Towards a Framework for Comparing Functional Magnetic Resonance Imaging Data Across Vendors and Models

National Institute of Mental Health

Peter J. Molfese[1], John Rodgers-Lee[2], Sean Marrett[3], Adam Thomas[2], Vinai Roopchansingh[3], Dylan Nielson[2], Jan Varada[3], J. Andrew Derbyshire[3], Peter A. Bandettini[13]

[1]Section On Functional Imaging Methods, [2]Data Science and Sharing Team, [3]FMRI Facility

Introduction

Neuroimaging studies are increasingly targeting diverse populations represented by different genetic, social, phenotypic, and geographical profiles (e.g. Rueckl et al., 2015), which are made more feasible through data collection at multiple sites using a mixture of equipment. These studies often involve both cognitive tasks (Gee et al., 2015; Brown et al., 2010, Yendiki et al., 2010) and/or resting state (Turner et al., 2013). Some recommendations for combining data from multi-site studies exist (FBIRN; Zou et al., 2005), and some portion of these methods have been implemented in the HCP (Van Essen et al., 2013). Here we compare new approaches for normalizing and combining data from different MRI platforms, which may help facilitate more large-scale multi-site studies.

Methods

In this study, we collected fMRI data consisting of a rotating checkerboard stimulus on eleven participants, each imaged using four different 3T fMRI machines located at the National Institutes of Health (Bethesda, MD) with different scanner platforms (GE Signs HDx, GE MR 750, Siemens Skyra, Siemens Prisma), and manufacturer multi-channel whole-head phased array coils. All data were collected using matched scanner parameters (TR=2.0s, TE 0.03, 33 axial slices at 3mm isotropic resolution).

Scanner	GE Signa HDx	GE MR-750	Siemens Prisma	Siemens Skyra
Head Coil	8-channel	32-channel	32-channel	32-channel
Software Version	15M4	DV22.0_V02	VD13D	VD13A

Data from each MRI scanner were analyzed in AFNI using a standard pipeline

This Standard pipeline was compared to successive processing pipelines:

A. The standard pipeline
B. Smoothing all images to a fixed FWHM + SFNR as a covariate
C. Additional distortion correction using blip-up/blip-down EPI
D. Regression of physiological signals (RetroTS)
E. Removal of signal from White Matter tissues (ANATICOR)

In each pipeline, group level analysis was conducted using a either a more traditional ANOVA (1), or a Linear Mixed Effects model (2-5; 3dLME; Chen et al., 2013) with voxel-wise regressors of SFNR.

References

Brown, G. G., Mathalon, D. H., Stern, H., Ford, J., Mueller, B., Greve, D. N., … Function Biomedical Informatics Research Network. (2011). Multisite Reliability of Cognitive BOLD Data. NeuroImage, 54(3), 2163–2175. http://doi.org/10.1016/j.neuroimage.2010.09.076

Chen, G., Saad, Z.S., Britton, J.C., Pine, D.S., Cox, R.W. (2013). Linear Mixed-Effects Modeling Approach to FMRI Group Analysis. NeuroImage 73:176-190. http://dx.doi.org/10.1016/j.neuroimage.2013.01.047

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI clustering in AFNI: False positive rates redux. Brain Connectivity, 7, 152-171. doi: 10.1089/brain.2016.0475

Gee, D. G., McEwen, S. C., Forsyth, J. K., Haut, K. M., Bearden, C. E., Addington, J., … Cannon, T. D. (2015). Reliability of an fMRI Paradigm for Emotional Processing in a Multisite Longitudinal Study. Human Brain Mapping, 36(7), 2558–2579. http://doi.org/10.1002/hbm.22791

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W., Bick, A., Frost, S. J., Hancock, R., Wu, D. H., Mencl, W. E., Dunabeitia, J. A., Lee, J., Oliver, M., Zevin, J. D., Hoeft, F., Carreiras, M., Tzeng, O. J. L., Pugh, K. R., & Frost, R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. PNAS, 112 (50), 15510-15515. http://doi.org/10.1073/pnas.1509321112

Turner, J. A., Damaraju, E., van Erp, T. G. M., Mathalon, D. H., Ford, J. M., Voyvodic, J., Mueller, B. A., Belger, A., Bustillo, J., Mcewen, S., Potkin, S. G., FBIRN, & Calhoun, V. D. (2013). A multi-site resting state fmri study on the amplitude of low frequency fluctuations in schizophrenia.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, for the WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: An overview. NeuroImage 80(2013):62-79.

Yendiki, A., Greve, D. N., Wallace, S., Vangel, M., Bockholt, J., Mueller, B. A., … Gollub, R. L. (2010). Multi-site characterization of an fMRI working memory paradigm: Reliability of activation indices. NeuroImage, 53(1), 10.1016/j.neuroimage.2010.02.084. http://doi.org/10.1016/j.neuroimage.2010.02.084

Zou, K. H., Greve, D. N., Wang, M., Pieper, S. D., Warfield, S. K., White, N. S., … For the FIRST BIRN Research Group. (2005). Reproducibility of Functional MR Imaging: Preliminary Results of Prospective Multi-institutional Study Performed by Biomedical Informatics Research Network. Radiology, 237(3), 781–789. http://doi.org/10.1148/radiol.2373041630

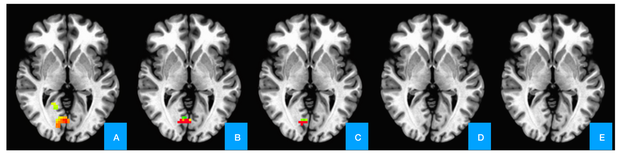

Figure 1 (above): Scanner differences in visual cortex to checkerboard stimulus in the 5 pipelines. Activation differences between scanners were non-significant at p=.001 in D and E.

Figure 2 (below): Activation of the visual cortex ROI, depicted in % signal change from baseline, for each MRI platform. While scanner differences were identified in early pipelines (A-C), those differences were not significant to the p<.001 (uncorrected) after accounting for physiological signals.

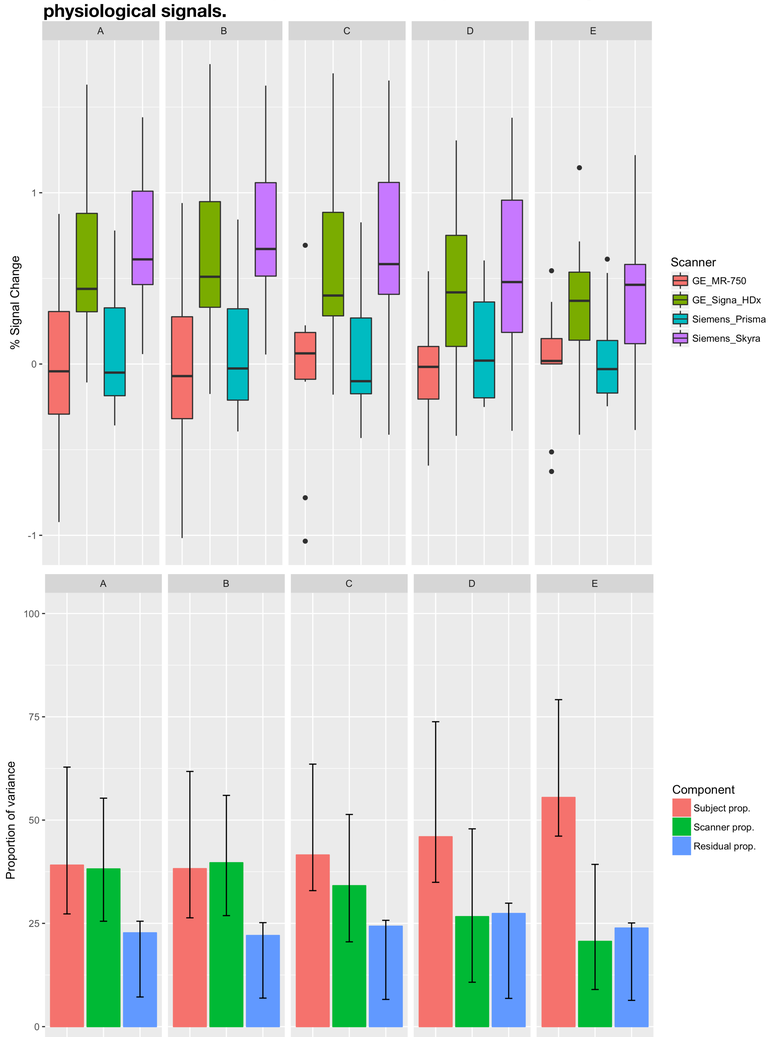

Figure 3 (above): Proportion of variance for Subject, Scanner, and Residual calculated for each processing pipeline (calculated for three MRI platforms). The confidence intervals were constructed using bootstrapping. 1000 bootstrapped samples for each version of the analysis were created by drawing from the original dataset randomly across all subject/scanner combinations. The LME model was fit on each bootstrapped sample to yield 1000 estimates of the proportions of variance. The error bars represent the 95% confidence interval for the estimates of the proportions of variance.

Results

A whole-brain repeated-measures ANOVA on the default processing pipeline identified a single cluster in visual cortex (Right Calcarine Gyrus) differentiating the MRI platforms at a threshold of p=.001 (uncorrected) made up of 85 voxels. This cluster size was smaller than what would survive cluster correction (Cox et al., 2017), and did not increase in size across subsequent analyses (B-E; Figure 1).

We then performed an ROI analysis, showing that with subsequent levels of analyses, the p-value of this activation decreased, corresponding to cluster size decreases. All analyses are corrected for sphericity using Greenhouse-Geisser, calculated using the Afex package in R.

Pipeline	F	P-value	Cluster Size
A	F(2.35, 23.46)=19.49	p<0.0001	85
B	F(2.23, 22.31)=22.31	p<0.0001	46
C	F(2.79, 25.09)=17.28	p<0.0001	25
D	F(2.44, 14.67)=7.56	p=.004	0
E	F(2.64, 15.85)=7.35	p=.003	0

Using the repeated-measures design, we fit a linear mixed effects model in which subject and scanner were specified as random effects. This is equivalent to the intraclass correlation (ICC) and yields the proportion of variance that can be attributed to each of the random effects, including the residual variance not captured by the model (depicted for each version of the analysis in Figure 3). Changes in proportion of variance did not reach statistical significance.

Conclusions

- We analyzed fMRI activation of a rotating checkerboard in the same 11 participants, each collected on four different scanner platforms
- Data were processed in parallel through separate pipelines, making use of “best practices” in fMRI analyses
- The inclusion of fixed-level smoothing, blip distortion correction, physiological regressors, and local white matter signal regression led to decreases in differences between MRI platforms
- Though non-significant, changes in proportion of variance accounted for by subject, scanner, and residual were observed
- Future studies
  - Should compare vendor reconstruction software to open source alternatives (i.e. Gadgetron).
  - Collect multiple sessions for each participant within each MR environment to better estimate variances of Subject and Scanner