Cluster Failure: fMRI's Big Shake-Up

Chris Hammill

September 13, 2016

Tens Of Thousands Of FMRI Brain Studies May Be Flawed

Bug in fMRI software calls 15 years of research into question

Popular pieces of software for fMRI were found to have false positive rates up to 70%

Science News *from research organiz*

Softwares for fMRI yield erroneous results

Cluster failure: Why fMRI inferences for spatial extent have inflated false positive rates

OOPSIE! —

Software faults raise questions about the validity of brain studies

Interpretation of functional MRI data called into question.

JOHN TIMMER - 7/1/2016, 2:55 PM

Figure 1:

So What Happened

- Eklund, Nichols, and Knutsson demonstrated standard fMRI statistical inference has badly inflated false positives rates
- Makes you wonder if exciting brain region X responding to stimulus Y finding was just a cherry-picked false positive.
- Highlighted that due to non-reproducible workflows, and poor data sharing, many of these finding could never be repeated with valid inference.

How Did We Get Here

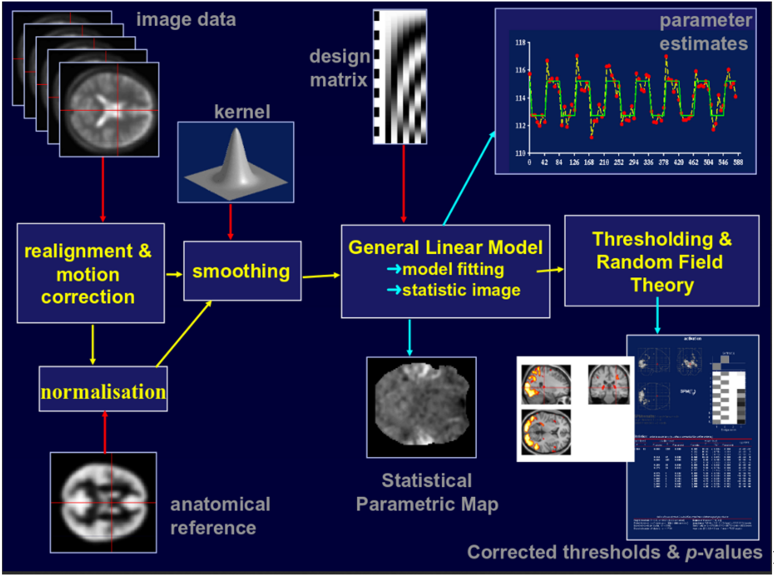
- fMRI is challenging to analyze
- Preprocessing steps widely used as black boxes
- Desire to use spatial information to determine signal significance
- Improperly specified models of spatial noise

About Group Task-Based fMRI

- Most fMRI seeks to measure brain activity by blood flow
- Blood oxygen level dependent (BOLD) contrast is used
- A time-series of volumes are acquired for each subject
- Stimuli are presented to the subject throughout the time series
- The BOLD signaled is modelled as a function of the stimuli
- The statistical association of the BOLD to the stimuli is compared across groups

Why Is This Tough

- Subjects move:
  - within subject each fMRI volume must be aligned to each-other
  - these must be aligned to a corresponding anatomical scan
  - these must be registered to a common space
- BOLD signal is sluggish
  - $\sim$ 2 seconds to start
  - $\sim$ 4-6 to peak
  - $\sim$ 10 to return to baseline so the stimulus time series is convolved with a function to match this behaviour
- Analyzing time series comes with it's own statistical challenges
  - how do we model temporal autocorrelation



[1]

[1] Borrowed from Nichols (2010)

Multiple Comparisons

- As with most imaging analysis, multiple comparisons is significant concern
- Solutions:
  1. Bonferroni: control your type one error rate by multiplying your results by the number of tests. This is equivalent to setting your type one error rate to α/n
  2. FDR (Benjamini-Hochberg): Order your p-values lowest to highest and accept or reject with increasing stringency α/i.
- But in low power situations this decreases sensitivity an unacceptable amount

Enter Spatial Models

- Signals with large spatial extent are probably more likely to be real than individual high intensity
- Question becomes, how do we analyze spatial extent, and how do we correct for multiple comparisons?
- Main Idea: threshold your data and use random field theory (RFT) results to assign a p-value to clusters based on their size
- Or: Assume some properties of the spatial distribution and generate a randomization distribution of cluster sizes, assign p-values from this.

The Problems

- When statistics maps aren't smooth enough, RFT p-values are biased
- RFT typically assumes a stationary noise distribution (uniform noise over the brain)
- Together these problems