Statistical issues in functional neuroimaging: independence, reliability, and dead fish

Disclaimer

Halloween horror story
Halloween horror story

“Across all datasets we found that 48% of studies in the fMRIDC archive had issues that prevented their reanalysis.”

Big Year in fMRI Stats

Vul et al. 2009
Kriegeskorte et al. 2009

Some Background
The Voxel

Cubical volume of measurement in MRI

The “black box” of neuroimaging

brainmap.org
Two Challenges

**Dimensionality**: the staggering number of measurements

**Noise**: the staggering amount of measurement error

Lingering Questions / Talk Outline

- How do you independently summarize groups of voxels?
- How important is multiple testing correction in fMRI?
- How reliable are the results from fMRI studies?

The Non-Independence Error
Regions-of-Interest (ROIs)

A Cited Example

The Non-Independence Error

\[ r_{\text{observed, observed}} = \frac{r_{AB} \cdot \sqrt{\text{reliability}_A \cdot \text{reliability}_B}}{1 + r_{AB} \cdot \sqrt{\text{reliability}_A \cdot \text{reliability}_B}} \]

Assuming reliability_{behavior} = 0.8 and reliability_{fMRI} = 0.7, expected correlations should rarely exceed 0.74.
The Non-Independence Error

Vul et al., 2009

The Issue

Ideal

Actual

Kriegeskorte et al., 2009

The Issue

- True Region
- Overfitted Region
- Independently Selected
Multiple Testing Correction

Piloting Experiments

Social Perspective Taking
The Atlantic Salmon

What Happened?

What would you conclude?
Presenting Our Findings

Importance

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Significant Vessels</th>
<th>Sens. Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Fluency</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Location Switching</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Task Switching</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Face: Main Effect</td>
<td>172</td>
<td>317</td>
</tr>
<tr>
<td>Face: Interaction</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Face Recognition</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Emotional Pictures</td>
<td>560</td>
<td>1269</td>
</tr>
<tr>
<td>Pain: Warning</td>
<td>121</td>
<td>118</td>
</tr>
<tr>
<td>Pain: Anticipation</td>
<td>14</td>
<td>34</td>
</tr>
<tr>
<td>Pain: Pain</td>
<td>38</td>
<td>349</td>
</tr>
</tbody>
</table>

Nichols and Hayasaka, 2003

Prevalence

<table>
<thead>
<tr>
<th>Journal</th>
<th>Count / Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Brain Mapping</td>
<td>14/57</td>
<td>25%</td>
</tr>
<tr>
<td>NeuroImage</td>
<td>67/260</td>
<td>26%</td>
</tr>
<tr>
<td>Cerebral Cortex</td>
<td>26/80</td>
<td>32%</td>
</tr>
<tr>
<td>JoCN</td>
<td>26/68</td>
<td>38%</td>
</tr>
<tr>
<td>SCAN</td>
<td>10/25</td>
<td>40%</td>
</tr>
</tbody>
</table>
**Our Arguments**

- A principled method should be used to threshold fMRI results.
- The practice of reporting results from a peak voxel should be avoided.

**Correction Methods**

- Composite Display
  - \( p(\text{uncorr}) < 0.001 \)
  - \( p(\text{FDR}) = 0.05 \)
  - Hybrid


**fMRI Reliability**

**The Big Question**

If you did a scan twice, how stable would values be?
Measuring Reliability

Overlap of significant voxels 
(Rombouts et al., 1997)

Measuring Reliability

\[ \text{ICC}(3, 1) = \frac{\text{BMS-EMS}}{\text{BMS} + (k-1)\text{EMS}} \]

Correlation of voxel values 
(Shrout and Fleiss, 1979)

Experiment Design

<table>
<thead>
<tr>
<th>Block Design</th>
<th>Event-Related Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episodic Recognition</td>
<td></td>
</tr>
<tr>
<td>Two-Back Working Memory</td>
<td></td>
</tr>
</tbody>
</table>
Results: Event-Episodic

For an event-related episodic recognition task...

= Histogram bins of ICC correlation values

Results: Across Conditions

Episodic / Block Design
Episodic / Event-Related Design
Working / Block Design
Working / Event-Related Design

= ICC correlation values of thresholded voxels

Results: Thresholded
Results: Six Months

Really, that low?

Literature Review

Mean Overlap (26 papers) = 0.29
(Dee overlap = 0.45, Jaccard overlap = 0.29)

Mean ICC (15 papers) = 0.47

**Overall Conclusions**

- Use a principled method to control false positives.
- There are many factors that influence fMRI reliability.
- The reliability of some fMRI results may be lower than many of us assume.

**Thanks to the Miller Lab, and others**

- Mike Miller
- George Wolford
- Abigail Baird
- Scott Guerin
- Christa-Lynn Donovan
- Elissa Aminoff
- Danielle King
- Amy Frithsen
Thank You

bennett @ psych.ucsb.edu
http://prefrontal.org