

GROUP ANALYSIS

AGGREGATING MULTIPLE SUBJECTS

- When we conduct multi-subject analysis we are trying to understand whether an effect is “significant” across a group of people.
- Whether something is significant depends on the variance we assess it against:

Classical statistical hypothesis testing proceeds by comparing the difference between the expected and hypothesized effect against the “yardstick” of variance.

[Holmes & Friston, 1998]



VARIANCE AT THE GROUP LEVEL

- ***Fixed Effects (FFX)***: is about the intra-subject variability. An effect is compared against the “yardstick” of the precision with which it can be measured (for each subject). The different subjects are considered to be “fixed.”
- ***Random Effects (RFX)***: is about the inter-subject variability. An effect is compared against the “yardstick” of how much variability there is across different subjects. The different subjects are considered to be a “random” sample from a greater population.
- ***Mixed Effects (MFX)***: is about intra-subject & inter-subject variability.



FIXED EFFECTS: INTRA-SUBJECT VARIABILITY

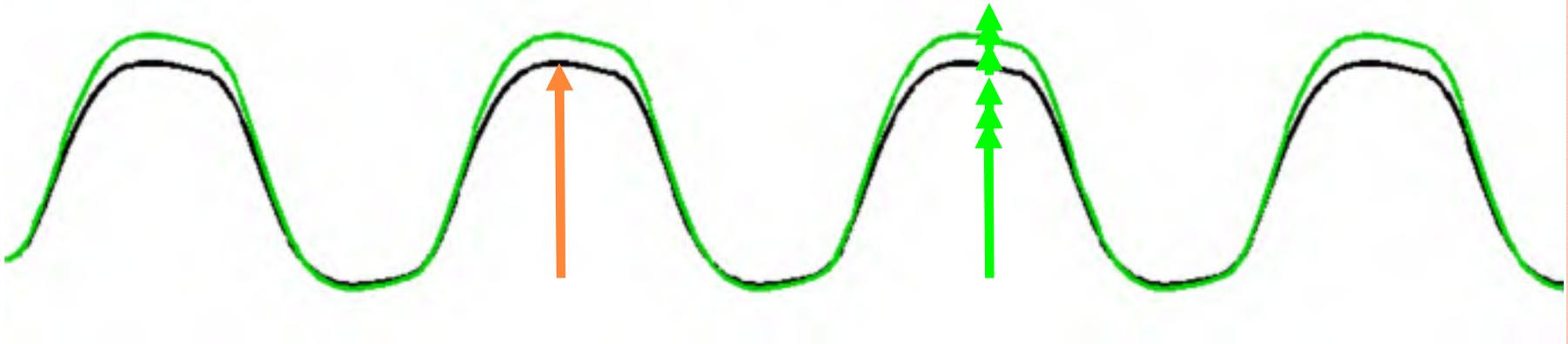


- Only variation (over sessions) is measurement error
- True Response magnitude is *fixed*



Adapted from T Nichols

RANDOM EFFECTS: INTER-SUBJECT VARIABILITY

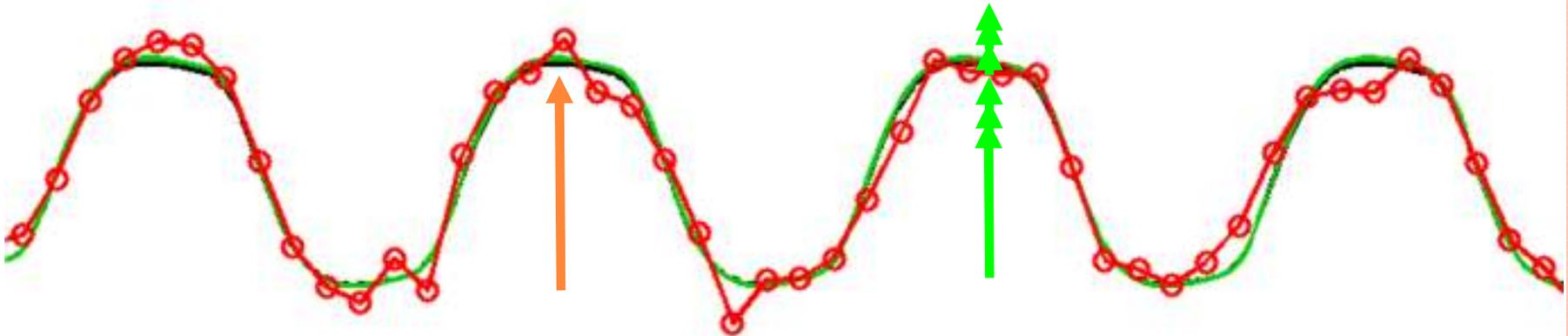


- Source of variation
 - Response magnitude
- Response magnitude is *random*
 - Each subject/session has random magnitude
 - But note, the population mean is *fixed*



Adapted from T Nichols

MIXED EFFECTS



- Two sources of variation
 - Measurement error
 - Response magnitude
- Response magnitude is *random*
 - Each subject/session has random magnitude
 - But note, the population mean is *fixed*



IN OTHER WORDS ...

FFX Model:

$$y_{ij} = d_i + \varepsilon_{ij}$$

Subj. effect Meas. error

$$\varepsilon_{ij} \sim (0, \sigma_w^2)$$



IN OTHER WORDS ...

FFX Model:

$$y_{ij} = d_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim (0, \sigma_w^2)$$

But d_i is a random variable!

$$d_i = d_{pop} + z_i$$

$$z_i \sim (0, \sigma_b^2)$$

Population
effect

Subj. variability
(around d_{pop})



IN OTHER WORDS ...

FFX Model:

$$y_{ij} = d_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim (0, \sigma_w^2)$$

But d_i is a random variable!

$$d_i = d_{pop} + z_i$$

$$z_i \sim (0, \sigma_b^2)$$

MFJ Model:

$$y_{ij} = d_{pop} + z_i + \varepsilon_{ij}$$

Population
effect

Subj. variability
(around d_{pop})

Meas. error



A HAIRY EXAMPLE

Question: Do M & F hair differ in length?

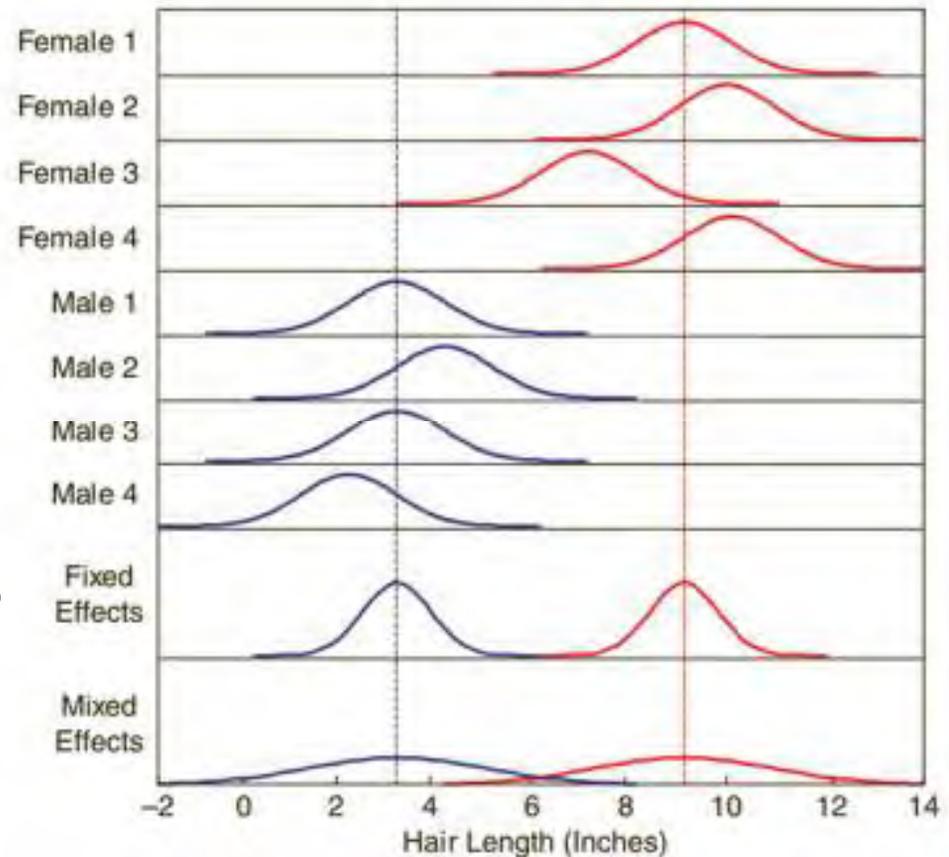
Experiment: Take 25 hairs from each of 8 Ss (4F, 4M)

$$[\sigma_w^2=1, \sigma_b^2=49]$$

$$\sigma_{FFX}^2: \frac{\sigma_w^2}{Nn} = \frac{1}{(4 \times 25)} = 0.01$$

$$\sigma_{RFX}^2: \frac{\sigma_b^2}{N} = \frac{49}{4} = 12.25$$

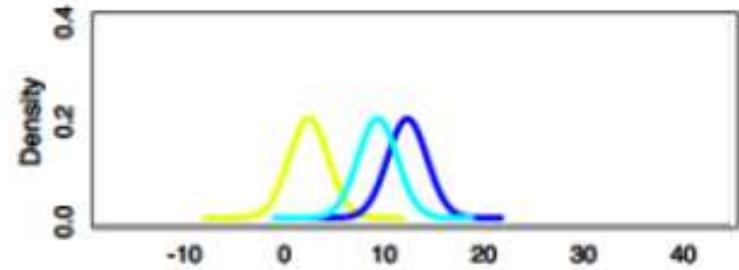
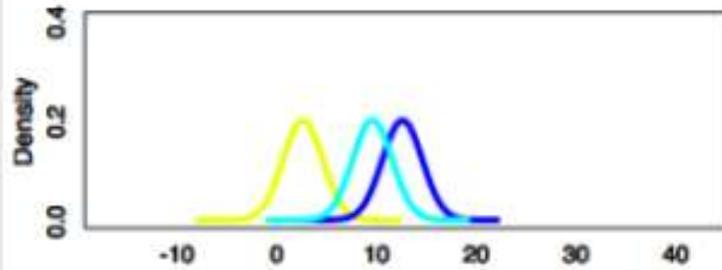
$$\sigma_{MFX}^2: \frac{\sigma_w^2}{Nn} + \frac{\sigma_b^2}{N} = \frac{49}{4} = 12.26$$



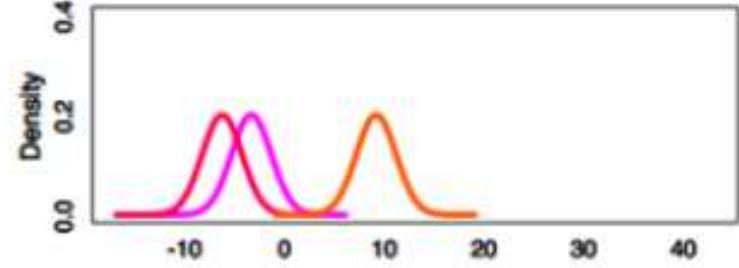
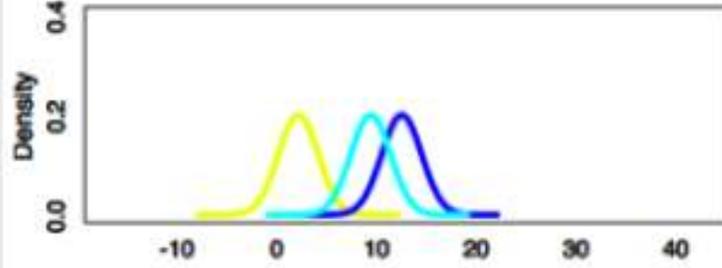
Fixed

Mixed

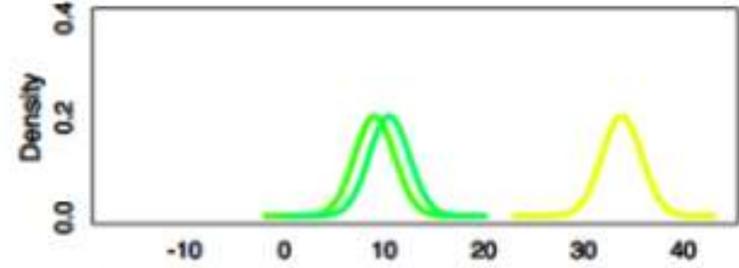
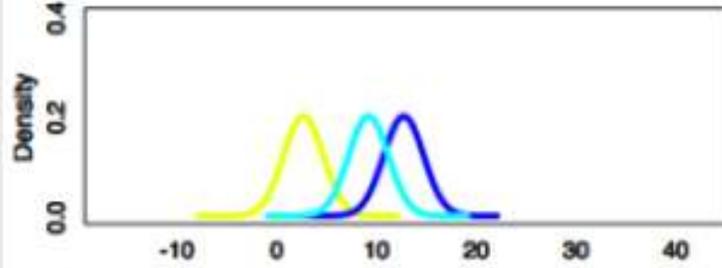
Sample 1



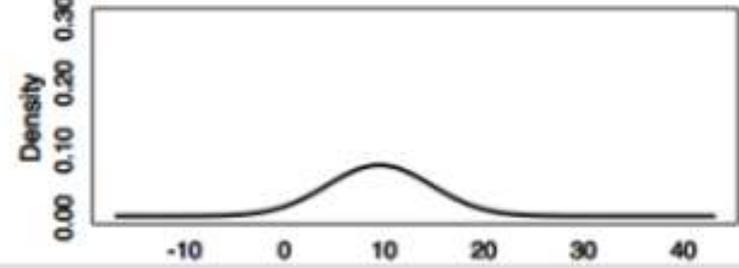
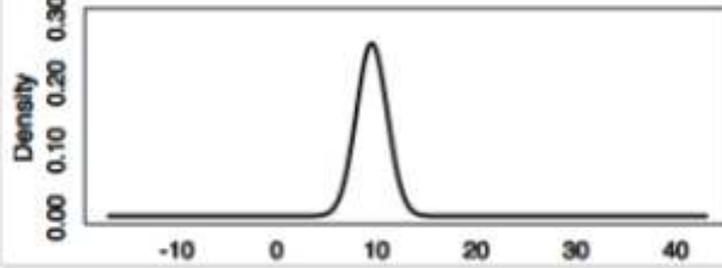
Sample 2



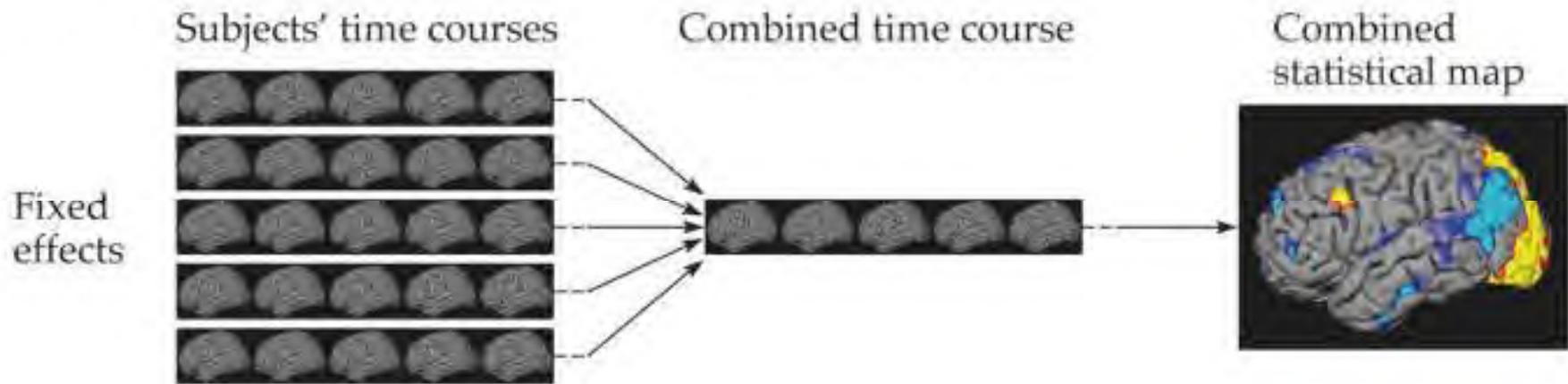
Sample 3



Group
Distribution



GROUP ANALYSIS STRATEGIES: FFX



$$\begin{pmatrix} Y_{1,1} \\ \vdots \\ Y_{1,20} \\ Y_{2,1} \\ \vdots \\ Y_{3,20} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_g + \begin{pmatrix} \epsilon_{1,1} \\ \epsilon_{1,2} \\ \epsilon_{1,3} \\ \epsilon_{2,1} \\ \vdots \\ \epsilon_{N,3} \end{pmatrix}, \quad \epsilon_{i,j} \sim N(0, \sigma_{win}^2)$$

Fixed effect

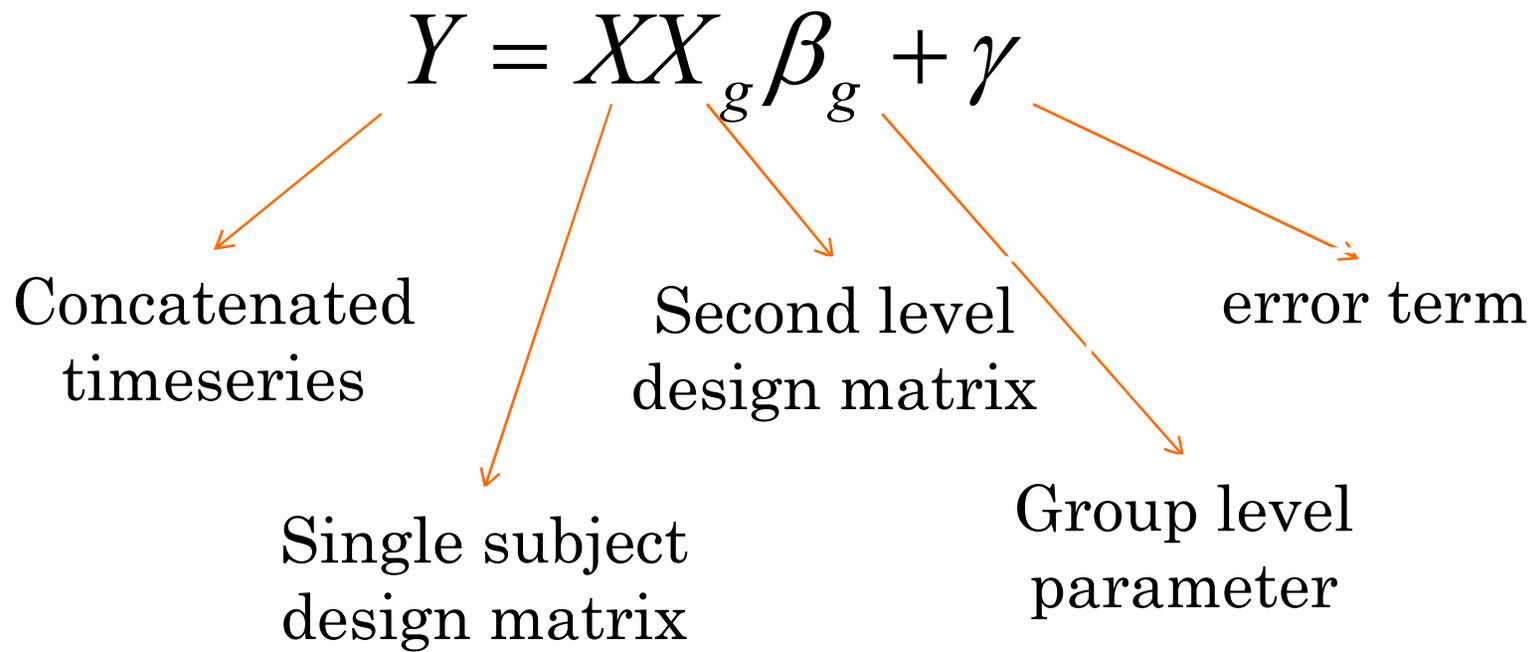
Residual error

FIXED V RANDOM

- Fixed isn't "wrong," just usually isn't of interest
- Fixed Effects Inference
 - "I can see this effect in this cohort"
- Random Effects Inference
 - "I can extend my inference to the population"

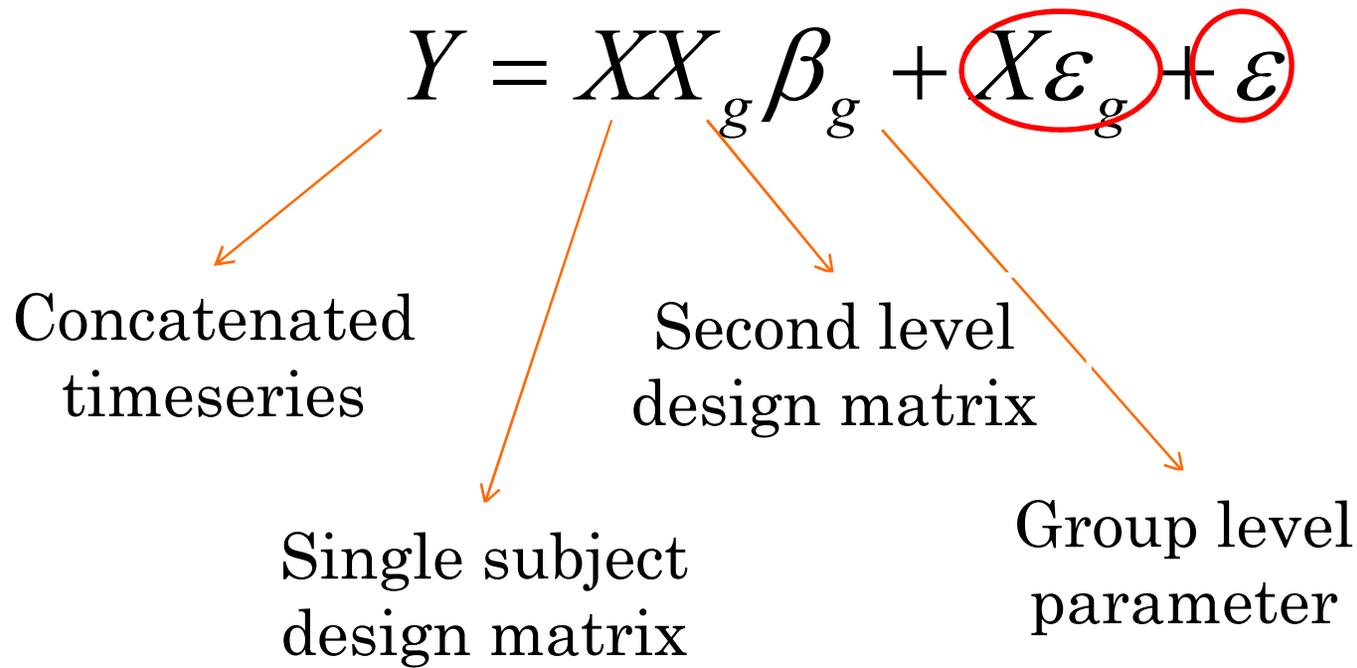
GROUP ANALYSIS STRATEGIES (I): “ALL-IN-ONE”

- Complete single-level GLM that relates various parameters of interest at the group level to the full set of (time series) data available

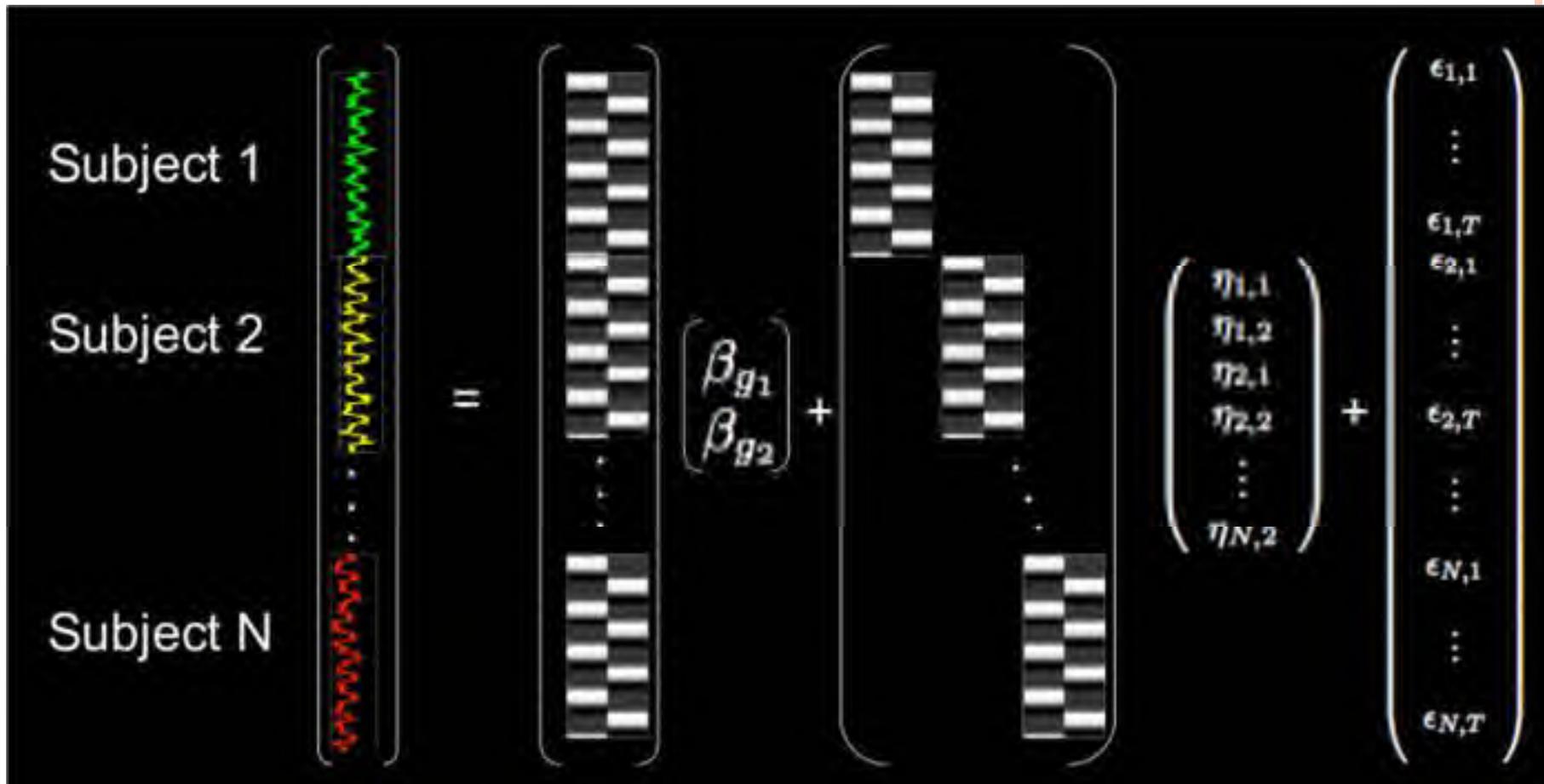


GROUP ANALYSIS STRATEGIES (I): “ALL-IN-ONE”

- Complete single-level GLM that relates various parameters of interest at the group level to the full set of (time series) data available



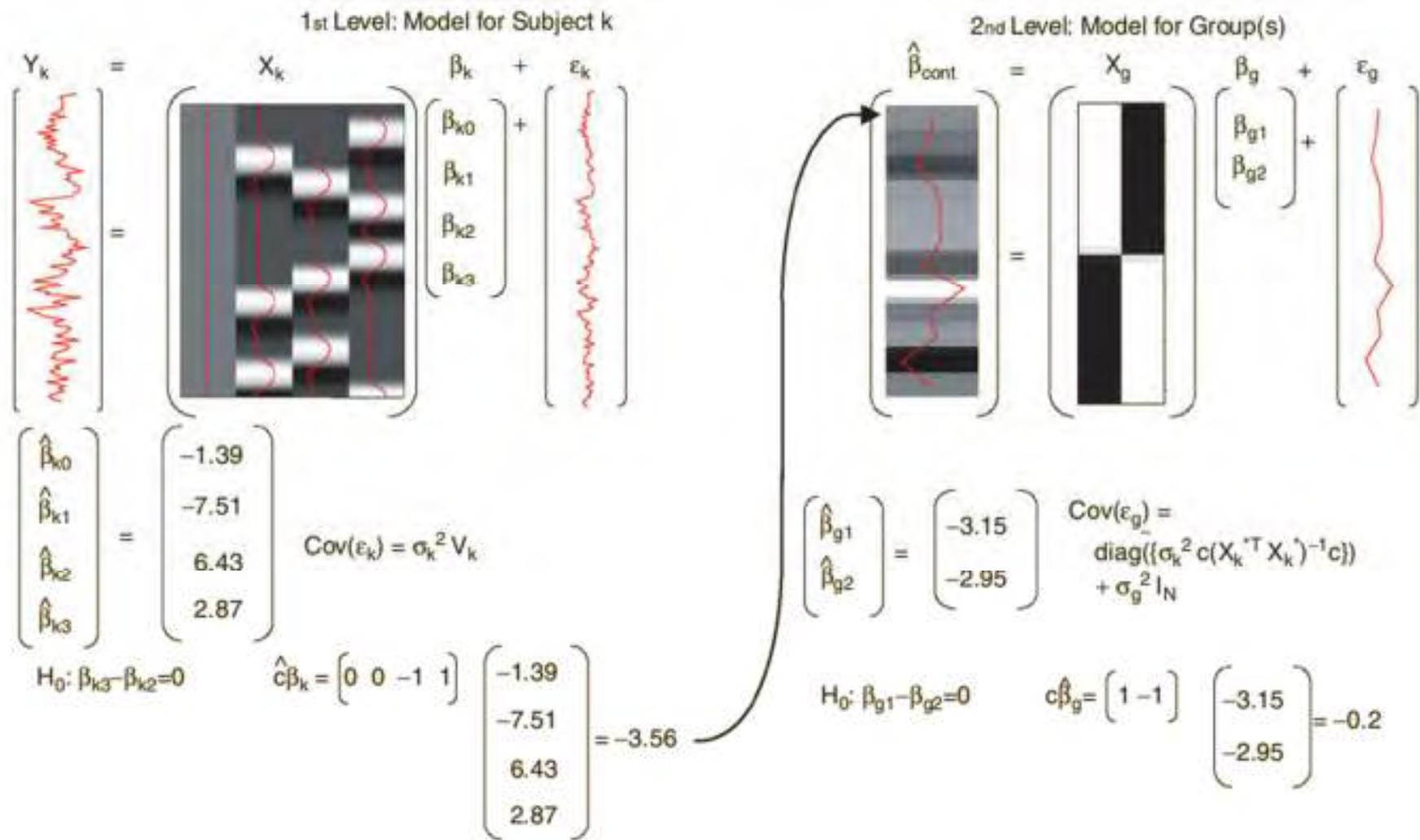
GROUP ANALYSIS STRATEGIES (I): “ALL-IN-ONE”



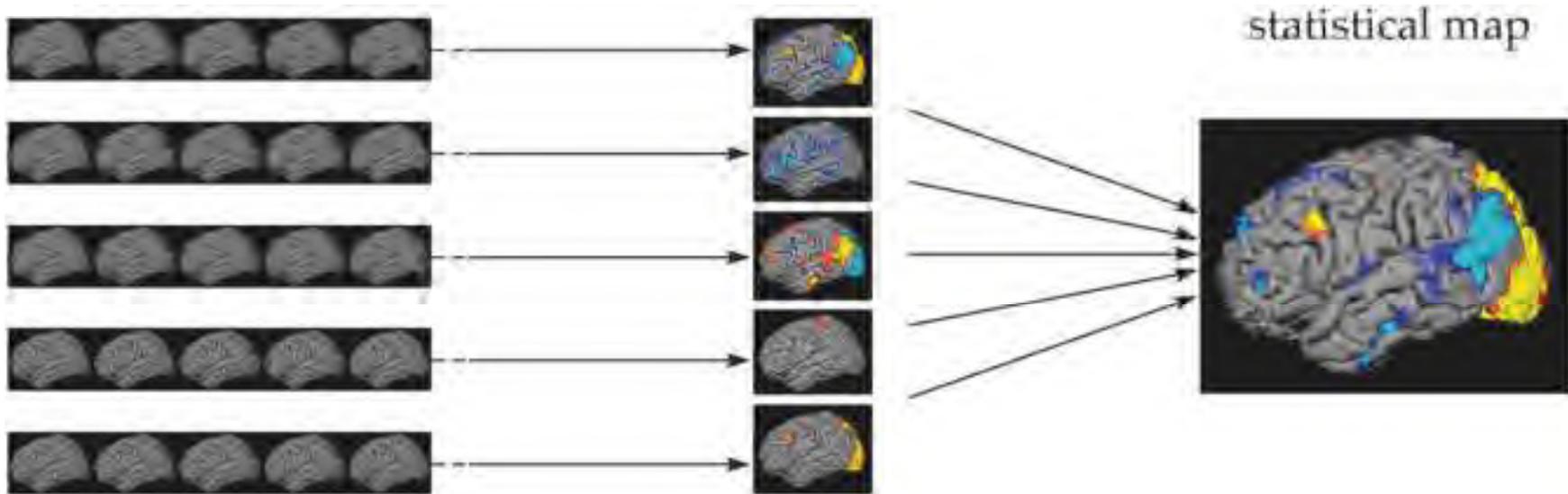
- Computationally intense approach
- What if you acquire 1 more dataset?



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH



FUNCTIONAL MAGNETIC RESONANCE IMAGING, Figure 12.20 © 2004 Sinauer Associates, Inc.



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

$$\begin{cases} Y = X\beta + \varepsilon & [1^{st} \text{ lvl } FFX] \\ \beta = X_g \beta_g + \varepsilon_g & [2^{st} \text{ lvl } RFX] \end{cases}$$

$$\hookrightarrow Y = XX_g \beta_g + X\varepsilon_g + \varepsilon$$



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

$$\begin{cases} Y = X\beta + \varepsilon & [1^{st} \text{ lvl } FFX] \\ \beta = X_g \beta_g + \varepsilon_g & [2^{st} \text{ lvl } RFX] \end{cases}$$

$$\begin{cases} Y = X\beta + \varepsilon & [1^{st} \text{ lvl } FFX] \\ \hat{\beta} = X_g \beta_g + \varepsilon_g^* & [2^{st} \text{ lvl } MFX] \end{cases}$$

$$\varepsilon_g^* = \frac{\sigma_w^2}{n} + \sigma_b^2$$



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

SPM (I): Assume homoscedastic 1st level variances and do an OLS

To maintain equivalence with all-in-one model

1. first level variances must be equal (σ_w^2)
2. First level X matrices must be the same (i.e., “balanced” for all subjects)



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

SPM (I): Assume homoscedastic 1st level variances and do an OLS.
FSL: Carry forward c/ β estimates *and* covariance matrix. (i.e., do a GLS)

$$RFX : \beta = X_g \beta_g + \varepsilon_g \quad \varepsilon_g \sim (0, \sigma_b^2 I)$$



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

SPM (I): Assume homoscedastic 1st level variances and do an OLS.
FSL: Carry forward c/ β estimates *and* covariance matrix. (i.e., do a GLS)

$$RFX : \beta = X_g \beta_g + \varepsilon_g \quad \varepsilon_g \sim (0, \sigma_b^2 V_g)$$

$$V_g = \begin{pmatrix} \sigma_{win_1}^2 + \sigma_g^2 & & 0 \\ & \dots & \\ 0 & & \sigma_{win_N}^2 + \sigma_g^2 \end{pmatrix}$$

GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

SPM (I): Assume homoscedastic 1st level variances and do an OLS.

FSL: Carry forward c/β estimates *and* covariance matrix. (i.e., do a GLS)

SPM (II): Estimate the covariance structure (ReML) from first level (only significant voxels) and carry that forward.



GROUP ANALYSIS STRATEGIES (II): THE SUMMARY STATISTIC APPROACH

SPM (I): Assume homoscedastic 1st level variances and do an OLS.

FSL: Carry forward c/β estimates *and* covariance matrix. (i.e., do a GLS)

SPM (II): Estimate the covariance structure (ReML) from first level (only significant voxels) and carry that forward.

Data:

Beckmann 03 (FSL): must weight variances

Friston 05 (SPM): OLS is robust to unequal variances

Mumford 09: OLS is robust, but only for 1 sample t-test, GLS always more optimal strategy.



RECAP

- i. FFX inferences are valid, but only with respect to the sample. May be of interest for single case studies, or small rare populations you can fully sample.
 - ii. MFX inferences are valid over the population you sample from because you are accounting for sampling variability. This is what you want to do for a typical group study.
 - iii. The Summary statistic approach is efficient. Run 1st levels independently, then combine the results. If you run 1 more subject, then you only have to re-run the group.
- 

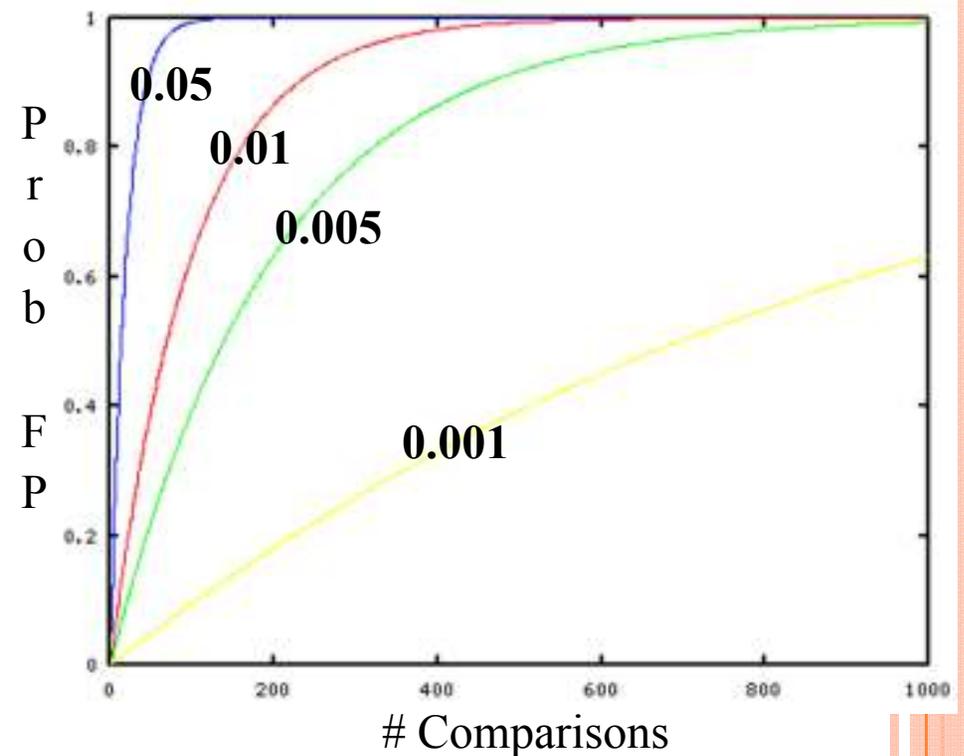
MULTIPLE COMPARISONS PROBLEM

- When you make 1 test, what is the probability that a positive result is, in fact, not true (i.e., false positive)
→ α (say, 5%)
- If we make 2 tests, what is the overall probability (i.e., ‘family-wise’ probability) of false positives?
→ $1-(1-\alpha)^2$ (at a nominal 5%: 9.75%)
- If we make n tests, what is the overall probability (i.e., ‘family-wise’ probability) of false positives?
→ $1-(1-\alpha)^n$

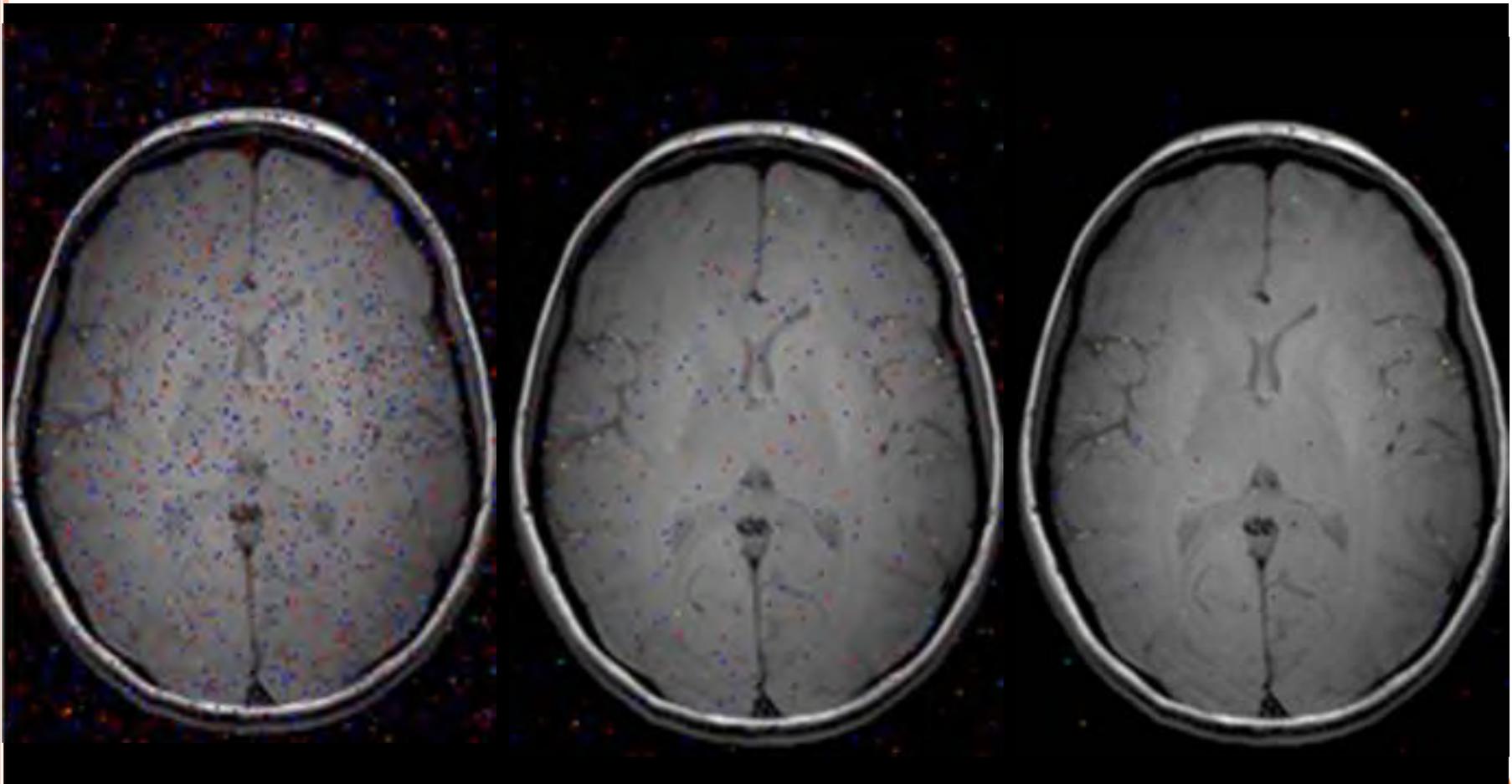


MULTIPLE COMPARISONS PROBLEM

- How many tests do we perform in fMRI analysis?
- Over (say) 100,000 null voxels, how many times will we incorrectly reject H_0 ?
- ~5,000 voxels (on average!)



FALSE ACTIVATIONS UNDER H0

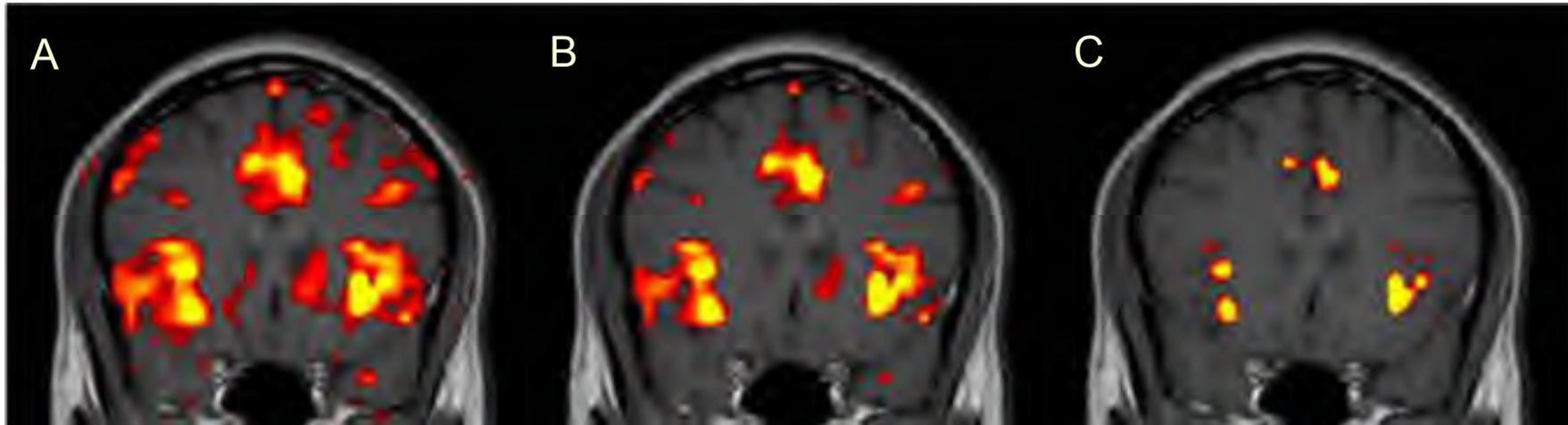


$P < 0.05$ (1682 voxels)

$P < 0.01$ (364 voxels)

$P < 0.001$ (32 voxels)

HOW MUCH CORRECTION?



A $t = 2.10, p < 0.05$ (uncorrected)

B $t = 3.60, p < 0.001$ (uncorrected)

C $t = 7.15, p < 0.05$,
Bonferroni Corrected

Poor Specificity

(risk of false positives)

Good Power

Good Specificity

Poor Power

(risk of false negatives)



LEVEL OF CORRECTION

- Voxel: There is a 5% chance that some voxel I see occurred by chance.
 - Highly specific (e.g., reject H_0 on each voxels), but conservative
- Cluster: There is a 5% chance that some cluster I see occurred by chance.
 - More sensitive, less (spatially) specific
- Set: There is a 5% chance that the number of clusters I see occurred by chance.



CORRECTION FOR MULTIPLE COMPARISONS

2 main strategies:

1. ***Family Wise Error (FWE)***: Control for the probability of ***any*** false positives (e.g., Bonferroni, Random Field Theory, Permutation)
2. ***False Discovery Rate (FDR)***: Control proportion of false positives ***among*** rejected tests



FWE (I): BONFERRONI

- Main idea: make each individual test more stringent, so overall you end up with your total (i.e., family-wise) ‘desired’ false positive rate.

$$\alpha_i^{Bonf} = \frac{\alpha_{FW}}{n} \rightarrow \sum_{i=1}^n P(T_i > \alpha_i | H_0) \leq \alpha_{FW}$$

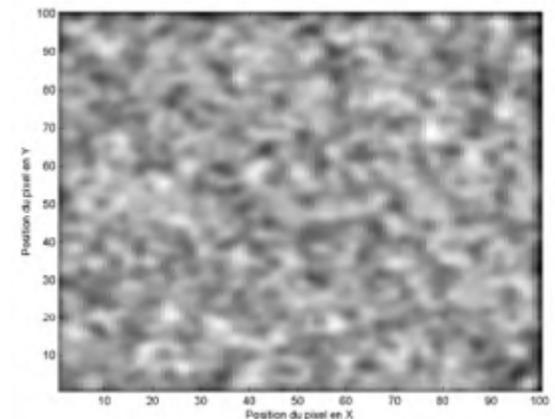
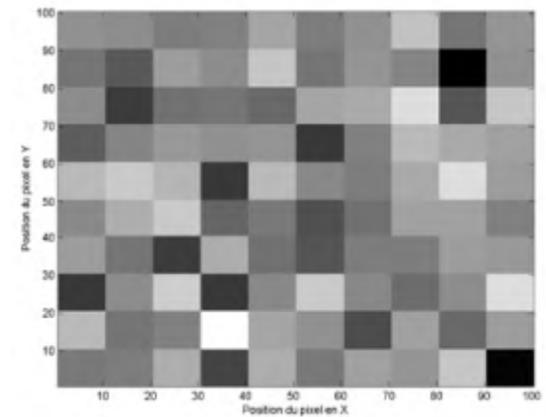
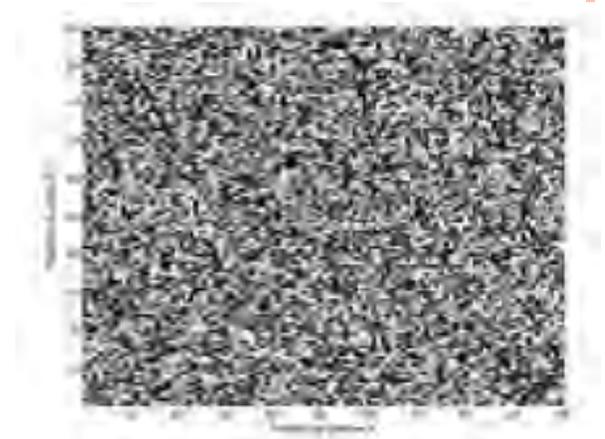
- For example:
 - Desired familywise false positive rate: $\alpha_{FW} = 0.05$
 - Total number of (independent) tests: 100,000
 - Then the Bonferroni-corrected false positive level *for each individual test* is:

$$\alpha_i^{Bonf} = \frac{\alpha_{FW}}{n} = \frac{0.05}{100,000} = 0.0000005$$



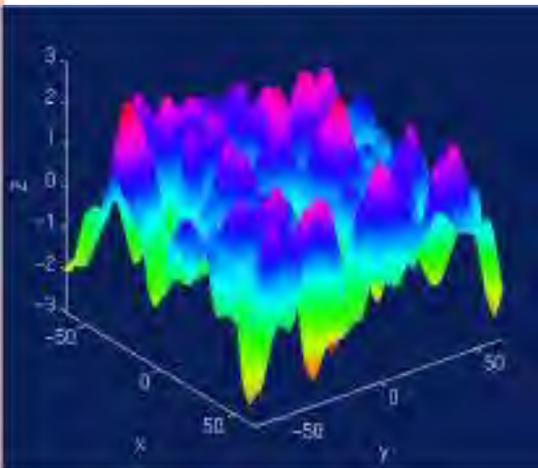
FWE (I): BONFERRONI

- Assumes independent tests
- FMRI data spatially correlated, so the number of independent tests is less than the number of voxels
 - Overly stringent
 - Increases Type II error
- Difficult to find what is n in order to calculate the correct α_{bonf}



FWE (II): RANDOM FIELD THEORY

- Allows to find a threshold in a set of data where it's not easy (or even impossible) to find the number of independent variables
- 3 step approach:
 - i. Estimate how smooth the data is (“resels”)
 - ii. Compute how many peaks would be above the threshold by chance (“Euler Characteristic”)
 - iii. Calculate the threshold that yields desired FWER



1. SMOOTHNESS PARAMETRIZATION

We can't compute the # of independent voxels, but we can compute the number of resolution elements (i.e. “resels”).

- RESELS – **R**esolution **E**lements

- 1 RESEL = $\text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z$

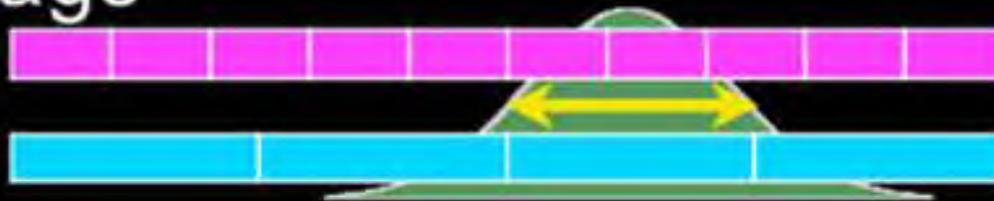
- RESEL Count R

- $R = V \sqrt{|\lambda|}$ ← *The only data-dependent part of $E(y_{in})$*

- Volume of search region in units of smoothness

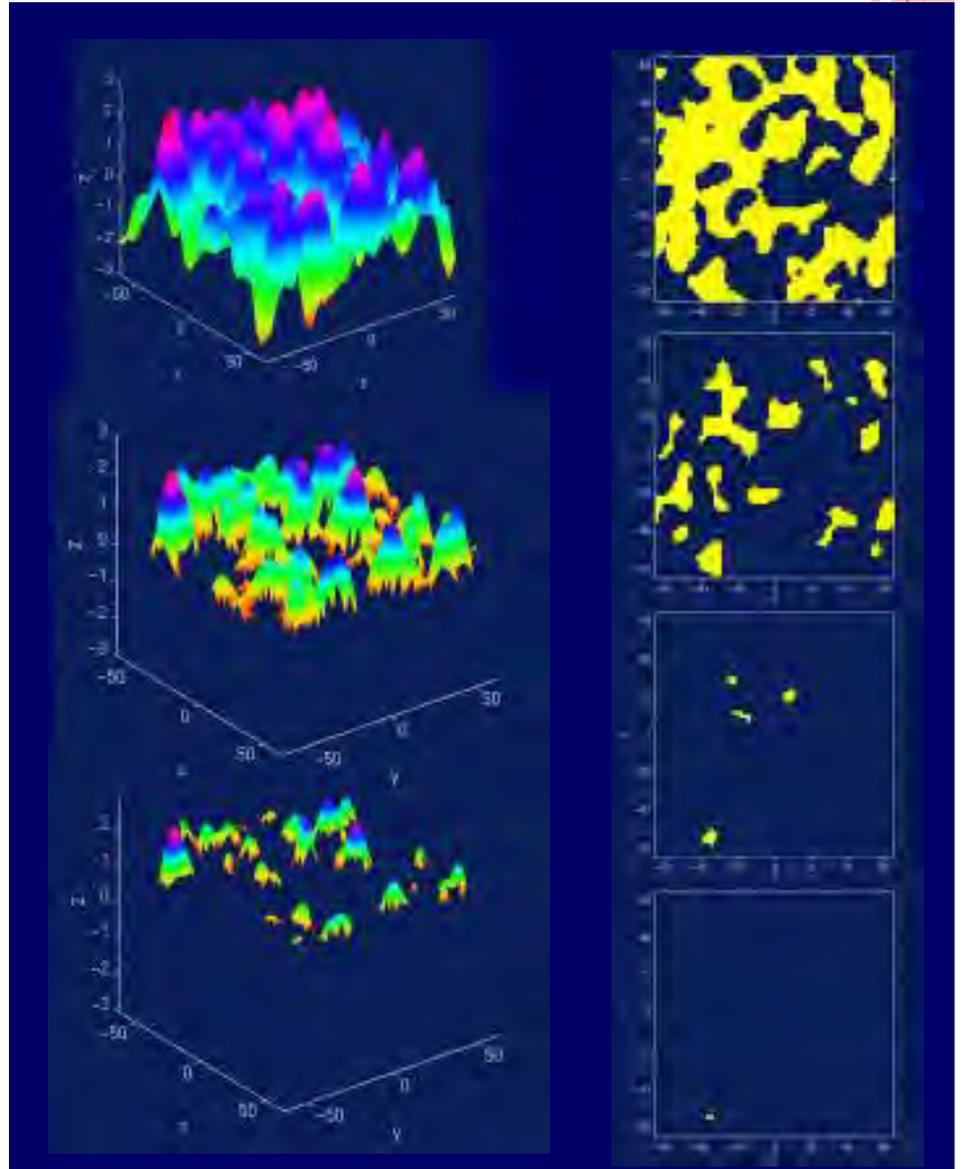
- Eg: 10 **voxels**, 2.5 voxel FWHM **smoothness**, 4 **RESELS**

- RESELS not # of independent ‘things’ in the image

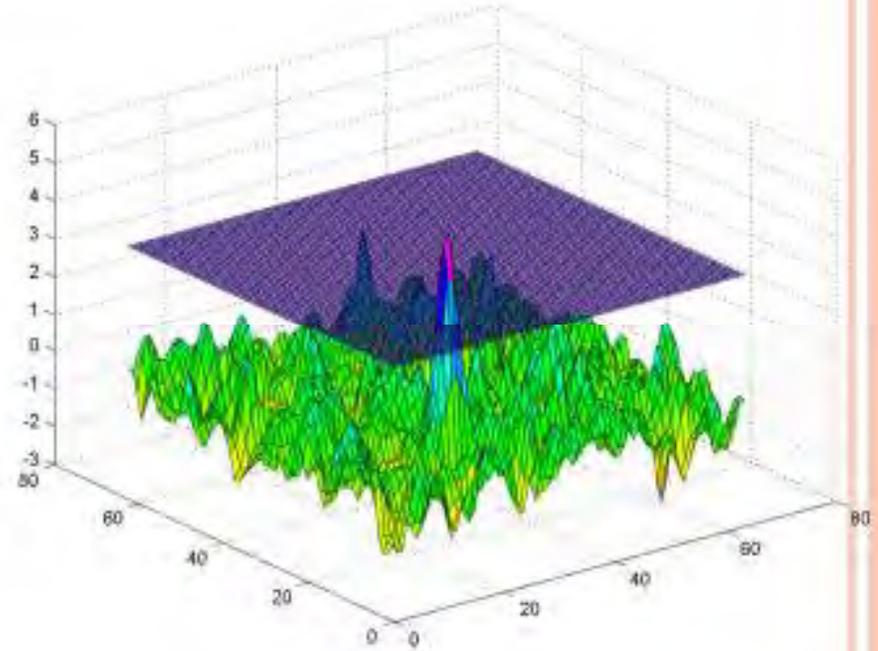
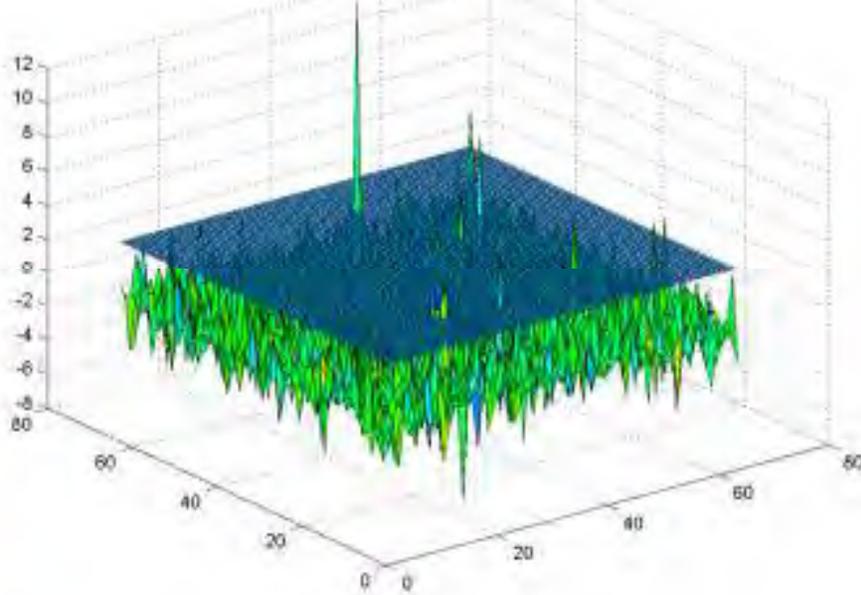


2. EULER CHARACTERISTIC

- Topological measure [χ]
- Threshold an image at u
- $EC = \#$ of blobs - $\#$ holes
- At high u :
 - $P(\text{blob}) = E[EC]$
- Under H_0 , $\alpha_{\text{FWE}} = E[EC]$



3. THRESHOLD



$$\alpha_{FW} = E[\chi] = R(4 \log_e 2)(2\pi)^{\frac{-3}{2}} Z e^{\frac{-1}{2} Z^2}$$

Given the smoothness of my data (R), what threshold (Z) do I need to set so that I have α_{FW} chance ($\sim E[EC]$) of having peak above threshold?



FALSE DISCOVERY RATE (FDR)

- FDR controls the expected proportion of false positive values among supra-threshold values (i.e., false claims v false tests):
- $p < 0.05$ FWE means: There is only a 5% chance any result is a false positive.
- $p < 0.05$ FDR means: No more than 5% of active voxels are false positives.

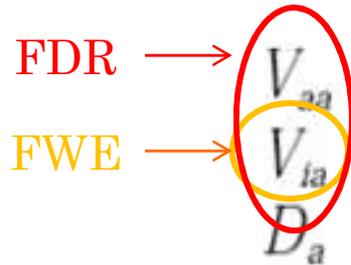


FALSE DISCOVERY RATE (FDR)

- FDR controls the expected proportion of false positive values among supra-threshold values (i.e., false claims v false tests):

Classifications of Voxels in V Simultaneous Tests

	Declared active (discovered)	Declared inactive (not discovered)
Truly active	V_{aa}	V_{ai}
Truly inactive	V_{ia}	V_{ii}
	D_a	D_i
		V



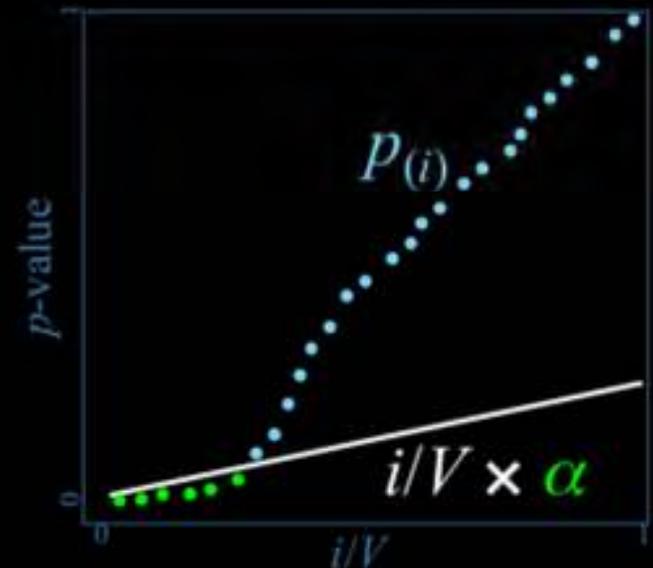
FALSE DISCOVERY RATE (FDR)

Benjamini & Hochberg Procedure

- Select desired limit α on FDR
- Order p-values, $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(V)}$
- Let r be largest i such that

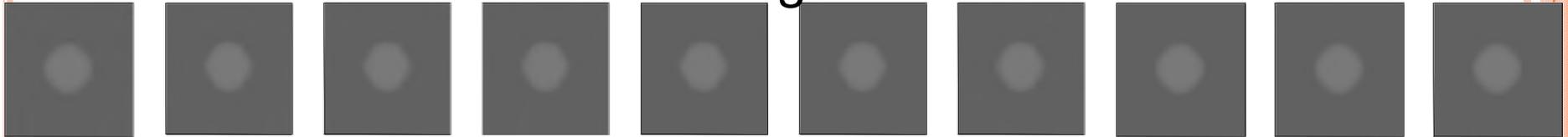
$$p_{(i)} \leq i/V \times \alpha$$

- Reject all hypotheses corresponding to $p_{(1)}, \dots, p_{(r)}$.

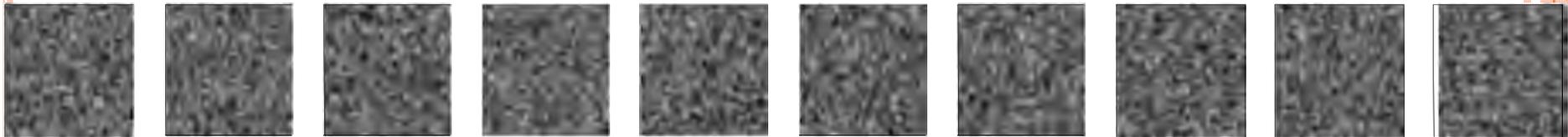


COMPARING CORRECTION METHODS

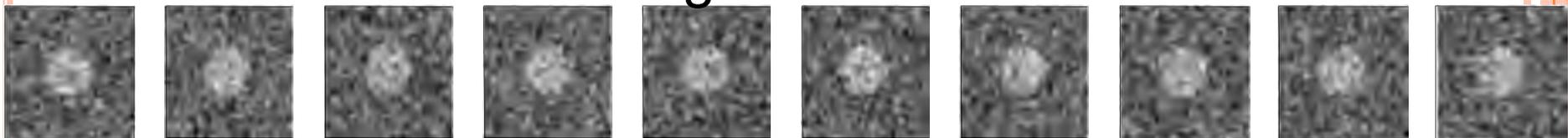
Signal



Noise



Signal+Noise



Exp
1

Exp
2

Exp
3

Exp
4

Exp
5

Exp
6

Exp
7

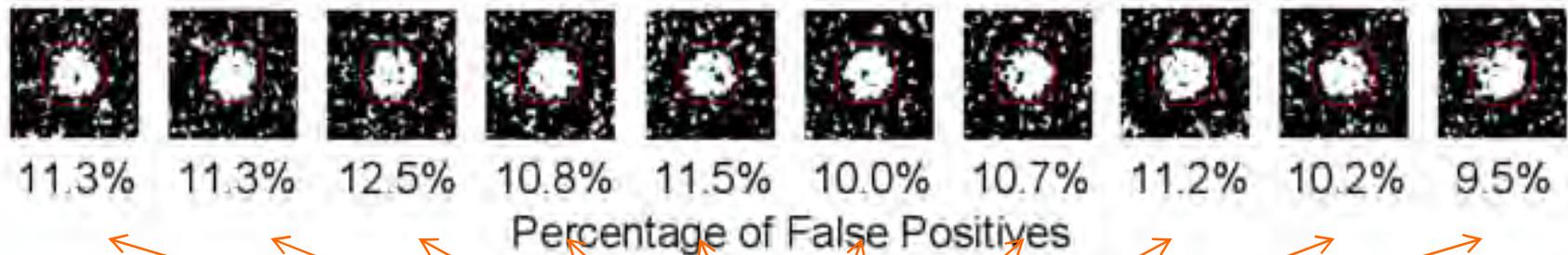
Exp
8

Exp
9

Exp
10

No CORRECTION ($\alpha = 0.1$)

Control of Per Comparison Rate at 10%

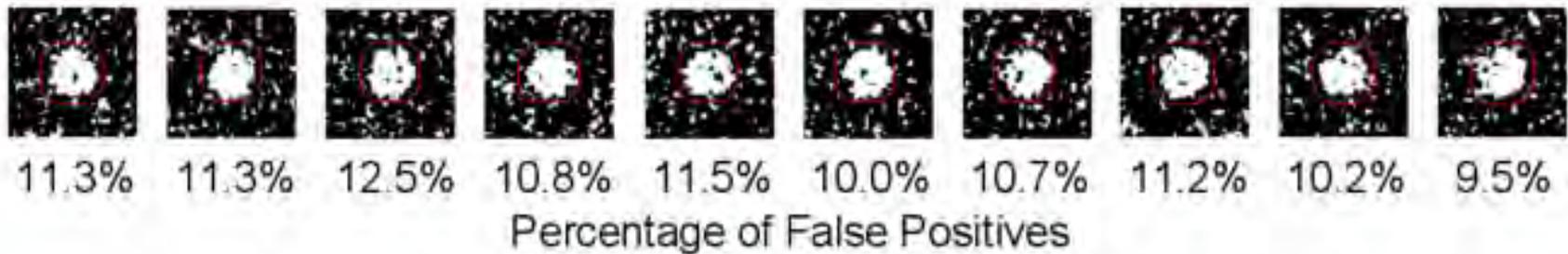


On average, 10% of the 'false' voxels are incorrectly declared "active."

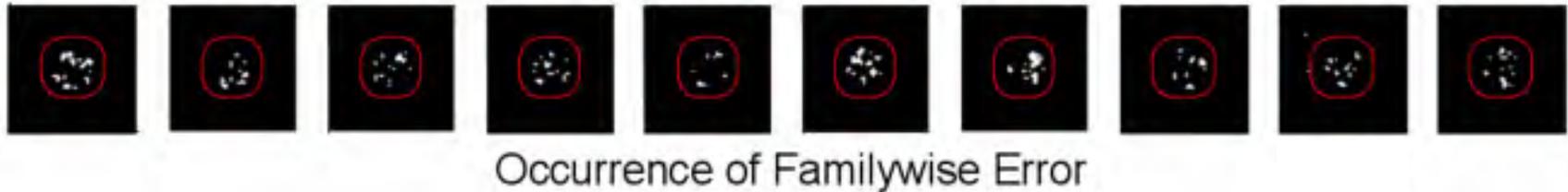
In *each* experiment we have about 10% false alarms

FWE ($\alpha = 0.1$)

Control of Per Comparison Rate at 10%

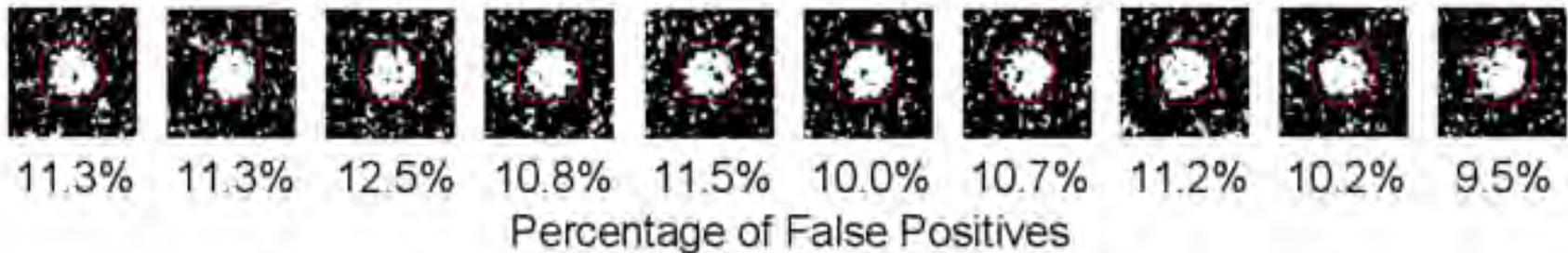


Control of Familywise Error Rate at 10%

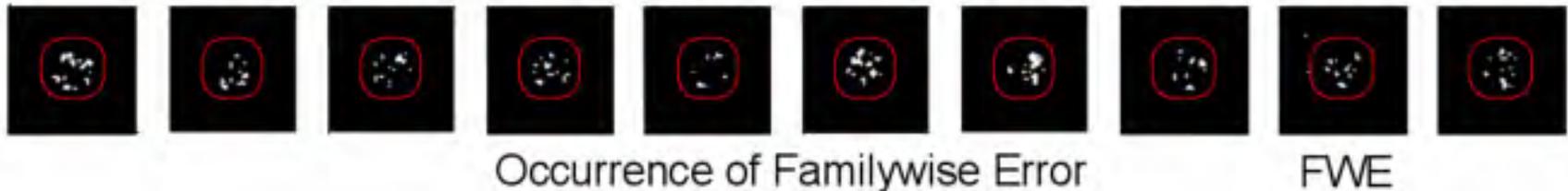


FDR ($\alpha = 0.1$)

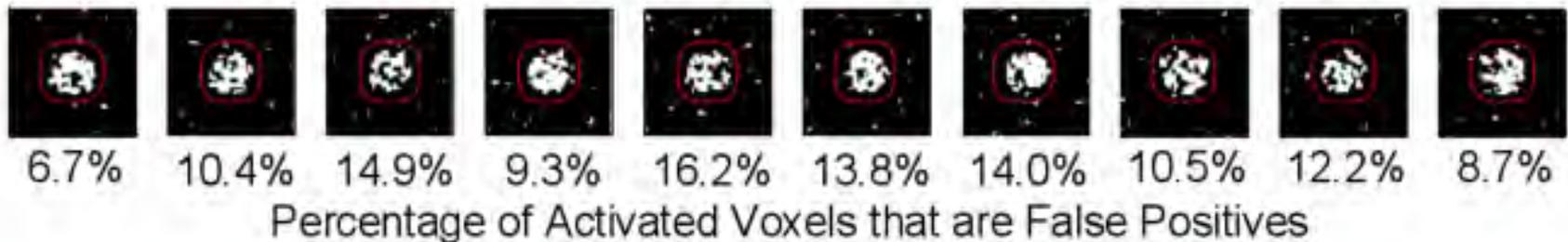
Control of Per Comparison Rate at 10%



Control of Familywise Error Rate at 10%



Control of False Discovery Rate at 10%



REAL DATA: FWE (RFT) v FDR

