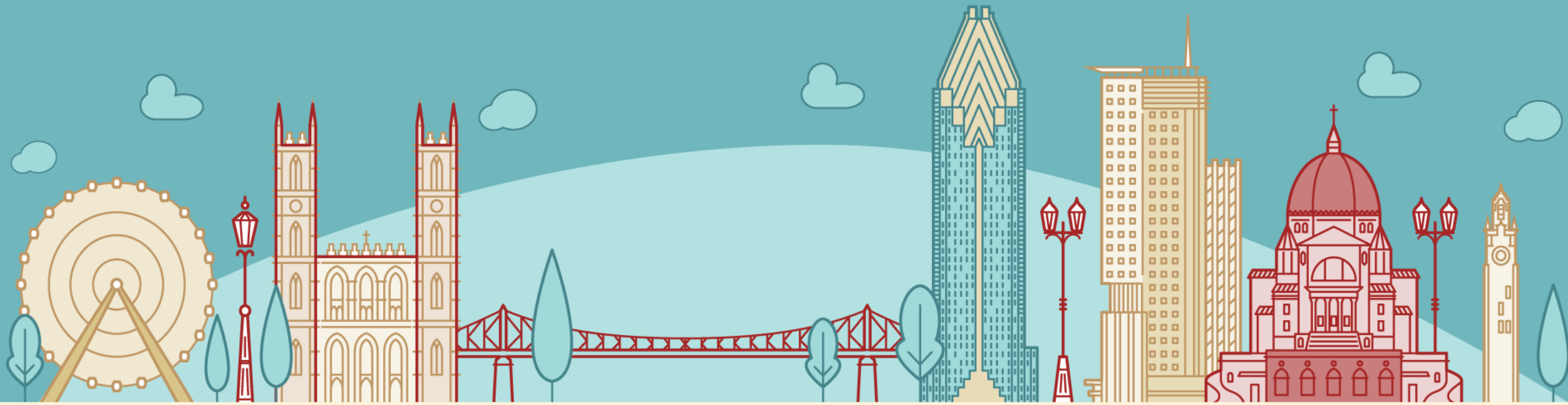
OHBM 2023

MONTRÉAL

JULY 22-26



I have no disclosures

OHBM 2023

Educational course
Beyond blobology: advances in statistical inference for neuroimaging

MONTRÉAL

Beyond blobology educational course

OHBM 2023

Morning theoretical sessions:

08:00 - 08:15 Walk-in with coffee

08:15 - 08:20 Organizers, General Introduction

08:20 - 08:40 Wouter Weeda, Classical cluster inference and its caveats

08:40 - 09:00 Stephanie Noble, Cluster failure or power failure?

09:00 - 09:40 Jelle Goeman FDP/TDP based methods primer

09:40 - 09:50 Coffee break (10min)

09:50 - 10:30 Alexandre Blain, Non-parametric templates

10:30 - 11:10 Samuel Davenport, TDP Inference in GLM

11:10 - 11:20 Coffee break (10min)

11:20 - 12:00 Thomas Nichols, Spatial inference via confidence sets

12:00 - 12:40 Amanda Mejia, Spatial Bayesian models

12:40 - 13:45 Lunch break (1hr05)

Afternoon practicals:

13:45 - 14:30 Xu Chen, All-resolutions Inference Framework

14:30 - 15:15 Alexandre Blain, Introducing the Sansouci package and Notip.

15:15 - 15:30 Tea break (15 min)

15:30 - 16:15 Samuel Davenport, TDP Inference in regression

16:15 - 17:00 Tom Maullin, Spatial inference via confidence sets

17:00- 17:15 Organizers, Closing statements

MONTRÉAL

Our mission today

OHBM 2023

Insert famous quote about blobs here

MONTRÉAL

"It is a mistake to think you can solve any major problems just with potatoes."

— Douglas Adams, Life, the Universe and Everything

OHBM 2023

Classical inference and its caveats

Wouter Weeda, Leiden University

MONTRÉAL

OHBM 2023

Part 1: Inference in functional MRI

MONTRÉAL

Inference in functional MRI

OHBM 2023

Classical inference and its caveats

MONTRÉAL

Inference in functional MRI

- For most functional MRI studies measured signal comes from distinct locations in the brain called voxels: a 3-dimensional grid of 3x3x3 mm cubes.

- Inference in functional MRI is done on each location (voxel) separately.

- The maps that you often see are the outcomes of this inference (usually in the form of a z or t-statistic indicating significance).

Inference in functional MRI

- Outcomes are usually statistics values for a contrast (e.g., rest versus task, or group A vs group B).

- These statistics can be converted into per voxel p-values and subsequently z-values.

- Note that p-values can also be obtained using permutations/randomization designs.

Inference in functional MRI

- How to decide which areas of the brain are ‘active’ (or more active in one condition than another)?

- We use classical null-hypothesis testing:
  - H_0 : voxel is not active
  - H_1 : voxel is active

- Calculate a p-value that conveys surprise of observing the data when H_0 is true

Inference in functional MRI

- In classical inference there are four possible scenarios:

	Statistical test H_0 = TRUE (test not significant)	**Statistical test H_0 = FALSE (test significant)**
Reality H_0 = TRUE (no effect)	Correct acceptance of H_0 (True negative)	False positive decision
Reality H_0= FALSE (genuine effect)	False negative decision	Correct rejection of H_0 (True positive)

Inference in functional MRI

- In the decision of whether H_0 can be rejected we allow some uncertainty:
  - An α% chance for falsely rejecting H_0 when H_0 is actually true (a false-positive result)
- Usually, α is set at 5%.
- So, if I would repeat my experiment multiple times, in about 5% of the cases I would falsely reject H_0

And now for the big problem

- In functional MRI studies we don't do just one test, we do over 200.000 in one experiment.
- This increases the chances of finding false-positive voxels dramatically (out of our 200.000 voxels 10.000 could be false-positives)
- If we don't control, we are in trouble. As pointed out by the dead salmon of Bennett et al. in 2009.

What do we want to control?

- When performing multiple hypothesis tests, we want to control the 'Family-wise error rate':
    - FWER: the chance of at least one false-positive result in a 'family' of tests. We want this chance to be lower than α.
- The family of tests is usually either all the voxels or clusters in the brain

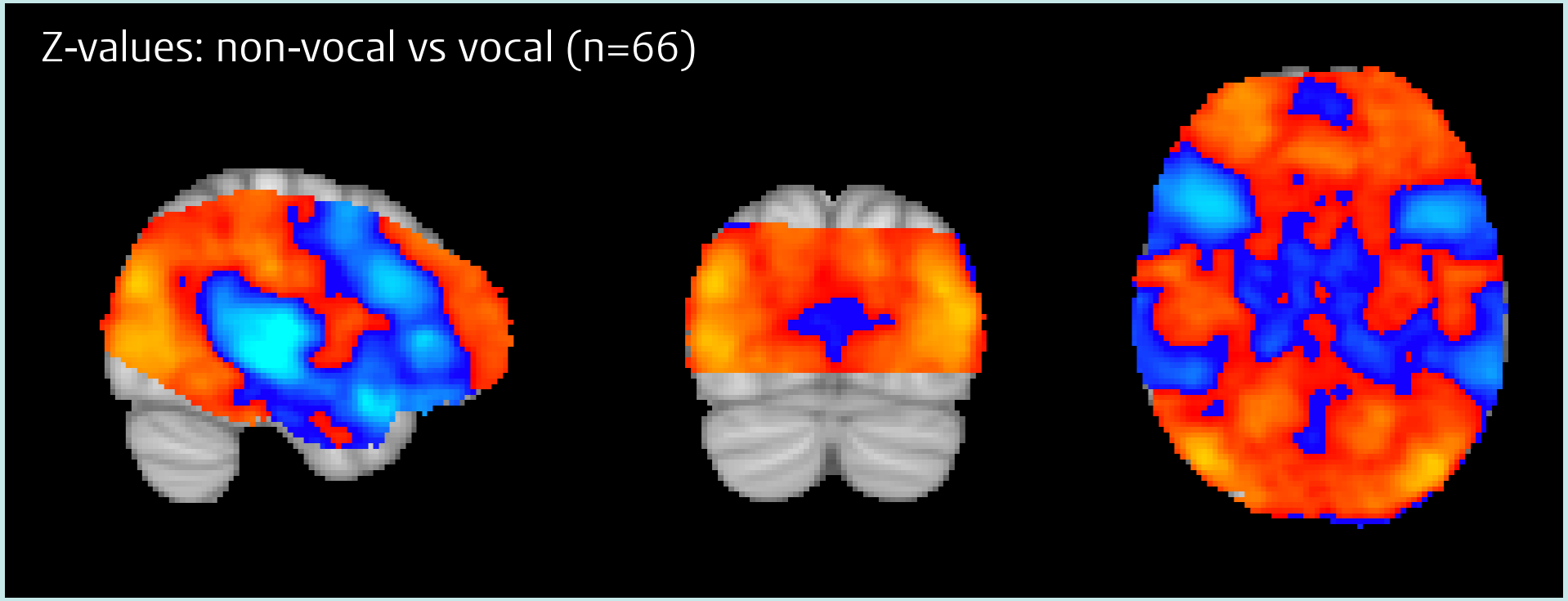
Quick recap

- The goal of fMRI inference is to decide for each voxel whether it is active or not (using a hypothesis test).
- For each test we allow a little uncertainty of whether our decision is the right one.
- When doing multiple tests, the chances of making a wrong decision somewhere in our ‘family’ of tests increases dramatically.
- The family-wise error rate (FWER) of our family of tests is what we want ‘controlled’.

Part 2: Multiple testing in functional MRI

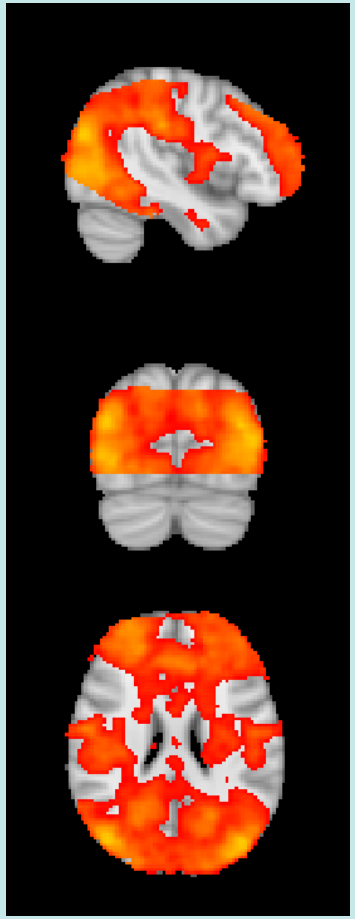
Let's start with an example

- Study on vocal and non-vocal sounds, Pernet et al., 2015



Let's start with an example

- In total 166.407 in-mask voxels.
- Focus only on positive values for now.
- Z-statistics indicate whether a voxel is more active in the *non-vocal* condition than in the *vocal* condition.
  - H_0= not active (z-value = 0)
  - H_1= active (z-value > 0)

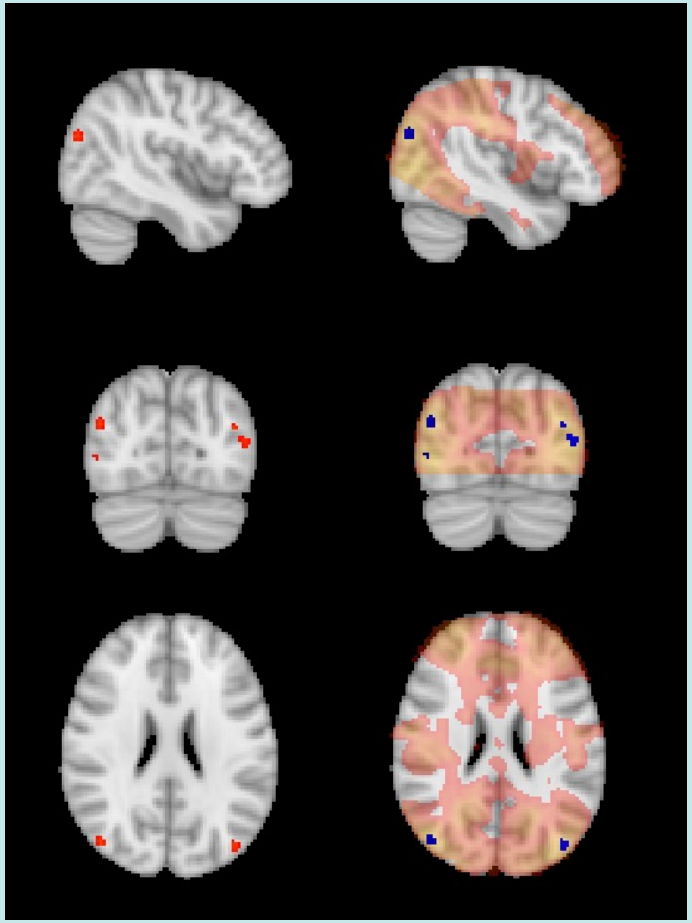


Voxel-wise error control

- Controls the FWER over all voxels in the brain (mask). Family = all voxels.
- Easiest method to control the FWER is Bonferroni correction.
- Calculated by setting the per-voxel α to be α / #voxels

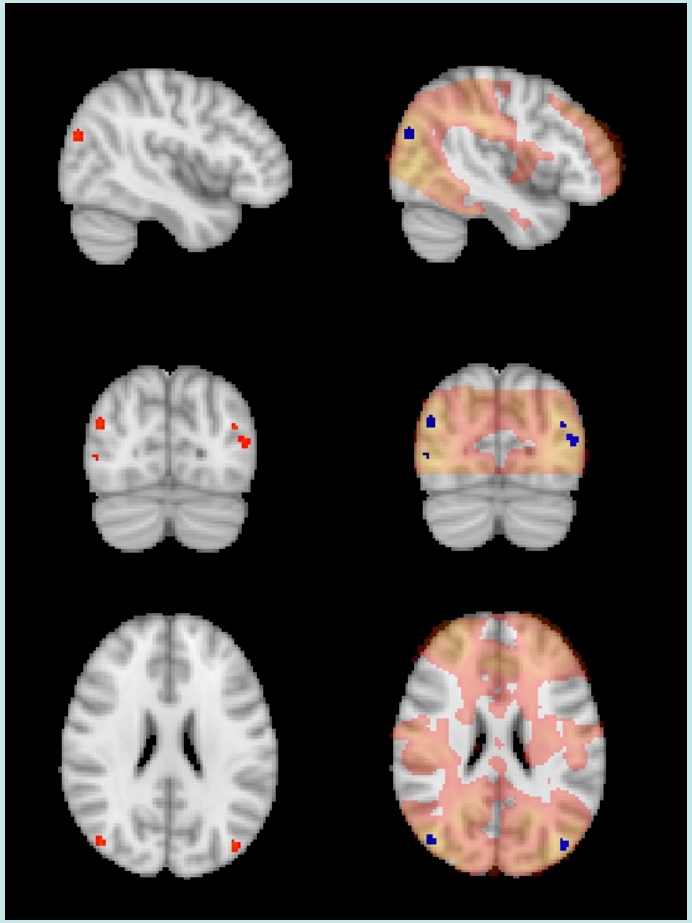
  .05 / 166407 = .0000003

- Usually not very powerful.



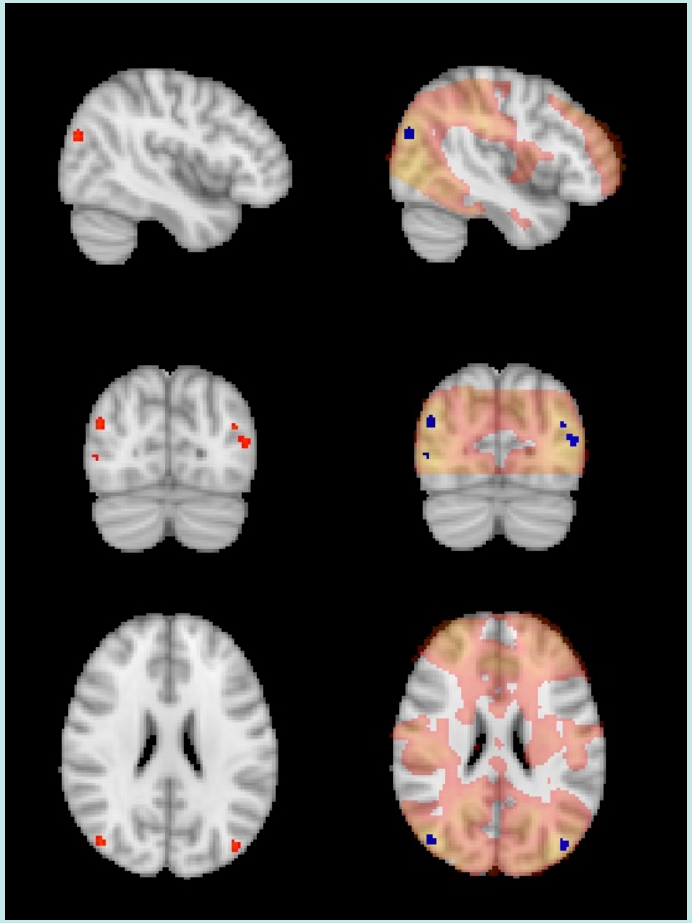
Voxel-wise error control

- But…
- Since our family is all voxels, we know exactly where the activation is!
- In other words: we have high spatial specificity.
- (because the chance of any of these voxels being a false-positive < 5%)



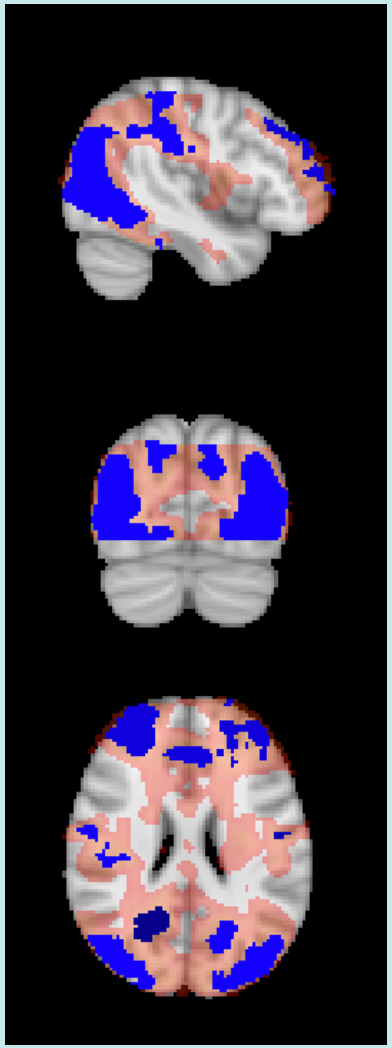
Voxel-wise error control

- Additional methods are available, for example:
  - Random Field Theory based peak-thresholding
  - Permutation based methods
  - False Discovery Rate based methods (more powerful, but weaker control of the FWER)



Cluster-extent approaches

- Usually, we are not interested in single-voxel activity per se. A more natural unit is a ‘cluster’ of voxels (which we will name ’blob’).
- A cluster or blob is defined as a contiguous/connected set of voxels.
- We control the number of false-positive blobs (our family in FWER is thus all possible blobs, not all voxels).



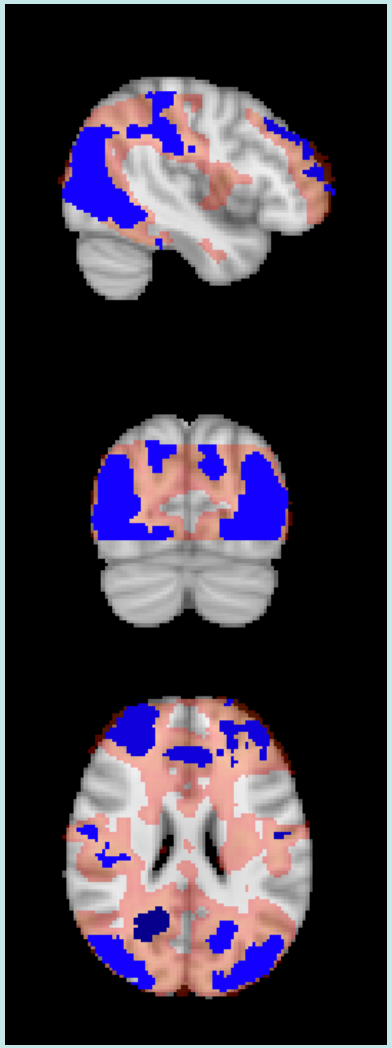
Cluster-extent approaches

- In practice, using a two-step approach:

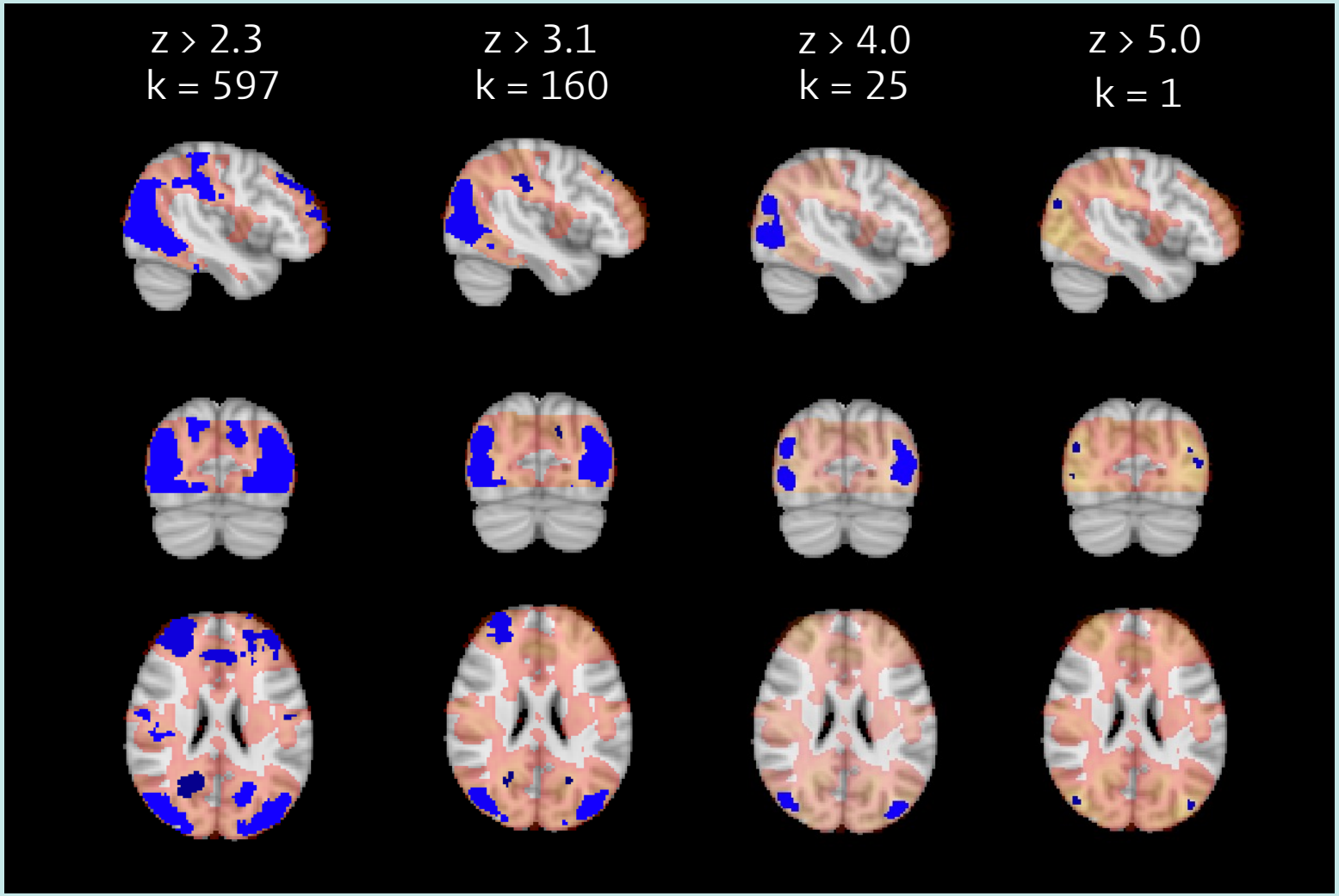
  1. Choose a ‘cluster-forming’ threshold *z* and estimate the size of all contiguous clusters above this threshold.

  Determine the minimum cluster size *k* that occurs by chance under the null (95%) given the smoothness of the data and the chosen threshold *z* (e.g., using RFT or permutations)

  2. Check which clusters are larger than *k* (all clusters that are larger are significant).



Cluster-extent approaches



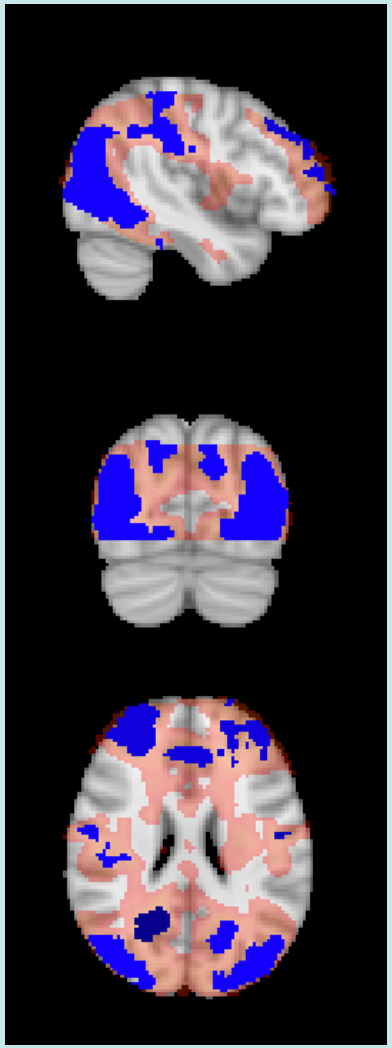
Cluster-extent approaches

- More powerful than voxel-wise approaches, but… more powerful in detecting activation, not in localizing it.

- Because of hypotheses being on the ‘cluster’ level:

  - Non-significant when cluster-extent is smaller than k
  - Significant when cluster-extent is larger than k

- No information about voxels within a cluster (clusters are large enough or not).



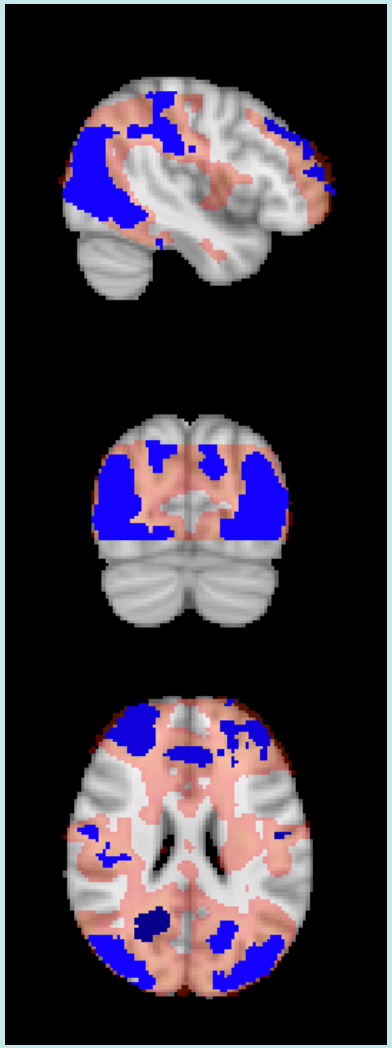
Cluster-extent approaches

- Formal way of stating this:

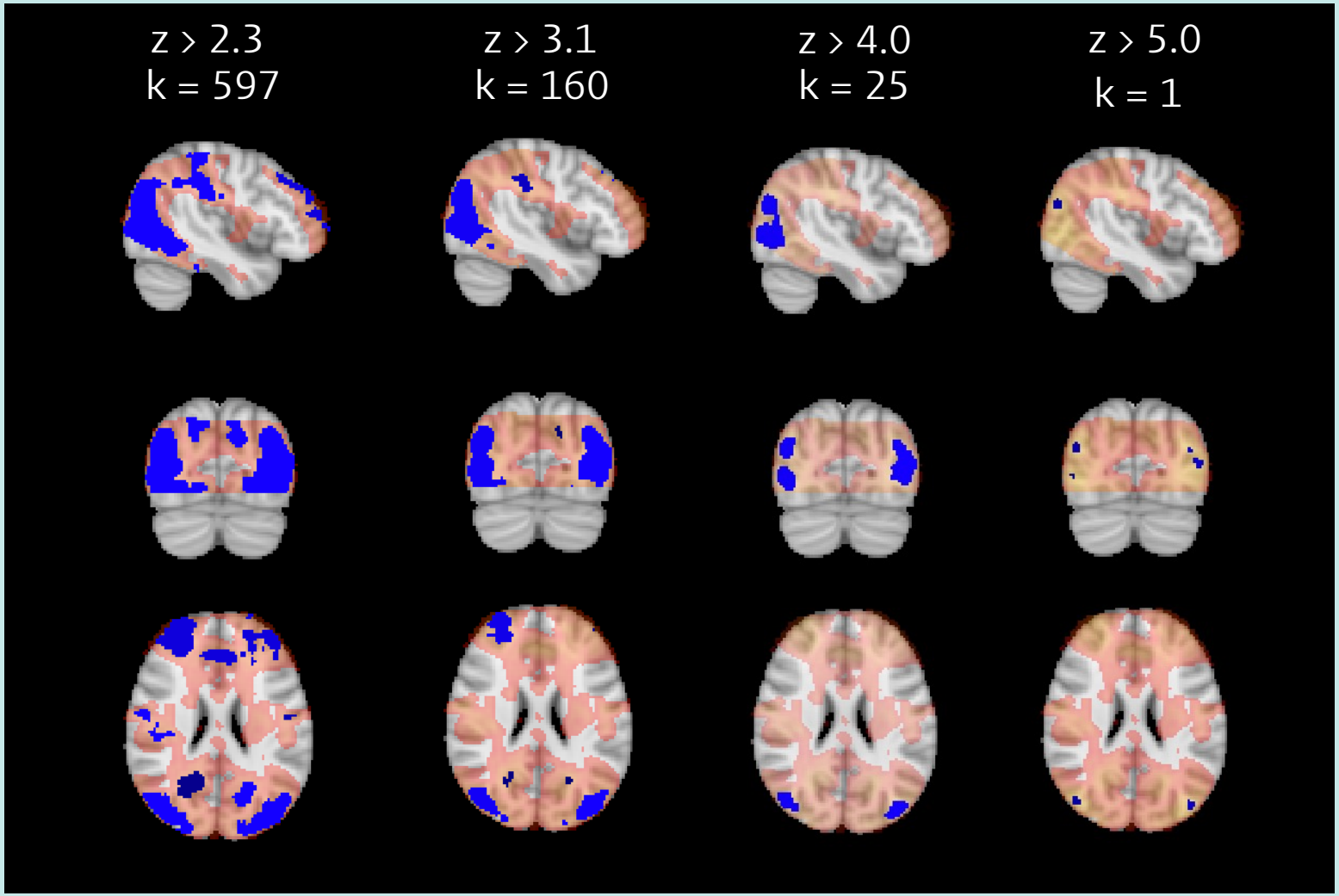
  H_0= no activation within a cluster
  H_1= at least one voxel active within a cluster

- So, the larger the cluster found, the less we know about activation within a cluster.

- This is called the Spatial specificity paradox (Woo et al., 2014, Lindquist & Mejia, 2015).



Cluster-extent approaches

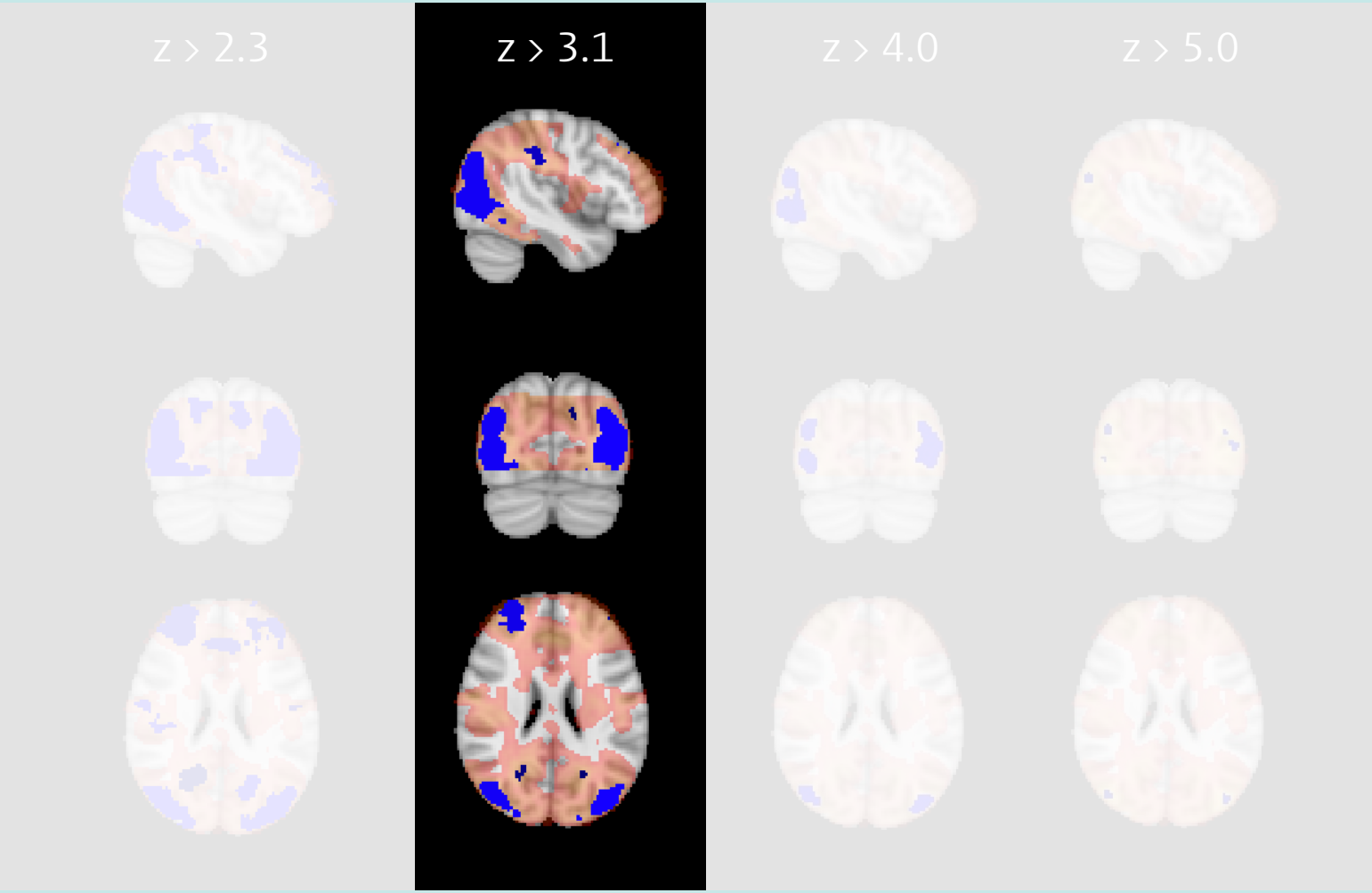


Left: more voxels but we know less about location

Right: less voxels but we know more about location

No formal way of deciding the optimal threshold.

Cluster-extent approaches



And we cannot run the analysis with another threshold after seeing the data without compromising our FWER.

Cluster-extent vs voxel-wise

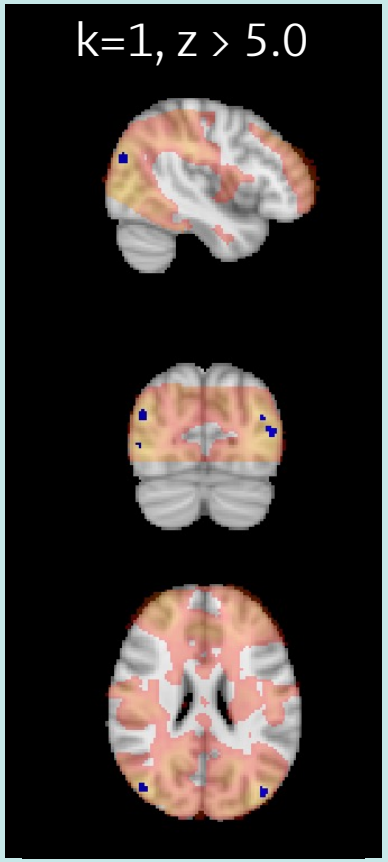
- The chance of finding at least one contiguous cluster of size at least *k*, given threshold *z*, under the null-hypothesis of no activation must be smaller than a criterion (e.g. 5%).

- Usually, researchers choose *z* and estimate *k* (e.g., using RFT, permutations, Monte-Carlo simulations).

- Setting *k* and estimating *z* is also possible:

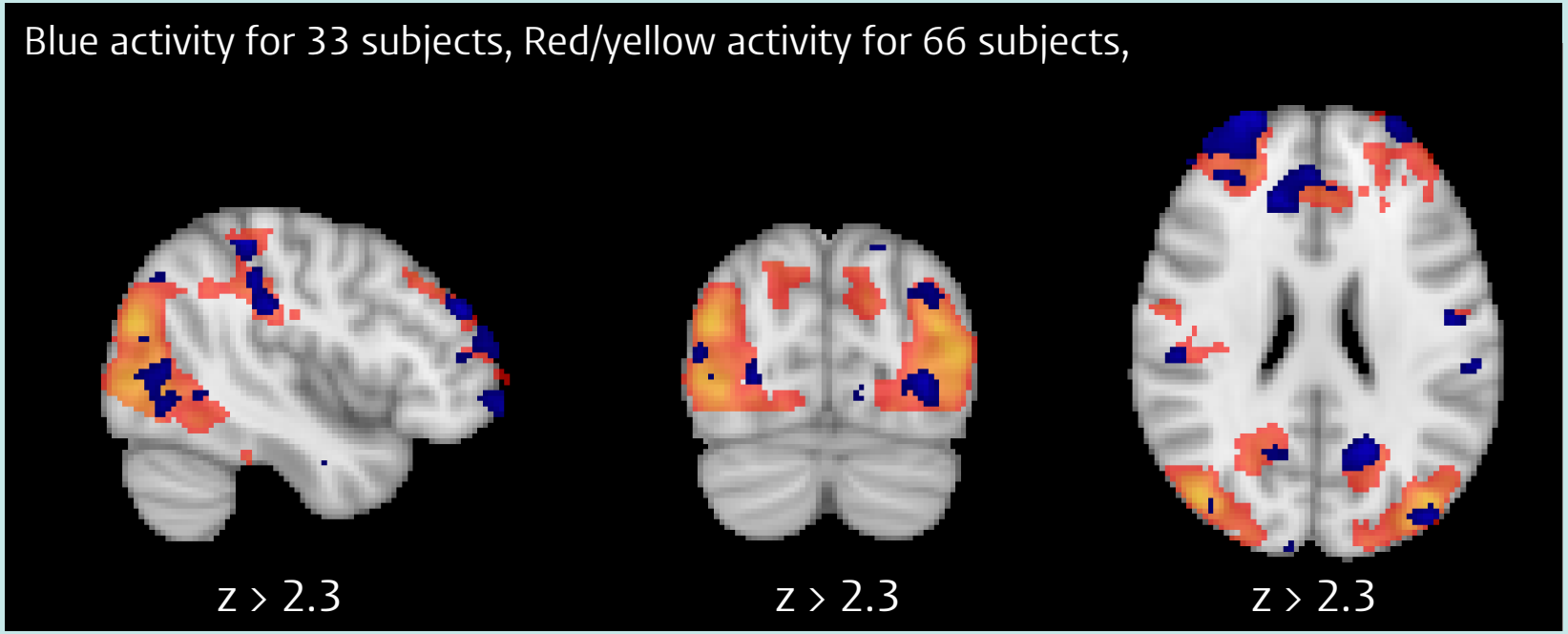
  - With k = 1, this reverts to a voxel-wise threshold.

  - With k > 1, this reverts to selecting a minimum cluster size one is interested in.



Another cluster-extent caveat

Another issues (more general with p-value based research): when the number of participants increases, the p-values get smaller, clusters get larger for the same threshold.



Summary part 2

- Voxel-wise approaches are usually less powerful in detecting activation but allow for proper localization of activity.

- Cluster-extent approaches are usually more powerful in detection activation but are less spatially specific (and it gets worse the larger the cluster).

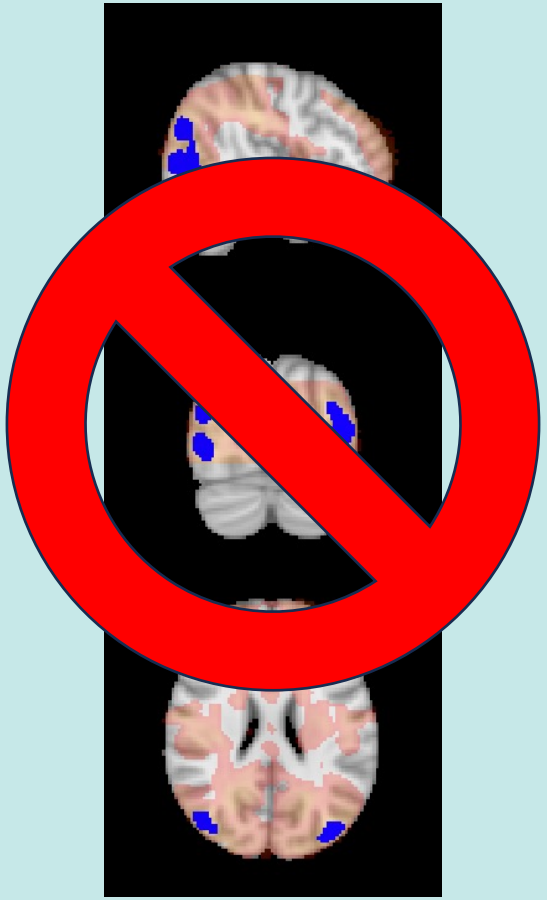
- No principled way to select the arbitrary threshold.

- With increasing *n*, cluster extent becomes too large with the same threshold.

Is all then lost?

- Cluster-extent approaches are a sub-optimal way to do inference, because statements like this cannot be answered using this approach:
  - “A large significant cluster contains a substantial number of active voxels.”
  - “A large significant cluster is a more substantial scientific finding than a small significant cluster.”
  - “Substantial overlap between a significant cluster and an anatomical brain area indicates evidence for the presence of activity in that anatomical brain area.”

Goeman et al., accepted



Of (educational) course not

- Today is about solutions to the problems of classical cluster-extent inference:
  - Approaches based on recent advances in multiple testing: true/false discovery proportions, closed testing, and joint error-rate control.
  - Approaches based on effect-sizes (instead of p-values)
  - Approaches based on (Bayesian) spatial modelling and surface (edge/vertex) data.

What to expect today

- We can actually quantify the number of ‘truly’ active voxels within a (or any cluster) in the data, without losing FWER control, increasing spatial specificity.

- We can leverage the information (aka spatial correlations) in the data to optimize this estimation.

- We can quantify spatial confidence bounds for clusters, alleviating the increasing *n* problem.

- We can use explicit statistical models to test spatial hypotheses.

Beyond blobology educational course

OHBM 2023

Morning theoretical sessions:

08:00 - 08:15 Walk-in with coffee

08:15 - 08:20 Organizers, General Introduction

08:20 - 08:40 Wouter Weeda, Classical cluster inference and its caveats

08:40 - 09:00 Stephanie Noble, Cluster failure or power failure?

09:00 - 09:40 Jelle Goeman FDP/TDP based methods primer

09:40 - 09:50 Coffee break (10min)

09:50 - 10:30 Alexandre Blain, Non-parametric templates

10:30 - 11:10 Samuel Davenport, TDP Inference in GLM

11:10 - 11:20 Coffee break (10min)

11:20 - 12:00 Thomas Nichols, Spatial inference via confidence sets

12:00 - 12:40 Amanda Mejia, Spatial Bayesian models

12:40 - 13:45 Lunch break (1hr05)

Afternoon practicals:

13:45 - 14:30 Xu Chen, All-resolutions Inference Framework

14:30 - 15:15 Alexandre Blain, Introducing the Sansouci package and Notip.

15:15 - 15:30 Tea break (15 min)

15:30 - 16:15 Samuel Davenport, TDP Inference in regression

16:15 - 17:00 Tom Maullin, Spatial inference via confidence sets

17:00- 17:15 Organizers, Closing statements

MONTRÉAL

Advanced blobology