Supplemental Material: FBIRN progress since June 2005

The summer of 2005 has been a rich and productive time-window for FBIRN researchers, who were able to fully exploit the collaborative environment and physical infrastructure created over the past three years. Multi-site, multi-disciplinary scientific developments depend on an ethos of mutually shared language, respect, and goals. Establishing this collaborative culture for a national consortium involves overcoming the competitive biases that successful investigators from diverse fields bring to the table, and solving the significant problems that are not present in smaller collaborations. The FBIRN investigators have made substantial scientific, professional, and financial commitments of time and effort to develop this culture. The result is an efficient collaboratory that has made significant additional progress since mid-June when the proposal was submitted. Through simultaneous development, testing and implementation across multiple and diverse sites, FBIRN offers an efficient and rapid development cycle for robust tools designed for use in the broader scientific community.

When the proposal was submitted, the FBIRN had collected approximately 40 subjects across four sites, in Step 4 of Series A. Those datasets formed the basis for much of the preliminary data sections. In the past two months, that dataset has tripled to over 140 subjects across 10 sites. This large, multi-site, standard dataset is critical in development of the FBIRN tools. This summer's progress ensures that we will meet our completion recruitment target in November. This report summarizes the progress made through the integrated efforts of FBIRN clinical, cognitive, statistics, calibration and information technology researchers over the past few months, namely:

1) Improvements in the multi-site data storage, access, and analysis.

2) Continued assessment of intersite variance using refined methods and expanded tools.

3) Reliability measures on the new multi-site dataset in both cognitive tasks.

4) Continued improvements in calibration methods and the application of calibration and statistical methods to the multi-site datasets.

The initial Series A, traveling subjects data also are being uploaded to the SRB, an essential step prior to being made publicly available to the research community. A public roll-out of tools and data is scheduled for the Society for Neuroscience Meeting, November 12-16 in Washington, DC.

1. Improvements in multi-site data storage, access, and analysis

Database Schema Deployment and Use: The release version (Human Imaging Database, HID 1.0) was deployed in mid-June and is being integrated into the standard BIRN software suite through the ROCKS distribution technology. Features and improvements in this release consist of support for double entry, backing tables for complete audit histories, standard interfaces for the distributed query mediators (BIRN-CC), and regression testing datasets. The HID is currently available for both PostgreSQL and Oracle databases. The first release of documentation (user manual, testing, etc) for the HID schema, the Graphical User Interface (GUI) interface, the clinical assessment tools CALM and GAME is available. **The HID is currently installed and deployed at all FBIRN sites and ready for data input.** SQL scripts containing protocol information for Series A, Step 3 (Traveling Subjects) and Step 4 (Multi-site Data Collection) studies have been issued to all sites. **All sites are on track to complete entry of Series A, Step 3 data in September 2005, as well as continually adding Step 4 clinical, behavioral, and imaging data to the databases.**

Graphical User Interface Improvements: The support for double data entry of assessment forms in the GUI Interface, including conflict resolution when the two entries disagree, has been completed. The current GUI interface consists of 17 online assessment entry forms that mimic the paper forms to improve data entry ease and minimizes entry errors. Each of these online assessment forms has been validated for consistency between the web-based form and the database entries. The development of these forms is the result of a close collaboration between the IT and Clinical researchers at all sites.

Additions to the query forms of the GUI include mediated query options which will send a query to all other available databases configured during GUI installation. With the addition of the standard interface for distributed queries added to the database (see above database section), **all of the deployed sites can be queried by configuring the GUI with site information during installation.** These additions serve to provide both simple distributed query functionality while the Mediator being developed by BIRN-CC is being tested, and an interface for real-time database status logging.

Upload Scripts Finalized: Improved human data upload scripts have been developed and are being tested. These scripts both upload files to the SRB in the standard directory structure, and automatically integrate the Uniform Resource Indentifier (URI) links to all files uploaded to the SRB into the database HID. URI links, which are not dependent on paths to image files but use unique identifiers in SRB space, are unaffected by moving of files between directories or physical storage locations. Additionally, the scripts create wrapper XML files for both format agnostic raw image access and additional paradigm information required for FIPS automated image processing. **These scripts are an important integration between the HID, SRB, XML, and FIPS technologies.**

FSL Image Processing Scripts (FIPS) improvements: Automated methods are critical to the analysis and handling of large datasets. There have been three areas of progress in FIPS development since the proposal submission. (1) First-level database integration. We have implemented a set of routines to perform automated first-level fMRI analysis. The routines run locally, extracting data from and storing data to the FBIRN hierarchy. The routines automatically find and analyze all the data sets that meet a certain criteria (eg, site or task paradigm). The output is stored in a way that allows it to be easily uploaded to the SRB. (2) Morphometry BIRN Integration. Routines have also been added to integrate the FIPS results with anatomical analyses as generated by MBIRN tools. These include functional/structural volumetric registration, sampling of fMRI results onto FreeSurfer-generated surfaces, and ROI analysis using regions automatically generated by Freesurfer (as used in the Statistical WG analyses). (3) fMRI Analysis Framework. As part of the first-level analysis, the user must define a First-Level Analysis Configuration, or FLAC. The FLAC defines certain analysis parameters (eg, spatial smoothing, model specification, etc). These are defined once and then used to analyze all data from all subjects and sites. This assures that the analyses are consistent.

Improvements to XML Tools: The epoch averaging analysis tools (as used in the Preliminary Path Analysis of Auditory Oddball Data section below) now support **frequency filtering** using Chebyshev filters. Contacts at Siemens have enabled FBIRN to account for discrepancies in **slice order** at the various Siemens sites, and to provide the correct information in the XML tools. We now specify the **phase-encode/frequency directions** (rows vs. columns) in the XML files. The XML tools now support the most recent version of the GE P-File structure (ver11) and can properly describe DTI data from scanners running GE Excite Software 12.0. We have provided **event extraction tools** for the output of Presentation software and any other software providing similar text-based tabular output, and improved the semi-automated conversion of proprietary E-Prime data files to text format (via "remote-control" of the program E-DataAid) for later use by event analysis tools.

In the analysis of the Auditory Oddball datasets below and briefly in other areas, more detail about these applications is given.

2. Assessment of Intersite Differences in Imaging Signal

One of the FBIRN's stated aims is to assess and correct for the intersite imaging variability. The sources of variability can be found in the equipment, the methods, and the paradigms used. The ability to correct for intersite differences is driven by measurement of those sources. The Series A, Step 3 traveling subject dataset is critical to this measurement; it has been uploaded to the SRB using the original version of the upload scripts. These data were downloaded by multiple sites for local analysis and exploration, which forms the basis for these results. The following summarizes our continued progress in these assessments.

Variance Components Analysis and Generalizability of Sensorimotor BOLD Data: A Demonstration of the FBIRN Image Processing Stream (FIPS). We presented initial analyses of the intersite variability in subject-specific BOLD signal in the various preliminary data sections of the proposal. Here we present an updated analysis of variance components and reliability coefficients for sensorimotor imaging data, obtained from the Series A, Step 3 traveling-subjects study. We used FIPS to calculate the mean amplitude of the hemodynamic response in three regions of interest (ROIs). Surfaced-based methods, developed by the Morphometry BIRN and incorporated into FIPS, were used to identify subject-specific cortical ROIs. A completely crossed design was analyzed treating subject, site, day, and run as random effects and hemisphere as a fixed effect. Estimates were obtained using software that relies on the method of moments, which allows the estimates of zero variance across days for most cases. Alternative estimation strategies are being explored.

The table below provides the percent of total variance explained by a given factor for the mean of an entire ROI, and for the 10% most active voxels within each ROI. The Subject X Site+ effect encompassed all higher-way effects, other than the residual, that included the Subject X Site interaction (e.g., Subject x Site x Day, Subject x Site x Run). In all ROIs the site effect was larger than the subject effect, which was the object of measurement. Subject x Site interactions accounted for some of the variance, implying that differences among subjects varied by site.

Table 1. % of Variance Explained – Sensorimotor Task

	ROI – Entire Region			ROI – Top 10% of Activated Voxels		
Effect	Auditory	Hand	Visual	Auditory	Hand	Visual
Subject	16.00	16.47	13.00	18.78	18.31	21.75
Site	46.75	28.52	38.07	43.03	21.03	43.83
Day	0.00	0.00	0.00	0.00	0.00	0.06
Run	4.49	0.29	0.00	0.44	0.12	0.10
Subject X Site	6.27	5.49	9.89	3.60	14.57	10.49
Subject X Site+	19.67	37.59	34.67	20.68	35.20	20.02
Residual	1.48	3.60	2.55	1.50	4.22	1.53

Generalizability coefficients, which measure consistency of the BOLD response across subjects, were excellent for the three cortical ROIs [Entire ROI: auditory = .93, visual = .88, and hand area = .87; Top 10%: auditory = .95, visual = .92, hand area = .89]. Dependability coefficients, which measure of the exchangeability of observations across levels of study facets, were smaller for the three ROIs [Entire ROI: auditory = .73, visual = .70, hand area = .75; Top 10%: auditory = 78; visual = .77; hand area = .81]. Site differences and interactions involving subjects and site accounted for much of the unwanted variance in the dependability coefficient. As indicated by the generalizability coefficients, BOLD data can be used to consistently measure between subject variance with a high degree of consistency. Yet **site differences in BOLD response reduce the exchangeability of BOLD data, biasing analyses where important subject variables, such as diagnostic group, interact with site.** These new analyses confirm the groundbreaking results presented in the proposal: **Intersite variability can exceed intersubject variability, thereby greatly limiting the value of multi-site imaging studies.** The proposal discusses methods to increase the generalizability and dependability of BOLD measures in multisite studies. The sections that follow show how these methods can improve the power to detect group differences.

Spike noise in FBIRN time series varies with vendor and field strength. Intersite differences in scanner spike noise stability (global mean image intensity outliers) were investigated using the Series A, Step 3 sensorimotor (SM) data from 5 traveling subjects (4). Spikes were detected on unprocessed images with the AFNI program 3dToutcount. The time series was detrended with a 2nd order polynomial using least-absolute-deviation (L1-norm) regression. The median absolute deviation of the time

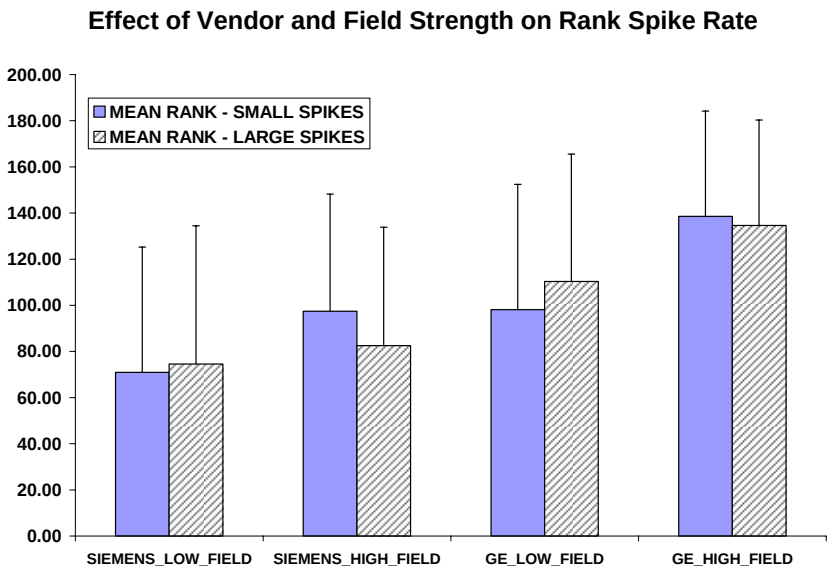


Fig. 1. The mean rank spike rate across the different vendors and field strengths, by thresholded spike size. There are two sites for each field strength of the Siemens scanners; two for the low-field GE; and three sites with high-field GE scanners.

series minus the trend was computed. Points that are far away from the trend were deemed "outliers". The degree of deviance of an outlier is defined by its "q" value. Q is analogous to p-value in least square regression and is an index of the probability of an event occurring by chance. The analysis was performed at two thresholds, q = 0.001 ("Small spikes") and q = 1E-14 ("Large spikes"). **Both field and vendor effects are demonstrated in the high threshold image spikes**. The vendor effect was by far the largest. Field: $F(1,176) = 4.4$, $p = 0.037$; Vendor: $F(1,176) = 28.3$, $p < 0.00001$. Figure 1 shows results demonstrating both site and vendor effects. These events can substantially affect postprocessing steps of motion correction, detrending, smoothing and statistical evaluation of time series data.

Because of these findings, the FBIRN FIPS processing stream now includes a spike detection step. Automated processing is required, as visually inspecting literally millions of images from a multi-site study is untenable. One of the sites was found to have faulty hardware during this scanning period and was subsequently repaired.

<u>Temporal Autocorrelation in FBIRN time series increases with field strength and by vendor.</u>

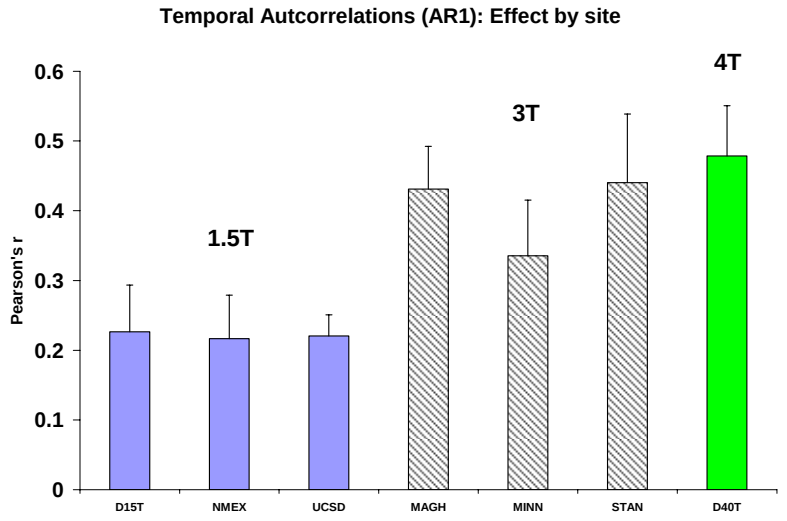


Fig. 2. Autocorrelation for seven sites at time lag 1. Significant site and field differences were observed: Site effects $F(6,44) = 25.4$; $p < 0.0001$. Field effects: $F(1,49) = 89.2$; $p < 0.0001$.

There is a growing awareness of the need to account for temporally autocorrelated residuals in fMRI time series analysis. The necessity for this depends on the strength of the autocorrelations. We have shown that **the intensity of temporal autocorrelations increases by vendor and in particular, by scanner field-strength across FBIRN sites** (5). The Series A, Step 3 sensorimotor (SM) data from 5 traveling subjects was used: A hemodynamic response function (HRF) was fit individually for each subject at each visit using deconvolution. The time series were regressed onto the reference vectors made by concatenating the fitted HRFs. The correlation structure

among residuals was modeled as an autoregressive process of degree 1 to degree 3. The autocorrelation was estimated using the Yule-Walker equations, at voxels in each subject's motor cortex activation. Measures were analyzed using a Mixed Model ANOVA, with either Field-strength (FIELD) or SITE as fixed effects and subject as a random effect. Measurements were transformed to linear scale before statistical analyses, and transformed back into correlations. Results in Fig. 2 show significant site and field strength effects. These confirmed vendor and field strength differences will be important to include in multi-site image data analysis.

Inter-site differences in HRF characteristics remain an issue for multi-site analyses. **Site-dependent HRF variations have implications for the sensitivity and intersite variability of all proposed fMRI analyses,** as discussed in the proposal**.** The HRF analyses presented in the Calibration WG preliminary data (see proposal section C.7) have now been augmented to include slice-timing correction for precision. The results remained nearly the same, with significant inter-site variation in the measured shape of the same individual's HRF: FWHM. F(site) = 5.9, df=6, 43.2; p = 0.0002; Rise time. F(site) = 3.9, df=6, 46; p = 0.003; Fall time. F(site) = 7.0, df=6, 43.1; p = 0.0001 (6).

3. Validation of the cognitive protocol at multiple sites

Since the submission of the revised proposal in June 2005, FBIRN has focused upon analyzing intermediate results from the Series A, Step 4 Auditory Oddball and Sternberg Item Recognition Paradigm (SIRP), collected at multiple sites on both patients and healthy controls. While over 140 subjects have been recruited to date, these analyses by necessity consider subsets of the dataset. The validation of the protocol for test-retest reliability and the expected group effects are required steps for multi-site imaging methodologies. (Only the Auditory Oddball data are presented here for space limitations.)

Series A, Step 4: Multi-site Data from the Auditory Oddball Task: Analysis Methods

The preliminary analyses reported here were conducted on the first 77 subjects for whom two visits were available. This group included 38 control subjects and 39 schizophrenics. The data were collected on 1.5 T, 3.0 T, and 4.0 T scanners and the scanners were manufactured by either Picker, Siemens, or General Electric. The data organization permitted drilling down along any combination of diagnosis, visit, site, field strength, vendor, region of interest (ROI), accuracy of response, and other variables. However, we have restricted our presentation here to only a few of these variables to simplify this report.

For these analyses, BIRN XML headers (BXH) were created for each series (corresponding to an experimental run) and for each imaging session. These BXH headers describe the data and header formats from each vendor so that the data could be read into processing programs without reformatting. The data from all contributing sites were first assessed by the FBIRN's Quality Assurance (QA) program and the resulting HTML log files were examined for each subject and visit. The QA program produced a XML file for run that provided information about the quality of each image volume in that run. The log files produced by Eprime for each experimental run were then read by a program that extracted timing information for stimulus and response events. The image data were then run through a pipeline that performed time slice adjustment, motion correct, coregistration of functional and structural data, normalization with a template brain, and smoothing. The smoothed normalized brains were then used for the remainder of the analyses reported here.

The experimental timing and quality assurance XML files were then used to extract epochs of data from each subject's time series of image volumes that were time-locked to the onset of stimulus events. Epochs that included an image volume with a mean intensity more than 3SD from the run's mean were likely marred by artifact and thus excluded. Averaged time epochs were then created for each stimulus condition (auditory target tone or auditory standard tone) for each visit of each subject. A statistical map was then created for each subject and visit representing the correlation of a template waveform with the averaged epochs for each voxel. These maps were then combined into group averages based upon other selection criteria (e.g., diagnosis, field strength, etc).

A grand mean map was created summarizing all 154 sessions. This map was then used to define functional regions of interest for several areas including the superior temporal gyrus (STG), middle frontal gyrus (MFG), anterior cingulate (ACG), and thalamus. These functional ROIs were then used as masks through which all 154 sessions were measured for average signal change and percentage of active voxels. These values were written to a database with associated metadata so that cross-tabulations could be computed.

Series A, Step 4: Multi-site Data from the Auditory Oddball Task: Results

Results from functional ROIs. The AO task was adopted primarily to activate temporal lobe regions, as these areas have been implicated in schizophrenia. We thus focus here on the STG ROI. The figure at right shows the mean signal change in this ROI for Controls and Patients for Visit 1 and Visit 2. Comparing visits, the ICC for the percent of active voxels for Controls was .55 (95% confidence interval: .29-.74) and for Patients was .62 (CI: .38-.78). **The difference in STG activation between Controls and Patients was significant (main effect, F=9.2, p=.003), confirming the strong effect of group that the paradigm was designed to reveal.** (See Figure 3.)

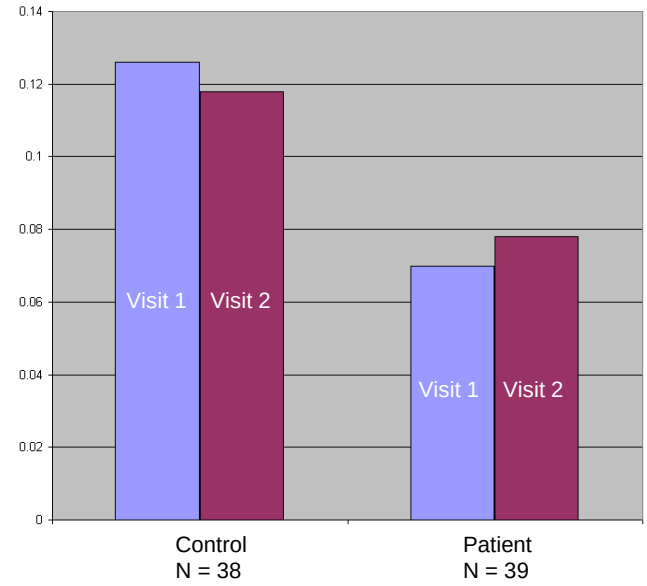


Fig. 3 Average STG signal change for auditory deviants, by groups and visit.

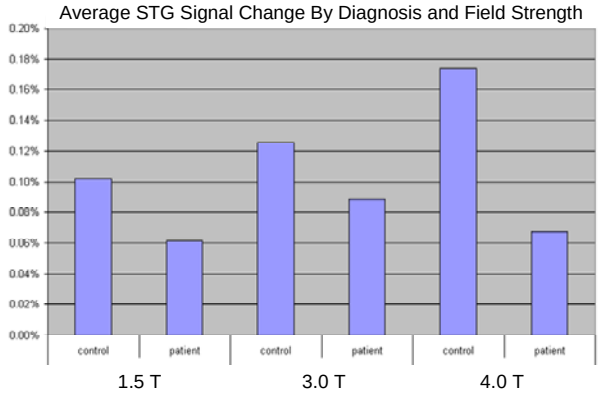


Fig. 4. Average STG signal change by group and scanner field strength.

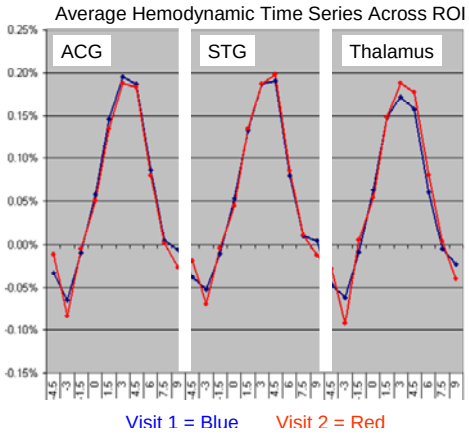


Fig. 5. The estimated hemodynamic response across subjects by region and visit.

The STG activation difference between Patients and Controls was present at all field strengths. As expected, **the average activation evoked by auditory targets increased with field strength (Fig 4).** The field strength effects for auditory targets were closely mirrored by increases in signal that occurred in the Breath Hold calibration task. On a subject-by-subject basis, the correlation between Breath Hold and Auditory Target signal changes in the STG was .31 (p=.007).

Hemodynamic responses. The average hemodynamic time course was calculated for each ROI. The average waveforms showed close correspondence between visit 1 and visit 2, as shown in Figure 7 for the ACG, STG, and thalamus. In Figure 5, the blue line indicates visit 1 and the red line indicates visit 2 averaged across all subjects and

sites for the identical voxels. **As expected from the Variance Components analyses of the Step 3 traveling subjects data, the effect of visit is small.**

Activation maps. A comparison of the activation maps evoked by Auditory Targets for Controls and Patients reveals **a greatly diminished response for the Patients.** For example, in the coronal image below (Figure 6), strong activation of the dorsolateral prefrontal cortex was obtained for Controls, but not Patients. Activation of auditory cortex was also greater for Patients than Controls, as can be seen in the axial figure (Figure 7). However, the activation by Auditory Targets was present in Patients, but at smaller amplitude and thus, with a corresponding smaller functional signal-to-noise. If the activation threshold for the Patients is reduced to 1/3rd of that of Controls (far right image), then activation of corresponding areas can be observed in the patient data.

This clearly identifies an opportunity for applying the multi-site calibration and correction methods in clarifying these group differences.

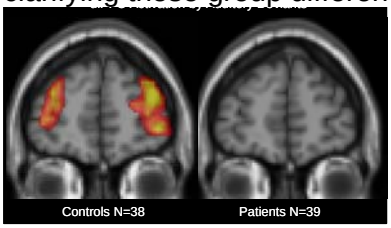


Fig. 6. DLPFC activation by Auditory Deviants.

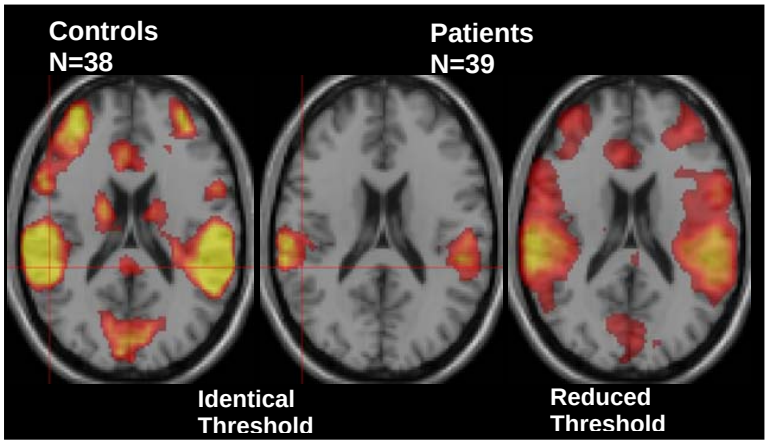
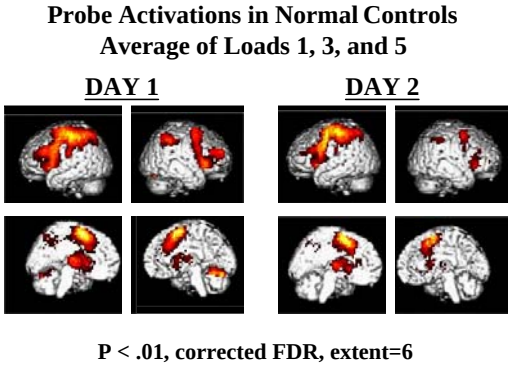


Fig. 7. Areas responding to the Auditory Deviants in controls and patients (at the same threshold) and in patients at a reduced threshold.

Series A, Step 4: Multi-site Data from the SIRP Task: Test-retest reliability



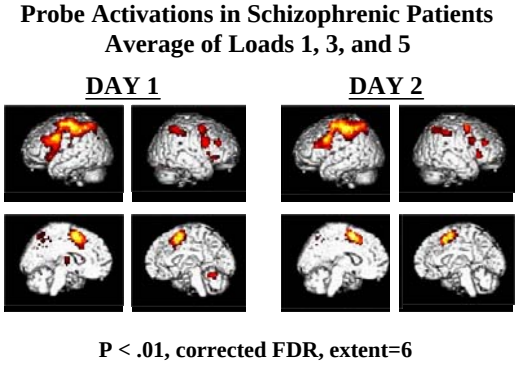


Fig. 8, left. Rendering of averaged activation for the various memory loads in the SIRP condition in controls across days.

Fig. 8, right. The same, for subjects with schizophrenia.

The preliminary analyses reported here were conducted on 56 subjects for whom two visits were available. This group included 29 control subjects and 27 patients with schizophrenia. The data were collected on 1.5, 3.0 and 4.0 Tesla scanners manufactured by Picker, Siemens, or General Electric. Many analyses were performed, however to simplify this report, we present only a few key results in the following.

Activations averaged across Memory Loads of 1, 3, and 5 items are shown in Figure 8 below separately for controls and patients for Day 1 and Day 2. Extensive DLPFC activation can be seen for both groups, in addition to ACG, parietal lobe structures and left motor cortex. There is a strong similarity between the group activations obtained on test (Day 1) and retest (Day 2).

Statistical assessment of test-retest reliability. To assess reliability of the SIRP activations in the DLPFC, we constructed left and right DLPFC masks using automated methods. We then extracted the average beta values (reflecting height of hemodynamic response activation) for the top 10% of voxels in the masks separately for each subject for the average activations across the three loads.

Test-retest reliability was assessed using Generalizability (G) Coefficients (see discussion above). The G-coefficients are presented in the table below, calculated separately for patients (n=26) and controls (n=29), as well as in the combined sample (N=55). The coefficients indicate modest to moderate test-retest reliability for the DLPFC activations associated with the SIRP memory probes.

	Left DLPFC	Right DLPFC
Sample	Average over Loads 1,3, 5	Average over Loads 1,3, 5
Normal Controls	.56	.68
Schizophrenic Patients	.51	.77
Combined Groups	.53	.72

The SIRP and Auditory Oddball data highlight the differences by task and diagnosis in activation of the same brain areas, such as DLPFC. Multiple tasks may be needed to develop and validate multi-site imaging methodologies that are robust across various cortical regions. The Cognitive researchers also performed additional analyses on a preliminary study that used the proposed dual SIRP and auditory mismatch task. We are prepared to discuss the latter results at the reverse site visit if requested.

4. Methods for removal of intersite variance

Given the measurements of individual reliability in the ICC's above, and the measures of intersite variability reported above, the reduction of intersite variability to improve the power of multi-site datasets is required. We have made significant progress on this through various means as applied to a subset of the Series A, Step 4 dataset.

Theoretical basis for proportional scaling using the Breathhold (BH) response. Since the proposal submission, a newer model has been developed that motivates the use of the breathhold (BH) task for calibration of confounding inter-subject vasoreactivity variations that can mask true intersubject neuronal differences. During task activation, the signal is given by

$$S_{act} = S_0[\, f_{act}^{\alpha-\beta} m^{\beta} - 1], \qquad (1)$$

where S_0 is a constant that depends on vasomotor reactivity and other local characteristics, $f_{act} = CBF_{act}/CBF_0$ is the fractional increase in flow relative to baseline, and $m = CMRO2_{act}/CMRO2_0$ is the fractional increase in metabolic rate of oxygen (1). The coupling between blood volume and blood flow may be characterized by $CBV \propto CBF^{\alpha}$. β is a nonlinearity parameter that accounts for diffusion and intravascular effects. During the BH task, CMRO2 does not change ($m = 1$), and the signal is

$$S_{BH} = S_0[\, f_{BH}^{\alpha-\beta} - 1]. \qquad (2)$$

Note that for a given brain region, the multiplicative constant S_0 is common between the two conditions of task activation and BH. This suggests that hemodynamic and other variations between brain regions and individuals encompassed in S_0 can be removed by dividing S_{act} by S_{BH} to obtained normalized values:

$$S_{norm} = [\, f_{act}^{\alpha-\beta} m^{\beta} - 1]/[\, f_{BH}^{\alpha-\beta} - 1]. \qquad (3)$$

Eq. 3 indicates that as long as the flow changes during BH and task activation are similar in a given region of cortex, the normalized values should have reduced variability when comparing different brain regions or in comparisons between individuals or groups of individuals because S_0 variations are removed. As calculations in Fig. 9 demonstrate, a nearly linear relationship is obtained that is characteristic of variations in metabolic changes ($n=(f_{act}-1)/(m-1)$), and that variations in F in fact have relatively small effect. Other work has shown that these flow variations are small (2,3).

The theory was validated by results presented in the following section (see Figure 11 below) and by results of a working memory (WM) study with children and adults at Stanford. Scaling the WM effect size for each subject by their BH task effect reduces the inter-subject variance resulting from uninteresting hemodynamic differences and thereby increases the group activation as shown in Fig. 10.

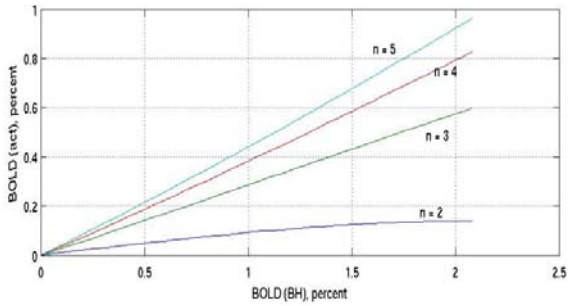


Fig. 9. BOLD signal for task activation vs. corresponding BH signal, for different $F \equiv f_{BH}/f_{act}$. A linear relationship holds over large variations in F.

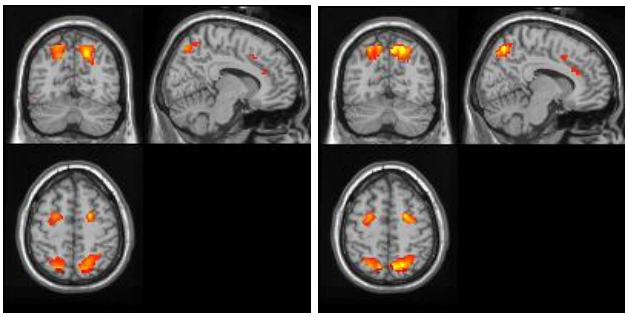


Fig. 10. Comparison of activation for 16 children and 16 adults in spatial working memory task without (left) and with (right) BH calibration. Greater group activation is demonstrated after calibration (9% > voxel volume). (M. Thomason & G. Glover)

<u>Increase in Between Group Effect Sizes through FBIRN Calibration in a multi-site dataset.</u> A central mission of the FBIRN program is to develop methods to reduce the impact of scanner differences when testing clinically relevant hypotheses using BOLD data collected from multiple sites. As shown above, FBIRN studies of BOLD signals obtained from the same subjects at different sites revealed a large impact of site on subject results. In the previous grant period, FBIRN investigators developed two calibration methods, proportional scaling and smoothing to a target full width half maxmimum (FWHM) degree of spatial correlation, in order to reduce site effects. The current analysis examined the impact of these calibration methods on the effect size associated with a diagnosis by site analysis of variance of BOLD data from the FBIRN auditory oddball task in an auditory ROI. Thirty subjects' data from multiple FBIRN sites in the Series A, Step 4 dataset were analyzed.

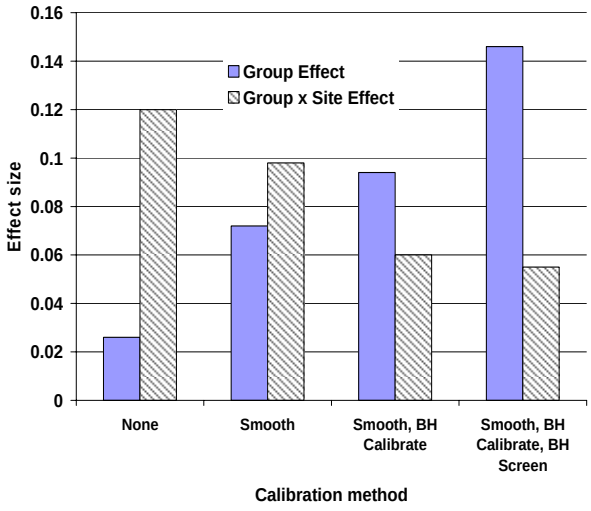


Fig. 11. The impact of calibration method on ANOVA observed effect sizes.

Four calibration methods, applied cumulatively, were compared: (1) no calibration – the native images from each scanner were analyzed; (2) smooth-to calibration -- for each run the mean image intensity from each scanner was normalized to the same value and all echoplanar images were smoothed to 8 mm FWHM; (3) breath-hold proportional scaling – in addition to smooth-to calibration all images were normalized by dividing a subject's BOLD response to the oddball task by their BOLD response to the breath-hold condition; (4) breath-hold scaling

with pre-screening – outliers in the breath-hold data (based on exploratory tools) were discarded then the calibration described in method 3 was performed. Eighteen healthy volunteers and 12 schizophrenia patients were studied at three FBIRN sites (two Siemens/one GE site; two 1.5T scanners/one 3.0T scanner). Effect sizes were measured by Cohen's f, which is directly related to observed power.

Figure 11 shows the impact of the four scaling conditions on Cohen's f. The no calibration condition (1) shows a small effect size of Diagnostic Group and a large effect of the Group by Site interaction. Clearly **generalizations about differences between groups are complicated by the variability of the group difference across sites.** The smooth-to calibration method (2) increased the effect of group and reduced the impact of the group by site effect, two desirable results. Adding proportional scaling (3) to the calibration pathway further improved the results. Once outliers (three subjects) were discarded, **the combined calibration method (4) revealed a large effect of group (Cohen's f > .40).** Data from one schizophrenia patient is instructive. This patient had a very large BOLD response to breath-hold and had one of the largest BOLD responses in either group on the oddball task. Only at the site where this subject was studied was the mean oddball response in patients greater than controls. Breath-hold proportional scaling helped to mitigate the adverse impact of data from this abnormally responsive patient. Although the calibration methods described above reduced the magnitude of the group by site interaction, it still remains moderately large. We anticipate that calibration methods proposed in the current application will further reduce the variation of diagnostic group by site.

Preliminary Path Analysis of Auditory Oddball Data. The analyses involving the STG results for the Auditory Oddball in patients and controls are initial tests of calibration methods using known, robust effects. The goal for multi-site imaging datasets is to reveal more subtle effects that cannot be found in a limited sample size. The circuitry analyses (as discussed in the Statistics Section D) require larger datasets. Continued progress estimating sample sizes needed for these analyses has been made using a subset of the Series A, Step 4 dataset this summer.

Following a deconvolution analysis to produce subject level data, areas of significant activation were located by voxelwise t-tests, using a volume cluster to control for Type I errors. The same dataset as discussed in the above calibration results was used. Not surprisingly, significant areas of activation included the auditory cortex and the sensorimotor hand area in each hemisphere. In this analysis a continuous region of significant BOLD response found in the middle and superior frontal gyri spanned the two hemispheres. Using knowledge of functional neuroanatomy and information from the correlation matrix involving the regions of interest (ROIs) described above, two classes of models were tested. The first class assumed that information from the auditory cortex flowed into the frontal ROI either directly from the auditory areas or indirectly through the hand area. This class of models treated the frontal cortex as a tertiary projection area for auditory information. These models did not fit the observed covariance matrix well.

The second class of models assumed that the frontal region had a primary modulatory role with no antecedent information flowing from other ROIs being modeled. An example of a well-fitting model from the second class appears in Figure 12 [$\chi^2(3)$ = 1.568, Adjusted Goodness of Fit Index = .90 (max = 1.0), error terms omitted in diagram to avoid clutter]. The frontal cortex modulated BOLD response not only in the two hand areas but also in the two auditory regions, directly in the left hemisphere

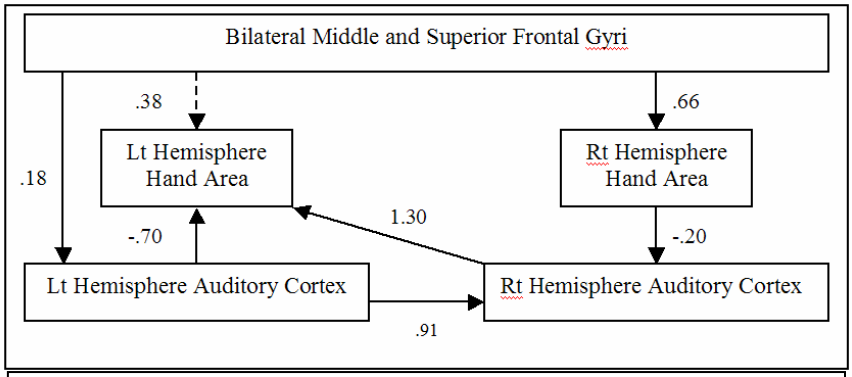


Fig. 12. Example model with a modulatory role for frontal cortex

and indirectly in the right. Although the path coefficients suggest several interesting functional relationships among the brain regions modeled, the coefficients are associated with some uncertainty that is expressed in their standard errors.

We investigated the question of how sample sizes influence the power of the χ^2 test to reject the null hypothesis that a path coefficient is zero against the alternative that the path coefficient is equal to the lower bound of the 95% confidence interval [CI] for the coefficient of interest. As an example, for the path from the frontal region to the left hand area (dotted arrow), the observed power to detect a significant improvement in model fit is .76 when comparing the alternative hypothesis that the coefficient is equal to the observed value of .38 versus the null hypothesis of zero. Using bootstrapped estimates of the standard error, the power to detect a lower bound of the 95% CI for this coefficient was estimated to be .30 for 50 subjects, .62 for 100 subjects, .67 for 150 subjects, .70 for 200 subjects, and .72 for 300 subjects within a single group. **Investigations of other paths indicated that sample sizes between 150 and 300 are required per group modeled to obtain a reasonably narrow CI and adequate power for auditory oddball path-coefficients.**

The above progress report for Summer 2005 is a selected representation of FBIRN efforts since the submission. We look forward to discussing the details of this work, other progress not reported here, and the full proposal on August 31.

Citations

1. Davis TL, Kwong KK, Weisskoff RM, Rosen BR. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc Natl Acad Sci U S A 1998;95(4):1834-1839.
2. Kastrup A, Kruger G, Neumann-Haefelin T, Glover GH, Moseley ME. Changes of cerebral blood flow, oxygenation, and oxidative metabolism during graded motor activation. Neuroimage 2002;15(1):74-82.

NEW or updated since submission

3. Thomason ME, Burrows BE, Gabrieli JD, Glover GH. Breath holding reveals differences in fMRI BOLD signal in children and adults. Neuroimage 2005;25(3):824-837.
4. Friedman L, Krenz D, Magnotta V, FIRST BIRN (2005) Isolated Spike-like Transients (ISLTs) in Functional MR Time series: Evidence for Vendor and Field Strength Effects on Rates. In: 11th Annual OHBM Meeting. Toronto, Ontario, Canada.
5. Friedman L, Krenz D, FIRST BIRN (2005) "Field-Strength and Scanner Vendor Effects on the Degree of Temporal Autocorrelation in fMRI Time series." In: 11th Annual OHBM Meeting. Toronto, Ontario, Canada.
6. Friedman L, Krenz D, FIRST BIRN (2005) Scanner and Field-Strength Effects on the Shape of the Hemodynamic Response Function. In: 11th Annual OHBM Meeting. Toronto, Ontario, Canada.
7. Lee, HJ, Turner, J., Potkin, SG., FIRST BIRN. (2005) Calibration Method of Functional Magnetic Resonance Images. Proceedings of the 2005 International Conference Imaging Science, Systems, and Technology: Computer Graphics (CISST'05), Las Vegas June 27-30 2005, pp 13-19.
8. Friedman, L., Zhao, Q, Cheng, H., Duensing R., Greve, D.N., Glover, G.H., A. Functional BIRN Consortium (2005): Multi-center fMRI Calibration with SMARTPHANTOM. 13th Annual Meeting of ISMRM, Miami Beach, FL.
9. Friedman, L., Krenz, D., Magnotta, V., Functional BIRN (2005): Isolated Spike-Like Transients (ISLTs) in Functional MR Time series: Evidence for Vendor and Field Strength Effects on Rates. 13th Annual Meeting of the ISMRM, Miami Beach, FL.
10. Kim, S., Smyth, P., Stern, H., Turner, J., FIRST BIRN. (2005) Parametric response surface models for analysis of multi-site fMRI data. Proceedings of the 8th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). *Lecture Notes in Computer Science*, Springer-Verlag, New York, in press.
11. Turner, J.A., Smyth, P., Fallon, J.F., Kennedy, J.L., Potkin, S.G., FIRST BIRN (2005). Imaging and genetics in schizophrenia. *Neuroinformatics,* in press*.*
12. Keator, D; Gadde, S; Grethe, J ; Taylor, D; FIRST BIRN; Potkin, S. A General XML Schema and Associated SPM Toolbox for Storage and Retrieval of Neuro-Imaging Results and Anatomical Labels. *Manuscript under review*.