Multisite Collaborations and Large Databases in Psychiatric Neuroimaging: Advantages, Problems, and Challenges

Godfrey Pearlson

Studying large numbers of subjects in psychiatric neuroimaging research is often necessary to generate sufficient statistical power to reveal more subtle differences missed in smaller analyses and to allow subtyping in what are often heterogeneous diagnostic groups. Such an approach is probably essential in the case of clinical studies dealing with groups that are difficult to recruit from a single site in large numbers. This issue is brought into even sharper relief where neuroimaging and genetics are both part of a particular study. Dealing with such high-dimensional datasets often implies the need for both multisite collaborations with regard to data collection and specialized statistical approaches with regard to data analysis. These issues deserve detailed discussion and careful examination as multisite imaging collaborations begin to proliferate in neuropsychiatric research. Recent large-scale multisite collaborations with neuroimaging as an essential component include the structural and functional imaging arms of the Biomedical Informatics Research Network (BIRN, http://www.nbirm.net/), the Alzheimer’s Disease Neuroimaging Initiative (ADNI, http://www.adni-info.org/), the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP, http://www.b-snip.org/), and The North American Prodrome Longitudinal Study (NAPLS, http://www.schizophreniaforum.org/new/detail.asp?id = 1339), all funded by National Institutes of Health.

As noted above, there are many pluses of multisite collaborations. Results gathered in diverse clinical populations are more likely to generate generalizable data and to provide adequately sized and powered samples. Planning such collaborations forces the issue of standardization of imaging collection and processing and the creation of centralized neuroinformatics repositories. This in itself is complex and probably one of the major reasons brain imaging is lagging behind large-scale genetic studies.

It is likely that multisite collaborations benefit by being vetted by a team of collaborating scientists from the beginning, although in the minus column they require more administrative oversight/infrastructure to ensure all aspects of the study are performed consistently. Naturally, this also leads to tool development and dissemination (which is one of the goals of Functional BIRN). Collections of large datasets inevitably lead to attempts to automate and increase efficiency of associated procedures including automated software to read image data and radiological header files and automatically report back departures from protocol. Each site must remain vigilant to fluctuations in image quality and stability not only day to day but also over longer time periods in terms of signal to noise. For example, one multisite study discovered functional imaging fluctuations, ultimately traced after a heroic search to the occasional passing of metro trains in a tunnel adjacent to the hospital.

Other problems associated with multisite collaborations include the fact that use of different platforms of neuroimaging equipment and other related site differences add undesirable heterogeneity, necessitating complex preparation in terms of sending both inert radiologic phantoms and traveling subjects, often the investigators, in the form of “human phantoms” from site to site to ensure maximum comparability, a strategy used both by BIRN and ADNI. The ADNI study, that involves over 100 participating sites in the United States, devoted much time and effort to the preparatory comparison of different magnetic resonance imaging (MRI) scanning platforms in order to obtain maximally comparable structural imaging sequences on different model scanners. This also involved working closely with the scanner manufacturers and obtaining their collaboration in the face of often closely guarded proprietary information.

Despite such precautions, multisite studies still need to add site as a covariate in analyses, advanced analysis approaches such as independent component analysis (ICA) employed in the Kim article in this issue (and see Meda et al 2008). The former approach may be able to filter out much of the multisite heterogeneity.

However, while such approaches can be used to identify and to minimize site effects, they cannot fully remove them. Experience from the BIRN study identifies another key element in the group comparison approach. Calhoun Vince 2008 (personal communication) found in most cases that between-group differences were consistent across site (few interactions of site and group) but that there were very significant main effect site differences.
This gives emphasis to the importance of collecting similar numbers of patients and controls at each location, which seems to control for many of the site differences.

With regard to image analysis, many studies have discovered the advantages of data analysis at a single designated site for the entire study. Another issue is that image analysis techniques used in comparing large numbers of subjects are necessarily constrained. Readers are likely to be familiar with voxel-based morphometry (VBM), an efficient whole-brain unbiased technique, to examine between-group structural MRI differences. VBM provides automated measures of highly localized regions that may not be investigated in hypothesis-based studies that employ more labor-intensive region of interest (ROI) measures that become increasingly impractical in large N studies. The fBIRN and MIND Clinical Imaging Consortium (MCIC) study (see Segal’s VBM article in this issue) examined 503 subjects combined across both fBIRN and MCIC, another multisite schizophrenia study.

VBM is useful to identify regional differences in the concentration/volume of gray or white matter between groups by increasing the anatomical range of volumetric comparisons. For example, in schizophrenia VBM studies first identified lesser gray matter concentrations in insular cortex, a finding that has been remarkably consistent across multiple VBM studies and which was subsequently confirmed by hand-drawn ROI measures. However, in VBM studies of schizophrenia, differences in more anatomically variable regions, such as inferior parietal lobule, are reported less consistently, although confirmed in ROI studies. According to a recent meta-analysis of VBM studies differences in the latter areas, though reproducible, are inconsistently reported due to the variations in size of smoothing kernels used across studies, the heterogeneity of schizophrenia itself, and failure of small samples to adequately represent the disease (Honea and Crow).

Another associated methodological approach in statistical parametric mapping version 2 is the use of optimized VBM, which is the normalization of gray matter (GM) segments to a study-specific template. This helps best represent the GM distribution found within a particular study sample, helps account for data inhomogeneities due to different scanners, and minimizes the degree of warping required for coregistration. On the deficit side, such normalization algorithms although reliable, make the VBM process less sensitive to group differences in shape or gray-white matter differentiation than many ROI methods and may misregister anatomical structures. The removal of global size differences in the normalization process is a caveat in all VBM studies, especially where brain atrophy occurs, as in Alzheimer disease. More recently, the new SPM5 unified segmentation algorithm (discussed in the Segall article in this issue) has replaced optimized VBM and removes the need for a study specific template.

There are other problems to tackle in analyzing large, complex datasets. The “curse of dimensionality” looms large when modeling high-dimensional discrete data, eg, hundreds of thousands of voxels and/or genes typically involved in a Genome Wide Association Study/functional imaging study and rapidly devolves to a needle in a haystack search, as the number of possible variable combinations balloons. This challenge has spawned new statistical techniques. For example, novel methods such as parallel ICA represent new approaches for analyzing multimodal data. A recent article (Liu et al. 2008) used this type of algorithm to identify simultaneously independent components of imaging and genetic modalities and the relationships between them. This issue is relevant to the Potkin single-nucleotide polymorphism/functional MRI article in this issue.

In choosing appropriate neurocognitive tasks for use in multisite studies, one can adopt several strategies. One choice is to focus on tasks that emphasize between-group differences in both behavior and brain function, eg, working memory tasks where, eg, schizophrenia patients (and in some cases their relatives) often perform markedly differently from healthy controls. Differences in brain response can then be compared, given suitable test designs, at either equal group performance or at a level equal task difficulty, an issue discussed in the second (sternberg item recognition paradigm) Potkin article in this issue.

An alternate approach is to choose simpler structured, robust tasks such as an auditory oddball paradigm, where schizophrenia and healthy control subjects both perform well and at similar accuracy levels but nonetheless show differences in blood oxygen level–dependent activation patterns or even paradigms utilizing no cognitive task at all, such as resting state designs examining the brain in an “idling state.” Such designs prove more straightforward to equate across different imaging sites.

The ability to combine imaging data across multiple sites is a critical advance. Success depends on careful preparation in order to maximize comparability with regard to methods, data collection, image analysis, and statistical analysis. Confounds, if not restrained, will undermine the advantages of multisite data collection. The growing use of large neuroimaging databases in psychiatric research will bring increasing sophisticated and powerful tools to bear on the questions of brain abnormalities associated with psychiatric disorders. These techniques will be increasingly important in helping determine etiopathology.