A Prospective Multi-Institutional Study of the Reproducibility of fMRI: A Preliminary Report from the Biomedical Informatics Research Network

Kelly H. Zou, PhD [1,3], Douglas N. Greve, PhD [2,3], Meng Wang, MSE [1,3], Steven D. Pieper, PhD [1,3], Simon K. Warfield, PhD [1,3], Nathan S. White, BS [2,3], Mark G. Vangel, PhD [2,3], Ron Kikinis, MD [1,3], William M. Wells, III, PhD [1,3], and FIRST BIRN [3]

[1] Surgical Planning Laboratory,
Department of Radiology, Brigham and Women's Hospital,
[2] Athinoula A. Martinos Center for Biomedical Imaging,
Department of Radiology, Massachusetts General Hospital,
Harvard Medical School
[3] Functional Imaging Research of Schizophrenia Testbed (FIRST),
Biomedical Informatics Research Network (BIRN)

Draft, 1-13-03

Abstract

Purpose

To investigate the effects of factors such as study site, field strength, visit and run on the reproducibility of the performance of a sensory-motor (SM) task by 5 healthy subjects in a prospective multi-institutional study.

Background

Functional MRI (fMRI) has significantly contributed to studies of the human brain. [1,2] Unfortunately, there exists high variability in the magnitude, spatial distribution, and statistical significance of fMRI maps. The Biomedical Informatics Research Network (BIRN) compared the fMRI signal to determine whether combining fMRI maps from different sites is meaningful. We used an estimation-maximization algorithm, Simultaneous Truth and Performance Level Estimation (STAPLE), [3,4] to characterize the variability in a sensory-motor (SM) task.

Methods

Eleven sites participated in the Functional Imaging Research of Schizophrenia Testbed BIRN study. Data was analyzed from 8 (5 1.5T and 3 3.0T scanners) of these sites. Five healthy right-handed males were scanned at each site in 2 visits. The SM task was performed for 4 of 10 runs per visit. A block design was used with 15-second epochs of alternating baseline (fixation) and task for 85 (plus 2 discarded) acquisitions per run. Subjects performed bilateral finger tapping on a dummy and an actual button box with a 3Hz audio cue and a reversing checkerboard. The subjects pressed buttons 1 through 4 in consecutive order and then back again using both hands simultaneously and in sync. Imaging was acquired with block-design echo-planar or spiral gradient echo (oblique axial, 64x64 matrix, 1 shot, 35 slices, 4mm, 3 sec, TE varied with strength). A bite bar

helped minimize head movement. Motion correction for each run was applied to middle time point using AFNI. Smoothing was based on FWHM 5mm. Registration was performed over runs. STAPLE optimally estimated a composite true 3D activation map over the runs per visit, visualized in Slicer. [5] Sensitivity and specificity were derived from voxel counts. [6] Significance of the factors was analyzed via summary statistics and linear models.

Results

Significant factors for sensitivity included field strength (p=0.002), subject (p=0.003), site (p=0.02) and manufacturer (p=0.02), and for specificity included subject (p=0.04) and run (p=0.04). At 3.0T, the mean sensitivity per subject ranged 0.57-0.76 (SD =0.17-0.26) while the mean specificity ranged 0.99-1.00 (SD=0.002-0.02). At 1.5T, the mean sensitivity only ranged 0.42 to 0.69 (SD=0.18-0.34) while the mean specificity ranged 0.95-1.00 (SD=0.004-0.22).

Conclusions

Subject: Moderate to high reproducibility within subject was found. Site: The variability across sites appeared less than that across subjects. 3.0T generally yielded greater activation and higher sensitivity than 1.5T. Run: Variable patterns over runs were observed, with less activation but more robust patterns during latter visits. Thus, a calibration plan may be possible to minimize the variability introduced by different sites.

References

1. Brannen JH et al. AJNR 2001; 22: 1711-8.
2. Machielsen WCM et al. HBM 2000; 9: 156-164.
3. Warfield SK et al. LNCS 2002; 2488: 290-297.
4. Zou KH et al. Acad Radiol 2003; 10: 1359-1368.
5. Gering D et al. JMRI 2001; 13: 967-975.
6. Genovese CR et al. Neuroimage 2002; 15: 870-878.

Figure 1. Mean activation percentage by subject and field strength.

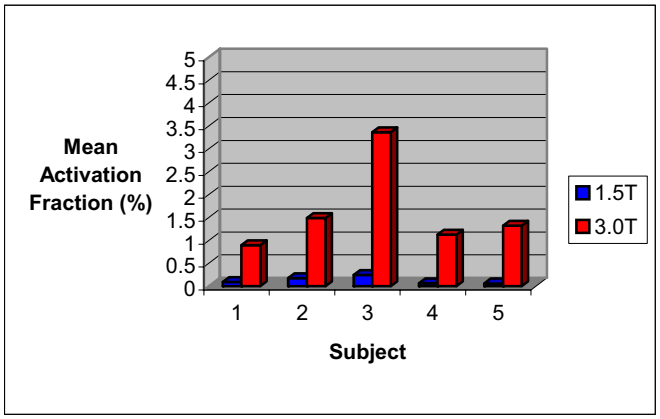

Figure 2. Mean sensitivity (true activation fraction) by subject and field strength.

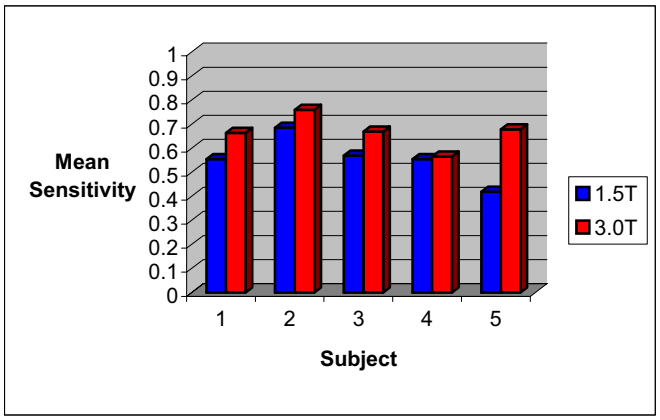

Figure 3. Mean specificity (true non-activation) by subject and field strength.