

Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements

Nicholas J. Tustison^{a,1}, Arno Klein^c, Philip A. Cook^b, Gang Song^b, Sandhitsu R. Das^b, Jeffrey T. Duda^b, Benjamin M. Kandel^b, Niels van Strien^c, James R. Stone^a, James C. Gee^b, Brian B. Avants^b

^aDepartment of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

^bPenn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA

^cSage Bionetworks, Seattle, WA

^dDepartment of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

Numerous studies have explored the relationship between cortical structure, brain development, cognitive function, and functional connectivity. The highly convoluted cortical topography makes manual measurements arduous and often impractical given the population sizes necessary for inferring statistical trends. Computational techniques have permitted large-scale studies as they provide localized measurements characterizing the cortex with little or no human intervention. Useful to the neuroscience community are publicly available tools, such as the popular surface-based FreeSurfer, which facilitate the testing and refinement of hypotheses. Further motivating the adoption of such tools are the availability of robust parameter sets which have been tuned to provide good performance during the execution of the different pipeline components. For these reasons we developed the volume-based Advanced Normalization Tools (ANTs) cortical thickness automated pipeline comprising well-vetted components such as SyGN (multivariate template construction), SyN (image registration), N4 (bias correction), Atropos (*n*-tissue segmentation), and DiReCT (cortical thickness). Although good repeatability is seen with both frameworks, such assessments of precision do not properly extend to inferences of accuracy or statistical modeling capabilities which is crucial. For evaluative purposes, four open data sets (IXI, Kirby, NKI, and Oasis), consisting of approximately 1200 images, were processed by both the ANTs and FreeSurfer pipelines using standard processing protocols including the recently proposed “Desikan-Killiany-Tourville” (DKT) atlas. Measures of repeatability are augmented with straightforward demographic-based measures that illustrate strong predictive performance with ANTs over FreeSurfer. In addition to using open image data sets, to further promote open science, scientific reproducibility, and the use of the proposed ANTs pipeline, all scripts and results have been made publicly available.

Keywords: advanced normalization tools, age prediction, cortical thickness, gender prediction, open science

1. Introduction

Imaging-based structural analysis of the brain plays a fundamental role in identifying the relationship between cortical morphology, disease, and cognition. Discriminative quantitative cortical measures have been demonstrated in conditional abnormalities such as Huntington’s disease [72, 71, 76], schizophrenia [65], bipolar disorder [58], Alzheimer’s disease and frontotemporal dementia [23, 21], Parkinson’s disease [44], Williams syndrome [86], multiple sclerosis [70], autism [13, 38], migraines [18], chronic smoking [50], alcoholism [30], cocaine addiction [61], Tourette syndrome in children [84], scoliosis in female adolescents [90], early-onset blindness [41], chronic pancreatitis [33], obsessive-compulsive disorder [79], ADHD [1], obesity [69], and heritable [67] and elderly [8] depression. Evidence of cortical thickness variation has also been found to be a function of age [48], gender [55], untreated male-to-female transsexuality [56], handedness [54, 2], intelli-

gence [78], athletic ability [93], meditative practices [52], musical ability [9, 31], tendency toward criminality [68], childhood sexual abuse in adult females [40], and Tetris-playing ability in female adolescents [36]. Additionally, recent studies demonstrate structural connectivity relationships using cortical thickness measures [96, 53, 39, 12]. Although these findings are subject to debate and interpretation [34], the availability of quantitative computational methods for extracting such information has proven invaluable for developing and refining fundamental neuroscience hypotheses.

Computational methods for analyzing the cortex may be broadly characterized as surface mesh-based or volumetric [74, 14]. Representative of the former is the FreeSurfer² cortical modeling software package [16, 28, 26, 27, 29] which owes its popularity to public availability, excellent documentation, good performance, and integration with other toolkits, such as the extensive FMRIB software library (FSL) [83]. Similar to other surface approaches (e.g. [19, 60, 59, 45]), the pial and white matter surfaces from individual subject MR data are modeled with polygonal meshes which are then used to determine

¹Corresponding author: PO Box 801339, Charlottesville, VA 22908; T: 434-924-7730; email address: ntustison@virginia.edu.

Partial support provided by US Army Medical Research and Materiel Command; Contract grant number: W81XWH-09-2-0055.

²<http://surfer.nmr.mgh.harvard.edu/>

local cortical thickness values based on a specified correspondence between the surface models.

Image volumetric (or meshless) techniques vary both in algorithmic terms as well as the underlying definition of cortical thickness. An early, foundational technique is the method of [42] in which the inner and outer surface geometry is used to determine the solution to Laplace's equation where thickness is measured by integrating along the tangents of the resulting field lines spanning the boundary surfaces. Subsequent contributions improved upon the original formulation. For example, in [98], an Eulerian PDE approach was proposed to facilitate the computation of correspondence paths. Extending the surface-based work of [59], the hybrid approach of [45] uses the discrete Laplacian field to deform the white matter surface mesh towards the pial surface. Although the Laplacian-based approach has several advantages including generally lower computational times and non-crossing correspondence paths, direct correlative assessments with histology are potentially problematic as the quantified distances are not necessarily Euclidean. Other volumetric algorithms employ coupled level sets [99], model-free intelligent search strategies either normal to the gray-white matter interface [74], or using a min-max rule [89]. Most relevant to this work is the DiReCT (Diffeomorphic Registration-based Cortical Thickness) algorithm proposed in [17] where generated diffeomorphic mappings between the white and pial matter surfaces are used to propagate thickness values through the cortical gray matter. A unique benefit of DiReCT is that it naturally estimates the boundaries of buried sulci by employing a diffeomorphic constraint on the probabilistic estimate of the gray matter and cerebrospinal fluid interface.

Although a variety of techniques exist for estimating cortical thickness from imaging data (of which only a fraction are cited), several common preprocessing components can be identified. The most fundamental of these include inhomogeneity correction, skull stripping, and n -tissue segmentation for differentiating the gray and white matters. For statistical analysis across large populations, construction of population-specific unbiased templates is also potentially beneficial [25]. In addition, intermediate steps might include a crucial registration component (e.g. propagating template-based tissue priors for improved segmentation).

Also increasing the complexity of cortical thickness studies is the need for large neuroimaging datasets such as that provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the increasing importance of fully automated and multiple modality brain mapping tools [94]. In the case of ADNI, the scale of such datasets is projected to increase over time as international projects join the effort to build large-scale AD-related neuroimaging resources. Currently, the National Institutes of Health (NIH) also mandates that any NIH-funded data resources, including MRI, must be released to the public. In contrast to ADNI, which provides standardized data acquisition protocols used across all sites, these smaller-scale projects are collected in an unstructured way. Therefore, neuroimage processing tools must reliably quantify even when there is a relative lack of quality control over the input data. While robustness is a goal shared by all software development targeted at

neuroscience research, very few methods have been thoroughly tested on large and unstructured neuroimaging datasets.

The general lack of availability of published algorithms [49] (not to mention critical preprocessing components) inhibits performing studies by external researchers and makes comparative evaluations difficult. For example, one recent evaluation study [14] compared FreeSurfer (a surface-based method) with two volumetric methods [42, 17]. Whereas the entire FreeSurfer processing pipeline has been made publicly available, refined by the original authors and other contributors, and described in great detail (specifically in terms of suggested parameters); both volumetric methods were implemented solely by the authors of the evaluation (not the actual algorithm developers) using unspecified parameters with relatively small, private data sets making the comparisons less than ideal (see [88] for further discussion concerning the issue of instrumentation bias and scientific reproducibility in the use and evaluation of software). Further complicating comparisons are distinct processing domains between volumetric and surface-based techniques and the potential for the introduction of bias [46].

A brief description of our proposed pipeline, which produces a volumetric cortical thickness map from an individual subject's T1-weighted MRI, is provided. Additionally, we note that it is freely available as part of the Advanced Normalization Tools (ANTs) software package. This includes all the necessary preprocessing steps consisting of well-vetted, previously published algorithms for bias correction [87], brain extraction [4], n -tissue segmentation [6], template construction [7], and image normalization [5]. More important, however, we provide explicit coordination between these components within a set of well-documented shell scripts which are also available in the ANTs repository where parameters have been tuned by ANTs developers (N.T. and B.A.). A fully functional 2-D example is provided at the github repository associated with this work.³

In this work, we demonstrate the use of the described framework in processing a large cohort of publicly available T1-weighted brain MRI data drawn from four well-known data sets totaling approximately 1200 images. For comparative evaluation we also process the same data using the standard FreeSurfer cortical thickness processing protocol. Using data subsets of repeated acquisitions, repeatability measures for both frameworks are reported as in previous work [14, 43]. However, as repeatability (or, more generally, *precision*) is not conceptually equivalent to *accuracy*, and given that ground-truth measurements are not available for these data, we use demographic-based predictive assessments to show that ANTs outperforms FreeSurfer-based thickness estimation for these data in well-studied cortical thickness and age/gender relationships.

Finally, we make available all data from both ANTs and FreeSurfer processing outcomes. This includes derived image data, processing scripts, and tabulated results. The availability of both the code and data permits the set of results described in this work to be fully reproducible. This permits other researchers to contrast their own results against this baseline pro-

³<http://www.github.com/KapowskiChronicles/>

cessing and to adapt the given volumetric pipeline for measuring cortical thickness with their own datasets.

2. Methods and Materials

2.1. ANTs volumetric-based cortical thickness estimation pipeline

The ANTs-based cortical thickness estimation workflow is illustrated in Figure 1. The steps are as follows:

1. initial N4 bias correction on input anatomical MRI,
2. brain extraction using a hybrid segmentation/template-based strategy,
3. alternation between prior-based segmentation and “pure tissue” posterior probability weighted bias correction,
4. DiReCT-based cortical thickness estimation, and
5. optional normalization to specified template and multi-atlas cortical parcellation.

Each component, including both software and data, is briefly detailed below with the relevant references for additional information.

The coordination of all the algorithmic components is encapsulated in the shell script `antsCorticalThickness.sh` with subcomponents delegated to `antsBrainExtraction.sh` and `antsAtroposN4.sh`. A representative script command is reproduced in Listing 1 for a single IXI subject to demonstrate the simplicity and mature status of what we propose in this work and a comparison with the analogous FreeSurfer command. Option descriptions are provided by invoking the help option, i.e. `'antsCorticalThickness.sh -h'`.

```
# Processing calls for subject IXI002-Guys-0828-T1

# ANTs
antsCorticalThickness.sh \
-a IXI/T1/IXI002-Guys-0828-T1.nii.gz \
-e IXI/template/T_template0.nii.gz \
-m IXI/template/T_template0ProbabilityMask.nii.gz \
-p IXI/template/Priors/priors%d.nii.gz \
-f IXI/template/T_template0ExtractionMask.nii.gz \
-o IXI/ANTSResults/IXI002-Guys-0828

# FreeSurfer
recon-all \
-i IXI/T1/IXI002-Guys-0828-T1.nii.gz \
-s IXI002-Guys-0828 \
-sd IXI/FreeSurferResults/ \
-all
```

Listing 1: Analogous ANTs and FreeSurfer command line calls for a single IXI subject which is representative of the calls used for each subject of the evaluation study.

2.1.1. Anatomical template construction

Normalizing images to a standard coordinate system reduces intersubject variability in population studies. Various approaches exist for determining the normalized space such as the selection of a pre-existing template based on a single subject, e.g. the Talairach atlas [85], or a publicly available averaged group of subjects, e.g. the MNI [15] or ICBM [62] templates. Additionally, mean templates constructed from labeled data can be used to construct spatial priors for improving segmentation

algorithms. The work of [7] explicitly models the geometric component of the normalized space during optimization to produce such mean templates. Coupling the intrinsic symmetry of SyN pairwise registration [5] and an optimized shape-based sharpening/averaging of the template appearance, Symmetric Group Normalization (SyGN) is a powerful framework for producing optimal population-specific templates.

The ANTs implementation of this technique is currently available as a shell script, `buildtemplateparallel.sh`. A generalized, multivariate version is also available as `antsMultivariateTemplateConstruction.sh`. Both scripts are distributed as part of the ANTs repository. The multivariate script permits the construction of multimodal templates (e.g. T1-weighted, T2-weighted, proton density MRI and fractional anisotropy). Both scripts accommodate a variety of computational resources for facilitating template construction. These computational resource possibilities include:

- serial processing on a single workstation,
- parallelized processing on a single workstation with multiple cores using pexec⁴,
- parallelized processing using Apple’s XGrid technology⁵,
- parallelized processing using Sun Grid Engine for cluster-based systems⁶, and
- parallelized processing using the Portable Batch System for cluster-based systems⁷.

For this work, database-specific templates were used during cortical thickness pipeline processing for both brain extraction and brain segmentation steps. Multivariate templates were constructed from the multimodal data sets. However, their usage was based on the fact that they had been built previously for other work and not because they provide a discernible advantage over univariate templates (i.e. T1-only) for the proposed workflow.

2.1.2. N4 bias field correction

Critical to quantitative processing of MRI is the minimization of field inhomogeneity effects which produce artificial low frequency intensity variation across the image. Large-scale studies, such as ADNI, employ perhaps the most widely used bias correction algorithm, N3 [81], as part of their standard protocol [10].

In [87] we introduced an improvement of N3, denoted as “N4”, which demonstrates a significant increase in performance and convergence behavior on a variety of data. This improvement is a result of an enhanced fitting routine (which includes

⁴<http://www.gnu.org/software/pexec/pexec.1.html>

⁵<https://developer.apple.com/hardwaredrivers/hpc/xgrid.intro.html>

⁶<http://www.oracle.com/technetwork/oem/grid-engine-166852.html>

⁷<http://www.pbsworks.com/>

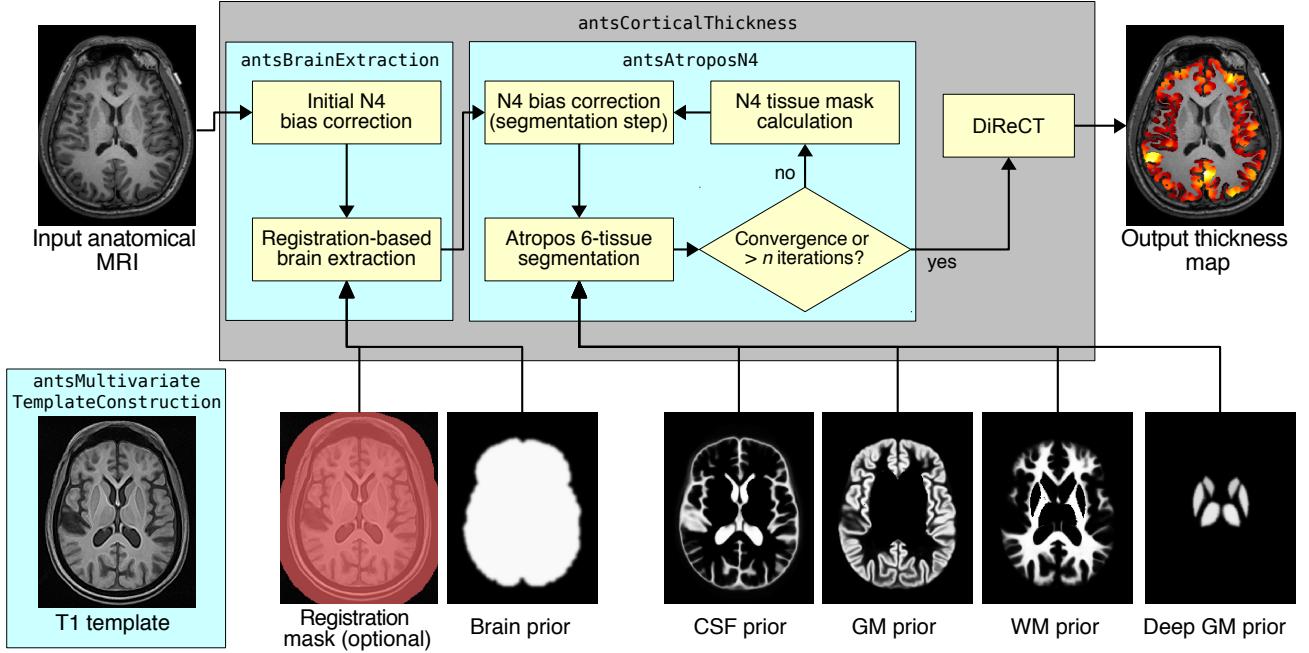


Figure 1: Illustration of the main components of the ANTs processing workflow containing all elements for determining cortical thickness. We also included the domain of operations for the selected scripts. Not shown are the probability maps for the brain stem and cerebellum priors.

multi-resolution capabilities) and a modified optimization formulation. For our workflow, the additional possibility of specifying a weighted mask in N4 permits the use of a “pure tissue” probability map (described below) calculated during the segmentation pipeline for further improvement of bias field estimation.

N4 is used in two places during the individual subject processing (cf Figure 1). Initially, it is used to generate an initial bias corrected image for use in brain extraction. The input mask is created by adaptively thresholding the background from the foreground using Otsu’s algorithm [66]. Following brain extraction, six-tissue (cerebrospinal fluid, cortical gray matter, white matter, deep gray matter, brain stem, and cerebellum) segmentation involves iterating between bias field correction using the current pure tissue probability map as a weight mask and then using that bias corrected image as input to the Atropos segmentation step (described below).

2.1.3. Brain extraction

Brain extraction using ANTs combines template building, high-performance brain image registration, and Atropos with topological refinements. An optimal template [7] is first constructed using labeled brain data (e.g. LPBA40). Template construction iterates between estimating the optimal template and registering each subject to the optimal template. Thus, the construction produces the transforms necessary to warp each subject’s labeling to the template space. We use these transformed labelings to create a probabilistic estimate of the brain mask for the template. In this work, we perform the additional step of building separate templates for each cohort and propagate the probabilistic mask to each cohort template using registration of

the T1-weighted templates (cf Figure 3). Further refinement include thresholding the warped brain probability map at 0.5 and dilated the resulting mask with a radius of two. Atropos is used to generate an initial three-tissue segmentation estimate within the mask region. Each of the three tissue labels undergo specific morphological operations which are then combined to create a brain extraction mask for use in the rest of the cortical thickness workflow using the script `antsBrainExtraction.sh`.

A comparison of an earlier version of our extraction methodology using open access brain data with publicly available brain extraction algorithms including AFNI’s `3dIntracranial` [92], FSL’s `BET2` [82], FreeSurfer’s `mri_watershed` [75], and BrainSuite [22] demonstrated that our combined registration/segmentation approach [4] performed comparable BrainSuite (tuned) and FreeSurfer. A visual comparison of results derived with the current ANTs brain extraction processing with FreeSurfer is seen in Figure 2.

2.1.4. Atropos six-tissue segmentation

In [6] we presented an open source *n*-tissue segmentation software tool (which we denote as “Atropos”) attempting to distill 20+ years of active research in this area particularly some of its most seminal work (e.g. [100, 3]). Specification of prior probabilities includes spatially varying Markov Random Field modeling, prior label maps, and prior probability maps typically derived from our template building process. Additional capabilities include handling of multivariate data, partial volume modeling [77], a memory-minimization mode, label propagation, a plug-n-play architecture for incorporation of novel likelihood models which includes both parametric and non-parametric models for both scalar and tensorial images,

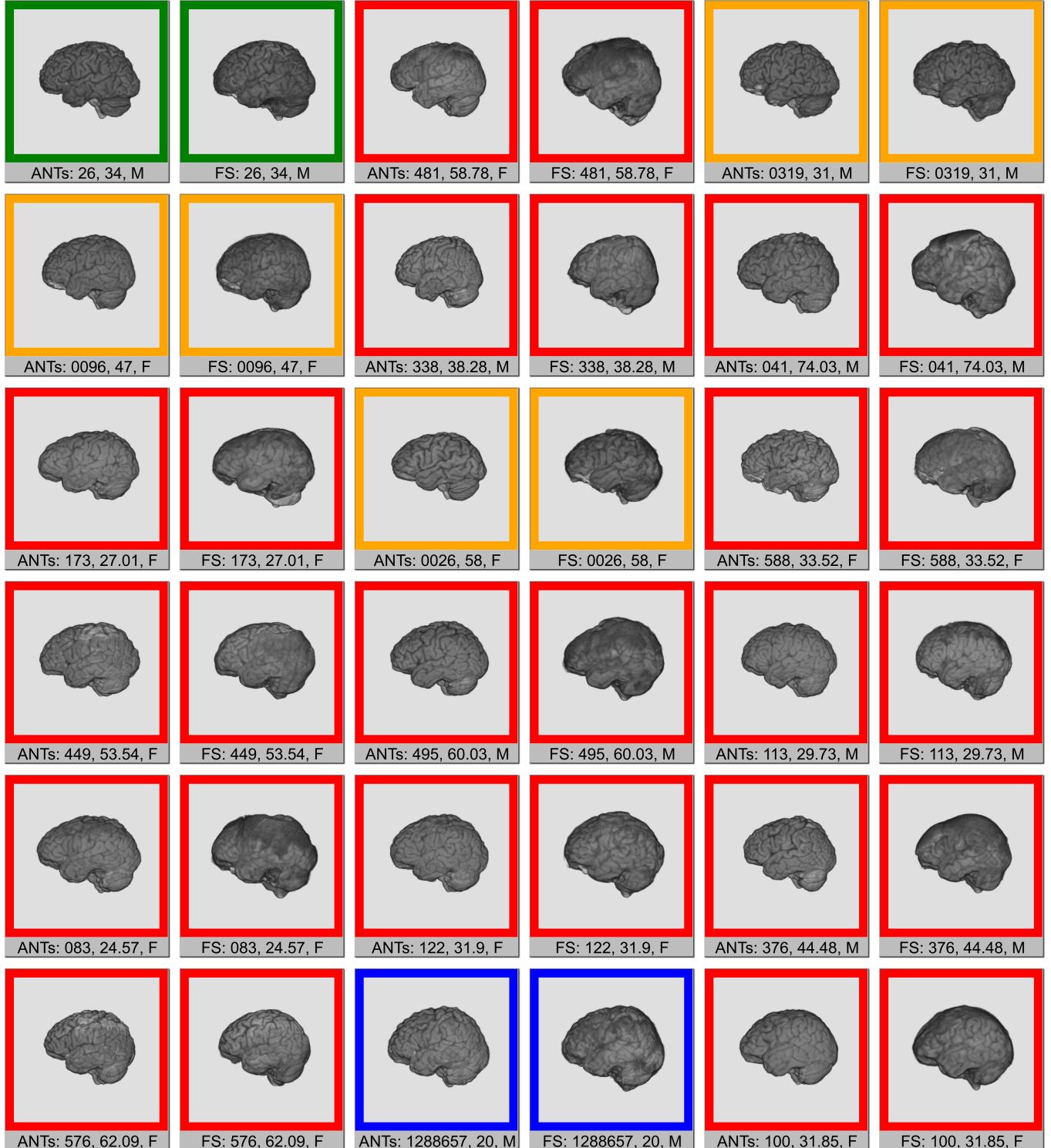


Figure 2: Random sampling of volumetric brain renderings from the four different cohorts (IXI = red, Kirby = green, NKI = blue, Oasis = orange) illustrating the qualitative difference between ANTs and FreeSurfer (“FS”) results which are arranged pairwise. Each brain was rigidly registered to the Oasis template for rendering purposes. Below each image we provide subject ID, age, and gender.

and alternative posterior formulations for different segmentation tasks.

Due to the important interplay between segmentation and bias correction, we perform multiple N4 \rightleftharpoons Atropos iterations. In order to better integrate Atropos and N4, we use a pure tissue probability weight mask generated from the posterior probabilities derived from the segmentation step. Given N labels and the corresponding N posterior probability maps $\{P_1, \dots, P_N\}$ produced during segmentation, the N4 weight mask is created at each N4 \rightleftharpoons Atropos iteration from

$$P_{\text{pure tissue}}(\mathbf{x}) = \sum_{i=1}^N P_i(\mathbf{x}) \prod_{j=1, j \neq i}^N (1 - P_j(\mathbf{x})). \quad (1)$$

One of the key insights of the original N3 development is the observation that inhomogeneities cause the intensity values of pure tissue peaks to spread in the intensity histogram as though convolved with a Gaussian. A core contribution of N3 is the proposed corrective of deconvolving the intensity histogram to accentuate the tissue peaks coupled with a spatial smoothing constraint. The pure tissue probability mask weights more heavily the voxels corresponding to pure tissue types (as determined by the segmentation) during the deconvolution process while minimizing the contribution of regions such as the gray/white matter interface where peak membership is ambiguous.

Atropos enables prior knowledge to guide the segmentation process where template-based priors are integrated into the optimization with a user-controlled weight. Modulating the likelihood and prior contributions to the posterior probability is essential for producing adequate segmentations. Atropos weights the likelihood and priors according to $P(x|y) \propto P(y|x)^{1-\alpha} P(x)^\alpha$ where α is a user-selected parameter which weights the tradeoff between the likelihood and priors terms. In this work, we chose a weighting of $\alpha = 0.25$ based on our extensive experimentation with different parameter weights.

Since cortical thickness estimation only requires the cortical gray and white matters, the deep gray and white matters (both labelings and posterior maps) are combined to a single “white matter” set. This new set combined with the cortical gray matter results are the only results from the segmentation step used in the DiReCT algorithm (described below).

2.1.5. DiReCT (aka KellySlater/KellyKapowski) cortical thickness estimation

DiReCT was introduced in [17] and made available in ANTs as the program *KellySlater*. Since then several improvements have been made and incorporated into the program *KellyKapowski*.⁸ Among the most significant advancements is that the more recent implementation is multi-threaded, written in rigorous ITK coding style, and has been made publicly

1) caudal anterior cingulate	2) caudal middle frontal
3) cuneus	4) entorhinal
5) fusiform	6) inferior parietal
7) inferior temporal	8) isthmus cingulate
9) lateral occipital	10) lateral orbitofrontal
11) lingual	12) medial orbitofrontal
13) middle temporal	14) parahippocampal
15) paracentral	16) pars opercularis
17) pars orbitalis	18) pars triangularis
19) pericalcarine	20) postcentral
21) posterior cingulate	22) precentral
23) precuneus	24) posterior anterior cingulate
25) rostral middle frontal	26) superior frontal
27) superior parietal	28) superior temporal
29) supramarginal	30) transverse temporal
31) insula	

Table 1: The 31 cortical labels (per hemisphere) of the DKT atlas.

available through ANTs complete with a unique user interface design developed specifically for ANTs tools.

2.2. Public data resources

The four public data sets which were processed using the previously described pipeline are IXI, Kirby, NKI, and Oasis. In addition, we used a subset of the MindBoggle-101 data⁹ labeled using the Desikan-Killiany-Tourville (DKT) protocol [47] to define the regions of interest (ROI). All five data sets are described below.

2.2.1. MindBoggle-101 data for ROI definitions

In [47] the authors proposed the DKT protocol—a modification of the popular Desikan protocol [20] for purposes of labeling consistency and facilitating FreeSurfer cortical classification of regions (which are given in Table 1). Since the regional thickness values defined by the DKT atlas are calculated as part of the current FreeSurfer processing protocol, this atlas provides the common measurement standard for comparison between ANTs and FreeSurfer.

The work of [47] also resulted in a publicly available set of 101 DKT-labeled T1-weighted brain images from different sources including a subset of 20 images from the Oasis data set. These 20 images are used to define the volumetric regions for each subject during the ANTs processing using multi-atlas label propagation [91]. Parameters used are found in the script `antsMultiLabeling.sh` which is also distributed with the ANTs toolkit. Cortical thickness values are averaged within each label for each subject using only the non-zero voxels from the cortical thickness map.

2.2.2. Public data for thickness estimation evaluation

We apply the same pipeline to diverse publicly available datasets collected from multiple sites and with a mixture of 3T

⁸Traditional academic discourse encountered in the published literature rarely contextualizes peculiarities such as algorithmic nomenclature. We briefly mention that this was the source of a rare disagreement between the first and last authors based, as many disagreements are, on a simple misunderstanding and not an affronting existential statement concerning a certain favorite sitcom of the first author’s youth.

⁹<http://mindboggle.info/data.html>

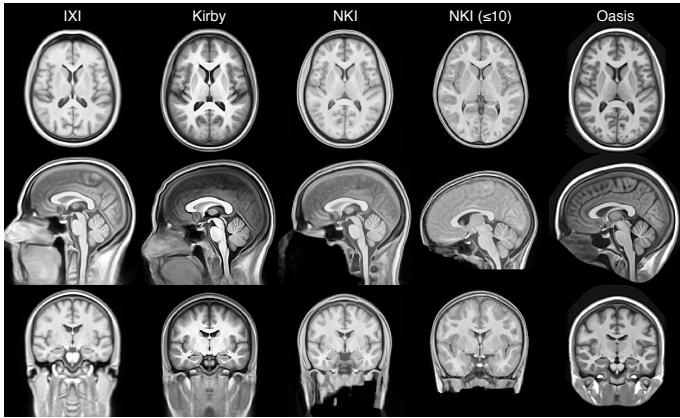


Figure 3: Population-specific templates for each of the four public data sets used for cortical thickness estimation generated using the script `antsMultivariateTemplateConstruction.sh`. The benefit of using such templates is obvious considering the variability in acquisition and data preparation (e.g. defacing protocols).

and 1.5T T1 brain images. Subjects in this dataset span the age range from 4 to 96 years old. This strategy tests robustness to variation in head position, brain shape, defacing, image contrast, inhomogeneity, imaging artifacts, and the broad variation in extracerebral tissue. Failure can occur in initial brain extraction, segmentation, registration, or bias correction, any of which will lead to an inaccurate cortical thickness measurement.

In total, we processed 1205 T1-weighted images from four different public data sets to obtain the corresponding cortical thickness maps. The descriptions of the four data sets are as follows:

IXI. Initially, we began with 581 T1-weighted images from the IXI¹⁰ data set of which all were processed but only 563 subjects (313 females, 250 males) were included in the post processing analysis due to missing demographic information preventing an accurate estimate of the age at the time of image acquisition. These data were imaged at three sites with several modalities acquired (T1-weighted, T2-weighted, proton density, magnetic resonance angiography, and diffusion tensor imaging). The database also consists of demographic information such as date of birth, date of scan, weight, height, ethnicity, occupation category, educational level, and marital status.

Kirby. The Multi-Modal MRI Reproducibility Resource¹¹, or more informally, the Kirby data set, was originally described in [51] consisting of 21 subjects (10 females, 11 males) and features a rich multiple modality and repeated acquisition schedule.

NKI. In support of open science, the 1,000 Functional Connectomes Project¹² was initiated on December 11, 2009 by various members of the MRI community seeking to form collaborative partnerships with imaging institutions for sharing well-

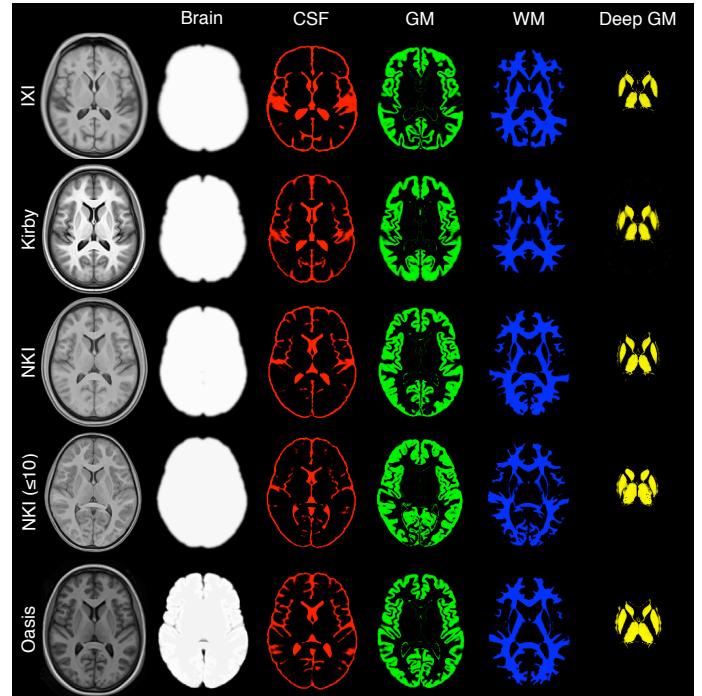


Figure 4: Axial slices from each of the five T1 templates including the corresponding probability masks used for brain extraction and brain tissue segmentation. Not shown are the prior probability maps for brain stem and cerebellum regions.

documented multimodal image sets accompanied by phenotypic data. One such contribution is the Nathan Klein Institute (NKI)/Rockland sample consisting of 186 T1-weighted images (87 females, 99 males).¹³

Oasis. The initial Open Access Series of Imaging Studies (OASIS)¹⁴ data set consisted of 433 T1-weighted images. All were processed although 100 subjects were excluded from analysis due to probable Alzheimer's disease ($CDR > 0$) and 20 subjects had repeat scans for 313 individual subjects included in the normal group statistical analysis (118 males, 195 females). Ages were between 18 and 96 and all subjects are right-handed.

2.3. Processing miscellany

Given the documented variability in FreeSurfer results with version number and Mac OS [35] (as we would expect with our own ANTs pipeline), all data was processed using the same ANTs and FreeSurfer versions on the same hardware platform. Processing was performed using the Linux (CentOS release 6.4) cluster at the University of Virginia¹⁵ using single-threading with a maximal requested memory footprint of 8 gb for ANTs and 4 gb for FreeSurfer. The development version of ANTs was used for processing (git commit tag: 69d3a5a6c7125ccf07a9e9cf6ef29f0b91e9514f, date

¹⁰<http://biomedic.doc.ic.ac.uk/brain-development/>

¹¹<http://www.nitrc.org/projects/multimodal/>

¹²http://fcon_1000.projects.nitrc.org

¹³Downloaded on September 22, 2012.

¹⁴<http://www.oasis-brains.org/>

¹⁵<http://www.uvacse.virginia.edu/>

Dec. 11, 2013). FreeSurfer version 5.3 x86_64 for CentOS was downloaded on 5 December, 2013 (“freesurfer-Linux-centos6_x86_64-stable-pub-v5.3.0”, release date: 15 May, 2013).

3. Evaluation

Traditional assessment approaches, such as manual labeling, are inadequate for evaluating large-scale performance. We therefore sought to minimize failure rate, quantify the repeatability of cortical thickness measures, and determine whether the ANTs pipeline reveals biologically plausible relationships between the cortex, gender,¹⁶ and age and how performance compares to the current de facto standard of FreeSurfer-derived thickness estimation. Collectively, these surrogate measurements allow us to establish data-derived relative performance standards. Additionally, for completeness, we include timing results as that factors into usability.

3.1. Repeatability

Repeat scans of 40 subjects (20 Kirby subjects and 20 Oasis subjects) were used to determine the repeatability of regional cortical thickness measurements, T . Similar to the assessment given in [43], we demonstrate this in terms of the percent variability error:

$$\varepsilon = \frac{|T_{\text{scan}} - T_{\text{rescan}}|}{0.5 \times (T_{\text{scan}} + T_{\text{rescan}})}. \quad (2)$$

Comparison of the ANTs and FreeSurfer percent variability errors for the 62 DKT regions for both the Oasis and Kirby reproducibility data sets are given in Figure 5. Although the variance is slightly greater for the set of ANTs measurements, statistical testing per cortical region (two-tailed paired t-test, corrected using false discovery rate) did not indicate non-zero mean differences for either approach for any region.

We also calculated the intraclass correlation coefficient (“ICC(2,1)” in the notation of [80]) to assess scan/rescan reliability. The ANTs thickness pipeline produced an ICC value of 0.98 whereas the FreeSurfer thickness pipeline yielded an ICC value of 0.97 indicating good scan/rescan reliability for both ANTs and FreeSurfer.

3.2. Age prediction assessment

Despite good repeatability with both ANTs and FreeSurfer, such measures do not provide an assessment of accuracy or even relative utility. For example, strong priors can yield good repeatability measures but potentially at the expense of data fidelity thus compromising the quality of models (statistical or otherwise) built from such results. Given that ground truth is not available for these data nor for the many studies looking

at brain morphology, an indirect method (or set of methods) is required for determining the quality of thickness estimation.

Previous research has used predictive modeling for comparing cortical thickness algorithms. For example, in [14], classification of healthy, semantic dementia, and progressive non-fluent aphasia categories using regional cortical thickness values was used to determine the predictive modeling capabilities of different cortical thickness processing protocols in 101 subjects. However, differential diagnosis of dementia [64] is not as straightforward as obtaining a subject’s age or gender and regressing that against cortical thickness—biological relationships which have been well-studied and reported in the literature.

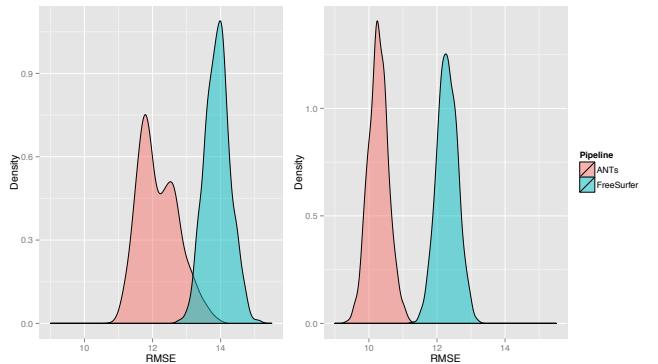


Figure 6: Age prediction RMSE distributions of linear (left) and random forest (right) models for the ANTs- and FreeSurfer-derived thickness values. For both prediction models ANTs RMSE error is lower.

For our first assessment, we modeled age versus regional cortical thickness values to determine which framework produces better predictive thickness estimates. We first subdivided the thickness data into training and testing subsets with an even split between the two subsets.¹⁷ We then used the training data to create two models for each pipeline: 1) standard linear regression and 2) random forests (a non-parametric machine learning technique) [11], for estimating age from both ANTs and FreeSurfer thickness values in the testing data.

The formula (in the notation of [95]) for the linear model is

$$AGE \sim VOLUME + \sum_{i=1}^{62} T(DKT_i) * GENDER \quad (3)$$

where $T(DKT_i)$ is the average thickness value in region DKT_i . Similarly, the random forest model was specified as a combination of all terms using the `randomForest`¹⁸ package in R with the default settings and 200 trees.

¹⁷We tried various training proportions between 10 and 90% (in increments of 10%) to see if that had an effect on relative performance for both age and gender prediction comparisons. Although age predictive capabilities for both pipelines showed improvement (gender prediction was mostly unaffected), the relative outcomes were the same. One interesting result is that predictive performance degenerated at the 10% level which translates into approximately 100 subjects being used for training raising concerns about the use of smaller cohorts for performance comparisons.

¹⁸<http://cran.r-project.org/package=randomForest>

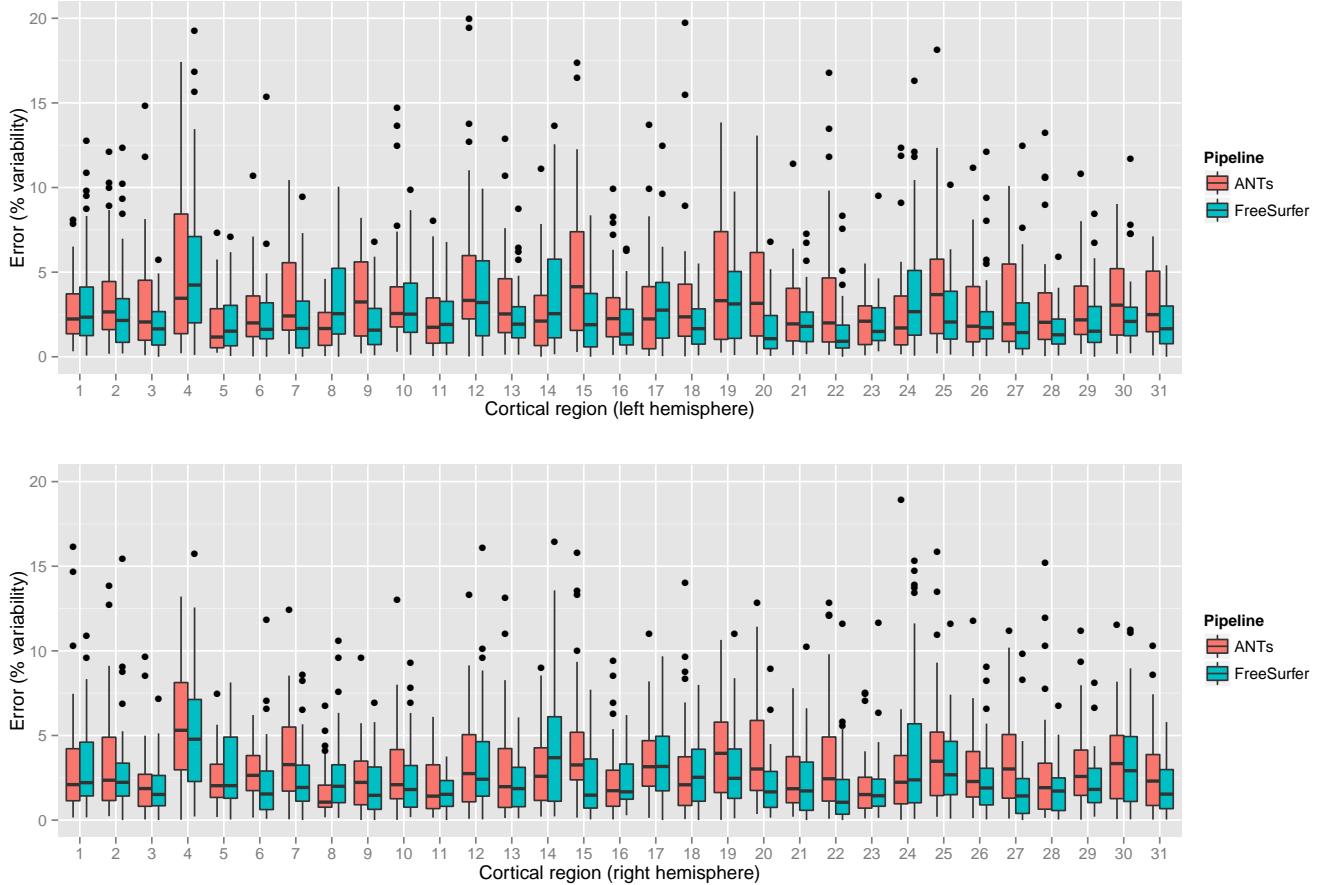


Figure 5: Percent error variability for both ANTs and FreeSurfer pipelines over the left and right hemispheres of both the Kirby and Oasis data subsets within the 62 regions defined by the Desikan-Killiany-Tourville atlas. Both methods demonstrate good repeatability qualities.

Table 2: Mean RMSE for age prediction (in years).

	Linear	Random Forest
ANTs	12.2	10.3
FreeSurfer	13.9	12.3

In order to ensure a fair comparison, the procedure described above consisting of training and testing steps was performed for $n = 1000$ permutations to elicit a performance distribution which we measure using the relative mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum (AGE_{true} - AGE_{predicted})^2}{N}}. \quad (4)$$

The resulting distributions are illustrated in Figure 6 with the linear model results displayed on the left and random forest results on the right. The RMSE value was lower with ANTs thickness values for both models with the ANTs-based random forest predictions performing the best. Mean RMSE values are provided in Table 2.

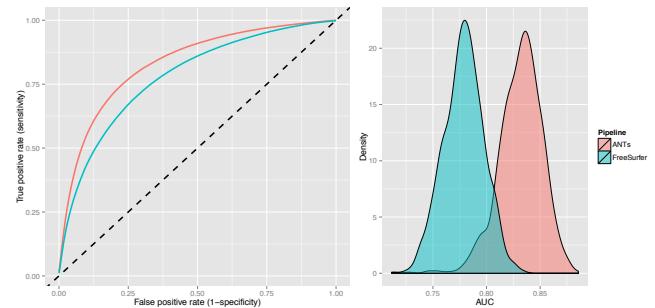


Figure 7: Average ROC curve and corresponding AUC distributions for gender prediction using ANTs and FreeSurfer thickness values. Values were averaged for 1000 permutations resulting in mean values of $ANTs_{AUC} = 0.83$ and $FreeSurfer_{AUC} = 0.78$.

3.3. Gender prediction assessment

We also perform a similar prediction assessment using gender as the regressand. The binomial generalized linear model is

$$GENDER \sim VOLUME + \sum_{i=1}^{62} T(DKT_i) * AGE \quad (5)$$

where $T(DKT_i)$ is the average thickness value in region DKT_i . We then characterized performance using a ROC curve for both methods (see Figure 7) where we averaged over 1000 permutations. The mean area under the curve (AUC) for both methods was also quantified with values of $\text{ANTs}_{AUC} = 0.83$ and $\text{FreeSurfer}_{AUC} = 0.78$.

3.4. Computation time

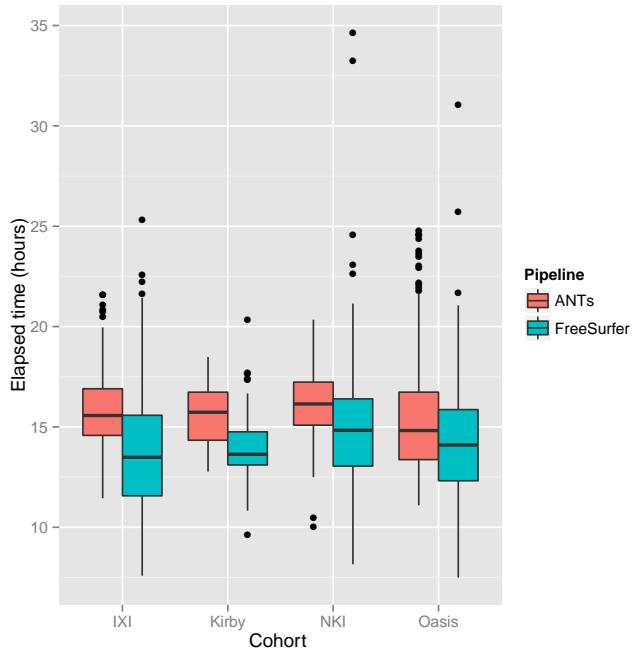


Figure 8: Elapsed time across data sets for ANTs and FreeSurfer processing.

All images underwent the ANTs and FreeSurfer pipeline processing using the computational cluster at the University of Virginia. Processing times varied approximately between 10–20 hours per subject for both pipelines for the entire cortical thickness estimation procedure although ANTs processing, on average, took slightly longer (cf Figure 8).

The propagation of the DKT labels to each subject using label fusion as described earlier was performed in parallel and took anywhere between 40 and 80 hours per subject for 16 serial image registrations and application of the joint label fusion algorithm [91]. Note that the script mentioned earlier `antsMafLabeling.sh` parallelizes the registration component which decreases the time for parallel computation platforms.

4. Discussion

In the absence of ground truth, we used repeatability and prediction of demographic variables to compare the ANTs and FreeSurfer cortical thickness pipelines as it applies to large-scale data. One very important issue which was not discussed in this work is quality control for ensuring proper pipeline processing. The time required to go through approximately

1200 sets of results ($\times 2$ for both pipelines) is enormous (not to mention the tedium). However, the first author did do this for the brain extraction step to ensure that expected intermediate results were being achieved for both pipelines. The only major failure for both pipelines was a FreeSurfer brain extraction of an IXI subject (IXI430-IOP-0990) shown in Figure ???. Also, we mention that three NKI subjects were not processed to completion with FreeSurfer (1713515, 18755434, and 2674565). These data were not included with the thickness compilation. Although researchers might quibble over processing minutiae such as the inclusion of too much (or not enough) of the meninges, we approached our evaluation using more objective criteria which concern all those engaged in this type of research. We are currently trying to develop methods to facilitate data inspection for quick quality assurance/quality control.

4.1. Repeatability of thickness measurements

The OASIS dataset and the Kirby dataset allow us to test whether the same thickness values emerge from T1-weighted neuroimages collected on the same subject but at different times of the day or over a time separation within a few weeks. Given confounds such as short-term alterations due to T1-weighted susceptibility to blood flow [32, 73, 97] and longitudinal variation in scanner conditions, this strategy is not ideal. However, related tools have looked at this question. An independent evaluation of the FreeSurfer pipeline shows good repeatability measurements [43]. The authors report FreeSurfer reproducibility in the range of 1.5 - 5% depending on the site and region of the brain. A similar study found good repeatability properties although the segmentation accuracy was worse than other popular segmentation methods [24]. The CLADA pipeline showed the ability to detect changes as small as 1 millimeter and showed good agreement with FreeSurfer [63].

Very recently, it was suggested that 3T MRI consistently overestimates cortical thickness [57]. Repeatability of thickness estimates in that study were in the range of 0.2 mm although the study design differs substantially from that used here. In summary, our results (though computed with a different cortical parcellation) are competitive. Finally, some users may choose to segment and register with ANTs and subsequently employ any alternative (e.g. surface-based) method for thickness estimation. Further work is needed by independent authors working on established pipelines (as in [57, 43]) in order to better compare surface-based and volume-based thickness reliability across different populations and age ranges.

4.2. Age and gender prediction

Although repeatability between ANTs and FreeSurfer is comparable, such measures are not as useful in determining the utility of the measuring software. That is the reason we use a training and testing paradigm to evaluate how well both frameworks produce measurements capable of predicting demographics which are well-known to correlate with cortical thickness. Additionally, these demographic measures are probably some of the easiest and most reliably obtained of all possible demographic measures used for this type of assessment. For

age prediction, we used both a linear model (due to its general ubiquity) and a random forest model (a non-parametric model to contrast with the linear approach) which showed overall good performance. Also, the linear and random forest models have the advantage of being interpretable. That is, the models reveal the specific predictors that are most valuable which will be explored in future work.

4.3. Computation time

Computation time for the registration and segmentation components of the ANTs pipeline are substantial. It is likely that nearly as reliable results can be obtained in much less time for many of the subjects in this study. However, our interest in maximizing reliability and quality led us to employ parameters in the registration, segmentation, and bias correction that are as robust as possible to differences in head position, the presence of large deformation between template and target brains and substantial inhomogeneity or artifact within the image content itself. Several subjects (e.g. NKI: 1898228, 1875434) provide examples of more difficult data from which we are able to extract meaningful segmentations and registrations despite the presence of a “garbage-in/garbage-out” problem. A subject of future study is determining an exact cut-off for the inclusion of such data. We do not investigate this issue here, which has concerned statisticians for over half a century [37].

5. Conclusions

Imaging biomarkers such as cortical thickness play an important role in neuroscience research. Extremely useful to researchers are robust software tools for generating such biomarkers. In this work we detailed our open source offering for estimating cortical thickness directly from T1 images and demonstrated its utility on a large collection of public brain data from multiple databases acquired at multiple sites. To our knowledge this study constitutes the largest collection of cortical thickness data processed in a single study. We expect that public availability of our tools and extensive tuning on the specified cohorts will prove useful to the larger research community. In this work, we only explored a portion of the potentially interesting investigations possible with these data. However, since all these data are publicly available, further work can be easily pursued by us or even other interested groups.

References

- [1] Almeida Montes, L. G., Prado Alcántara, H., Martínez García, R. B., De La Torre, L. B., Avila Acosta, D., Duarte, M. G., Mar 2012. Brain cortical thickness in ADHD: Age, sex, and clinical correlations. *J Atten Disord*.
- [2] Amunts, K., Armstrong, E., Malikovic, A., Hömke, L., Mohlberg, H., Schleicher, A., Zilles, K., Feb 2007. Gender-specific left-right asymmetries in human visual cortex. *J Neurosci* 27 (6), 1356–64.
- [3] Ashburner, J., Friston, K. J., Jul 2005. Unified segmentation. *Neuroimage* 26 (3), 839–51.
- [4] Avants, B. B., Klein, A., Tustison, N. J., Woo, J., Gee, J. C., 2010. Evaluation of an open-access, automated brain extraction method on multi-site multi-disorder data. *Human Brain Mapping*.
- [5] Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., Gee, J. C., Feb 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54 (3), 2033–44.
- [6] Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., Gee, J. C., Dec 2011. An open source multivariate framework for *n*-tissue segmentation with evaluation on public data. *Neuroinformatics* 9 (4), 381–400.
- [7] Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J. C., Feb 2010. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49 (3), 2457–66.
- [8] Ballmaier, M., Sowell, E. R., Thompson, P. M., Kumar, A., Narr, K. L., Lavretsky, H., Welcome, S. E., DeLuca, H., Toga, A. W., Feb 2004. Mapping brain size and cortical gray matter changes in elderly depression. *Biol Psychiatry* 55 (4), 382–9.
- [9] Bermudez, P., Lerch, J. P., Evans, A. C., Zatorre, R. J., Jul 2009. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex* 19 (7), 1583–96.
- [10] Boyes, R. G., Gunter, J. L., Frost, C., Janke, A. L., Yeatman, T., Hill, D. L. G., Bernstein, M. A., Thompson, P. M., Weiner, M. W., Schuff, N., Alexander, G. E., Killiany, R. J., DeCarli, C., Jack, C. R., Fox, N. C., ADNI Study, Feb 2008. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage* 39 (4), 1752–62.
- [11] Breiman, L., 2001. Random forests. In: *Machine Learning*, pp. 5–32.
- [12] Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., Evans, A. C., Oct 2008. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb Cortex* 18 (10), 2374–81.
- [13] Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., Evans, A. C., May 2005. Cortical thickness analysis in autism with heat kernel smoothing. *Neuroimage* 25 (4), 1256–65.
- [14] Clarkson, M. J., Cardoso, M. J., Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., Ourselin, S., Aug 2011. A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage* 57 (3), 856–65.
- [15] Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18 (2), 192–205.
- [16] Dale, A. M., Fischl, B., Sereno, M. I., Feb 1999. Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–94.
- [17] Das, S. R., Avants, B. B., Grossman, M., Gee, J. C., Apr 2009. Registration based cortical thickness measurement. *Neuroimage* 45 (3), 867–79.
- [18] DaSilva, A. F. M., Granziera, C., Snyder, J., Hadjikhani, N., Nov 2007. Thickening in the somatosensory cortex of patients with migraine. *Neurology* 69 (21), 1990–5.
- [19] Davatzikos, C., Bryan, N., 1996. Using a deformable surface model to obtain a shape representation of the cortex. *IEEE Trans Med Imaging* 15 (6), 785–95.
- [20] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., Killiany, R. J., Jul 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage* 31 (3), 968–80.
- [21] Dickerson, B. C., Bakkar, 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., Buckner, R. L., Mar 2009. 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 (3), 497–510.
- [22] Dogdas, B., Shattuck, D. W., Leahy, R. M., Dec 2005. Segmentation of skull and scalp in 3-D human MRI using mathematical morphology. *Hum Brain Mapp* 26 (4), 273–85.
- [23] Du, A.-T., Schuff, N., Kramer, J. H., Rosen, H. J., Gorno-Tempini, M. L., Rankin, K., Miller, B. L., Weiner, M. W., Apr 2007. Different regional patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia. *Brain* 130 (Pt 4), 1159–66.
- [24] Eggert, L. D., Sommer, J., Jansen, A., Kircher, T., Konrad, C., 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PLoS One* 7 (9), e45081.

- [25] Evans, A. C., Janke, A. L., Collins, D. L., Bailet, S., Aug 2012. Brain templates and atlases. *Neuroimage* 62 (2), 911–22.
- [26] Fischl, B., Dale, A. M., Sep 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A* 97 (20), 11050–5.
- [27] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A. M., Jan 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (3), 341–55.
- [28] Fischl, B., Sereno, M. I., Dale, A. M., Feb 1999. Cortical surface-based analysis. ii: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9 (2), 195–207.
- [29] Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M., Jan 2004. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14 (1), 11–22.
- [30] Fortier, C. B., Leritz, E. C., Salat, D. H., Venne, J. R., Maksimovskiy, A. L., Williams, V., Milberg, W. P., McGlinchey, R. E., Dec 2011. Reduced cortical thickness in abstinent alcoholics and association with alcoholic behavior. *Alcohol Clin Exp Res* 35 (12), 2193–201.
- [31] Foster, N. E. V., Zatorre, R. J., Oct 2010. Cortical structure predicts success in performing musical transformation judgments. *Neuroimage* 53 (1), 26–36.
- [32] Franklin, T. R., Wang, Z., Shin, J., Jagannathan, K., Suh, J. J., Detre, J. A., O'Brien, C. P., Childress, A. R., 2013. A VBM study demonstrating ‘apparent’ effects of a single dose of medication on T1-weighted MRIs. *Brain Structure and Function* 218 (1), 97–104.
- [33] Frøkær, J. B., Bouwense, S. A. W., Olesen, S. S., Lundager, F. H., Eskildsen, S. F., van Goor, H., Wilder-Smith, O. H. G., Drewes, A. M., Apr 2012. Reduced cortical thickness of brain areas involved in pain processing in patients with chronic pancreatitis. *Clin Gastroenterol Hepatol* 10 (4), 434–8.e1.
- [34] Gernsbacher, M. A., Jan 2007. Presidential column: The eye of the beholder. *Observer* 20 (1).
- [35] Gronenschild, E. H. B. M., Habets, P., Jacobs, H. I. L., Mengelers, R., Rozendaal, N., van Os, J., Marcelis, M., 2012. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS One* 7 (6), e38234.
- [36] Haier, R. J., Karama, S., Leyba, L., Jung, R. E., 2009. 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, 174.
- [37] Hampel, F. R., March 2001. Robust statistics: A brief introduction and overview. Research Report 94, Eidgenössische Technische Hochschule.
- [38] Hardan, A. Y., Muddasani, S., Vemulapalli, M., Keshavan, M. S., Minshew, N. J., Jul 2006. An MRI study of increased cortical thickness in autism. *Am J Psychiatry* 163 (7), 1290–2.
- [39] He, Y., Chen, Z. J., Evans, A. C., Oct 2007. Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb Cortex* 17 (10), 2407–19.
- [40] Heim, C. M., Mayberg, H. S., Mletzko, T., Nemeroff, C. B., Pruessner, J. C., Jun 2013. Decreased cortical representation of genital somatosensory field after childhood sexual abuse. *Am J Psychiatry* 170 (6), 616–23.
- [41] Jiang, J., Zhu, W., Shi, F., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., Jiang, T., Feb 2009. Thick visual cortex in the early blind. *J Neurosci* 29 (7), 2205–11.
- [42] Jones, S. E., Buchbinder, B. R., Aharon, I., Sep 2000. Three-dimensional mapping of cortical thickness using Laplace’s equation. *Hum Brain Mapp* 11 (1), 12–32.
- [43] Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wilfong, J., Roccatagliata, L., Nobili, F., Hensch, T., Tränker, 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., The PharmaCog Consortium, May 2013. Brain morphometry reproducibility in multi-center 3T MRI studies: A comparison of cross-sectional and longitudinal segmentations. *Neuroimage*.
- [44] Jubault, T., Gagnon, J.-F., Karama, S., Ptito, A., Lafontaine, A.-L., Evans, A. C., Monchi, O., Mar 2011. Patterns of cortical thickness and surface area in early Parkinson’s disease. *Neuroimage* 55 (2), 462–7.
- [45] Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab’bagh, Y., MacDonald, D., Lee, J. M., Kim, S. I., Evans, A. C., Aug 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage* 27 (1), 210–21.
- [46] Klein, A., Ghosh, S. S., Avants, B., Yeo, B. T. T., Fischl, B., Ardekani, B., Gee, J. C., Mann, J. J., Parsey, R. V., May 2010. Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 51 (1), 214–20.
- [47] Klein, A., Tourville, J., 2012. 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci* 6, 171.
- [48] Kochunov, P., Glahn, D. C., Lancaster, J., Thompson, P. M., Kochunov, V., Rogers, B., Fox, P., Blangero, J., Williamson, D. E., Sep 2011. Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan. *NeuroImage* 58 (1), 41–9.
- [49] Kovacevic, J., 2006. From the editor-in-chief. *IEEE Trans Imag Process* 15 (12).
- [50] Kühn, S., Schubert, F., Gallinat, J., Dec 2010. Reduced thickness of medial orbitofrontal cortex in smokers. *Biol Psychiatry* 68 (11), 1061–5.
- [51] 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., van Zijl, P. C. M., Feb 2011. Multi-parametric neuroimaging reproducibility: a 3-t resource study. *Neuroimage* 54 (4), 2854–66.
- [52] 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., Fischl, B., Nov 2005. Meditation experience is associated with increased cortical thickness. *Neuroreport* 16 (17), 1893–7.
- [53] Lerch, J. P., Worsley, K., Shaw, W. P., Greenstein, D. K., Lenroot, R. K., Giedd, J., Evans, A. C., Jul 2006. Mapping anatomical correlations across cerebral cortex (MACACC) using cortical thickness from MRI. *NeuroImage* 31 (3), 993–1003.
- [54] Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., Toga, A. W., Aug 2006. Hemispheric asymmetries in cortical thickness. *Cereb Cortex* 16 (8), 1232–8.
- [55] Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Woods, R. P., Deluca, H., Jancke, L., Toga, A. W., Apr 2006. Gender effects on cortical thickness and the influence of scaling. *Hum Brain Mapp* 27 (4), 314–24.
- [56] Luders, E., Sánchez, F. J., Tosun, D., Shattuck, D. W., Gaser, C., Vilain, E., Toga, A. W., Aug 2012. Increased cortical thickness in male-to-female transsexualism. *J Behav Brain Sci* 2 (3), 357–362.
- [57] Lüsebrink, F., Wollrab, A., Speck, O., Apr 2013. Cortical thickness determination of the human brain using high resolution 3t and 7t mri data. *NeuroImage* 70, 122–31.
- [58] Lyoo, I. K., Sung, Y. H., Dager, S. R., Friedman, S. D., Lee, J.-Y., Kim, S. J., Kim, N., Dunner, D. L., Renshaw, P. F., Feb 2006. Regional cerebral cortical thinning in bipolar disorder. *Bipolar Disord* 8 (1), 65–74.
- [59] MacDonald, D., Kabani, N., Avis, D., Evans, A. C., Sep 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* 12 (3), 340–56.
- [60] Magnotta, V. A., Andreasen, N. C., Schultz, S. K., Harris, G., Cizadlo, T., Heckel, D., Nopoulos, P., Flaum, M., Mar 1999. Quantitative in vivo measurement of gyration in the human brain: changes associated with aging. *Cereb Cortex* 9 (2), 151–60.
- [61] 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., Breiter, H. C., Oct 2008. Cortical thickness abnormalities in cocaine addiction—a reflection of both drug use and a pre-existing disposition to drug abuse? *Neuron* 60 (1), 174–88.
- [62] Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., Lancaster, J., Jun 1995. A probabilistic atlas of the human brain: theory and rationale for its development. the International Consortium for Brain Mapping (ICBM). *Neuroimage* 2 (2), 89–101.
- [63] Nakamura, K., Fox, R., Fisher, E., Jan 2011. CLADA: cortical longitudinal atrophy detection algorithm. *Neuroimage* 54 (1), 278–89.
- [64] Neary, D., Snowden, J., Mann, D., Nov 2005. Frontotemporal dementia. *Lancet Neurol* 4 (11), 771–80.

- [65] Nesvåg, R., Lawyer, G., Varnäs, K., Fjell, A. M., Walhovd, K. B., Frigessi, A., Jönsson, E. G., Agartz, I., Jan 2008. Regional thinning of the cerebral cortex in schizophrenia: effects of diagnosis, age and antipsychotic medication. *Schizophr Res* 98 (1-3), 16–28.
- [66] Otsu, N., 1979. A threshold selection method from gray-level histograms. *EEE Trans. Sys., Man., Cyber.* 9 (1), 62–66.
- [67] Peterson, B. S., Warner, V., Bansal, R., Zhu, H., Hao, X., Liu, J., Durkin, K., Adams, P. B., Wickramaratne, P., Weissman, M. M., Apr 2009. Cortical thinning in persons at increased familial risk for major depression. *Proc Natl Acad Sci U S A* 106 (15), 6273–8.
- [68] Raine, A., Laufer, W. S., Yang, Y., Narr, K. L., Thompson, P., Toga, A. W., Oct 2011. Increased executive functioning, attention, and cortical thickness in white-collar criminals. *Hum Brain Mapp*.
- [69] Raji, C. A., Ho, A. J., Parikhshak, N. N., Becker, J. T., Lopez, O. L., Kuller, L. H., Hua, X., Leow, A. D., Toga, A. W., Thompson, P. M., Mar 2010. Brain structure and obesity. *Hum Brain Mapp* 31 (3), 353–64.
- [70] Ramasamy, D. P., Benedict, R. H. B., Cox, J. L., Fritz, D., Abdelrahman, N., Hussein, S., Minagar, A., Dwyer, M. G., Zivadinov, R., Jul 2009. Extent of cerebellum, subcortical and cortical atrophy in patients with MS: a case-control study. *J Neurol Sci* 282 (1-2), 47–54.
- [71] Rosas, H. D., Hevelone, N. D., Zaleta, A. K., Greve, D. N., Salat, D. H., Fischl, B., Sep 2005. Regional cortical thinning in preclinical Huntington disease and its relationship to cognition. *Neurology* 65 (5), 745–7.
- [72] Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., van der Kouwe, A., Jenkins, B. G., Dale, A. M., Fischl, B., Mar 2002. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58 (5), 695–701.
- [73] Salgado-Pineda, P., Delaveau, P., Falcon, C., Blin, O., 2006. Brain t1 intensity changes after levodopa administration in healthy subjects: a voxel-based morphometry study. *British journal of clinical pharmacology* 62 (5), 546–551.
- [74] Scott, M. L. J., Bromiley, P. A., Thacker, N. A., Hutchinson, C. E., Jackson, A., Apr 2009. A fast, model-independent method for cerebral cortical thickness estimation using MRI. *Med Image Anal* 13 (2), 269–85.
- [75] Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., Fischl, B., Jul 2004. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22 (3), 1060–75.
- [76] Selemon, L. D., Rajkowska, G., Goldman-Rakic, P. S., Jan 2004. Evidence for progression in frontal cortical pathology in late-stage Huntington's disease. *J Comp Neurol* 468 (2), 190–204.
- [77] Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., Leahy, R. M., May 2001. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage* 13 (5), 856–76.
- [78] Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., Giedd, J., Mar 2006. Intellectual ability and cortical development in children and adolescents. *Nature* 440 (7084), 676–9.
- [79] Shin, Y.-W., Yoo, S. Y., Lee, J. K., Ha, T. H., Lee, K. J., Lee, J. M., Kim, I. Y., Kim, S. I., Kwon, J. S., Nov 2007. Cortical thinning in obsessive-compulsive disorder. *Hum Brain Mapp* 28 (11), 1128–35.
- [80] Shrout, P. E., Fleiss, J. L., Mar 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86 (2), 420–8.
- [81] Sled, J. G., Zijdenbos, A. P., Evans, A. C., Feb 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 17 (1), 87–97.
- [82] Smith, S. M., Nov 2002. Fast robust automated brain extraction. *Hum Brain Mapp* 17 (3), 143–55.
- [83] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobniak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., Matthews, P. M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 Suppl 1, S208–19.
- [84] Sowell, E. R., Kan, E., Yoshii, J., Thompson, P. M., Bansal, R., Xu, D., Toga, A. W., Peterson, B. S., Jun 2008. Thinning of sensorimotor cortices in children with Tourette syndrome. *Nat Neurosci* 11 (6), 637–9.
- [85] Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain: 3-Dimensional proportional system—An approach to cerebral imaging. Thieme.
- [86] 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., Reiss, A. L., Apr 2005. Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *J Neurosci* 25 (16), 4146–58.
- [87] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., Gee, J. C., Jun 2010. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 29 (6), 1310–20.
- [88] Tustison, N. J., Johnson, H. J., Rohlfing, T., Klein, A., Ghosh, S. S., Ibanez, L., Avants, B., 2013. Instrumentation bias in the use and evaluation of scientific software: Recommendations for reproducible practices in the computational sciences. *Frontiers in Neuroscience* 7 (162).
- [89] Vachet, C., Hazlett, H. C., Niethammer, M., Oguz, I., Cates, J., Whitaker, R., Piven, J., Styner, M., February 2011. Group-wise automatic mesh-based analysis of cortical thickness. In: Benoit M. Dawant, D. R. H. (Ed.), *SPIE Medical Imaging: Image Processing*.
- [90] Wang, D., Shi, L., Chu, W. C. W., Burwell, R. G., Cheng, J. C. Y., Ahuja, A. T., Jan 2012. Abnormal cerebral cortical thinning pattern in adolescent girls with idiopathic scoliosis. *Neuroimage* 59 (2), 935–42.
- [91] Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., Yushkevich, P. A., 2013. Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Analysis and Machine Intelligence*.
- [92] Ward, B. D., 1999. Intracranial segmentation. Tech. rep., Medical College of Wisconsin, <http://afni.nimh.nih.gov/pub/dist/doc/3dIntracranial.pdf>.
- [93] Wei, G., Zhang, Y., Jiang, T., Luo, J., 2011. Increased cortical thickness in sports experts: a comparison of diving players with the controls. *PLoS One* 6 (2), e17112.
- [94] 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., , A. D. N. I., Feb 2012. The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement* 8 (1 Suppl), S1–68.
- [95] Wilkinson, G. N., Rogers, C. E., 1973. Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 22 (3), 392–399.
- [96] Worsley, K. J., Chen, J.-I., Lerch, J., Evans, A. C., May 2005. Comparing functional connectivity via thresholding correlations and singular value decomposition. *Philos Trans R Soc Lond B Biol Sci* 360 (1457), 913–20.
- [97] Yamasue, H., Abe, O., Kasai, K., Suga, M., Iwanami, A., Yamada, H., Tochigi, M., Ohtani, T., Rogers, M. A., Sasaki, T., et al., 2007. Human brain structural change related to acute single exposure to sarin. *Annals of neurology* 61 (1), 37–46.
- [98] Yezzi, Jr, A. J., Prince, J. L., Oct 2003. An Eulerian PDE approach for computing tissue thickness. *IEEE Trans Med Imaging* 22 (10), 1332–9.
- [99] Zeng, X., Staib, L. H., Schultz, R. T., Duncan, J. S., Oct 1999. Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation. *IEEE Trans Med Imaging* 18 (10), 927–37.
- [100] Zhang, Y., Brady, M., Smith, S., Jan 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20 (1), 45–57.