

Large-scale cortical thickness quantification with Advanced Normalization Tools (ANTs)

Nicholas J. Tustison^{a,1}, 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

^cDepartment 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 reproducible localized measurements characterizing the cortex with little or no human intervention. Particularly useful to the neuroscience community are publicly available tools, such as the popular surface-based Freesurfer, which facilitate the testing and refinement of hypotheses. In this paper, we introduce 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) all developed as part of the ANTs open science effort. We employ cortical thickness as both an outcome measurement and as a reference for validation. Errors in any previous part of the pipeline will influence the biological plausibility of the final cortical thickness results including the statistical sensitivity of these measurements to demographic variables such as age and gender. Furthermore, we demonstrate, for the first time, that volumetric cortical thickness measurements are viable and sensitive in large-scale population that span age, gender and image acquisition platforms. We use four open data sets (IXI, Kirby, NKI, and Oasis), consisting of approximately 1200 images, in order to evaluate the ANTs cortical thickness pipeline. Because ground truth is unavailable in these large datasets, we use surrogate measurements to establish validity. We show that the pipeline has a low failure rate, has high cortical thickness precision, and produces reasonable “BrainAGE” measurements. DiReCT is also shown to reveal expected relationships between cortical thickness, age and gender, in addition to novel gender differences in predicting structural connectivity across the lifespan. To further promote open science and use of the proposed tools, all results and scripts used to produce the results are publicly available using open science venues. We also introduce quality assurance/control mechanisms, such as the *brain constellation map*, for quick assessment of performance on individual subjects. Out of over 1200 subjects, we found no severe registration, segmentation or cortical thickness failures using the proposed methodology.

Keywords: advanced normalization tools, cortical thickness, 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, 77], schizophrenia [65], bipolar disorder [59], Alzheimer’s disease and frontotemporal dementia [23, 21], Parkinson’s disease [46], Williams syndrome [88], multiple sclerosis [70], autism [14, 39], migraines [19], chronic smoking [51], alcoholism [30], cocaine addiction [62], Tourette syndrome in children [85], scoliosis in female adolescents [92], early-onset blindness [43], chronic pancreatitis [34], obsessive-compulsive disorder

[80], ADHD [1], obesity [69], and heritable [67] and elderly [9] depression. Evidence of cortical thickness variation has also been found to be a function of age [49], gender [56], untreated male-to-female transsexuality [57], handedness [55, 2], intelligence [79], athletic ability [96], meditative practices [53], musical ability [10, 31], tendency toward criminality [68], childhood sexual abuse in adult females [42], and Tetris-playing ability in female adolescents [37]. Additionally, recent studies demonstrate structural connectivity relationships using cortical thickness measures [99, 54, 41, 12]. Although these findings are subject to debate and interpretation [35], the availability of quantitative computational methods for extracting such information has proven invaluable for developing and refining fundamental neuroscience hypotheses.

Large neuroimaging datasets such as that provided by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) are increasing the importance of fully automated and multiple modality brain mapping tools [97]. The scale of such datasets will

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.

only increase over time as international ADNI 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, very few methods have been thoroughly tested on large and unstructured neuroimaging datasets.

Computational methods for analyzing the cortex may be broadly characterized as surface mesh-based or volumetric [75, 15]. Representative of the former is the Freesurfer² cortical modeling software package [17, 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) [84]. Similar to other surface approaches (e.g. [20, 61, 60, 47]), the pial and white matter surfaces from individual subject MR data are modeled with polygonal meshes which are then used to determine 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 [44] 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 [101], an Eulerian PDE approach was proposed to facilitate the computation of correspondence paths. Extending the surface-based work of [60], the hybrid approach of [47] 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 [103], model-free intelligent search strategies either normal to the gray-white matter interface [75], or using a min-max rule [91]. Most relevant to this work is the DiReCT (Diffeomorphic Registration-based Cortical Thickness) algorithm proposed in [18] 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 for exist for estimating cortical thickness from imaging data (of which only a fraction

are cited), several common preprocessing components may be identified. The most fundamental of these include inhomogeneity correction, skull stripping, and *n*-tissue segmentation for differentiating the gray and white matter. For statistical analysis across large populations, construction of population-specific unbiased templates is also potentially beneficial [24]. In addition, intermediate steps might include a crucial registration component (e.g. propagating template-based tissue priors for improved segmentation).

The general lack of availability of published algorithms [50] (not to mention critical preprocessing components), inhibits performing studies by external researchers and makes comparative evaluations difficult. For example, one recent evaluation study [15] compared Freesurfer (a surface-based method) with two volumetric methods [44, 18]. 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 making the comparisons less than ideal (see [90] for further discussion concerning the issue of instrumentation bias 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 [48].

In this work, we describe our cortical thickness pipeline which produces a cortical thickness map from an individual subject’s T1-weighted MRI. Additionally, 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 [89], brain extraction [5], *n*-tissue segmentation [7], template construction [8], and image normalization [6]. We also describe improvements made to the original DiReCT algorithm [18]. More important, however, we provide explicit coordination between these pipeline 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, viz., N.T. and B.A, and provide good performance across a number of data sets. Furthermore, we post all the derived image data, processing scripts, and a 2-D example with data and command line call using the online tools figshare³ and github⁴, respectively. 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 processing and to adapt the given volumetric pipeline for measuring cortical thickness with their own datasets. Below, we overview a framework that achieves an effective zero failure rate in processing a large collection of neuroimages culled from existing public data repositories.

²<http://surfer.nmr.mgh.harvard.edu/>

³<http://www.figshare.com>

⁴<http://www.github.com>

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 Oasis subject to demonstrate the simplicity and mature status of what we propose in this work. Out of all possible parameter combinations, extensive tuning produced a single set of parameters for processing all the data described in this work allowing other researchers easy access to the performance showcased in the following sections.

```
antsCorticalThickness.sh \
-d 3 \
-a Oasis/T1/OAS1_0001_MR1_mpr_n4_anon_sbj_111.nii.gz \
-e Oasis/template/T_template0.nii.gz \
-m Oasis/template/T_template0ProbabilityMask.nii.gz \
-p Oasis/template/Priors/priors%d.nii.gz \
-f Oasis/template/T_template0ExtractionMask.nii.gz \
-k 0 \
-o Oasis/Results/OAS1_0001_MR1_mpr_n4_anon_sbj_111 \
-w 0.25 \
-n 3
```

Listing 1: Command line call for a single Oasis subject which is representative of the calls used for each subject of the evaluation study. Option descriptions are provided by invoking the help option, i.e. ‘`antsCorticalThickness.sh -h`’.

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 [87], or a publicly available averaged group of subjects, e.g. the MNI [16] or ICBM [63] templates. Additionally, mean templates constructed from labeled data can be used to construct spatial priors for improving segmentation algorithms. The work of [8] explicitly models the geometric component of the normalized space during optimization to produce such mean templates. Coupling the intrinsic symmetry of SyN pairwise registration [6] 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 pexec⁵,
- parallelized processing using Apple’s XGrid technology⁶,
- parallelized processing using Sun Grid Engine for cluster-based systems⁷, and
- parallelized processing using the Portable Batch System for cluster-based systems⁸.

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 [82], as part of their standard protocol [11].

In [89], 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 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. In addition to its public availability through ANTs

⁵<http://www.gnu.org/software/pexec/pexec.1.html>

⁶<https://developer.apple.com/hardwaredrivers/hpc/xgrid.intro.html>

⁷<http://www.oracle.com/technetwork/oem/grid-engine-166852.html>

⁸<http://www.pbsworks.com/>

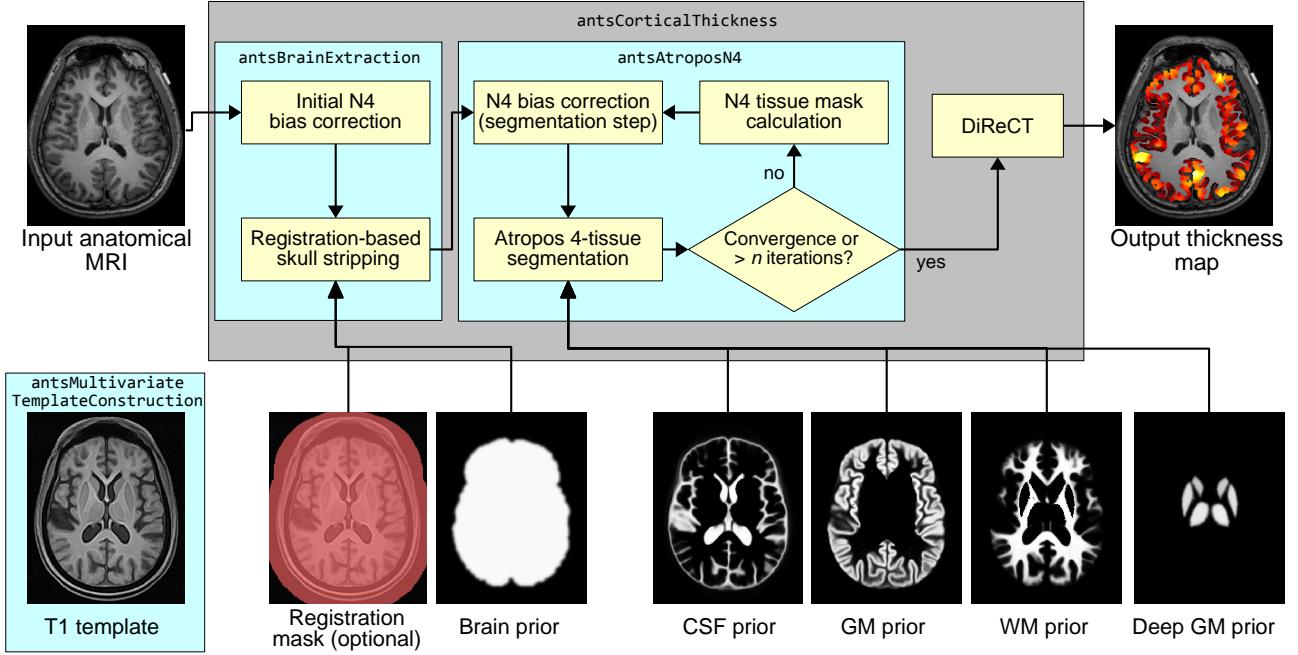


Figure 1: Illustration of the main components of the ANTs processing workflow containing all elements for determining cortical thickness. We also included the domain of operations for the selected scripts. Not shown is the optional subject to template registration.

and the Insight Toolkit, it has also been included in the popular open source Slicer software package for visualization and medical image computing [25].

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, four-tissue 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 [8] 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 2. 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.

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 [94], FSL’s BET2 [83], FreeSurfer’s mri_watershed [76], and BrainSuite [22] demonstrated that our combined registration/segmentation approach [5] performs at the top level alongside BrainSuite (tuned) and FreeSurfer.

2.1.4. Atropos four-tissue segmentation

In [7] 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. [104, 4]). 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 [78], 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, 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

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

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

$$P(x|y) \propto P(y|x)^{1-\alpha} P(x)^\alpha \quad (2)$$

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). Thus, accurate differentiation between white matter and deep gray matter is not necessary for accurate cortical thickness measurements.

2.1.5. DiReCT (aka KellySlater/KellyKapowski) Cortical Thickness Estimation

DiReCT was introduced in [18] and made available in ANTs as the program KellySlater. Since then several improvements have been made and incorporated into the program KellyKapowski.⁹ 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 available through ANTs complete with a unique user interface design developed specifically for ANTs tools.

⁹Traditional 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.

2.2. Public Data Resources

As mentioned previously, the four public data sets which were processed using the previously described pipeline are:

- IXI,
- Kirby,
- NKI, and
- Oasis.

In addition, we used the NIREP data set to produce cortical labels for each of the processed T1 images. All five data sets are described below.

2.2.1. NIREP Data for Cortical Labelings

The Nonrigid Image Registration Evaluation Project (NIREP¹⁰) is an ongoing framework for evaluating image registration algorithms [13]. The initial data set introduced into the project consists of 16 (8 male and 8 female) high resolution skull-stripped brain data with 32 cortical labels (cf Table 1) manually drawn using a published protocol. Reasons for using the NIREP data include public availability, high resolution T1-weighted images, and the number of subjects for use with our label propagation scheme.

Cortical label propagation to each individual subject for all data sets described below was performed using the PICSL multi-atlas joint label fusion algorithm described in [93] which recently led the competition at the MICCAI 2012 Grand Challenge on Multi-Atlas Labeling¹¹ and 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. Scatter plots for each region listed in Table 1 showing the cortical thickness for all cohort data are given in Figure 2.

2.2.2. Data for Pipeline Evaluation

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

¹⁰<http://www.nirep.org/>

¹¹https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_Page

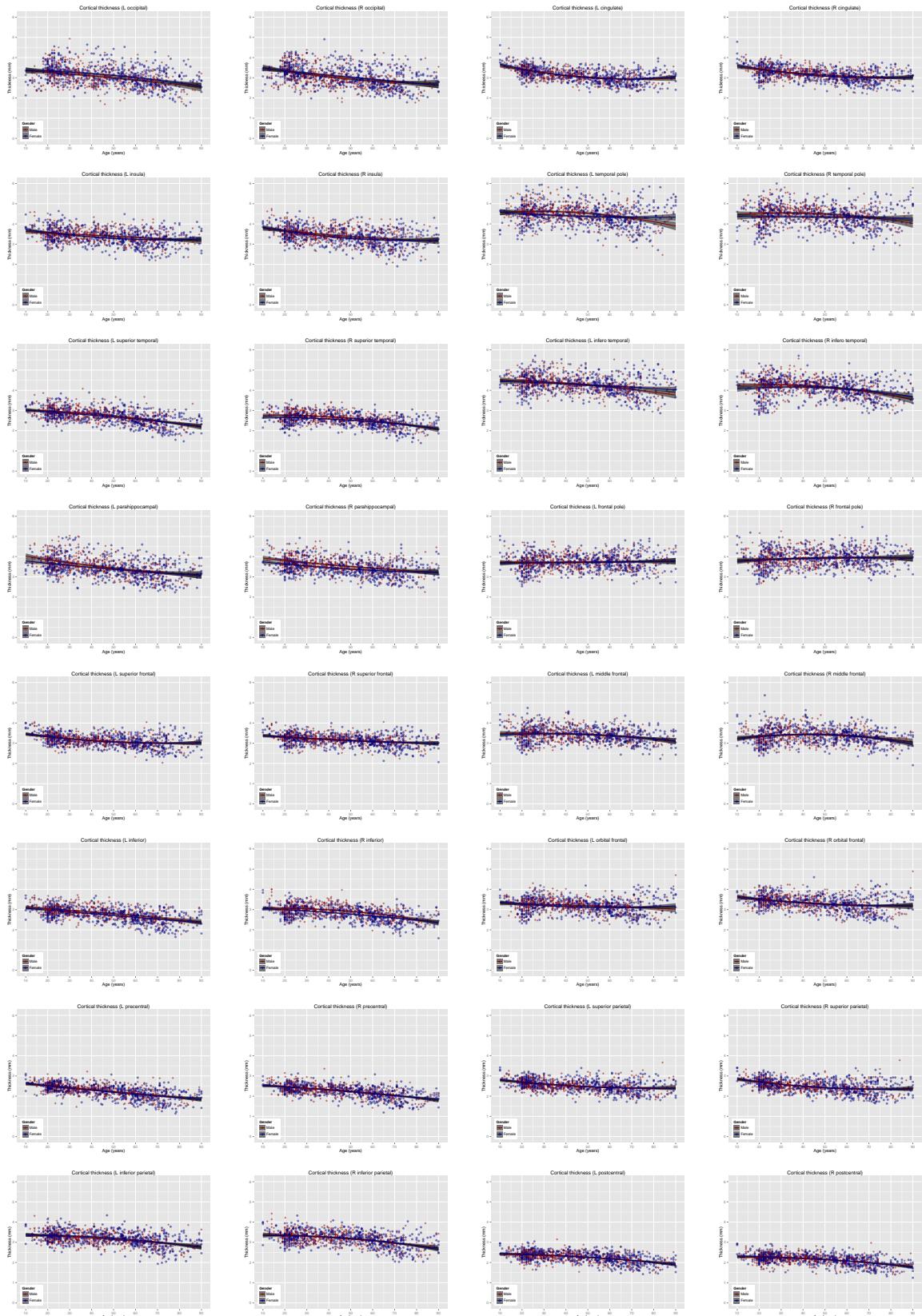


Figure 2: Age vs. thickness plots for all data ($T(REGION_i) \sim AGE + AGE^2$) where the regions are arranged left to right according to Table 1. Individual female subjects are rendered in blue, males in dark red. The R script to produce these plots are `ageThicknessPlots.R`. Data are located in the following csv files: `labelresultsI.csv` (IXI), `labelresultsK.csv` (Kirby), `labelresultsN.csv` (NKI), and `labelresultsO.csv` (Oasis).

1) L occipital lobe	2) R occipital lobe
3) L cingulate gyrus	4) R cingulate gyrus
5) L insula gyrus	6) R insula gyrus
7) L temporal lobe	8) R temporal lobe
9) L superior temporal gyr.	10) R superior temporal gyr.
11) L infero temporal region	12) R infero temporal region
13) L parahippocampal gyr.	14) R parahippocampal gyr.
15) L frontal pole	16) R frontal pole
17) L superior frontal gyrus	18) R superior frontal gyrus
19) L middle frontal gyrus	20) R middle frontal gyrus
21) L inferior gyrus	22) R inferior gyrus
23) L orbital frontal gyrus	24) R orbital frontal gyrus
25) L precentral gyrus	26) R precentral gyrus
27) L superior parietal lobule	28) R superior parietal lobule
29) L inferior parietal lobule	30) R inferior parietal lobule
31) L postcentral gyrus	32) R postcentral gyrus

Table 1: The 32 cortical NIREP labels. Regional thickness values across age are given in Figure 2.

IXI. Initially, we began with 581 T1-weighted images from the IXI¹² 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 Resource¹³, or more informally, the Kirby data set, was originally described in [52] 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 Project¹⁴ was initiated on December 11, 2009 by various members of the MRI community seeking to form collaborative partnerships with imaging institutions for sharing well-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).¹⁵

Oasis. The initial Open Access Series of Imaging Studies (OASIS)¹⁶ 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.

3. Evaluation

We cannot rely on traditional approaches such as manual labeling to evaluate large-scale performance. We therefore sought to minimize failure rate, quantify the repeatability of cortical thickness measures, test accuracy in BrainAGE [32], and determine whether the DiReCT pipeline reveals biologically plausible relationships between the cortex, gender,¹⁷ and age. Collectively, these surrogate measurements allow us to establish data-derived performance standards.

¹²<http://biomedic.doc.ic.ac.uk/brain-development/>

¹³<http://www.nitrc.org/projects/multimodal/>

¹⁴http://fcon_1000.projects.nitrc.org

¹⁵Downloaded on September 22, 2012.

¹⁶<http://www.oasis-brains.org/>

¹⁷We recognize the distinction often made between “sex” and “gender” (<http://www.who.int/gender/whatisgender/en/>). As the demographic information collected during the course of the imaging studies is presumably self-reported, we assume that most self-identify in terms of gender and, therefore, use the term “gender” in image data descriptions.

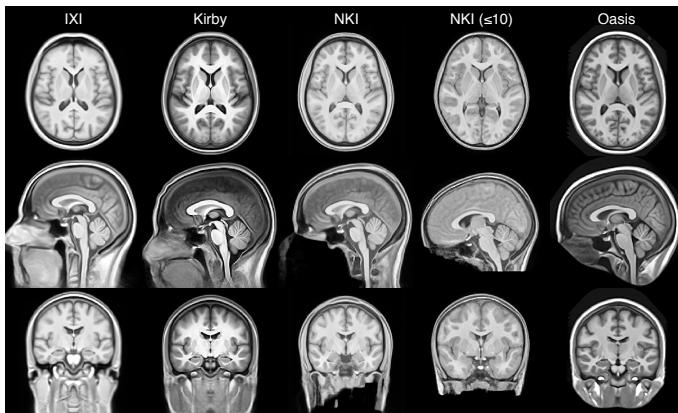
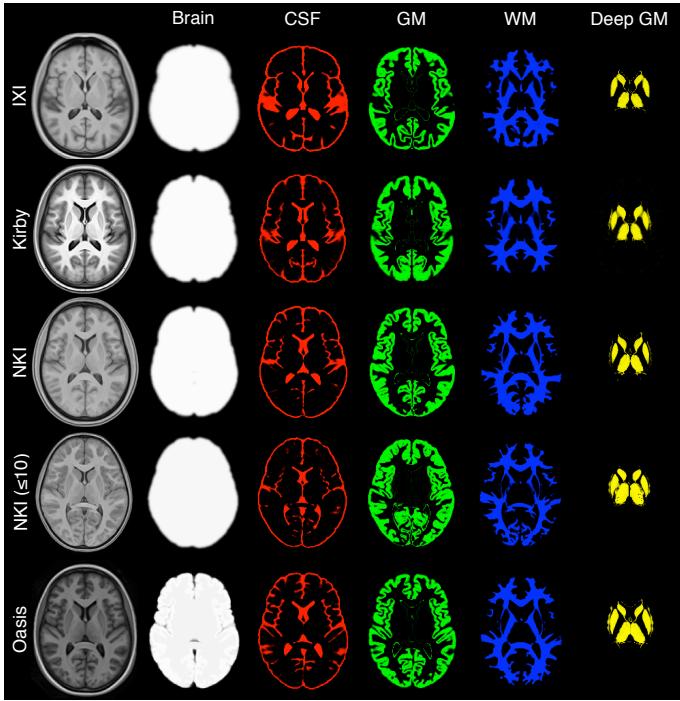


Figure 3: Population-specific templates for each of the four public data sets used for cortical thickness estimation generated using the script antsMultivariateTemplateConstruction.sh. The benefit of using such templates is obvious considering the variability in acquisition and data preparation (e.g.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



3.2. Reproducibility

Figure 4: Axial slices from each of the four T1 templates including the corresponding probability masks used for brain extraction and four-tissue brain segmentation.

3.1. Computation Time and Failure Rate

All images underwent the pipeline processing illustrated in Figure 1 using the computational cluster at the University of Virginia.¹⁸ Processing times varied approximately between 10–20 hours per subject for the entire cortical thickness estimation procedure. The propagation of the NIREP 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 [93].¹⁹ Average thickness values were tabulated per subject for each of the 32 NIREP labels. Brain volumes (cerebrum only excluding the brain stem and cerebellum) for each subject derived from the brain extraction step were also calculated and are used as the *VOLUME* variable measurement in the evaluations below. All these data were written to separate csv files corresponding to data set for subsequent analysis (also included with the scripts). Visual sample results from each data set are provided in Figure 5.

No obvious brain extraction failures were detected during the course of this study. Quality assessment was done manually by multiple rounds of visual inspection using ITK-SNAP [102]. Similarly, we detected no obvious segmentation errors. Brain constellation maps (detailed in a later section) provided a unified view of all thickness results for the entire cohort for a quick quality check.

¹⁸<http://www.uvacse.virginia.edu/itc-clusters/>

¹⁹All processing on the UVA cluster was set to be single-threaded with a maximum requested memory footprint of 8 GB.

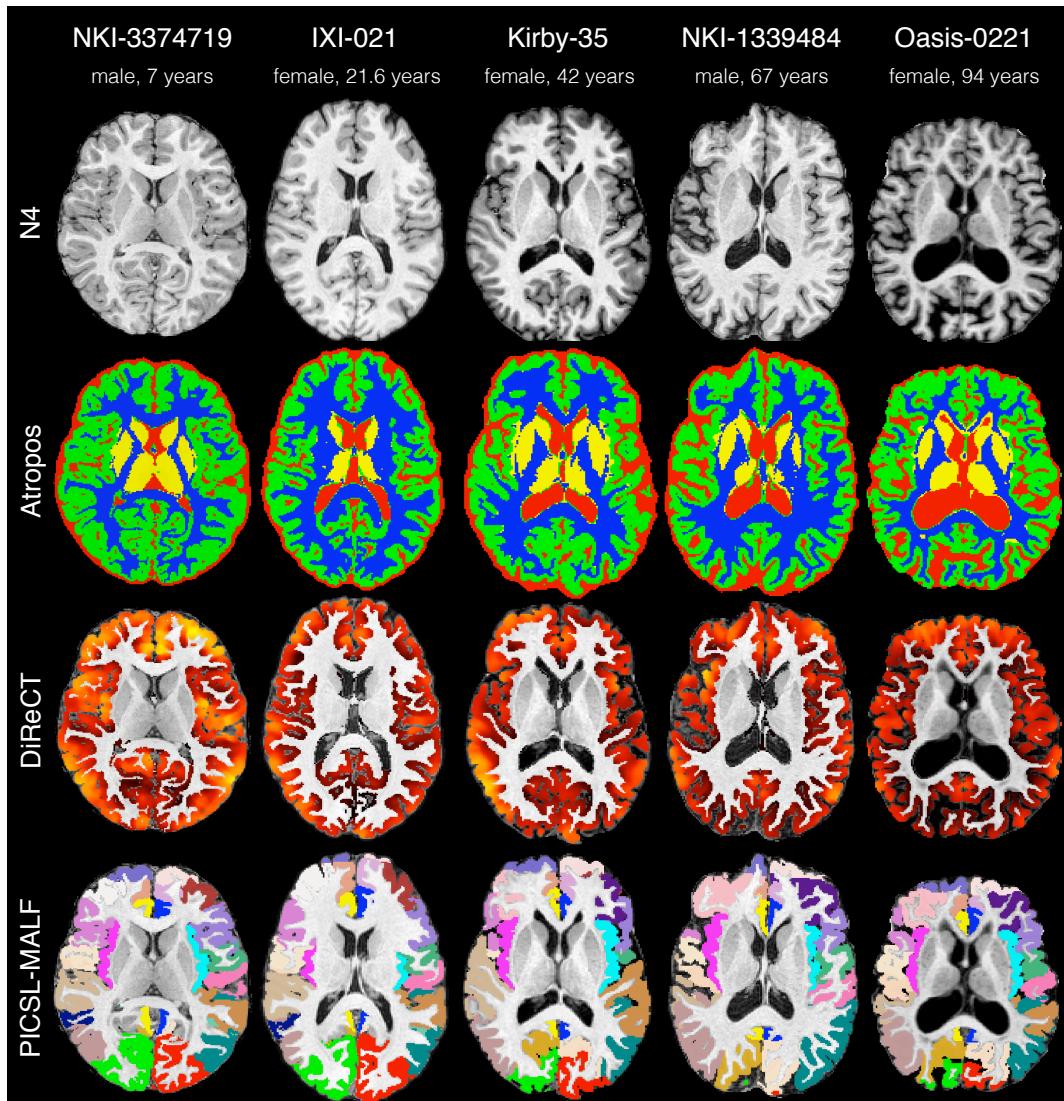


Figure 5: Sample results from each of the four data sets showing the N4 bias corrected images, 4-tissue segmentation, cortical thickness, and joint fusion label maps.

Region	Absolute Difference (mm)		Percent Variability Error	
	Left	Right	Left	Right
occipital	0.14 ± 0.14	0.16 ± 0.19	3.57 ± 3.8	4.13 ± 4.81
cingulate	0.07 ± 0.08	0.09 ± 0.09	2.18 ± 2.48	2.63 ± 2.9
insula	0.08 ± 0.05	0.09 ± 0.1	2.37 ± 1.6	2.54 ± 3.01
temporal pole	0.21 ± 0.14	0.19 ± 0.15	3.98 ± 3.04	3.89 ± 3.56
superior temporal	0.08 ± 0.07	0.07 ± 0.06	2.83 ± 2.31	2.54 ± 2.36
inferior temporal	0.2 ± 0.25	0.19 ± 0.24	4.34 ± 5.82	4.49 ± 6.38
parahippocampal	0.19 ± 0.21	0.17 ± 0.19	4.82 ± 5.51	4.29 ± 4.85
frontal pole	0.17 ± 0.17	0.17 ± 0.14	4.97 ± 4.97	4.42 ± 3.97
superior frontal	0.07 ± 0.06	0.08 ± 0.06	2.14 ± 1.91	2.45 ± 2.05
middle frontal	0.12 ± 0.12	0.09 ± 0.07	3.7 ± 3.71	2.96 ± 2.51
inferior	0.09 ± 0.09	0.09 ± 0.08	2.92 ± 3.15	3.19 ± 2.68
orbital frontal	0.11 ± 0.12	0.13 ± 0.12	3.19 ± 3.79	3.66 ± 3.69
precentral	0.07 ± 0.07	0.06 ± 0.05	2.52 ± 2.51	2.37 ± 1.91
superior parietal	0.07 ± 0.06	0.07 ± 0.05	2.45 ± 2.14	2.62 ± 1.67
inferior parietal	0.1 ± 0.09	0.09 ± 0.08	3.11 ± 2.67	2.82 ± 2.73
postcentral	0.09 ± 0.09	0.08 ± 0.05	3.56 ± 3.33	3.26 ± 2.25

Table 2: Mean absolute difference and percent variability error (\pm standard deviation) of repeated cortical measurements for both the Oasis and Kirby repeat scans. These differences were not statistically significant (two-tailed t -test with false discovery rate (FDR) multiple comparisons correction).

Region	Absolute Difference (mm)		Percent Variability Error	
	Left	Right	Left	Right
left.caudal.anterior.cingulate	0.09 ± 0.06	0.09 ± 0.07	2.89 ± 2.04	3.63 ± 3.02
left.cuneus	0.08 ± 0.07	0.29 ± 0.27	3.09 ± 3.14	5.33 ± 4.74
left.fusiform	0.08 ± 0.07	0.08 ± 0.06	1.96 ± 1.8	2.53 ± 2.16
left.inferior.temporal	0.17 ± 0.13	0.06 ± 0.04	3.5 ± 2.75	1.78 ± 1.21
left.lateral.ocipital	0.12 ± 0.08	0.14 ± 0.12	3.45 ± 2.44	4.17 ± 4.33
left.lingual	0.08 ± 0.07	0.19 ± 0.18	2.34 ± 1.96	5.45 ± 5.32
left.middle.temporal	0.12 ± 0.1	0.09 ± 0.08	3.2 ± 2.78	2.63 ± 2.39
left.paracentral	0.12 ± 0.11	0.08 ± 0.07	5.19 ± 4.47	2.78 ± 2.38
left.pars.orbitalis	0.09 ± 0.08	0.09 ± 0.1	2.95 ± 3.04	3.43 ± 3.87
left.pericalcarine	0.11 ± 0.09	0.1 ± 0.11	4.05 ± 3.36	4.46 ± 4.52
left.posterior.cingulate	0.08 ± 0.07	0.09 ± 0.1	2.66 ± 2.25	3.5 ± 3.92
left.precuneus	0.06 ± 0.04	0.11 ± 0.12	2.15 ± 1.47	2.63 ± 2.91
left.rostral.middle.frontal	0.13 ± 0.11	0.08 ± 0.07	4.39 ± 3.74	2.77 ± 2.58
left.superior.parietal	0.07 ± 0.06	0.09 ± 0.08	3.12 ± 2.89	2.96 ± 3.04
left.supramarginal	0.08 ± 0.06	0.08 ± 0.05	2.88 ± 2.34	3.22 ± 2.26
left.insula	0.12 ± 0.09	0.1 ± 0.1	3.06 ± 2.24	3.41 ± 3.64
right.caudal.middle.frontal	0.09 ± 0.08	0.05 ± 0.04	3.4 ± 3.31	2.15 ± 2.02
right.entorhinal	0.31 ± 0.26	0.11 ± 0.07	5.99 ± 4.18	2.43 ± 1.68
right.inferior.parietal	0.09 ± 0.06	0.17 ± 0.13	2.79 ± 1.73	3.77 ± 2.84
right.isthmus.cingulate	0.05 ± 0.05	0.08 ± 0.07	1.6 ± 1.47	2.39 ± 1.97
right.lateral.orbitofrontal	0.11 ± 0.09	0.07 ± 0.05	2.92 ± 2.52	1.98 ± 1.56
right.medial.orbitofrontal	0.15 ± 0.15	0.11 ± 0.1	3.95 ± 4.2	2.9 ± 3.03
right.parahippocampal	0.1 ± 0.08	0.11 ± 0.1	3.03 ± 2.41	5.07 ± 4.61
right.pars.opercularis	0.07 ± 0.06	0.11 ± 0.07	2.33 ± 2.3	3.56 ± 2.27
right.pars.triangularis	0.09 ± 0.08	0.11 ± 0.08	3.11 ± 3.06	4.1 ± 2.93
right.postcentral	0.09 ± 0.08	0.08 ± 0.06	4.27 ± 3.55	2.5 ± 1.96
right.precentral	0.09 ± 0.09	0.06 ± 0.05	3.59 ± 3.44	2.04 ± 1.91
right.rostral.anterior.cingulate	0.11 ± 0.1	0.13 ± 0.1	2.79 ± 3.2	4.19 ± 3.53
right.superior.frontal	0.08 ± 0.06	0.08 ± 0.06	2.89 ± 2.33	3.61 ± 2.99
right.superior.temporal	0.09 ± 0.09	0.09 ± 0.08	2.88 ± 3.33	3.17 ± 2.62
right.transverse.temporal	0.08 ± 0.06	0.11 ± 0.1	3.59 ± 2.65	2.85 ± 2.44

Table 3: ANTs CT.

Region	Absolute Difference (mm)		Percent Variability Error	
	Left	Right	Left	Right
left.caudal.anterior.cingulate	0.09 ± 0.09	0.07 ± 0.07	3.48 ± 3.26	3.02 ± 2.89
left.cuneus	0.04 ± 0.03	0.18 ± 0.15	1.81 ± 1.4	5.48 ± 4.68
left.fusiform	0.05 ± 0.05	0.06 ± 0.06	2 ± 1.74	2.42 ± 2.65
left.inferior.temporal	0.06 ± 0.05	0.09 ± 0.07	2.15 ± 1.97	3.47 ± 2.77
left.lateral.ocipital	0.04 ± 0.04	0.08 ± 0.06	2 ± 1.72	3.06 ± 2.36
left.lingual	0.05 ± 0.03	0.09 ± 0.07	2.19 ± 1.61	3.6 ± 2.71
left.middle.temporal	0.07 ± 0.05	0.12 ± 0.11	2.4 ± 1.87	4.18 ± 3.96
left.paracentral	0.06 ± 0.05	0.05 ± 0.04	2.34 ± 2.05	2 ± 1.66
left.pars.orbitalis	0.08 ± 0.07	0.05 ± 0.04	3.09 ± 2.68	2 ± 1.63
left.pericalcarine	0.06 ± 0.05	0.03 ± 0.03	3.27 ± 2.6	1.68 ± 1.55
left.posterior.cingulate	0.05 ± 0.04	0.04 ± 0.04	2.12 ± 1.7	1.65 ± 1.86
left.precuneus	0.05 ± 0.04	0.1 ± 0.09	2.12 ± 1.78	3.88 ± 3.68
left.rostral.middle.frontal	0.06 ± 0.05	0.06 ± 0.06	2.69 ± 2.15	2.46 ± 2.56
left.superior.parietal	0.05 ± 0.05	0.04 ± 0.03	2.28 ± 2.52	1.57 ± 1.18
left.supramarginal	0.05 ± 0.05	0.06 ± 0.06	2.19 ± 1.98	2.62 ± 2.29
left.insula	0.06 ± 0.04	0.08 ± 0.07	1.96 ± 1.47	3.23 ± 2.64
right.caudal.middle.frontal	0.07 ± 0.07	0.04 ± 0.03	2.96 ± 2.88	1.89 ± 1.59
right.entorhinal	0.2 ± 0.19	0.07 ± 0.06	5.83 ± 5.31	2.91 ± 2.29
right.inferior.parietal	0.05 ± 0.05	0.07 ± 0.05	2.22 ± 2.34	2.54 ± 2.03
right.isthmus.cingulate	0.06 ± 0.06	0.04 ± 0.04	2.67 ± 2.44	2.04 ± 1.82
right.lateral.orbitofrontal	0.07 ± 0.06	0.03 ± 0.02	2.51 ± 2.22	1.63 ± 1.02
right.medial.orbitofrontal	0.09 ± 0.08	0.06 ± 0.04	3.5 ± 3.31	2.09 ± 1.57
right.parahippocampal	0.11 ± 0.1	0.05 ± 0.05	4.2 ± 3.78	2.29 ± 2.03
right.pars.opercularis	0.06 ± 0.04	0.09 ± 0.06	2.29 ± 1.58	3.44 ± 2.35
right.pars.triangularis	0.06 ± 0.05	0.05 ± 0.04	2.64 ± 1.91	3.03 ± 2.42
right.postcentral	0.04 ± 0.04	0.06 ± 0.05	2.12 ± 1.8	2.3 ± 2.15
right.precentral	0.05 ± 0.06	0.05 ± 0.04	1.85 ± 2.28	1.98 ± 2.06
right.rostral.anterior.cingulate	0.12 ± 0.11	0.07 ± 0.05	4.45 ± 4.61	3.11 ± 2.3
right.superior.frontal	0.06 ± 0.05	0.04 ± 0.04	2.52 ± 2.14	1.96 ± 2.09
right.superior.temporal	0.05 ± 0.04	0.05 ± 0.04	1.82 ± 1.55	2.17 ± 1.67
right.transverse.temporal	0.09 ± 0.07	0.06 ± 0.05	3.49 ± 3.03	2.03 ± 1.66

Table 4: Freesurfer.

Region	Absolute Difference (mm)		Percent Variability Error	
	Left	Right	Left	Right
left.caudal.anterior.cingulate	0.06 ± 0.06	0.07 ± 0.06	2.1 ± 1.97	2.91 ± 2.93
left.cuneus	0.05 ± 0.06	0.14 ± 0.15	2.47 ± 2.4	3.82 ± 3.68
left.fusiform	0.06 ± 0.07	0.05 ± 0.06	1.69 ± 1.72	2.02 ± 2.28
left.inferior.temporal	0.09 ± 0.1	0.05 ± 0.04	2.47 ± 2.2	1.63 ± 1.19
left.lateral.ocipital	0.07 ± 0.08	0.1 ± 0.12	2.31 ± 2.07	3.53 ± 4.3
left.lingual	0.05 ± 0.06	0.13 ± 0.15	1.85 ± 1.78	4.65 ± 4.68
left.middle.temporal	0.09 ± 0.1	0.07 ± 0.07	2.89 ± 2.75	2.54 ± 2.13
left.paracentral	0.07 ± 0.06	0.05 ± 0.05	3.17 ± 2.91	2.02 ± 1.95
left.pars.orbitalis	0.06 ± 0.07	0.06 ± 0.09	2.55 ± 2.75	2.69 ± 3.8
left.pericalcarine	0.06 ± 0.06	0.05 ± 0.05	2.59 ± 2.32	2.57 ± 2.38
left.posterior.cingulate	0.05 ± 0.04	0.05 ± 0.06	1.79 ± 1.45	2.36 ± 2.38
left.precuneus	0.04 ± 0.04	0.08 ± 0.12	1.66 ± 1.41	2.37 ± 2.84
left.rostral.middle.frontal	0.09 ± 0.09	0.07 ± 0.07	3.5 ± 3.44	2.82 ± 3.1
left.superior.parietal	0.05 ± 0.06	0.07 ± 0.08	2.33 ± 2.7	2.49 ± 2.81
left.supramarginal	0.07 ± 0.06	0.05 ± 0.05	2.69 ± 2.36	2.29 ± 2.12
left.insula	0.07 ± 0.07	0.08 ± 0.09	1.97 ± 1.82	3.22 ± 3.56
right.caudal.middle.frontal	0.07 ± 0.07	0.04 ± 0.04	3.14 ± 3.13	1.94 ± 1.77
right.entorhinal	0.19 ± 0.13	0.06 ± 0.07	4.92 ± 2.92	1.84 ± 1.64
right.inferior.parietal	0.06 ± 0.06	0.1 ± 0.11	2.2 ± 1.93	2.84 ± 2.65
right.isthmus.cingulate	0.03 ± 0.03	0.06 ± 0.06	1.12 ± 0.92	2.15 ± 1.78
right.lateral.orbitofrontal	0.07 ± 0.09	0.04 ± 0.04	2.32 ± 2.56	1.51 ± 1.28
right.medial.orbitofrontal	0.1 ± 0.15	0.07 ± 0.1	3.18 ± 4.15	2.38 ± 3.05
right.parahippocampal	0.07 ± 0.08	0.06 ± 0.05	2.43 ± 2.58	2.98 ± 2.46
right.pars.opercularis	0.05 ± 0.05	0.08 ± 0.06	1.84 ± 1.89	3.05 ± 2.32
right.pars.triangularis	0.07 ± 0.08	0.07 ± 0.07	2.77 ± 3.12	3.11 ± 2.48
right.postcentral	0.05 ± 0.05	0.05 ± 0.04	2.35 ± 2.39	2.04 ± 1.43
right.precentral	0.05 ± 0.05	0.04 ± 0.05	2.17 ± 2.07	1.7 ± 1.72
right.rostral.anterior.cingulate	0.08 ± 0.1	0.08 ± 0.1	2.84 ± 3.19	3.16 ± 3.57
right.superior.frontal	0.07 ± 0.07	0.05 ± 0.05	2.97 ± 3.2	2.63 ± 2.51
right.superior.temporal	0.06 ± 0.08	0.06 ± 0.06	2.2 ± 3.32	2.42 ± 2.33
right.transverse.temporal	0.06 ± 0.06	0.07 ± 0.08	2.68 ± 2.62	1.81 ± 2.12

Table 5: ATITH.

Analysis	<i>r</i>	mean error (years)
Gray matter probability	0.92	6.4
Cortical thickness	0.90	7.25

Table 6: Correlation and mean error values for both the gray matter probability and cortical thickness *BrainAGE* evaluation.

Repeat scans of 40 subjects (20 Kirby subjects and 20 Oasis subjects) were used to determine the reproducibility of regional cortical thickness measurements.²⁰ Similar to the assessment given in [45], we show regional reproducible thickness measurements, T , in terms of the variability error:

$$\varepsilon = \frac{|T_{\text{scan}} - T_{\text{rescan}}|}{0.5 \times (T_{\text{scan}} + T_{\text{rescan}})}. \quad (3)$$

Error values (including absolute mean differences) for the 32 NIREP regions for both the Oasis and Kirby reproducibility data sets are given in Table 2. We also calculated the intra-class correlation coefficient (“ICC(2,1)” in the notation of [81]) to assess scan/rescan reliability which showed reliable agreement ($ICC = 0.98$). Additional regression testing exploring the effects of site, age, and gender demonstrated no statistically significant effect on regional mean thickness difference.

3.3. BrainAGE Evaluation

In [32], an estimation framework is presented for predicting apparent age from gray matter segmentation probability maps (denoted by the authors as *BrainAGE*). Given a normal age population spanning the age range of interest, the authors showed how kernel regression methods can be used to reliably estimate age. The basic processing pipeline includes gray matter segmentation from a subject’s T1, followed by affine registration to a common reference space (e.g. the MNI template), smoothing (8 mm FWHM), and downsampling (8 mm isotropic resolution). A principal components (PCA) model is constructed from the resulting aligned training image set. The images of both the training set and testing set are decomposed into the bases of the PCA model which form the feature set for relevance vector machine (RVM)-based learning and prediction, respectively.

We applied the BrainAGE framework to the gray matter probability maps derived from our pipeline. We also applied the same strategy to predicting age from our cortical thickness images.²¹ We randomly separated the images of each of the

four cohorts into approximately two equal subgroups (testing and training). Construction of the PCA model and decomposition of all images into the corresponding bases were performed on the training group using tools developed from the Insight Toolkit.²² We used the R package *kernlab*²³ package to train the RVM model and perform prediction. Results for both analyses are shown in Figure 6 (cf Figure 3 in [32]). The resulting predictions for both image sets are quite similar as demonstrated visually in Figure 3. The correlation coefficients and mean errors in Table 6 between the two approaches are also evidence of mutual corroboration.

3.4. Gender and Age Relationships with DiReCT Cortical Thickness

As discussed in the introduction, previous studies have demonstrated thickness variation with gender and age which is supported with the results of our study. Subdividing the cohort into training and testing subsets, we generate regression models from the training data for both gender and age and predict such biological relationships with the testing data. Stepwise model selection using Akaike Information Criterion (AIC) provides an optimal model with minimal parameters. This provides insight into regional thickness importance as well as the influence of confounds such as site.²⁴

We first tested the ability of regional thickness and cerebral volume to determine gender. In the notation of [98],²⁵ the initial binomial generalized linear model is

$$\begin{aligned} \text{GENDER} \sim & \text{AGE} + \text{VOLUME} + \sum_{i=1}^{32} T(\text{REGION}_i) \\ & + \sum_{i=1}^{32} T(\text{REGION}_i) : \text{AGE} \end{aligned} \quad (4)$$

where $T(\text{REGION}_i)$ is the average thickness value in REGION_i . The stepwise AIC model selection resulted in the pruned model

$$\begin{aligned} \text{GENDER} \sim & \text{VOLUME} + \sum_{i \in \alpha} T(\text{REGION}_i) + \\ & \sum_{j \in \beta} T(\text{REGION}_j) : \text{AGE} \end{aligned} \quad (5)$$

where the sets $\alpha = \{1, 2, 4, 7, 8, 10, 12, 14, 15, 19, 22, 24, 27, 29, 32\}$ and $\beta = \{7, 24, 27, 29\}$ (see Table 1). We then characterized the performance using a ROC curve (see Figure 7(a)) with $AUC = 0.86$ and 95% confidence interval = (0.84, 0.87).

Similarly, a testing/training data partitioning was used to generate a linear model and test age prediction by regressing

²⁰The R script used for this section is `reproducibility.R`. Data are located in the csv files `labelresultsK_pairwise.csv` (Kirby) and `labelresultsO_pairwise.csv` (Oasis).

²¹The R script used for this section is `brainAgeAnalysis.R`. Data are located in the following csv files:

- `trainingCorticalThicknessProjections.csv`,
- `testingCorticalThicknessProjections.csv`,
- `trainingBrainSegmentationPosterior2Projections.csv`, and
- `testingBrainSegmentationPosterior2Projections.csv`.

²²http://www.itk.org/Dxygen/html/classitk_1_1ImagePCADecompositionCalculator.htr

²³<http://rss.acs.unt.edu/Rdoc/library/kernlab/html/rvm.html>

²⁴The R scripts for this analysis are `genderThicknessRegression.R` and `ageThicknessRegression.R`. Data are located in the following csv files: `labelresultsI.csv` (IXI), `labelresultsK.csv` (Kirby), `labelresultsN.csv` (NKI), and `labelresultsO.csv` (Oasis).

²⁵For consistency with R and other computational packages, we replace ‘.’ with ‘:’ to denote all interaction terms.

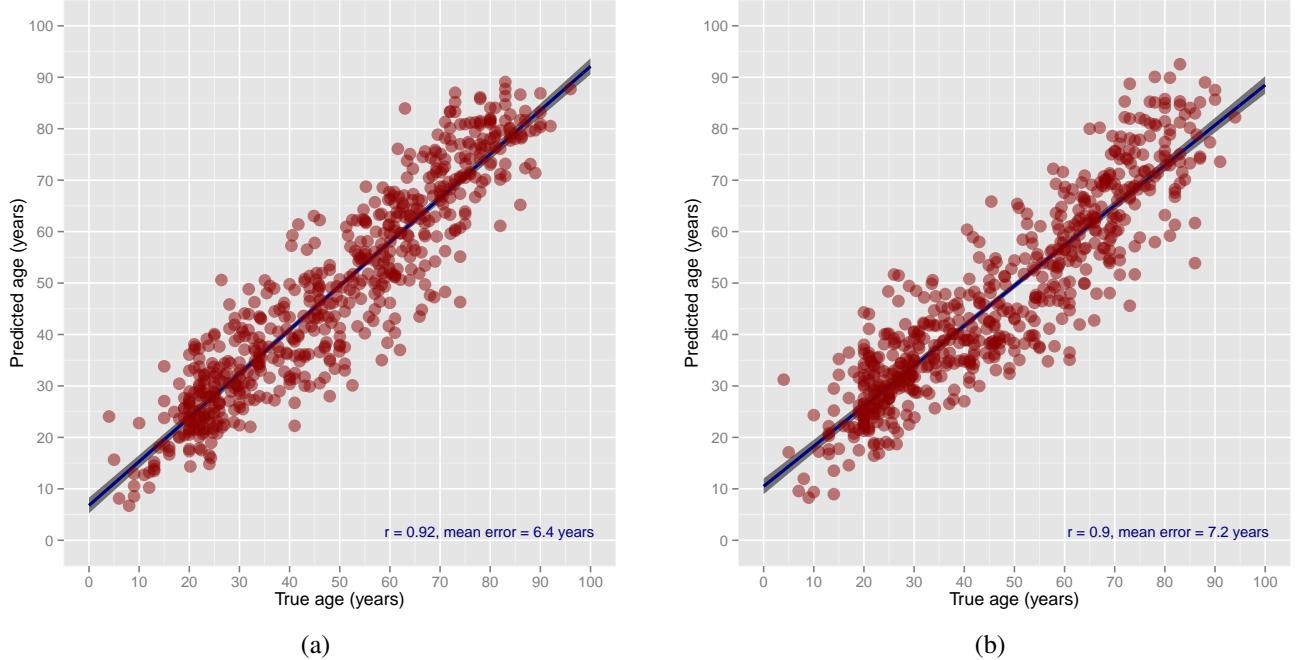


Figure 6: Results of RVM-based age prediction using (a) gray matter probability maps as in [32] and (b) cortical thickness maps both of which are derived from the previously described workflow.

on regional thickness, cerebral volume, gender, and site resulting in the formula

$$AGE \sim VOLUME + SITE + GENDER + \sum_{i=1}^{32} T(REGION_i) \quad (6)$$

with stepwise AIC selection producing the following model

$$AGE \sim SITE + \sum_{i \in \gamma} T(REGION_i) \quad (7)$$

where the set $\gamma = \{2, 3, 7, 9, 12, 13, 15, 16, 21, 23, 25, 26, 27, 28, 31\}$. The predicted age vs. true age plot is given in Figure 7(b) with Pearson correlation coefficient = 0.77 with a mean age error of 10 years.

3.5. Gender Structural Connectivity Across Age Using Cortical Thickness

As mentioned in the Introduction, cortical thickness has been used to determine structural connectivity relationships in the brain where strong correlations in regional cortical thickness values across subjects provide evidence for anatomical connectivity [41, 12, 40]. Specifically, networks of neuronal regions are thought to have small-world network properties [86] in which clustered subnetworks are sparsely connected to other such clusters. Measures such as the clustering coefficient (or local transitivity) and mean shortest path length [95], are used to characterize networks in terms of their small-worldness. Although the principal purpose of this work is to showcase the publicly available ANTs cortical thickness

pipeline and its performance on open data (and not necessarily explore the deeper neuroscience implications of the results), we use the compiled cortical thickness data to briefly sketch the temporal variation in gender-based small-world networks of the brain.²⁶ These results demonstrate potential future exploration of the neuroscience-related implications of our proposed pipeline complete with the requisite tools.

At each age between 10 and 90 years (in increments of 5), the weighted correlation matrix for each gender is calculated from the thickness residuals (modeling the imaging acquisition site and total brain volume as covariates). An undirected graph ($V \in \{\text{NIREP regions}\}, E \in \{\text{all NIREP pairings}\}$) is constructed from the correlation matrix where the graph density is specified at 25%, i.e. only the nodal adjacencies corresponding to the top 25% correlation values are used to create edges. The local transitivity for a given vertex, v_i , with k_i neighbors of the resulting graph is calculated from

$$\text{transitivity}(v_i) = \frac{|\{e_{jk} : v_j, v_k \in V, e_{jk} \in E\}|}{k_i(k_i - 1)/2}. \quad (8)$$

Informally, this quantifies the proportion of edges between the neighbors of v_i to the total number of possible edges in the neighborhood to quantify the proximity of the neighborhood to a complete graph.

Age-wise transitivity gender differences were calculated with statistical significance determined from permutation testing

²⁶The R scripts for this analysis are `genderStructuralConnectivity.R`. Data are located in the following csv files: `labelresultsI.csv` (IXI), `labelresultsK.csv` (Kirby), `labelresultsN.csv` (NKI), and `labelresultsO.csv` (Oasis).

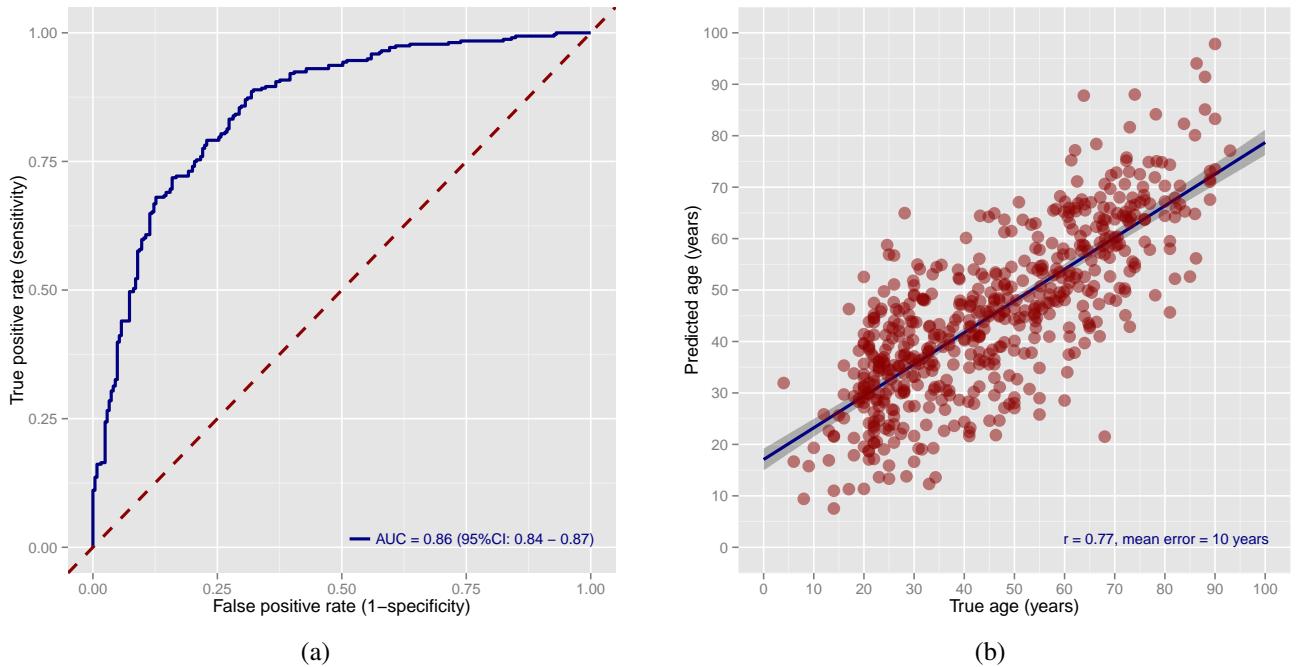


Figure 7: (a) ROC curve based on a gender prediction model using total brain volume and regional thickness values (coupled with cross terms with age—cf Eqn. 5). (b) Correlation plot for age prediction using regional thickness.

($n = 1000$ permutations). These statistically significant regional results are given in Table 7. We also visualize these connectivity networks for both female and male within the brain space and as phylogenetic radial trees in Figures 8 and 9, respectively.

3.6. Quality Assessment Measures

Considering the multiple components of the thickness protocol, there are several points at which a single subject processing failure can occur. Initial mis-registration of the template can produce erroneous brain extraction results which severely affects the remainder of the processing workflow. Similarly, incorrect segmentation results will negatively impact the cortical thickness measurements. However, determining which individual subjects did not process correctly is a time-consuming task. Although much of our quality assessment involved inspection of each individual subject overlaid with the various processed images, an additional tool (which we dub the *brain constellation map*) proved to be very useful as it provided a quick assessment of results over the entire cohort and allowed for immediate identification of problematic cases.²⁷ A star plot of the thickness residuals ($T(\text{REGION}_i) \sim \text{AGE} + \text{VOLUME} + \text{SITE}$) for each subject is in Figure 10. Subjects have been organized by increasing age and unique colors are assigned to each data set (IXI = red, Kirby = green, NKI = blue, Oasis = orange) which permits visualization of age distribution with data set. Additionally, for each subject we test for non-normality using the Shapiro-Francia normality test [73] which provides

an additional measurement for quality assessment. For our work, which is the largest cortical thickness study of which we are aware, non-normality (per Shapiro-Francia) of the cortical thickness residuals was an excellent indicator of processing failures during the tuning of the pipeline.

4. Discussion

In the absence of ground truth, we used data inspection, prediction of demographic variables and reproducibility to evaluate the ANTs cortical thickness pipeline as it applies to large-scale data. Each result complements the other by relying upon different information provided by the pipeline. Here, we discuss and contextualize each of the major evaluation points.

4.1. Computation Time and Failure Rate

Computation time for the registration and segmentation components of the 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, which has concerned statisticians for over half a century [38], here.

²⁷The script to create the brain constellation map is `makeThicknessStarsPlot.R`.

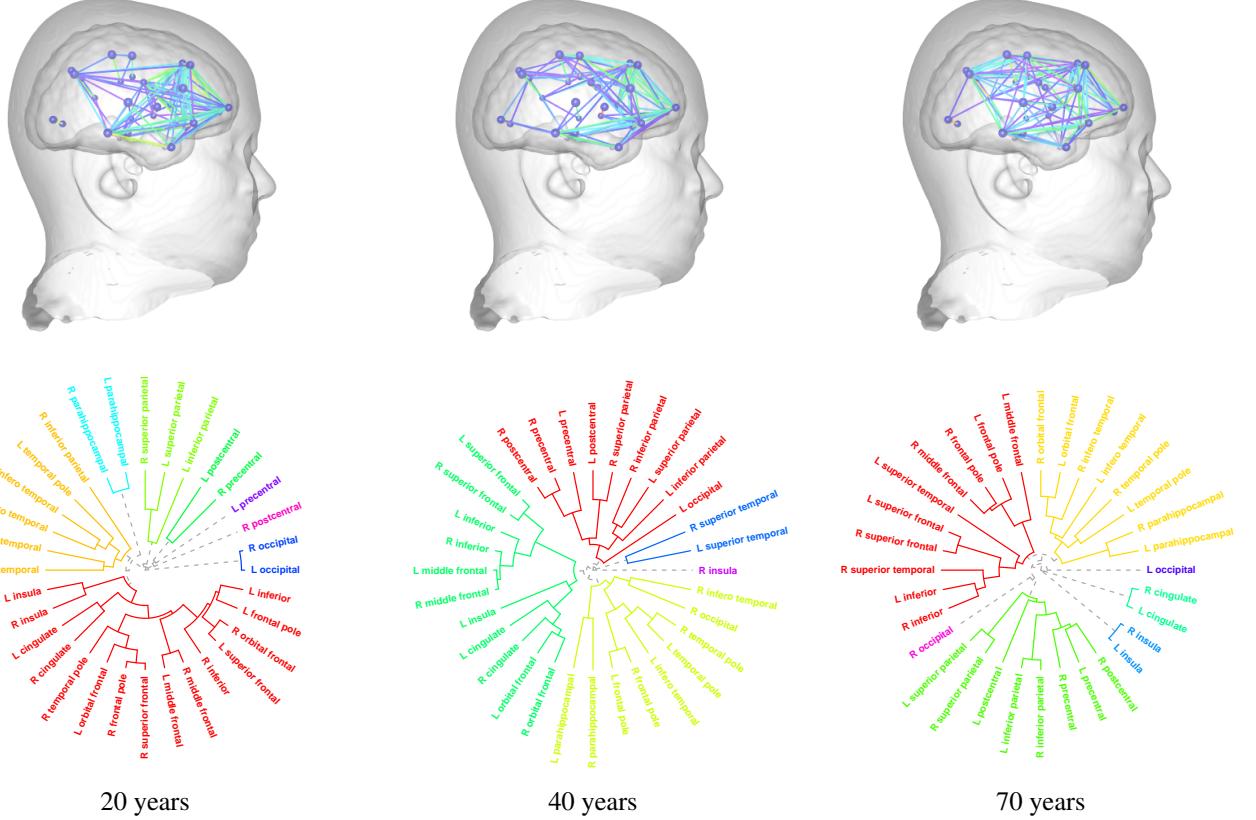


Figure 8: Visual illustration of the female thickness networks for ages 20, 40, and 70 years. Community relationships between regions are depicted both in brain space (top row) and as a radial phylogenetic tree where colors denote neighborhoods (bottom row).

4.2. Reproducibility 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 that T1-weighted images are susceptible to short-term alterations due to blood flow [33, 74, 100], this strategy is not ideal. However, related tools have looked at this question. An independent evaluation of the FreeSurfer pipeline shows good repeatability measurements [45]. The authors report FreeSurfer reproducibility in the range of 1.5 - 5% depending on the site and region of the brain. The CLADA pipeline showed the ability to detect changes as small as 1 millimeter and showed good agreement with FreeSurfer [64]. Very recently, it was suggested that 3T MRI consistently overestimates cortical thickness [58]. 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 with these methods. 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 [58, 45]) in order to better compare surface-based and volume-based thickness reliability across different

populations and age ranges.

4.3. BrainAGE

We used the ANTs cortical thickness pipeline as input to the BrainAGE algorithm implemented according to [32]. The original BrainAGE algorithm was evaluated on a population 19-86 years of age with n=650 and produced a mean error of 5 years with a correlation between the testing subject age and real age of 0.92. Our analysis, trained on half of our subjects and tested on the other half, produced an identical correlation and a mean absolute error of 6.4 years. While the age range in our study is overall similar, we have over 61 subjects younger than 20 years and several dozen subjects with probable Alzheimer’s Disease. Our study also draws data from multiple scan sites (even within the same cohort, such as IXI). While the true mechanisms underlying BrainAge are unknown, the primary driving forces from an image processing perspective are accurate affine registration and gray matter probability images with tissue-derived information that is relevant to the subject’s age. Therefore, this sub-study serves as validation of the findings by [32] as well as the probabilistic segmentation and affine registration components of our pipeline.

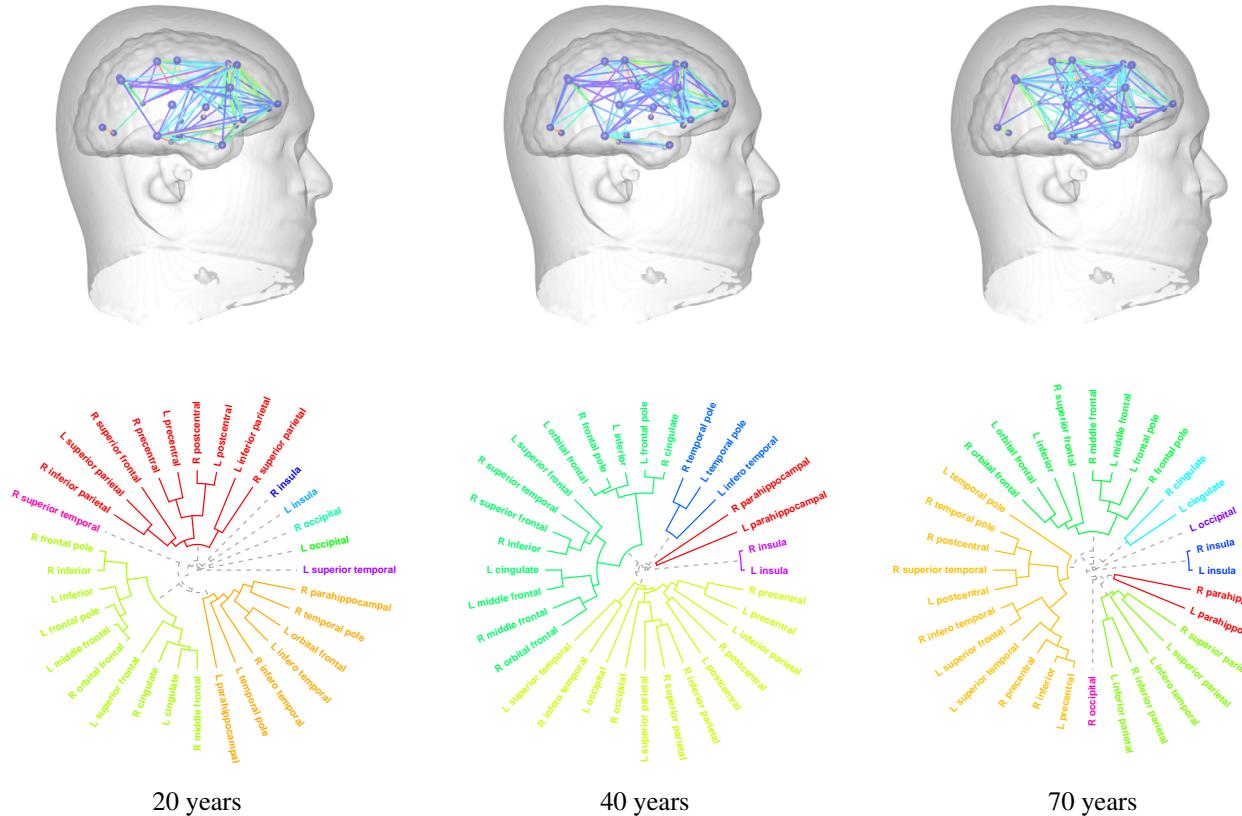


Figure 9: Visual illustration of the male thickness networks for ages 20, 40, and 70 years. Community relationships between regions are depicted both in brain space (top row) and as a radial phylogenetic tree where colors denote neighborhoods (bottom row).

4.4. General Linear Model Gender & Age Prediction

We also use a training and testing paradigm to evaluate whether a more traditional prediction model, general linear modeling, is capable of predicting age or gender from our data. The general linear model has the advantage, in comparison to the relevance vector machine, of being interpretable. That is, the models reveal the specific predictors that are most valuable.

4.4.1. Site

The site variable does not emerge as an important predictor. This suggests that our brain volume measurements (which are the most dominant predictor in the model) are not affected by data collection site. Thickness measurements also enter the model. However, there are several possible regions that may be used to improve the classification accuracy and these particular regions are not necessarily uniquely predictive. Several components that encode an interaction between age and thickness as predictors of age also consistently enter the model.

4.4.2. Age

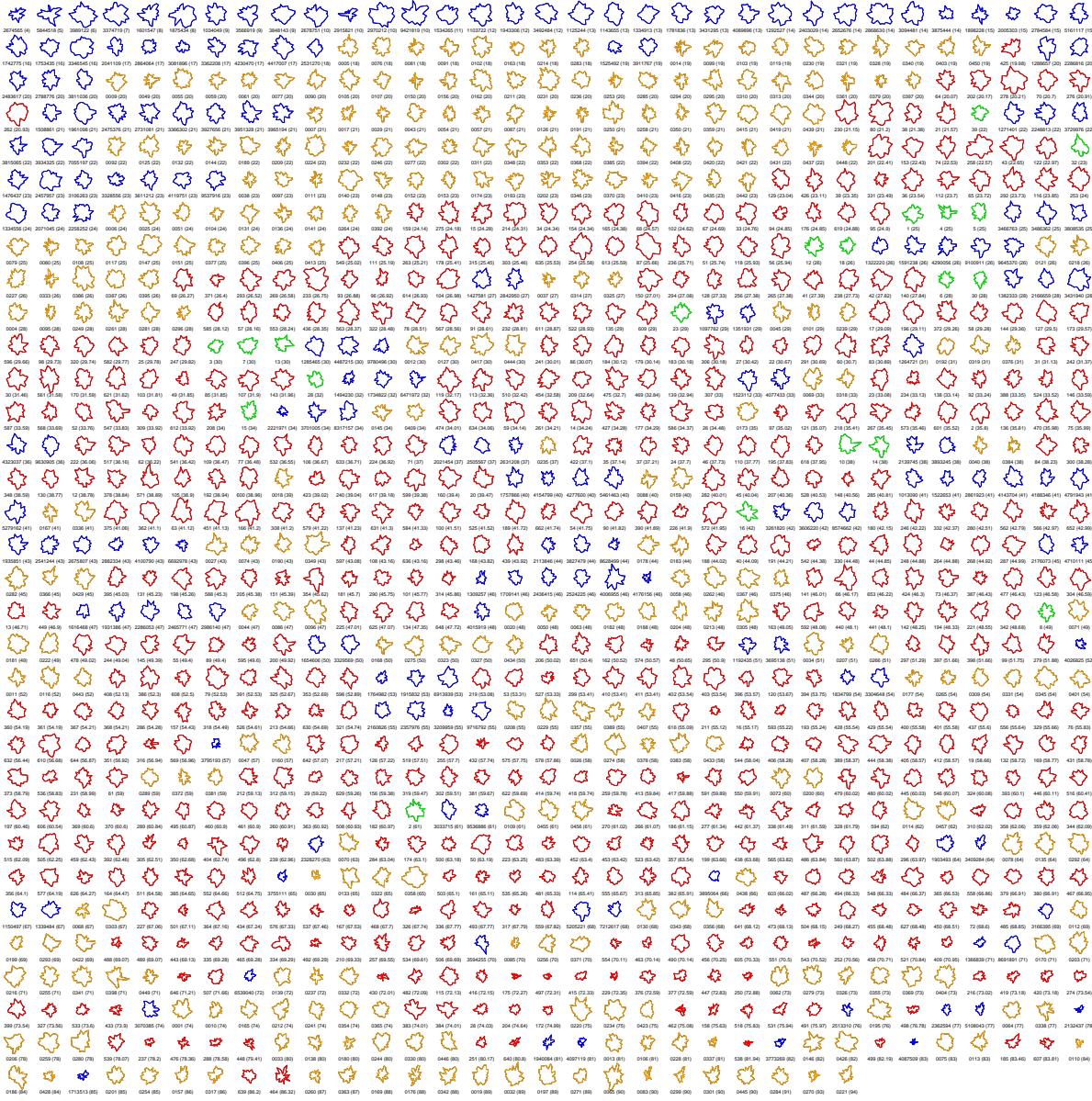
The interpretable linear regression model, based on cortical thickness, finds a correlation with age of 0.77 and a mean absolute error of 10 years. The idea, here, is not to produce the best possible prediction (that was the purpose of BrainAge) but

to identify whether putative regions associated with aging survive within our model and also to determine the degree to which data collection site impacts the prediction. Indeed, our variable selection process includes SITE in the model suggesting that our analysis pipeline may not completely remove the effect of SITE on thickness measurements. A second reason that SITE is included is that it does, indeed, covary with age. That is, the Kirby and NKI subjects are overall younger than the OASIS and IXI subjects. Thus, SITE is a confounding variable in this dataset. Despite this issue, regions previously implicated in aging are included in the model such as occipital lobe, cingulate cortex, several temporal regions, inferior frontal gyrus (Broca's area) and pre and post-central gyri. The superior and middle frontal gyri, known to be relatively preserved in aging, are absent from the model, as expected. Thus, despite a significant impact of SITE on age prediction, biologically plausible cortical regions emerge in our traditional regression model for age. Regions are in agreement, generally, with a previous large-scale study with a similar age range [36].

4.5. Network Small-Worldness

This result is more exploratory and shows the potential of large datasets such as this one in generating new descriptive and quantitative hypotheses about aging and gender.

Brain Constellation Map of Thickness Residuals



Age	Significant Regions	Transitivity Diff.	p-value
10	L superior frontal	0.41	0.016
	R superior parietal	0.40	0.042
15	L superior frontal	0.28	0.046
	R superior frontal	0.33	0.004
	L inferior	0.33	0.027
20	R superior frontal	0.31	0.013
25	—	—	—
30	L superior temporal	1.0	0.0
	L inferior	0.42	0.045
	L orbital frontal	0.5	0.003
	L superior parietal	0.6	0.003
35	R superior temporal	0.81	0.01
	L inferior parietal	0.27	0.039
40	R infero temporal	0.21	0.044
	R inferior parietal	0.68	0.001
45	R superior frontal	0.33	0.022
	R inferior	0.42	0.021
50	R parahippocampal	0.57	0.026
	R superior frontal	0.34	0.017
55	L occipital	0.56	0.026
	R occipital	0.87	0.009
60	R occipital	0.72	0.001
	R orbital frontal	0.44	0.016
65	—	—	—
70	—	—	—
75	R parahippocampal	0.83	0.021
	L frontal pole	0.40	0.021
	L inferior parietal	0.33	0.033
80	L middle frontal	0.39	0.022
	R inferior	0.53	0.001
	R inferior parietal	0.34	0.003
	R postcentral	0.30	0.018
85	R inferior parietal	0.26	0.017
90	—	—	—

Table 7: Regional gender differences in transitivity with age. Statistical significance was determined using permutation testing ($n = 1000$).

4.6. Brain Constellation Maps

Finally, there is no substitute for looking at one’s data. However, few researchers have the resources to support detailed inspection of over 1200 subjects with cortical thickness measurements spread across a three-dimensional volume. Therefore, we developed “constellation maps” which serve to quickly summarize the cortical thickness measurements across the entire population. These maps are derived from the full subjects by cortex matrix of thickness measurements across all subjects and all 32 NIREP regions. Each column (NIREP region) is normalized in the range 0 to 1 across all subjects. Because of this processing, the constellation map allows one to easily identify non-normal (normal in the statistical sense) distributions of thickness residuals in the population which, for the data analyzed in this study, proved to be indicative of processing failure during the software refinement stage.

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. We note that in addition to making the scripts, csv result files, and cortical thickness maps available to researchers, posterior probability images are also offered permitting additional quantification such as voxel-based morphometry [3].

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., Hömké, 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., Jun 2000. Voxel-based morphometry—the methods. *Neuroimage* 11 (6 Pt 1), 805–21.
- [4] Ashburner, J., Friston, K. J., Jul 2005. Unified segmentation. *Neuroimage* 26 (3), 839–51.
- [5] 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 dat. *Human Brain Mapping*.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [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] Christensen, G. E., Geng, X., Kuhl, J. G., Bruss, J., Grabowski, T. J., Pirwani, I. A., Vannier, M. W., Allen, J. S., Damasio, H., 2006. Introduction to the non-rigid image registration evaluation project (NIREP).

- In: Proceedings of the Third international conference on Biomedical Image Registration. WBIR'06. Springer-Verlag, Berlin, Heidelberg, pp. 128–135.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Dale, A. M., Fischl, B., Sereno, M. I., Feb 1999. Cortical surface-based analysis. i. segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–94.
- [18] Das, S. R., Avants, B. B., Grossman, M., Gee, J. C., Apr 2009. Registration based cortical thickness measurement. *Neuroimage* 45 (3), 867–79.
- [19] 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.
- [20] 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.
- [21] Dickerson, B. C., Bakkour, A., Salat, D. H., Feczkó, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., Sperling, R. A., Atri, A., Growdon, J. H., Hyman, B. T., Morris, J. C., Fischl, B., 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] Evans, A. C., Janke, A. L., Collins, D. L., Bailet, S., Aug 2012. Brain templates and atlases. *Neuroimage* 62 (2), 911–22.
- [25] Fedorov, A., Li, X., Pohl, K. M., Bouix, S., Styner, M., Addicott, M., Wyatt, C., Daunais, J. B., Wells, W. M., Kikinis, R., 2011. Atlas-guided segmentation of vervet monkey brain MRI. *Open Neuroimag J* 5, 186–97.
- [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 alcohol 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] Franke, K., Ziegler, G., Klöppel, S., Gaser, C., Alzheimer's Disease Neuroimaging Initiative, Apr 2010. Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: exploring the influence of various parameters. *Neuroimage* 50 (3), 883–92.
- [33] 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.
- [34] Frøkjær, 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.
- [35] Gernsbacher, M. A., Jan 2007. Presidential column: The eye of the beholder. *Observer* 20 (1).
- [36] Groves, A. R., Smith, S. M., Fjell, A. M., Tamnes, C. K., Walhovd, K. B., Douaud, G., Woolrich, M. W., Westlye, L. T., Oct 2012. Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure. *Neuroimage* 63 (1), 365–80.
- [37] 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.
- [38] Hampel, F. R., March 2001. Robust statistics: A brief introduction and overview. *Research Report* 94, Eidgenössische Technische Hochschule.
- [39] 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.
- [40] He, Y., Chen, Z., Evans, A., Apr 2008. Structural insights into aberrant topological patterns of large-scale cortical networks in alzheimer's disease. *J Neurosci* 28 (18), 4756–66.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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., Alessandrin, 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*.
- [46] 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.
- [47] Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'aghi, 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.
- [48] 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.
- [49] 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.
- [50] Kovacevic, J., 2006. From the editor-in-chief. *IEEE Trans Imag Process* 15 (12).
- [51] Kühn, S., Schubert, F., Gallinat, J., Dec 2010. Reduced thickness of medial orbitofrontal cortex in smokers. *Biol Psychiatry* 68 (11), 1061–5.
- [52] 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.
- [53] 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.
- [54] 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.
- [55] 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.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] 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.
- [63] 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.
- [64] Nakamura, K., Fox, R., Fisher, E., Jan 2011. CLADA: cortical longitudinal atrophy detection algorithm. *Neuroimage* 54 (1), 278–89.
- [65] Nesvág, R., Lawyer, G., Varnás, 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] Royston, P., Jan 1993. A pocket-calculator algorithm for the Shapiro-Francia test for non-normality: an application to medicine. *Stat Med* 12 (2), 181–4.
- [74] 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.
- [75] 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.
- [76] 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.
- [77] Selemion, 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.
- [78] 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.
- [79] 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.
- [80] 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.
- [81] Shrout, P. E., Fleiss, J. L., Mar 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86 (2), 420–8.
- [82] 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.
- [83] Smith, S. M., Nov 2002. Fast robust automated brain extraction. *Hum Brain Mapp* 17 (3), 143–55.
- [84] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, 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.
- [85] 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.
- [86] Sporns, O., Chialvo, D. R., Kaiser, M., Hilgetag, C. C., Sep 2004. Organization, development and function of complex brain networks. *Trends Cogn Sci* 8 (9), 418–25.
- [87] Talairach, J., Tournoux, P., 1988. Co-planar stereotaxic atlas of the human brain: 3-Dimensional proportional system—An approach to cerebral imaging. Thieme.
- [88] 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.
- [89] 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.
- [90] 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).
- [91] 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*.
- [92] 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.
- [93] 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*.
- [94] Ward, B. D., 1999. Intracranial segmentation. Tech. rep., Medical College of Wisconsin, <http://afni.nimh.nih.gov/pub/dist/doc/3dIntracranial.pdf>.
- [95] Watts, D. J., Strogatz, S. H., Jun 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (6684), 440–2.
- [96] 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.
- [97] 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.
- [98] 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.
- [99] 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.
- [100] 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.
- [101] 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.
- [102] Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., Gerig, G., Jul 2006. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31 (3), 1116–28.
- [103] 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.
- [104] 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.