

Foundations for Multi-Center Neuroimaging Studies

Gary H. Glover, Ph.D.

Department of Radiology, Stanford University

Introduction

Magnetic resonance imaging (MRI) data are increasingly being used as biomarkers for validation of efficacy in therapeutic, pharmaceutical and device development studies. Because of the need for large scale enrollment and diversity in the study population, as well as a desire to reduce the total duration of the imaging phase, multi-center studies are becoming increasingly popular. However, the ability to pool data from multiple sites is based on the assumption that there are no site effects- i.e., a phantom or random subject scanned at any of the sites must provide imaging data that can not be differentiated from that from any other site in the consortium- preferably as acquired or after a calibration. For this to be true, and based on the type of imaging study, careful attention to pertinent details is critical. This talk will introduce issues common to all types of multi-center MRI (MC-MRI) studies.

Issues in Multi-Center MRI

In any MC-MRI trial, one requires careful imaging design based on the goals for the imaging but considerate of the site diversity, and this engenders a number of questions:

1. What tissue/structure/functional parameters are being studied, and what precision is required to support or deny the null hypotheses of the study? This may sound trivial, but the effects of small variations in image bias/shading, contrast in grey/white/CSF tissues, SNR variations with coil types, reconstruction algorithm differences and many other site/vendor-specific characteristics may produce subtle changes in the end measurement of, e.g., cortical thickness, that are poorly understood or cannot be estimated a priori. After performing a critical sensitivity analysis to the extent possible based on available data, some pilot studies are recommended, and iteration of the design to minimize the impact of site variability may be necessary.
2. Does each site have the requisite hardware and software to obtain identical image data (perhaps with correction), if not identical acquisitions? How does one equilibrate pulse sequence choices across different vendors? For example, the 3D-MPRAGE pulse sequence is standard issue for several vendors but only an alternate sequence (3D-FSPGR) is available on others. Can parameters be found for each that allow the resulting images to be made equivalent for purposes of the study? Will some sites use receive-only head coils and others use T/R volume coils (not recommended!)? Are there differences in slice profiles due to different RF pulse designs, differences in artifacts from, e.g. crusher gradients, phase corrections, or eddy current characteristics, or differences in *effective* receive bandwidth? Is the geometric precision adequate and the same across all sites? How reproducible are data taken on different days at each site? In the case of functional studies, how critical are the characteristics of the ancillary equipment such as stimulus or response box equipment across sites? Very importantly, how many sites will be undergoing hardware or software upgrades during the course of the study, and what impact will the upgrade have on the resulting data?
3. A special case of hardware issues: Will multiple field strengths be included? If so, special caution must be exercised because of the inherent differences in SNR, tissue NMR parameters leading to contrast differences (e.g. for 3T vs. 1.5T, BOLD contrast has different proportions of

diffusion and susceptibility weighting and the longer T1 of blood may lead to better SNR in perfusion studies), differences in artifacts such as susceptibility-induced signal dropout in fMRI or geometric distortion in EPI-DTI, B1 heterogeneity, safety concerns for some patient populations, different proportions of cognitive, physiological and thermal noise and others. In the case where multiple field strengths are included, pilot studies will be required to determine if the site effects can be minimized, or properly accounted for if there is a specific intention to compare field effects.

4. Each site may have different types of personnel performing the scans and overseeing other aspects of the study. It will be important to have adequate training so that the operators understand the criticality of uniformity in patient/subject preparation and setup (e.g. head restraint method), slice prescription, verification of the proper protocol (do any special options need to be enabled/disabled each time or performed in a specific unusual order?). It may be important to have a script or checklist to ensure uniformity within and across the sites.

5. If the analysis is centralized, how are the imaging and meta data sent to the analysis center? What is the quality control for integrity of this process?

Recommendations

There are general recommendations that apply to all types of MC-MRI studies, but many more that are specific to each study type. Here we list some of the common elements.

1. The most important step that can be taken to minimize inter-site variability is to standardize every aspect of the image acquisition and analysis to the extent possible. Because of vendor-specific (possibly subtle) differences in pulse sequence characteristics, coil sensitivity profiles, localization methods, etc. it may be desired to restrict the study to a single vendor. However, even within a single vendor there can be significant differences that depend on software release and gradient/RF hardware options. If it is not practical to utilize only one scanner model, then minimize the variability as much as possible and calibrate or correct for remaining differences. The type of such mitigation to be employed of course depends on the specific study type. It is highly recommended that the same type of head coil be used (single/multiple channel, receive-only/Transmit-receive).

2. Certify each site's hardware for entry into the study. This will involve initial tests of all important characteristics such as SNR/CNR, stability, geometric precision, artifacts, and reproducibility. This will by definition require that standards for measurement and the specifications to be met are established and rigorously applied. For DTI studies, distortions due to eddy currents will be critical to evaluate, while for functional and perfusion studies scanner inter-image stability and spike noise are critical to evaluate. If a scanner does not initially perform to specification, service must be performed until it can. Experience with the fBIRN has shown that vendors are highly responsive to such requests for service, because of the visibility of failure in a large consortium.

3. Implement periodic Quality Assurance (QA) procedures that probe critical image characteristics, and upload such data to the central data repository. This process ensures that every site has an on-going validity check that can be used in cases when data from a site on a given day is questionable. Furthermore, with a log of each site's performance over time, it will be easy to determine if the scanner requires maintenance or recalibration, since its properly performing characteristics are known. Elements of a QA program can include measuring geometric precision, short term/long term stability, SNR, CNR, center frequency,

transmitter/receiver gains, artifacts and system/thermal noise, using a standardized phantom(s) and automated acquisition/analysis protocols.

4. After initial site qualification and QA, certify each site's ability to acquire the imaging and meta data properly under study conditions. This ideally requires that a coordinator visit each site while one or more scan sessions are performed for training and to ensure uniformity. The importance of this cannot be overestimated, because an external observer will be able to spot small details whose significance in the larger picture might not be appreciated by site personnel. During the fBIRN study's initial setup, for example, it was found that visual display systems had as much as a four-fold difference in lumen intensity across 10 sites, with some very dim and others uncomfortably bright. The study design may also utilize traveling subjects that visit every site, in order to quantify inter-site precision with actual human imaging data. While this process can be expensive, the information so obtained is extremely important and unavailable by other means.

5. Monitor quality of ongoing collection of data, either using automated analysis routines or using experienced observers at each site, or both. If there are questions about scan data integrity a year after it was acquired, there may be no good reason to disqualify that data unless it was flagged in a metadata entry (e.g. phase encoding direction was reversed, slice prescription was incorrect).

6. Perform data analysis centrally. In addition to being able to standardize and economize on the analysis effort, implementing a central data processing mechanism will have the added benefit of having oversight on the image quality from all sites and will be in a better position to spot inconsistencies quickly for remediation. Remnant inter-site differences may be more readily calibrated out or otherwise accounted for during a centralized analysis. Note that central analysis does not necessarily imply a central data repository.

Conclusions

The technical and ultimate biomedical success of multi-center MRI studies depends on rigorous attention to details, from inception of design to data analysis. Some studies have less rigorous requirements than others (arguably structural MRI vs. fMRI, e.g.), but in most cases pilot studies will be required in order to finalize the study design. In all cases, site qualification and standardization, continuing QA using phantoms and inspection of imaging data and uniformity of procedures across sites are critical. As multi-center MRI studies become increasingly implemented, further advances in minimizing site effects will be developed, such as vendor-available standardized pulse sequences and analysis software.

References

1. Olsrud J et al. A two-compartment gel phantom for optimization and quality assurance in clinical BOLD fMRI. *Magn Reson Imaging*. 2008 Feb;26(2):279-86.
2. Friedman L et al. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging*. 2006 Jun;23(6):827-39.
3. Automated analysis of multi site MRI phantom data for the NIHPD project. Fu L, Fonov V, Pike B, Evans AC, Collins DL. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*. 2006;9(Pt 2):144-51.
4. Multi-center validation of the transferability of the magnetic resonance T2* technique for the quantification of tissue iron. Tanner MA, He T, Westwood MA, Firmin DN, Pennell DJ;

Thalassemia International Federation Heart T2* Investigators. *Haematologica*. 2006
Oct;91(10):1388-91.