Longitudinal Lesion Expansion in Chronic Traumatic Brain Injury

Holly J. Freeman,[1] Alexander S. Atalay,[1] Jian Li,[1,2] Evie Sobczak,[1] Samuel B. Snider,[3] Holly Carrington,[4] Enna Selmanovic,[4,5] Ariel Pruyser,[4] Lisa Bura,[4] David Sheppard,[6] David Hunt,[7] Alan C. Seifert,[8] Yelena G. Bodien,[1,9] Jeanne M. Hoffman,[6] Christine L. Mac Donald,[7] Kristen Dams-O'Connor[4,10,*], Brian L. Edlow[1,2,*]

*co-senior authors

Affiliations
[1]Center for Neurotechnology and Neurorecovery, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
[2]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
[3]Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
[4]Department of Rehabilitation and Human Performance, Icahn School of Medicine at Mount Sinai, New York, NY
[5]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY
[6]Department of Rehabilitation Medicine, University of Washington School of Medicine, Seattle, WA
[7]Department of Neurological Surgery, University of Washington, Seattle, WA
[8]Biomedical Engineering and Imaging Institute, Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, NY
[9]Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, and Harvard Medical School, Charlestown, MA
[10]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY

Funding:
This study was supported by the NIH National Institute of Neurological Disorders and Stroke (RF1NS128961R01NS128961, RF1NS115268, U01NS086625), NIH Director's Office (DP2HD101400), and the Chen Institute MGH Research Scholar Award.

Correspondence:
Holly J. Freeman
101 Merrimac Street – Suite 310, Boston, MA 02114
Phone: (617) 724-6352
Email: hjfreeman@mgh.harvard.edu

Declarations:
Ethics approval This study was approved by the Institutional Review Boards at Mount Sinai School of Medicine and University of Washington School of Medicine.
Informed consent Informed consent was provided by participants or their surrogates, if participants lacked decision-making capacity.
Conflicts of interests None

Abstract

Traumatic brain injury (TBI) is a risk factor for neurodegeneration and cognitive decline, yet the underlying pathophysiologic mechanisms are incompletely understood. This gap in knowledge is in part related to the lack of analytic methods to account for cortical lesions in prior neuroimaging studies. The objective of this study was to develop a lesion detection tool and apply it to an investigation of longitudinal changes in brain structure among individuals with chronic TBI. We identified 24 individuals with chronic moderate-to-severe TBI enrolled in the Late Effects of TBI (LETBI) study who had cortical lesions detected by T1-weighted MRI at two time points. Initial MRI scans were performed more than 1-year post-injury and follow-up scans were performed 3.1 (IQR=1.7) years later. We leveraged FreeSurfer parcellations of T1-weighted MRI volumes and a recently developed super-resolution technique, SynthSR, to identify cortical lesions in this longitudinal dataset. Trained raters received the data in a randomized order and manually corrected the automated lesion segmentation, yielding a final lesion mask for each scan at each timepoint. Lesion volume significantly increased between the two time points with a median volume change of 3.2 (IQR=5.9) mL ($p<0.001$), and the increases significantly exceeded the possible variance in lesion volume changes due to manual tracing errors ($p < 0.001$). Lesion volume significantly expanded longitudinally in 23 of 24 subjects, with all FDR corrected p-values ≤ 0.02. Inter-scan duration was not associated with the magnitude of lesion growth. We also demonstrated that the semi-automated tool showed a high level of accuracy compared to “ground truth” manual lesion segmentation. Semi-automated lesion segmentation is feasible in TBI studies and creates opportunities to elucidate mechanisms of post-traumatic neurodegeneration.

Introduction

Traumatic brain injury (TBI) is a well-established risk factor for neurodegenerative diseases (Dams-O'Connor et al., 2020). The pathophysiologic mechanisms that link TBI to post-traumatic neurodegeneration (PTND) are not fully understood, though emerging evidence implicates a "polypathology" (Kenney et al., 2018) that includes axonal injury (Johnson et al., 2013), tau deposition (McKee et al., 2013), vascular injury (Dams-O'Connor et al., 2023; Sandsmark, Bashir, Wellington, & Diaz-Arrastia, 2019), and neuroinflammation (Johnson et al., 2013). An underexplored factor in the pathogenesis of PTND is the potential impact of focal lesions, such as cerebral contusions, which are amongst the most common lesions in individuals with TBI (Vande Vyvere et al., 2024). It is unknown whether focal lesion size or pathophysiologic characteristics evolve over time, and if so, whether this may have implications for clinical decline.

In addition to a paucity of longitudinal studies of chronic to moderate-to-severe TBI survivors, a primary barrier to elucidating the impact of lesions on PTND pathogenesis is methodological. Historically, lesions that disrupt the surface of the cerebral cortex have prevented MRI segmentation tools from segmenting the brain into its anatomic components (Merkley et al., 2008; Santhanam, Wilson, Oakes, & Weaver, 2019; Strangman et al., 2010). As a result, segmentation tools distributed with imaging analysis programs such as FreeSurfer (Fischl, 2012), FSL (Smith et al., 2004), and SPM (Friston et al., 1994) have been unable to robustly measure longitudinal lesion growth. Hence, patients with cortical lesions have typically been excluded from studies of cortical and subcortical volumetrics in individuals with TBI (Ding et al., 2008; Warner et al., 2010). Moreover, efforts at lesion segmentation have required substantial time by operators trained in human neuroanatomy (Diamond et al., 2020). In this context, the systematic study of lesions their role in PTND has not been performed.

To address this methodological barrier and knowledge gap, we performed a longitudinal MRI study of individuals with chronic TBI and leveraged recent innovations in machine learning image analysis (Iglesias et al., 2023; Iglesias et al., 2021) to create a new, semi-automated lesion segmentation tool. We tested the ability of this semi-automated lesion tool to detect longitudinal changes in lesion volume in individuals with chronic TBI enrolled in the Late Effects of TBI (LETBI) study (Edlow et al., 2018).

Methods

Participant selection

Between 2014 and 2023, a total of 305 participants were enrolled in the ongoing LETBI study (Edlow et al., 2018). Criteria to be included in the present longitudinal analysis required participants to have undergone two scanning sessions during consecutive study visits (≥2 years apart), each producing T1-weighted (T1w) multi-echo magnetization prepared gradient-recalled echo (MEMPRAGE) scans (van der Kouwe, Benner, Salat, & Fischl, 2008) with a resolution of 1 mm isotropic. Based on these criteria, 249 participants were excluded (n = 225 without consecutive timepoints, n = 24 due to sequence mismatch between timepoints). Of the remaining participants, 56 met these criteria and their T1w images were visually assessed by a trained rater as having lesions on both longitudinal imaging sessions (**Figure 1)**. Lesions were defined by

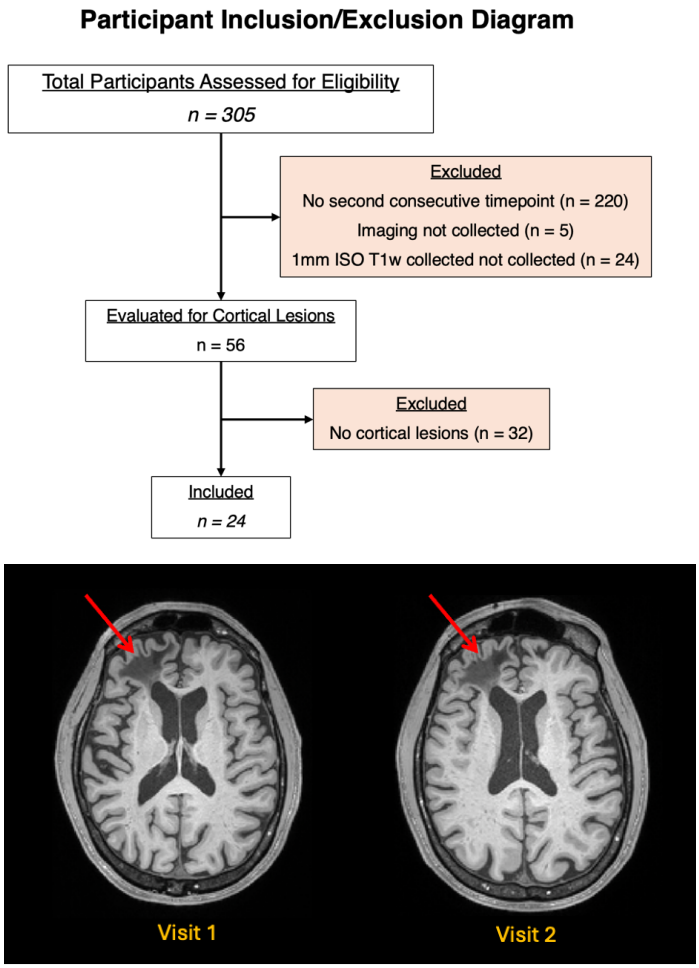

Figure 1: Evaluation of cohort for eligible participants (top). Example of visual review of the T1w images of both timepoints to evaluate for lesion presence. Lesioned area is marked with red arrow.

visible disruptions in grey matter, white matter, or the grey/white junction. These included areas showing ongoing demyelination or asymmetry compared to the opposite hemisphere (where applicable). All participants identified with a lesion in the first visit, also had a lesion identified in the second visit. After excluding 32 participants without cortical lesions, the final cohort thus consisted of 24 participants.

Data acquisition, quality assessment, and processing

T1w images were obtained using Siemens Skyra, Philips Achieva, and Philips Ingenia Elition X scanners, all operating at 3 Tesla field strength. The images were acquired at 1 mm isotropic resolution. Siemens Skyra scans used a repetition time (TR) of 2530 ms and echo times (TE) ranging from 1.79 ms to 7.37 ms. Philips Achieva scans used TRs ranging from 2530 ms and TEs ranging from 1.67 to 7.07 ms. Philips Ingenia Elition X scans used a TR of 2530 ms and a TE of 2.14 ms. Further information about the number of scans obtained from each scanner is provided in **Table 1.**

Table 1. MRI acquisition parameters for study participants

Manufacturer	Model	Field strength (T)	TR/TE (msec/msec)	Number of scans
Siemens	Skyra	3T	2530/1.79 – 7.37	18
Philips	Achieva	3T	2530 / 1.67 – 7.07	11
Philips	Ingenia Elition X	3T	2530 / 2.14	19

Despite these variations in sequence parameters, efforts were made to ensure uniformity and comparability across both subjects and scanning platforms. Additional sequence parameters for the T1w sequences on each scanner have been previously reported (Edlow et al., 2018).

Qualitative and quantitative data quality assessments were performed on the processed images of all 24 longitudinal lesion subjects at both timepoints. Visual quality assessments were based upon accuracy of FreeSurfer-generated surfaces (excluding those encompassing lesioned tissue) and segmentation of subcortical structures. Signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) were measured using the FreeSurfer tools 'wm-anat-snr' and 'mri_cnr', calculating SNR in white-matter and the average of the WM-GM and GM-CSF contrasts, respectively. While no subjects were excluded due to quality assessment measures, differences were observed between the SNR distributions of enrollment sites, as reported in **Table 2**.

The T1w images were then processed and the surfaces constructed using FreeSurfer v7.4 (Fischl, 2012). FreeSurfer processing involves motion correction, averaging of T1w images, removal of non-brain tissue, automated Talairach transformation, and segmentation of brain structures. It also includes intensity normalization, gray/white matter boundary tessellation, and topology correction. Further steps involve surface deformation, surface inflation, spherical atlas registration, cortical parcellation, and creation of curvature and sulcal depth maps. To robustly segment neuroanatomic structures in brains with heterogenous pathology, we used the Sequence Adaptive Multimodal SEGmentation (SAMSEG) tool (Cerri et al., 2021; Puonti, Iglesias, & Van Leemput, 2016) instead of the default automated segmentation (aseg) tool before FreeSurfer recon-all. The FreeSurfer reconstructions for all participants were completed successfully.

"Ground truth" segmentation

Ground truth segmentations for all participants were established through manual tracing performed by a neurologist who was blinded to subject identification and time point. The process involved loading each T1w image into the FreeSurfer image viewer, Freeview. A blank label volume was created using the same geometry as the T1w image. The neurologist then manually segmented each lesioned area using the voxel edit tool, ensuring accurate and detailed delineation of the lesions. All segmentations were performed on the same label volume, thus creating a single ground truth segmentation volume for each timepoint for all participants.

Table 2. Quantitative quality assessment of MRI data at each enrollment site.

Enrollment site	Longitudinal cohort	SNR	CNR
MSSM	n=9	23.30 +/- 5.73	0.93 +/- 0.19
UW	n=15	39.29 +/- 12.21	0.75 +/- 0.14
All sites	n=24	33.30 +/- 12.72	0.82 +/- 0.18

Longitudinal analyses for each time point were performed in the subject's native space. This method was selected instead of using the FreeSurfer longitudinal pipeline, which combines the two time points to generate a base image (Reuter, Schmansky, Rosas, & Fischl, 2012). The averaging process in the FreeSurfer pipeline would obscure the examination of lesion progression by blending the time points together, thus failing to capture dynamic changes in lesion size and location. By performing analyses in native space, we maintain the integrity of individual time point

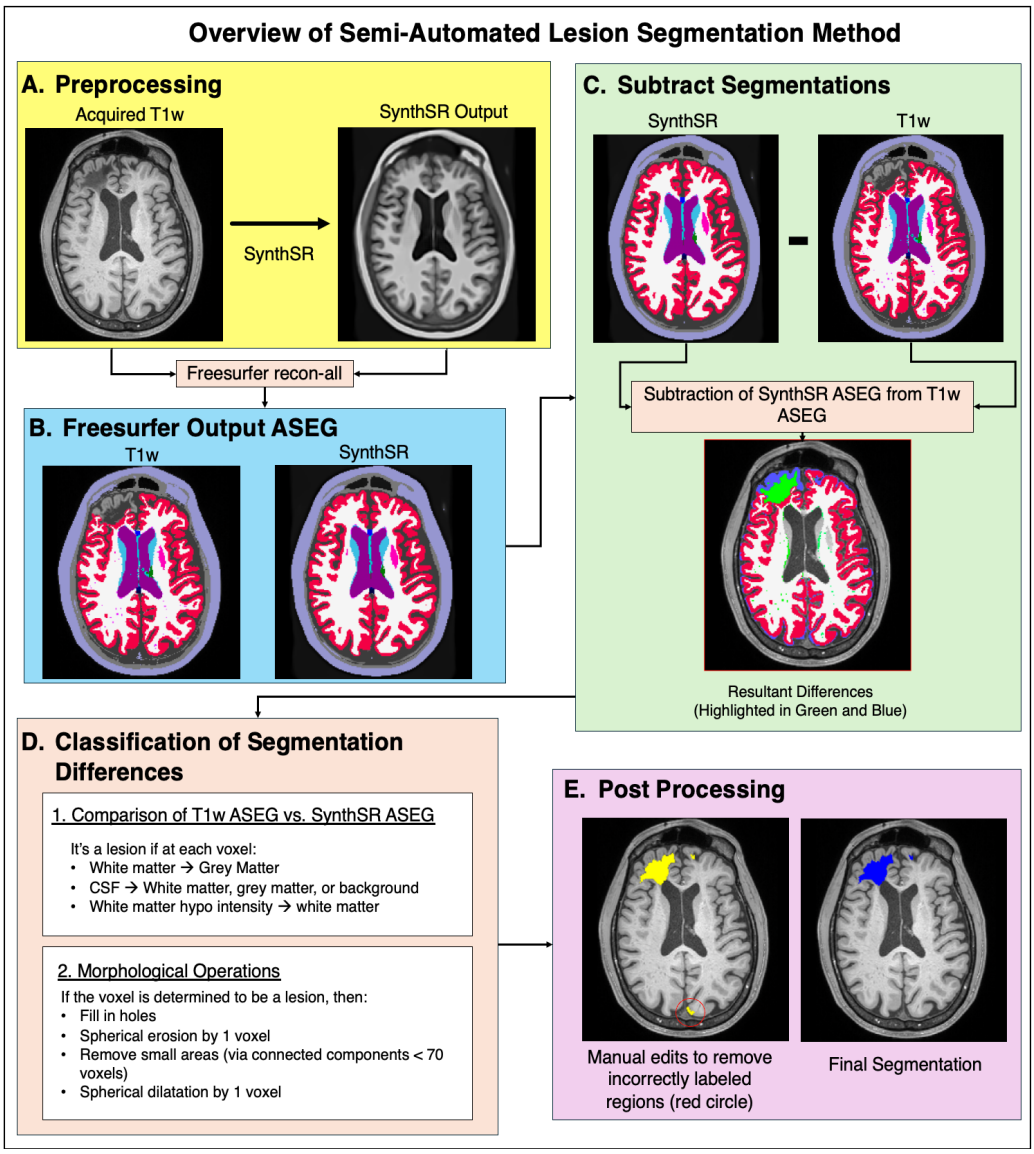

Figure 2: Overview of semi-automated lesion segmentation method. SynthSR images are generated for each acquired T1w image (A). Both images are then processed through Freesurfer recon-all, resulting in ASEG label volumes (B). The SynthSR ASEG is subtracted from the acquired T1w ASEG to highlight segmentation differences (C). These differences are classified voxel by voxel based on specific rules comparing the SynthSR and acquired T1w segmentation volumes. Identified lesions are then refined using morphological operations (D). Finally, the cleaned segmentation is reviewed for errors, including incorrectly labeled anatomy or missed lesions, and corrected to produce the final modified, semi-automated segmentation (E).

data, allowing for precise tracking of lesion growth and development over the study period without introducing registration artifacts.

Semi-automated lesion segmentation

To minimize time requirements and reduce false negatives (i.e., missed labeling) in manual tracing, we developed a novel method for semi-automated lesion segmentation. As illustrated in **Figure 2A**, we leveraged SynthSR (Iglesias et al., 2023; Iglesias et al., 2021), a publicly available tool (integrated within FreeSurfer) that turns a clinical MRI scan of any orientation, resolution and contrast into a 1 mm isotropic T1w image while inpainting lesions.

We applied SynthSR to T1w images for all participants, and we then repeated the FreeSurfer recon-all process on the synthesized images. We define lesional areas by comparing the SAMSEG labels from the synthesized image with those from the original T1w image (**Figure 2C**) using the following rules: at each voxel, it is a lesion if 1) the segmentation label changed from white matter (in the original T1w recon) to gray matter (in the SynthSR recon); or 2) from CSF to background/white matter/gray matter; or 3) from white matter hypo-intensity to white matter. These rules were determined heuristically based on the segmentation label changes inside the lesional areas from a subset of our samples. Subsequently, we applied morphological image processing (Soille, 2004) to remove false positives, reduce noise, and ensure the detected lesional areas are topologically correct, including hole filling, spherical erosion/dilation, and area opening. Successful application of the above-described pipeline facilitated the isolation of significant changes between the SynthSR-impainted volume and the original volume, yielding an initial automated lesion segmentation mask (**Figure 2E**). We then performed manual edits to enhance the accuracy of lesion segmentation boundaries, yielding a final semi-automated lesion mask.

Longitudinal changes in lesion volume

We hypothesized that there are detectable changes in lesion volume size when comparing Visit 1 to Visit 2 data (**Figure 3)**. Using the ground truth segmentation volumes to assess the statistical significance of changes in lesion volume between the visits, we examined the number of non-zero voxels in Visit 1 versus the number of non-zero voxels in Visit 2. We calculated the differences (Visit 2 - Visit 1) for each pair of measurements taken from the ground truth segmentation volumes. Statistical analysis was performed using the Wilcoxon Signed Rank test. This approach allowed for a direct comparison of how volumes varied between the two time points.

Evaluation of inter-rater reliability and null hypothesis construction

To establish inter-rater reliability for lesion segmentation, 10 randomly selected T1w images were analyzed by three raters who were blinded to the identity of the participant and the timepoint of the scan. Raters were provided with SynthSR-generated segmentation masks (i.e., the raw automated lesion mask, see **Figure 2E**) and instructed to revise the segmentations, creating the final semi-automated lesion mask. The raters' lesion masks were then compared against each

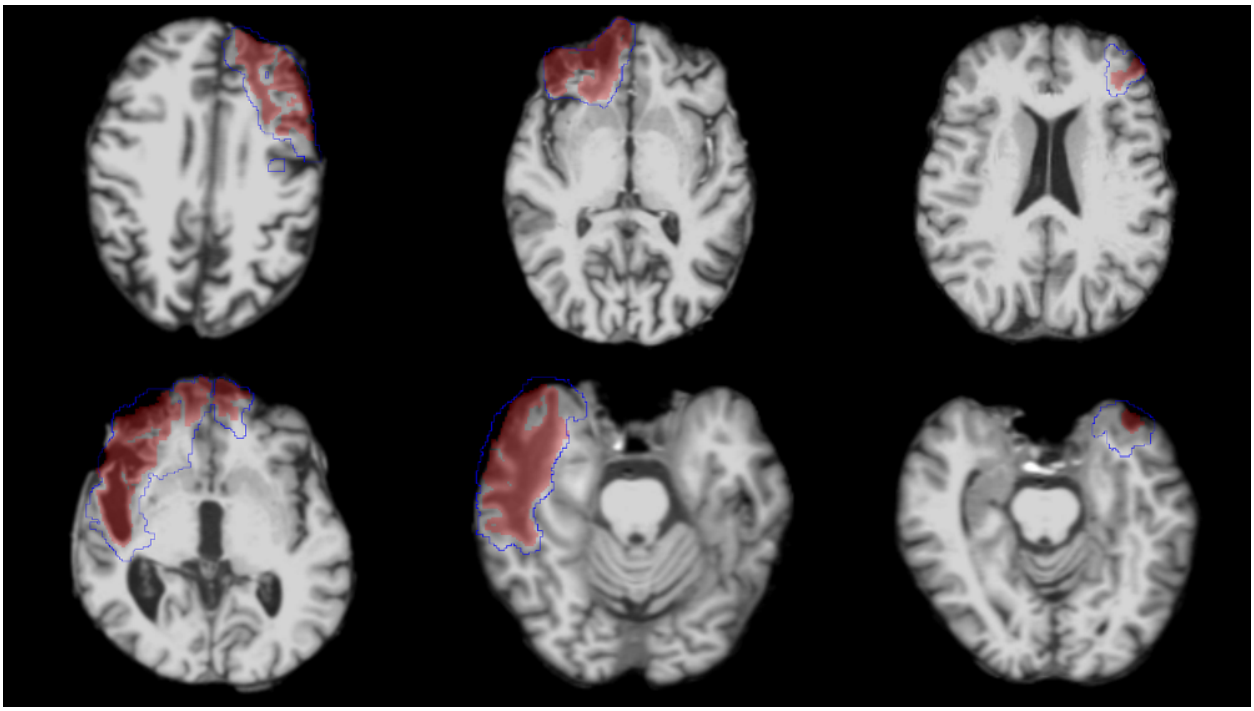

Figure 3: T1w images from Visit 1 of six representative subjects demonstrating the heterogeneous nature of the lesion size and varying locations. Images are overlayed with the ground truth lesion segmentations from visit 1 (red shaded regions) and visit 2 (blue outlined regions).

other to measure inter-rater variability. Specifically, we calculated the Dice coefficient between lesion masks from each pair of raters: Rater 1 versus Rater 2, Rater 1 versus Rater 3, and Rater 2 versus Rater 3. We aggregated the Dice scores from each pair and combined them into a single comprehensive list, encapsulating the full range of variability among all three raters. These Dice scores form a null distribution for statistical testing of the lesion expansion.

Evaluation of lesion expansion compared to inter-rater variability

To test if the longitudinal changes in lesion size is beyond the possible variability due to manual tracing errors, we measured Dice score between Visit 2 and Visit 1 for each subject, whereby larger expansion leads to a lower Dice score. We hypothesized that longitudinal lesion volume changes are larger in magnitude than inter-rater variability in lesion tracing (i.e., we expect that the Dice scores between two longitudinal time points are significantly lower than those in the null distribution derived from inter-rater testing). We tested this hypothesis using the Wilcoxon Rank Sum test, with a significance level of 0.05, to minimize the impact from outliers and the small sample size. To account for multiple comparisons, the p-values were adjusted using the False Discovery Rate (FDR).

Evaluation of lesion expansion as a function of time between scans

We examined the relationship between changes in lesion size (measured in voxels) and the interval between imaging sessions (measured in days). We determined the longitudinal change in lesion size by subtracting ground truth lesion size at Visit 1 from lesion size at Visit 2 for each

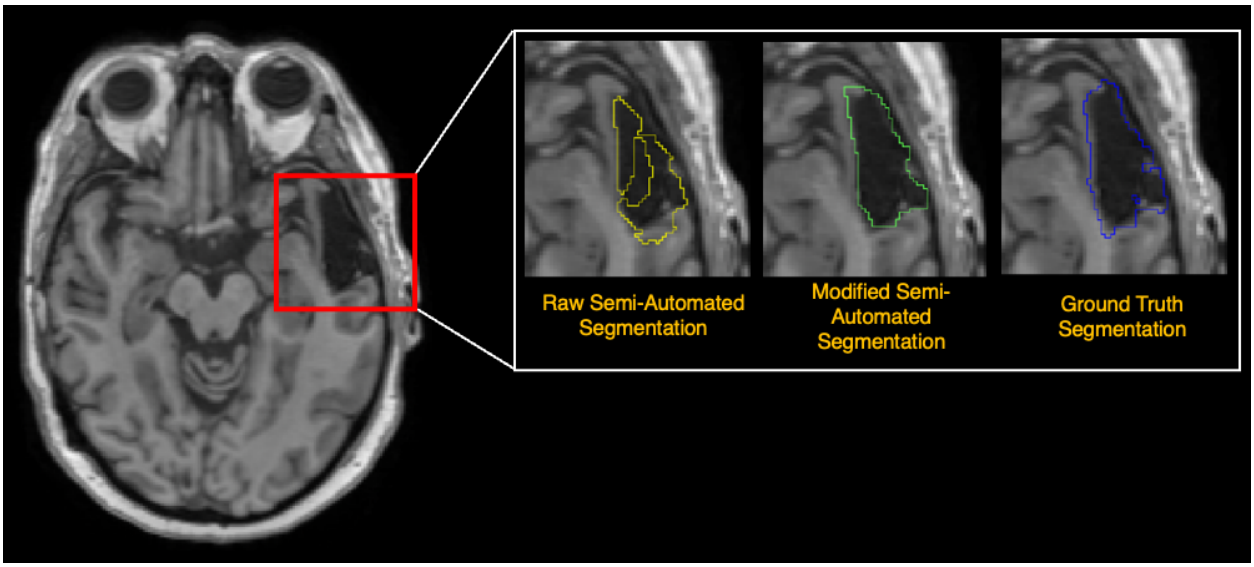

Figure 4: Zoomed in comparison of the raw output from semi-automated lesion segmentation method, against the modified semi-automated segmentation, which was revised by a trained rater, and the manually traced "ground truth" segmentation volume.

subject. Pearson correlation coefficient (R) and two-tailed p-value were computed to assess the strength and significance of any linear relationship between changes in lesion size and duration between study visits. We applied Ordinary Least Squares (OLS) regression analysis to further investigate how age, sex and interval between study visits relate to changes in lesion volume.

Comparison of semi-automated segmentation to ground truth measurements

To evaluate the accuracy of the semi-automated segmentations, trained raters made edits to each raw output of the semi - automated segmentation, as shown in **Figure 4**, creating modified semi-automated segmentation volumes.

An additional aim of the present study was to evaluate the performance of the semi-automated lesion segmentation tool by comparing the necessary edits (see examples in **Figure 2E** and **Figure 4)** to the raw segmentation volumes with those of the manually edited volumes. We compared the raw and modified segmentations the ground truth at both time points using Wilcoxon ranked-sum tests.

To analyze the variance between the modified semi-automated and ground truth segmentations, we evaluated dice scores and non-zero voxel counts. We further assessed this variance using Mean Absolute Error (MAE) and Pearson correlation coefficients. Lastly, we calcula we calculated dice scores for ground and semi-automated segmentations at both timepoints and analyzed them with the Wilcoxon ranked-sum test to determine if there were statistically significant differences between the methods.

Lastly, we calculated dice scores for ground and semi-automated segmentations at both timepoints and analyzed them with the Wilcoxon ranked-sum test to determine if there were statistically significant differences between the methods.

Results

Patient Characteristics

The 24 participants ranged in age from 33 to 73 years old, with a median age of approximately 57 years (IQR = 13.5). Nineteen were males. Full descriptive statistics information can be found in **Table 3.**

Table 3. Patient characteristics

n=24	M	SD	Range
Age (Years)	55.08	11.56	33 - 73
Visit Interval (Days)	1328.50	498.74	734-2366
Gender (Male:Female)	19:4		

Longitudinal changes in absolute lesion volume

Lesion segmentations were assigned voxel values of 1. Lesion sizes derived from ground truth segmentations at Visit 1 and Visit 2 ranged from 2,459 to 104,334 non-zero segmentation voxels. The Wilcoxon signed-rank test results yielded a statistic of 23.0 and a corresponding p-value of 0.00007, indicating a statistically significant difference in lesion volume between Visit 1 and Visit 2. The magnitude of longitudinal lesion volume change for each individual subject is illustrated in **Figure 5**.

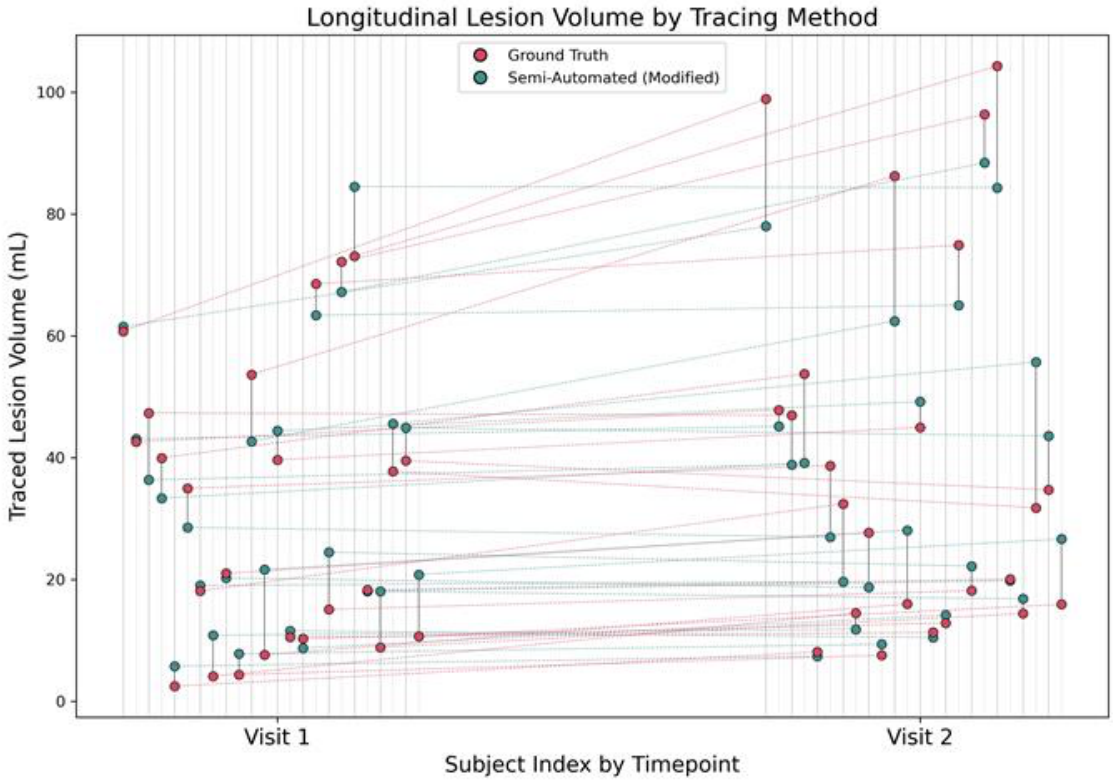

Figure 5: Overview of ground truth and semi-automated lesion tracing volumes across timepoints. For each subject, visit 1 and visit 2 ground truth and semi-automated lesion mask volumes are plotted according to subject index (sorted alphanumerically). Dotted lines indicated change in lesion volume in a single subject across timepoints. Ground truth and semi-automated tracings tend to show similar changes longitudinally, which are generally in the positive direction.

Inter-rater reliability

We first computed the mean Dice coefficient for each pair of raters, resulting in mean values of 0.73, 0.59, and 0.66, respectively. To obtain an overall measure of agreement, we calculated the average of these mean Dice coefficients, yielding an overall mean Dice coefficient of 0.66. This value indicates a moderate level of agreement among the raters, suggesting reasonable consistency in lesion segmentation across the three raters and highlighted the heterogeneity of lesion conditions represented in **Figure 6**.

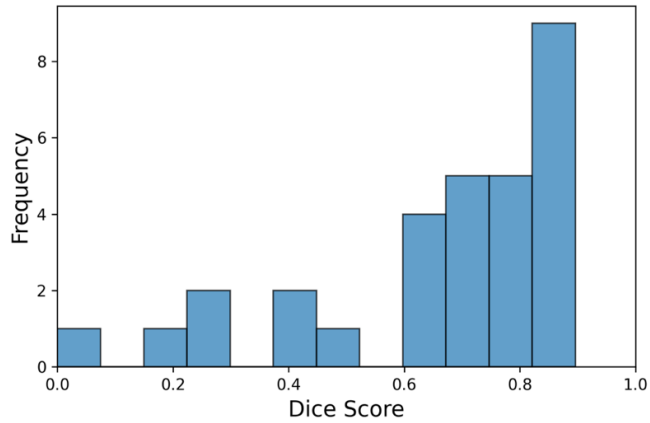

Figure 6: Distribution of inter-rater dice scores, forming the null distribution.

Changes in lesion Dice compared to null distribution

The comparison of each individual dice score and the interrater dice scores revealed statistically significant differences for 23 of 24 subjects (**Figure 7**).

All FDR corrected p-values for these comparisons were below the significance threshold of 0.05, indicating a significant deviation from the interrater scores. The findings from the Wilcoxon rank-sum test (**Figure 5**) revealed a statistically significant difference in group-level lesion volume changes when compared against the null distribution (test statistic: 3.887, z = 13.172, p = 0.0001). The 95% confidence interval for the test statistic ranged from 3.24 to 4.54, suggesting a robust and statistically significant deviation.

Lesion Size Variation Across Different Visit Intervals

Inter-visit intervals ranged from 727 to 2,366 days. Correlation analyses revealed no relationship between visit intervals and differences in lesion size (R = 0.17, p = 0.44; **Figure 8**). The regression model explained 10.7% of the variance in lesion volume change with an R-squared value of 0.11. After adjusting for number of predictors, the adjusted R-squared value was -0.027, indicating reduced explanatory power. None of the individual predictors: age, sex, or visit interval, were statistically significant. The intercept showed no significant effect on lesion volume change (p=0.71), while the coefficients for age (p=0.34), sex (p=0.26),

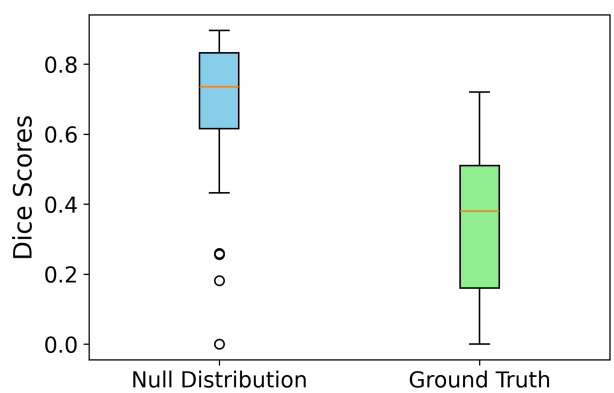

Figure 7: Box plots illustrating ground truth and inter-rater dice scores, demonstrating the distribution of the data. Scatter points, color-coded to represent calculated p-values from comparisons of individual dice scores against the null distribution, are overlaid on the ground truth data. This analysis highlights that statistically significant deviations from the null distribution were observed in 23 out of 24 participants.

and visit interval (p=0.996), lacked statistical significance. The overall F-statistic was 0.7954 (p=0.511), suggesting the model did not predict changes in lesion volume based on the variables examined. Together, these findings suggest the changes observed here are not expected or explained by age-related change or influenced by sex.

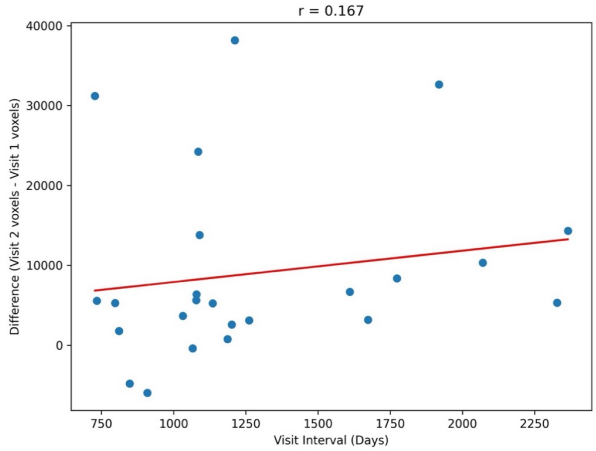

Figure 8: Correlation analysis examining the time between study visits and the calculated difference in lesion volume.

Semi-automated segmentation performance compared to ground truth segmentations

To assess volumetric differences in lesion volume, six Wilcoxon ranked-sum tests with FDR corrections to account for multiple comparisons were conducted to examine the relationship between raw, modified, and ground truth segmentation volumes across two time points. There were no statistically significant differences between raw and modified semi-automated volumes at Visit 1 (W = 262.0, p = 0.60) or Visit 2 (W = 244.0, p = 0.37). This suggests that the modifications made to the semi-automated segmentations did not result in significant changes in volume compared to the original raw segmentations. Furthermore, when compared to ground truth segmentations, neither the raw segmentations at Visit 1 (W = 296.0, p = 0.88) nor Visit 2 (W = 327.0, p = 0.88), nor the modified segmentations at Visit 1 (W = 260.0, p = 0.57) or Visit 2 (W = 292.0, p = 0.94) showed statistically significant differences. These findings suggest that both raw and modified semi-automated segmentations closely approximate the ground truth, indicating robustness in the segmentation process across both time points (**Figure 9**). Though time required to perform each segmentation method varied across scans depending on lesion burden, we estimated the time required to perform each method as 10-20 minutes per scan for lesion adjustment using the semi-automated method and 60-90 minutes for manual segmentation.

To provide additional insights into the variance between the modified semi-automated segmentations and ground truth segmentations, MAE was across both time points. At Visit 1 the MAE of 5686.29 (R = 0.95), signifies a very strong linear relationship and high concordance in lesion segmentation accuracy between the two methods. At Visit 2, the MAE increased to 8973.71 (R = 0.94), further substantiating a robust positive correlation but suggesting a higher average discrepancy at the second timepoint.

The results of the Wilcoxon ranked-sum test were used to determine whether there were statically significant differences between the dice scores of the modified semi-automated segmentations and ground truth segmentations. This analysis was conducted to assess both volumetric discrepancies and spatial overlap accuracy, as represented by Dice scores, between the modified segmentations and ground truth. The results yielded a test statistic of 373.00 and corresponding p-value of 0.0813 indicating no statistically significant difference between methods. This result is

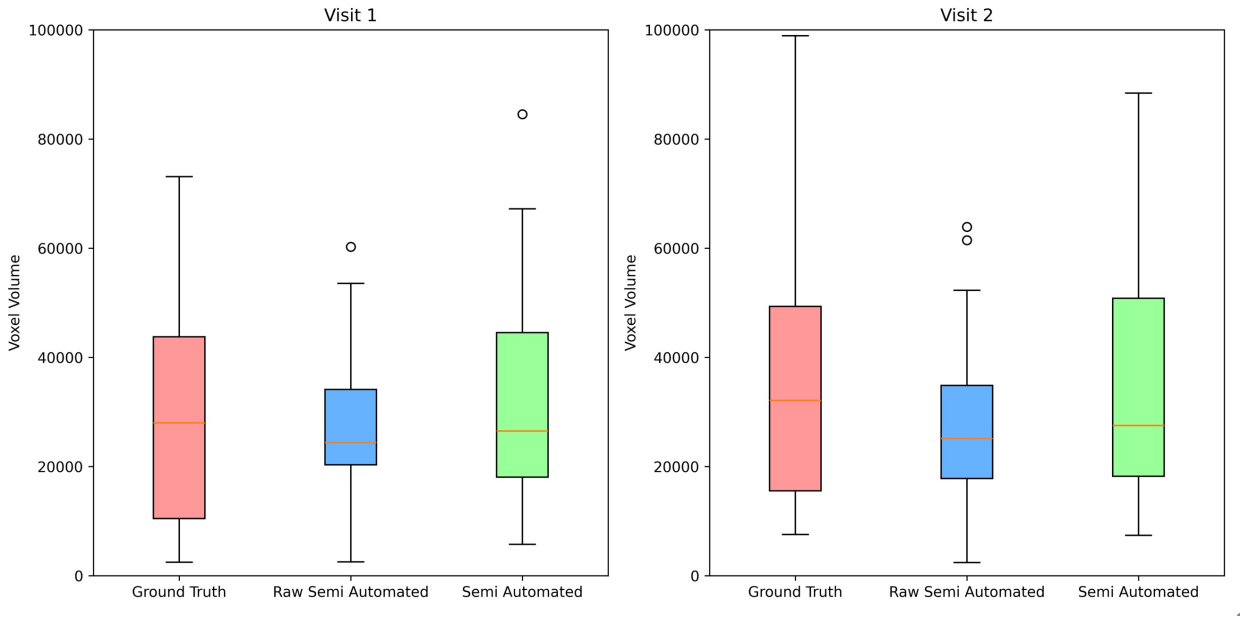

Figure 9: Comparative analysis of voxel volume distributions across different segmentation methods and time points. Box plots representing voxel volume for Ground Truth, Raw Semi Automated, and Semi Automated groups across two visits. The left column displays data for Visit 1, and the right column displays data for Visit 2. For each visit, three box plots illustrate the distribution of voxel volume for the respective groups: Ground Truth (light red), Raw Semi Automated (light blue), and Semi Automated (light green).

further illustrated in **Figure 10**, highlighting the performance of the semi-automated tool in accurately segmenting lesions when compared to ground truth.

Discussion

In this longitudinal MRI study of 24 individuals with chronic TBI, we demonstrate the feasibility and efficiency of a semi-automated lesion segmentation tool. Our findings indicate this FreeSurfer-based tool performs robustly against ground-truth manual tracings to segment neuroanatomic structures in the presence of lesions with improved efficiency, as compared to previously developed methods (Diamond et al., 2020). Further, in a proof-of-principle application of the semi-automated lesion tool, we provide initial evidence that cortical lesions continue to expand even beyond one-year post-injury, with 23 of 24 subjects experiencing lesion expansion. These observations raise the possibility that lesion expansion may be a contributing factor in PTND – a finding that will require confirmation in larger longitudinal studies with clinical-radiologic-pathological correlations. The semi-automated lesion tool thus creates new opportunities to investigate the role of cortical lesions in the pathogenesis of post-traumatic PTND.

The semi-automated lesion segmentation tool developed here builds upon recent innovations in machine learning-based imaging analysis, most notably SynthSR (Iglesias et al., 2023; Iglesias et al., 2021). What distinguishes this tool from previously developed tools are: 1) increased efficiency when compared to traditional manual tracing; 2) scalability for rapid initial segmentation of large datasets; and 3) improved accessibility and reproducibility, providing preliminary segmentations that are both consistent and highlight potentially impacted areas and offer manual

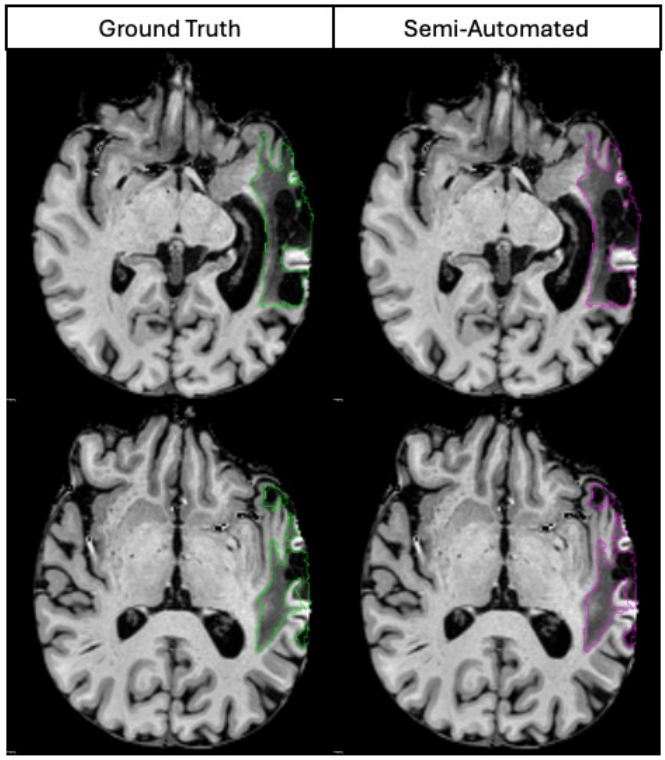

Figure 10: Application of the semi-automated lesion segmentation method showcases the accuracy attainable through comparison with ground truth segmentations, highlighting the method's performance in accurately delineating lesions.

raters a starting point for refinement of segmentations. The new semi-automated tool demonstrates strong performance characteristics against "ground-truth" manual lesion segmentations, as evidenced by the strong positive correlations observed between the two methods and lack of statistically significant changes between ground truth and raw segmentation volumes. These findings underscore the consistent and reliable performance of semi-automated segmentation compared to traditional manual tracing across both evaluation points in our study.

Importantly, the new method continues to require human input to refine and optimize the lesion's boundaries – a step that reflects the inherent challenge of training automated tools to detect traumatic lesions, which often have heterogeneous signal characteristics related to hemorrhagic and non-hemorrhagic components. Nonetheless, the time required for this manual step is far less than for our previously published lesion segmentation method (Diamond et al., 2020). Specifically, while the prior tool required manual creation of set points along the entire lesion surface, the new method requires only a small number of voxel-based edits in volumetric space.

The lesion expansion observed in this cohort is consistent with, and builds upon, the growing evidence base indicating that pathological processes in TBI persist and progress in the chronic setting, even beyond one year-post injury. Whether lesion expansion is attributable to chronic inflammation, gliosis, microvascular ischemia, or some combination of factors will require pathological-radiologic correlation analyses, which the LETBI study is designed to perform, given the premortem consent for autopsy provided by LETBI participants (Edlow et al., 2018). The absence of an association between lesions expansion and time between scans suggests that lesion expansion occurs at variable rates, though this preliminary observation will require future studies with larger sample sizes to confirm. The potential contribution of lesion expansion to the pathogenesis of PTND remains unknown and will require future studies with sufficiently large sample sizes to account for other risk factors, and protective factors. The short-term cognitive and functional correlates of lesion expansion is also an area for future inquiry.

Despite the promising findings from use of semi-automated segmentation tools utilized in this longitudinal MRI study, several limitations should be considered. The small sample size of 24 individuals with chronic TBI limits the generalizability of our results, necessitating larger cohorts

for validation. The follow-up period may also be insufficient to fully capture the long-term trajectory of lesion expansion and its implications for PTND. While the semi-automated tool improves efficiency, it still requires manual input for refining lesion boundaries, introducing potential variability and subjectivity. Additionally, the heterogeneous nature of the lesions, as demonstrated by a heatmap illustrating the locations of all lesions included in our study (**Figure 11**), including both hemorrhagic and non-hemorrhagic components, further complicates the segmentation process, as the tool may not uniformly handle all types of lesions with the same accuracy. Lastly, this study did not test for cognitive and functional correlates of lesion expansion – a crucial area for future research. Addressing these limitations will be essential for advancing our understanding of lesion dynamics in chronic TBI.

In summary, we developed and implemented a semi-automated lesion detection tool that accurately and efficiently identifies chronic lesions in patients with TBI. Further, we provide proof-of-principle evidence that this lesion segmentation tool can detect longitudinal lesion growth in individuals with chronic TBI. Future applications of this tool have the potential to elucidate the

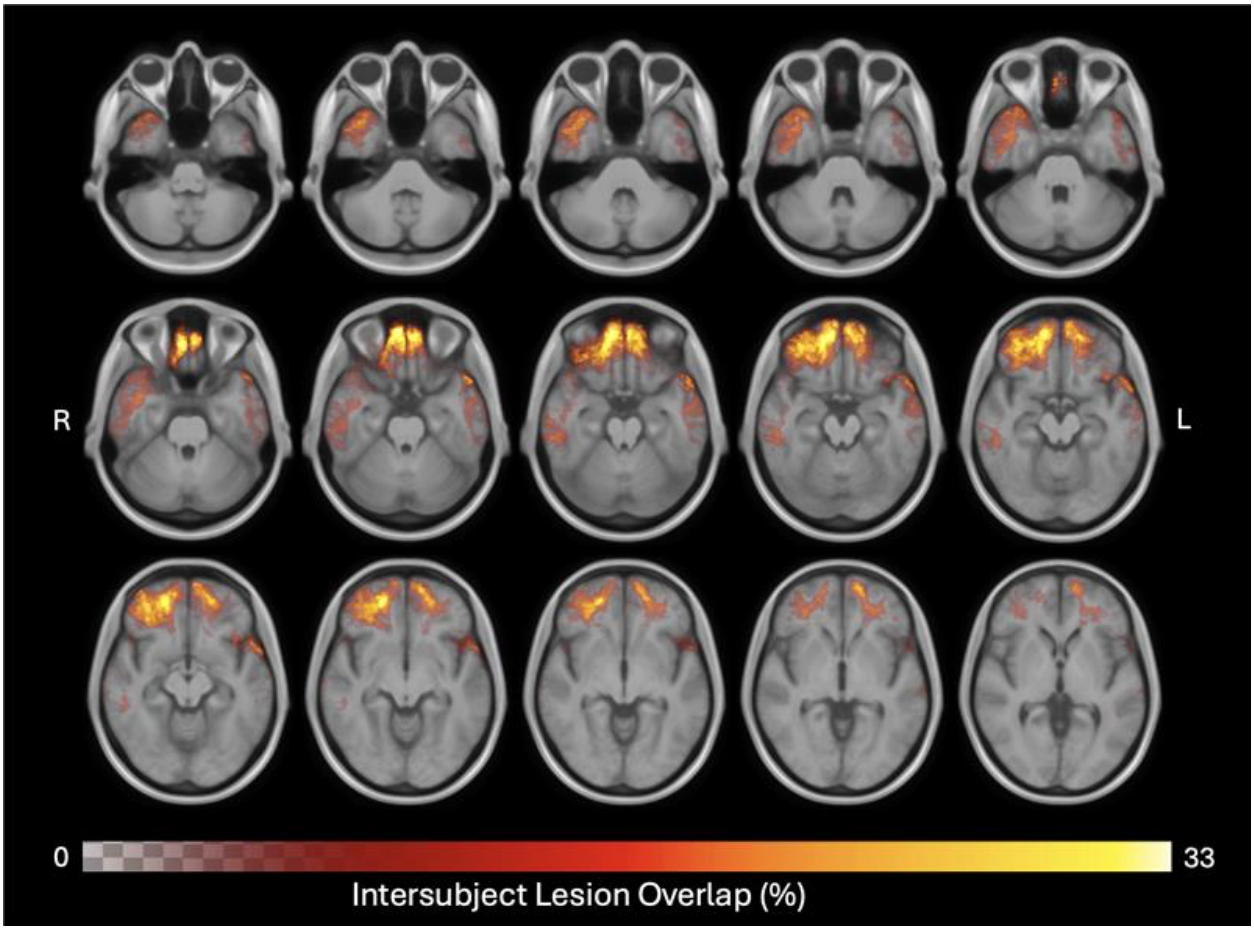

Figure 11: Volumetric distribution of ground truth lesions across timepoints. Heatmap of all 48 ground truth lesion tracings registered to MNI space and overlayed on anatomical average, revealing the most prevalent neuroanatomical positions of lesions in this cohort. Color and alpha of the heatmap is modulated by the percent of lesion traces in each voxel, with the maximum overlap observed being 33%. Lesions are most prevalent in the anterior temporal lobe and ventral anterior frontal lobe.

potential pathophysiologic links between lesion expansion and PTND. Ultimately, the integration of lesion segmentation into clinical MRI workflows also has the potential to inform preventive, diagnostic, prognostic, and therapeutic strategies in clinical care.

References

Cerri, S., Puonti, O., Meier, D. S., Wuerfel, J., Muhlau, M., Siebner, H. R., & Van Leemput, K. (2021). A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. *Neuroimage, 225*, 117471. doi:10.1016/j.neuroimage.2020.117471

Dams-O'Connor, K., Ketchum, J. M., Cuthbert, J. P., Corrigan, J. D., Hammond, F. M., Haarbauer-Krupa, J., . . . Miller, A. C. (2020). Functional Outcome Trajectories Following Inpatient Rehabilitation for TBI in the United States: A NIDILRR TBIMS and CDC Interagency Collaboration. *J Head Trauma Rehabil, 35*(2), 127-139. doi:10.1097/HTR.0000000000000484

Dams-O'Connor, K., Seifert, A. C., Crary, J. F., Delman, B. N., Del Bigio, M. R., Kovacs, G. G., . . . Folkerth, R. D. (2023). The neuropathology of intimate partner violence. *Acta Neuropathol, 146*(6), 803-815. doi:10.1007/s00401-023-02646-1

Diamond, B. R., Donald, C. L. M., Frau-Pascual, A., Snider, S. B., Fischl, B., Dams-O'Connor, K., & Edlow, B. L. (2020). Optimizing the accuracy of cortical volumetric analysis in traumatic brain injury. *MethodsX, 7*, 100994. doi:10.1016/j.mex.2020.100994

Ding, K., Marquez de la Plata, C., Wang, J. Y., Mumphrey, M., Moore, C., Harper, C., . . . Diaz-Arrastia, R. (2008). Cerebral atrophy after traumatic white matter injury: correlation with acute neuroimaging and outcome. *J Neurotrauma, 25*(12), 1433-1440. doi:10.1089/neu.2008.0683

Edlow, B. L., Keene, C. D., Perl, D. P., Iacono, D., Folkerth, R. D., Stewart, W., . . . Dams-O'Connor, K. (2018). Multimodal Characterization of the Late Effects of Traumatic Brain Injury: A Methodological Overview of the Late Effects of Traumatic Brain Injury Project. *J Neurotrauma, 35*(14), 1604-1619. doi:10.1089/neu.2017.5457

Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781. doi:10.1016/j.neuroimage.2012.01.021

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp, 2*(4), 189-210.

Iglesias, J. E., Billot, B., Balbastre, Y., Magdamo, C., Arnold, S. E., Das, S., . . . Fischl, B. (2023). SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci Adv, 9*(5), eadd3607. doi:10.1126/sciadv.add3607

Iglesias, J. E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gilberto Gonzalez, R., . . . Alzheimer's Disease Neuroimaging, I. (2021). Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *Neuroimage, 237*, 118206. doi:10.1016/j.neuroimage.2021.118206

Johnson, V. E., Stewart, J. E., Begbie, F. D., Trojanowski, J. Q., Smith, D. H., & Stewart, W. (2013). Inflammation and white matter degeneration persist for years after a single traumatic brain injury. *Brain, 136*(Pt 1), 28-42. doi:10.1093/brain/aws322

Kenney, K., Iacono, D., Edlow, B. L., Katz, D. I., Diaz-Arrastia, R., Dams-O'Connor, K., . . . Perl, D. P. (2018). Dementia After Moderate-Severe Traumatic Brain Injury:

Coexistence of Multiple Proteinopathies. *J Neuropathol Exp Neurol, 77*(1), 50-63. doi:10.1093/jnen/nlx101

McKee, A. C., Stern, R. A., Nowinski, C. J., Stein, T. D., Alvarez, V. E., Daneshvar, D. H., . . . Cantu, R. C. (2013). The spectrum of disease in chronic traumatic encephalopathy. *Brain, 136*(Pt 1), 43-64. doi:10.1093/brain/aws307

Merkley, T. L., Bigler, E. D., Wilde, E. A., McCauley, S. R., Hunter, J. V., & Levin, H. S. (2008). Diffuse changes in cortical thickness in pediatric moderate-to-severe traumatic brain injury. *J Neurotrauma, 25*(11), 1343-1345. doi:10.1089/neu.2008.0615

Puonti, O., Iglesias, J. E., & Van Leemput, K. (2016). Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. *Neuroimage, 143*, 235-249. doi:10.1016/j.neuroimage.2016.09.011

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage, 61*(4), 1402-1418. doi:10.1016/j.neuroimage.2012.02.084

Sandsmark, D. K., Bashir, A., Wellington, C. L., & Diaz-Arrastia, R. (2019). Cerebral Microvascular Injury: A Potentially Treatable Endophenotype of Traumatic Brain Injury-Induced Neurodegeneration. *Neuron, 103*(3), 367-379. doi:10.1016/j.neuron.2019.06.002

Santhanam, P., Wilson, S. H., Oakes, T. R., & Weaver, L. K. (2019). Accelerated age-related cortical thinning in mild traumatic brain injury. *Brain Behav, 9*(1), e01161. doi:10.1002/brb3.1161

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., . . . Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23 Suppl 1*, S208-219. doi:10.1016/j.neuroimage.2004.07.051

Soille, P. (2004). *Morphological image analysis*. Berlin: Springer.

Strangman, G. E., O'Neil-Pirozzi, T. M., Supelana, C., Goldstein, R., Katz, D. I., & Glenn, M. B. (2010). Regional brain morphometry predicts memory rehabilitation outcome after traumatic brain injury. *Front Hum Neurosci, 4*, 182. doi:10.3389/fnhum.2010.00182

van der Kouwe, A. J. W., Benner, T., Salat, D. H., & Fischl, B. (2008). Brain morphometry with multiecho MPRAGE. *Neuroimage, 40*(2), 559-569. doi:10.1016/j.neuroimage.2007.12.025

Vande Vyvere, T., Pisica, D., Wilms, G., Claes, L., Van Dyck, P., Snoeckx, A., . . . Parizel, P. M. (2024). Imaging Findings in Acute Traumatic Brain Injury: a National Institute of Neurological Disorders and Stroke Common Data Element-Based Pictorial Review and Analysis of Over 4000 Admission Brain Computed Tomography Scans from the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study. *J Neurotrauma*. doi:10.1089/neu.2023.0553

Warner, M. A., Youn, T. S., Davis, T., Chandra, A., Marquez de la Plata, C., Moore, C., . . . Diaz-Arrastia, R. (2010). Regionally selective atrophy after traumatic axonal injury. *Arch Neurol, 67*(11), 1336-1344. doi:10.1001/archneurol.2010.149