

## Precision Brain Morphometry: Feasibility and Opportunities of Extreme Rapid Scans

Jared A. Nielsen<sup>1,2</sup>, Ross W. Mair<sup>2,3</sup>, Justin T. Baker<sup>4,5</sup>, and Randy L. Buckner<sup>1,2,3,5</sup>

1. Department of Psychology, Harvard University, Cambridge, MA 02138
2. Center for Brain Science, Harvard University, Cambridge, MA 02138
3. Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Charlestown, MA 02129
4. Psychotic Disorders Division, McLean Hospital, Belmont, MA 02478
5. Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114

Abbreviated title: Extreme Rapid Scans

Key words: MRI, atrophy, volumetrics, T1, brain structure

Correspondence should be addressed to either of the following: Dr. Jared A. Nielsen, Email: [jarednielsen@fas.harvard.edu](mailto:jarednielsen@fas.harvard.edu) ; or Dr. Randy L. Buckner, Email: [randy\\_buckner@harvard.edu](mailto:randy_buckner@harvard.edu).

## Highlights

- High resolution structural MRI scans can be acquired in about a minute.
- Rapid structural scans yield stable automated morphometric measures.
- Averaged morphometric estimates from rapid scans do better than equivalent duration long scans.
- Cluster scanning across sessions can obtain a within-subject cortical thickness error of <10um.
- Rapid structural scans will save tremendous amounts of valuable scan time if widely adopted.

## Abstract

The traditional approach to achieve optimal structural brain morphometry is to increase spatial resolution and maintain a high signal-to-noise ratio within a single scanning session using long acquisitions. Here we explore alternative approaches that use multiple rapid structural scans and averaging to achieve stable morphometric estimates. Effects of sequence, resolution, acceleration, and number of scans within- and between-sessions were varied across six studies that included 1,025 scans. Single ~1-min magnetization-prepared rapid gradient echo (MP-RAGE) scans that use wave-controlled aliasing in parallel imaging (Wave-CAIPI) encoding yield morphometric estimates nearly as accurate as those obtained with traditional long scans. Averaging estimates from multiple rapid scans reduces error below traditional approaches using less total acquisition time. Test-retest reliability within- and between-sessions for ~2 min multi-echo MP-RAGE scans revealed that a source of error is the idiosyncratic positioning of the subject's head between sessions. Cluster scanning over multiple tightly-spaced sessions mitigates this effect and achieves cortical thickness estimates with less than 10 $\mu$ m global (30 $\mu$ m regional) error. Test-retest error for the hippocampus falls below 0.6%. Additional sources of error include coil heating when scans are acquired rapidly within the same session and change in console software between sessions. Rapid acquisitions allow novel experimental designs that can use cluster scanning to achieve precise longitudinal assessment within the individual, build in robustness by affording redundant acquisitions, and reduce the structural scanning burden to a minute when anatomical registration is the goal. The field should consider replacing long structural scans with fast alternatives.

## Introduction

Morphometric analysis of brain structure is central to the study of development and neurodegeneration, and remains the mainstay of MRI protocols exploring neuropsychiatric illness and individual differences (Lerch et al., 2017). Nearly all functional MRI studies acquire a structural MRI image for volumetric and surface-based anatomical registration, and many utilize the anatomical details for reference during analysis.

The scanning resources devoted to structural imaging are substantial. As an example, Harvard Center for Brain Science, a moderate load single human scanner research facility, acquired more than 2,000 structural scans in 2017 for a total of over 200 hours of scan acquisition. Across the numerous similar and larger facilities, the scanning time dedicated to structural scans exceeds many tens of thousands of hours for research protocols and millions of hours for clinical assessment. Given their central place in MRI research protocols, it is useful to ask what kinds of scans are typically acquired and then explore whether there are alternatives that can achieve the same goals more quickly.

In this paper we extensively explore rapid structural protocols to illustrate how these scans can replace and improve upon longer acquisitions. Another purpose for evaluating rapid structural scans is that they allow new strategies that sequentially repeat many acquisitions to increase morphometric precision and make the study tolerant to lost or low quality scans. Our work builds upon prior developments that have expanded how structural scanning can be approached in the context of morphometric measurement. We summarize these developments first and then introduce the goals of the present work.

### **Emergence of Extremely Rapid Structural Scans**

While unimaginable a decade ago, it is now possible to obtain structural scans usable for quantitative morphometry in about a minute. The path to a one-minute structural scan followed a sequence of advances over the past three decades. Initial anatomical images for brain morphometry (e.g., Jack et al., 1989; 1992) used the Spoiled Gradient Recalled Echo technique (SPGR) – also known as Fast Low Angle SHot (FLASH; Frahm et al., 1986; Crawley et al., 1988). These images provided good contrast but required targeted alignment to measure small structures. For example, when focused on

hippocampal measurement, image slices (4mm) were aligned perpendicular to the axis of the hippocampus to obtain consistent sampling between individuals (Jack et al., 1989).

Separate sagittal acquisitions were used to trace the contour of the inner table for measurement of the intracranial volume. Later 3D variants allowed multiple cortical regions with complex geometry to be sampled and manually traced (e.g., Raz et al., 1997 used 1.3mm resolution). A workhorse for many years, these anatomical images set the standard for quantitative morphometry and context for later improvements.

A key development was the magnetization-prepared rapid gradient echo technique, or MP-RAGE (Mugler and Brookeman, 1990; Runge et al., 1991). The MP-RAGE adds an inversion pulse to the rapid low-flip-angle readout train of the SPGR to provide images of neuroanatomy with high contrast between the gray and white matter. In a seminal study, Sheline et al. (1996) quantitatively analyzed MP-RAGE scans using stereology to show hippocampal volume is reduced in major depression. Further refinements to the MP-RAGE technique included acquiring multiple echoes with a wider acquisition bandwidth to reduce susceptibility distortion (multi-echo MP-RAGE, van der Kouwe et al., 2008) and the use of additional readout trains following inversion to produce multiple images with varying contrast levels (MP2RAGE, Marques et al., 2010). While these sequence variations allow for a wider set of image contrast properties, they do not address the lengthy time required for image acquisition, which often led to the use of non-isotropic voxels (e.g., Sheline et al., 1996 used 1x1x1.25mm voxels). MP-RAGE acquisition time is about 10 minutes for a 1mm<sup>3</sup> brain scan without the use of in-plane acceleration techniques.

Aided by the advent of parallel-array head coils (Wiggins et al., 2006), strategies for in-plane acceleration were a major breakthrough for reducing scan time. These techniques include SENsitivity Encoding (SENSE; Pruessmann et al., 1999) and Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA; Griswold et al., 2002). Two-fold in-plane acceleration has led to a roughly 6-minute scan with 1mm isotropic resolution to become the most widely used option for T1-weighted structural imaging. For example, the “FreeSurfer recommended” protocol for morphometry is of this type (<https://surfer.nmr.mgh.harvard.edu/fswiki>) and the Alzheimer’s Disease Neuroimaging Initiative (ADNI) adopted such variants in their ADNI-2 and ADNI-3 waves of data collection (Jack et al., 2010; Jack et al., 2015). The recently launched Adolescent Brain

Cognitive Development (ABCD) study uses such acquisitions ranging in duration from 5'38" (Philips scanner platform) to 7'12" (Siemens scanner platform)(ABCDstudy.org). Using a protocol that is likely close to the limit of in-plane acceleration, the UK Biobank Study, which is acquiring brain imaging data on 100,000 individuals, achieved a 4'54" scan (Miller et al., 2016).

The floor in speed improvement results because the noise-amplification (g-factor) in GRAPPA-accelerated scans yields a limiting return with higher acceleration factors. That is, as one accelerates further, the image SNR degrades beyond a usable level. One option within this sequence class, which we explore in this paper, is to speed scanning further by reducing acquisition resolution. For example, as discussed below, a 2'12" scan viable for automated morphometry can be obtained by reducing resolution to 1.2mm (Mair et al., 2012; Holmes et al., 2015). However, further advances are needed to achieve high-resolution with maintained SNR.

Reconstruction techniques employing Controlled Aliasing in Parallel Imaging Results in Higher Acceleration (CAIPIRINHA, Breuer et al., 2006) have led to another set of breakthroughs in acquisition time reduction in a variety of MRI neuroimaging modalities. Most notably, this has been adopted widely in the Simultaneous Multi-Slice (or multiband) technique for functional and diffusion imaging (Moeller et al., 2010; Setsompop et al., 2012) where multiple 2D-echo planar slices can be acquired simultaneously with little noise penalty. For 3D imaging techniques in high-array coils, such as those used for MP-RAGE structural scans, CAIPI-accelerations can be applied in multiple dimensions with much-reduced noise-amplification compared to conventional Cartesian accelerations, such as GRAPPA. Bilgic et al. (2015) pioneered the use of the Wave-CAIPI technique where non-cartesian corkscrew-trajectories through k-space coupled with CAIPI accelerations result in substantial reductions in the duration of 3D structural scans.

These technical developments fundamentally change options for rapid structural scanning. In a dramatic recent demonstration, Polak et al. (2018) achieved an MP-RAGE protocol based on wave-CAIPI that yielded a usable 1mm isotropic, 3D T1-weighted brain image in less than 90 seconds. The noise amplification (g-factor) was below 1.5 – a value less than that for four-fold GRAPPA acceleration (Wiggins et al., 2006). In the present

paper, we demonstrate the feasibility of extreme rapid scans for automated morphometric analysis.

### **Subtle Head Motion Influences Morphometry**

While structural scans have long been excluded for significant motion artifacts (e.g., Raz et al., 1997; Marcus et al., 2007), structural images were long thought to be relatively immune to subtle motion that did not manifest as visible artifacts in the images.

Quantitative structural measures have recently been demonstrated to be sensitive to subject motion even when artifacts such as ringing are minimally apparent. This is problematic given known differences in the tendency to move between age groups and between patients and controls.

Reuter et al. (2015) studied the effects of intentional periodic subject motion on morphometric measures as estimated by three widely available tools: voxel-based morphometry in the VBM8 toolbox (Gaser et al., 2014); cortical thickness analysis in Freesurfer 5.3 (Fischl et al., 1999a; 1999b); and percent brain volume change in FSL SIENA (Smith, 2002). Quantitative morphometric measures were reduced when participants moved more, even when the data were screened to include only those structural images that passed visual quality control. Tisdall et al. (2016) demonstrated that prospective motion correction, based on subject-position tracking using rapid volumetric navigators, could dramatically reduce the effects of motion on these morphometric measures, although not eliminate the effect entirely (See also Igata et al., 2017). In a recent study that targeted motion under typical acquisition procedures, Alexander-Bloch et al (2016) used motion estimates from functional MRI scans in the same sessions to show that children who tend to move during functional scans show biased morphometric estimates for their structural scans. These collective observations make clear that accuracy and stability of morphometric estimates depend on subjects remaining still.

Sensitivity to head motion has challenged the field to rethink approaches to structural scans. One way to reduce the impact of head motion is to reduce the duration of the scan and subject burden, thereby lessening the opportunity for movement. Another way to reduce the impact of motion is to track the level of movement during each acquisition, such as through the insertion of navigators in the main sequence (Welch et al.,

2002; van der Kouwe et al., 2006). External trackers, either with optical or NMR-based measures can also be used for this purpose. (Zaitsev et al., 2006; Aranovitch et al., 2018). In this manner, structural data can be scrubbed or matched across individuals in ways that parallel strategies now commonly adopted in the analyses of functional MRI data. Additionally, one can reacquire motion-contaminated portions of the structural acquisition (e.g., White et al., 2010; Tisdall et al., 2012; Tisdall et al., 2016). Such active motion-correction approaches are beyond the scope of this paper but offer a strategy to quantitatively monitor motion during the scan.

### **Rapid Scans May Be Viable for Quantitative Morphometry**

Automated quantitative morphometry may be viable for images that have lower resolution and less SNR than the presently and commonly used long-duration 1mm isotropic scans. We stumbled upon this possibility serendipitously. In a past effort, the need to reduce subject burden caused us to implement a 2'12" multi-echo MP-RAGE sequence that used 1.2mm isotropic voxels with a partial-fourier acquisition of 6/8 (reducing the true resolution even further) and a GRAPPA acceleration factor of 4 (the larger voxels bring back signal that is lost by the four-fold acceleration) (Holmes et al., 2015). The expectation was that the images would be good enough for reference and registration purposes but fall short when attempting to derive automated morphometric properties such as subcortical volume measures (e.g., hippocampus) and thickness. What was surprising is that morphometric measures from the rapid scans were strongly correlated with those derived from 1.0mm longer MP-RAGE acquisitions ( $r^2 > 0.9$  for most measures; Mair et al., 2012; Holmes et al., 2015).

The unexpected viability of rapid, low resolution scans for automated morphometry is surprising given the field's trend toward higher resolution images that maximize SNR by keeping acceleration low – and hence resulting in long scan times. For example, the Human Connectome Project employed T1- and T2-weighted high-resolution structural scans (0.7mm isotropic) that were specifically configured to explore the boundaries of detecting architectonic delineations (Van Essen et al., 2012) – but which took ~ 7-8 minutes to acquire. It is thus worthwhile to note that different strategies may be appropriate for different goals.

In many studies of morphometry, measurements are made of structures that are sampled across hundreds or thousands of individual acquisition voxels (e.g., hippocampus). Even when cortical thickness is estimated, the measures are typically an average across hundreds of vertices. For example, in the commonly used Desikan-Killiany atlas, surface regions range from approximately 6mm to 30mm in diameter (Desikan et al., 2006). What this means in practice is that morphometric measures for individual structures are sampled by large numbers of voxels even when those voxels are large (e.g., 1.2 mm). To the degree that the voxels align randomly within and along the boundaries of the structure being measured, the local errors of individual voxels are offset by other voxels that sample the same structure. In this manner, while any given voxel may be an uncertain estimate of the boundary of the structure or local thickness, the aggregate estimate of the structure or regional thickness may be quite stable (similar to the properties that make manual stereology viable; Mayhew, 1992).

## Overview of the Present Work

In this paper, we explore the viability and opportunities of extreme rapid T1-weighted structural scans. One thousand and twenty-five structural scans were acquired across 6 studies during 232 distinct MRI scanning sessions from 34 adult participants. We varied acquisition protocols to explore and pit traditional ~6 min MP-RAGE 1mm isotropic scans against newer varieties of accelerated multi-echo MP-RAGE and Wave-CAIPI MP-RAGE scans, and varieties of these newer sequences against one another. Various imaging parameters were tested (e.g., acquisition time, voxel size, acceleration factor). At the broadest level, the overarching goal of these studies was to determine whether rapid, low burden scans can replace traditional longer scans for morphometric analysis. Within this broad level goal, each study and set of analyses targeted more specific questions. In every instance we sought, and reported, features of rapid scans that might differ and be less desirable than traditional longer acquisitions for certain purposes.

A key feature of the design of several of the studies surrounded the possibility of averaging estimates from individual scans to obtain better, pooled estimates of morphometry. We refer to the possibility of pooling estimates across multiple sessions collected on the same day or across a few closely-spaced days as 'cluster scanning'. The

multiple sessions are not intended as separate longitudinal estimates but rather as an averaging strategy to obtain precise within-subject estimates. The idea is that a component of the estimated error term is due to the idiosyncratic position of the person and the person's state during the specific scanning session. By pooling estimates across sessions, the stability of the estimate is improved. Figure 1 illustrates the different sources of error that are captured across the various studies.

Another question tackled was, given a fixed amount of time and burden, what is the best strategy to stabilize the morphometric estimate? For example, given 6 minutes of scanning, is it better to collect one long traditional 1.0mm<sup>3</sup> MP-RAGE scan, 3 1.2mm<sup>3</sup> multi-echo MP-RAGE scans, or 6 1.2mm<sup>3</sup> Wave-CAIPI MP-RAGE scans? Do estimates from one or a few extremely rapid scans provide sufficient data for basic morphometric analyses? The series of studies and analyses were designed to help investigators answer these questions by illustrating the tradeoffs between multiple experimental choices that include newer rapid structural acquisitions and scans of lower resolution (1.2mm<sup>3</sup>) than typically acquired.

We found that rapid scans, as short as a minute in length, are appropriate for most purposes. Counterintuitively, rapid scans can reduce measurement error (when averaged) to a level lower than that achieved with longer scans. These observations will be discussed in the context of encouraging the field to consider adopting rapid structural scan protocols in place of longer variants.

---

Insert Figure 1 About Here

---

## Methods and Materials

### Participants

Thirty-four adults (25 female; 25.1 +/- 4.0 years, age range: 18-38 years) were enrolled across 6 studies (Table 1). Structural acquisitions were the entire focus of three studies (Studies 1, 3, and 4). Structural acquisitions were included in combination with

other functional data acquisitions for three studies (Studies 2, 5, and 6). The present paper only analyzes the structural scans. Functional data associated with Study 6 have been published previously (Braga & Buckner, 2017).

---

Insert Table 1 About Here

---

Participants for Studies 1-4 and 6 were recruited through advertisements on campus electronic bulletin boards and email lists or through flyers at local universities and community colleges. They did not have any history of neurological or psychiatric illness or ongoing use of psychoactive medications. Participants provided written informed consent in accordance with the guidelines set by the Institutional Review Board of Harvard University or Partners Healthcare and compensated. Participants in Study 5 were enrolled as part of the clinical outpatient service at McLean hospital and diagnosed with mental illness (bipolar illness, n=3 or schizoaffective disorder, n=2). These individuals used psychoactive medications throughout the duration of the study. Participants in Study 5 provided written informed consent in accordance with the guidelines set by the Institutional Review Board of Partners Healthcare and compensated.

Participants in 4 studies were scanned over the course of a day to a week (Studies 1-4). Participants in Study 5 were scanned between 5 and 14 times over a period of 8-15 months. Participants in Study 6 were scanned between 24 and 30 times over a period of 4-6 months.

### **MRI Data Acquisition**

Table 1 summarizes the type and number of structural brain scans that were acquired for each of the studies on a 3T Siemens Prisma MRI scanner (Siemens Healthcare, Erlangen, Germany) using either the vendor's 32-channel head or 64-channel head-neck phased-array coil. The 64-channel coil was used for all studies except those that targeted Wave-CAIPI sequences, which used the 32-channel coil because its form factor maximizes performance (see Table 1). The subjects either stared at a crosshair during the image

acquisition or viewed movie clips from Planet Earth without sound but with subtitles, in order to encourage them to remain awake and still during the image acquisition.

Individuals requiring vision correction either wore contact lenses or MRI compatible glasses in the scanner. Foam or inflatable cushions were used to immobilize the participant's head. Extensive instructions were provided before scanning to emphasize keeping the head still including description about the amount of movement that can affect scans (movements as small as the width of a dime) and the types of movements that can affect scans (e.g., swallowing, opening and closing mouth). Reminders, feedback, and motion check-ins occurred throughout the scanning session.

*Study 1: Feasibility of Extremely Rapid Structural Acquisition.* The goals of Study 1 were (1) to establish the feasibility of rapid structural scans and (2) to compare morphometric estimates derived from standard T1-weighted MP-RAGE scans with those derived from rapid acquisition protocols. Six subjects (S1-S6) underwent one MRI scanning session, during which a total of 28 T1-weighted images were acquired in counterbalanced order for four distinct scanning sequences. The number of acquisitions for each type varied to allow for comparisons of different scan types with matched total scanning duration, such that averaged estimates from multiple short sequences could be contrasted with standard, longer sequences. The greatest emphasis was placed on the shortest Wave-CAIPI scans to allow their variance properties to be examined fully.

The 4 acquisition types were: (1) 4 Standard (R=2) 6'12" scans (voxel size=1.0mm isotropic; TR=2530ms; TI=1100ms; TE=2.9ms; flip angle=7°; field of view=256×256×176mm; in-plane GRAPPA acceleration=2), (2) 2 Standard (R=4) 3'22" scans (voxel size=1.0mm isotropic; TR=2200ms; TI=1100ms; TE=2.9ms; flip angle=7°; field of view=256×256×176mm; in-plane GRAPPA acceleration=4), (3) 2 Wave-CAIPI (R=2x2) 2'45" scans (voxel size=1.0mm isotropic; TR=2530ms; TI=1100ms; TE=3.3ms; flip angle=7°; field of view=256×256×176mm; CAIPI acceleration=2x2), and (4) 20 Wave-CAIPI (R=3x3) 1'16" scans (voxel size=1.0mm isotropic; TR=2530ms; TI=1100ms; TE=3.3ms; flip angle=7°; field of view=256×256×192mm; CAIPI acceleration=3x3). Siemens WIP sequence 1084 (Polak et al., 2018) was used to acquire the Wave-CAIPI scans. For S1, the scanning console froze yielding 17 Wave-CAIPI (R=3x3) scans.

*Study 2: Testing the Limits of Extremely Rapid Structural Acquisitions.* Study 1 established the feasibility of rapid, structural data acquisitions and comparability to standard longer protocols. In Study 2, we tested the viability of an extremely rapid (1'01") version of the Wave-CAIPI T1-weighted MP-RAGE protocol that used acceleration combined with slightly enlarged voxel size (1.2mm<sup>3</sup>) to approach a one minute scan duration. Six subjects (S7-S12) underwent a single scanning session, during which two sequences were used in counterbalanced order to acquire a total of 28 scans per subject: (1) 4 Standard (R=2) 6'12" scans (as described in the first scanning protocol) and (2) 24 Wave-CAIPI (R=3x3) 1'01" scans (voxel size=1.2mm isotropic; TR=2530ms; TI=1100ms; TE=3.25ms; flip angle=7°; field of view=250×250×173mm; CAIPI acceleration=3x3).

*Study 3: Exploration of Within- Versus Between-Session Acquisition.* The goal of Study 3 was to explore within- and between-session error in morphometric estimates for a rapid T1-weighted structural scan. This study used multiple sequential acquisitions of a brief, accelerated multi-echo MP-RAGE scanning sequence acquired across three distinct MRI scanning sessions on the same day. The acquisition protocol parallels that used by the low-burden, Brain Genomics Superstruct Project (GSP) (Holmes et al., 2015) previously found to yield morphometric measures that correlated with those from longer acquisitions (Mair et al., 2012).

For 4 subjects (S14-S17), 20 scans were acquired per session for a total of 60 scans per subject. For one subject (S13), 16 scans were acquired per session (48 total). The acquisitions all used a T1-weighted multi-echo MP-RAGE (R=4) 2'23" scan (voxel size=1.2mm isotropic; TR=2200ms; TI=1100ms, TE=1.57, 3.39, 5.21, 7.03ms; flip angle=7°; field of view=230x230x173mm, in-plane GRAPPA acceleration=4; phase partial Fourier=6/8). To measure and provide the opportunity for analysis of the effects of movement, volumetric navigators were embedded within the acquisition sequences (Tisdall et al., 2012). The embedded navigators require 4 dummy TR periods to achieve a steady state, and so their use extends the scan duration from the shortest achievable duration, which is 2'12" for the scan parameters used (see Study 5). Note the volumetric navigators were only used to track the motion, in this instance, not to update the field-of-view positioning and correct the image acquisition for subject motion in real time, as they were originally intended (Tisdall et al., 2012).

#### *Study 4: Head-to-Head Comparison of 1.0mm<sup>3</sup> and 1.2mm<sup>3</sup> Acquisition Resolution*

*Within- Versus Between-Sessions.* The fourth study pitted two forms of the same multi-echo MP-RAGE sequence against one another to understand their relative ability to provide estimates of morphometry when pooling estimates within- and between-sessions. One rapid version (2'23") used lower resolution (1.2mm<sup>3</sup> voxel size) while the other employed a more traditional 1.0mm<sup>3</sup> voxel size and longer scan duration (5'58").

Six subjects (S18-S23) each underwent 2 scanning sessions on the same day. Fourteen scans were acquired per session for a total of 28 scans per subject, with order of scan type counterbalanced within each session: (1) 8 multi-echo MP-RAGE (R=2) 5'58" scans (voxel size=1.0mm isotropic; TR=2420ms; TI=1100ms; TE=1.69, 3.55, 5.41, 7.27ms; flip angle=7°; field of view=256×256×176mm; in-plane GRAPPA acceleration=2) and (2) 20 multi-echo MP-RAGE (R=4) 2'23" scans (as described in Study 3).

#### *Study 5: Feasibility of Rapid Structural Acquisitions in Patients Over Extended Longitudinal Epochs.*

A target use of rapid, low-burden structural imaging is to measure longitudinal change in patient populations. Study 5 explored the feasibility of rapid structural scans by acquiring 2 scans per session in 5 outpatient clinical participants (S24-S28) who were scanned over 8 to 15 months across upwards of 14 distinct MRI sessions. For each session, 2 T1-weighted multi-echo MP-RAGE scans were acquired (as described for Study 3, although the acquisition time was 2'12" because the sequence did not acquire volumetric navigators). By having multiple sessions and multiple acquisitions per session, over an extended period of time, these data were able to reveal differences in morphometric estimates between scans within the same session as well as between sessions.

Midway through Study 5's data acquisition, the MRI console software was upgraded from VD13D to VE11C. The scans for all other participants in this paper were acquired consistently with either the VD13D or the VE11C software version, with no differences within- or between-subjects. While unanticipated, the switch in scanner console version allowed us to explore the effect of a software upgrade on morphometric measures since each participant had multiple sessions of data before and after the software transition.

*Study 6: Testing the Limits of Between-Session Averaging of Rapid Structural Acquisitions.* The goal of Study 6 was to test the limits of averaging across sessions to achieve highly precise within-subject morphometric estimates. Six subjects (S29-S34) were scanned across 24 to 30 distinct sessions with a single rapid, T1-weighted structural scan acquired in each session using the same multi-echo MP-RAGE (R=4) 2'12" acquisition protocol as described for Study 3, although the acquisition time was 2'12" because the sequence did not acquire volumetric navigators. The scanning sessions occurred over 3-7 months.

## Image Processing and Morphometry

All T1-weighted MP-RAGE images were analyzed with version 5.3 of the FreeSurfer software suite (Fischl et al., 1999; Dale et al., 1999; <http://surfer.nmr.mgh.harvard.edu/>) employing both volumetric and surface-based analyses.

Volume-based processing included the following steps: (1) creation of an unbiased template space and image for each subject using robust, inverse consistent registration (Reuter et al., 2010); (2) alignment of each image to its subject-specific template space and image; (3) intensity normalization; (4) skull stripping (Segonne et al., 2004); and (5) segmentation into neuroanatomical structures (Fischl et al., 2002). From the segmented image, the volume of the subcortical structures was calculated (e.g., thalamus, caudate, putamen, hippocampus, amygdala, and corpus callosum).

Surface-based processing created a model of the white-matter surface (i.e., boundary between the gray matter of the cerebral cortex and the underlying white matter) and the pial surface (i.e., boundary between the gray matter of the cerebral cortex and the overlaying pia matter) (Fischl et al., 1999; Dale et al., 1999). The surface models consisted of points or vertices that form a mesh made up of triangles. The white-matter and pial surfaces were determined in subject-specific template space. The two cortical surfaces in the subject-specific template space were used to estimate a number of morphometric measures including: (1) Mean Cortical Thickness or distance between the white-matter and pial surfaces as the mean of ~120,000 vertices per hemisphere or for all ~240,000 vertices pooled across both hemispheres (Fischl & Dale, 2000), (2) Regional Cortical Thickness measures for the 68 cortical regions in the Desikan-Killiany atlas (Desikan et al., 2006;

Fischl et al., 2004); (3) Cortical Surface Area of the white-matter surface mesh; (4) Total Brain Volume (TBV) that included the entire brain volume except voxels labeled as brain stem, ventricles, cerebrospinal fluid, choroid plexus, or background; (5) Gray Matter (GM) volume that included volume inside the pial surface and excluded volume inside the white-matter surface and volume in the cortical ribbon that is not part of the cerebral cortex (e.g., hippocampus); and (6) White Matter (WM) volume that included all volume within the white-matter surface and excluded non-white matter voxels.

Estimated Total Intracranial Volume (eTIV), another global morphometric measure, was estimated in FreeSurfer using the parameters of registration of each individual to the template as extended from Buckner et al. (2004). Note that the template used here was a custom template constructed previously from 700 multi-echo MP-RAGE scans, with the intention of approximating contrast properties of the present acquisitions (as used by Holmes et al., 2015).

### **Image Signal-to-Noise Ratio (SNR)**

SNR in the T1-weighted images was calculated using the MRIQC software package (Esteban, et al., 2017). While this package provides multiple SNR values, we report the SNR\_wm as the best representative value for the images considered here. SNR\_wm is calculated by dividing the mean signal throughout the White Matter mask of each image by the standard-deviation of the signal in the mask region. We chose this metric over measures of mean signal divided by background noise estimates as the noise floor around the head in the periphery of images is increasingly non-uniform in multi-channel reconstructions from highly-parallel array coils and when parallel imaging acceleration techniques are used. This is particularly the case for the wave-MPRAGE scans. True SNR calculations (Triantafyllou et al., 2011) require complex k-space data which was not available in most cases.

### **Error Estimates in Morphometric Measures and Cluster Scanning**

Error estimates for morphometric measures included estimates derived from individual scans and also average estimates from groups of scans (i.e., calculate the morphometric estimate from multiple scans individually and then average the

morphometric estimates, as opposed to averaging multiple scans and then calculating the morphometric estimate from the averaged scan). This allowed the potential of different strategies that average estimates within- and between-sessions to be explored. For example, given 6 minutes of scanning time, how can the time best be allocated to obtain the most stable estimate of a morphometric measure? In the instance of a single traditional longer structural acquisition, the obtained estimate is typically the measurement estimated from a single scan. In the newly explored space of rapid structural acquisitions, that estimate can come from the average of up to 6 scans. For this reason, we computed estimates of morphometric measures from individual images as well as estimates that averaged independent estimates from multiple images.

The averaging approach was determined by the question at hand in each specific analysis, as described in the context of specific results and plots. In several instances, the continuum of averaging was considered by asking how estimates improve as the data went from including only an individual image to averaging over larger and larger numbers of images. A second critical factor explored the effects of whether the images used to derive the average estimate came from within the same session or averaged morphometric estimates from images collected across sessions, enabling an approach we call cluster scanning (where multiple acquisitions are obtained across distinct scanning sessions). Cluster scanning is intended to reduce the error term associated with idiosyncratic features and the head position of an isolated scanning session and improve the precision of morphometric estimates, given that studies of longitudinal change must necessarily compare estimates between separate scanning sessions.

After processing the T1-weighted MP-RAGE scans, we calculated the error in the morphometric measures derived from scans acquired within a single scanning session and between scanning sessions (Figure 1). For within-session error, the scans from the same scanning session were randomly assigned to two groups. Then, the morphometric measures within each group of scans were averaged. The absolute difference of the two averages was calculated as the error. This process was done 10,000 times to create a distribution of the absolute differences in the raw measures and also a distribution of the percent error. Depending on the analysis and number of scans acquired during a session, the number of morphometric measures included in the average estimate varied. For the

between-session error, we followed the same process but randomly assigned scans from one session to one group and then randomly assigned scans from another session to a second group before averaging the morphometric estimates and subtracting the two groups of morphometric estimates. For cluster error, we again followed the same process but randomly assigned scans from multiple sessions to one group and then randomly assigned other scans from the same multiple sessions to a second group before averaging and subtracting the two groups of morphometric estimates.

A limited number of scans were acquired for some of the sequence types. For example, we only collected four Standard MP-RAGE scans to approximately match the total imaging time for the 20 Wave-CAIPI MP-RAGE scans (Standard: 24'48" total imaging time; Wave: 25'20" total imaging time). When calculating the error in the morphometric measures, we iterate through the process 10,000 times. In the case of the four Standard MP-RAGE images, we recognize that 10,000 iterations are not necessary given there are only six combinations to compare. However, in order to keep the methods consistent across the acquisition types, we ran 10,000 iterations for error analyses regardless.

## Statistical and Error Analyses

We explored sources of error in morphometric measures and how to reduce the error. First, the effect of gradient coil heating on intracranial volume was tested. We calculated the correlation between the order in which scans (multi-echo MP-RAGE) were acquired during the scanning session and eTIV. Second, the effect of a cortical thickness outlier was tested. The individual participant surfaces were registered to the *fsaverage* template surface for all of the following vertexwise analyses. The analysis was done in 5 subjects (S13-S17) across every vertex of the *fsaverage* template surface to identify outliers in cortical thickness. Outliers were defined as a cortical thickness estimate from a single vertex that is more than three standard deviations greater than or less than the single vertex's mean cortical thickness estimate for a given subject. Third, the effect of smoothing on cortical thickness error estimates was tested with 5mm, 10mm, and 20mm full-width half-maximum Gaussian smoothing kernels. The analysis was done in 5 subjects (S13-S17) at a single vertex of the *fsaverage* template surface in the left superior temporal gyrus. Finally, the amount of error in cortical thickness estimates was measured in each vertex of

the *fsaverage* template surface. Before the error was calculated, a 10mm smoothing kernel was applied to the map of cortical thickness, and then at each vertex the cortical thickness estimates from 12 scans were randomly sampled and averaged (4 scans per session from 3 sessions). The error was then calculated as the difference between 2 groups of averaged cortical thickness estimates. This process was done 10,000 times at each vertex. The average error from the 10,000 iterations is displayed at each vertex of the cerebral cortex.

## Visualization

The scatter, box, and line graphs were generated in RStudio version 1.0.136 with the ggplot2 package version 2.2.1 (Wickham, 2009). The coronal and axial slices and surface overlays were generated in freeview as part of the FreeSurfer software suite. The images of the inflated brain surfaces were displayed in the Connectome WorkBench with wb\_view (Marcus et al., 2011).

## Results

### Rapid Structural Acquisitions Are Feasible and Yield Stable Morphometric Estimates

The most fundamental result of these studies is that extremely rapid structural acquisitions are viable for many purposes and often better than longer acquisitions. This central observation was reinforced across all of our analyses and generalized to every variant of rapid acquisition we tested. Focusing on Study 1, Figure 2 illustrates corresponding image sections from a Standard long acquisition MP-RAGE and an accelerated Wave-CAIPI MP-RAGE acquisition taking just over one minute. These structural volumes were analyzed separately to determine the inner and outer boundaries of the cortical surface and various global and local morphometric estimates. Three results emerged. First, the Wave images were clear, of high quality, and demonstrated contrast properties by which the gray-white boundary could be easily discerned by eye as well as all subcortical structures. That is, to visual impression, all relevant features detectable in the longer Standard acquisitions could be detected in the rapid Wave acquisitions. Second, the Wave acquisitions had predictably lower SNR resulting in a grainier appearance in the white matter. For study 1, image SNR ranged from 10.9 to 13.8 for the Wave scans as

compared to 17.5 to 24.4 for the Standard long scans. Third, the automated quantitative morphometric estimates derived from the rapid Wave scans were stable when single scans were used. Estimates became highly stable when multiple scans were averaged, often surpassing estimates from Standard scans.

---

Insert Figure 2 About Here

---

Specifically, global morphometric measures derived from the Standard and Wave images are stable across separate scans within the same session including for the extremely rapid acquisitions (Figure 3). There were differences in the absolute value of estimates between sequences, serving as a reminder that all of these forms of T1-weighted sequence have idiosyncratic contrast properties that influence the absolute values of morphometric estimates. While the pattern was not perfectly consistent, the general trend for cortical thickness was for the Wave scans to yield estimates that were slightly smaller (about 1%) than the Standard scan estimates.

---

Insert Figure 3 About Here

---

Human cortical thickness markedly varies between regions, a feature originally determined in postmortem samples and confirmed in the earliest analyses of regional cortical thickness estimated from MRI (Fischl et al., 2000). To determine if rapid acquisitions can recapitulate regional differences that are found for longer Standard acquisitions, regional thickness estimates were determined for the 34 regions of the Desikan-Killiany atlas (Desikan et al., 2006) for individual Standard and Wave scans from Study 1. For all subjects, there was high correspondence between the two sequences. Figure 4 plots representative correlations of regional estimates ( $r^2 = 0.93\text{--}0.97$ ). There were also systematic differences. The slope of the best-fit line was lower than that of the

identity line: as thickness increased, the Wave scans tended to slightly underestimate regional thickness as compared to the Standard scans.

---

Insert Figure 4 About Here

---

The remaining results of Study 1 concerned quantifying the level of error associated with morphometric estimates from each acquisition type. Error estimates were constructed between individual scans as well as averages from multiple scans. This is a particularly critical question because rapid scans allow multiple scans to be acquired in the same amount of time as a longer scan. Rapid acquisitions may be optimal even when more time is available since average estimates may be more stable than an estimate from a single high-contrast scan. For these results, the amount of error in the morphometric measures was calculated as illustrated in Figure 1's example of within-session error. Error was considered as the difference in value from one set of scans versus another set of independently analyzed scans. In cases where there was only one scan compared against another, error was simply the difference in estimate between the two scans. When multiple scans were considered, the difference in the two average estimates was calculated, and this process was repeated 10,000 times to provide the best estimate of the error. Note that the morphometric estimates were always derived from individual scans and then averaged afterwards.

Figure 5 compares the error in the global morphometric measures when holding the amount of imaging time that goes into the error estimate similar across the Wave and Standard images (5'04" for 4 Wave scans versus a single 6'12" Standard scan). The average error for three global morphometric measures—eTIV, TBV, and Mean Cortical Thickness—are smaller in the majority of the subjects when the estimates are constructed from multiple rapid acquisitions.

Figure 6 explores the number of Wave scans needed to equal or better the amount of test-retest error derived from Standard scans. The first observation is that the error from a single Wave scan (of just over one minute) is almost as low as the estimates from

the much longer Standard scans for most global morphometric measures for most subjects. As estimates from Wave scans are averaged, they show an advantage for most measures and subjects over the longer Standard scan. For example, the average amount of Wave imaging time needed for the six subjects in Study 1 to be equal to or less than the error from one Standard scan was 2'57" for eTIV, 5'42" for TBV, and 4'26" for Mean Cortical Thickness. The other feature to note is that there is no fixed rule across the subjects. For example, the estimated error for Wave scans for Mean Cortical Thickness in S3 never surpassed the error achieved by Standard scans. By contrast, a single Wave scan did better than the Standard scan for S2's eTIV error estimate.

Figure 7 further explores the contrast of multiple Wave scans against a single Standard scan for multiple commonly used morphometric measures. The data from Studies 1 and 2 are presented in the plots on the left (Study 1) and right (Study 2) sides of the figure. As with the results in Figure 5, this number of scans makes the Wave scanning time about the same (slightly less) than the time allotted to the Standard scan. Results are plotted for estimates of global structures that would be expected to be the most stable, subcortical structures that span those known to have high test-retest reliability (hippocampus) and those with lower reliability (amygdala), and local regional thickness. In each case, other than Cortical Surface Area in Study 1 and TBV, WM, and cortical thickness of the left superior frontal gyrus in Study 2, the median amount of error calculated from subjects is smaller for the Wave scans than the Standard scans, even though the total scanning duration for the multiple Wave scans is more than a minute less than the Standard scan in Study 1 and 6 seconds less in Study 2. The result that morphometric measures from multiple rapid acquisitions can be averaged and provide stable estimates allows for approaches that use 'cluster scanning'.

---

Insert Figures 5, 6 and 7 About Here

---

## Cluster Scanning Allows for Precise Morphometric Estimates Within Individuals

Cluster scanning involves collecting multiple independent acquisitions across closely timed sessions to obtain a highly stable estimate of brain morphometry. Results of Studies 3 and 4 suggest cluster scanning may lower test-retest error estimates. However, the tradeoff of collecting more scans within the same session where the subject's head position is fixed versus scans collected between sessions is not known. To calculate within- and between-session error, we followed the procedure outlined in Figure 1 by comparing estimates that computed the test-retest error from (1) multiple acquisitions all collected from within the same session (Within), (2) estimates from scans acquired in different sessions (Between), or (3) estimates from scans that pooled acquisitions from separate sessions (Cluster). In all cases the exact same number of estimates were averaged and compared. The difference between conditions was only where the scans came from. The Within condition test-retest should be the lowest as all estimates being averaged and compared are collected from the same continuous sessions; the open question was how much worse the estimates would be when they came from scans obtained in different sessions (Between condition) or utilized pooling of scans collected across sessions (Cluster condition).

Figure 8 illustrates the results of Study 3. As evidence of the broad stability of the measures, the three conditions often yielded similar error estimates. There were two exceptions. First, for the estimate of eTIV, the Between-session error estimate was notably worse than the other two. eTIV is a highly stable morphometric estimate used in modeling the meaning of other local estimates so this finding is itself important. Second, while the estimates were generally comparable, multiple sporadic examples of larger errors occurred in the Between-session condition (e.g., TBV for S14 and S17, and Mean Cortical Thickness for S17). These results indicate that, while morphometric estimates are generally stable within- and between-sessions, there is a relevant error term that becomes evident between sessions for some measures and in some instances. This is a relevant feature of morphometric error because of the general practice in longitudinal studies to scan an individual once at each time point. Between-session error was analyzed further for independent data from Study 4.

### Insert Figure 8 About Here

---

Figure 9 illustrates the results of Study 4. Study 4 also contrasted a rapid acquisition protocol that used  $1.2\text{mm}^3$  resolution with a longer acquisition at higher resolution ( $1.0\text{mm}^3$ ). The two protocols were set up such that five low-resolution acquisitions could be acquired in the same amount of time as two high-resolution acquisitions. This allowed us to contrast the effects of acquisition resolution while holding the amount of scan time constant (by comparing estimates from five low-resolutions scans to two high-resolution scans). Results again reveal that the estimated test-retest error was low across all conditions including both resolutions. As before, for the estimate of eTIV the Between-session error was notably worse than the error term from scans pooled across sessions (the Cluster condition). The increase in the Between-session error term was sporadically apparent in the other estimates (e.g., TBV for S22, and Mean Cortical Thickness for S19 and S21).

---

### Insert Figure 9 About Here

---

Figure 10 explores how different numbers of scans affect the error term for two representative subjects (S13 and S17) from Study 3. The key observation is that, even when many scans contribute to the average estimate, the between-session error cannot be eliminated. This is an important result as it indicates that the amount of additional scanning within a session quickly reaches a diminishing return. Generally, the greatest decline in error occurred during the first few additional acquisitions. As more scans were included, the decline in error tapered off. The decline in error for the within-session error and the cluster error follows a similar trajectory to each other because the cluster error incorporates morphometric estimates from different sessions thereby removing the between-session error. In Study 6 we explored the limits of estimates from cluster scanning by pooling estimates from numerous scans acquired across independent sessions.

---

Insert Figure 10 About Here

---

Figure 11 uses data from Study 3 to illustrate how smoothing may affect the optimal number of scans to collect within a session. Different smoothing kernels were applied to the maps of cortical thickness and then the amount of cluster error in a single vertex of the superior temporal gyrus was measured. The smoothing kernel incorporates the cortical thickness estimates of surrounding vertices, thereby increasing SNR but also likely decreasing the sensitivity to detect a difference or change in local anatomy. Figure 11 illustrates that the larger the smoothing kernel (i.e., the more surrounding vertices that provide information to the cortical thickness estimate) the lower the amount of cluster error. When smoothing is coupled with multiple image acquisitions, the amount of error is further decreased. The greatest gain in error removal occurs when going from no smoothing to 5mm smoothing (48% decrease in error), and then the gain in error removal decreases with each subsequent size of smoothing kernel (37% decrease from 5mm to 10mm and 28% decrease from 10mm to 20mm). These results do not necessarily indicate a larger smoothing kernel should be used, as sensitivity issues factor into that decision which may vary depending on the size of the regions possessing the thickness effect, but it does illustrate that the benefit of increased numbers of scans is affected by the intended processing procedures.

---

Insert Figure 11 About Here

---

Figure 12 illustrates the error in cortical thickness across the full cortical mantle. The local estimates were derived from Study 3 and were constructed by averaging 12 multi-echo MP-RAGE scans (4 scans per session from 3 sessions) for each thickness estimate. Error is plotted in absolute cortical thickness as well as in percent relative to the

local thickness. The amount of error was found to vary from region-to-region across the cortex, and the regional patterns of error also varied from one subject to the next.

Sensorimotor cortex, visual cortex, and lateral temporal regions have the smallest amount of absolute error ( $<20 \mu\text{m}$ ). However, in the case of the sensory and motor regions, they are also the regions with the thinnest cortex. When the cortical thickness of a region is taken into account and the amount of error is depicted in percent error, the lateral temporal regions and fusiform gyrus have the smallest percentage of error ( $<1.0\%$ ). The insula, the cingulate cortex near the border with the corpus callosum, the inferior parietal lobule, and the ventral and medial surfaces of the prefrontal cortex consistently have the largest amount of absolute error ( $>50 \mu\text{m}$ ) and percent error ( $>2.0\%$ ).

---

Insert Figure 12 About Here

---

### **Rapid Structural Acquisitions are Feasible in Patients**

A target use of rapid structural acquisitions are patients, children and older adults where reducing the overall session length is desirable as well as the length of individual scans to mitigate motion. The results of Study 5 demonstrate the feasibility of using rapid acquisitions to obtain stable morphometric estimates from patients over extended periods of time (Figure 13). Generally, the scans were tolerated exceptionally well and in no session across the patients (86 scans across 43 sessions) were scans lost due to technical issues or complete motion failure. Given motion is a problem always and in patients in particular, this result is likely due to the comfort of these well-practiced individuals. But it may also reflect that the scans are of short duration and well tolerated. Quantitatively, Mean Cortical Thickness estimates were stable over time within individuals, with several notable exceptions that shed insight into the sources of error as reported in the next section.

---

Insert Figure 13 About Here

---

## Multiple Independent Sources of Error Influence Morphometric Error

As we explored and refined the cluster scanning approach, we noticed several sources of error in our morphometric measures that could be distinguished. The longitudinal data from Study 5 were particularly valuable for detecting sources of error because two rapid acquisitions were acquired during each of many independent sessions that spanned an 8-15-month period. This allowed within-session differences in estimates to be assessed in relation to differences between sessions. Moreover, a console software upgrade occurred midway through the study. By plotting Mean Cortical Thickness across scans (both within and between sessions), Figure 13 illustrates three distinct sources of error.

First, a clear decrease in the Mean Cortical Thickness for 4 of the patients (S24, S26-S28) coincides with the software upgrade. The decrease in Mean Cortical Thickness for the fifth patient (S25) due to the software upgrade likely also exists, although it is not as obvious because of the second source of error we noticed.

The second source of error, head motion during the scan, is especially apparent when comparing the two images from S25's fifth scanning session. Upon visual inspection of the raw images, the first image is a high-quality image without noticeable artifact, whereas the second image is blurry and marked with ringing, which is indicative of head motion. Head motion is known to reduce cortical thickness estimates (Reuter et al., 2015) and is the likely the cause of the decreased Mean Cortical Thickness in the fifth scanning session's second image.

The third source of error is likely due to different head positions within the head coil. Mean Cortical Thickness estimates for S27 are stable within sessions 6-8 but are markedly different across the three sessions. Although the between-session effect is only consistently seen in eTIV in Studies 3 and 4 (Figures 8, 9), the Mean Cortical Thickness differences across sessions in S27 further suggest that head placement within the head coil or another between-session source of variance contaminates morphometric measures broadly. The eTIV for sessions 6-8 corroborate the likely effect of head placement, as there are also clear between-session differences in the estimates (data not shown).

An unexpected but quite subtle source of error was detected in Study 3, which involved numerous (16 to 20) sequential rapid multi-echo MP-RAGE scans within each session. Figure 14 illustrates this effect. The eTIV estimate demonstrated an apparent increase over the course of the session likely due to frequency drift because of coil heating. This effect is probably present in all morphometric measures, with eTIV being the only measure stable enough for the effect to manifest. The average effect from the 5 subjects in Study 3 is a 0.009% increase per consecutive image acquired. The effect may be specific to the rapid multi-echo MP-RAGE scans, as we do not see the coil heating effect when the rapid multi-echo MP-RAGE scans were interleaved with the Standard multi-echo MP-RAGE scans (Study 4).

---

Insert Figure 14 About Here

---

The final source of error occurred during the processing stage of morphometric estimation. We searched for morphometric outliers by comparing the cortical thickness estimates at each vertex in the cerebral cortex. The first session of data was explored for each of the 5 subjects from Study 3 separately. One example outlier was found in the supramarginal gyrus of S13 (Figure 15). The estimate from the 11<sup>th</sup> image was flagged as an outlier, as it was  $>3$  standard deviations above the vertex's cortical thickness. By inspecting the pial and white-matter surfaces for each of the 16 scans, we found that the pial surface of the 11<sup>th</sup> scan did not follow an invagination that the surfaces from other scans made. The effect of the cortical thickness outlier increased the amount of error by  $\sim 45\%$ . This result highlights the sensitivity of morphometric errors to the uncommon effects of outliers. There was nothing apparent about this particular image that would elicit the outlier as compared to the many other scans from the same individual. Thus, while the error is certainly the result of an interaction of the image and process, to this point we cannot identify a simple cause. Notably, this form of error was extremely rare in our analyses of rapid structural acquisitions bolstering confidence in their robustness but reminding that sporadic artifacts can affect the stability of morphometric estimates.

---

Insert Figure 15 About Here

---

## Cluster Scanning Achieves Mean Cortical Thickness Estimates With <10 $\mu$ m Error

Our final results explored the limits of cluster scanning to obtain precise morphometric estimates. For this analysis, the estimated error for multiple morphometric measures was calculated for the 6 subjects from Study 6 who each were scanned across 24 to 30 independent sessions over a relatively brief period (3-7 months). Figure 16 illustrates the test-retest error as a function of the number of independent scans contributing to the measure. In all instances, following the typical square-root-of-N averaging rule of thumb, the error term dropped markedly initially and then further averaging yielded diminishing return once 5 or 6 scans contributed to the estimate. For all individuals, the error in Mean Cortical Thickness falls below 10  $\mu$ m when 12 independent scans are utilized (representing under 30 min of total scanning duration). For S30 and S32 the error is near 5  $\mu$ m representing 0.20%. For hippocampal volume, the estimated error was under 0.6% for 4 of the 6 subjects. The remaining 2 subjects have more than twice as much mean percent error due to outlier values in 1-2 scans. The availability of multiple scans allows for excluding of outliers using simple, consistent procedures.

As an example of a procedure for handling outliers, we recalculated the effect of the number of scans on the error term using the median instead of the mean. The median is a less frequently used measure of central tendency that is minimally sensitive to infrequent outliers. It also has the benefit of not requiring a large number of samples to apply an outlier criterion (e.g.,  $>3SD$ ) which is not powerful when small sample sizes are used to estimate the distribution, such as is the case here.

In the hippocampal mean percent error plot of Figure 16, the curves for S30 and S32 never resolve to the amount of error in the other four subjects when mean percent error is employed. In contrast, the error for both subjects in the median percent error curves is initially larger than the error for the other four subjects but is dramatically reduced after 3 scans are included in the error calculations. Despite the jagged appearance of some of the

median curves due to the averaging of estimates when an even number of scans is considered, the plots highlight the value of calculating the median as an effective strategy at mitigating the effect of outliers but only when including more than two scans. Table 2 lists the median estimated error for common morphometric metrics estimated from 1, 3, 6, and 12 independently collected scans. Other more sophisticated procedures are possible to further reduce the error term if assumptions are made; the use of median values shows a simple, intuitive example of how multiple scans can be used to stabilize estimates even when outliers are present.

---

Insert Figure 16 and Table 2 About Here

---

## Discussion

Results of all six studies converge to make clear that rapid structural sequences are appropriate for many research purposes. For anatomical registration and basic morphometry, a single scan of one to two minutes in duration is viable. We achieved our fastest morphometry-ready scan (1'01") by combining new parallel imaging approaches (specifically Wave-CAIPI) with a slight reduction in resolution to 1.2mm<sup>3</sup>. For most common uses, we suspect this one-minute scan is adequate. For applications where detecting morphometric change is the goal, acquiring multiple rapid structural scans may have benefits over acquiring a single long scan including reducing the error of the estimate and affording tolerance to motion-induced scan loss.

In discussing these results, focus is first placed on the basic quantitative properties of the rapid structural scans that were explored. We then discuss new study designs that are made possible by rapid scans, including the possibility of detecting clinical change within individual people over periods of less than a year using 'cluster scanning'. We conclude by discussing limitations of rapid structural scans and offer a practical suggestion for how the research community might explore transitioning to rapid scans.

## Rapid Structural Scans Are Viable for Morphometric Analysis

Rapid structural scans are viable for morphometric analysis possessing sufficient SNR for automated measurement of subcortical volumes (e.g., hippocampus) and cortical thickness. While the images appear grainier due to their reduced SNR, the test-retest error in morphometric analysis from rapid scans ranged from comparable to twice as small as that of a standard scan, while taking up to 1/5<sup>th</sup> of the scan time. Thus, if one can only acquire a single scan, without strong time constraints, the longer traditional scan is the best option. If time is tightly constrained, a single rapid scan will likely be sufficient for most purposes. In practice, scanning time is the commodity and multiple short scans can be imaged sequentially in the same amount of time as a longer scan so the question becomes one of tradeoffs and goals.

Broadly we found that, for an equal amount of time, a series of rapid structural scans did as well or better than a longer standard scan in terms of reducing test-retest error. Figure 6 illustrates this point by comparing the reduction in test-retest error when more scanning time is devoted to the newer rapid Wave-CAIPI MP-RAGE versus a standard scanning protocol. Benefits of averaging estimates from multiple rapid scans were observed for global morphometric measures as well as subcortical volumes and regional thickness measures (Figure 7). The estimated regional cortical thickness values from the Wave scans were highly correlated with those obtained from the standard scans ( $r^2 = 0.93 - 0.97$ ). However, the strong relation between estimates from different scan types does not mean they are interchangeable.

Absolute morphometric values obtained from different scan types were not the same. While estimates are stable within an individual for a scan type, Figure 3 illustrates that the values differ between scan types. This is likely due to a combination of how contrast properties and SNR influence the determination of gradients and registration that are the basis of automated morphometry. Another factor could be differences between how the contrast properties of the acquired images interact with the contrast properties of the reference atlases (see Buckner et al., 2004 for discussion). When acquisition resolution varies (Studies 2 and 4) partial volume effects will also contribute to biases in the absolute value of the estimates (Zaretskaya et al., 2018). In practice this means that, at this level of precision, analyses must occur between scans of the same type. Furthermore, these results

also are a reminder that tissue volumes, boundaries, and transitions (gray to white matter) are not measured directly but rather are estimated indirectly from MRI signal properties affected by sequence, resolution, and analysis procedure.

An informative subtle difference is noted when contrasting regional thickness estimates between scan types in Figure 4. Figure 4 shows that the values from the extremely rapid Wave scans underestimate regional thickness estimates of the longer standard scans. That is, while the correlations between the two scan types approach 1 ( $r^2 = 0.93 - 0.97$ ), the values from the Wave scans fall on a line that has a slope of less than 1 (for 4 of the 6 subjects). This bias is again a reminder that values are estimates and not true measures. But more critically, the difference in slope may indicate a reduction in sensitivity to detect change in thickness. As the cortex changes thickness over time, a reasonable assumption is that the value will move along the lines plotted in Figure 4. Estimates from the Wave scans will move less for a corresponding true change than estimates from the standard scans. For the majority of the data (for thickness estimates between 1.7mm and 3.5mm), the data fall extremely close to the identity line so this effect may be negligible, but it is a caution that test-retest error is not the only factor that will contribute to sensitivity to detect change over time.

### **New Study Designs Are Possible Including Cluster Scanning**

Our results suggest that for many studies long scans can be replaced with shorter variants. Such a change will have a major impact on the field by reducing subject burden and shortening session length (or allowing space for other scans). For example, consider a non-compliant child who must be scanned while sleeping or only stays still for short epochs. The ability to capture a structural scan rapidly could be the difference between success and failure. Short scans can be repeated multiple times throughout a session adding redundancy for failed or motion-contaminated acquisitions. Beyond substituting scans or taking the median values (Table 2), completely new study designs are possible with rapid scans.

One new design possibility is a strategy we call 'cluster scanning' – a name chosen to emphasize that numerous scans are acquired near to one another in time (see also 'Burst Designs' in behavioral research; e.g., Sliwinski, 2008). In longitudinal analysis, a goal is to

reduce the error term of the morphometric measure as much as possible while maintaining sensitivity to change. We have shown that averaging morphometric estimates from many scans reduces the error of the measurement. We have also demonstrated that a source of error is the idiosyncratic positioning or state of the subject at the time of scanning, which can present a barrier to obtaining the increased benefit of averaging more estimates (e.g., see the plot of eTIV error in Figure 10). To reduce this error, one may need to scan the individual across positions and/or states given that the estimate will be compared to a distant longitudinal estimate on another day in another position. Cluster scanning mitigates this error by acquiring many scans across sessions. Estimates from two separate sessions are better than one and presumably more are better than fewer, so the practical limitations of obtaining multiple scans will come into play in adopting cluster scanning.

Existing large-scale studies offer insight into possible ways cluster scanning can be adopted. Confronted with the practical barrier of what children can comfortably tolerate, the recently launched Human Connectome Project in Development divided the imaging modalities between two separate sessions with each session under one hour (Somerville et al., submitted). Long duration T1- and T2-weighted structural scans were included in the first session only. The present results suggest that multiple rapid scans spread across the two separate sessions may be more optimal for certain goals.

Cluster scanning may allow detection of developmental or disease-related change in individuals over modest intervals. As a research tool, this affords the ability to examine effects in small numbers of participants. In the translational arena, the approach could allow for early signs of efficacy in treatment trials, such as might occur in the initial safety or exploratory phases of human trials where limited numbers of individuals are enrolled. Previously determined estimates for morphometric change are useful to anchor expectations about what might be possible. Change in regional cortical thickness during typical adolescence may be as much as 300-400 microns over multiple years (e.g., Tamnes et al., 2010; Raznahan et al., 2011). In psychiatric illness, cortical differences can be 200 microns or more (estimated from Cannon et al., 2002; Kuperberg et al., 2003; Goldman et al., 2009). Neurodegeneration can be more substantial. In a recent exploration of many morphometric measures, whole-brain atrophy in Alzheimer's disease was found to be

about triple (~1.5% per year) that of older controls (~0.5% per year) (Cash et al., 2015). Hippocampal volume loss is similarly dramatic.

Our present results suggest that three or four half-hour MRI sessions can accommodate 30-40 or more separate morphometric measures. After data loss due to motion / quality screening, the within-individual measurement error for whole-brain atrophy may be below 0.2%, for hippocampal atrophy below 0.5%, and regional cortical thickness below 0.5% (<15  $\mu$ m). These measures are positioned to detect within-subject change in less than a year. However, this should be considered a hypothesis because sensitivity to change has not been directly tested in the present work. Our error estimates suggest such designs may be possible and should be explored going forward.

A final study design that becomes possible with fast acquisition time is to combine acquisitions. In the typical MRI scanning session, one image modality sequentially follows the next (e.g., the T1-weighted structural scan is followed by a functional MRI resting-state scan). Alternatives are possible. For example, the signal fluctuations that serve as the basis of resting-state functional MRI evolve slowly over minutes. For this reason, fast scanning methods in functional MRI have been used to achieve high sampling rates and high resolution, but the overall scan duration typically remains long (Ugurbil et al., 2013; Harms et al., submitted). Given that there may be physiological reasons that the functional MRI sampling must extend over time, efficiency can be gained by scanning the multiple modalities in combination during a single scan. For example, a functional MRI protocol could be constructed that was 6 minutes in length that acquired whole-brain coverage every second or so. If the functional acquisition took only 600-700 msec, then the remaining time between brain repetitions could be used to acquire portions of the T1- and T2-weighted structural scans. At the end of the 6 minutes, the full set of measures would be complete: a 6-min functional MRI scan, a T1-structural, and a T2-structural. With fingerprinting methods (Ma et al., 2013; 2018), it is not beyond reach to imagine a scan session that is 5 or 10 min in total length that acquires sufficient information to reconstruct any structural scan weighting as well as functional data that continuously samples the functional fluctuations over the entire session.

## Limitations of Rapid Structural Scans and the Present Methods

In discussing the ways rapid structural scans are as good as (or sometimes better than) longer duration traditional scans, it is critical to make clear that there are tradeoffs. In addition to the SNR reductions that individual fast scans necessarily possess compared to their equivalent longer counterparts, there are also clear limitations of trading off resolution for speed. The ability of 1.2mm structural scans to yield stable estimates of global hippocampal and amygdala volumes is unlikely to generalize to estimation of subdivisions of these structures where high-resolution images have been used previously (e.g., Ekstrom et al., 2009; Saygin et al., 2017). A resolution tradeoff will be generally problematic for delineation of fine anatomical boundaries and substructure, and altogether preclude examination of organization of structures of the cerebellum which possesses a thin, geometrically complex cortex.

As a particularly informative example of the use of resolution, The Human Connectome Project has recently released a large, unique dataset of young adults (Van Essen et al., 2012). These protocols acquire high-resolution T1- and T2-weighted structural scans (0.7mm isotropic) specifically designed to explore boundaries of architectonic transitions. Such resolution is usable in many ways, for example, in estimating signal associated with white-matter cortical myelin content (Glasser et al., 2011). Reduction in resolution may not be desirable especially as methods for estimating signal properties at different depths through the cortex are refined (Shafee, Buckner and Fischl, 2015). In expanding the Human Connectome Project to children and other populations where motion and burden are limitations, images with slightly reduced resolution (0.8mm isotropic) and prospective motion correction (Tisdall et al, 2012) were chosen (see Somerville et al., submitted). For goals linked to estimation of large structures or regional thickness, multiple shorter acquisitions spread over sessions may be optimal. For delineation of fine anatomical details of cortical myelin context, longer high resolution scans may be optimal. This is a reminder that there are many solutions to the challenges of structural acquisition with different options appropriate for different goals.

As another example of the limitations of trading off scan time against resolution, we revisit the trial design described above as an opportunity – that of the quick scan to obtain data in young children. For basic morphometric measures as analyzed extensively in this paper, that choice is a good one. But what if the clinically-relevant structural target shifts

from global morphometric measures to subtle local cortical dysplasias? Dysplasias come in different types (Tassi et al., 2002) and can be extremely subtle deformations of the cortical mantle that require local estimation, such as in autism spectrum disorders (e.g., Casanova et al., 2013; Stoner et al., 2014). Higher resolution acquisitions may be required for exploring certain forms of local dysplasias. This is a reminder that choice of lower resolution scans may have lasting consequences on what kinds of analyses are possible, those that are imagined and those that emerge over time.

Another limitation of the explored approaches is that many sequence strategies were simply not explored. Some of these can be explored on their own, while others may interact with the opportunity of acquiring extremely rapid structural scans. For example, we alluded to but did not explore the opportunities of active prospective motion correction (Tisdall et al., 2012). We also did not explore the possibility of obtaining structural scans with multiple contrast properties. In a most basic form, T1- and T2-weighted structural images can be combined (e.g., Glasser et al., 2011; Alfaro-Almagro et al., 2018); in elaborated forms, such as employed by recently developed magnetic resonance fingerprinting, quantitative structural images can be reconstructed or modeled across a range of contrast properties (Ma et al., 2013; 2018). These newer techniques hold promise for further revolutionizing quantitative morphometry.

A final limitation of our paper is linked to data analysis. We explored only a portion of the ways multiple images can be combined. For example, there may be opportunities to jointly estimate morphometric measures that obtain separate estimates from each scan in contrast to fewer estimates from averaged scans. Image averaging increases contrast, while potentially blurring the images. It seems likely that we have not found the optimal averaging procedure that may differ by the brain structure and targeted measure.

## **Conclusions and Practical Suggestion for Adoption**

Rapid structural scans are viable and can reduce scan session length when used as a replacement for longer traditional structural scans. Acquiring multiple structural scans within and between sessions can reduce morphometric error estimates lower than is achievable with a single longer scan, opening a variety of design possibilities including

‘cluster scanning’. However, these properties do not necessarily mean they will become widely adopted.

Rapid structural scans are grainier and have lower SNR. They simply do not look as good. We hope that the quantitative properties extensively described in this paper encourage users to consider their tremendous value in reducing scan burden and potential value in improving quantitative morphometric estimates. However, knowing that it is difficult to replace time-tested protocols that have worked successfully for so many purposes, we offer a final suggestion: rapid structural scans are so short that they can be tagged onto existing protocols without replacing traditional structural acquisitions. By tagging them onto existing sessions, their viability as a replacement to longer traditional scans can be explored with minimal risk.

## Acknowledgments

We thank the Harvard Center for Brain Science neuroimaging facility and FAS Division of Research Computing. André van der Kouwe, Kawin Setsompop, Berkin Bilgic, Wei Liu, Stephen Cauley, Yulin Chang, and Larry Wald provided advice and assisted in implementing the MRI sequences. Specifically, the Wave-CAIPI MP-RAGE sequence is a “Works in Progress” package available on request from Siemens under a research agreement. We also thank Mark Eldaief, Matt Hutchison, Rodrigo Braga, Laura Farfel, Lauren DiNicola, Garth Coombs, Erin Phelgar, Marisa Marotta, and Arpi Youssoufian for help with data collection. This work was supported by NIH grant P50MH106435, Kent and Liz Dauten, and involved the use of instrumentation supported by the NIH Shared Instrumentation Grant Program (S10OD020039) and sequence developments supported by the Center for Functional Neuroimaging Technologies (P41EB015896).

## References

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., Raznahan, A., 2016. Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. *Hum Brain Mapp* 37, 2385-2397.

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R., Griffanti, L., Douaud, G., Sotiroopoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Dragonu, I., Matthews, P.M., Miller, K.L., Smith, S.M., 2018. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* 166, 400-424.

Aranovitch, A., Haeberlin, M., Gross, S., Dietrich, B.E., Wilm, B.J., Brunner, D.O., Schmid, T., Luechinger, R., Pruessmann, K.P., 2018. Prospective motion correction with NMR markers using only native sequence elements. *Magn Reson Med* 79, 2046-2056.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. *NeuroImage* 26, 839-851.

Bilgic, B., Gagoski, B.A., Cauley, S.F., Fan, A.P., Polimeni, J.R., Grant, P.E., Wald, L.L., Setsompop, K., 2015. Wave-CAIPI for highly accelerated 3D imaging. *Magn Reson Med* 73, 2152-2162.

Braga, R.M., Buckner, R.L., 2017. Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron* 95, 457-471 e455.

Breuer, F.A., Blaimer, M., Mueller, M.F., Seiberlich, N., Heidemann, R.M., Griswold, M.A., Jakob, P.M., 2006. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med* 55, 549-556.

Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage* 23, 724-738.

Cannon, T.D., Thompson, P.M., van Erp, T.G., Toga, A.W., Poutanen, V.P., Huttunen, M., Lonnqvist, J., Standerskjold-Nordenstam, C.G., Narr, K.L., Khaledy, M., Zoumalan, C.I.,

Dail, R., Kaprio, J., 2002. Cortex mapping reveals regionally specific patterns of genetic and disease-specific gray-matter deficits in twins discordant for schizophrenia. *Proc Natl Acad Sci U S A* 99, 3228-3233.

Casanova, M.F., El-Baz, A.S., Kamat, S.S., Dombroski, B.A., Khalifa, F., Elnakib, A., Soliman, A., Allison-McNutt, A., Switala, A.E., 2013. Focal cortical dysplasias in autism spectrum disorders. *Acta Neuropathol Commun* 1, 67.

Cash, D.M., Frost, C., Iheme, L.O., Unay, D., Kandemir, M., Fripp, J., Salvado, O., Bourgeat, P., Reuter, M., Fischl, B., Lorenzi, M., Frisoni, G.B., Pennec, X., Pierson, R.K., Gunter, J.L., Senjem, M.L., Jack, C.R., Jr., Guizard, N., Fonov, V.S., Collins, D.L., Modat, M., Cardoso, M.J., Leung, K.K., Wang, H., Das, S.R., Yushkevich, P.A., Malone, I.B., Fox, N.C., Schott, J.M., Ourselin, S., 2015. Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage* 123, 149-164.

Crawley, A.P., Henkelman, R.M., 1988. A comparison of one-shot and recovery methods in T1 imaging. *Magn Reson Med* 7, 23-34.

Crawley, A.P., Wood, M.L., Henkelman, R.M., 1988. Elimination of transverse coherences in FLASH MRI. *Magn Reson Med* 8, 248-260.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9, 179-194.

Desikan, R.S., Segonne, 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., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968-980.

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., 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, 497-510.

Ekstrom, A.D., Bazih, A.J., Suthana, N.A., Al-Hakim, R., Ogura, K., Zeineh, M., Burggren, A.C., Bookheimer, S.Y., 2009. Advances in high-resolution imaging and computational unfolding of the human hippocampus. *NeuroImage* 47, 42-49.

Esteban, O., Birman, D., Schaer, M., Koyejo, O.O., Poldrack, R.A., Gorgolewski, K.J., 2017. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* 12, e0184661.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A* 97, 11050-11055.

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., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341-355.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9, 195-207.

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8, 272-284.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14, 11-22.

Frahm, J., Haase, A., Matthaei, D., 1986. Rapid three-dimensional MR imaging using the FLASH technique. *J Comput Assist Tomogr* 10, 363-368.

Gaser, C., 2014. Voxel-based morphometry extension to SPM8.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci* 31, 11597-11616.

Goldman, A.L., Pezawas, L., Mattay, V.S., Fischl, B., Verchinski, B.A., Chen, Q., Weinberger, D.R., Meyer-Lindenberg, A., 2009. Widespread reductions of cortical thickness in schizophrenia and spectrum disorders and evidence of heritability. *Arch Gen Psychiatry* 66, 467-477.

Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 47, 1202-1210.

Harms, M.P., Somerville, L.H., Ances, B.M., Andersson, J., Barch, D.M., Bastiani, M., Bookheimer, S.Y., Brown, T.B., Buckner, R.L., Burgess, G.C., Coalson, T.S., Chappell,

M.A., Dapretto, M., Douaud, G., Fischl, B., Glasser, M.F., Greve, D.N., Hodge, C., Jamison, K.W., Jbabdi, S., Kandala, S., Li, X., Mair, R.W., Mangia, S., Marcus, D., Mascali, D., Moeller, S., Nichols, T.E., Robinson, E.C., Salat, D.H., Smith, S.M., Sotropoulos, S.N., Terpstra, M., Thomas, K.M., Tisdall, M.D., Ugurbil, K., van der Kouwe, A., Woods, R.P., Zollei, L., Van Essen, D.C., Yacoub, E., 2018. Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage* 183, 972–984.

Holmes, A.J., Hollinshead, M.O., O'Keefe, T.M., Petrov, V.I., Fariello, G.R., Wald, L.L., Fischl, B., Rosen, B.R., Mair, R.W., Roffman, J.L., Smoller, J.W., Buckner, R.L., 2015. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Sci Data* 2, 150031.

Jack, C.R., Jr., Barnes, J., Bernstein, M.A., Borowski, B.J., Brewer, J., Clegg, S., Dale, A.M., Carmichael, O., Ching, C., DeCarli, C., Desikan, R.S., Fennema-Notestine, C., Fjell, A.M., Fletcher, E., Fox, N.C., Gunter, J., Gutman, B.A., Holland, D., Hua, X., Insel, P., Kantarci, K., Killiany, R.J., Krueger, G., Leung, K.K., Mackin, S., Maillard, P., Malone, I.B., Mattsson, N., McEvoy, L., Modat, M., Mueller, S., Nosheny, R., Ourselin, S., Schuff, N., Senjem, M.L., Simonson, A., Thompson, P.M., Rettmann, D., Vemuri, P., Walhovd, K., Zhao, Y., Zuk, S., Weiner, M., 2015. Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2. *Alzheimers Dement* 11, 740-756.

Jack, C.R., Jr., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., Killiany, R.J., Decarli, C.S., Dale, A.M., Carmichael, O.W., Tosun, D., Weiner, M.W., Alzheimer's Disease Neuroimaging, I., 2010. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement* 6, 212-220.

Jack, C.R., Jr., Petersen, R.C., O'Brien, P.C., Tangalos, E.G., 1992. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology* 42, 183-188.

Jack, C.R., Jr., Twomey, C.K., Zinsmeister, A.R., Sharbrough, F.W., Petersen, R.C., Cascino, G.D., 1989. Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. *Radiology* 172, 549-554.

Kuperberg, G.R., Broome, M.R., McGuire, P.K., David, A.S., Eddy, M., Ozawa, F., Goff, D., West, W.C., Williams, S.C., van der Kouwe, A.J., Salat, D.H., Dale, A.M., Fischl, B., 2003.

Regionally localized thinning of the cerebral cortex in schizophrenia. *Arch Gen Psychiatry* 60, 878-888.

Lerch, J.P., van der Kouwe, A.J., Raznahan, A., Paus, T., Johansen-Berg, H., Miller, K.L., Smith, S.M., Fischl, B., Sotropoulos, S.N., 2017. Studying neuroanatomy using MRI. *Nat Neurosci* 20, 314-326.

Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J.L., Duerk, J.L., Griswold, M.A., 2013. Magnetic resonance fingerprinting. *Nature* 495, 187-192.

Ma, D., Jiang, Y., Chen, Y., McGivney, D., Mehta, B., Gulani, V., Griswold, M., 2018. Fast 3D magnetic resonance fingerprinting for a whole-brain coverage. *Magn Reson Med* 79, 2190-2197.

Mair, R.W., van der Kouwe, A.J., Benner, T., Fischl, B., Buckner, R.L., 2012. Quantitative Comparison of Extremely Rapid Structural Data Acquisition Compared to Conventional MPRAGE. *Proceedings of the 20th Annual Meeting of ISMRM*, Melbourne, Australia.

Marcus, D.S., Harwell, J., Olsen, T., Hodge, M., Glasser, M.F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S.W., Van Essen, D.C., 2011. Informatics and data mining tools and strategies for the human connectome project. *Front Neuroinform* 5, 4.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci* 19, 1498-1507.

Marques, J.P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.F., Gruetter, R., 2010. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage* 49, 1271-1281.

Mayhew, T.M., 1992. A review of recent advances in stereology for quantifying neural structure. *J Neurocytol* 21, 313-328.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotropoulos, S.N., Andersson, J.L., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 19, 1523-1536.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* 63, 1144-1153.

Mugler, J.P., 3rd, Brookeman, J.R., 1990. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn Reson Med* 15, 152-157.

Polak, D., Setsompop, K., Cauley, S.F., Gagoski, B.A., Bhat, H., Maier, F., Bachert, P., Wald, L.L., Bilgic, B., 2018. Wave-CAIPI for highly accelerated MP-RAGE imaging. *Magn Reson Med* 79, 401-406.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med* 42, 952-962.

Raz, N., Gunning, F.M., Head, D., Dupuis, J.H., McQuain, J., Briggs, S.D., Loken, W.J., Thornton, A.E., Acker, J.D., 1997. Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. *Cereb Cortex* 7, 268-282.

Raznahan, A., Lerch, J.P., Lee, N., Greenstein, D., Wallace, G.L., Stockman, M., Clasen, L., Shaw, P.W., Giedd, J.N., 2011. Patterns of coordinated anatomical change in human cortical development: a longitudinal neuroimaging study of maturational coupling. *Neuron* 72, 873-884.

Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. *NeuroImage* 53, 1181-1196.

Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J., Fischl, B., 2015. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage* 107, 107-115.

Runge, V.M., Kirsch, J.E., Thomas, G.S., Mugler, J.P., 3rd, 1991. Clinical comparison of three-dimensional MP-RAGE and FLASH techniques for MR imaging of the head. *J Magn Reson Imaging* 1, 493-500.

Segonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22, 1060-1075.

Setsompop, K., Cohen-Adad, J., Gagoski, B.A., Raji, T., Yendiki, A., Keil, B., Wedeen, V.J., Wald, L.L., 2012a. Improving diffusion MRI using simultaneous multi-slice echo planar imaging. *NeuroImage* 63, 569-580.

Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012b. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn Reson Med* 67, 1210-1224.

Shafee, R., Buckner, R.L., Fischl, B., 2015. Gray matter myelination of 1555 human brains using partial volume corrected MRI images. *NeuroImage* 105, 473-485.

Sheline, Y.I., Wang, P.W., Gado, M.H., Csernansky, J.G., Vannier, M.W., 1996. Hippocampal atrophy in recurrent major depression. *Proc Natl Acad Sci U S A* 93, 3908-3913.

Sliwinski, M.J., 2008. Measurement-Burst Designs for Social Health Research. *Social and Personality Psychology Compass* 2, 245-261.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 17, 479-489.

Somerville, L.H., Bookheimer, S.Y., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Dapretto, M., Elam, J.S., Gaffrey, M.S., Harms, M.P., Hodge, C., Kandala, S., Kastman, E.K., Nichols, T.E., Schlaggar, B.L., Smith, S.M., Thomas, K.M., Yacoub, E., Van Essen, D.C., Barch, D.M., submitted. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5-21 year olds. *NeuroImage*.

Stoner, R., Chow, M.L., Boyle, M.P., Sunkin, S.M., Mouton, P.R., Roy, S., Wynshaw-Boris, A., Colamarino, S.A., Lein, E.S., Courchesne, E., 2014. Patches of disorganization in the neocortex of children with autism. *N Engl J Med* 370, 1209-1219.

Tamnes, C.K., Ostby, Y., Fjell, A.M., Westlye, L.T., Due-Tonnessen, P., Walhovd, K.B., 2010. Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure. *Cereb Cortex* 20, 534-548.

Tassi, L., Colombo, N., Garbelli, R., Francione, S., Lo Russo, G., Mai, R., Cardinale, F., Cossu, M., Ferrario, A., Galli, C., Bramerio, M., Citterio, A., Spreafico, R., 2002. Focal cortical dysplasia: neuropathological subtypes, EEG, neuroimaging and surgical outcome. *Brain* 125, 1719-1732.

Tisdall, M.D., Hess, A.T., Reuter, M., Meintjes, E.M., Fischl, B., van der Kouwe, A.J., 2012. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn Reson Med* 68, 389-399.

Tisdall, M.D., Reuter, M., Qureshi, A., Buckner, R.L., Fischl, B., van der Kouwe, A.J.W., 2016. Prospective motion correction with volumetric navigators (vNavs) reduces the bias and variance in brain morphometry induced by subject motion. *NeuroImage* 127, 11-22.

Triantafyllou, C., Polimeni, J.R., Wald, L.L., 2011. Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils, *NeuroImage* 55, 597–606.

Ugurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotropoulos, S.N., Jbabdi, S., Andersson, J.L., Behrens, T.E., Glasser, M.F., Van Essen, D.C., Yacoub, E., Consortium, W.U.-M.H., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage* 80, 80-104.

van der Kouwe A.J.W., Benner T., Dale A.M., 2006. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med* 56, 1019–1032.

van der Kouwe, A.J.W., Benner, T., Salat, D.H., Fischl, B., 2008. Brain morphometry with multiecho MPRAGE. *NeuroImage* 40, 559-569.

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., Consortium, W.U.-M.H., 2012. The Human Connectome Project: a data acquisition perspective. *NeuroImage* 62, 2222-2231.

Welch E.B., Manduca A., Grimm R.C., Ward H.A., Jack C.R., 2002. Spherical navigator echoes for full 3d rigid body motion measurement in MRI. *Magn Reson Med* 47, 32–41.

White, N., Roddey, C., Shankaranarayanan, A., Han, E., Rettmann, D., Santos, J., Kuperman, J., Dale, A., 2010. PROMO: Real-time prospective motion correction in MRI using image-based tracking. *Magn Reson Med* 63, 91-105.

Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wiggins, G.C., Triantafyllou, C., Potthast, A., Reykowski, A., Nittka, M., Wald, L.L., 2006. 32-channel 3 Tesla receive-only phased-array head coil with soccer-ball element geometry. *Magn Reson Med* 56, 216-223.

Zaitsev M., Dold C., Sakas G., Hennig J., Speck O., 2006. Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *NeuroImage* 31, 1038-1050.

Zaretskaya, N., Fischl, B., Reuter, M., Renvall, V., Polimeni, J.R., 2018. Advantages of cortical surface reconstruction using submillimeter 7 T MEMPRAGE. *NeuroImage* 165, 11-26.

## Tables

Table 1. Acquisition details across the 6 studies.

Study	Subjects	n	# of Sessions	Scans Per Session	Sequence	Voxel Size (mm <sup>3</sup> )	TA (min, sec)	R
1	S1-S6 <sup>a</sup>	6	1	4	Standard	1.0	6'12"	2
				2	Standard	1.0	3'22"	4
				20 <sup>b</sup>	Wave-CAIPI	1.0	1'16"	3x3
				2	Wave-CAIPI	1.0	2'45"	2x2
2	S7-S12 <sup>a</sup>	6	1	4	Standard	1.0	6'12"	2
				24	Wave-CAIPI	1.2	1'01"	3x3
3	S13-S17	5	3	20 <sup>c</sup>	Multi-echo	1.2 <sup>d</sup>	2'23" <sup>e</sup>	4
4	S18-S23	6	2	10	Multi-echo	1.2 <sup>d</sup>	2'23" <sup>e</sup>	4
				4	Multi-echo	1.0	5'58"	2
5	S24-S28	5	5-14	2	Multi-echo	1.2 <sup>d</sup>	2'12"	4
6	S29-S34	6	24-30	1	Multi-echo	1.2 <sup>d</sup>	2'12"	4

Notes: Time of acquisition (TA), acceleration factor (R). <sup>a</sup>32-channel head coil rather than 64-channel head coil; <sup>b</sup>One subject had 17 scans per session; <sup>c</sup>One subject had 16 scans per session; <sup>d</sup>Phase partial Fourier: 6/8; <sup>e</sup>Scan time included 11 seconds to accommodate interspersed acquisition of motion navigators.

Table 2. Estimates of achieved test-retest error using cluster scanning within individuals.

Morphometric Measure	1 Scan	3 Scans	6 Scans	12 Scans
eTIV	0.20% (0.17 - 0.25%) [3047 mm <sup>3</sup> ]	0.13% (0.10 - 0.16%) [1962 mm <sup>3</sup> ]	0.09% (0.06 - 0.11%) [1301 mm <sup>3</sup> ]	0.06% (0.04 - 0.08%) [886 mm <sup>3</sup> ]
TBV	0.60% (0.40 - 0.72%) [6669 mm <sup>3</sup> ]	0.40% (0.26 - 0.51%) [4401 mm <sup>3</sup> ]	0.27% (0.19 - 0.35%) [2985 mm <sup>3</sup> ]	0.18% (0.15 - 0.22%) [2042 mm <sup>3</sup> ]
GM	0.99% (0.64 - 1.41%) [4584 mm <sup>3</sup> ]	0.66% (0.40 - 0.86%) [3027 mm <sup>3</sup> ]	0.46% (0.26 - 0.67%) [2086 mm <sup>3</sup> ]	0.35% (0.18 - 0.61%) [1606 mm <sup>3</sup> ]
WM	0.59% (0.40 - 0.97%) [2720 mm <sup>3</sup> ]	0.39% (0.26 - 0.65%) [1807 mm <sup>3</sup> ]	0.27% (0.17 - 0.45%) [1242 mm <sup>3</sup> ]	0.20% (0.11 - 0.35%) [925 mm <sup>3</sup> ]
Cortical Surface Area	0.49% (0.37 - 0.67%) [825 mm <sup>2</sup> ]	0.34% (0.24 - 0.51%) [571 mm <sup>2</sup> ]	0.24% (0.17 - 0.37%) [399 mm <sup>2</sup> ]	0.18% (0.13 - 0.30%) [304 mm <sup>2</sup> ]
Mean Cortical Thickness	0.93% (0.65 - 1.32%) [23 µm]	0.61% (0.40 - 0.86%) [15 µm]	0.43% (0.26 - 0.61%) [11 µm]	0.33% (0.13 - 0.45%) [8 µm]
AD Cortical Thickness	1.36% (1.07 - 2.07%) [37 µm]	0.96% (0.69 - 1.51%) [26 µm]	0.67% (0.43 - 1.07%) [18 µm]	0.51% (0.30 - 0.87%) [14 µm]
Hippocampus	2.17% (1.04 - 4.12%) [185 mm <sup>3</sup> ]	1.00% (0.68 - 1.30%) [84 mm <sup>3</sup> ]	0.62% (0.42 - 0.80%) [52 mm <sup>3</sup> ]	0.44% (0.23 - 0.56%) [37 mm <sup>3</sup> ]
Thalamus	4.59% (2.57 - 7.65%) [656 mm <sup>3</sup> ]	2.55% (1.83 - 4.06%) [371 mm <sup>3</sup> ]	1.57% (1.26 - 2.45%) [231 mm <sup>3</sup> ]	1.02% (0.44 - 1.51%) [153 mm <sup>3</sup> ]
Caudate	2.36% (0.80 - 4.14%) [168 mm <sup>3</sup> ]	1.63% (0.54 - 3.00%) [115 mm <sup>3</sup> ]	1.13% (0.38 - 2.17%) [80 mm <sup>3</sup> ]	0.85% (0.25 - 1.83%) [60 mm <sup>3</sup> ]
Putamen	3.52% (2.85 - 4.23%) [385 mm <sup>3</sup> ]	2.26% (1.94 - 3.01%) [248 mm <sup>3</sup> ]	1.52% (1.23 - 2.20%) [166 mm <sup>3</sup> ]	0.94% (0.62 - 1.77%) [102 mm <sup>3</sup> ]
Amygdala	5.00% (3.79 - 6.46%) [166 mm <sup>3</sup> ]	3.17% (2.24 - 4.36%) [105 mm <sup>3</sup> ]	2.16% (1.45 - 3.00%) [72 mm <sup>3</sup> ]	1.60% (0.90 - 2.58%) [54 mm <sup>3</sup> ]
Corpus Callosum	1.42% (0.56 - 2.2%) [43 mm <sup>3</sup> ]	0.88% (0.44 - 1.42%) [27 mm <sup>3</sup> ]	0.58% (0.33 - 0.94%) [18 mm <sup>3</sup> ]	0.40% (0.28 - 0.48%) [12 mm <sup>3</sup> ]

Left Posterior Cingulate	2.29% (1.81 - 2.62%) [61 $\mu$ m]	1.62% (1.18 - 1.87%) [43 $\mu$ m]	1.16% (0.79 - 1.43%) [31 $\mu$ m]	0.97% (0.66 - 1.41%) [26 $\mu$ m]
Left Fusiform	1.80% (1.63 - 1.95%) [51 $\mu$ m]	1.15% (0.91 - 1.35%) [32 $\mu$ m]	0.79% (0.61 - 0.99%) [22 $\mu$ m]	0.58% (0.31 - 0.83%) [16 $\mu$ m]
Left Superior Frontal	1.66% (1.27 - 2.25%) [46 $\mu$ m]	1.19% (0.82 - 1.56%) [33 $\mu$ m]	0.84% (0.53 - 1.10%) [23 $\mu$ m]	0.60% (0.33 - 0.85%) [17 $\mu$ m]
Left Lateral Occipital	1.56% (1.28 - 2.00%) [33 $\mu$ m]	1.03% (0.86 - 1.39%) [22 $\mu$ m]	0.69% (0.54 - 0.94%) [15 $\mu$ m]	0.49% (0.33 - 0.59%) [10 $\mu$ m]
Left Inferior Parietal	1.80% (1.39 - 2.48%) [44 $\mu$ m]	1.27% (0.96 - 1.91%) [31 $\mu$ m]	0.91% (0.68 - 1.39%) [22 $\mu$ m]	0.76% (0.43 - 1.15%) [19 $\mu$ m]
Left Superior Temporal	1.52% (1.05 - 2.27%) [44 $\mu$ m]	1.00% (0.67 - 1.27%) [29 $\mu$ m]	0.70% (0.44 - 0.92%) [20 $\mu$ m]	0.50% (0.26 - 0.74%) [14 $\mu$ m]

Notes: Total Intracranial Volume (eTIV), Total Brain Volume (TBV), Gray Matter volume (GM), White Matter volume (WM). The descriptive statistics are presented for Study 6's 6 subjects in the following format: mean % error (range of mean % error), [mean absolute error] (when using the median procedure described in the text and illustrated in the third column of Figure 16).

## Figure Legends

**Figure 1.** Calculating error in morphometric measures derived from images acquired within a single scanning session, across scanning sessions, and using the cluster scanning approach. An example error analysis is depicted for Gray Matter volume (blue cortical ribbon highlighted in axial slices). For the within-session error estimate (yellow boxes), the morphometric measures for scan 1 and scan 2 of session 1 are averaged. The morphometric measures for scan 3 and scan 4 of session 1 are also averaged. Then, the error is calculated by comparing the two average estimates. For the between-session error estimate (blue boxes), the average estimates are calculated from scans 1 and 2 of two separate sessions before comparing the two average estimates. For the cluster scanning approach (red boxes), the average estimates are calculated from scans acquired from two different sessions before comparing the two average estimates.

**Figure 2.** Extremely rapid structural scans are feasible for quantitative morphometric analysis. Images show a Standard 6'12" MP-RAGE T1-weighted acquisition (Standard; left) in contrast to a rapid 1'16" Wave-CAIPI T1-weighted acquisition (Wave; right) from Study 1. Voxel resolution is 1.0mm isotropic for both. Cortical surfaces estimated from common automated procedures using FreeSurfer (see text) illustrate the white (red) and pial (yellow) surfaces. Coronal and axial slices from a single representative subject are displayed. The axial insets in the middle of the figure demonstrate the close correspondence between the cortical surfaces estimated from the two approaches. The subject's left is displayed on the left.

**Figure 3.** Global morphometric measures are reliable for rapid structural scans within individuals. eTIV (Total Intracranial Volume), TBV (Total Brain Volume), and Mean Cortical Thickness independently estimated for 28 Standard and Wave-CAIPI MP-RAGE scans are illustrated for S3 and S4 from Study 1. All data were acquired within the same session. The scan number corresponds with the order that the scans were acquired during the scanning session. Within an acquisition type, measures are stable across separate scans including for

the rapid acquisitions. Between acquisition types there are systematic differences in the morphometric measures.

**Figure 4.** Regional cortical thickness is correlated between Standard and Wave scans. The scatter plots compare regional cortical thickness estimated from a single Standard scan (x-axis) and a single Wave scan (y-axis) for subjects. Each data point represents the cortical thickness for one of the 34 left hemisphere regions in the Desikan-Killiany atlas (Desikan et al., 2006). Regional cortical thickness estimates from the Wave scans are highly correlated with those from the Standard scans ( $r^2 = 0.93-0.97$  for the 6 plots) but demonstrate a slight underestimate as evidenced by the best fit line's (red) slope falling below the dashed identity line. Note that the smallest values falling off the identity line (e.g., S4) are all near visual cortex (lingual gyrus, pericalcarine, and the cuneus).

**Figure 5.** Global morphometric error estimates are smaller for averaged estimates from rapid scans as compared to Standard long acquisitions. The percent error in the global morphometric measures from Figure 3 was calculated for 6 subjects who underwent a single scanning session to compare the Standard and Wave images. The amount of scan time was held roughly constant between the scan types by taking the mean percent difference between random pairs of Standard scans (6'12") or random averaged sets of 4 Wave scans (5'04" by averaging estimates from 4 individual 1'16" scans). This process was repeated 10,000 times for each morphometric measure of the respective image types within each subject. The data points represent the mean percent error from the 10,000 iterations. The error bars represent the standard deviation in the percent error from the 10,000 iterations. The error estimates for 12 of 18 values were smaller for the averaged estimates from Wave scans as compared to the Standard scan.

**Figure 6.** The relationship between error in the global morphometric measures and the amount of scan time for Standard and Wave scans. The percent error (y-axis) in relation to the amount of scan time (x-axis) for representative subjects (S1-S3) for the morphometric measures in Figure 3. For both Standard and Wave scans, the percent error in global morphometric measures decreases as more imaging time is included in the morphometric

averages. However, the percent error from the averaged Wave scan estimates is often smaller than that of the Standard scans. For reference, the dotted horizontal lines represent the error estimated for a single Wave image (the limit of a short duration scan), a single Standard scan (typical acquisition), and averaged from two Standard scans (the limit when averaging multiple long scans). Strikingly, one or a few rapid Wave scans is sufficient for achieving a highly-stable estimate.

**Figure 7.** Error estimates are smaller for global, subcortical, and regional morphometric measures derived from averages of rapid scans. The plots on the left side of the figure compare Standard scans to rapid Wave scans (Study 1, 1.0mm<sup>3</sup> voxel size, 1'16"), whereas the plots on the right compare Standard scans to extremely rapid Wave scans (Study 2, 1.2mm<sup>3</sup> voxel size, 1'01"). The horizontal lines of the box plot represent the 25th percentile, median, and 75th percentile in percent error, respectively. The whiskers extend to the most extreme data points that fall within 1.5 \* interquartile range above the 75th percentile or below the 25th percentile. Each data point represents the mean percent error for one of the subjects who underwent a single scanning session to compare Standard and Wave scans. The mean percent error was based on comparisons of single scans for the Standard scans (plots on left and right), four scans for the Wave (1.0mm<sup>3</sup>) scans (plots on left), and six scans for the Wave (1.2mm<sup>3</sup>) scans (plots on right). The regions depicted in the lower panel are representative left hemisphere regions of the Desikan-Killiany atlas (Desikan et al., 2006). Abbreviations: Total Intracranial Volume (eTIV), Total Brain Volume (TBV), Gray Matter volume (GM), White Matter volume (WM), Cortical Surface Area (Surf.), Mean Cortical Thickness (Thick.), thalamus volume (Thal.), caudate volume (Caud.), putamen volume (Put.), hippocampus volume (Hipp.), amygdala volume (Amyg.), corpus callosum volume (CC), posterior cingulate cortex (Post. Cing.), fusiform gyrus (Fusi.), superior frontal gyrus (Sup. Front.), lateral occipital cortex (Lat. Occ.), inferior parietal cortex (Inf. Par.), superior temporal gyrus (Sup. Temp.).

**Figure 8.** Error estimates are smaller within-session as compared to between-session. Between-session percent error estimates (red) are plotted for the global morphometric measures for subjects in Study 3 who were scanned multiple times per session and across

multiple sessions. The within-session estimate (yellow) compares only data within a single session, whereas the cluster scanning approach (blue) combines data across sessions thereby removing the between-session effects. The most likely source of the between-session error is different head positions within the head coil from session to session. We refer to combined estimates that average across sessions as cluster scanning because they pool data across multiple sessions and average across head position and other session-specific contributors to the error term.

**Figure 9.** Cluster scanning reduces error in global morphometric measures. Data from the 6 subjects in Study 4 scanned 28 times across 2 sessions are plotted to illustrate percent error. In the majority of subjects, the between-sessions error (dark and light blue) is larger than the cluster error (red and pink) for the eTIV and sporadically so for TBV and Mean Cortical Thickness. The larger between-sessions error is likely due to the subject's different head position within the head coil from session to session. The between-sessions error is similar for both  $1.0\text{mm}^3$  (dark blue) and  $1.2\text{mm}^3$  (light blue) multi-echo MP-RAGE images. No clear difference in between-session error or cluster error exists for Total Brain Volume or Mean Cortical Thickness.

**Figure 10.** Cluster scanning reduces error in global morphometric measures as a function of the number of scans included in the error estimate. Percent error estimates are plotted for two representative subjects from Study 3. As the number of scans included in the averaged morphometric measure (x-axis) increases, the amount of error (y-axis) decreases. The decrease in error occurs for both within-session and between-session error and across all three global morphometric measures. The error estimated from the cluster scanning approach follows the decline in within-session error because scans from all three sessions are included in the average morphometric measure, thereby removing the between-session error and leaving only the within-session error.

**Figure 11.** Effect of smoothing on cortical thickness error. Error estimates for a single vertex of the left superior temporal gyrus (red circle on inflated lateral view of left hemisphere) are illustrated for different levels of spatial smoothing. Plots for subjects (S13-

S17) display the decrease in error for cortical thickness (y-axis) as the level of smoothing increases (colored lines) and the number of estimates averaged together increases (x-axis). The lines depict the cluster error and include at least one scan from each of the three scanning sessions. Therefore, the x-axis begins at 3 scans and increases by intervals of 3.

**Figure 12.** Error in cortical thickness across the cerebral cortex. The amount of error was calculated at each vertex of the cerebral cortex for two individuals (S13 and S16). Before the error was calculated, a 10mm smoothing kernel was applied to the map of cortical thickness, and then the cortical thickness estimates from 12 images were randomly sampled and averaged (four images per session from three sessions). The error was then calculated as the difference between two groups of averaged cortical thickness estimates. This process was done 10,000 times at each vertex. The average error from the 10,000 iterations is displayed at each vertex of the cerebral cortex's left hemisphere. The smallest amount of error in cortical thickness is found in sensorimotor cortex, visual cortex, and lateral temporal regions ( $<20 \mu\text{m}$ ), whereas the largest amount is found in the insula, the cingulate cortex near the border with the corpus callosum, the inferior parietal lobule, and the ventral and medial surfaces of the prefrontal cortex ( $>50 \mu\text{m}$ ,  $>2.0\%$ ). The lateral temporal regions and fusiform gyrus consistently have the smallest percentage of error ( $<1.0\%$ ) when the absolute thickness of a region is taken into account.

**Figure 13.** Sources of error in morphometric measures from patients. Data from clinical outpatients who were scanned regularly over time in Study 5 allow multiple sources of error to be visualized. Each graph represents data from one individual (S24-S28). Mean Cortical Thickness is plotted from each scan, with scans from the same session connected by a line. The vertical dashed line represents the timing of the software upgrade. The software upgrade resulted in a decrease in estimated Mean Cortical Thickness as highlighted in the plot for S24. The within-session variability in the sessions leading up to the software upgrade is low compared to the change that occurs at the time of the software upgrade. The plot for S25 highlights what is likely the effect of motion on Mean Cortical Thickness, as evidenced by the high within-session variability. The plot for S27 highlights

what is likely the effect of different head positions across sessions on Mean Cortical Thickness, as evidenced by the low within-session variability and the measurable between-session variability. These are independent sources of error that can impact stability of longitudinal change estimates in morphometric analysis.

**Figure 14.** Gradient coil heating can affect morphometric measures. Data from individual subjects that each had numerous sequential scans from Study 3 illustrate an unexpected additional source of measurement error. The scatter plots show the relationship between the order in which scans (multi-echo MP-RAGE) were acquired during the session (x-axis) and eTIV (y-axis). eTIV appears to increase throughout the session likely due to the heating of the gradient coils that occurs when multiple images are acquired in rapid succession. The effect is shown for representative subjects (S15-S17) and appeared in other subjects as well. The effect is small and only apparent in highly stable morphometric measures but is likely present in other measures.

**Figure 15.** Effect of incorrect surface estimation on estimated cortical thickness. As another example of potential sources of error, a cortical thickness outlier was detected in a region of the left supramarginal gyrus (red region on inflated lateral view of left hemisphere) for a single subject, S13. The regional cortical thickness for the 11th of 16 scans acquired during the subject's scanning session (3.78mm; red data point on the scatter plot of the upper right panel) was more than three standard deviations above the mean regional cortical thickness of the 16 scans ( $3.43 \pm 0.11$ mm). The outlier in the cortical thickness estimate was due to the pial surface failing to follow an invagination in the left supramarginal gyrus (orange arrow, inset of middle panel). The error in cortical thickness decreased by ~45% when the outlier was removed from the error calculation (bottom plots).

**Figure 16.** Cluster scanning achieves extremely stable morphometric estimates. The absolute and percent error graphs for eTIV, TBV, Mean Cortical Thickness (Cort. Thick.), the mean cortical thickness derived from a mask of Alzheimer's Disease atrophy (AD Thick.; mask from Figure 3 of Dickerson et al., 2009), and hippocampal volume show how

the between-session error decreases as more scans are included in the error calculation. Each line represents an individual from Study 6 (S29-S34). The first column represents the mean absolute error. The second column represents the mean percent error. The third column represents the percent error when the median, rather than mean estimate, is calculated for each point. In instances where the data are stable (no outliers) the mean achieves a better and more consistent result. In instances where infrequent outliers are present in the data (e.g., hippocampal volume), the median percent error is able to achieve lower, more stable values. Each individual was scanned between 24 and 30 times over a 3-7-month period. Each scanning session included only one rapid multi-echo MP-RAGE scan. Therefore, all of the data points represent between-session error.

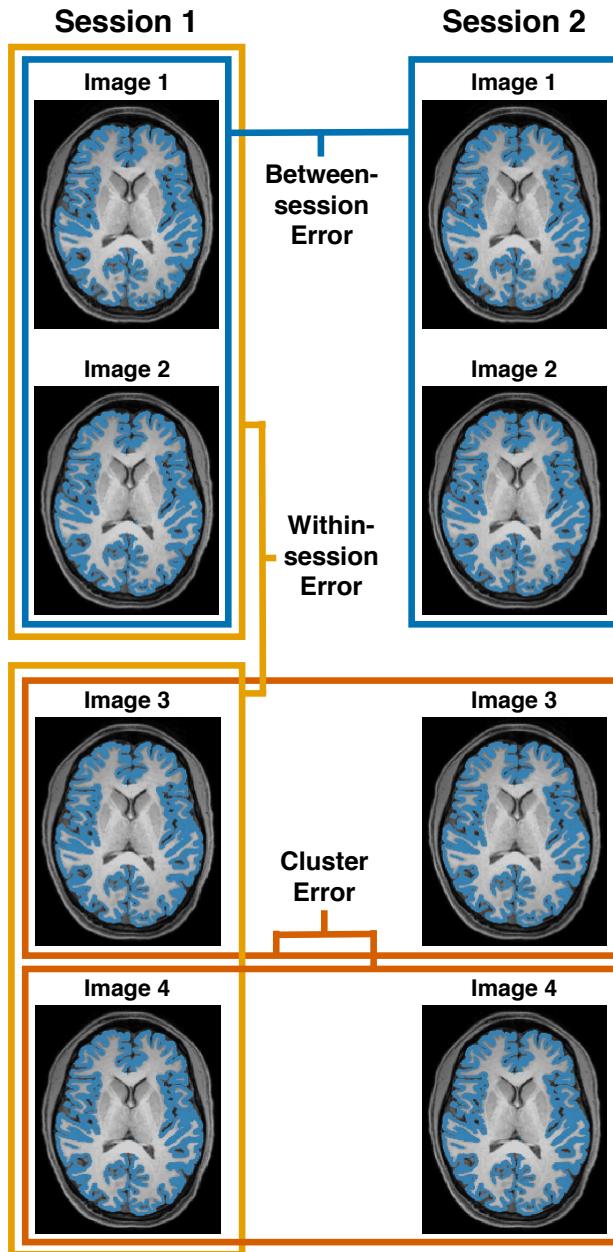


Figure 1

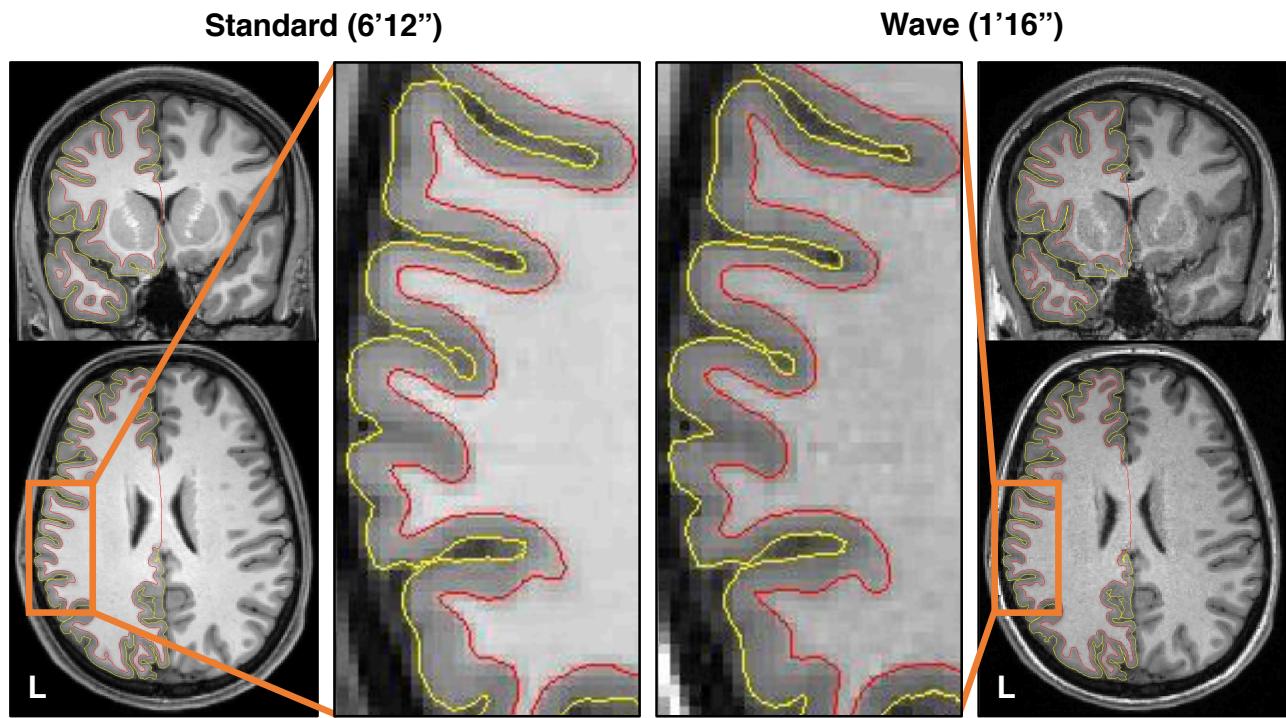


Figure 2

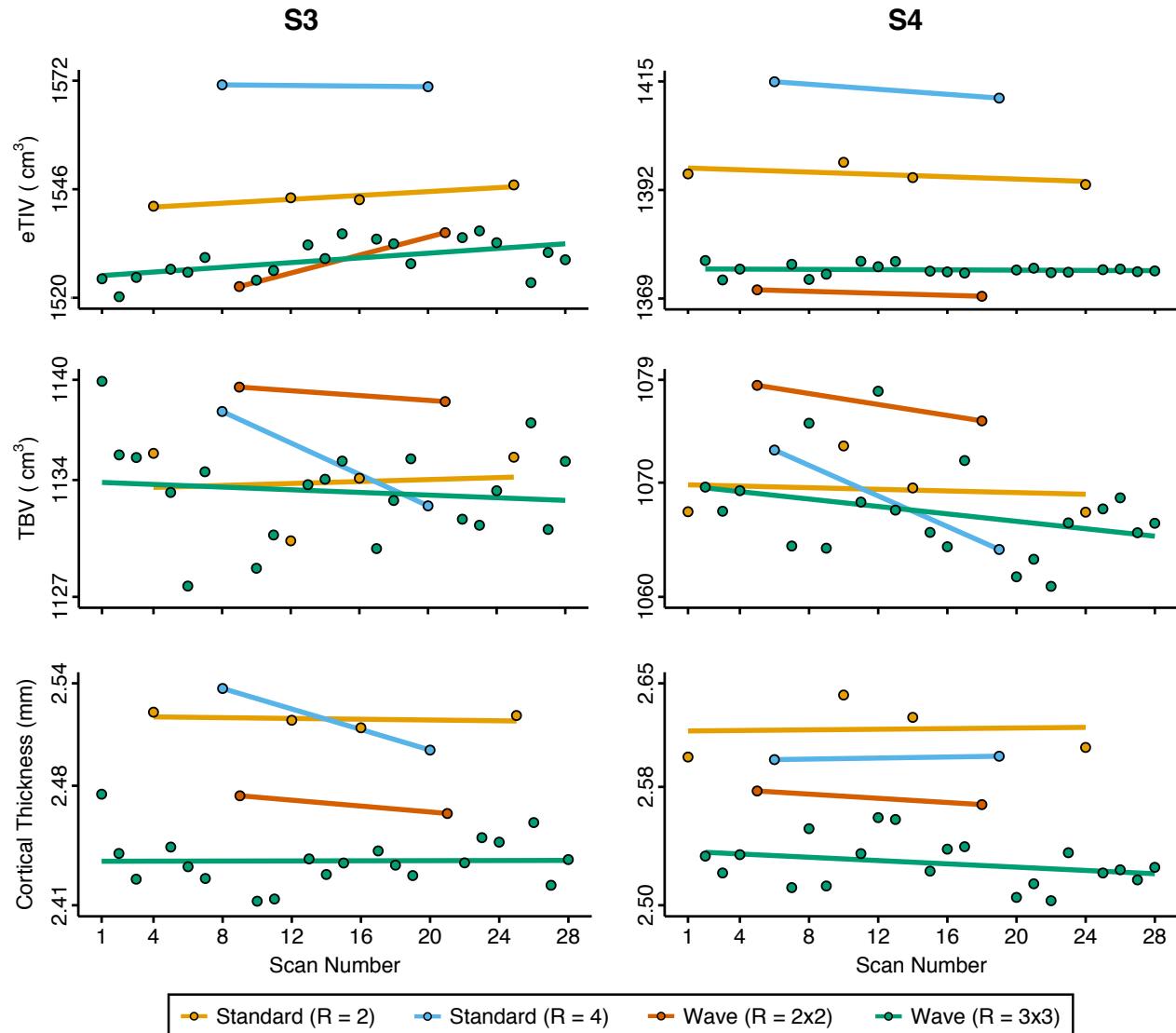


Figure 3

### Regional Cortical Thickness Correlations

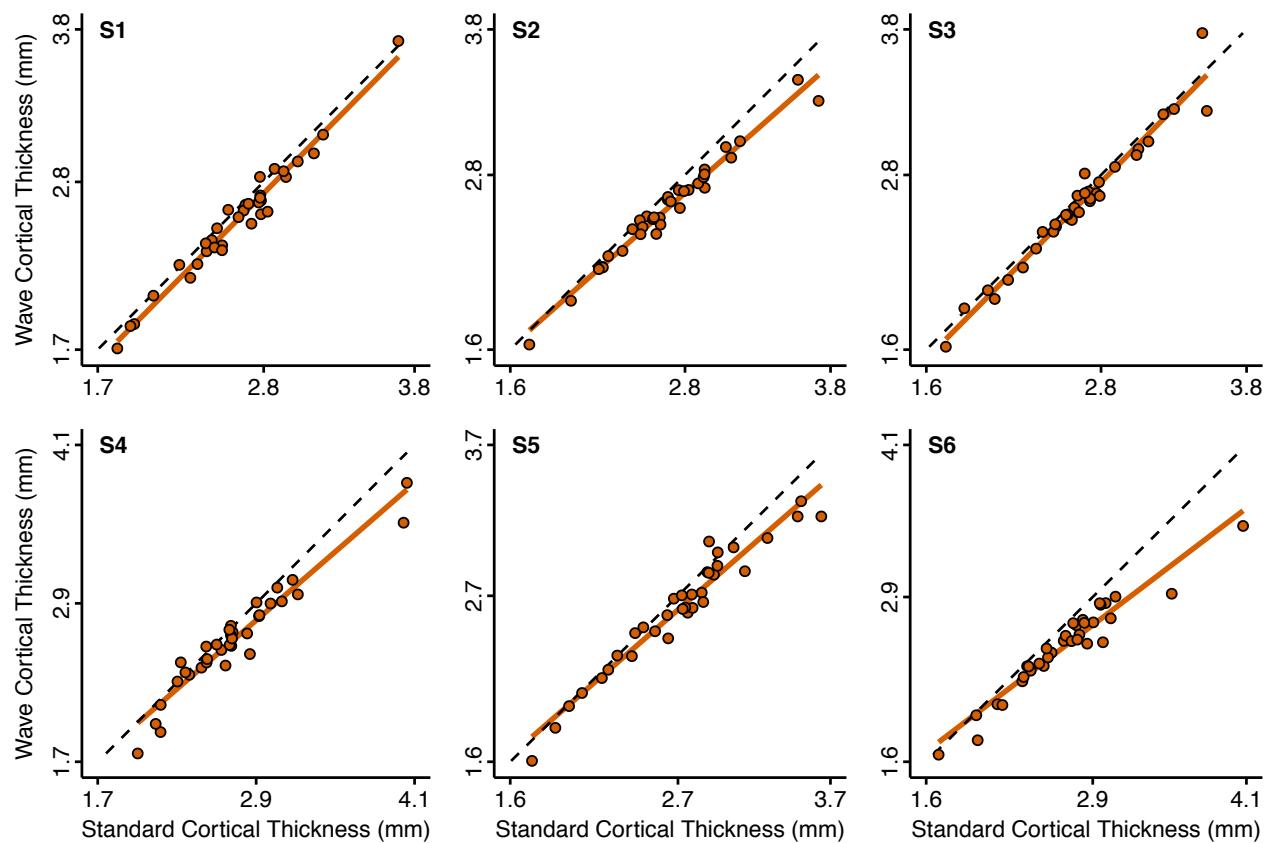


Figure 4

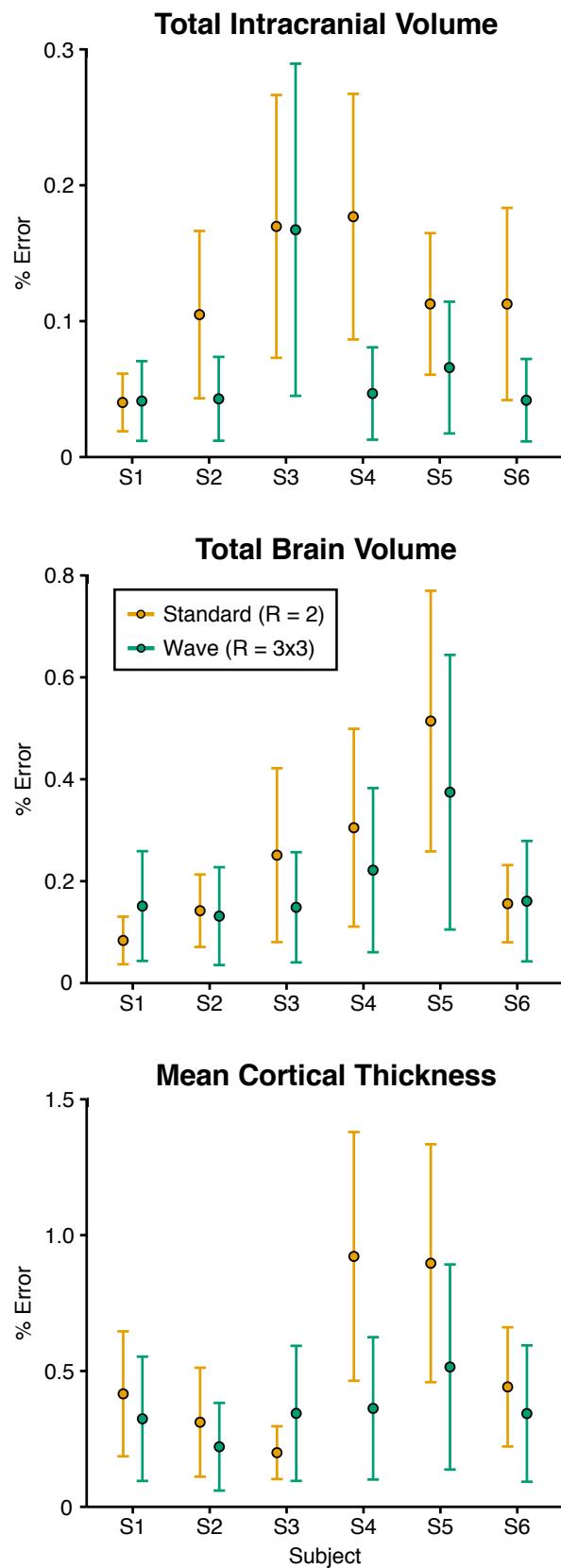


Figure 5

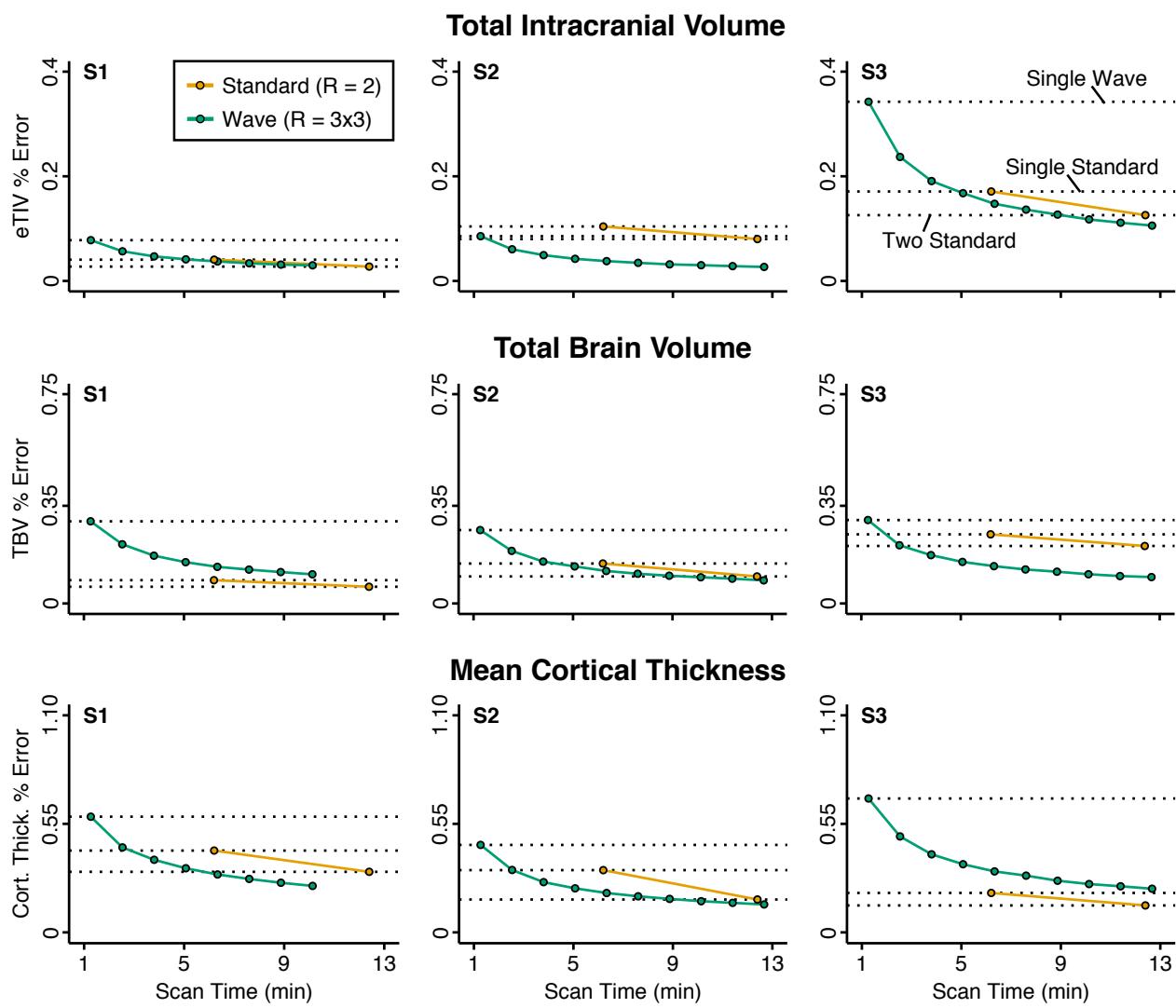


Figure 6

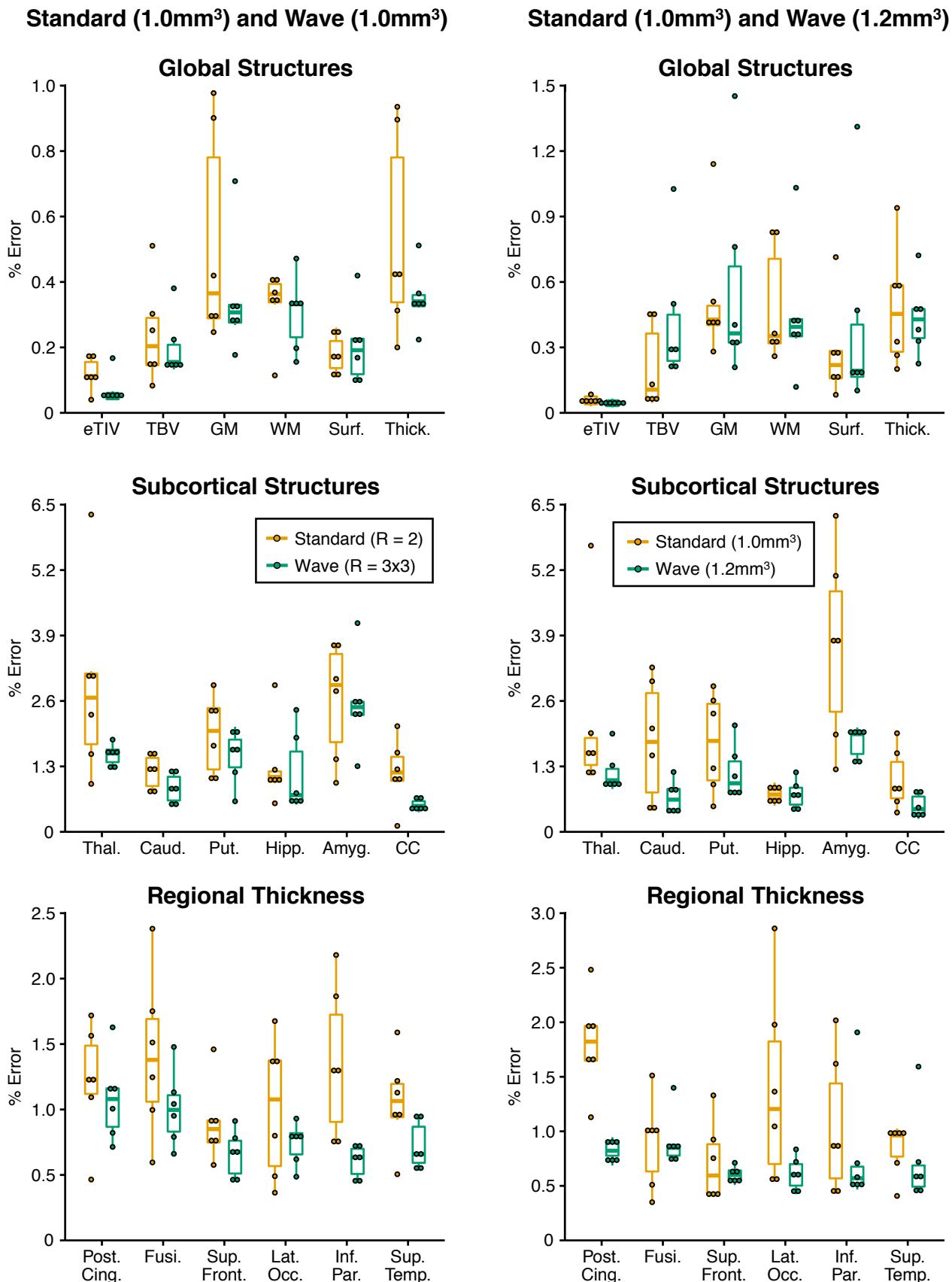


Figure 7

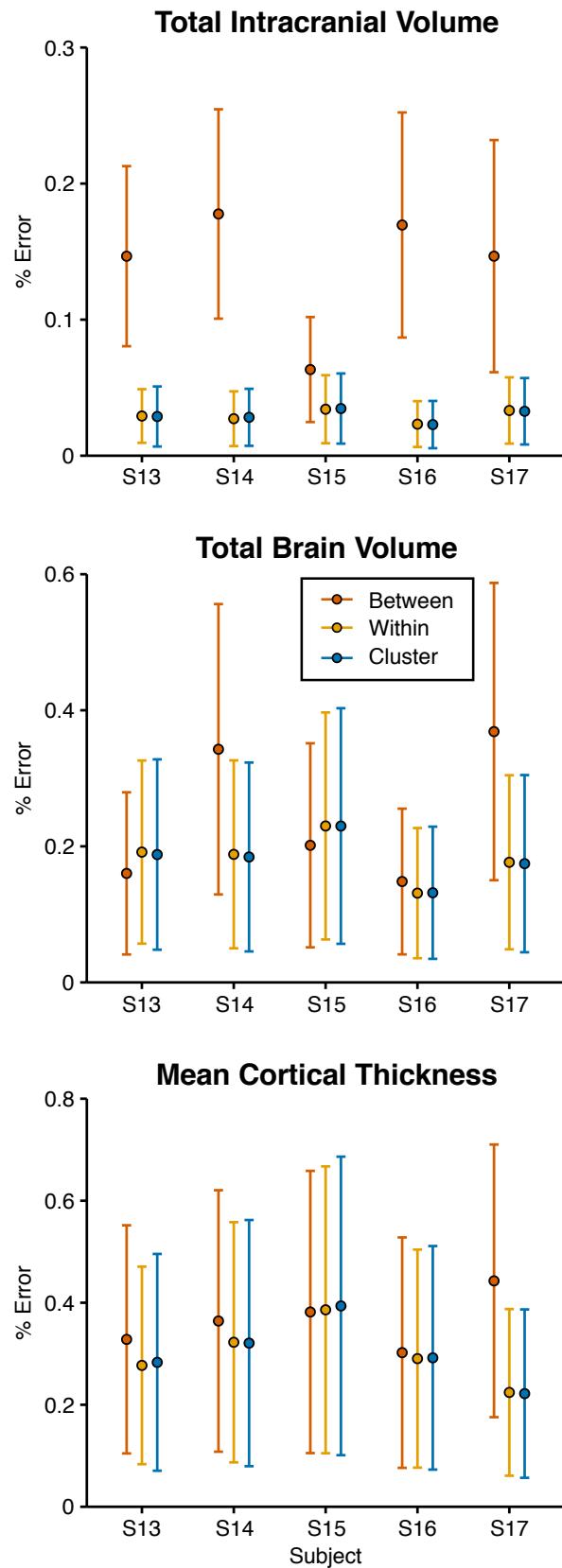


Figure 8

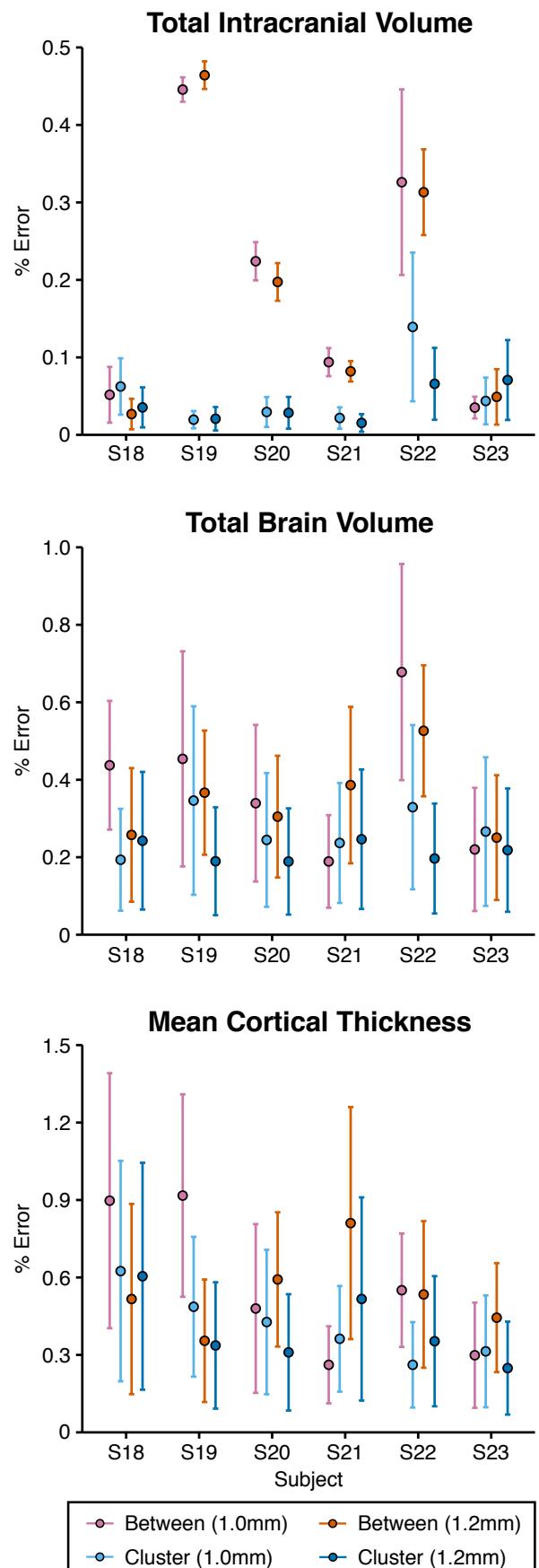


Figure 9

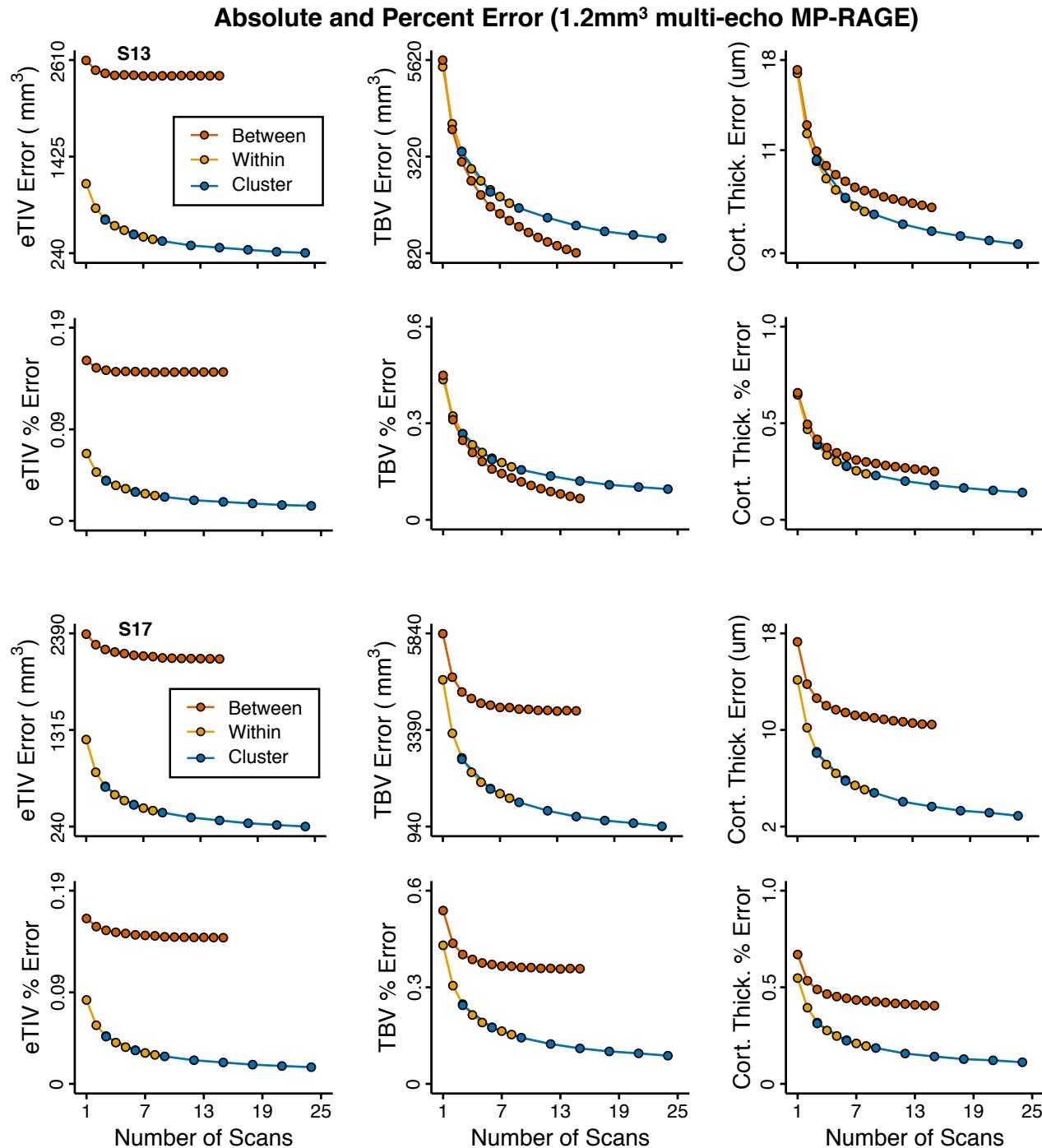


Figure 10

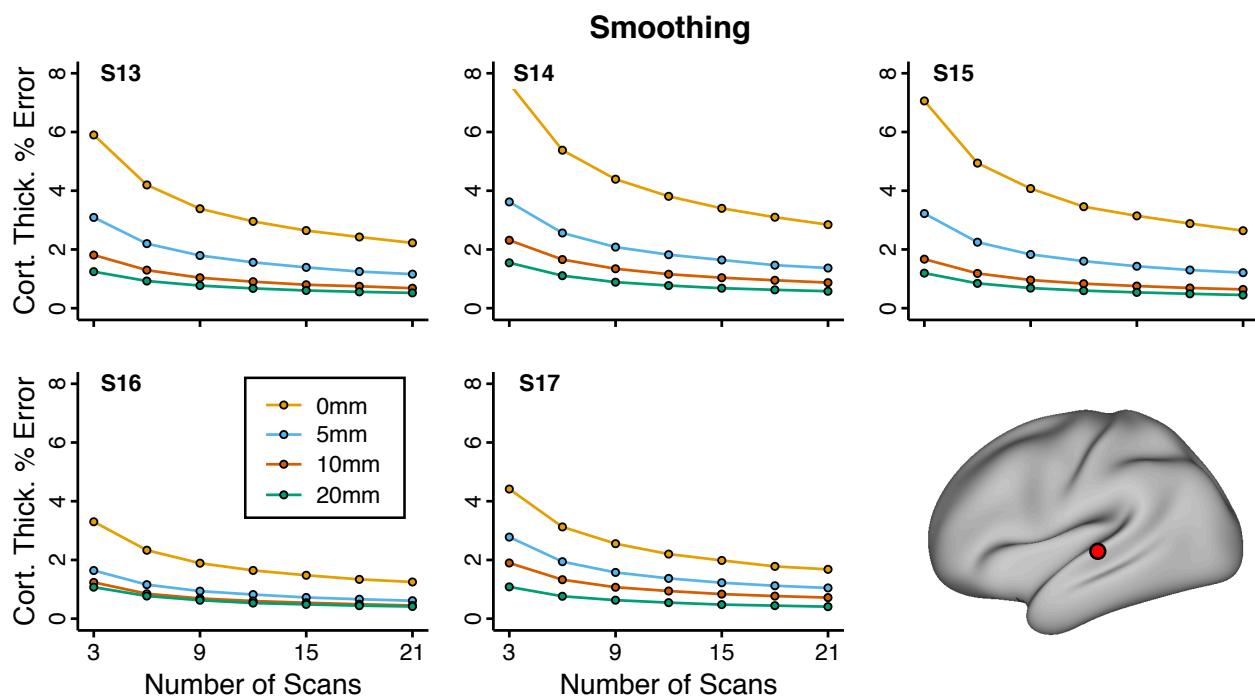


Figure 11

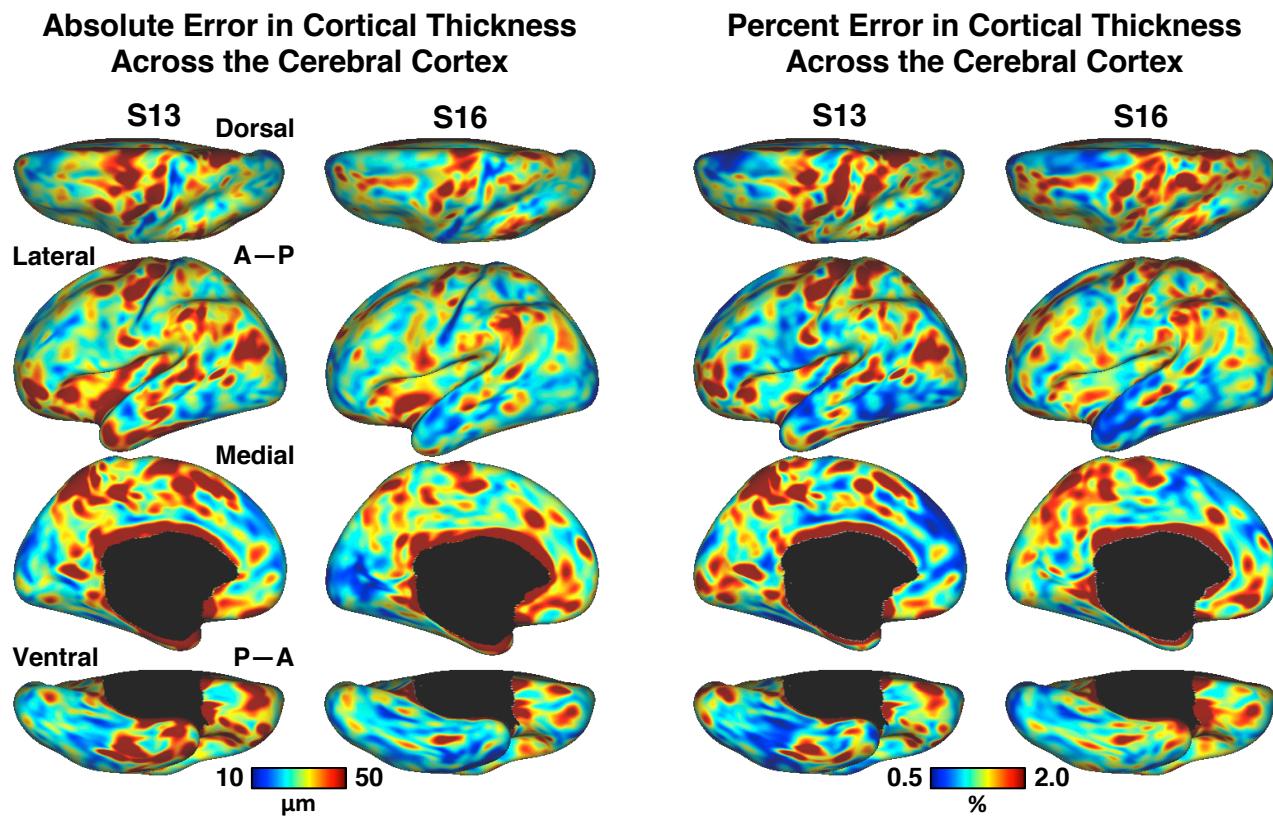


Figure 12

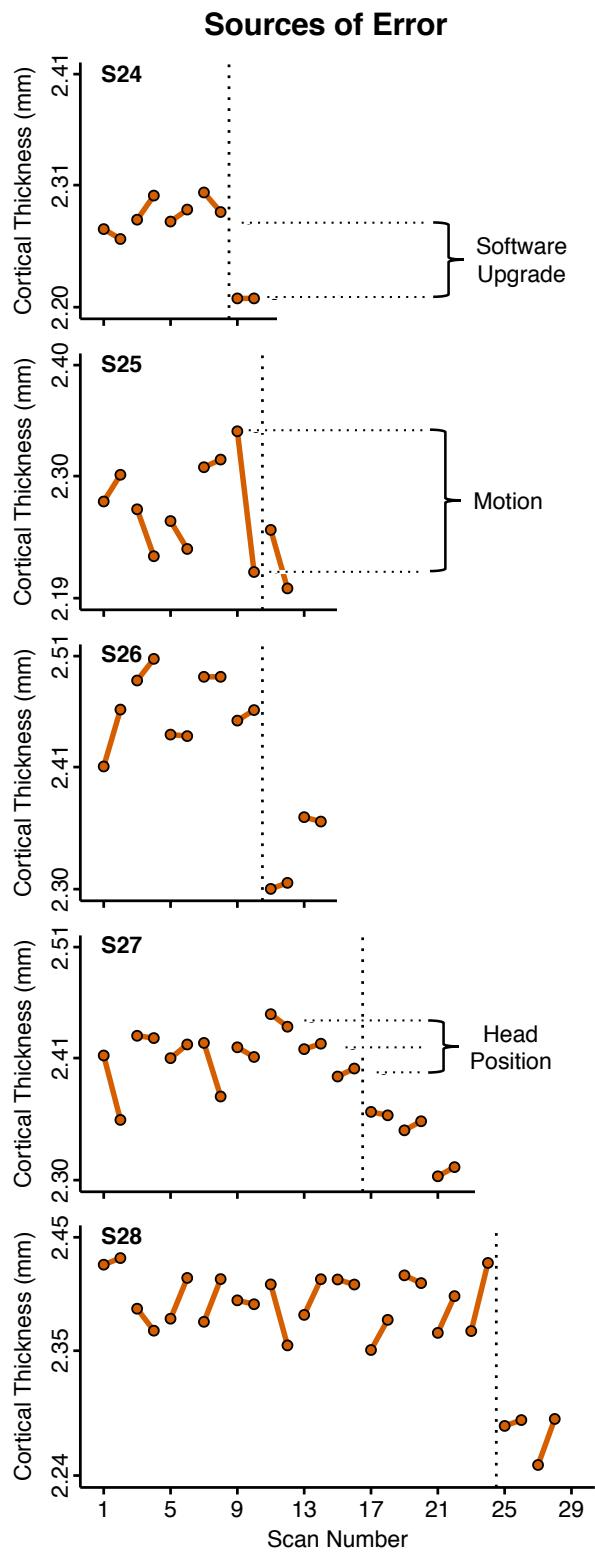


Figure 13

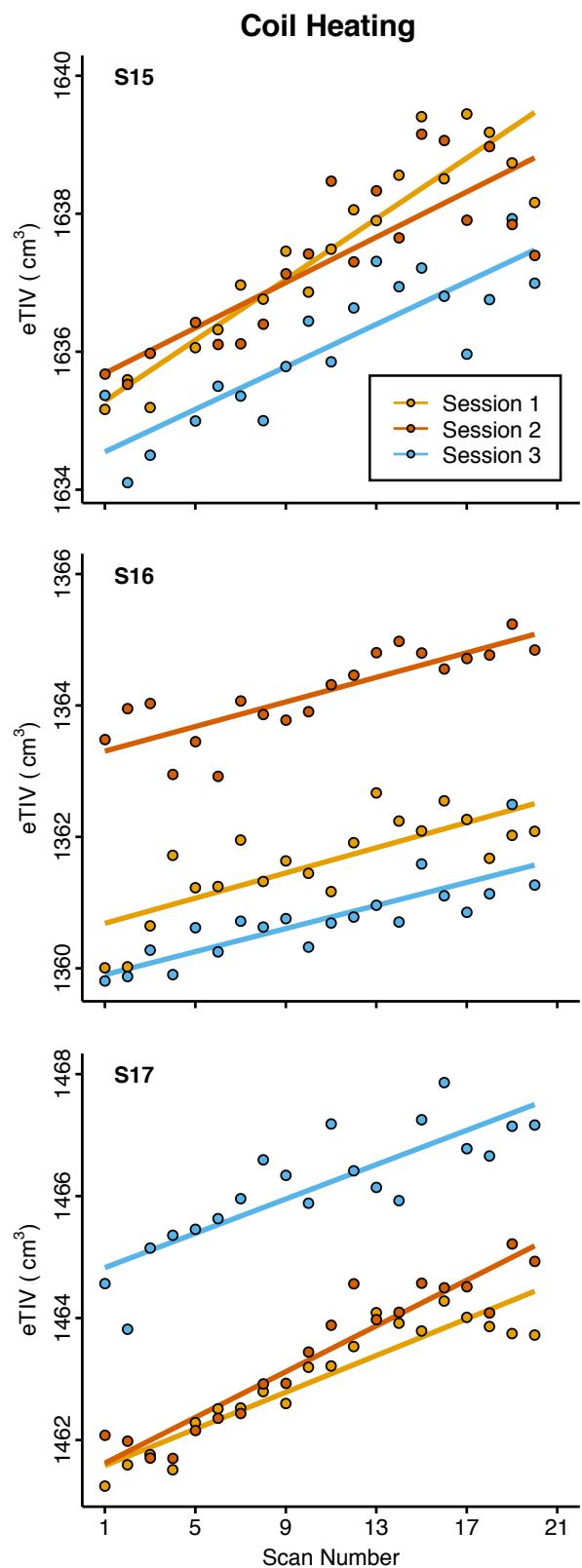


Figure 14

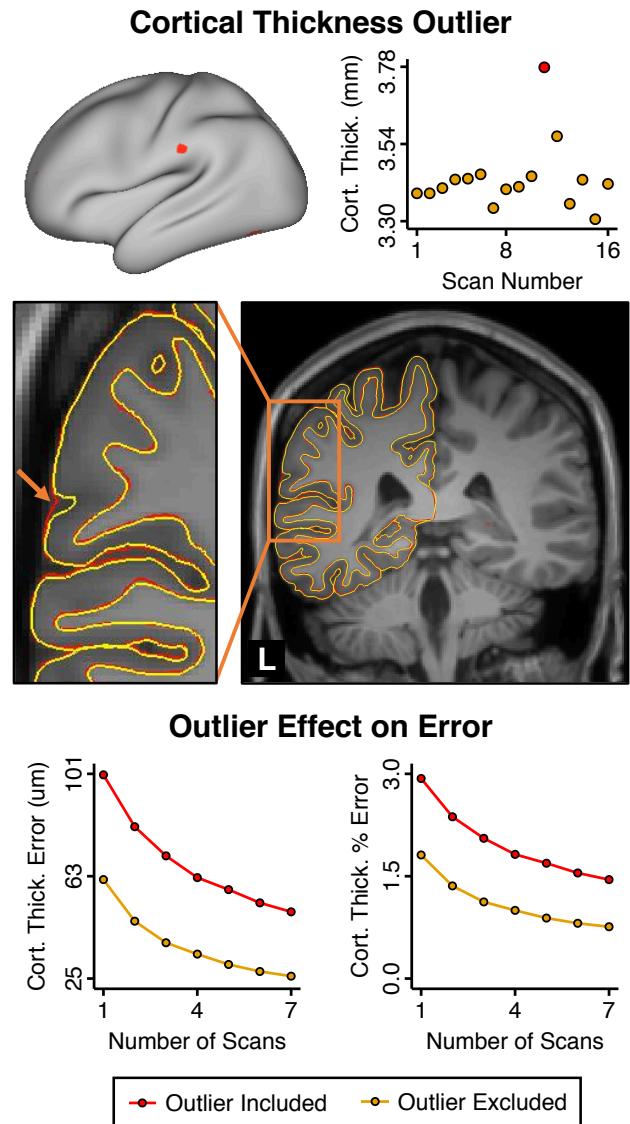


Figure 15

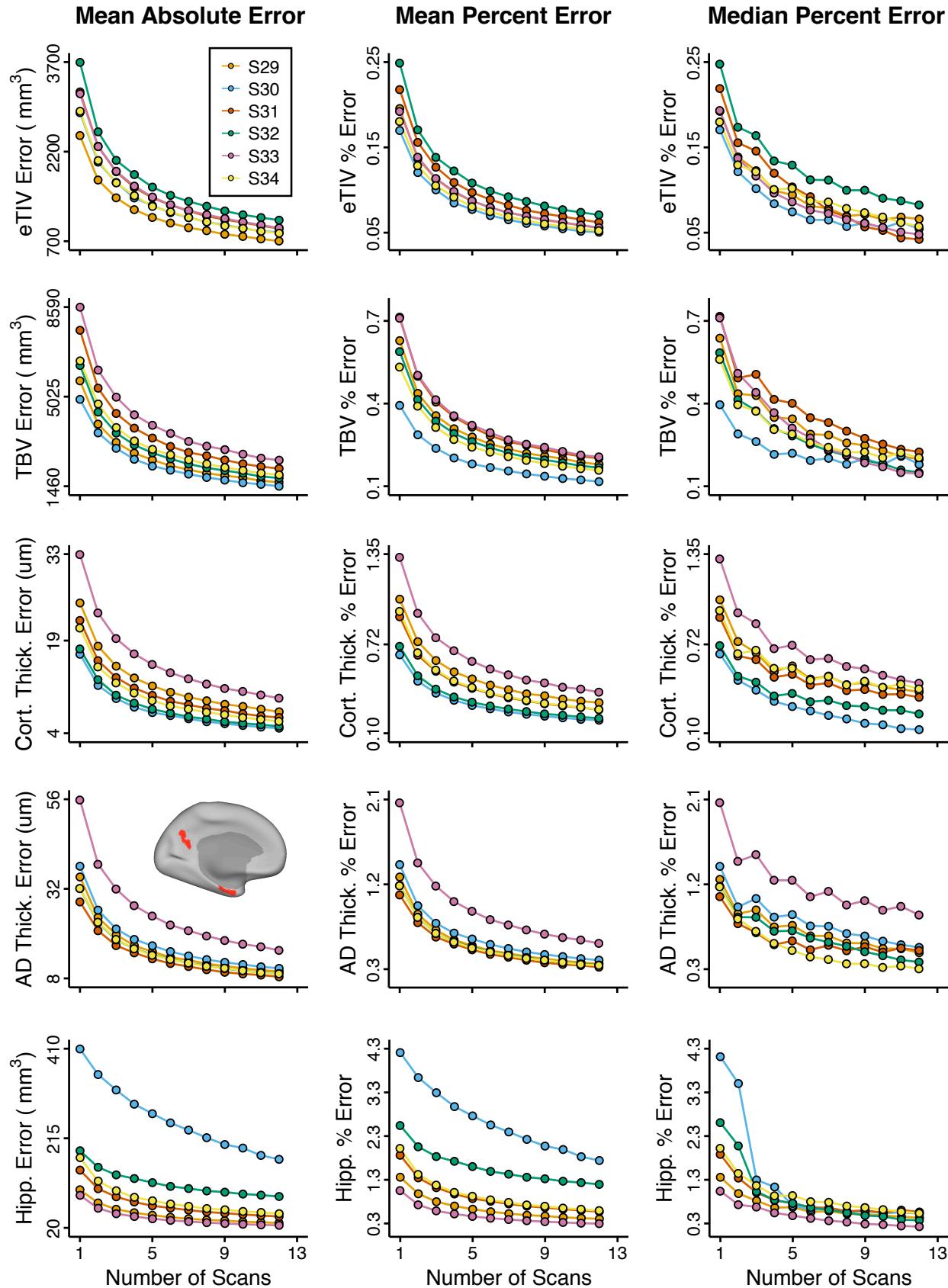


Figure 16