Test-retest reliability of cortical thickness measures: Effects of pulse sequence, geometry, and parallel acquisition

J. S. Wonderlick[1,2], D. A. Ziegler[1,2], A. Bakkour[3,4], A. van der Kouwe[3,5], C. Triantafyllou[2], S. Corkin[1,2,3], B. C. Dickerson[3,6]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[2]Athinoula A. Martinos Imaging Center at the McGovern Institute, Massachusetts Institute of Technology
[3]MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging
Departments of [4]Psychiatry, [5]Radiology, and [6]Neurology, Massachusetts General Hospital, Harvard Medical School

Introduction

BACKGROUND

- Development of sophisticated tools for the analysis of MRI data has led many researchers to adopt rapid, semi-automated methods for obtaining morphometric measurements, such as cortical thickness and volume.
- Technological advances in MRI hardware and acquisition techniques have resulted in a push toward obtaining higher resolution images and the use of multispectral sequences for improved contrast.
- Further, the use of parallel acquisition techniques, such as GRAPPA, have enabled substantially faster collection of structural neuroimaging data, while decreasing the adverse effects of motion artifact.

  This advance is especially beneficial to researchers studying subjects prone to motion, such as patients with movement disorders or children.
- While these advances are largely welcomed, it is important to understand how differences in image geometry, sequence parameters, and the use of parallel acquisition techniques influence subsequent processing steps and, ultimately, morphometric measurements.
- To address this question, we examined the reliability of cortical thickness, volume, and surface area measures as a function of different acquisition parameters and imaging techniques.

Methods

PARTICIPANTS

- 5 young adults (21.4 + 3.8 years; 1 male / 4 female)
- 5 older adults (66.4 ± 12.4 years; 3 male / 2 female)
- Two scanning sessions approximately two weeks apart

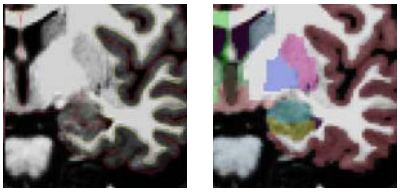
PARAMETERS AND SEQUENCES

- Scans collected on a Siemens 3T Trio scanner with 12-channel TIM head coil
- Sequences were permutations of T1-weighted Magnetization-Prepared RApid Gradient Echo (MP-RAGE)
- For the multiecho MP-RAGE, a final volume was generated from the RMS average of all echoes collected
- Parallel acquisition implemented by Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA)
- Two scans were collected per sequence during each session

PREPROCESSING

- Surface and morphometric measures generated with FreeSurfer and associated Montreal Neurological Institute (MNI) tools
- Multiple within-session acquisitions of each sequence were averaged after motion correction to generate a single volume for reconstruction
- Volumes underwent semi-automated processing for Talairach transformation, skull stripping, subcortical registration and segmentation, and white/pial surface generation[1-4]
- We derived measures of global cortical thickness and volumes of cortical grey matter (GM), white matter (WM), hippocampus, brainstem, putamen, thalamus, amygdala, and caudate

Sample anatomical data



Left: WM/GM (yellow) and GM/CSF (red) boundaries.
Right: automatic segmentation of cortical gray matter (red), white matter (white), caudate (blue-gray), putamen (pink), globus pallidus (light blue), hippocampus (brown), and amygdala (teal)

RELIABILITY MEASURES

- We calculated percentage difference scores and intra-class correlation coefficients (ICC)[5] to assess reliability of measures across sequences and scanning sessions (test-retest). ICCs were calculated as two-way mixed effects models of absolute agreement between measures.

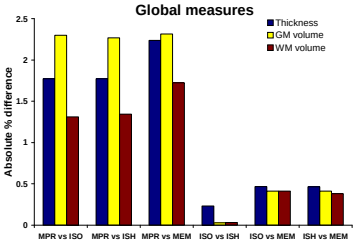
Sequences and scanning parameters

Sequence	TR (ms)	TI (ms)	TE (ms)	Flip angle	Bandwidth (Hz/pixel)	voxel size (mm)	iPAT
Anisotropic MP-RAGE (MPR)	2530	1100	3.39	7°	195	1.3 x 1.0 x 1.3	No
Isotropic MP-RAGE (ISO)	2530	1100	3.48	7°	195	1.0 x 1.0 x 1.0	No
Accelerated isotropic MP-RAGE (ISH)	2530	1100	3.48	7°	195	1.0 x 1.0 x 1.0	Yes (x2)
Multiecho MP-RAGE (MEM)	2530	1100	1.58 + (n x 1.74), n = 0,..,3	7°	698	1.0 x 1.0 x 1.0	Yes (x2)

Results

Inter-sequence

Global measures

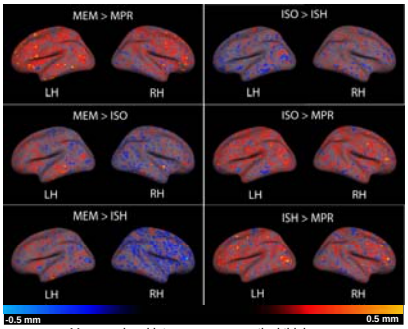


Intra-class correlation coefficients

	MPR v ISO	MPR v ISH	MPR v MEM	ISO v ISH	ISO v MEM	ISH v MEM
Thickness (left)	0.819	0.839	0.766	0.980	0.967	0.972
Thickness (right)	0.824	0.858	0.785	0.985	0.984	0.968
GM volume (left)	0.936	0.940	0.934	0.994	0.997	0.996
GM volume (right)	0.933	0.948	0.937	0.994	0.997	0.994
WM volume (left)	0.985	0.985	0.969	0.999	0.994	0.995
WM volume (right)	0.985	0.985	0.969	0.999	0.994	0.995

Percentage difference scores and ICCs showed lower reliability of thickness and volume measures for comparisons between anisotropic and isotropic sequences, relative to comparisons among the three isotropic sequences.

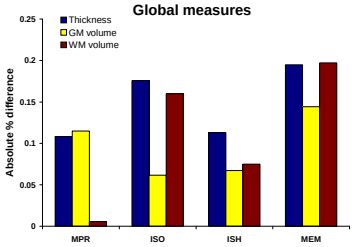
Regional thickness differences



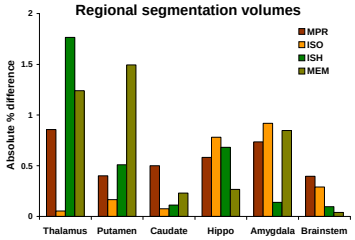
Mean regional inter-sequence cortical thickness differences mapped to averaged surfaces

Test-retest

Global measures



Regional segmentation volumes



Above: Differences in volume between scanning sessions varied regionally and by the type of sequence used.
Below: Examination of test-retest ICCs, however, revealed a high level of absolute agreement across sequences for most structures.

Intra-class correlation coefficients

	MPR	ISO	ISH	MEM
Left hemisphere				
Global thickness	0.970	0.993	0.966	0.982
Volumes				
Gray matter	0.991	0.997	0.988	0.997
White matter	0.999	0.999	0.999	0.997
Caudate	0.994	0.996	0.997	0.994
Brainstem	0.981	0.986	0.984	0.992
Putamen	0.969	0.984	0.976	0.800
Hippocampus	0.938	0.974	0.975	0.988
Thalamus	0.929	0.971	0.894	0.852
Amygdala	0.869	0.910	0.851	0.868
Right hemisphere				
Global thickness	0.971	0.989	0.975	0.980
Volumes				
Gray matter	0.995	0.996	0.989	0.996
White matter	1.000	0.999	0.999	0.997
Caudate	0.996	0.997	0.992	0.991
Hippocampus	0.981	0.971	0.967	0.981
Brainstem	0.981	0.986	0.984	0.992
Putamen	0.981	0.986	0.960	0.899
Thalamus	0.931	0.948	0.972	0.956
Amygdala	0.916	0.983	0.964	0.978

Summary & Conclusions

- Selection of voxel size had the largest impact on cortical thickness outcomes

  Voxel size had less impact upon WM and GM volume measures, which showed high reproducibility

  All GM volume ICCs > 0.988; all WM volume ICCs > 0.997
- GRAPPA acceleration had little effect on the reliability of thickness or volume measures
- Test-retest reliability of cortical thickness, GM volume, and WM volume measures was high across all sequences

  All percentage differences < 0.2%; all ICC values > 0.96
- Test-retest reliability was slightly less for structures defined by automatic segmentation

References

1. Dale AM, Fischl B, Sereno MI. (1999). *NeuroImage 9*: 179-194.
2. Fischl B, Sereno MI, Dale AM. (1999). *NeuroImage 9*: 195-207.
3. Fischl B. et al. (2001). *IEEE Trans Med Imaging 20*: 70-80.
4. Fischl B, et al. (2004). *Cereb Cortex 14*: 11-22.
5. Shrout PE & Fleiss JL. (1979). *Psych Bull 86*: 420-428.

Acknowledgments

We would like to thank Steven Shannon, Sheeba Arnold Anteraper, and Nicholas Harrington for technical assistance during data collection.
This research was supported by NIH grants AG021525 & T32 GM007484.
The Athinoula A. Martinos Imaging Center at the McGovern Institute is supported in part by a generous gift from Pat and Lore McGovern.
The Athinoula A. Martinos Center for Biomedical Imaging at MGH is supported by the National Center for Research Resources (P41RR14075), the Mental Illness and Neuroscience Discovery (MIND) Institute, and the National Alliance for Medical Image Computing (NAMIC) (NIBIB U54 EB005149).

Author contact information: **Julien Wonderlick**
jswonder@mit.edu
MIT 46-5127, 77 Mass. Ave
Cambridge MA 02139, USA