*11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, 2005*

# Test-retest reliability assessment for longitudinal MRI studies: A comparison of the effects of different $T_1$-weighted protocols, scanner platforms, and field strengths on semi-automated hippocampal volume measures

Brian Quinn[1,2], Elizabeth Fenstermacher[1,3], Xiao Han[1,2], Jenni Pacheco[1,2], Silvester Czanner[1,2], Andre van der Kouwe[1,2], Paul Maguire[4], David Raunig[4], Marilyn Albert[5], Nikos Makris[1,3,6], Rahul Desikan[1,7], Ron Killiany[1,7], Brad Dickerson[1,3], Bruce Fischl[1,2]

[1]MGH/MIT/HMS Athinoula Martinos Center for Biomedical Imaging, Charlestown, MA
Departments of [2]Radiology and [3]Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[4]Pfizer Global Research & Development, Groton, CT
[5]Division of Cognitive Neuroscience, Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD
[6]MGH Center for Morphometric Analysis, Charlestown, MA
[7]Department of Anatomy and Neurobiology, Boston University, Boston, MA, USA

## Introduction:

Longitudinal MRI studies offer the potential to quantify changes in brain structure over time in neurodegenerative diseases, such as Alzheimer's disease [1]. Such morphometric measures hold promise as biomarkers of disease progression for use in clinical trials of putative disease-modifying therapies, and include the volume of anatomical regions of interest (ROI; e.g., hippocampal formation (HF)). Better understanding of sources of variance in morphometric measures from longitudinal MRI studies—which may relate to raw MRI data (e.g., intensity), measurement protocols, or brain tissue due to disease or non-disease factors (e.g., subject hydration status)—should enable method development to minimize noise, thus increasing power to detect biologic effects of interest. Noise minimization is a difficult challenge in single-site longitudinal studies, and is compounded by multiple sites (i.e., multi-center clinical trials). To date, variance estimates for morphometric measures, such as brain or HF volume, have been derived from multiple measurements of the same scan or of two scans of the same subjects acquired on one day. This may underestimate variance in actual longitudinal studies in which MRI data are acquired at multiple timepoints separated by weeks to months. Therefore, we undertook this study to estimate variance in morphometric measures when healthy older subjects are scanned at two week intervals on the same scanner and on different scanners. This initial analysis focused on reproducibility of semi-automated derivation of HF volume. An accompanying abstract presents our investigation of image intensity reproducibility.

## Methods:

Fifteen healthy older subjects were each scanned at four different visits, ~two weeks apart, twice on 1.5T Siemens Sonata, once on 1.5T GE Signa, and once on 3T Siemens Trio. At each visit, two MP-RAGE volumes were collected. With the Siemens scanners, three multi-flip angle multi-echo FLASH scans were acquired [2]. Semi-automated computation of HF volume was employed for test-retest comparison, using four variations of a previous method [3]: (1) absolute HF volume calculation using MP-RAGE volumes, (2) #1 corrected using partial volume estimation for each ROI border voxel, (3) #2 corrected with gradient unwarping, and (4) #1-3 using multi-echo FLASH synthesized volumes.

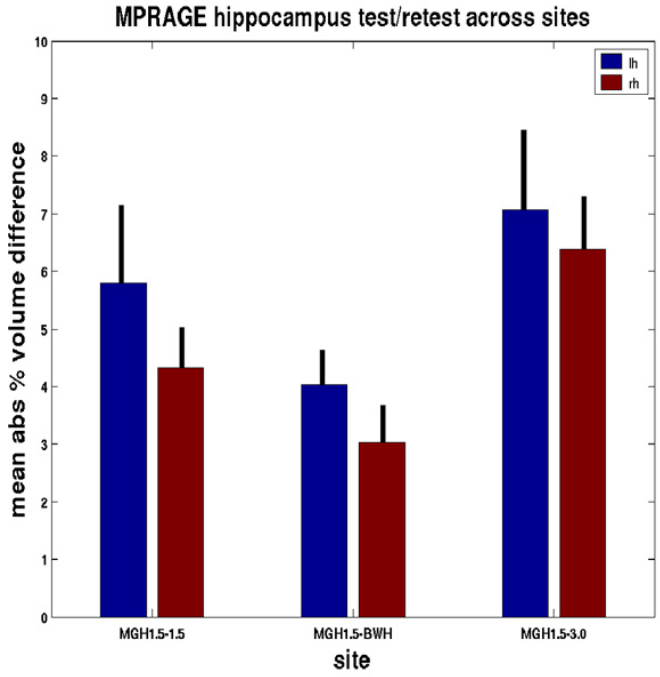
## Results and discussion:

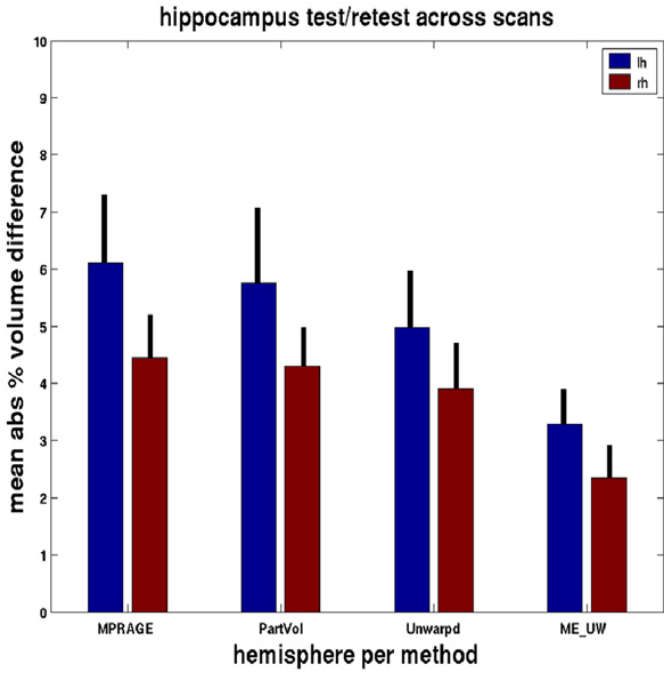
A Figure summarizes the results. Of the methods tested, unwarped, multi-echo FLASH volumes were least variable. In the intra-scanner comparison, the intra-class correlation coefficient for HF measures was >.90 and percent volumetric difference was 3.2% for left and 2.2% for right HF.

**Conclusions**:
These results suggest that, for longitudinal morphometric studies, the HF can be labeled efficiently and reliably using a semi-automated algorithm when subjects are scanned two weeks apart using the same scanner and using different scanners or field strengths. The methods outlined herein may contribute to additional improvements in reliability. Work is ongoing to analyze these data with respect to other subcortical and cortical regions of interest.

**References:**
[1] Fox NC & Schott JM. Lancet 2004;363:392-4
[2] Fischl B et al. NeuroImage 2004; 23 Suppl 1:S69-84
[3] Fischl B et al. Neuron 2002;33:341-55

**Acknowledgements**: This study was supported by Pfizer, Inc., the NIA (K23-AG22509 & P01-AG04953), and the NCRR BIRN Morphometry Project (U24-RR021382).


MPRAGE hippocampus test/retest across sites
mean abs % volume difference
lh
rh
MGH1.5-1.5
MGH1.5-BWH
MGH1.5-3.0
site


hippocampus test/retest across scans
mean abs % volume difference
lh
rh
MPRAGE
PartVol
Unwarpd
ME_UW
hemisphere per method