

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) consisting of 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 training/prediction 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 (e.g., Alzheimer’s disease and frontotemporal dementia [1, 2], Parkinson’s disease [3], Williams syndrome [4], multiple sclerosis [5], autism [6], migraines [8], chronic smoking [9], alcoholism [10], cocaine addiction [11], schizophrenia [12], bipolar disorder [13], autism [6], marijuana use in adolescents [14], Tourette syndrome in children [15], scoliosis in female adolescents [16], heart failure [17], early-onset blindness [18], chronic pancreatitis [19], obsessive-compulsive disorder [20], ADHD [21], obesity [22], heritable [23] and elderly [24] depression, age [25], gender [26], untreated male-to-female transsexuality [27], handedness [28], intelligence [29], athletic ability [30], meditative practices [31], musical ability [32, 33], musical instrument playing [34], tendency toward criminality [35], childhood sexual abuse in adult females [36], and Tetris-playing ability in female adolescents [37]). 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., [38–40]) and volumetric techniques (e.g., [41–44]). Of noted significance, and representing the former, is the well-known and highly utilized FreeSurfer software package [45–49].

In inferring developmental processes, many of these studies employ cross-sectional population sampling strategies despite the potential for confounding effects [50]. Large-scale studies involving longitudinal image acquisition of a targeted subject population, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [51], 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 [52]. For example, FSL’s SIENA (Structural Image Evaluation, using Normalization, of Atrophy) framework [53] for detecting atrophy between two time points 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 [54] in the context of hippocampal volumetric change where it was shown that interpolation-induced artifacts can artificially inflate effect size [55]. These insights have since been used for making specific recommendations with respect to longitudinal image data processing [52, 56, 57].

In a series of papers [52, 58] the authors motivated the design and implementation of the longitudinal FreeSurfer variant partly inspired by these earlier insights and encapsulated by 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 [59] and has been used in a variety of studies including pediatric cortical development [60], differential development in Alzheimer’s disease and fronto-temporal dementia [61], and fatigue in the context of multiple sclerosis [62].

In [63], 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.¹ This proposed ANTs-based pipeline has since been directed at a variety of neuroimaging research topics including mild cognitive impairment and depression [64], short term memory in mild cognitive impairment [65], and aphasia [66]. 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 use a machine learning-based training/prediction paradigm which demonstrates that the recommended longitudinal processing approach leads to improved predictive diagnostic accuracy over the alternative strategies.

¹<https://github.com/stnava/ANTs>

Methods and materials

ADNI-1 imaging data

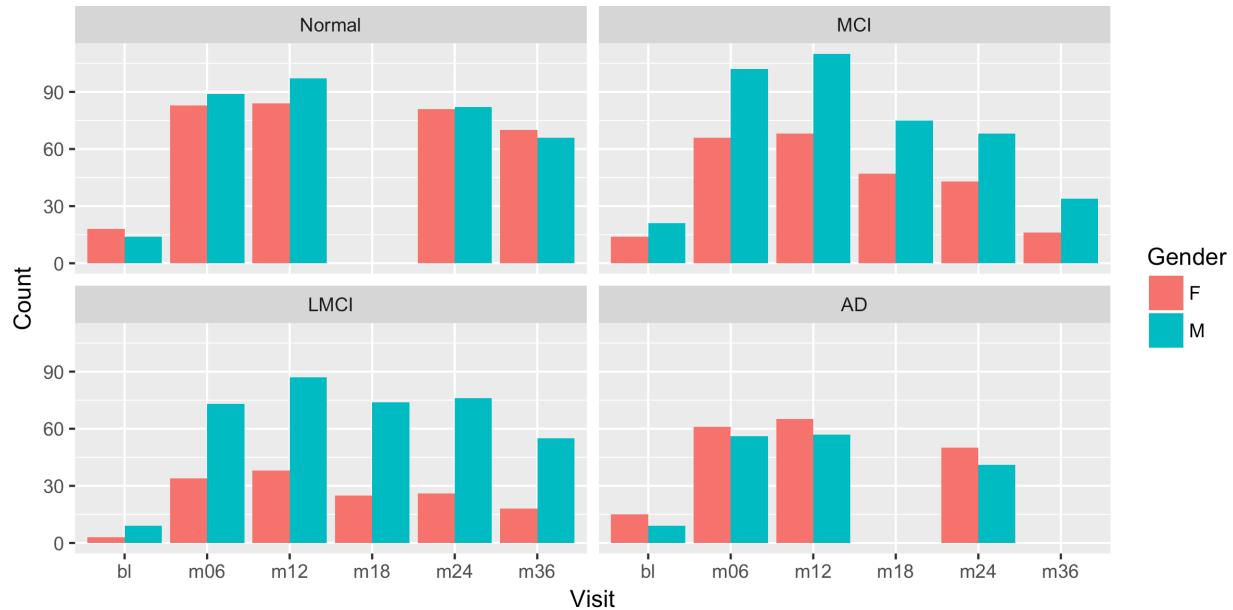


Figure 1: Demographic breakdown of the number of ADNI-1 subjects by diagnosis i.e., normal, mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), and Alzheimer’s disease (AD). Within each panel we plot the number of subjects (by gender) per visit—baseline (“bl”) and n months (“mn”).

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 [67] 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 [68]. 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 [63] (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 available² 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. A further breakdown of demographic information is given in Figure X.

²<https://github.com/ntustison/CrossLong>

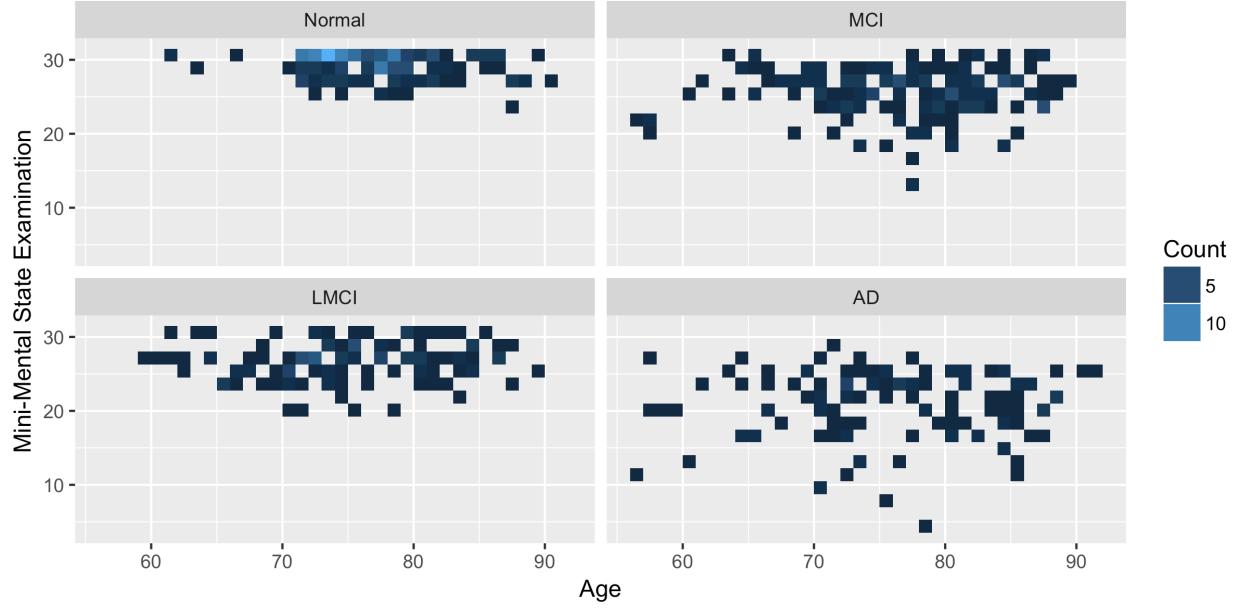


Figure 2: Age vs. Mini-mental examination (MMSE) scores for the ADNI-1 subjects by diagnosis.

Similarly, in Figure X, we show the 2-D distribution of Age vs. mini-mental examination (MMSE) scores taken at the month 12 visit across diagnoses for the subjects analyzed.

ANTs cortical thickness

Cross-sectional processing

A thorough discussion of the ANTs cross-sectional thickness estimation framework was previously discussed in [63]. As a brief review, given a T1-weighted brain MR image, processing comprises the following major steps (cf Figure 1 of [63]):

1. N4 bias correction [69],
2. brain extraction [70],
3. Atropos n -tissue segmentation [71], and
4. cortical thickness estimation [43].

ROI-based quantification is achieved through the use of the joint label fusion approach of [72] and the use of the MindBoggle-101 data labeled using the Desikan–Killiany–Tourville (DKT) protocol [73] consisting of 31 labels per hemisphere (cf Table 1). This pipeline has since been enhanced by the implementation [74] of a patch-based denoising algorithm [75] as an optional preprocessing step and multi-modal integration capabilities (e.g., joint T1- and T2-weighted processing).

Table 1: The 31 cortical labels (per hemisphere) of the Desikan-Killiany-Tourville atlas. The ROI abbreviations from the R `brainGraph` package are given 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)	

For evaluation, 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 [73] 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 [76], MMRR [77], NKI [78], and OASIS [79]) were used to predict age and gender using linear and random forest [80] models. The resulting regional statistics (including cortical thickness, surface area [81], volumes, and Jacobian determinant values) were made available online.³ These include the corresponding FreeSurfer measurements which are also publicly available for research inquiries (e.g., [82]). Since publication, this framework has been used in a number of cross-sectional studies (e.g., [83–85]).

Unbiased longitudinal processing

³<https://github.com/ntustison/KapowskiChronicles>

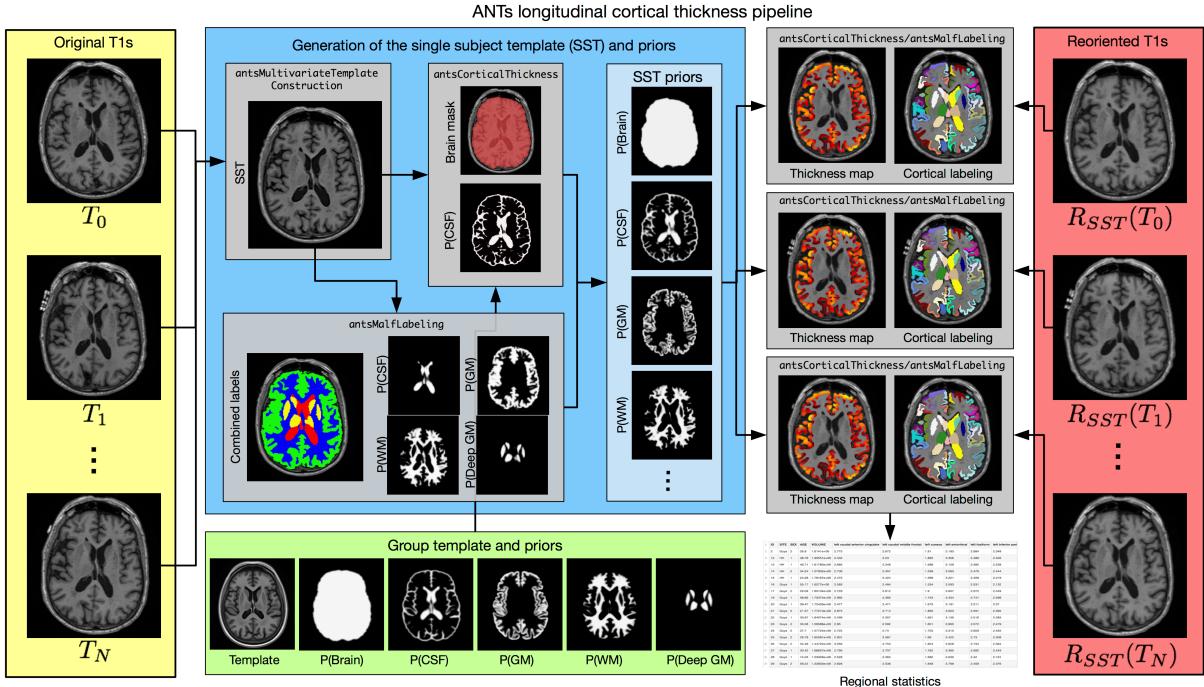


Figure 3: Diagrammatic illustration of the ANTs longitudinal cortical thickness pipeline for a single subject with N time points. From the N original T1-weighted images (left column, yellow panel) and the group template and priors (bottom row, green panel), the single-single subject template (SST) and 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 (denoted as “Longitudinal 2” in the text). Optionally, one can rigidly transform the time-point images prior to segmentation and cortical thickness estimation (right column, red panel) which we refer to “Longitudinal 1”. For regional thickness values, regional labels can be propagated to each image using a given atlas set and cortical parcellation scheme.

Given certain practical limitations (e.g., subject recruitment and retainment), as mentioned 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 [52, 54, 58] and, to our knowledge, regionally varying interpolation effects not previously made explicit.

Given N time-point T1-weighted MR images, a group template, and group template prior probability maps (described below), the longitudinal pipeline consists of the following steps:

1. (Offline): Creation of the group template.
2. Creation of the 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 X 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 the population data [86]. 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 [63]. A brief overview of this process is also provided in the next section describing the single-subject template. Canonical views of the ADNI-1 template and corresponding auxiliary images are given in Figure X.

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 [86] 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

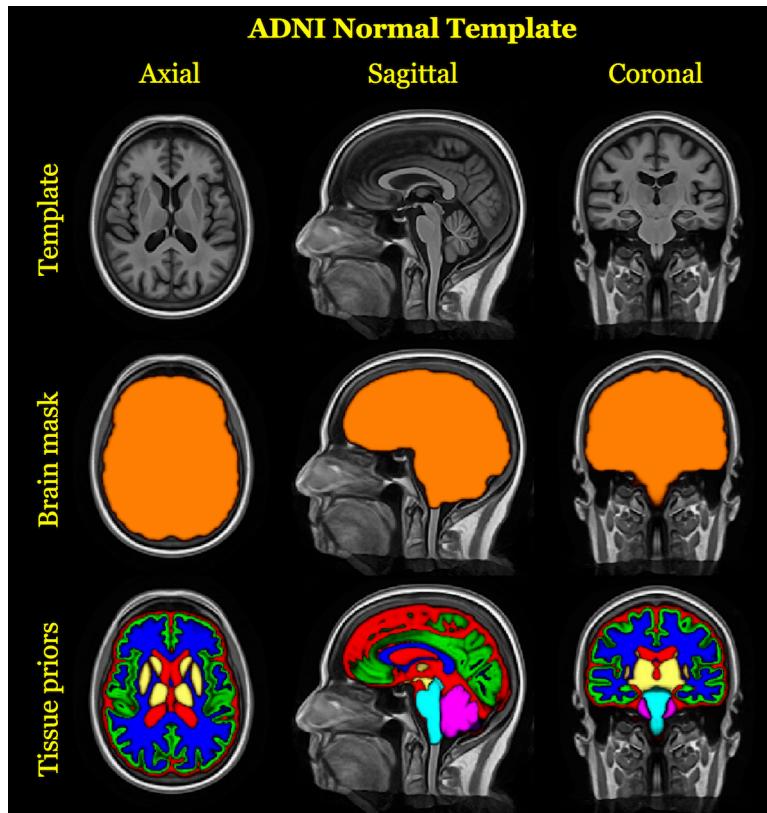


Figure 4: Top row: Canonical views of the template created from 52 cognitively normal subjects of the ADNI-1 database. The prior probability mask for the whole brain (middle row) and the six tissue priors (bottom row) are used to “seed” each single-subject template for creation of a probabilistic brain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

have been expertly annotated [73], a multi-atlas joint label fusion step (JLF) [72] is performed to create individualized probability maps for all tissue types. The five 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 of the original time series images.

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 [52].

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 [86]. 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 [63]. 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 [71]. Regarding the 4-D brain segmentation, the possible benefit is potentially outweighed by the possibility of “over-regularization” [52] whereby smoothing across time reduces detection ability of large time point changes. Additionally, it is less than straightforward to accommodate irregular temporal sampling such as the acquisition schedule of the ADNI-1 protocol.

However, during the course of this work we discovered that reorienting each time point image to the SST has significant detrimental measurement effects in which interpolation bias induces artificial anatomical changes and these changes correlate significantly with specific regions. Since we are measuring the thickness of the highly convoluted cortex involving measurements on the order of $2 - 8\text{mm}$ from images with $\sim 1\text{mm}^3$ voxels, these artificial changes significantly effect clinically related criteria of measurement quality such as confidence intervals and predictability. This is discussed further in the following sections.

Joint label fusion and pseudo-geodesic for large cohort labeling. Cortical thickness ROI-based analyses are performed using joint label fusion [72] 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

As described above, given a set of longitudinal data, there are possible options for cortical thickness processing. These are

- **Cross-sectional.** Process each subject’s time point independently using the cross-sectional pipeline originally described in [63].
- **Longitudinal 1.** Rigidly transform each subject to the SST and then segment and estimate cortical thickness in the space of the SST.
- **Longitudinal 2.** Segment and estimate cortical thickness in the native space.

In this section we describe how we compare each pipeline for processing longitudinal data.

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 () 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, [59] 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 [87] and [88]. 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 [89] only hint at in their exposition of the longitudinal FreeSurfer stream.

As previously noted we observed yearly cortical thickness measurements from sixty-two 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

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

Specification of parameters in the above prior distributions reflect commonly accepted diffuse priors [REFERENCE]. 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 [87]. The posterior distribution of r^k was summarized via the posterior median where the posterior distributions were obtained using the Stan probabilistic programming language [90].

Entorhinal cortical thickness variability

As a further assessment of utility as a biomarker, we used LME models and cortical thickness measurements of the entorhinal cortex to demonstrate how these variability criteria relate to potential scientific analyses. First, we used model () 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 variability 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 [63] incorporated a statistical modeling approach for predicting basic subject demographics (i.e., age and gender) from the summary DKT regional cortical thickness values. We use a similar evaluation strategy in this work with the ADNI-1 data. We build 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 use 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 which circumvents missing data issues. Thus, for each subject, we have a single set

of 62 slope coefficients, $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_{62}\}$, and a single diagnosis of cognitively normal, MCI, LMCI, or AD (cf Figure X). Note that the diagnosis for each subject did not change over the course of the image acquisition schedule. 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 [REFERENCE].

For model construction, we used extreme gradient boosting, which is a well-performing, out-of-the-box classifier implemented in the XGBoost⁴ package for the R project. Although there are many options available for classification, we chose this particular technique due to its successful deployment in several Kaggle data science challenges⁵ and its being the winner of the 2016 John M. Chambers Statistical Software Award.⁶ An additional advantage is that such techniques provide a feature importance quantity which can be reviewed for clinical plausibility.

⁴<http://xgboost.readthedocs.io/>

⁵<http://blog.kaggle.com/>

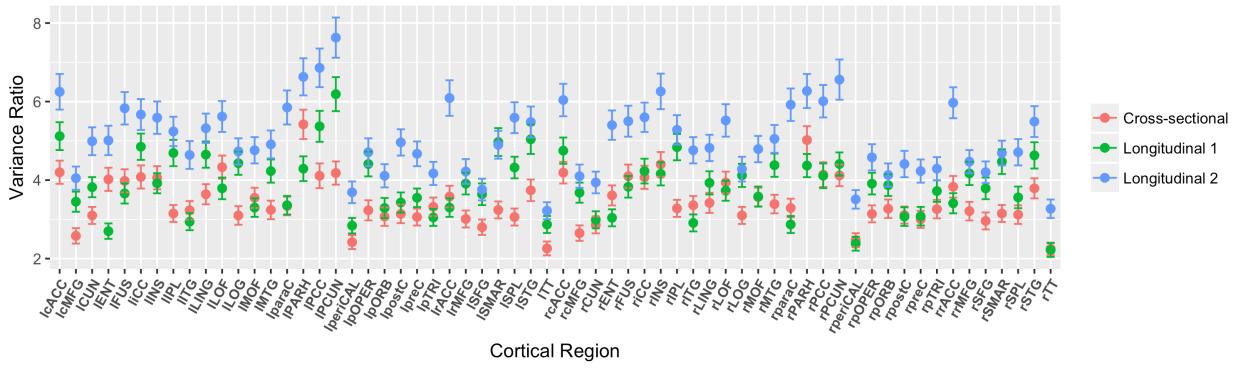
⁶<http://stat-computing.org/awards/jmc/winners.html>

Results

Regional within-subject and between-subject variance

This section presents processing method performance in 62 brain regions with respect to the variance ratio defined in (2) before demonstrating some ways in which different aspects of variability affect confidence in prediction and estimation.

Figure (5) provides resulting 95% credible intervals for the distributions of region specific variance ratios $r^k = \tau_k/\sigma_k$. Each processing method has its own color: blue, the color of the second longitudinal method, runs along the upper portions of the figure, rarely intersecting with green, the color of the first longitudinal method; green and orange, the color corresponding to the cross-sectional method, mix with each other nearer to the middle and bottom of the figure. The placement of the methods with respect to each other is meaningful. The higher the method, the larger the variance ratio corresponding to that method, and hence the greater discriminative capacity for the data corresponding to that processing method. Therefore, Figure (5) may be considered as evidence for method Longitudinal 2 [NEED TO COME UP WITH LABELS FOR THESE] providing higher quality data than that provided by the other methods. Looking past the strong performance of the second longitudinal method, the first longitudinal method and the cross-sectional method perform similarly to one another: Longitudinal 1 performs better in some regions, and the cross-sectional method performs better in others.



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 (0). 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 corresponding to worst performance are colored red.

Table 2: Processing methods and EC statistical results

Method	$\frac{\tau}{\sigma}$	CI width ¹	σ^2	Variance ²	$\sigma^2 + \tau^2$	CI width ³
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
Longitudinal 1						
<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
Longitudinal 2						
<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

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 (0). 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, the second longitudinal method 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 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 second Longitudinal

method performs best, and the first performs worst.

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 the second longitudinal method is 0.81, whereas the distances for the first longitudinal and cross-sectional methods are 0.75 and 0.70, respectively. Figure (6) displays the normalized results.

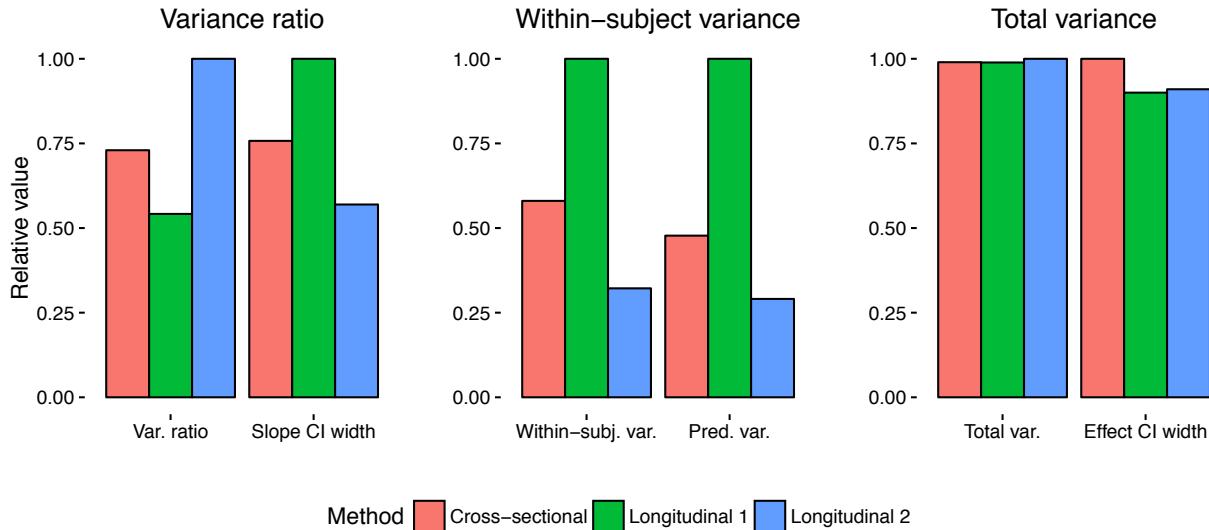


Figure 6: Aspects of model variance (see Section ??) are compared with credible interval sizes and variance in predictions. Values are normalized by the largest quantity, and processing methods are distinguished by color and ordering. On the left, the variance ratio $r = \tau/\sigma$ is compared to the width of credible interval for the slope term of model (0). 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.

(SUBSECTION TO REMOVE) Comparison of ANTs cortical thickness pipelines

As described previously, we employ a number of criteria to evaluate the performance of the ANTs cortical thickness longitudinal pipeline with the original cross-sectional workflow described previously [63]. Further, we look at two variants of the longitudinal pipeline: the first where each time point image is reoriented to

the SST prior to segmentation and cortical thickness estimation (denoted as “longitudinal 1”) and the second where each time point is processed in native space (denoted as “longitudinal 2”).

Within-subject/between-subject variability

Maximizing the between-subject variability while simultaneously minimizing the within-subject variability minimizes the overlap between groups implying a more significant effect of the condition separating the groups. In Figure X, we provide a general overview of how each of these three measurements compare across the three different ANTs pipeline variants over the 62 DKT regions. This illustrates the superiority of the longitudinal 2 pipeline where processing occurs in native space. It is also interesting to note that longitudinal 1 is an improvement over the cross-sectional processing in that, overall, it reduces both types of variability.

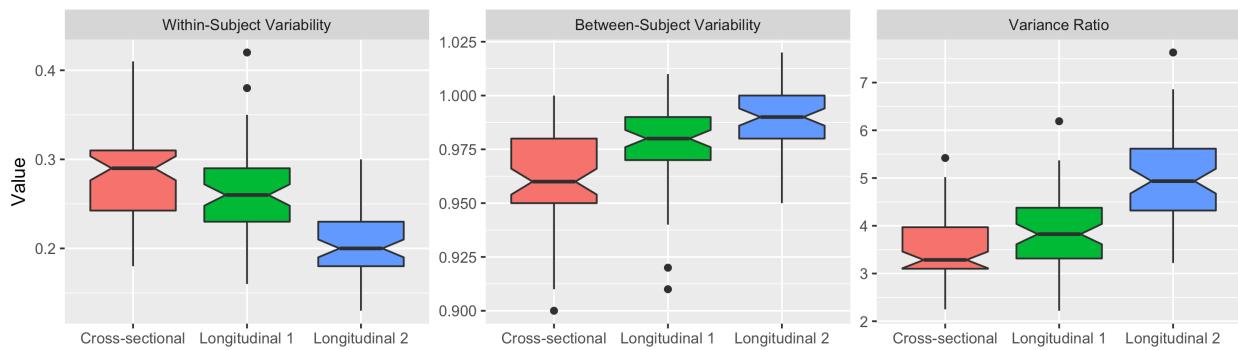


Figure 7: Notched box plots showing the distribution of the within-subject variability, between subject variability, and ratio of the between-subject variability and within-subject variability for each of the 62 DKT regions. Note that the “better” measurement maximizes this latter ratio.

Further insights can be gleaned by looking at the ratio measurements per region (see Figure X). On a per region basis, longitudinal 2 outperforms both the cross-sectional and longitudinal 1 pipelines. Although longitudinal 1 is superior for the majority of the DKT regions over cross-sectional processing, exceptions include 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 anterior 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.

Clinically-based assessment of measurement quality

A clinically-based prediction strategy was performed to evaluate the quality of the cortical thickness measurements produced by each pipeline. 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

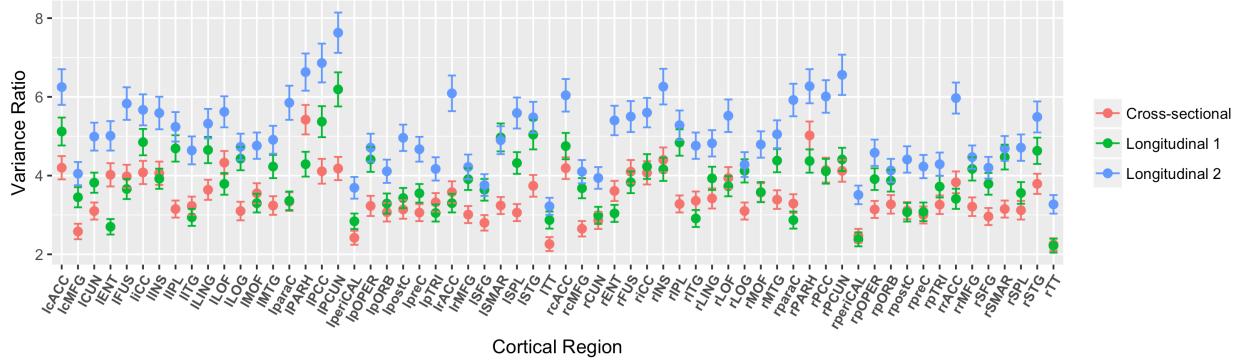


Figure 8: Per region quantities of the variance ratio (between-subject and within subject variability). These values are plotted spatially in Figure X.

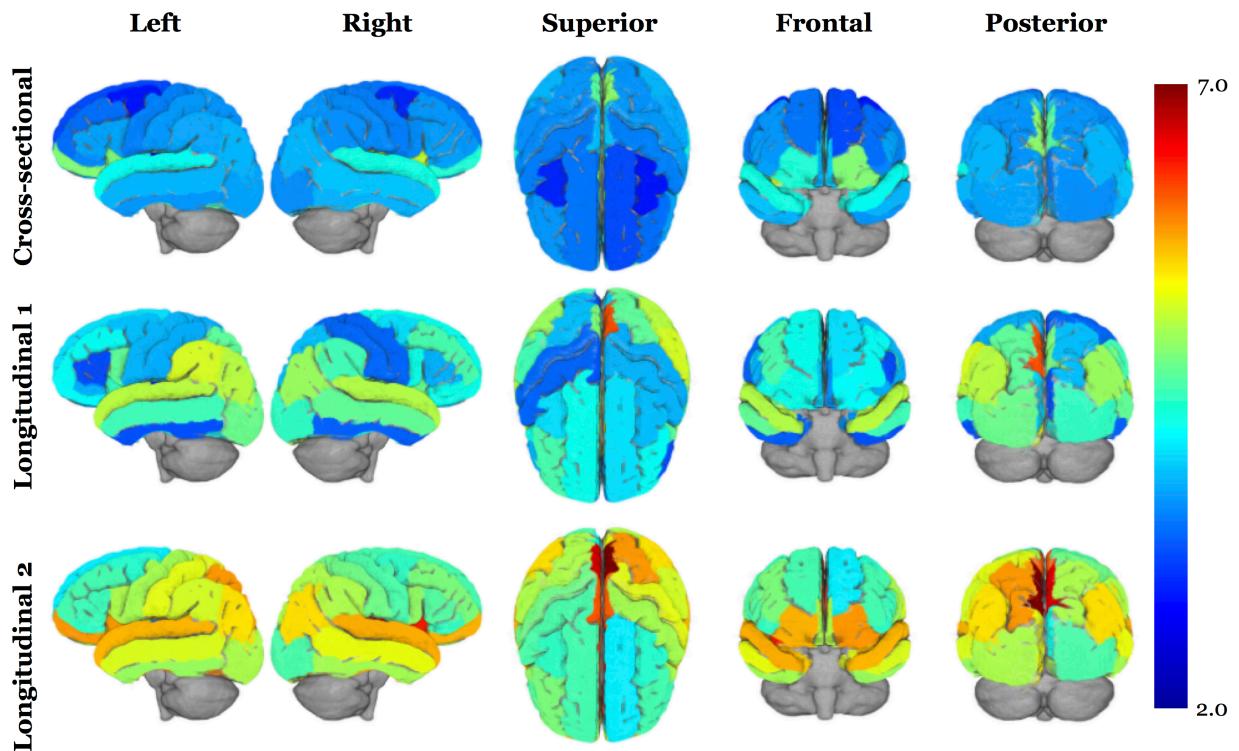


Figure 9: 3-D volumetric rendering of the variance ratio values plotted in Figure X.

in some regions versus others which should be reflected in the assigned diagnostic category.

For each of 500 permutations, an extreme gradient boosting model [91] was constructed from 90% of the cortical thickness slope data corresponding to each pipeline. Based on previous experience, parameters of these models were chosen from the default parameters which included a gradient step of 0.6 and 100 iterations (or trees). The resulting models were then used to predict on the remaining 10% of the data. The resulting accuracy distributions are plotted in Figure X and compared statistically using Tukey's range test which indicated increasing performance Cross-sectional < Longitudinal 1 < Longitudinal 2. These models also provide means for assessing feature "importance" through the "Gain" values which describes "the improvement in accuracy brought by a feature to the branches [of the tree or iteration] it is on."

Interpolation effects associated with Longitudinal 1 processing

The subset of regions where the longitudinal 1 < cross-sectional in terms of the variance ratio caused us to investigate this issue further. Considering the only difference between longitudinal 1 and longitudinal 2 is the reorientation to the SST, our exploration focused on the effects of interpolation. Specifically, we suspected that interpolation artificially changes the anatomy (i.e., volume and surface area) which results in additive noise to the thickness measurements and that this effect varies spatially. To test this, we used the 20 DKT atlases [92] which were sampled from the OASIS data set. An optimal mean/shape template [86] was created from this cohort to which each T1-weighted image was rigidly registered. Using the resulting rigid transform, the label map of 62 regions was warped to the space of the template (similar in spirit to the protocol characterizing longitudinal 1) using nearest neighbor interpolation (check on this). Calculation of the volumes (by summing up the volume of each voxel) and surface areas (using a well-performing digitally-based surface estimator [81]) for each of the 62 regions were calculated both before and after transformation to the template. We then computed the percent change for each of the two anatomical measures and correlated those with the per subject variability which is plotted in Figure X.

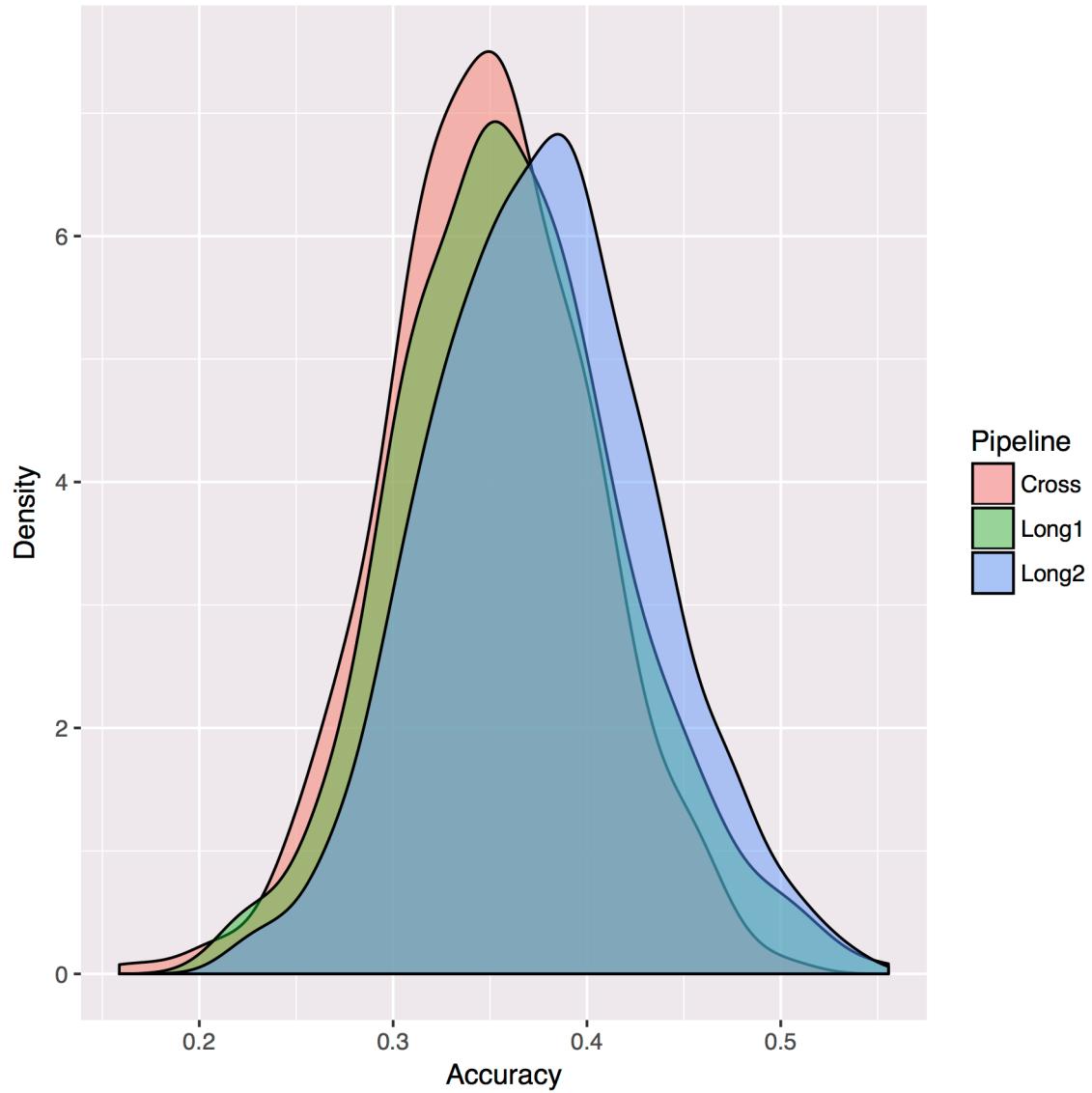


Figure 10: Density plot of the accuracy for each of the three pipeline choices. Accuracy was determined from the confusion matrices that were calculated for each of the 500 iterations. Using Tukey multiple comparisons of means at a 95% family-wise confidence level, the adjusted p-values were: Long1 – Cross = 0.067, Long2 – Cross < 1e-6, Long2 – Long1 = 0.00019.

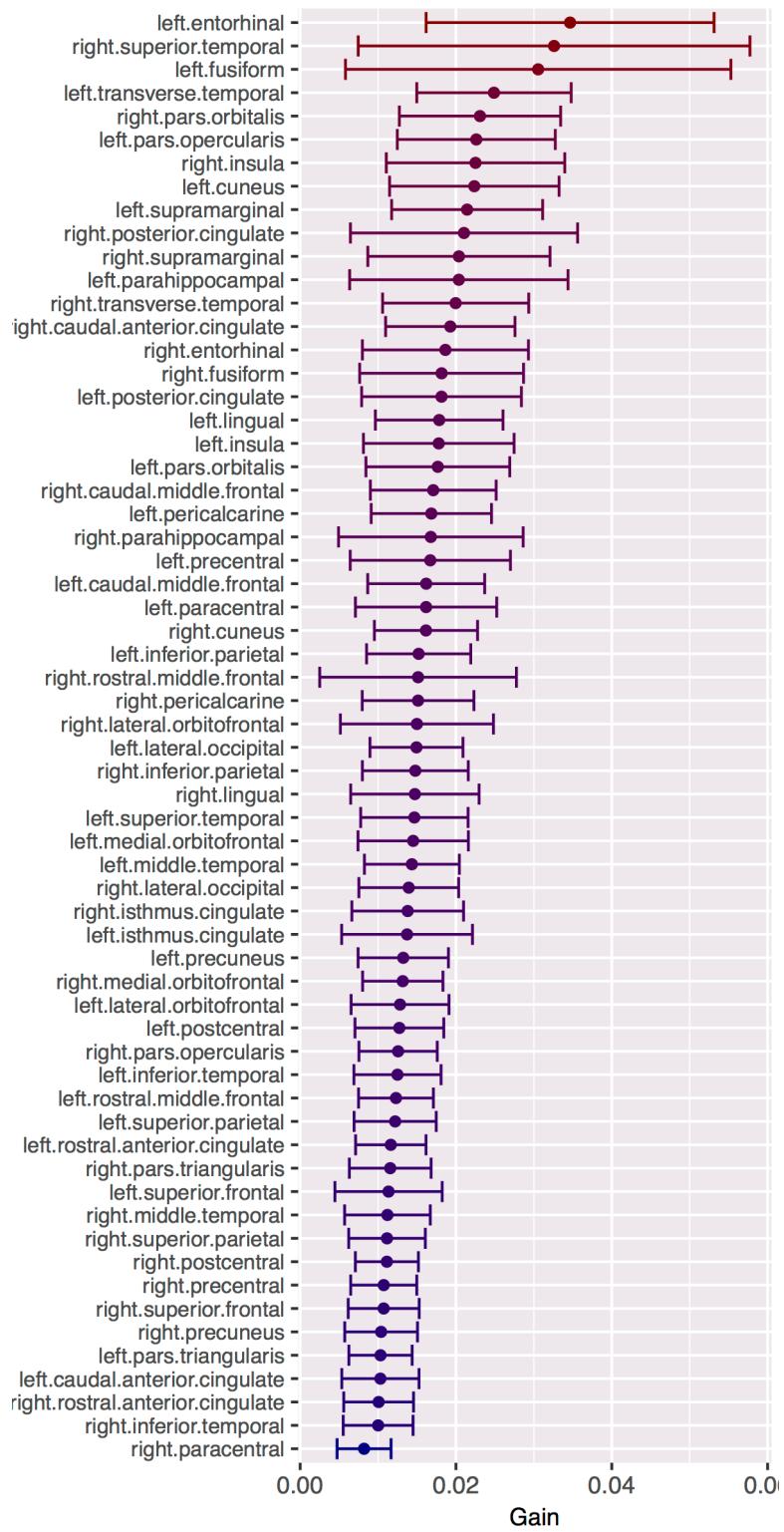


Figure 11: The “gain” measurements of the extreme gradient boosting models for the Longitudinal 2 processing.

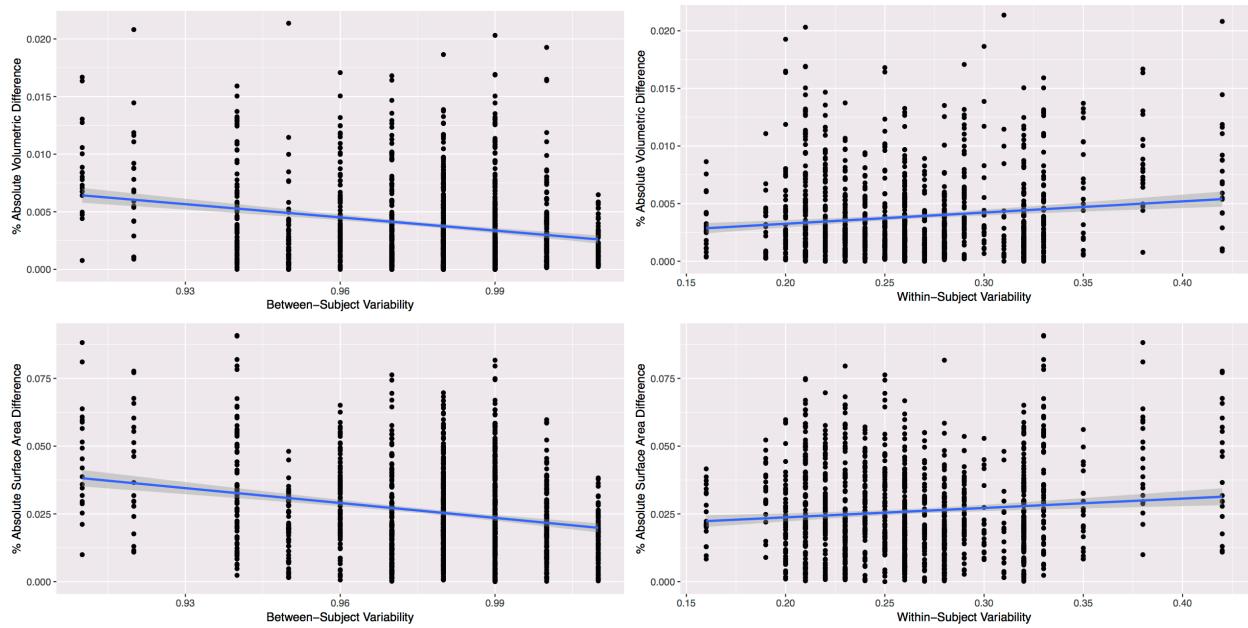


Figure 12: The between-subject variability (left) and within-subject variability (right) versus the volumetric (top) and surface area (bottom) measurements. Note that each of these correlations are significant ($p < 0.001$)

Discussion

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. Jubault, T., Gagnon, J.-F., Karama, S., Ptito, A., Lafontaine, A.-L., Evans, A. C., and Monchi, O. “**Patterns of Cortical Thickness and Surface Area in Early Parkinson’s Disease**” *Neuroimage* 55, no. 2 (2011): 462–7. doi:10.1016/j.neuroimage.2010.12.043
4. Thompson, P. M., Lee, A. D., Dutton, R. A., Geaga, J. A., Hayashi, K. M., Eckert, M. A., Bellugi, U., Galaburda, A. M., Korenberg, J. R., Mills, D. L., Toga, A. W., and Reiss, A. L. “**Abnormal Cortical Complexity and Thickness Profiles Mapped in Williams Syndrome**” *J Neurosci* 25, no. 16 (2005): 4146–58. doi:10.1523/JNEUROSCI.0165-05.2005
5. Ramasamy, D. P., Benedict, R. H. B., Cox, J. L., Fritz, D., Abdelrahman, N., Hussein, S., Minagar, A., Dwyer, M. G., and Zivadinov, R. “**Extent of Cerebellum, Subcortical and Cortical Atrophy in Patients with MS: A Case-Control Study**” *J Neurol Sci* 282, no. 1-2 (2009): 47–54. doi:10.1016/j.jns.2008.12.034
6. Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. “**Cortical Thickness Analysis in Autism with Heat Kernel Smoothing**” *Neuroimage* 25, no. 4 (2005): 1256–65. doi:10.1016/j.neuroimage.2004.12.052
7. Hardan, A. Y., Muddasani, S., Vemulapalli, M., Keshavan, M. S., and Minshew, N. J. “**An MRI Study of Increased Cortical Thickness in Autism**” *Am J Psychiatry* 163, no. 7 (2006): 1290–2. doi:10.1176/appi.ajp.163.7.1290
8. DaSilva, A. F. M., Granziera, C., Snyder, J., and Hadjikhani, N. “**Thickening in the Somatosensory Cortex of Patients with Migraine**” *Neurology* 69, no. 21 (2007): 1990–5. doi:10.1212/01.wnl.0000291618.32247.2d
9. Kühn, S., Schubert, F., and Gallinat, J. “**Reduced Thickness of Medial Orbitofrontal Cortex in**

Smokers” *Biol Psychiatry* 68, no. 11 (2010): 1061–5. doi:10.1016/j.biopsych.2010.08.004

10. Fortier, C. B., Leritz, E. C., Salat, D. H., Venne, J. R., Maksimovskiy, A. L., Williams, V., Milberg, W. P., and McGlinchey, R. E. “**Reduced Cortical Thickness in Abstinent Alcoholics and Association with Alcoholic Behavior**” *Alcohol Clin Exp Res* 35, no. 12 (2011): 2193–201. doi:10.1111/j.1530-0277.2011.01576.x
11. Makris, N., Gasic, G. P., Kennedy, D. N., Hodge, S. M., Kaiser, J. R., Lee, M. J., Kim, B. W., Blood, A. J., Evins, A. E., Seidman, L. J., Iosifescu, D. V., Lee, S., Baxter, C., Perlis, R. H., Smoller, J. W., Fava, M., and Breiter, H. C. “**Cortical Thickness Abnormalities in Cocaine Addiction—a Reflection of Both Drug Use and a Pre-Existing Disposition to Drug Abuse?**” *Neuron* 60, no. 1 (2008): 174–88. doi:10.1016/j.neuron.2008.08.011
12. Nesvåg, R., Lawyer, G., Varnäs, K., Fjell, A. M., Walhovd, K. B., Frigessi, A., Jönsson, E. G., and Agartz, I. “**Regional Thinning of the Cerebral Cortex in Schizophrenia: Effects of Diagnosis, Age and Antipsychotic Medication**” *Schizophr Res* 98, no. 1-3 (2008): 16–28. doi:10.1016/j.schres.2007.09.015
13. Lyoo, I. K., Sung, Y. H., Dager, S. R., Friedman, S. D., Lee, J.-Y., Kim, S. J., Kim, N., Dunner, D. L., and Renshaw, P. F. “**Regional Cerebral Cortical Thinning in Bipolar Disorder**” *Bipolar Disord* 8, no. 1 (2006): 65–74. doi:10.1111/j.1399-5618.2006.00284.x
14. Jacobus, J., Squeglia, L. M., Meruelo, A. D., Castro, N., Brumback, T., Giedd, J. N., and Tapert, S. F. “**Cortical Thickness in Adolescent Marijuana and Alcohol Users: A Three-Year Prospective Study from Adolescence to Young Adulthood**” *Dev Cogn Neurosci* 16, (2015): 101–9. doi:10.1016/j.dcn.2015.04.006
15. Sowell, E. R., Kan, E., Yoshii, J., Thompson, P. M., Bansal, R., Xu, D., Toga, A. W., and Peterson, B. S. “**Thinning of Sensorimotor Cortices in Children with Tourette Syndrome**” *Nat Neurosci* 11, no. 6 (2008): 637–9. doi:10.1038/nn.2121
16. Wang, D., Shi, L., Chu, W. C. W., Burwell, R. G., Cheng, J. C. Y., and Ahuja, A. T. “**Abnormal Cerebral Cortical Thinning Pattern in Adolescent Girls with Idiopathic Scoliosis**” *Neuroimage* 59, no. 2 (2012): 935–42. doi:10.1016/j.neuroimage.2011.07.097
17. Kumar, R., Yadav, S. K., Palomares, J. A., Park, B., Joshi, S. H., Ogren, J. A., Macey, P. M., Fonarow, G. C., Harper, R. M., and Woo, M. A. “**Reduced Regional Brain Cortical Thickness in Patients with Heart Failure**” *PLoS One* 10, no. 5 (2015): e0126595. doi:10.1371/journal.pone.0126595
18. Jiang, J., Zhu, W., Shi, F., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., and Jiang, T. “**Thick Visual Cortex in**

the Early Blind” *J Neurosci* 29, no. 7 (2009): 2205–11. doi:10.1523/JNEUROSCI.5451-08.2009

19. Frøkjær, J. B., Bouwense, S. A. W., Olesen, S. S., Lundager, F. H., Eskildsen, S. F., Goor, H. van, Wilder-Smith, O. H. G., and Drewes, A. M. **“Reduced Cortical Thickness of Brain Areas Involved in Pain Processing in Patients with Chronic Pancreatitis”** *Clin Gastroenterol Hepatol* 10, no. 4 (2012): 434–8.e1. doi:10.1016/j.cgh.2011.11.024
20. Shin, Y.-W., Yoo, S. Y., Lee, J. K., Ha, T. H., Lee, K. J., Lee, J. M., Kim, I. Y., Kim, S. I., and Kwon, J. S. **“Cortical Thinning in Obsessive Compulsive Disorder”** *Hum Brain Mapp* 28, no. 11 (2007): 1128–35. doi:10.1002/hbm.20338
21. Almeida Montes, L. G., Prado Alcántara, H., Martínez García, R. B., De La Torre, L. B., Avila Acosta, D., and Duarte, M. G. **“Brain Cortical Thickness in ADHD: Age, Sex, and Clinical Correlations”** *J Atten Disord* (2012): doi:10.1177/1087054711434351
22. Raji, C. A., Ho, A. J., Parikshak, N. N., Becker, J. T., Lopez, O. L., Kuller, L. H., Hua, X., Leow, A. D., Toga, A. W., and Thompson, P. M. **“Brain Structure and Obesity”** *Hum Brain Mapp* 31, no. 3 (2010): 353–64. doi:10.1002/hbm.20870
23. Peterson, B. S., Warner, V., Bansal, R., Zhu, H., Hao, X., Liu, J., Durkin, K., Adams, P. B., Wickramaratne, P., and Weissman, M. M. **“Cortical Thinning in Persons at Increased Familial Risk for Major Depression”** *Proc Natl Acad Sci U S A* 106, no. 15 (2009): 6273–8. doi:10.1073/pnas.0805311106
24. Ballmaier, M., Sowell, E. R., Thompson, P. M., Kumar, A., Narr, K. L., Lavretsky, H., Welcome, S. E., DeLuca, H., and Toga, A. W. **“Mapping Brain Size and Cortical Gray Matter Changes in Elderly Depression”** *Biol Psychiatry* 55, no. 4 (2004): 382–9. doi:10.1016/j.biopsych.2003.09.004
25. Kochunov, P., Glahn, D. C., Lancaster, J., Thompson, P. M., Kochunov, V., Rogers, B., Fox, P., Blangero, J., and Williamson, D. E. **“Fractional Anisotropy of Cerebral White Matter and Thickness of Cortical Gray Matter Across the Lifespan”** *Neuroimage* 58, no. 1 (2011): 41–9. doi:10.1016/j.neuroimage.2011.05.050
26. Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Woods, R. P., Deluca, H., Jancke, L., and Toga, A. W. **“Gender Effects on Cortical Thickness and the Influence of Scaling”** *Hum Brain Mapp* 27, no. 4 (2006): 314–24. doi:10.1002/hbm.20187
27. Luders, E., Sánchez, F. J., Tosun, D., Shattuck, D. W., Gaser, C., Vilain, E., and Toga, A. W. **“Increased Cortical Thickness in Male-to-Female Transsexualism”** *J Behav Brain Sci* 2, no. 3 (2012): 357–362.

doi:10.4236/jbbs.2012.23040

28. Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., and Toga, A. W. “**Hemispheric Asymmetries in Cortical Thickness**” *Cereb Cortex* 16, no. 8 (2006): 1232–8. doi:10.1093/cercor/bhj064
29. Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., and Giedd, J. “**Intellectual Ability and Cortical Development in Children and Adolescents**” *Nature* 440, no. 7084 (2006): 676–9. doi:10.1038/nature04513
30. Wei, G., Zhang, Y., Jiang, T., and Luo, J. “**Increased Cortical Thickness in Sports Experts: A Comparison of Diving Players with the Controls**” *PLoS One* 6, no. 2 (2011): e17112. doi:10.1371/journal.pone.0017112
31. Lazar, S. W., Kerr, C. E., Wasserman, R. H., Gray, J. R., Greve, D. N., Treadway, M. T., McGarvey, M., Quinn, B. T., Dusek, J. A., Benson, H., Rauch, S. L., Moore, C. I., and Fischl, B. “**Meditation Experience Is Associated with Increased Cortical Thickness**” *Neuroreport* 16, no. 17 (2005): 1893–7.
32. Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. “**Neuroanatomical Correlates of Musicianship as Revealed by Cortical Thickness and Voxel-Based Morphometry**” *Cereb Cortex* 19, no. 7 (2009): 1583–96. doi:10.1093/cercor/bhn196
33. Foster, N. E. V. and Zatorre, R. J. “**Cortical Structure Predicts Success in Performing Musical Transformation Judgments**” *Neuroimage* 53, no. 1 (2010): 26–36. doi:10.1016/j.neuroimage.2010.06.042
34. Hudziak, J. J., Albaugh, M. D., Ducharme, S., Karama, S., Spottswood, M., Crehan, E., Evans, A. C., Botteron, K. N., and Brain Development Cooperative Group. “**Cortical Thickness Maturation and Duration of Music Training: Health-Promoting Activities Shape Brain Development**” *J Am Acad Child Adolesc Psychiatry* 53, no. 11 (2014): 1153–61, 1161.e1–2. doi:10.1016/j.jaac.2014.06.015
35. Raine, A., Laufer, W. S., Yang, Y., Narr, K. L., Thompson, P., and Toga, A. W. “**Increased Executive Functioning, Attention, and Cortical Thickness in White-Collar Criminals**” *Hum Brain Mapp* (2011): doi:10.1002/hbm.21415
36. Heim, C. M., Mayberg, H. S., Mletzko, T., Nemeroff, C. B., and Pruessner, J. C. “**Decreased Cortical Representation of Genital Somatosensory Field After Childhood Sexual Abuse**” *Am J Psychiatry* 170, no. 6 (2013): 616–23. doi:10.1176/appi.ajp.2013.12070950
37. Haier, R. J., Karama, S., Leyba, L., and Jung, R. E. “**MRI Assessment of Cortical Thickness and Functional Activity Changes in Adolescent Girls Following Three Months of Practice on a Visual-**

Spatial Task” *BMC Res Notes* 2, (2009): 174. doi:10.1186/1756-0500-2-174

38. 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
39. 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 Gyrification in the Human Brain: Changes Associated with Aging**” *Cereb Cortex* 9, no. 2 (1999): 151–60.
40. 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
41. 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
42. 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.
43. 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
44. 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):
45. 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
46. 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
47. Fischl, B. and Dale, A. M. “**Measuring the Thickness of the Human Cerebral Cortex from Magnetic Resonance Images**” *Proc Natl Acad Sci USA* 97, no. 20 (2000): 11050–5. doi:10.1073/pnas.200033797
48. Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrave, 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.
49. 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.
50. 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
51. 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.
52. 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
53. 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.
54. 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**” *Neuroimage* 50, no. 2 (2010): 434–45. doi:10.1016/j.neuroimage.2009.12.007
55. 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
56. 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
57. 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
58. 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
59. 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
60. 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
61. 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
62. 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
63. 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
64. 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
65. 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
66. 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
67. 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.
68. 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
69. 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
70. 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):
71. 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
72. 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
73. 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
74. 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):
75. 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

76. Available at <http://brain-development.org/ixi-dataset/>

77. 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

78. Available at http://fcon_1000.projects.nitrc.org/indi/pro/nki.html

79. Available at <http://www.oasis-brains.org>

80. Breiman, L. “**Random Forests**” *Machine learning* (2001): 5–32.

81. Lehmann, G. and Legland, D. “**Efficient N-Dimensional Surface Estimation Using Crofton Formula and Run-Length Encoding**” *Insight Journal* (2012):

82. 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

83. 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

84. 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

85. 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

86. 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

87. Verbeke, G. and Molenberghs, G. “**Linear Mixed Models for Longitudinal Data**” (2009):
88. Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. “**Applied Longitudinal Analysis**” 998, (2012):
89. 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
90. 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):
91. Available at <http://xgboost.readthedocs.io/en/latest/>
92. 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