New Procedures for False Discovery Control

Christopher R. Genovese

Department of Statistics

Carnegie Mellon University

http://www.stat.cmu.edu/~genovese/

Elisha Merriam

Department of Neuroscience

University of Pittsburgh

This work partially supported by NIH Grant 1 R01 NS047493-01 and NSF Grant SES 9866147..

Multiple Testing In fMRI

- In fMRI, we attach statistics to brain locations – voxels, triangles, regions – but our inferences combine many locations.
- For m simultaneous hypothesis tests, get p-values $P_1, \ldots, P_m$. Consider rejecting null hypotheses i with $P_i \leq t$.
- For any rejection threshold t, classify results as follows:

	H_0 Retained	H_0 Rejected	Total
H_0 True	TN	FD	T_0
H_0 False	FN	TD	T_1
Total	N	D	m

Mnemonics: T/F = True/False, D/N = Discovery/Nondiscovery

- Want to find t that yields both sensitivity (small FN) and specificity (small FD), but *these goals are in conflict*.

  Must find a desirable balance.

How to Choose a Threshold?

- Uncorrected Testing (Control Per-Comparison Type I Error)
  - high sensitivity, low specificity (equivalently: high power, many type I errors)
  - bounds probability of false discovery for *each* test separately
  - may be useful for small ROIs

- Strong Control of Familywise Type I Error Rate (FWER)
  - low sensitivity, high specificity (equivalently: low power, few type I errors)
  - guarantees $\mathrm{P}\{FD > 0\} \leq \alpha$ for any subset of true nulls
  - methods: traditional (e.g., Bonferroni), resampling, random field

- Move the Slider Until the Pattern Looks Right
  - Uhhh, no.

More Power To You

- Strong control of FWER is effective but often want more power.

  The problem is how to get more power while remaining principled.

- Alternative criteria:
  - Cluster-size Threshold (Forman et al. 1996): Use spatial information to separate signal from noise.
  - Control k-FWER: Fix k and guarantee $\mathrm{P}\Big\{FD > k\Big\} \leq \alpha$.
  - False Discovery Control: Bound ratio FD/D in some way.

    Leading example (Benjamini and Hochberg 1995): guarantee that

$$\mathrm{FDR} \equiv \mathrm{E}\left(\frac{FD}{D + 1\Big\{D = 0\Big\}}\right) \leq \alpha.$$

    FDR is called the *False Discovery Rate*.

Controlling k-FWER

- k-FWER generalizes FWER by not penalizing some fixed number of errors.
- Several approaches developed (van der Laan et al. 2004, Romano and Lehmann 2004, Finner and Roters 2002, Sarkar 2002, Genovese and Wasserman 2002), but easiest is augmentation approach of van der Laan et al. (2004).
- Begin with FWER-controlling procedure and reject k additional null hypotheses.

  Typically use the next k smallest p-values, but can split k in any way – by slice, anatomical region, ... whatever.
- But wait, shouldn't we keep k small relative to the total number of rejections (e.g., active voxels)?

  Ah, that's where False Discovery Control comes in.

The Benjamini-Hochberg (BH) Procedure

- Sort the m p-values $0 \equiv P_{(0)} < P_{(1)} < \cdots < P_{(m)}$.

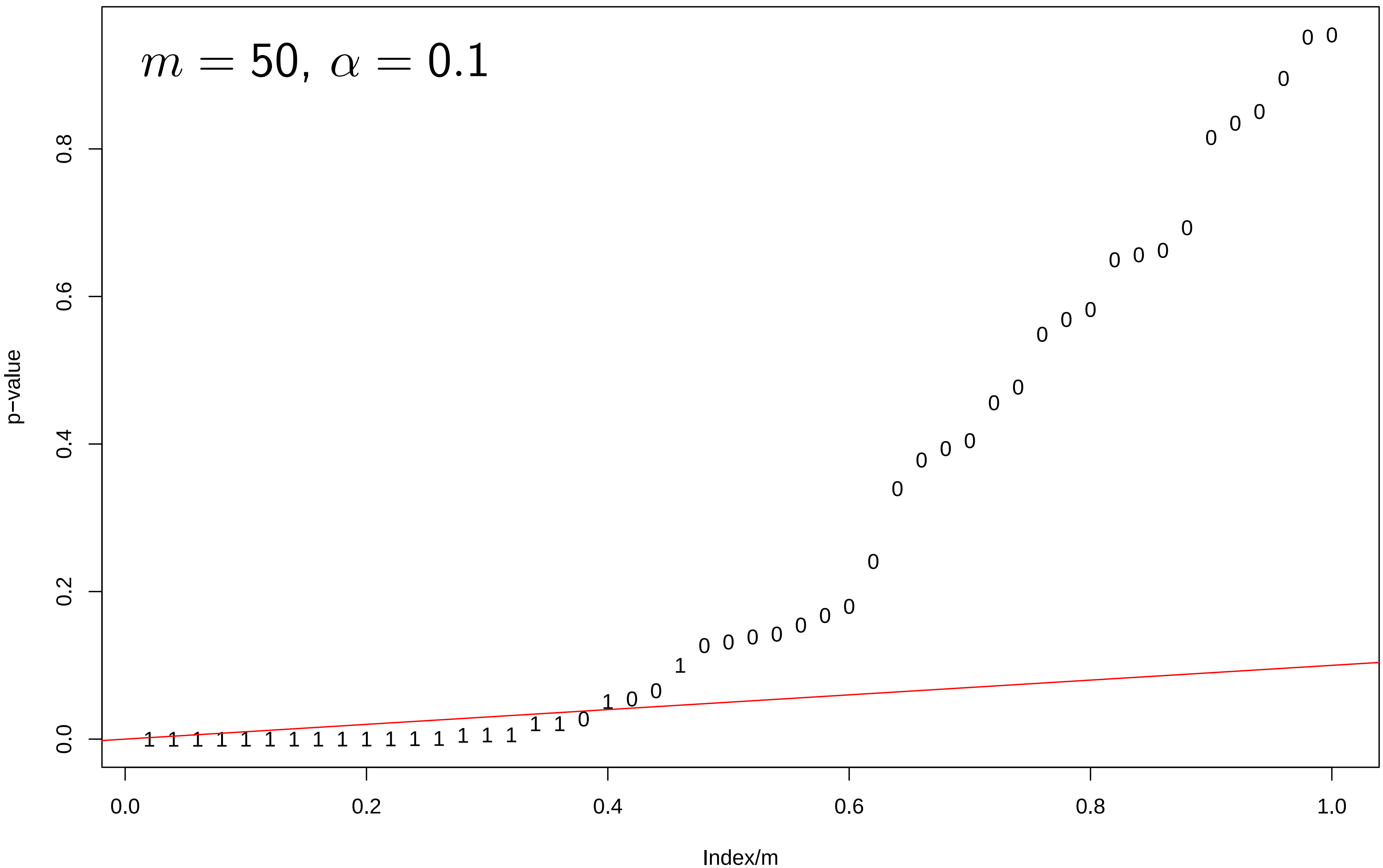
  The BH procedure has rejection threshold T_{BH} where

$$T_{\mathrm{BH}} = \max \left\{ P_{(i)} \colon \ P_{(i)} \leq \alpha \frac{i}{m},\ 0 \leq i \leq m \right\}.$$

  Almost trivial to compute.

- BH method guarantees that FDR $\leq \dfrac{T_0}{m}\,\alpha$.
- Proven for independent and "positive regression dependent" p-values. Hard to break in general. Distribution-free variant is simple but very conservative.
- Yields more power than Bonferroni (FWER control) and fewer Type I errors than uncorrected testing.
- BH method overcontrols FDR by factor T_0/m.

The Benjamini-Hochberg Procedure (cont'd)



Issues for fMRI

- Performance and Interpretation
  - How do we choose α? k?
  - How do we interpret the bounds? Do these bounds meet our needs?
  - How can we get more power?
- Dependence
  - Are fMRI p-values positive regression dependent? How do we test this?
  - How robust is the BH method to dependence assumptions?
- Smoothing
  - How should the data be smoothed?
- Spatial Structure
  - These methods ignore location information. How can we make regions rather than voxels the unit of inference?

Adaptive FDR Control

- If we knew the proportion of true nulls T_0/m, we could get $FDR = \alpha$ by using BH with $\beta = \alpha/(T_0/m)$.
- Instead: plug in an estimate. Several approaches developed (Benjamini and Hochberg 2000, Storey 2003, Genovese and Wasserman 2003, Benjamini et al. 2004).
- Easiest is Benjamini et al. (2004):

A. Apply BH procedure at level α. If nothing rejected, STOP.

B. Define $\mu_j = (m + 1 - j)/(1 - P_{(j)})$ for $1 \leq j \leq m$.

C. Let $k_* = \min\{j \geq 2 : \mu_j > \mu_{j-1}\}$.

D. Let $\widehat{T}_0/m = \min\left(1, \lceil \mu_{k_*} \rceil / m\right)$.

E. Apply BH procedure at level $\beta = \alpha/(\widehat{T}_0/m)$.

Controlling False Discovery Exceedance

- When controlling FDR, it is tempting to use the results to bound the # of false discoveries, but $D \cdot \alpha$ need not bound FD.

- Goal: Find a procedure that keeps FD/D low with high probability.

- Useful to define the False Discovery Proportion (FDP) at threshold t by

$$\mathsf{FDP}(t) = \frac{FD(t)}{D(t) + 1\{D(t) = 0\}}.$$

- Note: $\mathsf{FDR}(t) = \mathsf{E}(\mathsf{FDP}(t))$

Confidence Thresholds

- In practice, it is useful to be able to make a confidence statement about FDP.

- A $(\alpha, 1-\gamma)$ *confidence threshold* T is a statistic that satisfies

$$\mathsf{P}\Big\{\mathsf{FDP}(T) \leq \alpha\Big\} \geq 1-\gamma.$$

  This implies that FDR $\leq \alpha(1-\gamma)+\gamma$, though this is conservative.

- Can obtain such a procedure via a confidence envelope (Genovese and Wasserman 2003, 2004), a "step-down" testing proceudre (Romano and Lehmann 2004), or by augmenting a FWER-controlling procedure (van der Laan et al. 2004).

Confidence Envelopes

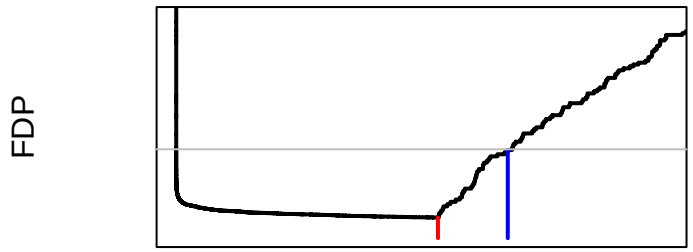
- A $1-\gamma$ confidence envelope for FDP is a random function $\overline{\mathrm{FDP}}(t)$ on $[0,1]$ such that

$$\mathrm{P}\Big\{\mathrm{FDP}(t) \le \overline{\mathrm{FDP}}(t) \text{ for all } t\Big\} \ge 1-\gamma.$$

- Given such an envelope, we can construct confidence thresholds.

  Two special cases have proved useful.

  - *Fixed-ceiling:* $T = \sup\{t\colon\ \overline{\mathrm{FDP}}(t) \le \alpha\}$.
  - *Minimum-envelope:* $T = \sup\{t\colon\ \overline{\mathrm{FDP}}(t) = \min_t \overline{\mathrm{FDP}}(t)\}$.



Exact Confidence Envelopes

- Let $\overline{\text{FDP}}$ maximize FDP over all subsets that look Uniform according to a hypothesis test.
- Can compute in linear time for suitable uniformity tests.
- Traditional uniformity tests have little power in the direction needed here.
- In contrast, using the kth order statistic as a one-sided test statistic — the $P_{(k)}$ test — has what we need:
  - For small k, sensitive to departures that have a large impact on FDP. Good "power."
  - Computing the confidence envelopes is linear in m.
  - Can choose k automatically.

Data Example

- Monkeys exhibit *visual remapping* in parietal cortex

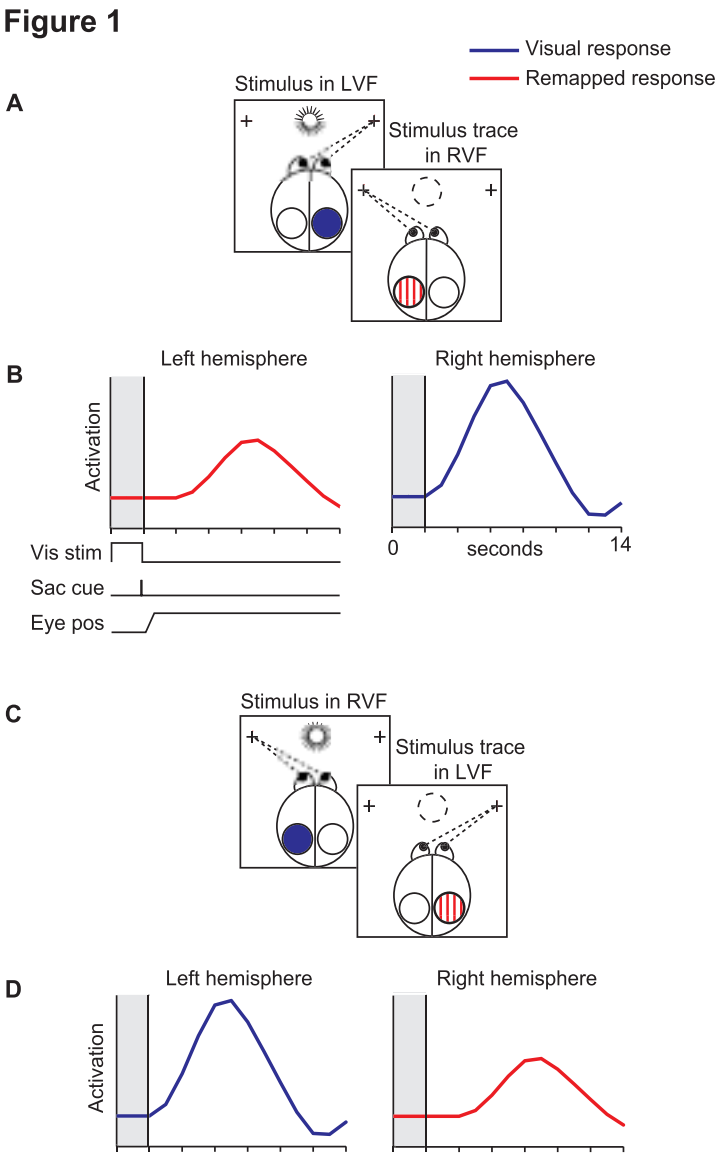
  When the eyes move so that the receptive field of a neuron lands on a previously stimulated location, the neuron fires even though no stimulus is present.

  Implies transformation in neural representation with eye movements. (Duhamel et al. 1992)

- Seek evidence for remapping in human cortex.

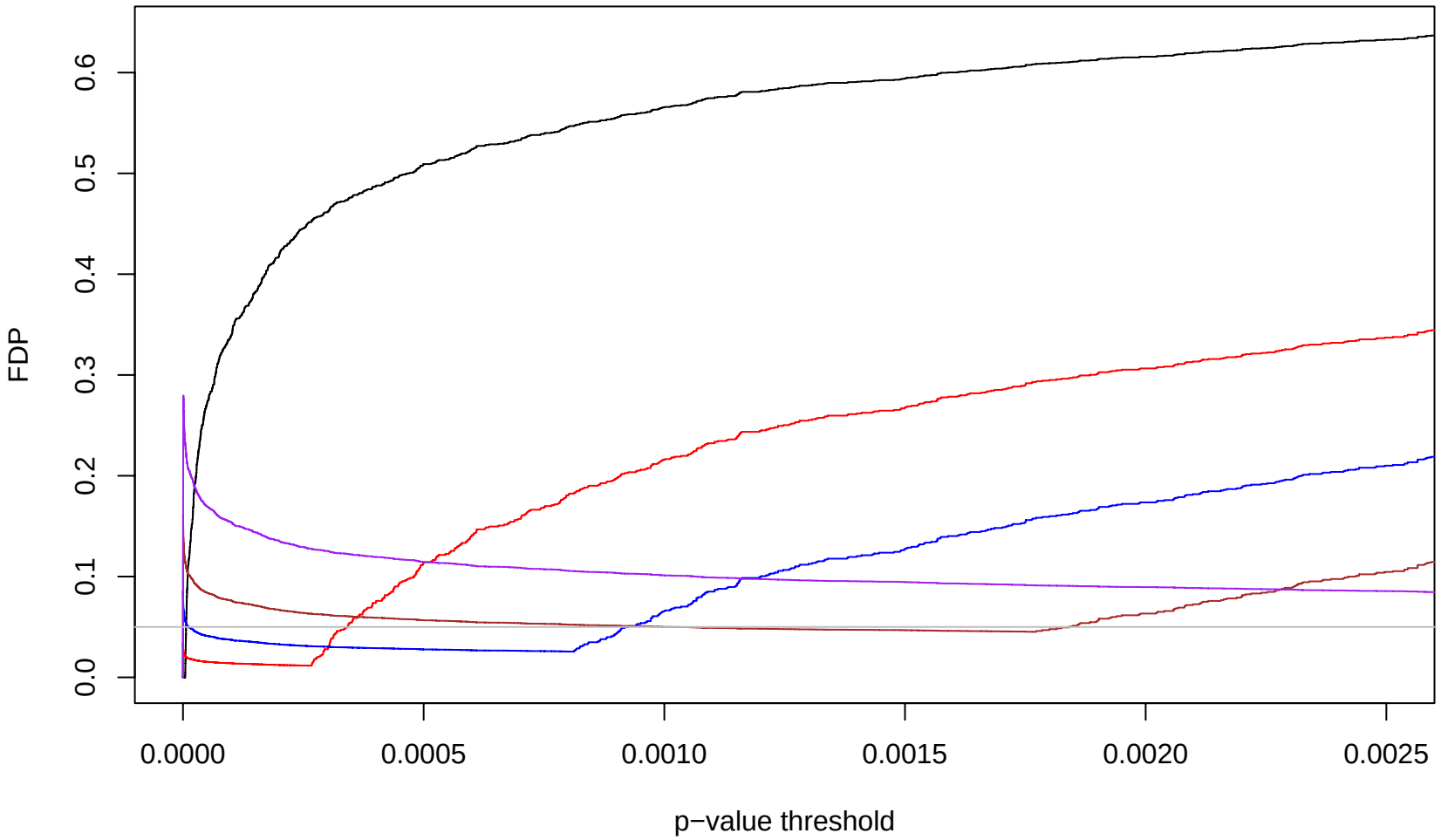
- See Merriam, Genovese, and Colby (2003). *Neuron*, 39, 361–373 for more details.

- EPI-RT acquisition, TR 2s, TE 30ms, 20 oblique slices, 3.125mm $\times$ 3.125mm $\times$ 3mm voxels.



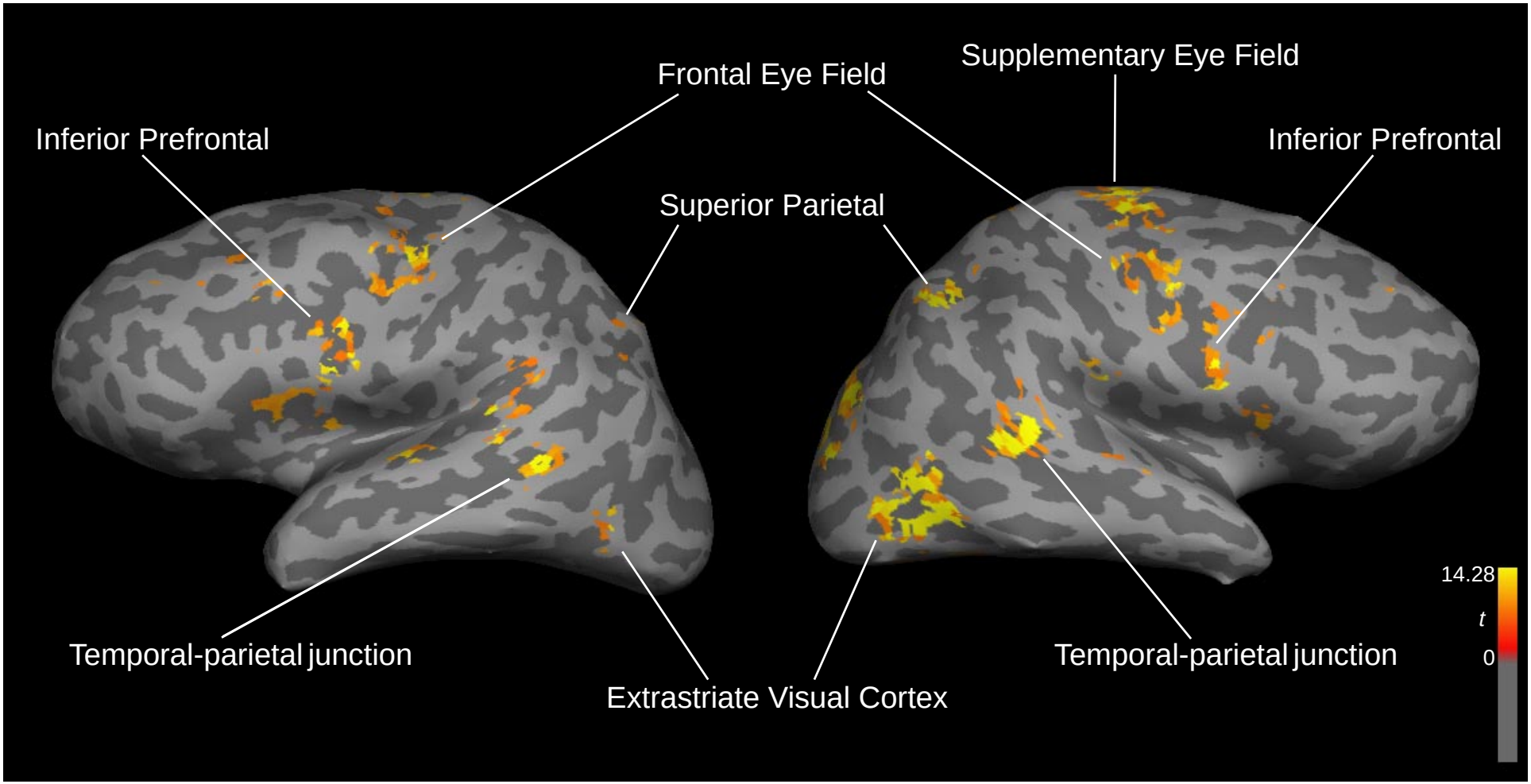
Results: $P_{(k)}$ 90% Confidence Envelopes

For $k = 1, 10, 25, 50, 100$, with 0.05 FDP level marked.



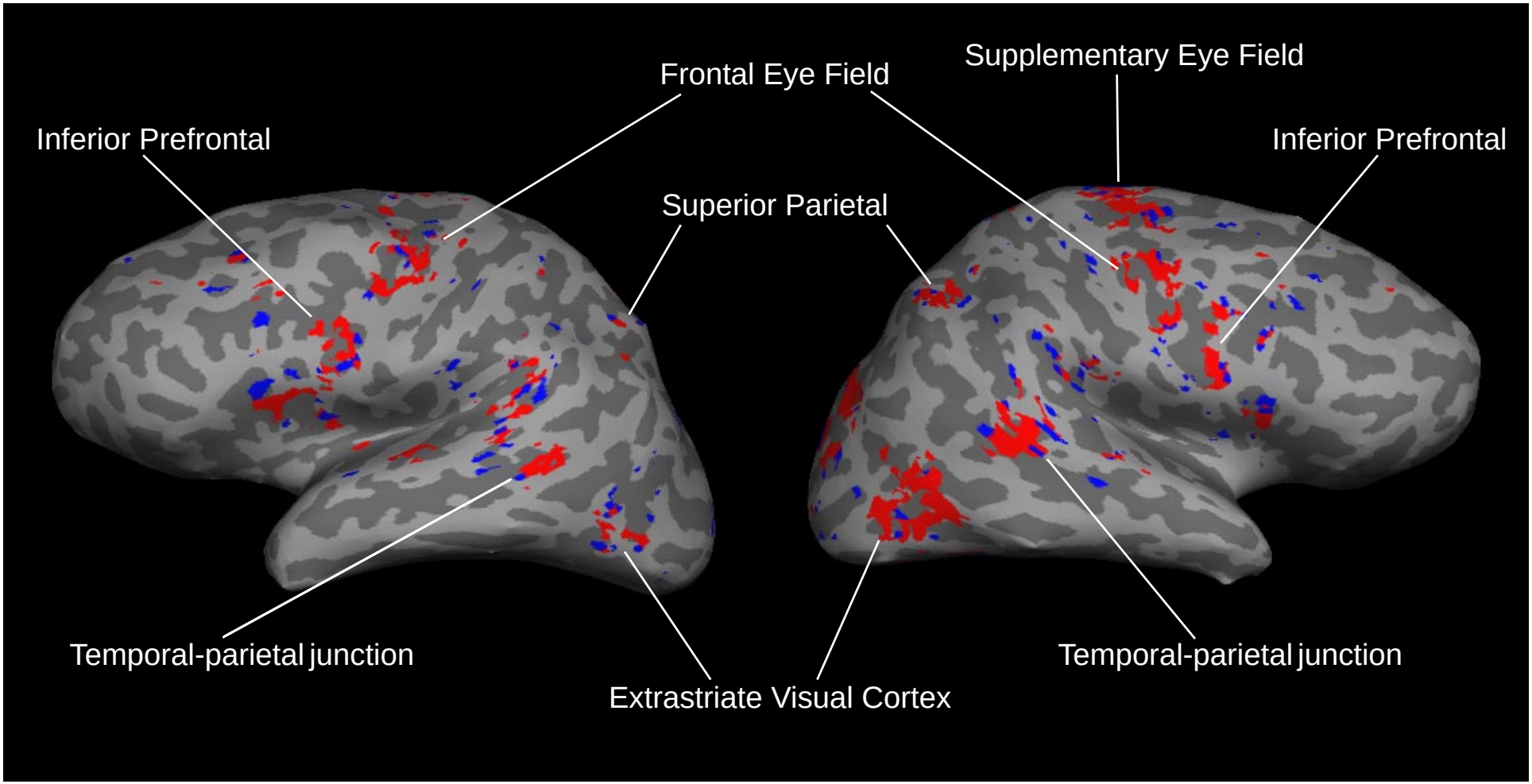
Results: Confidence Threshold

FDP bound: $\alpha = 0.05$ and Confidence Level: $1 - \gamma = 0.9$



Results: Confidence Threshold versus BH

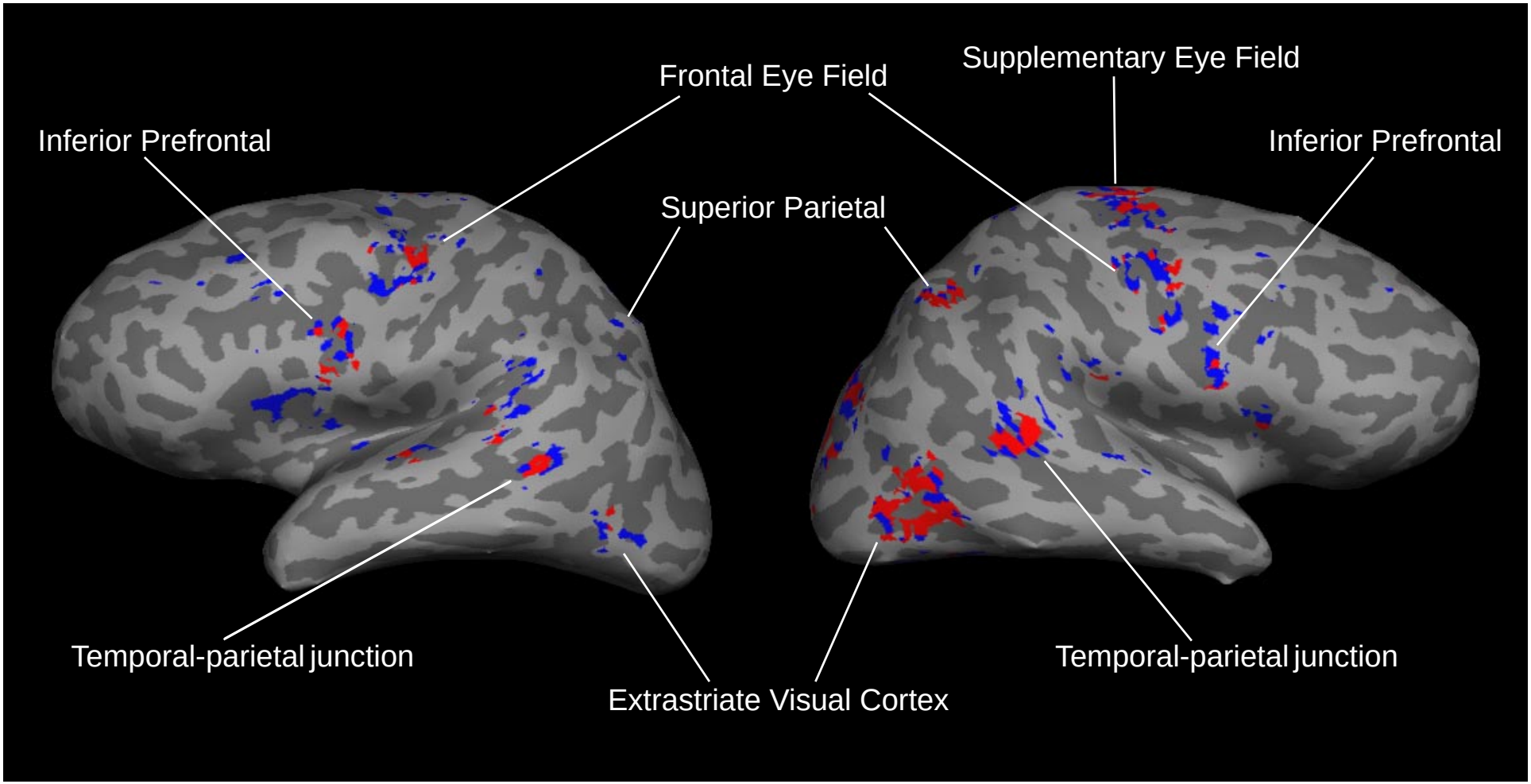
FDP bound: $\alpha = 0.05$ and Confidence Level: $1 - \gamma = 0.9$



Red = BH + Confidence Blue = BH only

Results: Confidence Threshold versus Bonferroni

FDP bound: $\alpha = 0.05$ and Confidence Level: $1 - \gamma = 0.9$



Red = Confidence + Bonferroni Blue = Confidence only

False Discovery Control for Random Fields

- Multiple testing methods based on the excursions of random fields are widely used in functional neuroimaging (e.g., Worsley 2003, Worsley et al. 2002, Cao and Worsley 2001).

- False Discovery Control extends to this setting under the same basic assumptions (Pacifico Perone, Genovese, et al. 2004).

- For a set S and a random field $X = \{X(s)\colon\ s \in S\}$ with mean function $\mu(s)$, use the realized value of X to test the collection of one-sided hypotheses

$$H_{0,s} : \mu(s) = 0 \text{ versus } H_{1,s} : \mu(s) > 0.$$

Let $S_0 = \{s \in S :\ \mu(s) = 0\}$.

False Discovery Control for Random Fields

- Define a spatial version of FDP by

$$\text{FDP}(t) = \frac{\text{Volume}(S_0 \cap \{s \in S : X(s) \geq t\})}{\text{Volume}(\{s \in S : X(s) \geq t\})}.$$

- Approach analogous to confidence thresholds:

  i. Find a confidence superset U such that $\text{P}\{U \supset S_0\} \geq 1 - \gamma$.

  ii. Compute a confidence envelope for FDP:

$$\overline{\text{FDP}}(t) = \frac{\text{Volume}(U \cap \{s \in S : X(s) > t\})}{\text{Volume}(\{s \in S : X(s) > t\})},$$

  iii. Select threshold from confidence envelope.

- With appropriate choice of α (FDP bound) and $1 - \gamma$ (confidence level), can control FDR or FDP exceedance.

Controlling the Proportion of False Regions

- We can use this to make inferences at the region level. (See also Taylor 2004 for another good idea on this.)
- Say a region R is false at tolerance ϵ if more than an ϵ proportion of its area is in null set S_0:

$$\text{error}(R) = \frac{\lambda(R \cap S_0)}{\lambda(R)} \geq \epsilon.$$

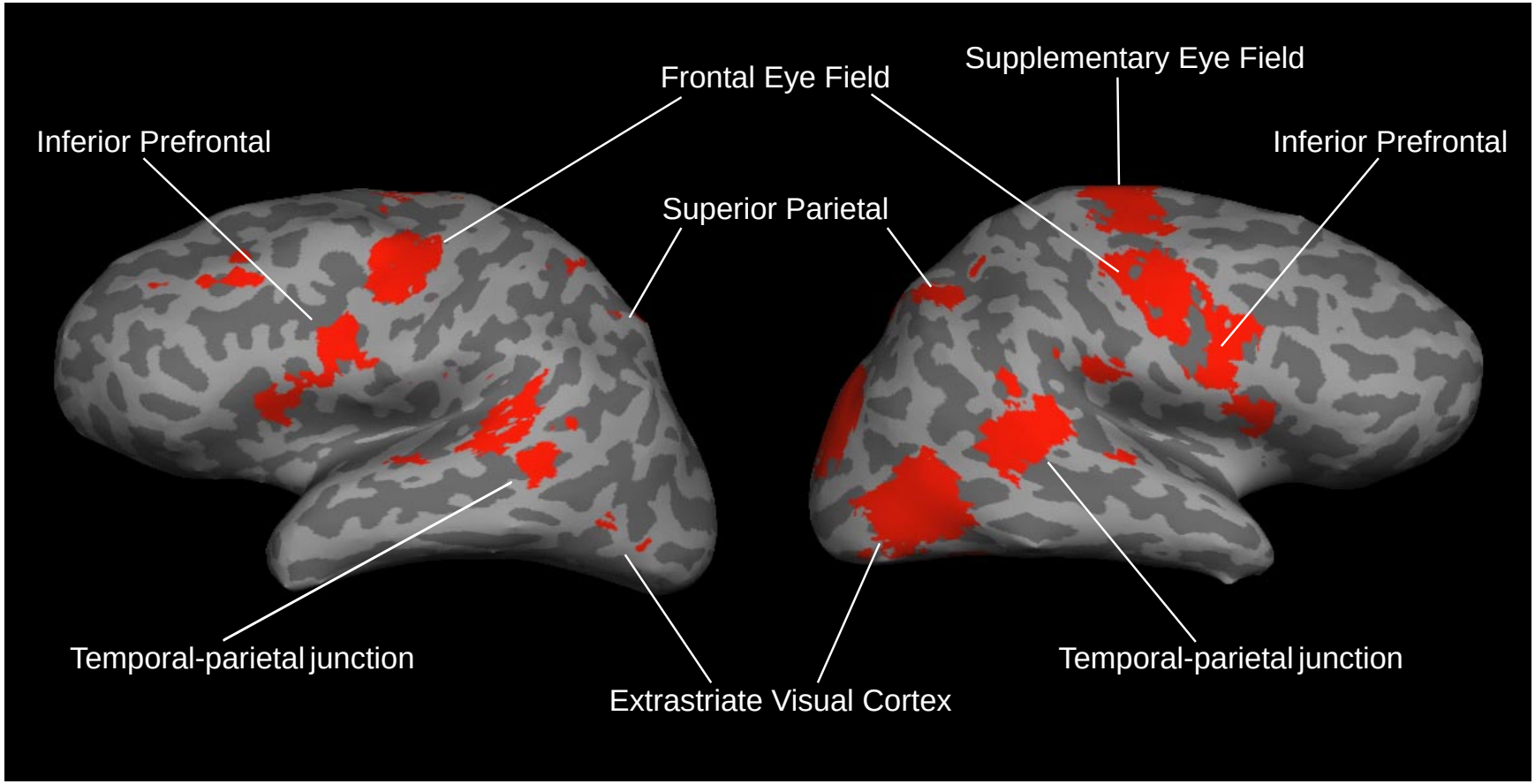
- Decompose the t-level set of X into its connected components $C_{t1}, \ldots, C_{tk_t}$.
- For each level t, let $\xi(t)$ denote the proportion of false regions (at tolerance ϵ) out of k_t regions.
- Then,

$$\overline{\xi}(t) = \frac{\#\left\{1 \leq i \leq k_t : \ \frac{\lambda(C_{ti} \cap U)}{\lambda(C_{ti})} \geq \epsilon\right\}}{k_t}$$

gives a $1 - \gamma$ confidence envelope for ξ.

Results: False Region Control Threshold

$\mathrm{P}\{\text{prop'n false regions} \leq 0.1\} \geq 0.95$ where false means null overlap $\geq 10\%$



Take Home Points

- Though multiple testing is only one among many problems in fMRI, recent research has produced a variety of useful error criteria and efficient procedures.
- Confidence thresholds have advantages for False Discovery Control.

  In particular, we gain a stronger inferential guarantee with little effective loss of power.
- Dependence tends to make all these procedures more conservative, but the bounds are still valid for practically relevant dependence.
- False Discovery Control procedures are available for smoothed or unsmoothed data.
- For spatial applications, we care about clusters/regions/areas not voxels. Current methods are extending to such region-based inferences.