Reproducibility of automated brain morphometric estimates derived from multi-site structural MRI studies

J. Jovicich[1], S. Czanner[1], X. Han[1], D. Greve[1], J. Pacheco[1], E. Busa[1], A. van der Kouwe[1], B. Fischl[1,2], A. Dale[3]

[1]Radiology, 1MGH/MIT/HMS Athinoula Martinos Center for Biomedical Imaging, Charlestown, MA, United States, [2]Computer Science & Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, United States, [3]Departments of Neurosciences and Radiology, University of California San Diego, San Diego, CA, United States

INTRODUCTION

Neuroimaging longitudinal studies of change over time are becoming increasingly a standard element of clinical neuropsychiatric research [1,2]. These studies create the imperative to characterize and correct technological sources of variance that limit image reproducibility in high-resolution structural MRI studies, thus enabling quantitative, platform-independent, multi-site evaluation. In previous work we studied the reproducibility of image intensity with and without distortion correction from gradient non-linearity effects [3]. Here we extend that work to study the variability of morphometry results derived from the same data (cortical thickness and subcortical volume estimates) obtained using the Freesurfer toolkit [4], which is freely available to the research community.

MATERIALS AND METHODS

Data acquisition: Five sites with clinical 1.5T whole body scanners used in regular functional and structural MRI studies participated in this study: a) two GE Signa sites with Cardiac Resonator Module gradient coils; b) one GE Signa site with Brain Resonator Module gradient coils; c) one Siemens Sonata site; and d) one Picker Eclipse site. Five healthy volunteers were scanned twice at each of these sites using a 3D-spoiled gradient echo volume (TR=20 ms, TE=6 ms, 256x192, 1.3 mm thick 124 sagittal slabs, FOV 25 cm, flip angles 30^0, 20^0, 5^0 and 3^0, 8 min 12 sec acquisition per flip angle) and using the vendors' standard head RF coil. We call this protocol single-echo FLASH, or SEF. Each subject (1 female, 4 males, and average age 39) was scanned twice on each site, in different sessions. For within-site repetitions the average time interval for rest-retest scans was 19 days (minimum time 1 day, maximum time 8 months), and for across-site repetitions the average time interval was 8 months (minimum 2 months, maximum 15 months).

Morphometry estimations: Subcortical structure segmentation of over 20 basic structures were computed both manually and automatically using the atlas-driven method described in [5]. We have rebuilt a two-channel atlas for SEF data (one channel for each of the flip-angles 30^0 and 5^0) using manually labeled data from other 11 subjects scanned at the Siemens Sonata site. Cortical reconstruction was computed on the test/retest scans using the FreeSurfer tools [4]. A linear discriminant analysis (LDA) method [6] was used to synthesize one scalar image with optimal gray/white CNR from the original 4-dimensional multi-flip angle inputs, which is then processed by FreeSurfer. Both cortical and subcortical measures were evaluated using the original data and the 3D distortion corrected data [3].

Reproducibility evaluation: For the subcortical variability, we computed for each subject the mean and standard deviation of the subcortical volumes across the sites. For the thickness variability, we computed for each subject the mean across the whole brain of the absolute point-wise thickness difference measured across the various sites

RESULTS AND DISCUSSION

The mean cortical thickness cross-site variability, averaged across subjects (for each subject averaged across the brain), is 0.13 mm. Given that the mean cortical thickness across the whole brain is approximately 2 mm, this represents a reproducibility error of less than 7%. Power analysis shows that with this measurement error the sample sizes required to detect (significance 0.5, statistical power 0.9) thickness difference effects (deviation from a mean of 2mm) of 20%, 10% and 5% will be 8, 24 and 86 subjects, respectively. The subcortical volume cross-site variability analysis is underway, but preliminary results show that the variability given by the automated method is significantly better than that given by the manual segmentations. The reproducibility varies across structures due to differences in signal-to-noise ratio, but on average the variability is fewer than 10%. Gradient unwarping did not significantly improve cortical thickness or subcortical volume reproducibility in this subject population and with this acquisition protocol. The complete analysis will allow us to compare within-site with across-site reproducibility errors, from which sample size predictions can be made for the detection of various size effects.

REFERENCES:

[1] Fox NC & Schott JM Lancet. 2004 Jan 31;363:392-4

[2] Ashburner J et al., Lancet Neurol. 2003 Feb,2:79-88

[3] Jovicich et al., NeuroImage (2006 in press) (Unwarping code in: http://www.nbirn.net/Resources/Downloads/GradientNonLinearity/index.htm)

[4] https://surfer.nmr.mgh.harvard.edu/

[5] Fischl B. et al., NeuroImage 2004; 23 Supp 1, S69-84.

[6] Duda R et al. Pattern Classification, John Wiley & Sons, Inc., 2001

ACKNOWLEDGEMENTS: This research was supported by a grant (#U24 RR021382) to the Morphometry Biomedical Informatics Research Network (BIRN, http://www.nbirn.net), that is funded by the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH).