICH Score A Simple, Reliable Grading Scale for Intracerebral Hemorrhage”. In: *Stroke* 32.4 (Apr. 1, 2001), pp. 891–897.
- [9] 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.
- [10] 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), pp. 1717–1722.
- [11] 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.
- [12] 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.
- [13] Craig S. Anderson et al. “Intensive blood pressure reduction in acute cerebral haemorrhage trial (INTERACT): a randomised pilot trial”. In: *The Lancet Neurology* 7.5 (2008), 391399.
- [14] Antihypertensive Treatment of Acute Cerebral Hemorrhage (ATACH) investigators. “Antihypertensive treatment of acute cerebral hemorrhage *. [Miscellaneous Article]”. In: *Critical Care Medicine February 2010* 38.2 (2010), pp. 637–648.
- [15] 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.
- [16] Christopher Rorden et al. “Age-specific CT and MRI templates for spatial normalization”. In: *NeuroImage* 61.4 (July 16, 2012), pp. 957–965.
- [17] 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.

- [18] 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.
- [19] 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).
- [20] T. Brott et al. “Measurements of acute cerebral infarction: a clinical examination scale.” In: *Stroke* 20.7 (July 1, 1989), pp. 864–870.
- [21] 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.
- [22] G. Teasdale and B. Jennett. “Assessment and prognosis of coma after head injury”. In: *Acta Neurochirurgica* 34.1 (Mar. 1, 1976), pp. 45–55.
- [23] 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), pp. 627–634.
- [24] M Bergstrom et al. “Variation with time of the attenuation values of intracranial hematomas”. In: *Journal of computer assisted tomography* 1.1 (1977), pp. 57–63.
- [25] 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.
- [26] John Ashburner and Karl J. Friston. “Unified segmentation”. In: *NeuroImage* 26.3 (July 1, 2005), pp. 839–851.
- [27] John Muschelli III. “brainR: Interactive 3 and 4d Images of High Resolution Neuroimage Data”. In: *R Journal* 6.1 (2014).
- [28] Hirotugu Akaike. “Information theory and an extension of the maximum likelihood principle”. In: *Second international symposium on information theory*. Akademinai Kiado, 1973, 267281.
- [29] 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.
- [30] John Rankin. “Cerebral vascular accidents in patients over the age of 60. II. Prognosis.” In: *Scottish medical journal* 2.5 (1957), 200215.
- [31] J. C. van Swieten et al. “Interobserver agreement for the assessment of handicap in stroke patients.” In: *Stroke* 19.5 (May 1, 1988), pp. 604–607.
- [32] A. Eloyan et al. “Health effects of lesion localization in multiple sclerosis: Spatial registration and confounding adjustment.” submitted.
- [33] Chris Rorden and Matthew Brett. “Stereotaxic Display of Brain Lesions”. In: *Behavioural Neurology* 12.4 (2000), pp. 191–200.
- [34] Stephen M. Smith. “Fast robust automated brain extraction”. In: *Human Brain Mapping* 17.3 (2002), 143155.
- [35] Mark Jenkinson et al. “FSL”. In: *NeuroImage* 62.2 (Aug. 15, 2012), pp. 782–790.

7 Figure Legends

Figure 1. ICH engagement prevalence. A non-spatial histogram of proportion of patients ((A)) show the majority of voxels have a low prevalence (median: 3%), while some voxels ($V = 5685$) have a high prevalence over 40%. Voxels with no prevalence are excluded. In ((B)), these proportions are represented in a 3D histogram (right side of image is left side of brain) overlaid on an MRI T1 template. More hemorrhages are in the left side of the brain. ICH is also more localized in the middle of the brain, with few extensions in the anterior and posterior areas. The interactive version of this figure is located at http://muschellij2.github.io/CT_Pipeline/index.html.

Figure 2. P-value maps for the voxel-wise NIHSS models. In the unadjusted (panel (A)) or sex-adjusted (not shown) linear models, voxels with the smallest p-values appear medial near the lateral ventricles. In age-adjusted (panel (B)), total baseline ICH volume-adjusted (not shown), or age and volume-adjusted (panel (C)), the p-values relating to NIHSS scores appear higher (more bluish). The Wilcoxon rank-sum test ((D)) have smaller p-values for the same area compared to the unadjusted linear model.

Figure 3. Highest Predictive Region (HPR) Analysis. Panels (A) and (B) correspond to the HPR for the top-performing model for NIHSS and GCS scores. The HPR in ((A)) represents a p-value threshold of .0100 (19047 voxels) for the voxel-wise p-value of ICH on NIHSS. The HPR in ((B)) represents 1000 with the lowest p-values for the voxel-wise ICH on GCS score regressions, corresponding to a p-value threshold of .0002. Panels (C) and (D) plot the relationship of the HPR coverage and severity score. The red line represents a linear fit and the blue line represents a LOESS fit. The larger the HPR coverage the higher (more severe stroke) NIHSS score and the lower (deeper unconsciousness) the GCS score.

8 Tables

Variable (N = 111)	N (%) or Mean (SD)
Age in Years: Mean (SD)	60.8 (11.2)
Sex: Female	35 (31.5%)
NIHSS Score: Mean (SD)	22.1 (8.7)
GCS Score: Mean (SD)	10.0 (3.0)
ICH Volume: Mean (SD)	37.4 (20.1)
IVH Volume: Mean (SD)	3.2 (6.3)
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 Pallidus	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
NIHSS Score					
Location Model		0.129	0.178	18.60	8.116
1000	.0005	0.236	0.265	2.47	7.598
2000	.0009	0.234	0.263	2.81	7.610
2422	.0010	0.247	0.275	0.98	7.545
3000	.0013	0.244	0.272	1.46	7.562
19047	.0100	0.254	0.282	0.00	7.511
47736	.0500	0.248	0.276	0.77	7.538
GCS Score					
Location Model		0.069	0.120	20.51	2.914
1000	.0002	0.214	0.243	0.00	2.677
2000	.0004	0.213	0.242	0.09	2.678
3000	.0006	0.212	0.241	0.25	2.680
4669	.0010	0.212	0.241	0.23	2.680
22858	.0100	0.191	0.221	3.17	2.716
52368	.0500	0.166	0.197	6.44	2.757

Table 2. Table of model-fit measures for GCS and NIHSS scores

Area	Population Prevalence	NIHSS HPR	GCS HPR
CSF	7.9	10.9	4.2
Insular	7.6		
Superior temporal gyrus	5.5		
Putamen	4.8	4.3	
External capsule	3.9		
Superior corona radiata	3.7	11.0	27.9
Precentral gyrus	3.3		
Precentral WM	3.1		1.3
Superior temporal WM	3.1		
Posterior limb of internal capsule	3.0	12.0	3.9
Thalamus		10.1	33.9
Caudate nucleus		8.4	9.6
Anterior limb of internal capsule		6.8	
Globus pallidus		6.0	
Superior longitudinal fasciculus		4.5	5.9
Outside brain mask		3.6	
Postcentral WM			6.7
Posterior corona radiata			3.1
Supramarginal WM			1.1

Table 3. Distribution of the top 10 areas of engagement

9 Figures

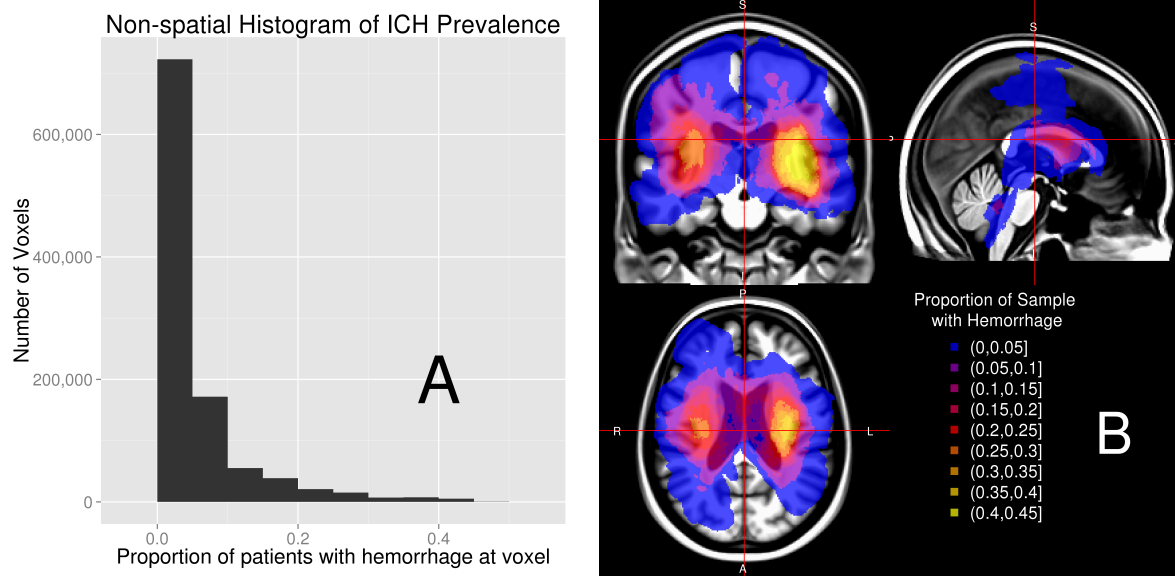


Figure 1

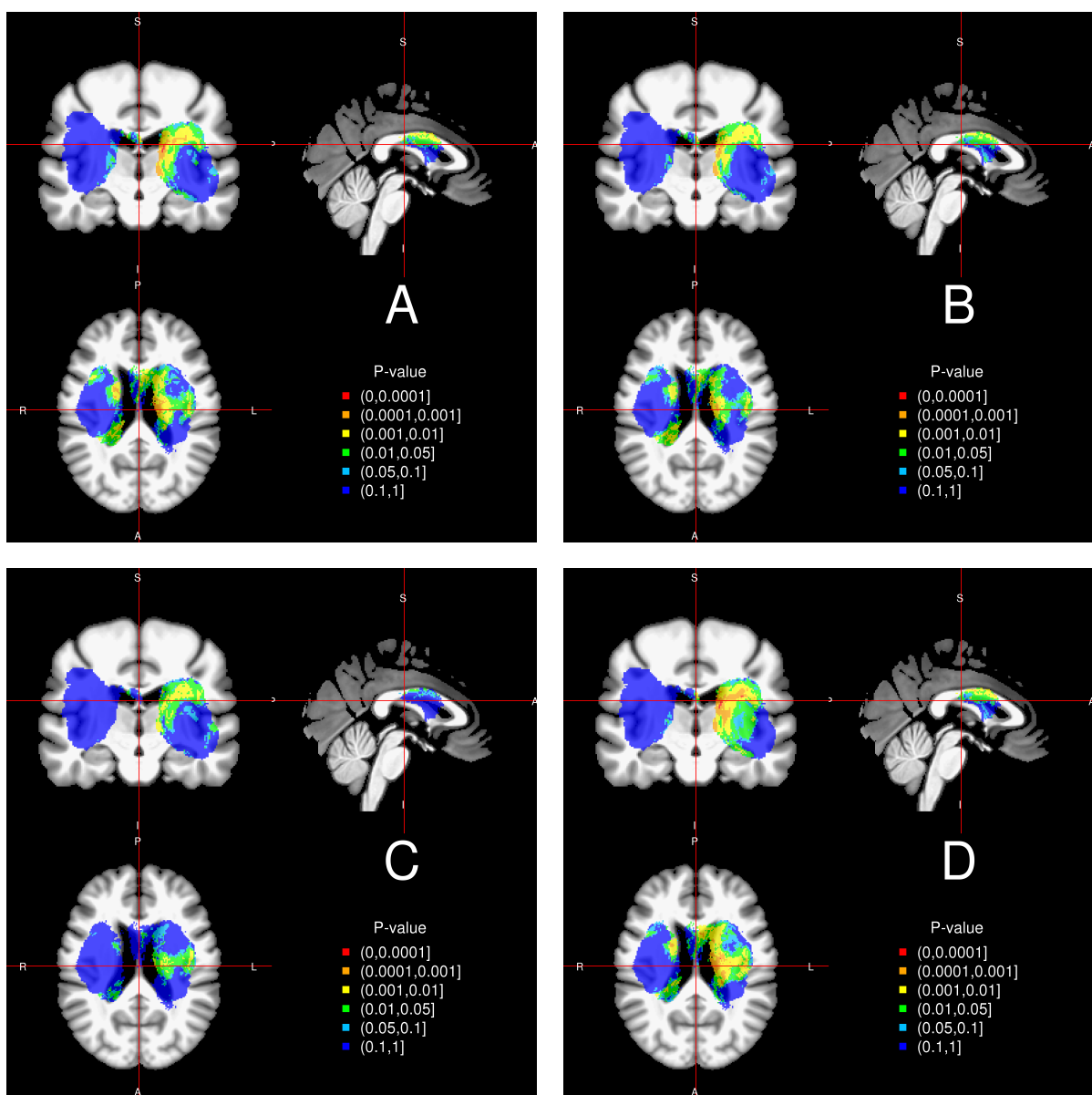


Figure 2

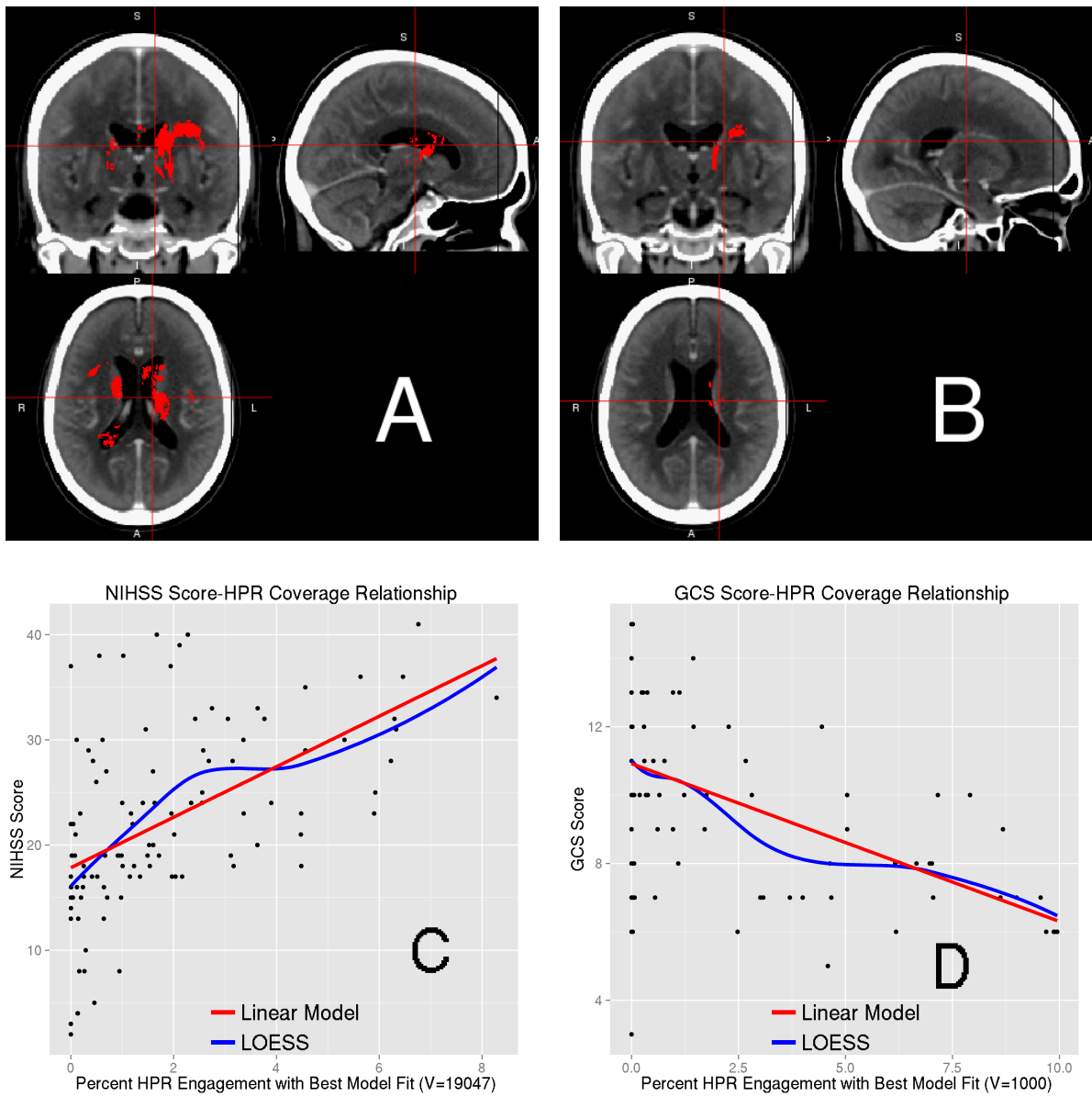


Figure 3

10 Supplemental Material

10.1 Supplemental Methods

10.1.1 Image Processing: Brain Extraction, Reorientation, Registration

The hemorrhage was excluded when registering the CT scan to the template using manual ICH segmentations. Binary ICH masks had voxels set 1 if the voxel was hemorrhage, and 0 otherwise.

CT images were processed as follows:

1. Export from OsiriX to DICOM format
2. Gantry tilt corrected (if applicable) using MATLAB script (<http://bit.ly/11tIM8c>)
3. Converted to the Neuroimaging Informatics Technology Initiative (NIfTI) format using `dcm2nii` (2009 version, MRIcro [33])
4. Brain extraction tool (BET) [34], a function of FSL (v5.0.4) [35] was applied
5. The data was aligned to the anterior-posterior commissure line (<http://bit.ly/1gUqMDw>).
6. The Clinical toolbox [16] was applied.

An image with its brain-extracted counterpart is displayed in Supplemental Figure IB.

We thresholded the smoothed hemorrhage mask: let s_{ij} be the smoothed mask for person i , and s_{ij} be voxel j of that mask; let v_{ij} represents voxel j of the thresholded image. The was thresholded using the rule [16]:

$$v_{ij} = \begin{cases} 1 & \text{if } s_{ij} \geq \frac{\min(S_i) + \max(S_i)}{2} \\ 0 & \text{if } s_{ij} < \frac{\min(S_i) + \max(S_i)}{2} \end{cases}$$

These binary ICH masks were used in analysis.

10.1.2 Calculating Region Engagement from the Eve Atlas

We also calculated the percent engagement of regions for the best-performing HPR for the NIHSS and GCS score analyses. More explicitly, let k denote the brain region (e.g. Putamen) and let $\sum_k v_k$ represent the sum of the voxels for an image (HPR or population ICH image) in that brain region. These p_k represent the percent that brain region engages the ICH compared to other regions (Table 3):

$$p_k = \frac{\sum_k v_k}{\text{Sum of Image}}$$

We also calculated brain region engagement with the population ICH or HPR images (Table II):

$$r_k = \frac{\sum_k v_k}{\text{\#Number of Voxels in Region}}$$

These percentages (r_k) are at a region level; p_k are at an image level.

10.1.3 Image Registration

We present the registration results (Figure I): manually segmented blood in the original space with the hemorrhage mask in pink (panel A), brain-extracted image (panel B), registered image and hemorrhage mask in template space (panel C), and the ICH mask on the template (panel D). These images represent one patient with variable slice thickness with a large hemorrhage.

Though variable slice thickness is present, the transformation morphs the image into the full space of the template; therefore, non-linear registration seems to reasonably account for variable slice thickness, likely by non-uniform scaling. We also see gross brain features remain relatively unchanged, but large deformations of tissue, due to ICH, appear well preserved by registration.

[Figure I here.]

10.2 Supplemental Tables

	NIHSS Score		GCS Score	
	HPR Coverage	Reader-Based	HPR Coverage	Reader-Based
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)
HPR Coverage per 10%	2.0 (1.1, 2.8)		-0.4 (-0.6, -0.2)	
Reader-Based Location				
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 I. Severity Score Regression Models for HPR-Based and Reader-Classified Location

Area	Population Engagement	NIHSS HPR	GCS HPR
Globus Pallidus: Total	20.3	40.0	0.0
Globus Pallidus: Right	14.8	34.8	0.0
Globus Pallidus: Left	25.2	44.7	0.0
Putamen: Total	23.3	6.6	0.0
Putamen: Right	17.5	3.8	0.0
Putamen: Left	29.2	9.4	0.0
Thalamus: Total	7.9	8.9	1.7
Thalamus: Right	6.8	3.1	0.0
Thalamus: Left	9.1	14.6	3.4

Table II. Thalamus, Putamen and Globus Pallidus ICH Engagement by the population and NIHSS/GCS HPR

10.3 Supplemental Figures and Figure Legends

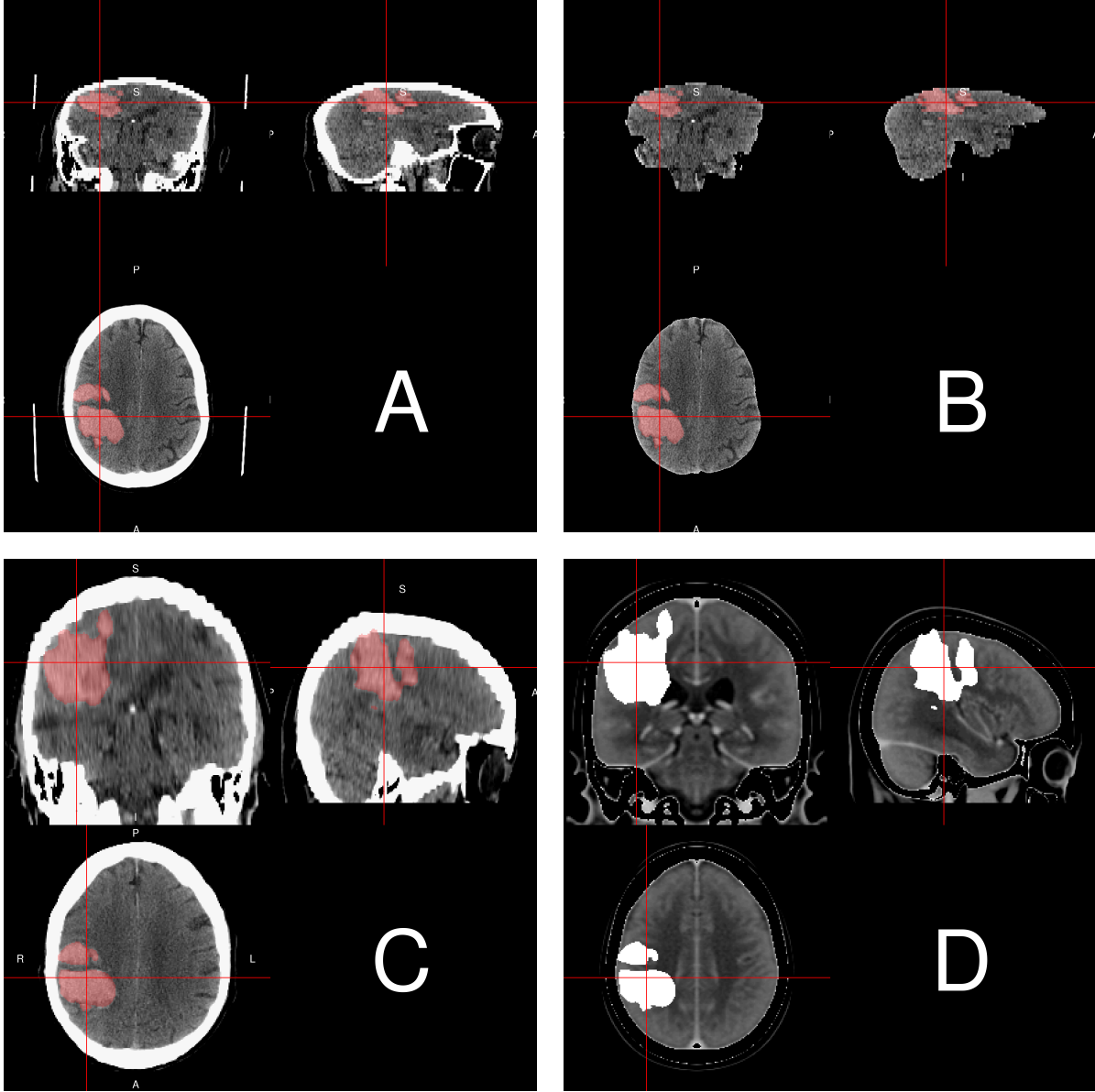


Figure I. Steps of Image Processing. The native space image (A), with hemorrhage overlaid in pink, shows variable slice thickness: the top of the brain has smaller voxel sizes (thinner slices) than the bottom of the brain. The brain-extracted image used for AC-PC alignment is in panel (B). The template-registered image ((C)) shows 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. Panel (D) shows the overlaid hemorrhage mask on the CT template.

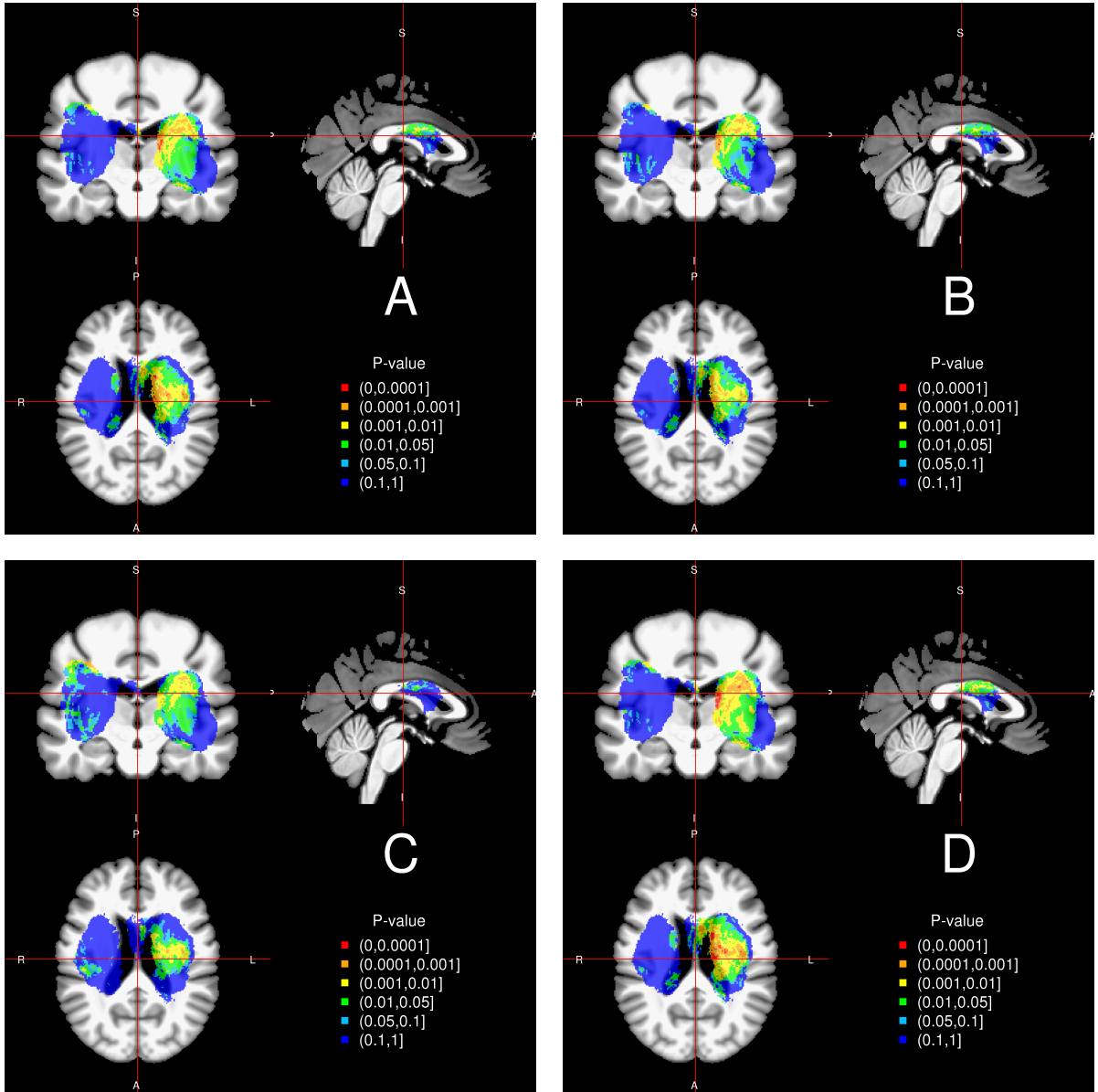


Figure II. P-value maps for the voxel-wise GCS models. In the unadjusted linear model (panel (A)) and after adjusting for sex (not shown, similar to (A)) there are voxels with the smallest p-values seem to be near the lateral ventricles. In models adjusting for age ((B)), or total baseline ICH volume (not shown, similar to (D)), or both ((C)), the p-value relating to GCS scores appear higher (more bluish). Also, we see the Wilcoxon rank-sum test have smaller p-values for the same area compared to the unadjusted linear model ((D)).

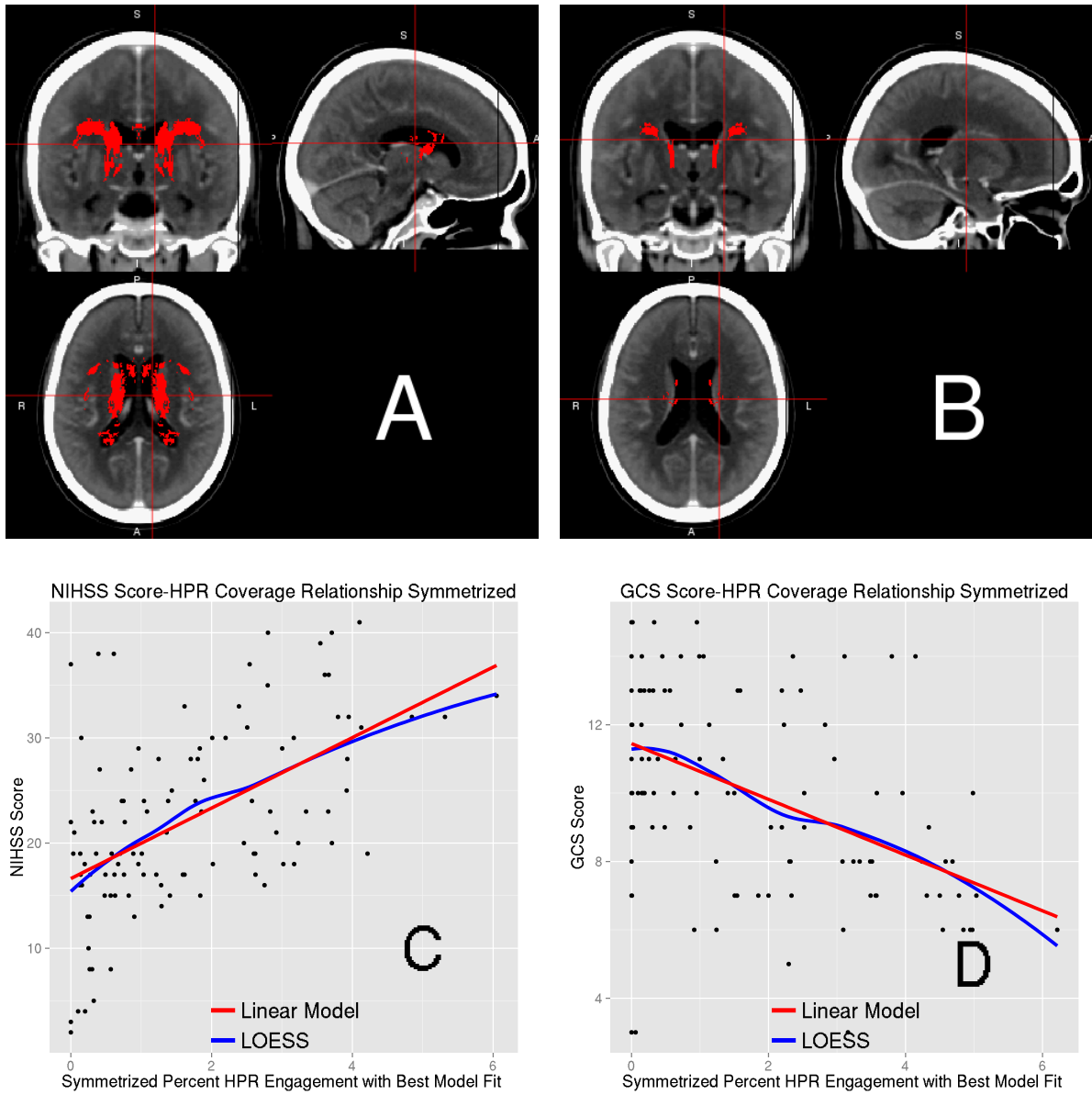


Figure III. Highest Predictive Region (HPR) Analysis with Symmetrized HPR. HPR were symmetrized by including voxels in the HPR and the voxel on the contra-lateral side. Panels (A) and (B) correspond to the symmetrized HPR for NIHSS (Figure 3(A)) and GCS (Figure 3(B)), respectively. Panels (C) and (D) plot the symmetrized HPR coverage–severity score relationship. The larger the coverage the higher (more severe stroke) NIHSS score and the lower (deeper unconsciousness) the GCS score.