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

Peter J. Molfese¹, John Rodgers-Lee², Sean Marrett³, Adam Thomas³, Vinal Roopchansingh⁴, Dylan Nelson⁵, Jan Varada⁶, J. Andrew Derbyshire⁷, Peter A. Bandettini⁸

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

Introduction

Neuroimaging studies are increasingly targeting diverse populations represented by different genetic, social, phenotypic, and geographical profiles (e.g. Rueckl et al., 2018), 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 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 Prisma, 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).

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 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.

Introduction

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 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.

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<0.001 (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 Afx package in R.

Using the repeated-measures design, we fit a linear mixed models effect model in which subject and scanner were specified as random effects. This is equivalent to the intracluster 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

References

Table 1: Comparison of processing pipelines and their performance metrics.

<table>
<thead>
<tr>
<th>Pipeline</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Signs Hdx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE MR 750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siemens Prisma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siemens Skyra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head Coil</td>
<td>8-channel</td>
<td>32-channel</td>
<td>32-channel</td>
<td>32-channel</td>
<td></td>
</tr>
<tr>
<td>Software Version</td>
<td>15M4</td>
<td>DV22.0_V02</td>
<td>VD13D</td>
<td>VD13A</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: (above): Scanner differences in visual cortex to checkerboard stimulus in the 2 pipelines. Activation differences between scanners were non-significant at p<0.001 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<0.001 (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 bootstrap 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 each bootstrapped sample to yield 1000 estimates of the proportions of variance.

Figure 4: (below): Comparison of processing pipelines and their performance metrics.