



nature *neuroscience*

Analyzing functional imaging studies

Functional magnetic resonance imaging (fMRI) has been used to measure brain activity for about a decade, and positron emission tomography (PET) only a few years longer. Thus, it is no surprise that reasonable people continue to disagree about appropriate experimental design and analysis. How do we maintain consistent criteria for publication under these circumstances? Although it would be premature to set absolute requirements without broadly accepted minimum standards in the field, we can explain the factors that we weigh in judging one paper against another.

Some standards, of course, are common to excellent science on any topic. To be published in *Nature Neuroscience*, papers should be exciting to specialists, as well as interesting to people outside the field. The research should be hypothesis-driven; for imaging studies, this means asking questions like “Is the hippocampus involved in retrieval of episodic memories?” rather than “What happens in the brain when subjects play chess?” Some interesting studies test hypotheses derived from experimental animal work, but it is ideal to take advantage of the unique abilities of human subjects, such as language.

Well-designed imaging studies allow scientists to ask questions about basic cognitive processes, rather than identifying networks of brain regions activated by a series of tasks. Such research relies on the authors’ ability to isolate the cognitive process of interest, and so the sophistication of the behavioral design is crucial. Experimental conditions should differ along the fewest possible dimensions, preferably only the parameter under study. In particular, variations in task difficulty or attention levels between conditions can lead to inappropriate conclusions. Many papers are rejected based on poorly chosen control tasks, or because control subjects are not well matched with experimental subjects for age, IQ or other relevant characteristics. Imaging studies are strengthened by correlations between behavioral performance and brain activation, particularly when these correlations can be demonstrated on single trials or for individual subjects. Psychophysical stimuli that can be quantified and experimentally manipulated along a continuum are well suited to parametric studies, in which degree of activation is correlated with changing values of a variable such as stimulus strength. If differences between conditions are reported, the conditions should be compared against each other directly, rather than showing that one condition differs from a baseline condition and the other does not.

For most applications, fMRI is the technique of choice. The number of PET scans per subject is limited by the use of radioactive contrast agents, whereas fMRI is non-invasive. In addition, PET has a lower temporal resolution than fMRI, and a lower theoretical limit on spatial resolution (although in practice fMRI data are often smoothed to a similar spatial resolution). However, PET allows researchers to study particular neurotransmitter systems with radiolabeled compounds, such as raclopride for D₂ dopamine receptors. In addition, because fMRI has different sensitivity in

different brain regions, PET is sometimes used to study areas where fMRI is relatively insensitive.

Any functional imaging experiment generates gigabytes of data, and reducing this massive dataset to areas of activation in a published figure involves a series of analysis choices that can influence the study’s conclusions. Researchers can ask different questions with the same data by using fixed- or random-effects analysis to identify areas of activation. The fixed-effects model determines whether a particular activation is significant by comparing its magnitude against the variability across scans for each subject. This is essentially a case study method; it gives highly reliable effects for individual subjects, but the results do not generalize beyond the individuals tested. In contrast, the random-effects model determines significance by comparing activation levels against the variability between subjects. This method requires more subjects to produce a significant effect, but allows conclusions about the population from which the subjects were drawn.

There is little consensus on appropriate statistical significance thresholds, which vary considerably across studies. A *p*-value of 0.01 means by definition that 1 in 100 tests will give a false positive result. Because imaging studies require comparing many thousands of voxels, researchers often correct for multiple comparisons, using a more stringent criterion to avoid getting false positives by chance. Increased specificity, however, comes at the cost of reduced sensitivity to real activations. This trade-off can be minimized by region-of-interest analysis, in which researchers define the areas to be studied *a priori*, based on previous literature. This approach increases sensitivity by reducing the number of voxels tested, making the correction for multiple comparisons smaller. Such analysis is strongly hypothesis-driven, but it can also lead authors to ignore activity in other parts of the brain, thus oversimplifying their story. In addition, some researchers report significance levels without showing the magnitude or time course of their activations, making it difficult for readers to evaluate the relative strength of activity in different conditions.

Comparing activity across subjects is problematic because individual brains are shaped differently. Brain anatomy can be identified in individual subjects by structural MRI, and then aligned with functional activation maps. Alternatively, approaches such as retinotopic mapping can be used to identify brain regions; such functionally defined areas show somewhat different anatomical positioning between subjects. Both approaches are widely used, although their results are difficult to compare.

Much of this uncertainty may stem from a fundamental problem: we do not fully understand the biological basis of functional imaging signals. Responses vary with blood flow and oxygenation, and although such changes presumably relate to local energy use, and thus to electrical signals in neurons, the precise relationships among these parameters remain unclear. Further research on the neural activity underlying functional imaging activations may lead to more principled choices in data analysis and interpretation.