bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The ANTs Longitudinal Cortical Thickness Pipeline

Nicholas J. Tustison[1,2], Andrew J. Holbrook[3], Brian B. Avants[4†], Jared M. Roberts[2], Philip A. Cook[4], Zachariah M. Reagh[2], James R. Stone[1], Daniel L. Gillen[3], and Michael A. Yassa[2] for the Alzheimer's Disease Neuroimaging Initiative*

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

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

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

[4]Department of Radiology, University of Pennsylvania, Philadelphia, PA

Corresponding author:

Nicholas J. Tustison

211 Qureshey Research Lab

Irvine, CA 92697-3800

ntustison@virginia.edu

† Currently employed by Biogen (Cambridge, MA).

*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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Longitudinal studies of development and 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 quantifying neurostructural changes. Longitudinal-specific strategies for acquisition and processing have potentially significant benefits including more consistent estimates of intra-subject measurements while retaining predictive power. 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 in these data that the single-subject template construction and native subject-space processing advantageously localizes data transformations and mitigates interpolation artifacts which results in a simultaneous minimization of within-subject variability and maximization of between-subject variability immediately estimable from a longitudinal mixed-effects modeling strategy. 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 narrower confidence intervals for determining cross-sectional effects. These concepts are first illustrated and explored in the entorhinal cortex. This evaluation strategy is then extended to the entire cortex, as defined by the Desikan-Killiany-Tourville labeling protocol, where comparisons are made with the popular cross-sectional and longitudinal FreeSurfer processing streams.

Keywords: Advanced Normalization Tools, entorhinal cortex, FreeSurfer, interpolation, longitudinal mixed-effects, longitudinal processing

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 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 and others have since been used for making specific recommendations with respect to longitudinal image data processing [17, 21–23].

In [17, 24], 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 [25] and has been used in a variety of studies including pediatric cortical development [26], differential development in Alzheimer's disease and fronto-temporal dementia [27], and fatigue in the context of multiple sclerosis [28].

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We introduced the Advanced Normalization Tools (ANTs) cortical thickness framework in [29] 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 [30]. This proposed ANTs-based pipeline has since been directed at a variety of neuroimaging research topics including mild cognitive impairment and depression [31], short term memory in mild cognitive impairment [32], and aphasia [33]. Other authors extended the general framework to non-human studies [34, 35].

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 previously outlined research illustrating the negative impact of interpolation effects on study results, we show that the common practice of reorienting individual time point images to a single-subject template for unbiased processing induces interpolation 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.

2 Methods and materials

2.1 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 [36] 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 [37]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

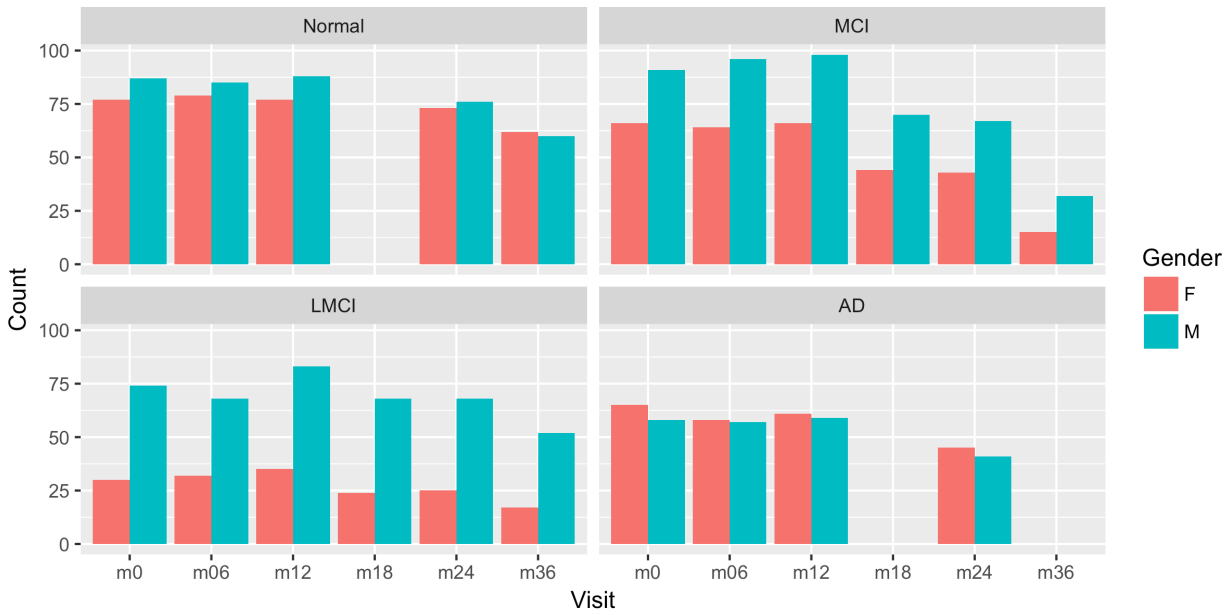

Figure 1: Demographic breakdown of the number of ADNI-1 subjects by diagnosis i.e., normal, mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). Within each panel we plot the number of subjects (by gender) per visit—baseline ("bl") and n months ("mn").

The ADNI-1 data was downloaded in May of 2014 and first processed using the ANTs cross-sectional cortical thickness pipeline [29] (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 in the github repository associated with this work [38]), 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 with one or more follow-up image acquisition appointments. Further breakdown of demographic information is given in Figures 1 and 2 to provide additional perspective on the data used for this work.

2.2 ANTs cortical thickness

2.2.1 Cross-sectional processing

A thorough discussion of the ANTs cross-sectional thickness estimation framework was previously discussed in [29]. As a brief review, given a T1-weighted brain MR image, processing comprises the

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

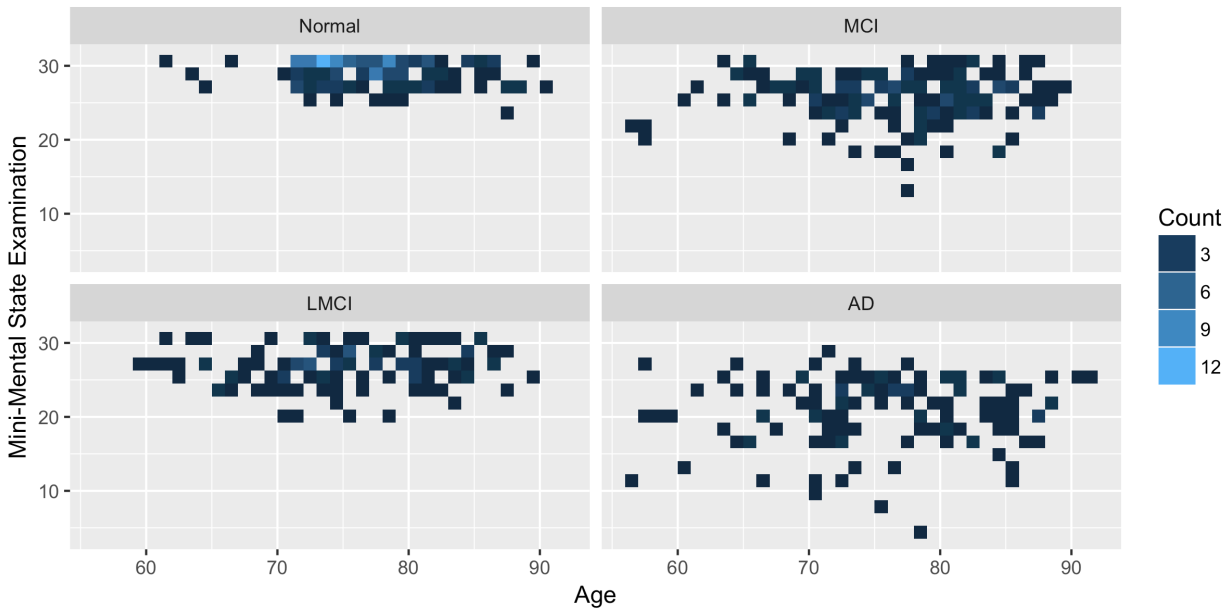

Figure 2: Age vs. Mini-mental examination (MMSE) scores for the ADNI-1 subjects by diagnosis providing additional demographic characterization for the subjects processed for this study.

following major steps (cf Figure 1 of [29]):

1. N4 bias correction [39],
2. brain extraction [40],
3. Atropos n-tissue segmentation [41], and
4. cortical thickness estimation [8].

ROI-based quantification is achieved through joint label fusion [42] of the cortex coupled with the MindBoggle-101 data. These data use the Desikan–Killiany–Tourville (DKT) labeling protocol [43] to parcellate each cortical hemisphere into 31 anatomical regions (cf Table 1). This pipeline has since been enhanced by the implementation [44] of a patch-based denoising algorithm [45] as an optional preprocessing step and multi-modal integration capabilities (e.g., joint T1- and T2-weighted image processing).

For evaluation, voxelwise regional thickness statistics were summarized based on the DKT parcellation scheme. Test-retest error measurements were presented from a 20 cohort subset of both the OASIS [46] and MMRR [47] data sets and compared with the corresponding FreeSurfer thickness values. Further evaluation employed a training/prediction paradigm whereby DKT regional cortical

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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) rosterior 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`)	

thickness values generated from 1205 images taken from four publicly available data sets (i.e., IXI [48], MMRR, NKI [49], and OASIS) were used to predict age and gender using linear and random forest [50] models. The resulting regional statistics (including cortical thickness, surface area [51], volumes, and Jacobian determinant values) were made available online [52]. These include the corresponding FreeSurfer measurements which are also publicly available for research inquiries (e.g., [53]). Since publication, this framework has been used in a number of studies (e.g., [54–56]).

2.2.2 Unbiased longitudinal processing

Given certain practical limitations (e.g., subject recruitment and retainment), as mentioned earlier, many researchers employ cross-sectional acquisition and processing strategies for studying develop-

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

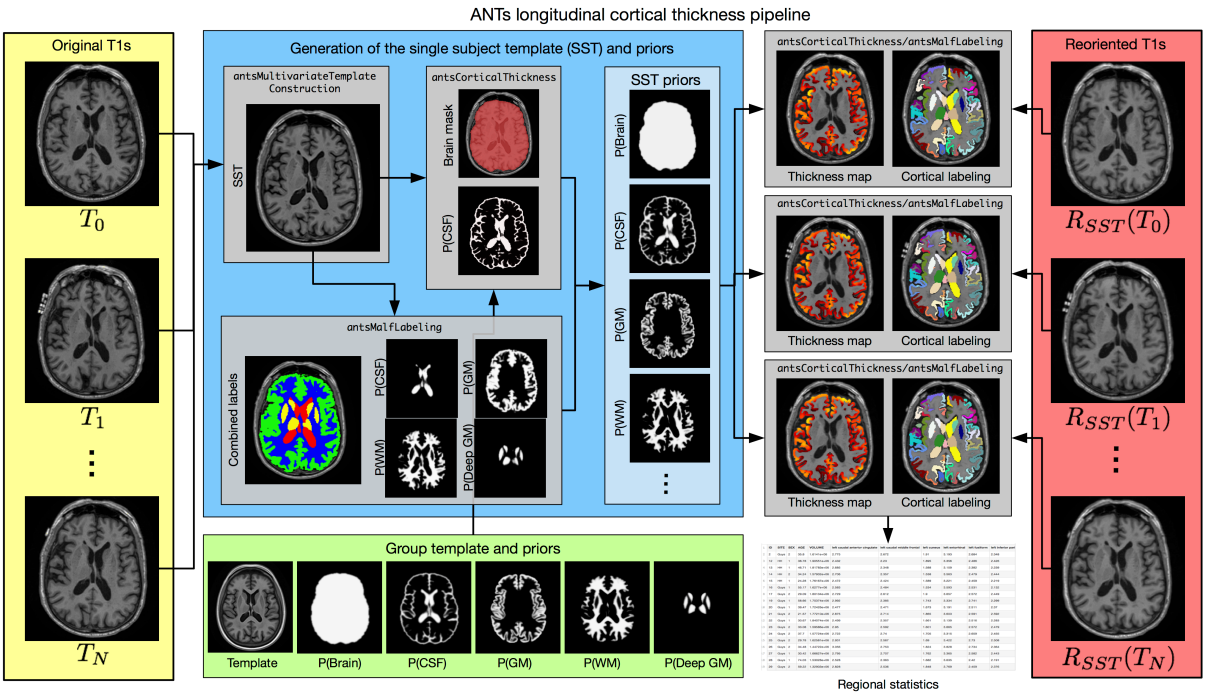

Figure 3: Diagrammatic illustration of the ANTs longitudinal cortical thickness pipeline for a single subject with N time points. From the N original T1-weighted images (left column, yellow panel) and the group template and priors (bottom row, green panel), the single-single subject template (SST) and 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 are propagated to each image using a given atlas set (with cortical labels) and joint label fusion.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

mental 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, 24].

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. This processes the SST based on the group template.
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 – with SST as reference template – to each individual time-point image. This processes individual time points based on the SST yet also allows concatenation of transforms to the group template.
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 [57]. 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 [29]. 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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

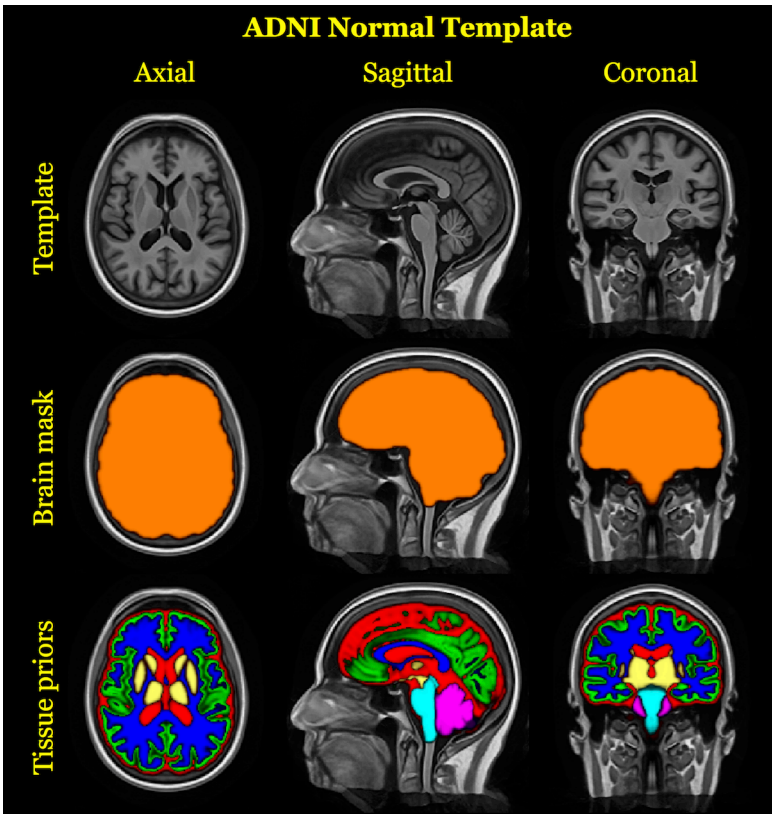

Figure 4: Top row: Canonical views of the template created from 52 cognitively normal subjects of the ADNI-1 database. The prior probability mask for the whole brain (middle row) and the six tissue priors (bottom row) are used to "seed" each single-subject template for creation of a probabilistic brain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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

protocol [57] 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 and made publicly available [43], a multi-atlas joint label fusion step (JLF) [42] 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 the 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., $P(CSF) = \max\left(P_{Seg}(CSF), P_{JLF}(CSF)\right)$, 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. The first step for subject-wise processing involves creating the SST from all the time points for that individual [57]. 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 [29]. However, in the longitudinal variant, the SST serves this purpose. We thus map the SST and its priors to the native space of each time point where individual-level 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 an option allowing for rotation of the individual time points to the SST would be of benefit, similar to FreeSurfer, in reducing variability, minimizing or eliminating possible orientation bias, and possibly permitting a 4-D segmentation given that the underlying Atropos segmentation implementation is dimensionality-agnostic [41]. Regarding the 4-D brain segmentation, any possible benefit is potentially outweighed by the

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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.

Joint label fusion and pseudo-geodesic for large cohort labeling. Cortical thickness ROI-based analyses are performed using joint label fusion [42] 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 (e.g., [58]). 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 for estimating regional labels for each image.

2.3 Statistical evaluation

Based on the above ANTs pipeline descriptions, there are three major variants for cortical thickness processing of longitudinal data. We denote these alternatives as:

- **ANTs Cross-sectional** (or **ANTs Cross**). Process each subject's time point independently using the cross-sectional pipeline originally described in [29].
- **ANTs Longitudinal-SST** (or **ANTs SST**). Rigidly transform each subject to the SST and then segment and estimate cortical thickness in the space of the SST.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

- **ANTs Longitudinal-native** (or **ANTs Native**). Segment and estimate cortical thickness in the native space.

For completeness, we also include a comparison with both the cross-section and longitudinal FreeSurfer v5.3 streams respectively denoted as "FreeSurfer Cross-sectional" (or "FS Cross") and "FreeSurfer Longitudinal" (or "FS Long").

Possible evaluation strategies could employ manual measurements in the histological [59] or virtual [60] domains but would require an inordinate labor effort for collection to be comparable with the size of data sets currently analyzed. Other quantitative measures representing "reliability", "reproducibility", or, more generally, "precision" can also be used to characterize such tools. For example, [61] used FreeSurfer cortical thickness measurements across image acquisition sessions to demonstrate improved reproducibility with the longitudinal stream over the cross-sectional stream. In [62] comparisons for ANTs, FreeSurfer, and the proposed method were made using the range of measurements and their correspondence to values published in the literature. However, none of these precision-type measurements, per se, indicate the utility of a pipeline-specific cortical thickness value as a potential biomarker. For example, Figure 8 in [29] confirms what was found in [62] which is that the range of ANTs cortical thickness values for a particular region exceeds those of FreeSurfer. However, for the same data, the demographic predictive capabilities of the former was superior to that of the latter. Thus, better assessment strategies are necessary for determining clinical utility. For example, the intra-class correlation (ICC) coefficient used in [29] demonstrated similarity in both ANTs and FreeSurfer for repeated acquisitions despite the variance discrepancy between both sets of measurements. This is understood with the realization that the ICC takes into account both inter-observer and intra-observer variability.

2.3.1 Regional within-subject and between-subject variance

A summary measure related to the ICC statistic [63] is used to quantify the relative performance of these cross-sectional and longitudinal ANTs pipeline variants and a comparison with their FreeSurfer comparisons. Specifically, we use longitudinal mixed-effects (LME) modeling to quantify pipeline-specific between-subject and within-subject variabilities with the intuition that comparative performance is determined by maximizing the ratio between the former and the latter. Such a quantity

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

implies greater within-subject reproducibility while distinguishing between patient sub-populations (e.g., Alzheimer's disease diagnosis). 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 LME model (1) outlined below.

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, [25] 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 [63] and [64]. 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 [65] 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 ith individual's cortical thickness measurement corresponding to the kth region of interest at the time point indexed by j. Under the Bayesian paradigm we utilized a model of the form

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

Specification of variance priors to half-Cauchy distributions reflects commonly accepted practice in the context of hierarchical models [66]. 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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

region k, the quantity of interest is thus the ratio

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

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. The posterior distribution of r^k was summarized via the posterior median where the posterior distributions were obtained using the Stan probabilistic programming language [67].

2.3.2 Practical implications of the variance ratio, a case study

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 [68–71]. In the ADNI sample, EC cortical thickness was the most powerful measure of structural change both in MCI and AD subjects [72]. EC thinning was also found to precede and predict hippocampal atrophy [73] and to predict conversion to AD with the greatest accuracy [74]. Thus, we chose the EC to be the target of an additional focused analysis to determine the relative utility of the different ANTs pipelines for measuring thickness in this particular region. Our choice of EC for comparative performance assessment is motivated both by its selective vulnerability to neurodegenerative processes as well as the difficulty of image segmentation in that particular region.

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 sub-populations 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.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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:

$$ECCT_i = \beta_0 + \beta_1 AD_i + \epsilon_i \,. \tag{3}$$

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.

3 Results

Based on the evaluation design described in the previous section, we compare the performance of the five processing approaches (FS Cross, FS Long, ANTs Cross, ANTs SST, and ANTs 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 for these different pipelines.

3.0.1 Cortical within-subject and between-subject thickness variability

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 five pipelines. The superior method is designated by larger variance ratios and has the greater discriminative capacity for the data corresponding to that processing method.

ANTs Native has the highest ratio variance across most of the 62 regions over the other methods. It rarely overlaps with ANTs SST and never with ANTs Cross. In contrast to the majority of FreeSurfer regional ratio variances (from both FS Cross and FS Long) which are smaller than those of the ANTs

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

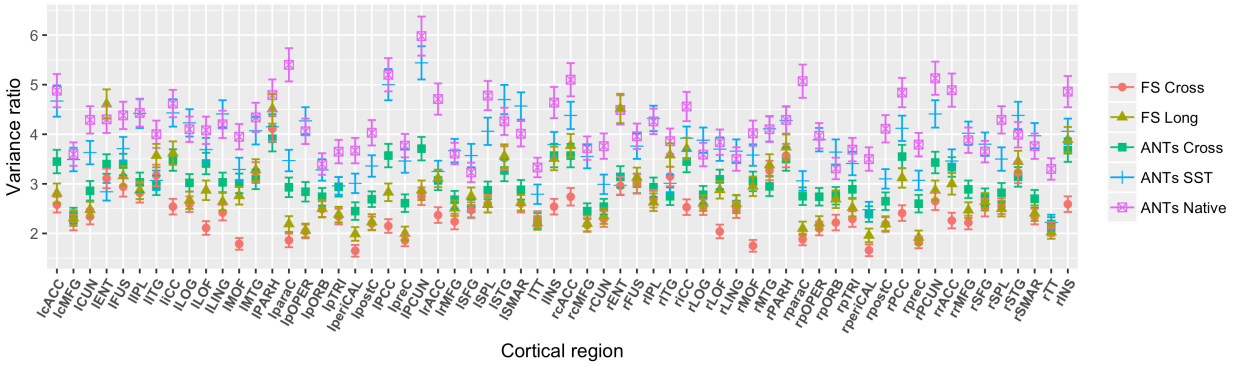

Figure 5: 95% credible intervals of the region-specific variance ratios $r^k = \tau_k/\sigma_k$ are presented for each processing method. The ANTs Longitudinal-native method dominates the others across the majority of regions: its point estimates (posterior medians) are greater than those of the other processing methods except for the left and right EC values in FreeSurfer Long (although there is significant overlap in the credible intervals in those regions). These results also suggest that longitudinal processing is to be preferred for both packages.

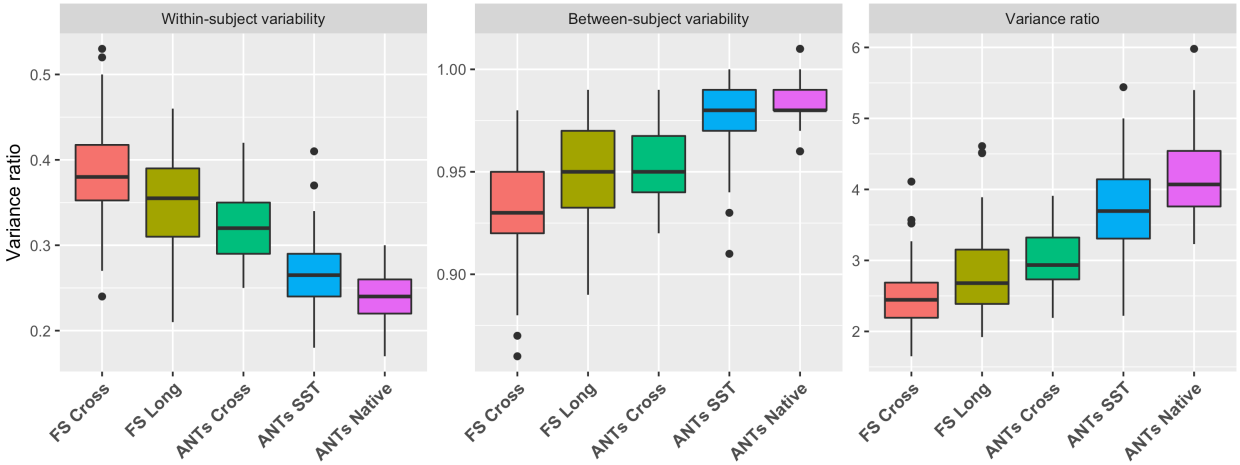

Figure 6: Box plots showing the distribution of the within-subject variability, between subject variability, and ratio of the between-subject variability and within-subject variability for each of the 62 DKT regions. Note that the "better" measurement maximizes this latter ratio.

pipelines, FS Long has larger ratio values for the EC region with the only overlap in the credible intervals with ANTs Native.

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 box plots. These relative distributions show that both between-subject and within-subject quantities contribute to the dispar-

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

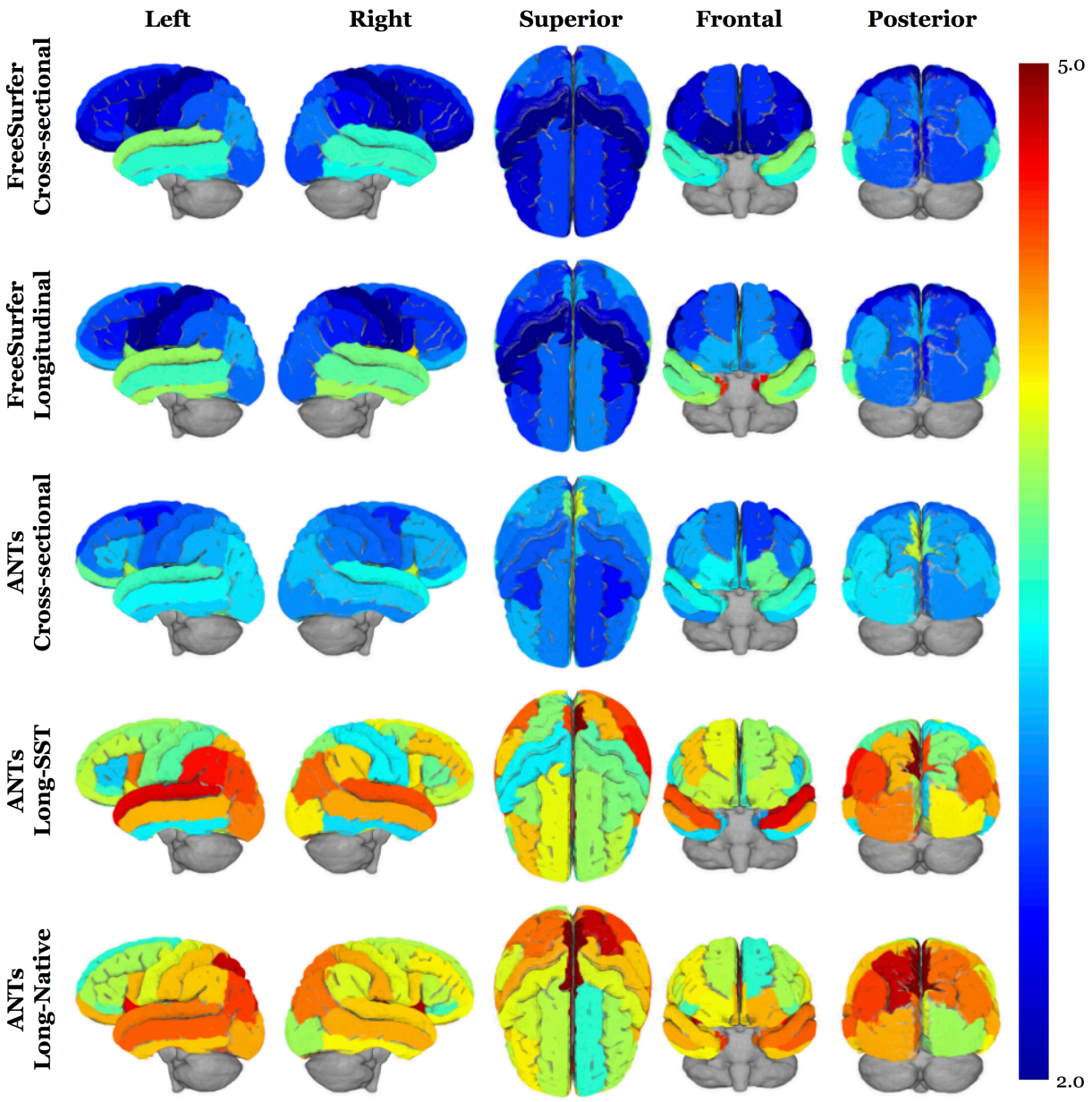

Figure 7: 3-D volumetric rendering of the regional variance ratio values on the generated ADNI template. The higher variance ratios indicate greater between-subject to within-subject variability.

ities 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. Therefore, Figures 5, 6, and 7 may be considered as evidence for ANTs Native providing higher quality data than those provided by the other methods.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3.0.2 Case study: entorhinal cortical thickness variability

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						
unnormalized	3.25	0.0020	0.09	0.097	1.01	0.41
normalized	0.75	0.81	0.83	0.76	1	1
Long-SST						
unnormalized	2.91	0.0025	0.11	0.128	1.00	0.42
normalized	0.68	1	1	1	0.99	1
Long-native						
unnormalized	4.36	0.0017	0.05	0.058	1.01	0.39
normalized	1	0.67	0.48	0.46	1	0.93

Here we present results from the entorhinal cortical thickness case study (Section 2.3.2). We demonstrate the way in which different arrangements of within-subject and between-subject variability influence statistical inference and prediction.

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 EC 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 corresponding 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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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, ANTs 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 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 ANTs Native method performs best whereas ANTs SST performs worst.

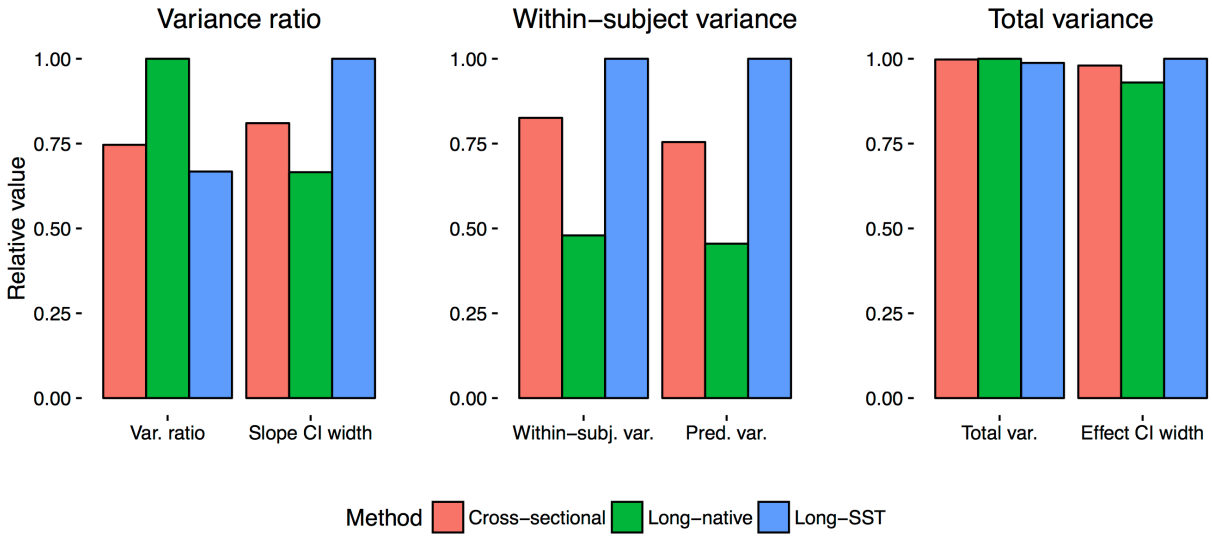

Figure 8: Aspects of model variance are compared with credible interval sizes and variance in predictions for the ANTs-based pipelines. Values are normalized by the largest quantity, and processing methods are distinguished by color and ordering. On the left, the variance ratio $r = \tau/\sigma$ is compared 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 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 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

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 ANTs Native is 0.81, whereas the distances for ANTs SST and ANTs Cross are 0.75 and 0.70, respectively. Figure 8 displays the normalized results.

4 Discussion

Herein we detailed the ANTs longitudinal cortical thickness framework which is 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 [29] and two longitudinal variants. One of the variants, ANTs 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. ANTs 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 discriminative capacity. In these terms, ANTs Native outperformed all other pipeline variants including both the FreeSurfer longitudinal and cross-sectional streams. Regional disparities between the ANTs Native and SST pipelines point to increases in both between-subject and within-subject variances which might be due to interpolation artifacts. In other words, interpolation potentially has a systematic but regionally varying effect. Investigation of this issue is the subject of future research.

One very interesting finding was the superior performance of FS Long in the EC regions where the variance ratios was slightly larger than those of ANTs Native where the credible intervals have significant overlap. Given the small volume and indistinguishability from surrounding structures, segmentation of the EC can be relatively difficult [75]. This segmentation complexity has led to EC-specific [76] and related [77] strategies for targeted regional processing. For this work, we wanted to avoid

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

such tuning and simply employ off-the-shelf input parameters and data. Future work will explore refining input template priors in these problematic regions for ANTs-based estimation of cortical thickness.

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 [29]. Both assessments 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 (cf Appendix). 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.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5 Appendix

5.1 Implementation overview

The script `antsLongitudinalCorticalThickness.sh` performs cortical thickness estimation for a longitudinal image series from a single subject. The following principal steps are performed:

1. A single-subject template (SST) is created from all the time point images.
2. The tissue prior probability images are generated for the SST. These six tissues are label 1: CSF, label 2: cortical gray matter, label 3: white matter, label 4: deep gray matter, label 5: brain stem, and label 6: cerebellum. Prior probability creation involves the following steps:
    1. The SST is passed through `antsCorticalThicknes.sh`.
    2. The brain extraction posterior for the SST is created by smoothing the brain extraction mask created during 2a.
    3. If labeled atlases are not provided, we smooth the posteriors from 2.1 to create the SST segmentation priors, otherwise we use the `antsJointFusion` program to create a set of posteriors using the script `antsCookTemplatePriors.sh`.
3. Using the SST + priors, each subject is processed through the `antsCorticalThickness.sh` script.

A typical command line call is:

```
antsLongitudinalCorticalThickness.sh \
              -d ${imageDimension} \
              -e ${brainTemplate} \
              -m ${brainExtractionProbabilityMask} \
              -p ${brainSegmentationPriors}
              -o ${outputPrefix}
              ${anatomicalImages[@]}
```

5.2 Input parameters

- `imageDimension`: dimensionality of the input images. Can handle 2 or 3 dimensions.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

- `brainTemplate`: the group template. We have made several publicly available along with the prior tissue and brain extraction images (https://figshare.com/articles/ANTs_ANTsR_Brain_Templates/915436).
- `brainExtractionProbabilityMask`: prior probability image for the whole brain corresponding to the `brainTemplate`.
- `brainSegmentationPriors`: prior probability images for the six brain tissues mentioned above. These files are specified with the relevant labels, e.g., `prior1.nii.gz`, `prior2.nii.gz`, `prior3.nii.gz`, `prior4.nii.gz`, `prior5.nii.gz`, and `prior6.nii.gz`. The command line argument is specified in C-style formatting, e.g., `prior%d.nii.gz`.
- `anatomicalImages`: the time point images for a single subjects.
- other optional input parameters are available. `antsLongitudinalCorticalThickness -h` provides a listing of the full set of parameters, their descriptions, and other help information.

5.3 Output

In the specified output directory, the following subdirectories are created:

- `${outputPrefix}SST`
- `${outputPrefix}${anatomicalImagesPrefix[0]}`
- `${outputPrefix}${anatomicalImagesPrefix[1]}`
- `${outputPrefix}${anatomicalImagesPrefix[2]}`
- ...

Each subdirectory contains the output of `antsCorticalThickness.sh` applied to the corresponding image. Output consists of the following files:

- `BrainExtractionMask`: Brain extraction mask in subject space.
- `BrainNormalizedToTemplate`: Extracted brain image normalized to the template space.
- `BrainSegmentation0N4`: Input to the segmentation algorithm. It is not brain extracted, but is bias-corrected. If multiple images are used for segmentation, there will be `BrainSegmentation1N4` and so on. The brain extracted version of this is `ExtractedBrain0N4`.
- `BrainSegmentation`: Segmentation image, one label per tissue class. The number of classes

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

is determined by the input priors.

- `BrainSegmentationPosteriors1`: Posterior probability of class 1. A similar image is produced for all classes. The numbering scheme matches the input priors.
- `CorticalThickness`: Cortical thickness image in subject space.
- `CorticalThicknessNormalizedToTemplate`: Cortical thickness image in template space.
- `ExtractedBrainON4`: Brain-extracted version of `BrainSegmentationOON4`.
- `SubjectToTemplate1Warp`, `SubjectToTemplateOGenericAffine.mat`: Transforms to be used when warping images from the subject space to the template space.
- `SubjectToTemplateLogJacobian`: Log of the determinant of the Jacobian, quantifies volume changes in the subject to template warp.
- `TemplateToSubjectOWarp`, `TemplateToSubject1GenericAffine.mat`: Transforms to be used when warping images from the template to the subject space.

In addition to these files, the SST subdirectory contains additional warps, suffixed "SubjectToGroupTemplateWarp.nii.gz" and "SubjectToTemplateOGenericAffine.mat", that can be used to warp each time point image to the group template. These are a combination of the subject to SST warp, and the SST to group template warp. Also included are the SST brain and tissue prior probability images.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgments

Additional support to N.T. and M.Y. provided by NIMH R01 MH102392 and NIA R21 AG049220, P50 AG16573.

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.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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., Feczko, 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 Gyrification 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 Thick-**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ness 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 Pub-**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

lished 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**" *Neuroimage* 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. Avants, B., Cook, P. A., McMillan, C., Grossman, M., Tustison, N. J., Zheng, Y., and Gee, J. C. "**Sparse Unbiased Analysis of Anatomical Variance in Longitudinal Imaging**" *Med Image Comput Comput Assist Interv* 13, no. Pt 1 (2010): 324–31.

22. 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

23. 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

24. Reuter, M. and Fischl, B. "**Avoiding Asymmetry-Induced Bias in Longitudinal Image**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Processing" *Neuroimage* 57, no. 1 (2011): 19–21. doi:10.1016/j.neuroimage.2011.02.076

25. 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

26. 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

27. 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

28. 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

29. 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

30. Available at https://github.com/stnava/ANTs

31. 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

32. 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

33. 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**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Primary Progressive Aphasia" *J Neurol* 263, no. 10 (2016): 1927–38. doi:10.1007/s00415-016-8221-1

34. Pagani, M., Damiano, M., Galbusera, A., Tsaftaris, S. A., and Gozzi, A. "**Semi-Automated Registration-Based Anatomical Labelling, Voxel Based Morphometry and Cortical Thickness Mapping of the Mouse Brain.**" *Journal of neuroscience methods* 267, (2016): 62–73. doi:10.1016/j.jneumeth.2016.04.007

35. Majka, P., Chaplin, T. A., Yu, H.-H., Tolpygo, A., Mitra, P. P., Wójcik, D. K., and Rosa, M. G. P. "**Towards a Comprehensive Atlas of Cortical Connections in a Primate Brain: Mapping Tracer Injection Studies of the Common Marmoset into a Reference Digital Template.**" *The Journal of comparative neurology* 524, no. 11 (2016): 2161–2181. doi:10.1002/cne.24023

36. 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.

37. 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

38. Available at https://github.com/ntustison/CrossLong

39. 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

40. 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):

41. 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*

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

9, no. 4 (2011): 381–400. doi:10.1007/s12021-011-9109-y

42. 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

43. 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

44. 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):

45. 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

46. Available at http://www.oasis-brains.org

47. 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

48. Available at http://brain-development.org/ixi-dataset/

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

50. Breiman, L. "**Random Forests**" *Machine learning* (2001): 5–32.

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

52. Available at https://github.com/ntustison/KapowskiChronicles

53. 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

54. Price, A. R., Bonner, M. F., Peelle, J. E., and Grossman, M. "**Converging Evidence for the**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Neuroanatomic Basis of Combinatorial Semantics in the Angular Gyrus" *J Neurosci* 35, no. 7 (2015): 3276–84. doi:10.1523/JNEUROSCI.3446-14.2015

55. 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

56. 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

57. 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

58. Tustison, N. J., Shrinidhi, K. L., Wintermark, M., Durst, C. R., Kandel, B. M., Gee, J. C., Grossman, M. C., and Avants, B. B. "**Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR**" *Neuroinformatics* 13, no. 2 (2015): 209–25. doi:10.1007/s12021-014-9245-2

59. Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., Kouwe, A. van der, Jenkins, B. G., Dale, A. M., and Fischl, B. "**Regional and Progressive Thinning of the Cortical Ribbon in Huntington's Disease**" *Neurology* 58, no. 5 (2002): 695–701.

60. Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., Goff, D., West, W. C., Williams, S. C. R., Kouwe, A. J. W. van der, Salat, D. H., Dale, A. M., and Fischl, B. "**Regionally Localized Thinning of the Cerebral Cortex in Schizophrenia**" *Arch Gen Psychiatry* 60, no. 9 (2003): 878–88. doi:10.1001/archpsyc.60.9.878

61. Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Nobili, F., Hensch, T., Tränkner, A., Schönknecht, P., Leroy, M., Lopes, R., Bordet, R., Chanoine, V., Ranjeva, J.-P., Didic, M., Gros-Dagnac, H., Payoux, P., Zoccatelli, G., Alessandrini, F., Beltramello, A., Bargalló, N., Blin, O., Frisoni, G. B., and The Pharma-Cog Consortium. "**Brain Morphometry Reproducibility in Multi-Center 3T MRI Studies:**

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A Comparison of Cross-Sectional and Longitudinal Segmentations" *Neuroimage* (2013): doi:10.1016/j.neuroimage.2013.05.007

62. Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., and Keshavan, A. "**Mindboggling Morphometry of Human Brains**" *PLoS Comput Biol* 13, no. 2 (2017): e1005350. doi:10.1371/journal.pcbi.1005350

63. Verbeke, G. and Molenberghs, G. "**Linear Mixed Models for Longitudinal Data**" (2009):

64. Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. "**Applied Longitudinal Analysis**" 998, (2012):

65. 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

66. 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.

67. 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):

68. 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.

69. Braak, H. and Braak, E. "**Neuropathological Stageing of Alzheimer-Related Changes**" *Acta Neuropathol* 82, no. 4 (1991): 239–59.

70. 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

71. Yassa, M. A. "**Ground Zero in Alzheimer's Disease**" *Nat Neurosci* 17, no. 2 (2014): 146–7. doi:10.1038/nn.3631

72. Holland, D., Brewer, J. B., Hagler, D. J., Fennema-Notestine, C., Fenema-Notestine, C., Dale, A.

bioRxiv preprint doi: https://doi.org/10.1101/170209; this version posted July 30, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

73. 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

74. 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

75. Price, C. C., Wood, M. F., Leonard, C. M., Towler, S., Ward, J., Montijo, H., Kellison, I., Bowers, D., Monk, T., Newcomer, J. C., and Schmalfuss, I. "**Entorhinal Cortex Volume in Older Adults: Reliability and Validity Considerations for Three Published Measurement Protocols**" *J Int Neuropsychol Soc* 16, no. 5 (2010): 846–55. doi:10.1017/S135561771000072X

76. Fischl, B., Stevens, A. A., Rajendran, N., Yeo, B. T. T., Greve, D. N., Van Leemput, K., Polimeni, J. R., Kakunoori, S., Buckner, R. L., Pacheco, J., Salat, D. H., Melcher, J., Frosch, M. P., Hyman, B. T., Grant, P. E., Rosen, B. R., Kouwe, A. J. W. van der, Wiggins, G. C., Wald, L. L., and Augustinack, J. C. "**Predicting the Location of Entorhinal Cortex from MRI**" *Neuroimage* 47, no. 1 (2009): 8–17. doi:10.1016/j.neuroimage.2009.04.033

77. Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., Kouwe, A. J. W. van der, Wald, L. L., Van Leemput, K., McKee, A. C., Fischl, B., and Alzheimer's Disease Neuroimaging Initiative. "**Predicting the Location of Human Perirhinal Cortex, Brodmann's Area 35, from MRI**" *Neuroimage* 64, (2013): 32–42. doi:10.1016/j.neuroimage.2012.08.071