

The ANTs Longitudinal Cortical Thickness Pipeline

Nicholas J. Tustison^{1,2}, Andrew J. Holbrook³, Brian B. Avants⁴, Jared M. Roberts², Philip A. Cook⁵, James R. Stone¹, Daniel L. Gillen³, and Michael A. Yassa² for the Alzheimer's Disease Neuroimaging Initiative*

¹Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

²Department of Neurobiology and Behavior, University of California, Irvine, Irvine, CA

³Department of Statistics, University of California, Irvine, Irvine, CA

⁴Biogen, Cambridge, MA

⁵Department of Radiology, University of Pennsylvania, Philadelphia, PA

Corresponding author:

Nicholas J. Tustison

211 Qureshey Research Lab

Irvine, CA 92697-3800

ntustison@virginia.edu

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

Large-scale longitudinal studies of developmental progression or disease in the human brain have motivated the acquisition of large neuroimaging data sets and the concomitant development of robust methodological and statistical tools for insight into potential neurostructural changes. Longitudinal-specific strategies for acquisition and processing have potentially significant benefits including the reduction of the inter-subject confound associated with cross-sectional studies. In this work, we introduce the open-source Advanced Normalization Tools (ANTs) cortical thickness longitudinal processing pipeline and its application on the first phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) comprising over 600 subjects with multiple time points from baseline to 36 months. We demonstrate that the single-subject template construction and native subject-space processing advantageously localizes data transformations and reduces interpolation artifacts and is the preferred processing strategy with respect to simultaneous minimization of within-subject variability and maximization of between-subject variability, respectively. It is further shown that optimizing these dual criteria leads to greater scientific interpretability in terms of tighter confidence intervals in calculated mean trends, smaller prediction intervals, and tighter confidence/credible intervals for determining cross-sectional effects. A complementary machine learning evaluation provides additional evidence of the benefits of this framework.

Keywords: ANTs, Alzheimer's disease, bias, cortical thickness, interpolation, longitudinal processing

Introduction

Quantification of brain morphology significantly facilitates the investigation of a wide range of neurological conditions with structural correlates, including neurodegenerative conditions such as Alzheimer’s disease [1, 2]. Essential for thickness quantification are the many computational techniques which have been developed to provide accurate measurements of the cerebral cortex. These include various mesh-based (e.g., [3–5]) and volumetric techniques (e.g., [6–9]). Of noted significance, and representing the former, is the well-known and highly utilized FreeSurfer software package [10–14].

In inferring developmental processes, many studies employ cross-sectional population sampling strategies despite the potential for confounding effects [15]. Large-scale studies involving longitudinal image acquisition of a targeted subject population, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [16], are designed to mitigate some of the relevant statistical issues. Analogously, much research has been devoted to exploring methodologies for properly exploiting such studies and avoiding various forms of processing bias [17]. For example, FSL’s SIENA (Structural Image Evaluation, using Normalization, of Atrophy) framework [18] for detecting atrophy between longitudinal image pairs avoids a specific type of processing bias by transforming the images to a midspace position between the two time points. As the authors point out “[i]n this way both images are subjected to a similar degree of interpolation-related blurring.” Consequences of this “interpolation-related blurring” were formally analyzed in [19] in the context of hippocampal volumetric change where it was shown that interpolation-induced artifacts can artificially create and/or inflate effect size [20]. These insights have since been used for making specific recommendations with respect to longitudinal image data processing [17, 21, 22].

In [17, 23] the authors motivated the design and implementation of the longitudinal FreeSurfer variant inspired by these earlier insights and the overarching general principle of “treat[ing] all time points exactly the same.” It has since been augmented by integrated linear mixed effects modeling capabilities [24] and has been used in a variety of studies including pediatric cortical development [25], differential development in Alzheimer’s disease and fronto-temporal dementia [26], and fatigue in the context of multiple sclerosis [27].

In [28], we introduced the Advanced Normalization Tools (ANTs) cortical thickness framework which leverages various pre-processing, registration, segmentation, and other image analysis tools that members of the ANTs and Insight Toolkit (ITK) open-source communities have developed over the years and disseminated publicly [29]. This proposed ANTs-based pipeline has since been directed at a variety of neuroimaging research topics including mild cognitive impairment and depression [30], short term memory in mild cognitive impairment [31], and aphasia [32]. In this work, we introduce the longitudinal version of the ANTs cortical thickness pipeline and demonstrate its utility on the publicly available ADNI-1 data set. In addition, we demonstrate that certain longitudinal processing choices have significant impact on measurement quality in terms of within-subject and between subject variances which, in turn, heavily impacts the scientific interpretability of results. Similar to other research illustrating the negative impact of interpolation effects on study results, we show that a common practice for unbiased processing induces a different set of problematic artifacts which guides processing choices for the proposed ANTs longitudinal pipeline. These choices for the ADNI-1 data produce tighter confidence intervals in calculated mean trends, smaller prediction intervals, and less varied confidence/credible intervals for discerning cross-sectional effects. To explore these findings in a more clinically oriented context, we also use a machine learning-based train-

ing/prediction paradigm which demonstrates that the recommended longitudinal processing approach leads to improved predictive diagnostic accuracy over the alternative strategies, including cross-sectional processing.

Methods and materials

ADNI-1 imaging data

The strict protocol design, large-scale recruitment, and public availability of the Alzheimer’s Disease Neuroimaging Initiative (ADNI) makes it an ideal data set for evaluating the ANTs longitudinal cortical thickness pipeline. An MP-RAGE [33] sequence for 1.5 and 3.0 T was used to collect the data at the scan sites. Specific acquisition parameters for 1.5 T and 3.0 T magnets are given in Table 1 of [34]. As proposed, collection goals were 200 elderly cognitively normal subjects collected at 0, 6, 12, 24, and 36 months; 400 MCI subjects at risk for AD conversion at 0, 6, 12, 18, 24, and 36 months; and 200 AD subjects at 0, 6, 12, and 24 months.

The ADNI-1 data was downloaded in May of 2014. The data was first processed using the ANTs cross-sectional cortical thickness pipeline [28] (4399 total images). Data was then processed using two variants of the ANTs longitudinal stream (described in the next section). In the final set of csv files (which we have made publicly available [35] in the github repository associated with this work), we only included time points for which clinical scores (e.g., MMSE) were available. In total, we included 186 elderly cognitive normals, 178 MCI subjects, 128 LMCI subjects, and 123 AD subjects. Further breakdown of demographic information is given in Figures 1 and 2 to provide additional perspective on the data used for this work.

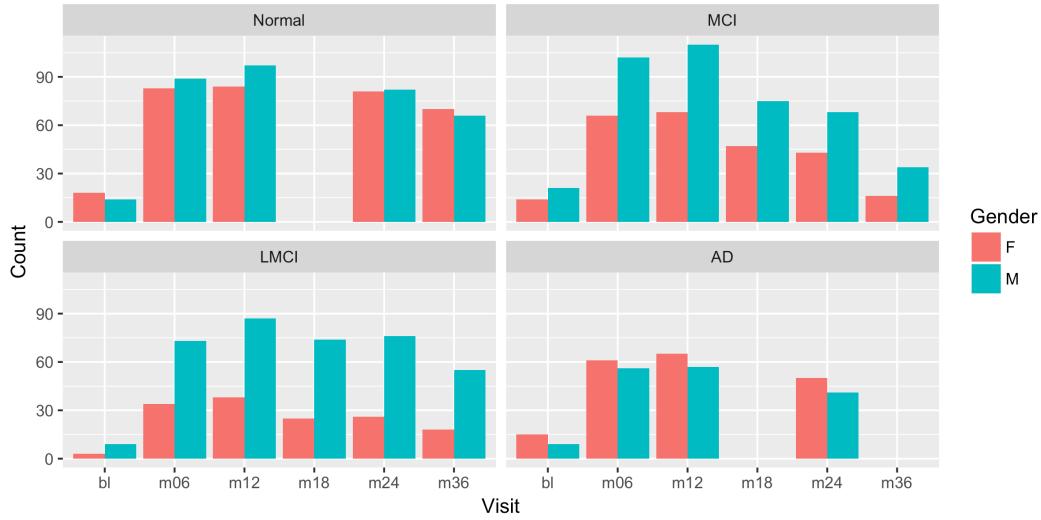


Figure 1: Demographic breakdown of the number of ADNI-1 subjects by diagnosis i.e., normal, mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). Within each panel we plot the number of subjects (by gender) per visit—baseline (“bl”) and n months (“mn”).

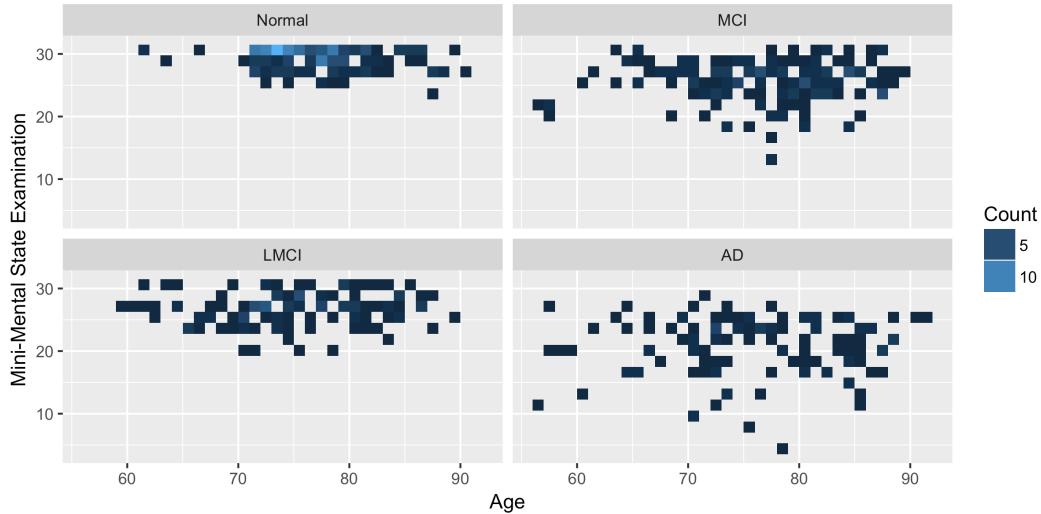


Figure 2: Age vs. Mini-mental examination (MMSE) scores for the ADNI-1 subjects by diagnosis.

ANTs cortical thickness

Cross-sectional processing

A thorough discussion of the ANTs cross-sectional thickness estimation framework was previously discussed in [28]. As a brief review, given a T1-weighted brain MR image, processing comprises the following major steps (cf Figure 1 of [28]):

1. N4 bias correction [36],
2. brain extraction [37],
3. Atropos n -tissue segmentation [38], and
4. cortical thickness estimation [8].

ROI-based quantification is achieved through the use of the joint label fusion approach of [39] and the use of the MindBoggle-101 data labeled using the Desikan–Killiany–Tourville (DKT) protocol [40] consisting of 31 labels per hemisphere (cf Table 1). This pipeline has since been enhanced by the implementation [41] of a patch-based denoising algorithm [42] as an optional preprocessing step and multi-modal integration capabilities (e.g., joint T1- and T2-weighted processing).

For evaluation, voxelwise regional thickness statistics were summarized based on the DKT parcellation scheme. Test-retest error measurements were presented from a cohort of 20 atlases taken from the OASIS data set which had been manually labeled [40] and compared with the corresponding FreeSurfer thickness values. Further evaluation employed a training/prediction paradigm whereby DKT regional cortical thickness values generated from 1205 images taken from four publicly available data sets (i.e., IXI [43], MMRR [44], NKI [45], and OASIS [46]) were used to predict age and gender using linear and random forest [47] models. The resulting regional statistics (including cortical thickness, surface area [48], volumes, and Jacobian determinant values) were made available online [49]. These include the corre-

Table 1: The 31 cortical labels (per hemisphere) of the Desikan-Killiany-Tourville atlas. The ROI abbreviations from the R brainGraph package are given in parentheses and used in later figures.

| | |
|-------------------------------------|---------------------------------------|
| 1) caudal anterior cingulate (cACC) | 17) pars orbitalis (pORB) |
| 2) caudal middle frontal (cMFG) | 18) pars triangularis (pTRI) |
| 3) cuneus (CUN) | 19) pericalcarine (periCAL) |
| 4) entorhinal (ENT) | 20) postcentral (postC) |
| 5) fusiform (FUS) | 21) posterior cingulate (PCC) |
| 6) inferior parietal (IPL) | 22) precentral (preC) |
| 7) inferior temporal (ITG) | 23) precuneus (PCUN) |
| 8) isthmus cingulate (iCC) | 24) rostral anterior cingulate (rACC) |
| 9) lateral occipital (LOG) | 25) rostral middle frontal (rMFG) |
| 10) lateral orbitofrontal (LOF) | 26) superior frontal (SFG) |
| 11) lingual (LING) | 27) superior parietal (SPL) |
| 12) medial orbitofrontal (MOF) | 28) superior temporal (STG) |
| 13) middle temporal (MTG) | 29) supramarginal (SMAR) |
| 14) parahippocampal (PARH) | 30) transverse temporal (TT) |
| 15) paracentral (paraC) | 31) insula (INS) |
| 16) pars opercularis (pOPER) | |

sponding FreeSurfer measurements which are also publicly available for research inquiries (e.g., [50]). Since publication, this framework has been used in a number of studies (e.g., [51–53]).

Unbiased longitudinal processing

Given certain practical limitations (e.g., subject recruitment and retainment), as men-

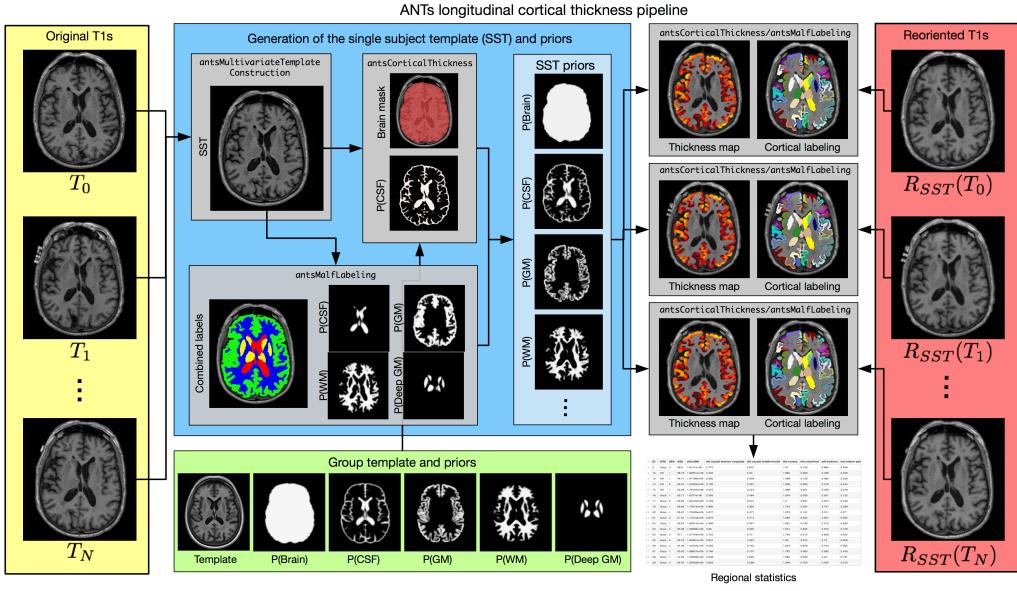


Figure 3: Diagrammatic illustration of the ANTs longitudinal cortical thickness pipeline for a single subject with N time points. From the N original T1-weighted images (left column, yellow panel) and the group template and priors (bottom row, green panel), the single-single subject template (SST) and auxiliary prior images are created (center, blue panel). These subject-specific template and other auxiliary images are used to generate the individual time-point cortical thickness maps, in the individual time point's native space (denoted as “Longitudinal-native” in the text). Optionally, one can rigidly transform the time-point images prior to segmentation and cortical thickness estimation (right column, red panel)> This alternative processing scheme is referred to as “Longitudinal-SST”. For regional thickness values, regional labels can be propagated to each image using a given atlas set and cortical parcellation scheme.

tioned earlier, many researchers employ cross-sectional acquisition and processing strategies for studying developmental phenomena. Longitudinal studies, on the other hand, can significantly reduce inter-subject measurement variability. The ANTs longitudinal cortical thickness pipeline extends the ANTs cortical thickness pipeline for longitudinal studies which takes into account various bias issues previously discussed in the literature [17, 19, 23].

Given N time-point T1-weighted MR images (and, possibly, other modalities) and representative images to create a population-specific template and related images, the longitudinal pipeline consists of the following steps:

1. (Offline): Creation of the group template and corresponding prior probability images.
2. Creation of the unbiased single-subject template (SST).
3. Application of the ANTs cross-sectional pipeline to the SST.
4. Creation of the SST prior probability maps.
5. (Optional): Rigid transformation of each individual time point to the SST.
6. Application of the ANTs cross-sectional pipeline to each individual time-point image.
7. Joint label fusion to determine the cortical ROIs for analysis.

An overview of these steps is provided in Figure 3 which we describe in greater detail below.

ADNI group template, brain mask, and tissue priors. Prior to any individual subject processing, the group template is constructed from representative population data [54]. For the ADNI-1 processing described in this work, we created a population-specific template from 52 cognitively normal ADNI-1 subjects. Corresponding brain and tissue prior probability maps for the CSF, gray matter, white matter, deep gray matter, brain stem, and cerebellum were created as described in

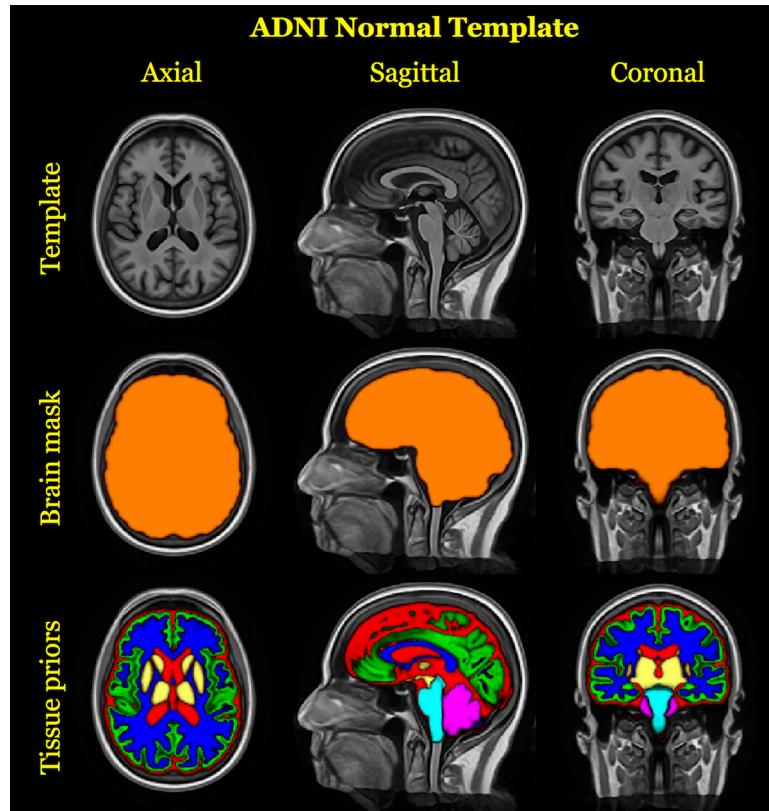


Figure 4: Top row: Canonical views of the template created from 52 cognitively normal subjects of the ADNI-1 database. The prior probability mask for the whole brain (middle row) and the six tissue priors (bottom row) are used to “seed” each single-subject template for creation of a probabilistic brain mask and probabilistic tissues priors during longitudinal processing.

[28]. A brief overview of this process is also provided in the section concerning creation of the single-subject template. Canonical views of the ADNI-1 template and corresponding auxiliary images are given in Figure 4.

Single-subject template, brain mask, and tissue priors. With the ADNI-1 group template and prior probability images, each subject undergoes identical processing. First, an average shape and intensity single subject template (SST) is created from all time point images using the same protocol [54] used to produce the ADNI-1 group template. Next, six probabilistic tissue maps (cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), deep gray matter (striatum + thalamus), brain stem, and cerebellum) are generated in the space of the SST. This requires processing the SST through two parallel workflows. First, the SST proceeds through the standard cross-sectional ANTs cortical thickness pipeline which generates a brain extraction mask and the CSF tissue probability map, $P_{Seg}(CSF)$. Second, using a data set of 20 atlases from the OASIS data set that have been expertly annotated [40], a multi-atlas joint label fusion step (JLF) [39] is performed to create individualized probability maps for all six tissue types. Five of the JLF probabilistic tissue estimates (GM, WM, deep GM, brain stem, and cerebellum) and JLF CSF estimate, $P_{JLF}(CSF)$, are used as the SST prior probabilities after smoothing with a Gaussian kernel (isotropic, $\sigma = 1mm$) whereas the CSF SST tissue probability is derived as a combination of the JLF and segmentation CSF estimates, i.e., $P(CSF) = \max(P_{Seg}(CSF), P_{JLF}(CSF))$, also smoothed with the same Gaussian kernel. Finally, $P(CSF)$ is subtracted out from the other five tissue probability maps. The final version of the SST and auxiliary images enable unbiased mappings to the group template, subject-specific tissue segmentations, region of interest volumes and cortical thickness maps for each time point of the original longitudinal image series.

Individual time point processing. In the FreeSurfer longitudinal stream, each

time-point image is processed using the FreeSurfer cross-sectional stream. The resulting processed data from all time points is then used to create a mean, or median, single-subject template. Following template creation, each time-point image is rigidly transformed to the template space where it undergoes further processing (e.g., white and pial surface deformation). This reorientation to the template space “further reduce[s] variability” and permits an “implicit vertex correspondence” across all time points [17].

The ANTs longitudinal workflow shares some common aspects of its FreeSurfer analog but differs in others as outlined above. The first step for subject-wise processing involves the creation of an optimal mean shape/intensity template from all the time points [54]. For the cross-sectional ANTs processing, the group template and auxiliary images are used to perform tasks such as individual brain extraction and n -tissue segmentation prior to cortical thickness estimation [28]. However, for the longitudinal variant, the group template is used to create the SST auxiliary images. We then map the SST and corresponding probabilistic tissue maps to the native space of each time point where segmentation and cortical thickness is estimated. Note that this unbiased longitudinal pipeline is completely agnostic concerning ordering of the input time-point images, i.e., we “treat all time points exactly the same.”

During the initial development of this work, it was thought that rotating the individual time points to the SST would be of benefit, similar to FreeSurfer, in reducing variability, minimizing or eliminating possible orientation bias, and permitting a 4-D segmentation given that the underlying Atropos segmentation implementation is dimensionality-agnostic [38]. Regarding the 4-D brain segmentation, any possible benefit is potentially outweighed by the occurrence of “over-regularization” [17] whereby smoothing across time reduces detection ability of large time-point changes. Additionally, it is less than straightforward to accommodate irregular temporal sam-

pling such as the acquisition schedule of the ADNI-1 protocol.

Joint label fusion and pseudo-geodesic for large cohort labeling. Cortical thickness ROI-based analyses are performed using joint label fusion [39] and whatever cortical parcellation scheme is deemed appropriate for the specific study. The brute force application of the joint label fusion algorithm would require N pairwise registrations for each time-point image where N is the number of atlases used. This would require a significant computational cost for a relatively large study such as ADNI. Instead, we use the “pseudo-geodesic” approach for mapping atlases to individual time point images. The transformations between the atlas and the group template are computed offline. With that set of transforms, we are able to concatenate a set of existing transforms from each atlas through the group template, to the SST, and finally to each individual time point.

Statistical evaluation

Given the above description, there are three major processing choices for using ANTs to perform a longitudinal cortical thickness study. Given its commonality, we could simply use the cortical thickness processing [28]. This has the advantage of requiring significantly less computation but does not address the various issues due to general cross-sectional sampling alluded to earlier. The second processing option is to re-orient each time-point image to the SST similar to FreeSurfer which addresses such issues as orientation bias. The third option is to process each time-point image in its native space while maintaining the use of the SST and probability images. Although the latter is susceptible to possible orientation bias, it minimizes the effects of interpolation artifacts. We denote these alternatives as:

- **Cross-sectional.** Process each subject’s time point independently using the

cross-sectional pipeline originally described in [28].

- **Longitudinal-SST.** Rigidly transform each subject to the SST and then segment and estimate cortical thickness in the space of the SST.
- **Longitudinal-native.** Segment and estimate cortical thickness in the native space.

In this section we describe the statistical framework for evaluating each option.

Regional within-subject and between-subject variance

To quantify the relative performance of these cross-sectional and longitudinal processing methods as a biomarker we considered a summary measure related to intra-class correlation. Specifically, we said that one processing method outperforms the other when it does a better job minimizing within-subject variability and maximizing between-subject variability in cortical thickness measurements. Such a quality implies greater within-subject reproducibility while distinguishing between patient subpopulations. As such this will amount to higher precision when cortical thickness is used as a predictor variable or model covariate in statistical analyses upstream. This criterion is immediately estimable from the longitudinal mixed-effects model (1) outlined below.

Longitudinal mixed-effect (LME) models comprise a well-established and widely used class of regression models designed to estimate cross-sectional and longitudinal linear associations between quantities while accounting for subject specific trends. As such, these models are useful for the analysis of longitudinally collected cohort data. Indeed, [24] provide an introduction to the mixed-effects methodology in the context of longitudinal neuroimaging data and compare it empirically to competing methods such as repeated measures ANOVA. For more complete treatments of the subject matter, see [55] and [56]. LME models are also useful for estimating and comparing within-subject and between-subject variability after conditioning

out systematic time trends in longitudinally measured data. In the context of the current investigation, by fitting simple LME models to the data resulting from cross-sectional and longitudinal processing techniques, we are able to quantify the relative performance of each approach with respect to within-subject, between-subject, and total variability in a way that [57] only hint at in their exposition of the longitudinal FreeSurfer stream.

As previously noted we observed a longitudinal sampling of cortical thickness measurements from 62 separate regions of interest. To assess the above variability-based criteria while accounting for changes that may occur through the passage of time, we used a Bayesian LME model for parameter estimation. Let Y_{ij}^k denote the i^{th} individual's cortical thickness measurement corresponding to the k^{th} region of interest at measurement j . Under the Bayesian paradigm we utilized a model of the form

$$\begin{aligned} Y_{ij}^k &\sim N(\alpha_i^k + \beta^k t, \sigma_k^2) \\ \alpha_i^k &\sim N(\alpha_0^k, \tau_k^2) \quad \alpha_0^k, \beta^k \sim N(0, 10) \quad \sigma_k, \tau_k \sim \text{Cauchy}^+(0, 5) \end{aligned} \tag{1}$$

Specification of variance priors to half-Cauchy distributions reflects commonly accepted practice in the context of hierarchical models [58]. In this model, τ_k represents the between-subject standard deviation, and σ_k represents the within-subject standard deviation, conditional upon time. For each region k , the quantity of interest is thus the ratio

$$r^k = \frac{\tau_k}{\sigma_k}, \quad k = 1, \dots, 62. \tag{2}$$

This ratio is at the heart of classical statistical discrimination methods as it features both in the ANOVA methodology and in Fisher's linear discriminant analysis. These

connections are important since the utility of cortical thickness as a biomarker lies in the ability to discriminate between patient sub-populations with respect to clinical outcomes. It is also similar to the intra-class correlation coefficient [55]. The posterior distribution of r^k was summarized via the posterior median where the posterior distributions were obtained using the Stan probabilistic programming language [59].

Entorhinal cortical thickness variability

The entorhinal cortex (EC) is one of the earliest regions to exhibit tau pathology in the Alzheimer's brain and is one of the first regions to show signs of neurodegenerative change [60–63]. In the ADNI sample, EC cortical thickness was the most powerful measure of structural change both in aMCI and AD brains [64]. EC thinning was also found to precede and predict hippocampal atrophy [65] and to predict conversion to AD with the greatest accuracy [66]. Thus, we chose the EC to be the target of an additional focused analysis to determine the relative utility of the different pipelines for measuring thickness in this particular region. It is worthy of note that the EC also presents a special challenge as its medial location in the brain makes it very difficult to segment from neighboring vasculature. Thus, it is particularly vulnerable to segmentation errors which can inflate its thickness. Our choice of EC as a region to demonstrate proof of concept was thus motivated both by its selective vulnerability to neurodegenerative processes as well as the difficulty of its segmentation.

As a further assessment of utility as a biomarker, we used LME models and cortical thickness measurements of the EC to demonstrate how these variability criteria relate to potential scientific analyses. First, we used model (1) to show that a greater ratio of between-subject to within-subject variability results in tighter confidence and credible intervals on the slope parameter β . This result indicates more confidence with respect to mean trends over time that are of common interest when comparing subpopulations of patients. Second, we showed that smaller within-subject variabil-

ity corresponds to smaller prediction intervals when predicting a subject's cortical thickness levels at future visits. This is important when considering regional cortical thickness measures as candidate biomarkers. Third, we use a simple linear regression model to compare the relationship between total variance and uncertainty with respect to cross-sectional effects. To do so, we regress baseline cortical thickness in the entorhinal cortex (EC) over baseline AD diagnostic status:

$$ECCT_i = \beta_0 + \beta_1 AD_i + \epsilon_i . \quad (3)$$

In general, lower total variability corresponds to tighter confidence/credible intervals for cross-sectional covariate effects, and hence higher certainty when evaluating linear associations between quantities such as cortical thickness and AD status. If total variability is similar across processing methods, we would expect to see credible intervals of roughly the same size.

Diagnostic prediction via extreme gradient boosting

As mentioned earlier, an important component of our previously reported cross-sectional comparative evaluation [28] incorporated a statistical modeling approach for predicting basic subject demographics (i.e., age and gender) from the summary DKT regional cortical thickness values. We used a similar evaluation strategy in this work with the ADNI-1 data. We built statistical models from cortical thickness values for predicting the ADNI-specified diagnosis. However, instead of using the regional cortical thickness values for all time points directly, we used the longitudinal cortical thickness values for each DKT region of each subject to compute a subject-specific, region-specific rate-of-thickness-change measurement generated from simple linear regression of the available data. Thus, for each subject, we calculated a single set of 62 slope coefficients, $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_{62}\}$, and a single diagnosis of cognitively nor-

mal, MCI, LMCI, or AD. Note that the diagnosis for each subject did not change over the course of the image acquisition schedule permitting this particular evaluative strategy.

This classification scenario yields the following model:

$$DIAGNOSIS \sim \sum_{k=1}^{62} \zeta_k \quad (4)$$

The motivating idea is that regional thinning is accelerated in some regions versus others which should be reflected in the assigned diagnostic category [64, 65].

For model construction, we used extreme gradient boosting, which is a well-performing, out-of-the-box classifier implemented in the XGBoost [67] package for the R project. Although there are many options available for classification, we chose this particular technique due to our recent interest in other projects based on its various successes in other fields.¹ An additional advantage is that XGBoost provides the “gain” quantity which describes “the improvement in accuracy brought by a feature to the branches [of the tree or iteration] it is on [67].” These can be reviewed for clinical plausibility of the results.

Results

Based on the evaluation design described in the previous section, we compare the performance of the three processing approaches (Cross-sectional, Longitudinal-SST, and Longitudinal-native) as applied to the ADNI-1 data. Specifically, we demonstrate how the variance ratio defined in Equation (2) illustrates ways in which different aspects of variability affect confidence in prediction and estimation.

¹<http://stat-computing.org/awards/jmc/winners.html>

Regional within-subject and between-subject variance

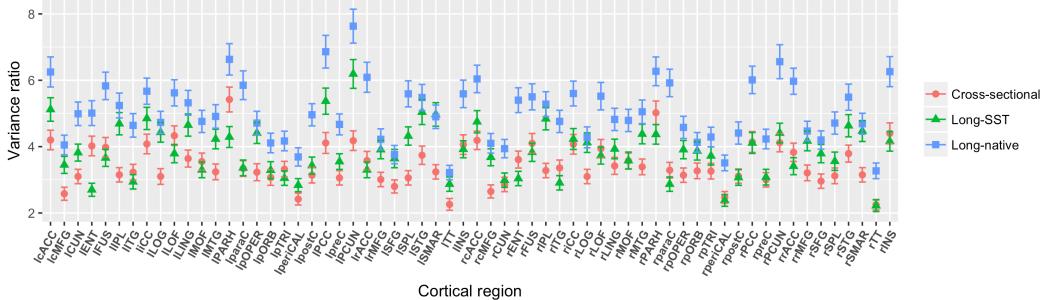


Figure 5: 95% credible intervals of the region-specific variance ratios $r^k = \tau_k/\sigma_k$ are presented for each processing method. The Longitudinal-native method dominates the others: its point estimates—posterior median—are greater than those of the other processing methods; and its credible intervals scarcely overlap with those of Longitudinal-SST and never overlap with those quantities calculated from the cross-sectional method. These results suggest that the cortical thickness values obtained using Longitudinal-native have greater discriminative capacity than the thickness values obtained using the other two methods.

Our first evaluation strategy was to use LME models to quantify the between-subject and within-subject variance with the expectation that maximizing the former while minimizing the latter optimizes measurement quality in terms of prediction and confidence intervals. Figure 5 provides the resulting 95% credible intervals for the distributions of region-specific variance ratios $r^k = \tau_k/\sigma_k$ for each of the three methods. The placement of the methods with respect to each other is meaningful. The superior method is designated by larger variance ratios and has the greater discriminative capacity for the data corresponding to that processing method.

Longitudinal-native has the highest ratio variance across all 62 regions over the two alternative methods. It rarely overlaps with Longitudinal-SST and never with Cross-sectional. In contrast, even though Longitudinal-SST has a greater number of superior ratio values across the regions relative to the Cross-sectional, exceptions include

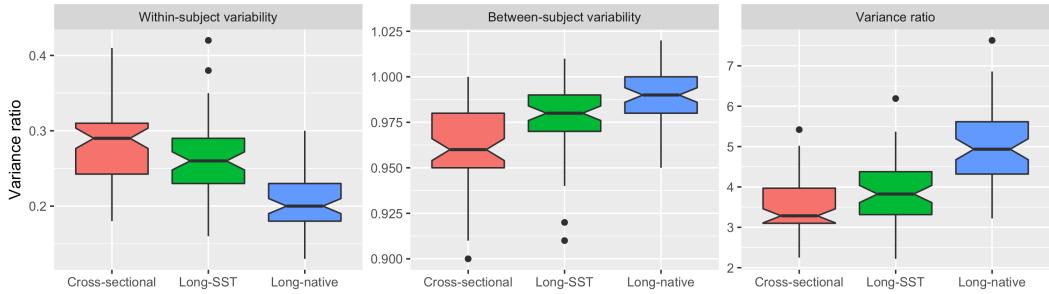


Figure 6: Notched box plots showing the distribution of the within-subject variability, between subject variability, and ratio of the between-subject variability and within-subject variability for each of the 62 DKT regions. Note that the “better” measurement maximizes this latter ratio.

the left and right entorhinal, left and right fusiform, left and right inferior temporal, left and right paracentral, left and right parahippocampal, left and right posterior cingulate, left lateral orbitofrontal, left medial orbitofrontal, left pars triangularis, right cuneus, right insula, right isthmus cingulate, right pericalcarine, right posterior cingulate right precentral, right postcentral, and right transverse temporal.

Therefore, Figures 5, 6, and 7 may be considered as evidence for method Longitudinal-native providing higher quality data than that provided by the other methods. The plot in Figure 6 shows a relative summary of all the regional quantities for all three variance measurements (within-subject, between-subject, and variance ratio) via notched box plots. These relative distributions show that both between-subject and within-subject quantities contribute to the disparities in the ratio evaluation metric. Finally, we overlay the variance ratio values on the corresponding regions of a 3-D rendering of the ADNI template (Figure 7) to provide an additional visual comparison between the methods.

Entorhinal cortical thickness variability

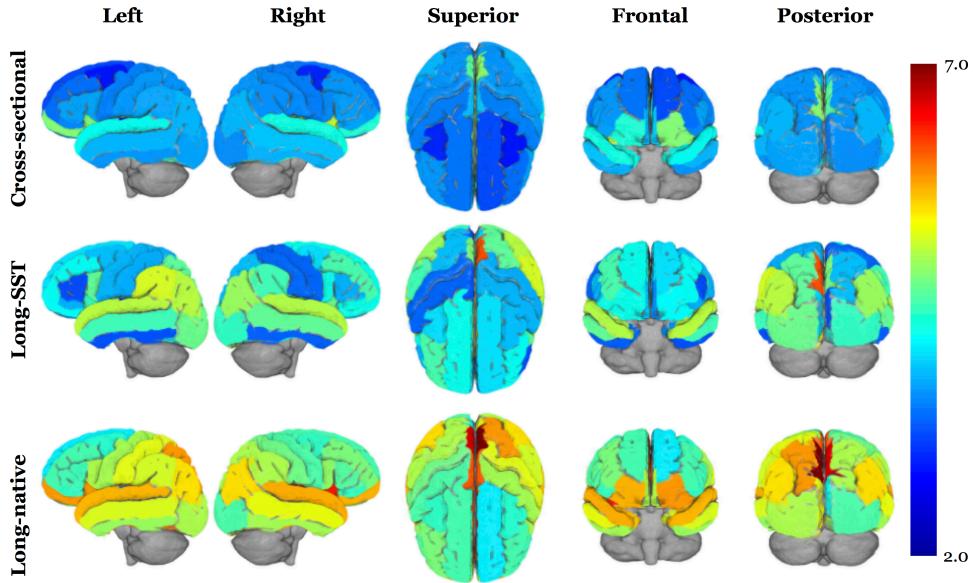


Figure 7: 3-D volumetric rendering of the regional variance ratio values on the generated ADNI template. The higher variance ratios indicate greater between-subject to within-subject variability.

Data quality translates directly to quality of statistical results and the scientific conclusions derived therefrom. Hence, data with good variance and precision properties will benefit statistical analyses in multiple ways. To demonstrate these benefits, we focus on data from the entorhinal cortex and present three different aspects of variability and their statistical upshots. Table 2 presents different aspects of model variability and shows their relationships to uncertainty in prediction and estimation. Model variability is shown in terms of point estimates (posterior medians) for different functions of the variance terms from Model (1). Predictive and estimation uncertainty takes the form of credible interval widths and predictive variance. The larger these quantities, the more uncertainty, and hence the less definite the scientific conclusions reached. Both raw and normalized results are presented. For each quantity, the cells corresponding to highest performance are colored green, and those corre-

Table 2: Processing methods and EC statistical results.

| Method | $\frac{\tau}{\sigma}$ | CI width | σ^2 | Variance | $\sigma^2 + \tau^2$ | CI width |
|------------------------|-----------------------|----------|------------|----------|---------------------|----------|
| Cross-sectional | | | | | | |
| <i>unnormalized</i> | 3.82 | 0.0026 | 0.07 | 0.07 | 1.02 | 0.46 |
| <i>normalized</i> | 0.73 | 0.76 | 0.58 | 0.48 | 0.99 | 1 |
| Long-SST | | | | | | |
| <i>unnormalized</i> | 2.83 | 0.0034 | 0.11 | 0.15 | 1.02 | 0.41 |
| <i>normalized</i> | 0.54 | 1 | 1 | 1 | 0.99 | 0.90 |
| Long-native | | | | | | |
| <i>unnormalized</i> | 5.23 | 0.0019 | 0.04 | 0.04 | 1.03 | 0.42 |
| <i>normalized</i> | 1 | 0.57 | 0.32 | 0.29 | 1 | 0.91 |

sponding to worst performance are colored red.

On the left of Table 2, the variance ratio is presented alongside the width of the credible interval corresponding to the slope parameter β from Model (1). In general, a higher ratio of between-subject and within-subject variances implies greater precision when estimating trends and associations through time. As expected from the previous results regarding the ratio of between- and within-subject variability, Longitudinal-native yields the smallest credible interval on the slope parameter.

In the middle of Table 2, within-subject variability is presented alongside predictive variance, i.e., the median for each subject-specific empirical variance when predicting EC cortical thickness 6 months out from the last observation. As might be expected these two quantities track closely to each other, since prediction variability is an amalgam of within-subject variability and uncertainty in model parameters. Again, the Longitudinal-native method performs best whereas Longitudinal-SST performs worst.

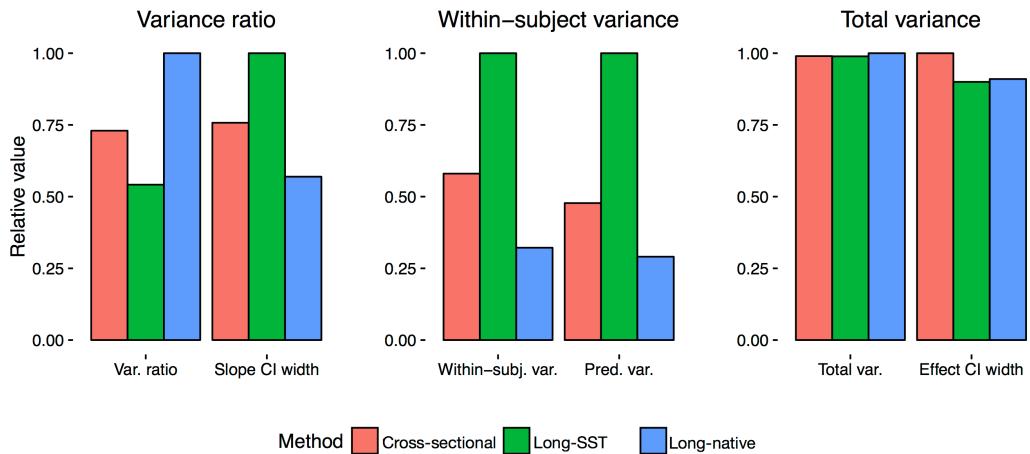


Figure 8: Aspects of model variance are compared with credible interval sizes and variance in predictions. Values are normalized by the largest quantity, and processing methods are distinguished by color and ordering. On the left, the variance ratio $r = \tau/\sigma$ is compared to the width of credible interval for the slope term of Model (1). In the middle, within-subject variance σ^2 is compared to predictive variance. On the right, total variance $\sigma^2 + \tau^2$ is compared to width of credible interval for the cross-sectional association of AD status with EC cortical thickness.

Finally, the right side of Table 2, compares total variance to the width of credible intervals pertaining to the cross-sectional association of AD diagnosis and EC cortical thickness as modeled in Equation (3). As total variance rises, so too does uncertainty in cross-sectional effects. However, all three processing methods achieve roughly the same amount of total variability, so no trend is visible. It is interesting to observe that for this particular example the lower bound of the second longitudinal is farther from the null effect of zero when compared to the other two approaches. That is, despite having marginally greater total variance, the distance from zero for the credible interval corresponding to Longitudinal-native is 0.81, whereas the distances for Longitudinal-SST and Cross-sectional are 0.75 and 0.70, respectively. Figure 8

displays the normalized results.

Diagnostic prediction via extreme gradient boosting

A clinically-based prediction strategy was performed to evaluate the quality of the cortical thickness measurements produced by each method. The rate of thickness change determined over the set of subject imaging visits was used as a feature set for predicting diagnosis. The basic idea is that regional thinning is accelerated in some regions versus others which should be reflected in the assigned diagnostic category.

Using the model described by Equation (4), we compared the classification capabilities of each of the three pipelines. For each of N iterations, the data was randomly split 90/10 (i.e., 90% training and 10% testing) and used to construct three diagnostic classification models from the training data, one for each pipeline. The testing data portion and the corresponding prediction sets were used to construct confusion matrices from which diagnostic accuracy was calculated. Chosen XGBoost model parameters deviating from the default were: number of trees = number of iterations (“nrounds”) = 100 and gradient step (“eta”) = 0.3, based on parameter tuning on a small data subset over all three pipelines.

The resulting accuracy distributions are plotted in Figure 9 and compared statistically using Tukey’s range test which indicated increasing performance Cross-sectional < Longitudinal-SST < Longitudinal-native. These models also provide means for assessing feature importance through the “gain” values. The gain plot is given in Figure 10. The information offered by the gain plot is consistent with what we know about selective vulnerability of brain regions to AD-related neurodegeneration [2]. For example, medial temporal regions such as the entorhinal and parahippocampal cortices are near the top offering the most diagnostic information, whereas regions such as paracentral cortices are near the bottom, offering very little diagnostic information.

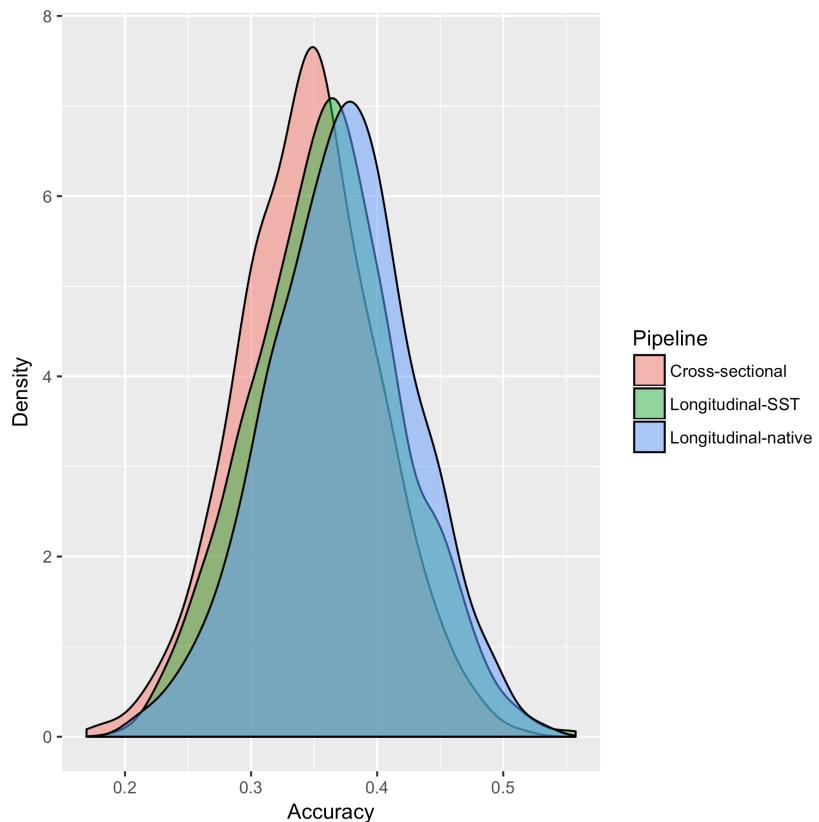


Figure 9: Density plot of the accuracy for each of the three pipeline choices. Accuracy was determined from the confusion matrices that were calculated for each of the 1000 iterations. Using Tukey multiple comparisons of means at a 95% family-wise confidence level, the adjusted p-values were: Longitudinal-SST – Cross-sectional $< 1e - 4$, Longitudinal-native – Cross-sectional $< 1e - 6$, Longitudinal-native – Longitudinal-SST $< 1e - 6$.

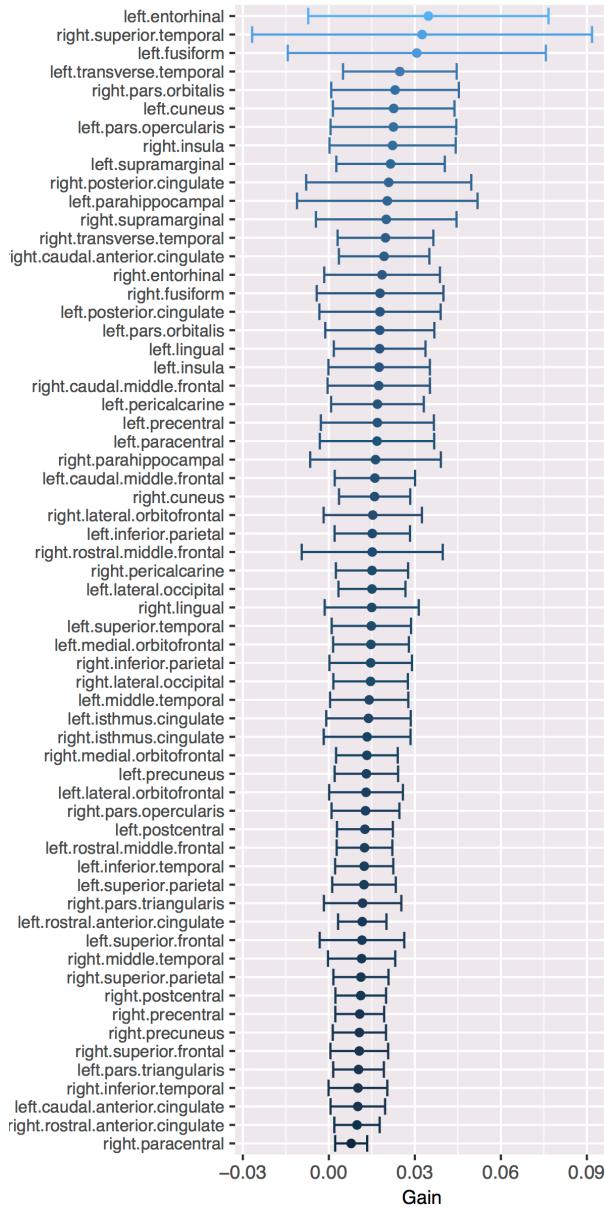


Figure 10: The “gain” measurements of the extreme gradient boosting models for the Longitudinal-native processing.

Discussion

Herein we detailed the ANTs longitudinal cortical thickness framework which was designed to take advantage of longitudinal data acquisition protocols while accounting for the various bias issues that have been associated with processing such data. Over 600 subjects from the well-known longitudinal ADNI-1 data set with diagnoses distributed between cognitively normal, MCI, LMCI, and AD were processed through the original ANTs cross-sectional framework [28] and two longitudinal variants. One of the variants, Longitudinal-SST, is similar to the FreeSurfer longitudinal stream in that each time-point image is reoriented to an unbiased single-subject template for subsequent processing. Longitudinal-native, in contrast, estimates cortical thickness in the native space while also using tissue prior probabilities generated from the SST.

Comparative assessment utilized LME models to determine the between-subject to within-subject variance ratios over the 62 regions of the brain defined by the DKT parcellation scheme where higher values indicate greater scientific interpretability. In these terms, Longitudinal-native was clearly the best performer. Longitudinal-SST outperformed Cross-sectional over the majority of regions. Regional disparities point to increases in both between-subject and within-subject variances which might be an additional interpolation artifact. In other words, interpolation potentially has a systematic but regionally varying effect. Investigation of this issue is the subject of future research.

Additional assessments included similar variance quantification in the EC and diagnostic prediction using extreme gradient boosting models. The former evaluation was motivated by the prominence of the EC as a biomarker in AD progression whereas the latter coupled a simplistic assumption of AD progression with modern machine learning techniques similar in spirit to what we did in our previous work [28]. Both assess-

ments supported the findings of the first assessment in demonstrating the superiority of Longitudinal-native. These findings promote longitudinal analysis considerations and motivates such techniques over cross-sectional methods for longitudinal data despite the increase in computational costs.

The longitudinal thickness framework is available in script form within the ANTs software library along with the requisite processing components. All generated data used for input, such as the ADNI template and tissue priors, are available upon request. As previously mentioned, we also make available the csv files containing the regional thickness values for all three pipelines.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Du, A.-T., Schuff, N., Kramer, J. H., Rosen, H. J., Gorno-Tempini, M. L., Rankin, K., Miller, B. L., and Weiner, M. W. “**Different Regional Patterns of Cortical Thinning in Alzheimer’s Disease and Frontotemporal Dementia**” *Brain* 130, no. Pt 4 (2007): 1159–66. doi:10.1093/brain/awm016
2. Dickerson, B. C., Bakkour, A., Salat, D. H., Feczkó, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., Sperling, R. A., Atri, A., Growdon, J. H., Hyman, B. T., Morris, J. C., Fischl, B., and Buckner, R. L. “**The Cortical Signature of Alzheimer’s Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and Is Detectable in Asymptomatic Amyloid-Positive Individuals**” *Cereb Cortex* 19, no. 3 (2009): 497–510. doi:10.1093/cercor/bhn113
3. MacDonald, D., Kabani, N., Avis, D., and Evans, A. C. “**Automated 3-D Extraction of Inner and Outer Surfaces of Cerebral Cortex from MRI**” *Neuroimage* 12, no. 3 (2000): 340–56. doi:10.1006/nimg.1999.0534
4. Magnotta, V. A., Andreasen, N. C., Schultz, S. K., Harris, G., Cizadlo, T., Heckel, D., Nopoulos, P., and Flaum, M. “**Quantitative in Vivo Measurement of Gyration in the Human Brain: Changes Associated with Aging**” *Cereb Cortex* 9, no. 2 (1999): 151–60.
5. Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab’bagh, Y., MacDonald, D., Lee, J. M., Kim, S. I., and Evans, A. C. “**Automated 3-D Extraction and Evaluation of the Inner and Outer Cortical Surfaces Using a Laplacian Map and Partial Volume Effect Classification**” *Neuroimage* 27, no. 1 (2005): 210–21. doi:10.1016/j.neuroimage.2005.03.036
6. Zeng, X., Staib, L. H., Schultz, R. T., and Duncan, J. S. “**Segmentation**

and Measurement of the Cortex from 3-D MR Images Using Coupled-Surfaces Propagation” *IEEE Trans Med Imaging* 18, no. 10 (1999): 927–37. doi:10.1109/42.811276

7. Jones, S. E., Buchbinder, B. R., and Aharon, I. “**Three-Dimensional Mapping of Cortical Thickness Using Laplace’s Equation**” *Hum Brain Mapp* 11, no. 1 (2000): 12–32.
8. Das, S. R., Avants, B. B., Grossman, M., and Gee, J. C. “**Registration Based Cortical Thickness Measurement**” *Neuroimage* 45, no. 3 (2009): 867–79. doi:10.1016/j.neuroimage.2008.12.016
9. Vachet, C., Hazlett, H. C., Niethammer, M., Oguz, I., Cates, J., Whitaker, R., Piven, J., and Styner, M. “**Group-Wise Automatic Mesh-Based Analysis of Cortical Thickness**” *SPIE medical imaging: Image processing* (2011):
10. Dale, A. M., Fischl, B., and Sereno, M. I. “**Cortical Surface-Based Analysis. I. Segmentation and Surface Reconstruction**” *Neuroimage* 9, no. 2 (1999): 179–94. doi:10.1006/nimg.1998.0395
11. Fischl, B., Sereno, M. I., and Dale, A. M. “**Cortical Surface-Based Analysis. II: Inflation, Flattening, and a Surface-Based Coordinate System**” *Neuroimage* 9, no. 2 (1999): 195–207. doi:10.1006/nimg.1998.0396
12. Fischl, B. and Dale, A. M. “**Measuring the Thickness of the Human Cerebral Cortex from Magnetic Resonance Images**” *Proc Natl Acad Sci U S A* 97, no. 20 (2000): 11050–5. doi:10.1073/pnas.200033797
13. Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Kouwe, A. van der, Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. M. “**Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain**” *Neuron* 33, no. 3 (2002): 341–

55.

14. Fischl, B., Kouwe, A. van der, Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. **“Automatically Parcellating the Human Cerebral Cortex”** *Cereb Cortex* 14, no. 1 (2004): 11–22.
15. Kraemer, H. C., Yesavage, J. A., Taylor, J. L., and Kupfer, D. **“How Can We Learn About Developmental Processes from Cross-Sectional Studies, or Can We?”** *Am J Psychiatry* 157, no. 2 (2000): 163–71. doi:10.1176/appi.ajp.157.2.163
16. Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., Trojanowski, J. Q., and, Alzheimer’s Disease Neuroimaging Initiative. **“The Alzheimer’s Disease Neuroimaging Initiative: A Review of Papers Published Since Its Inception.”** *Alzheimers Dement* 8, no. 1 Suppl (2012): S1–68.
17. Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. **“Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis”** *Neuroimage* 61, no. 4 (2012): 1402–18. doi:10.1016/j.neuroimage.2012.02.084
18. Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., and De Stefano, N. **“Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis”** *Neuroimage* 17, no. 1 (2002): 479–89.
19. Yushkevich, P. A., Avants, B. B., Das, S. R., Pluta, J., Altinay, M., Craige, C., and Alzheimer’s Disease Neuroimaging Initiative. **“Bias in Estimation of Hippocampal Atrophy Using Deformation-Based Morphometry Arises from Asymmetric Global Normalization: An Illustration in ADNI 3 T MRI Data”** *Neu-*

roimage 50, no. 2 (2010): 434–45. doi:10.1016/j.neuroimage.2009.12.007

20. Thompson, W. K., Holland, D., and Alzheimer's Disease Neuroimaging Initiative. “**Bias in Tensor Based Morphometry Stat-ROI Measures May Result in Unrealistic Power Estimates**” *Neuroimage* 57, no. 1 (2011): 1–4. doi:10.1016/j.neuroimage.2010.11.092
21. Fox, N. C., Ridgway, G. R., and Schott, J. M. “**Algorithms, Atrophy and Alzheimer's Disease: Cautionary Tales for Clinical Trials**” *Neuroimage* 57, no. 1 (2011): 15–8. doi:10.1016/j.neuroimage.2011.01.077
22. Hua, X., Hibar, D. P., Ching, C. R. K., Boyle, C. P., Rajagopalan, P., Gutman, B. A., Leow, A. D., Toga, A. W., Jack, C. R., Jr, Harvey, D., Weiner, M. W., Thompson, P. M., and Alzheimer's Disease Neuroimaging Initiative. “**Unbiased Tensor-Based Morphometry: Improved Robustness and Sample Size Estimates for Alzheimer's Disease Clinical Trials**” *Neuroimage* 66, (2013): 648–61. doi:10.1016/j.neuroimage.2012.10.086
23. Reuter, M. and Fischl, B. “**Avoiding Asymmetry-Induced Bias in Longitudinal Image Processing**” *Neuroimage* 57, no. 1 (2011): 19–21. doi:10.1016/j.neuroimage.2011.02.076
24. Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., Sabuncu, M. R., and Alzheimer's Disease Neuroimaging Initiative. “**Statistical Analysis of Longitudinal Neuroimage Data with Linear Mixed Effects Models**” *Neuroimage* 66, (2013): 249–60. doi:10.1016/j.neuroimage.2012.10.065
25. Wierenga, L. M., Langen, M., Oranje, B., and Durston, S. “**Unique Developmental Trajectories of Cortical Thickness and Surface Area**” *Neuroimage* 87, (2014): 120–6. doi:10.1016/j.neuroimage.2013.11.010
26. Landin-Romero, R., Kumfor, F., Leyton, C. E., Irish, M., Hodges, J. R.,

and Piguet, O. “**Disease-Specific Patterns of Cortical and Subcortical Degeneration in a Longitudinal Study of Alzheimer’s Disease and Behavioural-Variant Frontotemporal Dementia**” *Neuroimage* (2016): doi:10.1016/j.neuroimage.2016.03.032

27. Nourbakhsh, B., Azevedo, C., Nunan-Saah, J., Maghzi, A.-H., Spain, R., Pelletier, D., and Waubant, E. “**Longitudinal Associations Between Brain Structural Changes and Fatigue in Early MS**” *Mult Scler Relat Disord* 5, (2016): 29–33. doi:10.1016/j.msard.2015.10.006

28. Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., Strien, N. van, Stone, J. R., Gee, J. C., and Avants, B. B. “**Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements**” *Neuroimage* 99, (2014): 166–79. doi:10.1016/j.neuroimage.2014.05.044

29. Available at <https://github.com/stnavar/ANTs>

30. Fujishima, M., Maikusa, N., Nakamura, K., Nakatsuka, M., Matsuda, H., and Meguro, K. “**Mild Cognitive Impairment, Poor Episodic Memory, and Late-Life Depression Are Associated with Cerebral Cortical Thinning and Increased White Matter Hyperintensities**” *Front Aging Neurosci* 6, (2014): 306. doi:10.3389/fnagi.2014.00306

31. Das, S. R., Mancuso, L., Olson, I. R., Arnold, S. E., and Wolk, D. A. “**Short-Term Memory Depends on Dissociable Medial Temporal Lobe Regions in Amnestic Mild Cognitive Impairment**” *Cereb Cortex* 26, no. 5 (2016): 2006–17. doi:10.1093/cercor/bhv022

32. Olm, C. A., Kandel, B. M., Avants, B. B., Detre, J. A., Gee, J. C., Grossman, M., and McMillan, C. T. “**Arterial Spin Labeling Perfusion Predicts Longitudinal Decline in Semantic Variant Primary Progressive Aphasia**” *J Neurol* 263, no.

10 (2016): 1927–38. doi:10.1007/s00415-016-8221-1

33. Mugler, J. P., 3rd and Brookeman, J. R. “**Three-Dimensional Magnetization-Prepared Rapid Gradient-Echo Imaging (3D MP RAGE)**” *Magn Reson Med* 15, no. 1 (1990): 152–7.

34. Jack, C. R., Jr, Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., and Weiner, M. W. “**The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI Methods**” *J Magn Reson Imaging* 27, no. 4 (2008): 685–91. doi:10.1002/jmri.21049

35. Available at <https://github.com/ntustison/CrossLong>

36. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. “**N4ITK: Improved N3 Bias Correction**” *IEEE Trans Med Imaging* 29, no. 6 (2010): 1310–20. doi:10.1109/TMI.2010.2046908

37. Avants, B. B., Klein, A., Tustison, N. J., Woo, J., and Gee, J. C. “**Evaluation of an Open-Access, Automated Brain Extraction Method on Multi-Site Multi-Disorder Data**” (2010):

38. Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., and Gee, J. C. “**An Open Source Multivariate Framework for *n*-Tissue Segmentation with Evaluation on Public Data**” *Neuroinformatics* 9, no. 4 (2011): 381–400. doi:10.1007/s12021-011-9109-y

39. Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A. “**Multi-Atlas Segmentation with Joint Label Fusion**” *IEEE Trans Pattern Anal*

Mach Intell 35, no. 3 (2013): 611–23. doi:10.1109/TPAMI.2012.143

40. Klein, A. and Tourville, J. “**101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol**” *Front Neurosci* 6, (2012): 171. doi:10.3389/fnins.2012.00171
41. Tustison, N. J. and Herrera, J. M. “**Two Luis Miguel Fans Walk into a Bar in Nagoya → (Yada, Yada, Yada) → an ITK-Implementation of a Popular Patch-Based Denoising Filter**” *Insight Journal* (2016):
42. Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., and Robles, M. “**Adaptive Non-Local Means Denoising of MR Images with Spatially Varying Noise Levels**” *J Magn Reson Imaging* 31, no. 1 (2010): 192–203. doi:10.1002/jmri.22003
43. Available at <http://brain-development.org/ixi-dataset/>
44. Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A. D., Bogovic, J. A., Hua, J., Chen, M., Jarso, S., Smith, S. A., Joel, S., Mori, S., Pekar, J. J., Barker, P. B., Prince, J. L., and Zijl, P. C. M. van. “**Multi-Parametric Neuroimaging Reproducibility: A 3-T Resource Study**” *Neuroimage* 54, no. 4 (2011): 2854–66. doi:10.1016/j.neuroimage.2010.11.047
45. Available at http://fcon_1000.projects.nitrc.org/indi/pro/nki.html
46. Available at <http://www.oasis-brains.org>
47. Breiman, L. “**Random Forests**” *Machine learning* (2001): 5–32.
48. Lehmann, G. and Legland, D. “**Efficient N-Dimensional Surface Estimation Using Crofton Formula and Run-Length Encoding**” *Insight Journal*

(2012):

49. Available at <https://github.com/ntustison/KapowskiChronicles>
50. Hasan, K. M., Mwangi, B., Cao, B., Keser, Z., Tustison, N. J., Kochunov, P., Frye, R. E., Savatic, M., and Soares, J. “**Entorhinal Cortex Thickness Across the Human Lifespan**” *J Neuroimaging* 26, no. 3 (2016): 278–82. doi:10.1111/jon.12297
51. Price, A. R., Bonner, M. F., Peelle, J. E., and Grossman, M. “**Converging Evidence for the Neuroanatomic Basis of Combinatorial Semantics in the Angular Gyrus**” *J Neurosci* 35, no. 7 (2015): 3276–84. doi:10.1523/JNEUROSCI.3446-14.2015
52. Wisse, L. E. M., Butala, N., Das, S. R., Davatzikos, C., Dickerson, B. C., Vaishnavi, S. N., Yushkevich, P. A., Wolk, D. A., and Alzheimer’s Disease Neuroimaging Initiative. “**Suspected Non-AD Pathology in Mild Cognitive Impairment**” *Neurobiol Aging* 36, no. 12 (2015): 3152–62. doi:10.1016/j.neurobiolaging.2015.08.029
53. Betancourt, L. M., Avants, B., Farah, M. J., Brodsky, N. L., Wu, J., Ashtari, M., and Hurt, H. “**Effect of Socioeconomic Status (SES) Disparity on Neural Development in Female African-American Infants at Age 1 Month**” *Dev Sci* (2015): doi:10.1111/desc.12344
54. Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., and Gee, J. C. “**The Optimal Template Effect in Hippocampus Studies of Diseased Populations**” *Neuroimage* 49, no. 3 (2010): 2457–66. doi:10.1016/j.neuroimage.2009.09.062
55. Verbeke, G. and Molenberghs, G. “**Linear Mixed Models for Longitudinal Data**” (2009):
56. Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. “**Applied Longitudinal Analysis**” (2011):

ysis” 998, (2012):

57. Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. “**Within-Subject Template Estimation for Unbiased Longitudinal Image Analysis**” *Neuroimage* 61, no. 4 (2012): 1402–18. doi:10.1016/j.neuroimage.2012.02.084
58. Gelman, A. and others. “**Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)**” *Bayesian analysis* 1, no. 3 (2006): 515–534.
59. Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. “**Stan: A Probabilistic Programming Language**” *J Stat Softw* (2016):
60. Hyman, B. T., Van Hoesen, G. W., Damasio, A. R., and Barnes, C. L. “**Alzheimer’s Disease: Cell-Specific Pathology Isolates the Hippocampal Formation**” *Science* 225, no. 4667 (1984): 1168–70.
61. Braak, H. and Braak, E. “**Neuropathological Stageing of Alzheimer-Related Changes**” *Acta Neuropathol* 82, no. 4 (1991): 239–59.
62. Van Hoesen, G. W., Hyman, B. T., and Damasio, A. R. “**Entorhinal Cortex Pathology in Alzheimer’s Disease**” *Hippocampus* 1, no. 1 (1991): 1–8. doi:10.1002/hipo.450010102
63. Yassa, M. A. “**Ground Zero in Alzheimer’s Disease**” *Nat Neurosci* 17, no. 2 (2014): 146–7. doi:10.1038/nn.3631
64. Holland, D., Brewer, J. B., Hagler, D. J., Fennema-Notestine, C., Fenema-Notestine, C., Dale, A. M., and Alzheimer’s Disease Neuroimaging Initiative. “**Subregional Neuroanatomical Change as a Biomarker for Alzheimer’s**

Disease” *Proc Natl Acad Sci U S A* 106, no. 49 (2009): 20954–9.

65. Desikan, R. S., Sabuncu, M. R., Schmansky, N. J., Reuter, M., Cabral, H. J., Hess, C. P., Weiner, M. W., Biffi, A., Anderson, C. D., Rosand, J., Salat, D. H., Kemper, T. L., Dale, A. M., Sperling, R. A., Fischl, B., and Alzheimer’s Disease Neuroimaging Initiative. “**Selective Disruption of the Cerebral Neocortex in Alzheimer’s Disease**” *PLoS One* 5, no. 9 (2010): e12853. doi:10.1371/journal.pone.0012853
66. Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack, C. R., Jr, Feldman, H. H., Bokde, A. L. W., Alexander, G. E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., and North American Alzheimer’s Disease Neuroimaging Initiative (ADNI). “**Prediction of Conversion from Mild Cognitive Impairment to Alzheimer’s Disease Dementia Based Upon Biomarkers and Neuropsychological Test Performance**” *Neurobiol Aging* 33, no. 7 (2012): 1203–14. doi:10.1016/j.neurobiolaging.2010.10.019
67. Available at <http://xgboost.readthedocs.io/>