Title
Optimization of Brain Segmentation in Multiple Sclerosis Patients

Author
Zhu, Alyssa

Publication Date
2013

Peer reviewed|Thesis/dissertation

Optimization of Brain Segmentation in Multiple Sclerosis Patients

by

Alyssa Zhu

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Imaging

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

ACKNOWLEDGMENTS

My first and foremost thanks go to my advisor Dr. Roland Henry for giving me the opportunity to take on so many new tasks and learn from them. I would also like to thank Eduardo Caverzasi and Valentina Panara for their radiologic assessments without which proper analysis would not be possible. Thank you to Dr. Alastair Martin, Dr. Tracy Richmond McKnight, and Dr. Pratik Mukherjee for taking the time to act as my thesis committee. And additional thanks to UCSF and the MSBI program for providing both an amazing education and wonderful connections that will hopefully never fade.

DEDICATION

I would like to dedicate this to my family: my mother for her never-ending nurturing care, my father for always cultivating an analytical mind, and my older brother for always looking out for me.

Optimization of Brain Segmentation in Multiple Sclerosis Patients

Alyssa Zhu

ABSTRACT

Multiple sclerosis is an idiopathic, autoimmune disease that affects the central nervous system (CNS). Imaging studies have shown that gray matter volume, rather than whole brain or white matter volume, acts as the best imaging biomarker for MS progression. Previous studies were performed via cross-sectional analysis of each time point and then interrogating the difference between values. Because of variability inherent in software tools, the population of cross-sectional analysis studies is dependent on the segmentation program being utilized with smaller standard deviations allowing for smaller subject populations, particularly when the tissue volume difference being studied is small in comparison those standard deviations. Longitudinal analysis aims to minimize that variability and give more accurate segmentation results. Segmentation in MS is also plagued by the presence of white matter lesions, whose T1 hypointensities can result in the tissue being misclassified as gray matter. Two longitudinal programs that have been validated for healthy controls and patients with Alzheimer's disease – aBEAT and FreeSurfer – were explored by retrospectively analyzing 7 sets of longitudinal data both cross-sectionally and longitudinally. A comparison between programs revealed that FreeSurfer produced more accurate both segmentation and anatomical parcellation results. Quantitative analysis of gray matter volumes also showed FreeSurfer to be superior to aBEAT with FreeSurfer's cross-sectional processing yielding the

smoothest transition from time point to time point. The investigation into cortical thicknesses obtained by FreeSurfer, on the other hand, yielded slightly conflicting results between R^2 values and observed longitudinal trends. Further analysis of both longitudinal processing and lesion segmentation is required to evaluate the usefulness of currently available longitudinal processing programs and to avoid the need for manual segmentation respectively. Currently, cross-sectional segmentation is the optimal method for longitudinal brain volume analysis as longitudinal segmentation programs have proven inferior rather than superior to their cross-sectional counterparts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Multiple Sclerosis

Multiple sclerosis (MS) is a debilitating, autoimmune disease that affects the central nervous system (CNS) resulting in lesions and cerebral atrophy (National Multiple Sclerosis Society). It is idiopathic with several working theories regarding the interplay between the immune system, neurodegeneration, and genetics. Diagnosis comes from a combination of medical history and neurologic exam, MRI, visual evoked potential, spinal tap, and/or blood test. A physician's assessment includes medical history to identify potential symptoms caused by MS and scores that gauge physical, mental, or emotional functions. Physicians will then often turn to MR images to confirm or further support the diagnosis, looking for the presence of CNS lesions, though it should be noted that early stages of MS may not show lesions (National Multiple Sclerosis Society). A spinal tap can show immune system activity indicative of MS in the cerebrospinal fluid (CSF), whereas blood test results are used to eliminate other possibilities. There are four disease courses: relapsing-remitting (RRMS), primary progressive (PPMS), secondary progressive (SPMS), progressive-relapsing (RPMS). The courses are differentiated by the progression of a patient's symptoms, i.e. is there a period of recovery and if so does the patient recover to baseline. For example, patients with RRMS can recover back to baseline after acute attacks, whereas patients with PRMS will not. One course can develop into another – e.g. RRMS to SPMS – or it can stand on its own (PPMS). As there is no cure, various treatments are employed in combination, e.g. working to affect disease course or

manage symptoms. In order to gauge the progression of the disease and the efficacy of treatments, tissue quantification via imaging is being employed and explored as a biomarker.

MS was traditionally viewed as a white matter disease due to MRI's higher sensitivity for white matter volume changes compared to gray matter changes (Horakova et al., 2012). Previous hardware and pulse sequences resulted in low resolution images and, with cortical thickness being on the order of 2-3 mm, MRI's sensitivity to gray matter changes was extremely susceptible to partial volume effects. Additionally, white matter injury was found to be partially independent to gray matter injury (Fillipi et al., 2010), making it impossible to study white matter effects in lieu of those of gray matter. With the improvement in MRI technology, gray matter has emerged as the more indicative of the two as an imaging biomarker for MS progression (Honce, 2013). Gray matter atrophy, being less susceptible to inflammation (Horakova et al., 2012), fluctuates less than whole brain and white matter atrophy (Medscape Review). It is detectable early in the disease, accelerates over time (Honce, 2013), and may accumulate to a greater extent than white matter atrophy (Anderson et al., 2009). Gray matter atrophy has also been associated with MS clinical disability and has been found to correlate more strongly with disability than lesion volumes and white matter atrophy (Honce, 2013). Cortical lesion loads – a measure that combines lesion sizes and abundance – have been proposed to be a superior metric than gray matter atrophy, but detection requires special pulse sequences such as double inversion recovery (DIR) and phase sensitive inversion recovery (PSIR). These sequences benefit from higher field strengths (Honce, 2013).

While DIR and PSIR improve cortical lesion detection by hundreds of percent, measurements are less reliable and results less reproducible between sites compared to those of gray matter atrophy (Honce, 2013).

Segmentation

For many years, longitudinal studies analyzing changes in brain tissue volumes were performed by cross-sectionally processing all images in a longitudinal data set and then observing trends in or differences between values obtained of each individual time point. A pitfall of that methodology is the inherent variability of processing algorithms. For example, tissue volume quantification of images from two different time points may change either because of atrophy or because of the standard deviation of the method. The standard deviation of the volume measurement in the absence of atrophy determines the sample size needed to achieve a statistical significance ($p = 0.05$). A study by de Boer et al. (2010) investigated the differences between utilizing various cross-sectional segmentation programs for longitudinal analysis and extrapolated the approximate study sizes needed by each program to obtain results with statistical significance. FSL's FAST segmentation program (Zhang et al., 2001), which is implemented in Siena and SienaX (Smith et al., 2002), had the best reproducibility and would therefore require the smallest subject population out of the studied methods. In contrast, SPM5 (Ashburner and Friston, 2005) had the lowest reproducibility and accuracy for brain segmentation and would therefore need a lot more subjects.

Longitudinal processing programs have mostly been developed and validated in context of Alzheimer's disease by comparing brain volumes of healthy controls and patients with mild cognitive impairment or Alzheimer's disease. This processing involves considering scans of all time points together in a 4D data set to minimize the variability in segmentation from scan to scan. In literature, papers putting forth their new, successful algorithms for longitudinal segmentation often use results from Consistent Longitudinal Alignment and Segmentation for Serial Image Computing (CLASSIC) and to a degree FreeSurfer's longitudinal pipeline for comparison to their own and therefore validation of their programs. The Adult Brain Extraction and Analysis Toolbox (aBEAT) was one such program (Dai et al., 2013) that appeared to provide better results than either of the previous two.

An additional consideration for tissue segmentation in MS is the appearance of white matter lesions, which appear as hypointensities on T_1 weighted images and are therefore often misclassified as gray matter (Honce, 2013). Some methods like SienaX allow for the integration of a lesion mask, which will prevent voxels identified as lesions from being classified as gray matter. However, that requires that lesion masks be initially segmented, and the gold standard approach to identifying lesions is manual, which requires a lot of time and effort. Though some methods consider combining information from other image contrasts (FLAIR or T_2 images) in addition to T_1 images to properly segment lesions, resulting lesion masks have been found to be inadequate and still require manual editing. Lesion segmentation has been a persistent problem, and in 2008 an MS lesion segmentation challenge was held to

compare lesion segmentation algorithms (Styner et al., 2008). While submissions from the challenge were promising, automated lesion segmentation remains elusive.

This study aimed to evaluate two segmentation programs – one commonly used (FreeSurfer) and one recently released (aBEAT) – to assess their applicability for large scale cross-sectional and longitudinal processing pipelines. The analysis aimed to determine whether an existing pipeline was adequate for longitudinal segmentation studies or whether there were substantive deficiencies with current approaches. This determination is critical to the efficacy and practicality of longitudinal studies that aim to utilize the imaging results as a disease biomarker.

METHODS

Subjects

Subjects consisted of seven patients (4 men and 3 women, average age 44.7±8.0 at baseline) with diagnosed MS who showed focal symptoms, i.e. attacks were limited to one or a few limbs. As part of the Expression, Proteomics, Imaging, Clinical (EPIC) study, subjects had been scanned at yearly intervals for a total of 6 scans per patient with the exception of one who missed follow-up scans for years 3 and 4. Inversion recovery spoiled gradient echo (IR-SPGR) T1-weighted images (TR = 7.496 ms, TE = 1.652 ms, TI = 400 ms, FOV = 24x24 cm^2, matrix = 256 x 192, flip angle = 15°, NEX = 1) were acquired at University of California, San Francisco's China Basin Campus on

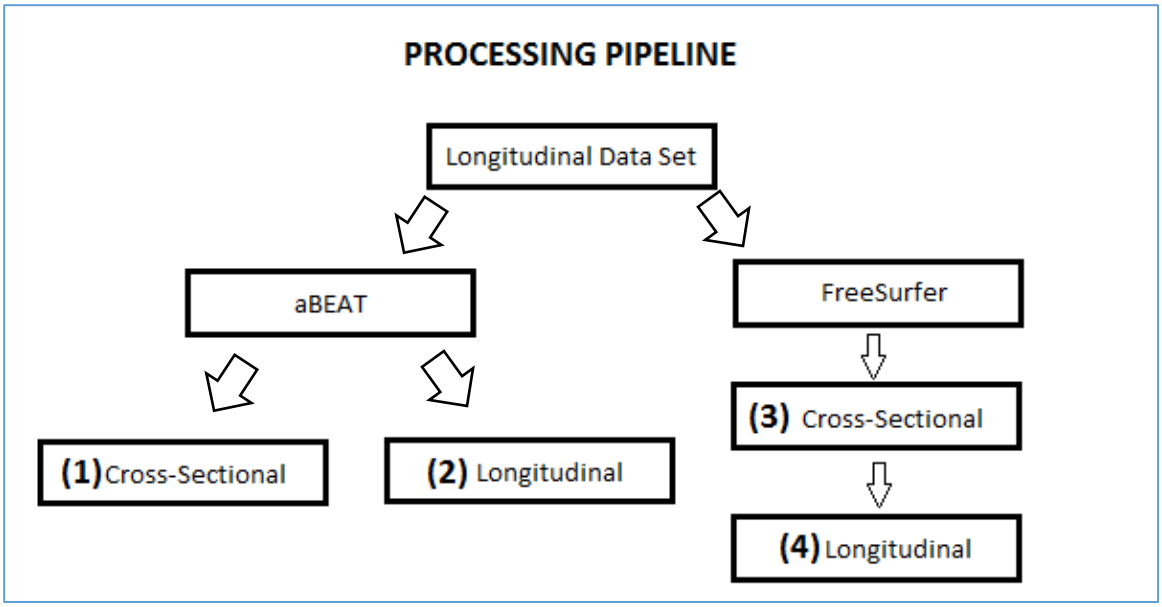

Figure 1: Processing Pipeline. Seven longitudinal data sets were retrospectively analyzed using two segmentation programs: aBEAT and FreeSurfer. Using aBEAT, images were segmented cross-sectionally and longitudinally. In FreeSurfer, the data was segmented individually first followed by longitudinal processing using the cross-sectional results to create a base image and then to initialize the 4D processing algorithms. Altogether, segmentation of each image resulted in four volume values.

a 3T GE Signa EXCITE scanner. These T_1 images served as the segmentation program inputs.

Tissue Segmentation and Anatomical Parcellation

The first longitudinal segmentation program tested was University of Pennsylvania's Multiplicative Intrinsic Component Optimization (MICO). While validation of the program was not complete, a lack of references for the program and discouraging preliminary results led to its abandonment. The next two programs tested were aBEAT and FreeSurfer. The longitudinal data sets were segmented using both cross-sectional and longitudinal processes for both programs (Figure 1).

Processing in aBEAT consisted of preprocessing, brain extraction, segmentation, and anatomical parcellation based on the Colin27 atlas mask (Dai et al., 2013).

1. Preprocessing consisted of image orientation to aBEAT's coordinate system and non-parametric non-uniform intensity normalization (N3) bias field correction to remove intensity inhomogeneities. To remove intra-subject intensity variations, longitudinal preprocessing also included intensity histogram matching of each image to the baseline image.
2. In 4D brain extraction, all images were affinely registered to a common space, then a brain probability map was warped onto the registered subject images and was used to remove non-brain voxels. If the brain extraction contained parts of the skull, the manual mask editor was used to create a new mask, which would then be used to perform a final brain extraction.
3. aBEAT's longitudinal segmentation first involved cross-sectional segmentation of all images with a spatial cortical thickness constraint, followed by 4D registration to align all segmentation results. Iterative 4D segmentation then occurred with the algorithm taking three terms into consideration: fitting of intensity distributions for white matter, gray matter, and CSF; a spatial cortical thickness constraint; and a temporal cortical thickness restraint. It should be noted that, while longitudinal segmentation in aBEAT did involve an initial cross-sectional segmentation of each image, the inputs for the longitudinal processing pipeline were the raw images.

4. In 4D anatomical labeling, longitudinal images were registered to their group-mean image based on four key points with distinctive features. The deformation field was also calculated for registering the Colin27 atlas to the group mean image. A combination of the two deformation pathways resulted in the mapping of Colin27's 90 regions of interest (ROIs) – 45 per hemisphere – to each time point image.
5. ROI analysis was then performed using the aBEAT graphical user interface (GUI), which allows the user to select subject(s), tissue type, and ROI for analysis.

In FreeSurfer, cross-sectional processing is a prerequisite to longitudinal processing (Reuter and Fischl, 2012), i.e. the inputs for longitudinal processing were previously segmented images. Individual processing was completed at each time point, and a median image was established as a base or template for longitudinal processing. Longitudinal preprocessing consisted of resampling to the base voxel space, followed by individual N3 bias field corrections. The base was then affinely registered to the Talairach coordinate system (Talairach and Tournoux, 1988). A longitudinal brain mask was created by taking the union of the registered brain masks of all time points to avoid eliminating any brain-containing voxels. Subcortical structure segmentations from each time point's cross-sectional processing were fused together with longitudinal weighting of each voxel's intensity determining the anatomical structure label to which the voxel will be assigned. Surfaces reconstruction, cortical atlas registration, and parcellations on the other hand were initialized using the base segmentation. Unlike aBEAT and its imposed cortical thickness constraints,

FreeSurfer allows its segmentation and surface reconstruction procedures to evolve freely.

Method Assessment

aBEAT and FreeSurfer's segmented and anatomically parcellated images were registered back to their corresponding T_1 space using FSL's FLIRT (Jenkinson et al., 2002) and FreeSurfer's mri_convert command respectively. The output being analyzed was then overlaid on the T_1 image at which point accuracy was qualitatively assessed by radiologist (Eduardo Caverzasi) and radiology resident (Valentina Panara) review. Lesion masks were obtained from FreeSurfer results by selecting for the white matter hypointensities label. The lesion masks were then overlaid on its corresponding T_1 image and reviewed by a radiologist (E.C.).

Quantitative analysis of the segmentation results was analyzed by examining the smoothness of the gray matter volume profiles over time. As aging is thought to result in a non-linear decrease in brain volume (Scahill et al., 2003) and MS can be thought of as an accelerated aging process, each subject's results from each segmentation process was fit with a quadratic equation and then ranked by R^2 values. A higher R^2 value indicated smoother transition between the gray matter volumes at each time point, which was preferred. Data from the subject with only four time points was excluded from this part of the analysis due to the limited number of points for the quadratic fit. Volumes were also averaged across subjects for each time point and segmentation method to gauge how the programs dealt with the population as a

whole. Cortical thickness values obtained from FreeSurfer were also analyzed with quadratic fits and averaging.

RESULTS

A comparison between programs showed FreeSurfer to prevail as the more accurate of the two for both segmentation and anatomical parcellation. aBEAT and FreeSurfer showed similar results for cortical segmentation; however, aBEAT showed an inability to completely segmenting the basal ganglia (Figure 2(b)). In regards to the anatomical parcellation, aBEAT showed inaccuracy in its labeling, as ROIs did not adhere to the surfaces of sulci and gyri (Figure 3(b)). aBEAT also showed an inability to deal with lesions, classifying them as gray matter (Figure 4(c)). FreeSurfer showed more accurate results overall, which can be seen in Figures 2(a), 3(a), and 4(b) respectively. Figure 4(b) for example shows FreeSurfer's white matter hypointensity segmentation, which correctly views lesions not as gray matter but as the white matter abnormalities.

Averaged volumes across subjects showed similar trends between processing pipelines, with the longitudinal processing having quantified less total gray matter for both programs. FreeSurfer's cross-sectional and longitudinal results paralleled each other, and aBEAT's two volume profiles were nearly identical (Figure 5). Quantitative analysis of gray matter volumes showed FreeSurfer to be superior to aBEAT with FreeSurfer's cross-sectional processing yielding the smoothest results (Table 1). In fact, gray matter volume profiles of the longitudinal data sets show

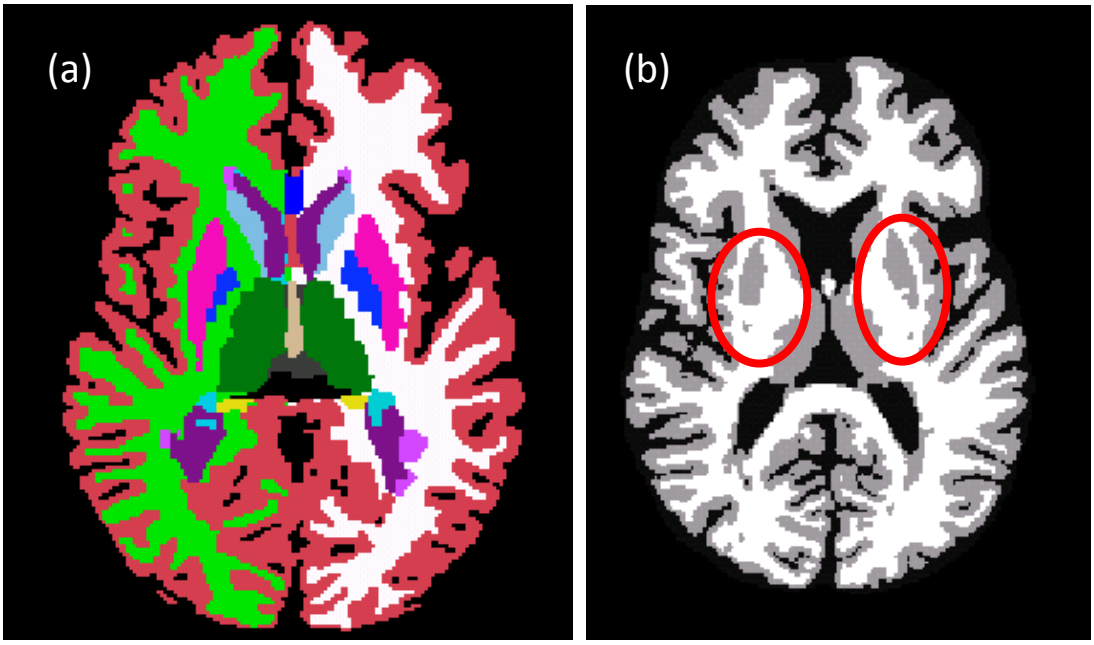

Figure 2: Segmentation Results. (a) FreeSurfer's subcortical segmentation algorithm is weighted across all time points to ensure that voxels are consistently labeled. This leads to the more accurate basal ganglia segmentation seen here. (b) While aBEAT is very stringent about its cortical thickness restraints, there are no such considerations for other parts of the brain. The incomplete segmentation of the left and right putamen – circled in red – demonstrate perhaps a need for more than just histogram-based segmentation for non-cortical regions as well.

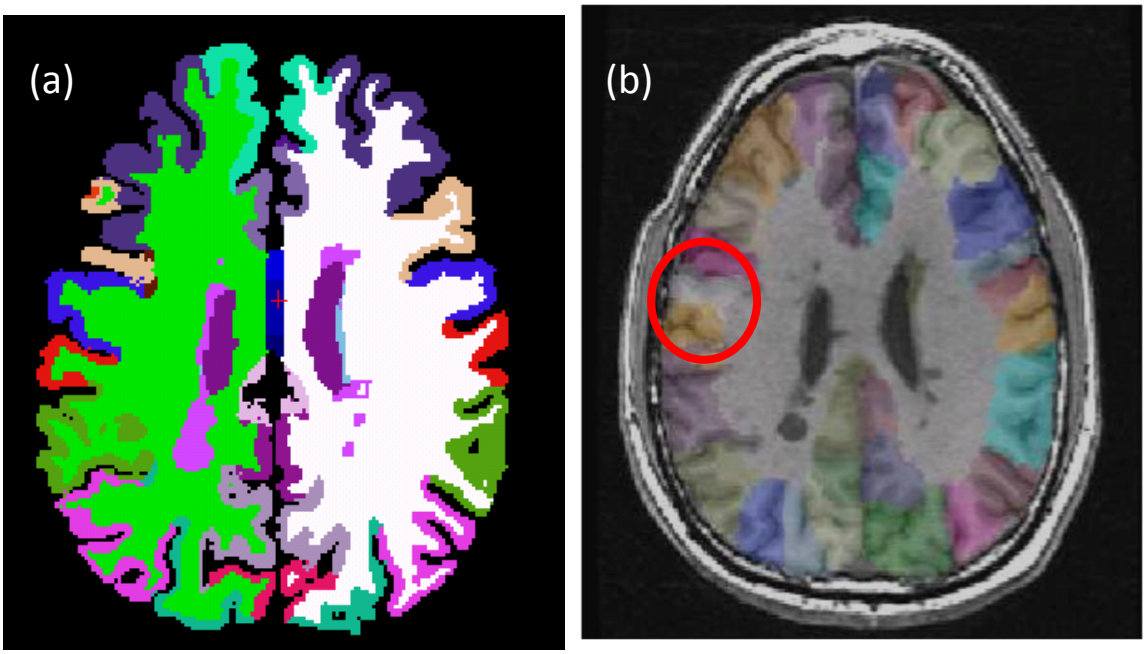

Figure 3: Anatomical Parcellation Results. (a) FreeSurfer's cortical segmentation shows a strict adherence to the surfaces of the sulci and gyri. (b) The aBEAT anatomical mask overlaid on the subject's T_1 image shows that the left precentral gyrus – circled in red – is colored inaccurately with both orange and white thereby labeling the posterior portion as the postcentral gyrus instead.

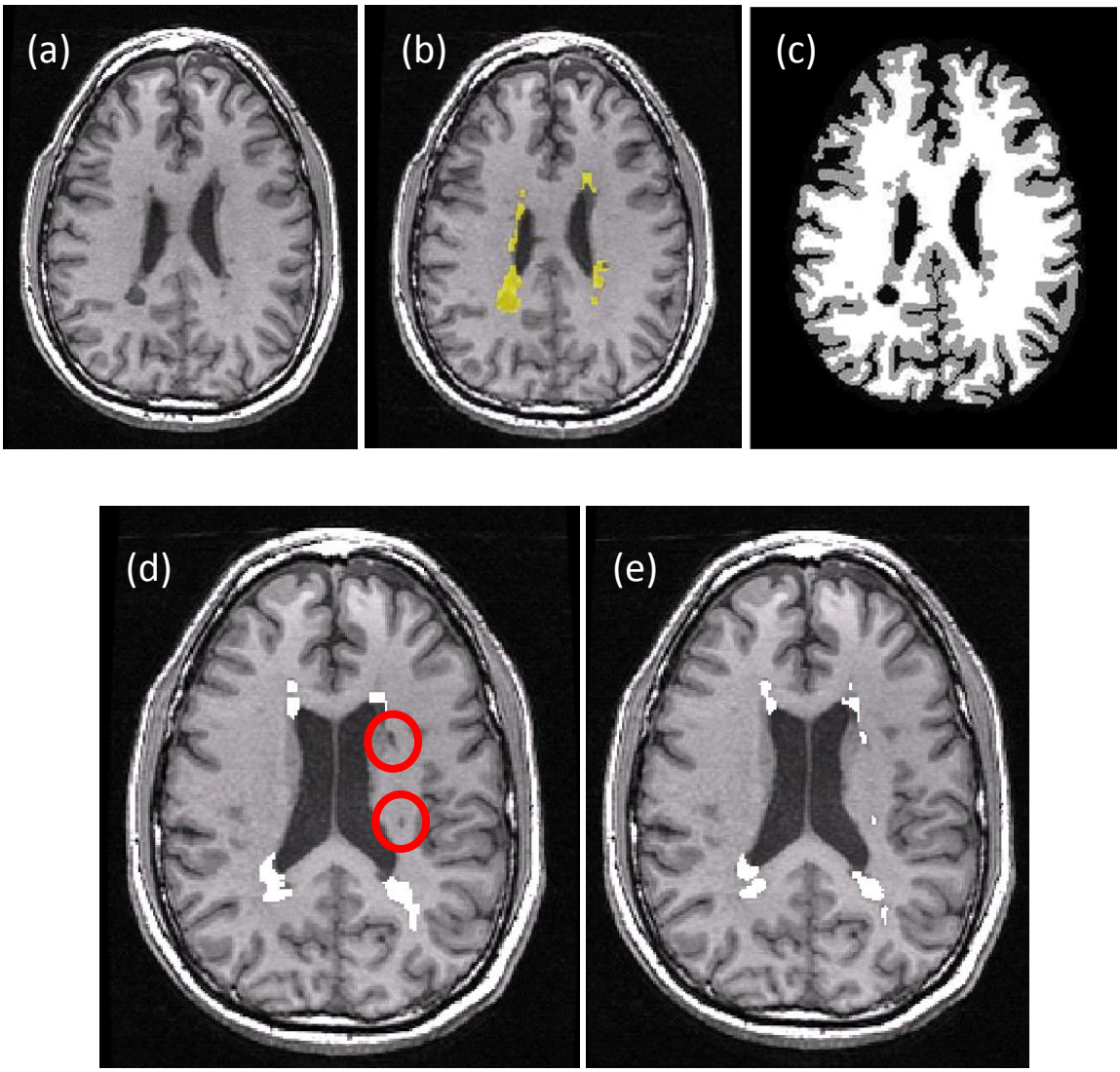

Figure 4: Lesion Segmentation Results. (a) A subject's baseline T1 image. (b) The white matter hypointensities label mask extracted from FreeSurfer's parcellation of (a), which covers the lesions. (c) aBEAT's segmentation of (a), showing that lesions were either misclassified as gray matter or CSF. (d) A different axial slice showing some missed lesions by FreeSurfer, compared to (e) in which lesions were manually segmented.

better smoothness with cross-sectional processing in both aBEAT and FreeSurfer. The investigation into cortical thicknesses obtained by FreeSurfer, on the other hand, yielded conflicting results. Smoothness analysis showed the cross-sectional cortical thickness profiles to be better (R^2 of 0.53 vs. 0.47 for longitudinal processing),

whereas graphically viewing the average cortical thicknesses across time showed less fluctuation with longitudinal processing (Figure 6).

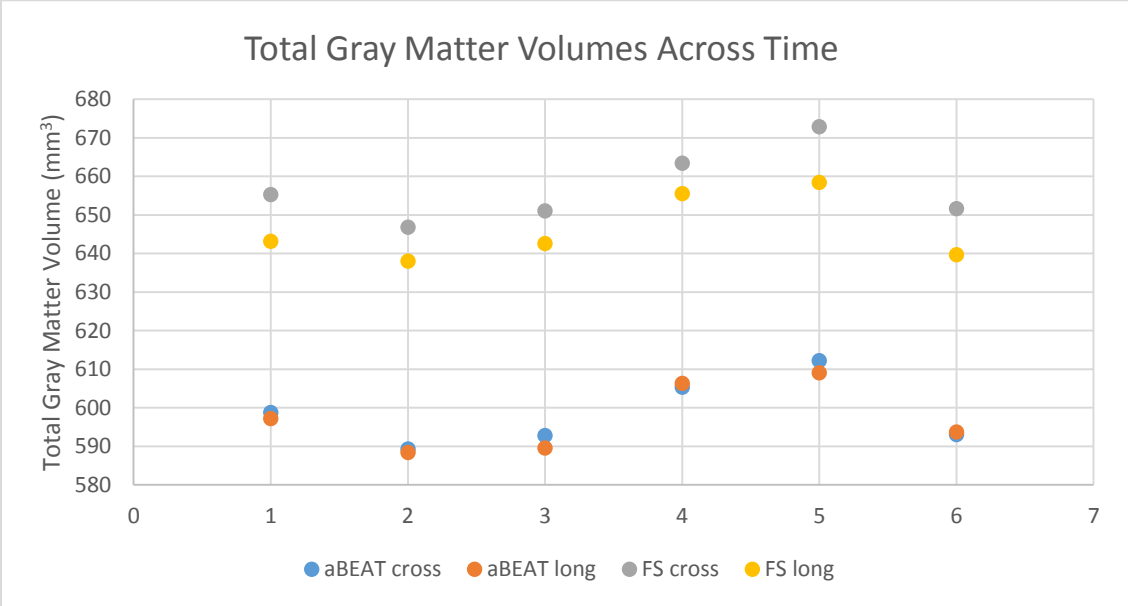

Figure 5: Total Gray Matter Volumes. The average volumes across subjects for each time point and segmentation procedure are shown here. The FreeSurfer cross-sectional and longitudinal results parallel each other, whereas the aBEAT two are nearly identical. Though the quantification of the two programs may be different (approximately 50 mm^3 apart), the volume profile of all four pipelines are extremely similar.

Method	R^2
aBEAT cross-sectional	0.55
aBEAT longitudinal	0.50
FreeSurfer cross-sectional	0.75
FreeSurfer longitudinal	0.56

Table 1: Smoothness Evaluation. The R^2 values of every quadratic fit were averaged and are displayed here. The values show FreeSurfer's cross-sectional analysis to be the smoothest by far, i.e. the least prone to jerky fluctuations between time points.

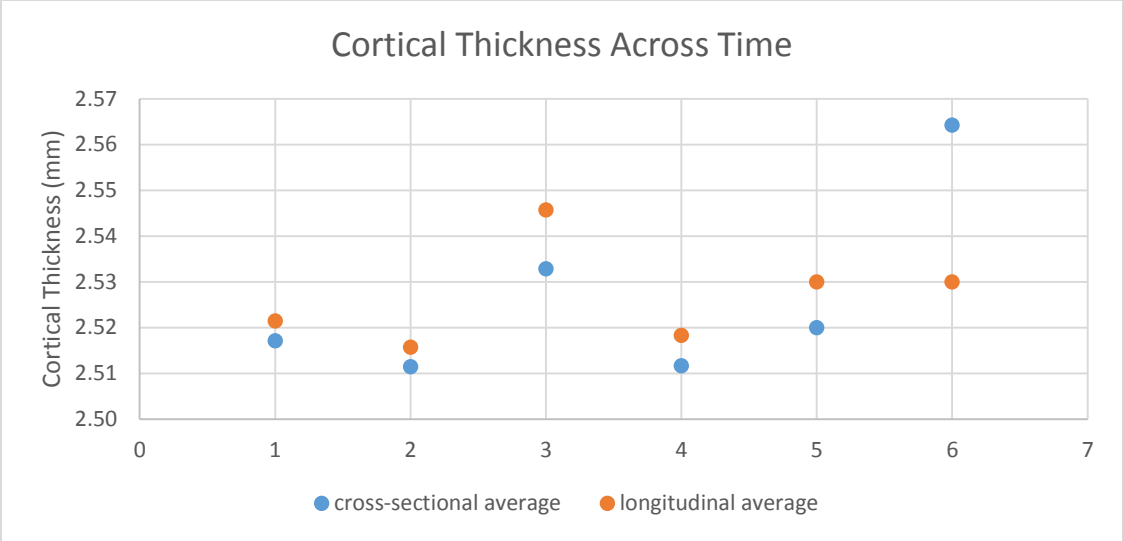

Figure 6: Cortical Thickness Values. With the exception of the last time point, the FreeSurfer-derived cortical thicknesses of the two pipelines nearly paralleled each other. The cross-sectional's sixth time point is most likely influenced by the standard deviations as cortical thickness is on the order of 0.4 – 0.9 mm.

DISCUSSION

Segmentation and Parcellation Results

By its own admission, aBEAT does not have a surface-based ROI analysis function, and while that is in reference to surface reconstruction tools, it is evident in the anatomical parcellation results. As can be seen in Figure 3B, the Colin27 anatomical labels mask used by aBEAT does not adhere to the sulci and gyri of the T_1 image that it is overlaid upon. The inaccurate segmentation of lesions and regions such as the basal ganglia poses a problem in analysis that would be relevant to MS. Lesion volumes being quantified as gray matter is problematic in that total lesion volume varies widely from patient to patient and that as the disease progresses lesion volume increases whereas gray matter is expected to atrophy, thereby decreasing sensitivity to gray matter volume changes. As the basal ganglia is involved with motor control, it

is a region of interest for MS and therefore complete, consistent segmentation of this specific ROI is important.

Thus far evaluation shows that longitudinal processing for gray matter segmentation and atrophy analysis is not beneficial despite its intent to be so. The R^2 values that correspond to smoothness showed FreeSurfer's cross-sectional analysis provided the smoothest or least stochastic changes between yearly image acquisitions. The difference in results found in the FreeSurfer cortical thickness and gray matter volume analyses are likely due to the difference in nature of each metric. Average cortical thickness is approximately 2.5 mm (Figure 6), and standard deviations within any given cortical region can range from 0.4 mm to 0.9 mm. This variability and outliers in particular can hugely sway the results. Total gray matter volume, on the other hand, is a singular measure that is much larger in comparison. Therefore the variability of a cross-sectional segmentation program is more likely to have a larger effect on cortical thickness measurements than gray matter atrophy. Longitudinal processing may therefore prove more useful for cortical thickness analysis rather than that of gray matter volume. This is due to the fact that longitudinal processing was intended in part for better researching subtle disease effects that may be hidden or convoluted by a program's inherent variability (Reuter and Fischl, 2012).

Though atrophy was expected, both the gray matter volume and cortical thickness results showed unexpected increases at various time points (Figures 5 and 6). Those increases were consistent between programs and can either indicate that the MS subject population for this study is unusual or that all the segmentation

methods, though validated on healthy controls and patients suffering from neurodegenerative diseases, suffer inaccuracies when faced with MS. The latter is unlikely to be true, however, as the gray matter metric fluctuations over time were consistent between programs and between cross-sectional and longitudinal segmentation processing results. This indicates that the fluctuations are data-driven and are possibly due to presently uncharacterized physiologic changes.

Additional Considerations

The sample size is notably small; however, this study was not intended to derive any new meaning from the results but rather to determine best practices in a larger cohort. Even with a larger sample size, validation of such segmentation methods proves to be challenging, as all accepted, known trends of brain volume changes in MS are based on imaging. Therefore, correlating segmentation results to previously observed and currently accepted trends will be in essence to correspond the results with the image processing pipeline used to discover that trend and not necessarily what is actually happening physiologically. A similar study completed by Durand-Dubief et al. (2012) evaluated methods for measuring brain volume loss in MS patients, though they only studied cross-sectional segmentation programs, and likewise made a point that it was examining robustness of methods, not accuracy. The lack of a validated gold standard that establishes truth is a fundamental limitation that presently does not have a solution.

An important attribute of the programs is how automatable they are, especially when the need for large scale batch processing arises. FreeSurfer operates

on command lines, whereas aBEAT is a GUI-based program. GUIs require constant human attention, and particularly with aBEAT additional attention is needed for quality control. It was not uncommon for one of the processing steps in aBEAT to fail for one out of a patient's six scans. In such instances the program would continue on and display no error or warning. Additionally, FreeSurfer calculates more measures (e.g. volume, thickness, curvature, etc), whereas aBEAT only calculates volumes, though cortical thickness can be derived with additional post-processing in MATLAB. It should be noted, however, that aBEAT is much newer than FreeSurfer and that it plans to incorporate surface-based ROI analysis and cortical thickness calculations into its next version. The only aspect in which aBEAT is presently preferable to FreeSurfer is computational time. Reported processing time for FreeSurfer is 20 to 40 hours per subject per time point for cross-sectional processing on an AMD Opteron 64bit 2.5GHz processor. aBEAT uses a parallel computing strategy and takes 6.7 hours for longitudinal processing of a single subject with four time points on 8 CPU cores (Intel Xeon, 2.4 GHz) in Linux operating system, and cross-sectional processing takes even less time (Dai et al., 2013).

Future Directions

As all images were obtained from the same scanner, future directions include further validating the processing pipeline with the inclusion of images from other scanners and sites. Durand-Dubief et al. (2012) previously found FreeSurfer's cross-sectional processing to show "significant differences in [brain volume] percentages between… 2 sites". Batch processing will also be explored as the sequential processing of the

lab's entire data set (3000+ images with more incoming) via FreeSurfer would take years. Therefore, alternate grid options are being considered. Additional future work with regards to FreeSurfer lies in further analysis of the longitudinal segmentation in cortical thickness analysis and verifying if it truly is or is not beneficial in a larger cohort.

Additional research is also needed in lesion segmentation. Manual segmentation is a slow and painstaking process that often holds up the rest of image processing pipeline that requires pre-made lesion masks. FreeSurfer shows promise, though it is limited by its single input of a T_1-weighted image, as some lesions are only seen on T_2 or FLAIR images, and even then approximately 33% of pathologically identified white matter lesions are missed (Honce, 2013). Investigation into manual segmentation is currently ongoing to determine inter-operator variability as well as the differences between lesion segmentation based on T_2 vs. FLAIR images. Despite ongoing efforts, manual lesion segmentation followed by cross-sectional segmentation remains the current, optimal approach.

CONCLUSION

We analyzed two software programs for their ability to automatically segment brain tissue. FreeSurfer proved to be more accurate than aBEAT for tissue segmentation and anatomical parcellation. In addition, FreeSurfer's cross-sectional analysis pipeline showed the best results for analyzing gray matter volume changes in the seven longitudinal data sets in this study. Longitudinal analysis did not appear to

provide a significant improvement in gray matter volume analysis for either of the two programs, though FreeSurfer's longitudinal pipeline showed some promise in cortical thickness analysis. Lesion segmentation remains an ongoing issue to be addressed with further research, which also includes further exploration of new longitudinal segmentation programs as they arise. As it stands, optimal segmentation in MS patients requires cross-sectional processing with manual lesion segmentation.

REFERENCES

1. Anderson VM, Fisniku LK, Altmann DR, Thompson AJ, Miller DH. *MRI measures show significant cerebellar gray matter volume loss in multiple sclerosis and are associated with cerebellar dysfunction*. Mult Scler. 2009 Jul; 15(7): 811-7.
2. Ashburner, J., Friston, K.J. *Unified Segmentation*. Neuroimage 2005; 26: 839-851.
3. Ceccarelli, A., Bakshi, R., Neema, M. *MRI in Multiple Sclerosis: A Review of the Current Literature*. Curr Opin Neurol. 2012; 25(4): 402-409.
4. Dai, Y., Wang, Y., Wang, L., Wu, G., Shi, F., Shen, D., ADNI. *aBEAT: a toolbox for consistent analysis of longitudinal adult brain MRI.* PLoS One. 2013; 8(4): e60344.
5. de Boer, R., Vrooman, H.A., Ikram, M.A., Vernooij, M.W., Breteler, M.M.B., van der Lugt, A., Niessen, W.J. *Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods.* Neuroimage. 2010 Jul 1;51(3):1047-56.
6. Durand-Dubief et al. *Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques*. AJNR Am J Neuroradiol. 2012 Nov; 33(10): 1918-24.
7. Fillipi, M., Rocca, M.A. *MR imaging of gray matter involvement in multiple sclerosis: implications for understanding disease pathophysiology and monitoring treatment efficacy.* AJNR Am J Neuroradiol. 2010 Aug; 31(7): 1171-1177.
8. Honce, *Gray Matter Pathology in MS: Neuroimaging and Clinical Correlations.* Multiple Sclerosis International. 2013; Volume 2013, Article ID 627870, 16 pages. doi:10.1155/2013/627870.
9. Horakova, D., Kalincik, T., Dusankova, J.B., Dolezal, O. *Clinical correlates of grey matter pathology in multiple sclerosis.* BMC Neurology 2012, 12:10. doi: 10.1186/1471-2377-12-10.
10. Jenkinson, M., Bannister, P., Brady, J. M. and Smith, S. M. *Improved Optimisation for the Robust and Accurate Linear Registration and Motion Correction of Brain Images*. NeuroImage, 2002; 17(2), 825-841.
11. National Multiple Sclerosis Society. *National Multiple Sclerosis Society.* Retrieved August 29, 2013, from http://www.nationalmssociety.org
12. Reuter, M., Fischl, B. *Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis*. NeuroImage 2012; 61(4): 1402-1418.
13. Scahill, R.I., Frost, C., Jenkins, R., Whitewell, J.L., Rossor M.N., Fox, N.C. *A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging*. Arch Neurol. 2003; 60(7):989-994.
14. Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N. *Accurate, robust and automated longitudinal and cross-sectional brain change analysis*. NeuroImage 2002; 17(1), 479-489.
15. Styner, M., Lee, J., Chin, B., Chin, M.S., Commowick, M., Tran, H-H., Jewells, V., Warfield, S. *3D Segmentation in the Clinic: A Grand Challenge II: MS Lesion Segmentation*. 2008.

16. Talairach, J., Tournoux, P. *Co-planar Sterotaxic Atlas of the Human Brain: 3-Dimensional Proportional System – an Approach to Cerebral Imaging*. Thieme Medical Publishers, New York, NY, 1988.
17. Zhang, Y., Brady, M., Smith, S. *Segmentation of brain MR mages through a hidden Markov random field model and the expectation-maximization algorithm*. IEEE Trans Med Imag 2001; 20(1): 45-47.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

9/9/13
Date