

---

# Association and prediction of phenotypic traits from neuroimaging data using a multi-component mixed model excluding the target vertex

Baptiste Couvy-Duchesne  
SPIE 2021 Conference

[baptiste.couvy@icm-institute.org](mailto:baptiste.couvy@icm-institute.org)  
<https://github.com/baptisteCD>

@BaptisteCouvy 

# Brain mapping

## Aim:

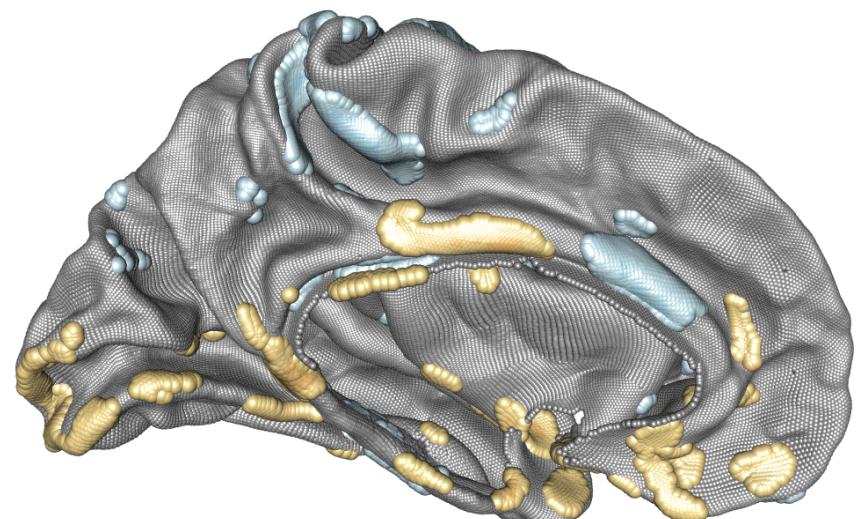
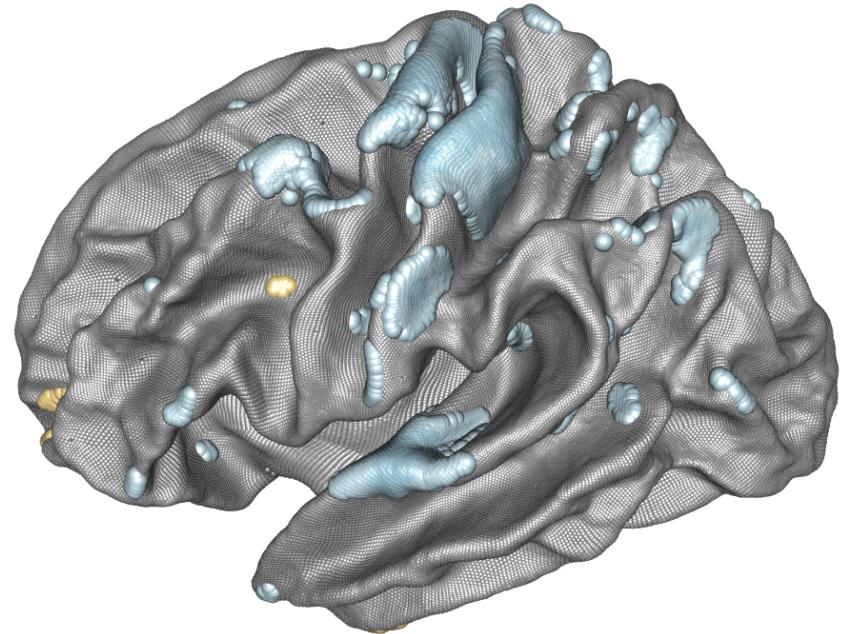
- Identify brain regions associated with a trait/disease

## Advantages:

- Direct insight into brain networks relevant for trait/disease
- Interpretable prediction models from associated regions

## Challenges:

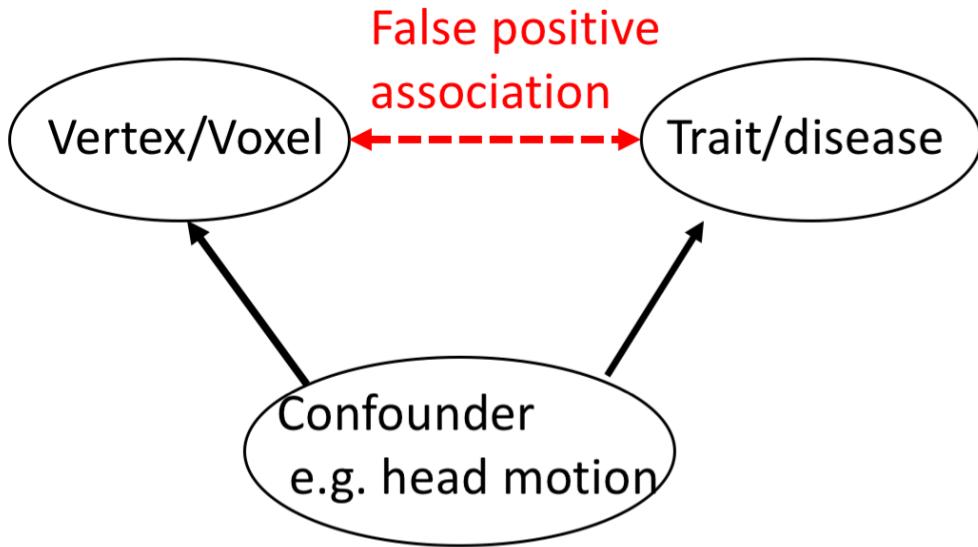
- High dimensional data i.e. “finding needle(s) in haystack”
- All confounders not known
- Complex correlation pattern of brain feature likely unrelated to trait/disease of interest



# Confounders in high-dimensional data

3

a)



Traditional confounder  
paradigm - case of a single  
brain feature

# Some evidence of this problem

> Neuroimage. 2012 Nov 1;63(2):754-9. doi: 10.1016/j.neuroimage.2012.06.052. Epub 2012 Jul 6.

## The impact of aging on gray matter structural



**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT | SUBMIT | NEWS & NOTES | ALI

Search

bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be considered as practice/health-related behavior, or be reported in news media as established information.

New Results

[Comment on this paper](#)

[Previous](#)

### Removal of Scanner Effects in Covariance Improves Multivariate Pattern Analysis in Neuroimaging Data

Posted December 14, 2020.

Andrew A. Chen, Joanne C. Beer, Nicholas J. Tustison, Philip A. Cook, Russell T. Shinohara, Haochang Shou, the Alzheimer's Disease Neuroimaging Initiative

doi: <https://doi.org/10.1101/858415>

This article is a preprint and has not been certified by peer review [what does this mean?].

[Download PDF](#)

[Supplementary Material](#)

[XML](#)

[Revision Summary](#)

Age-dependent covariance between structural brain features

Seen at OHBM last year - covariance structure dependent on scanner type

- Covariance pattern influenced by known confounders (age, sex, scanner) which may be accounted for
- What about unmeasured (e.g. head motion, pathologies...) or unknown confounders?

# Linear Mixed Models (LMM) : a solution?

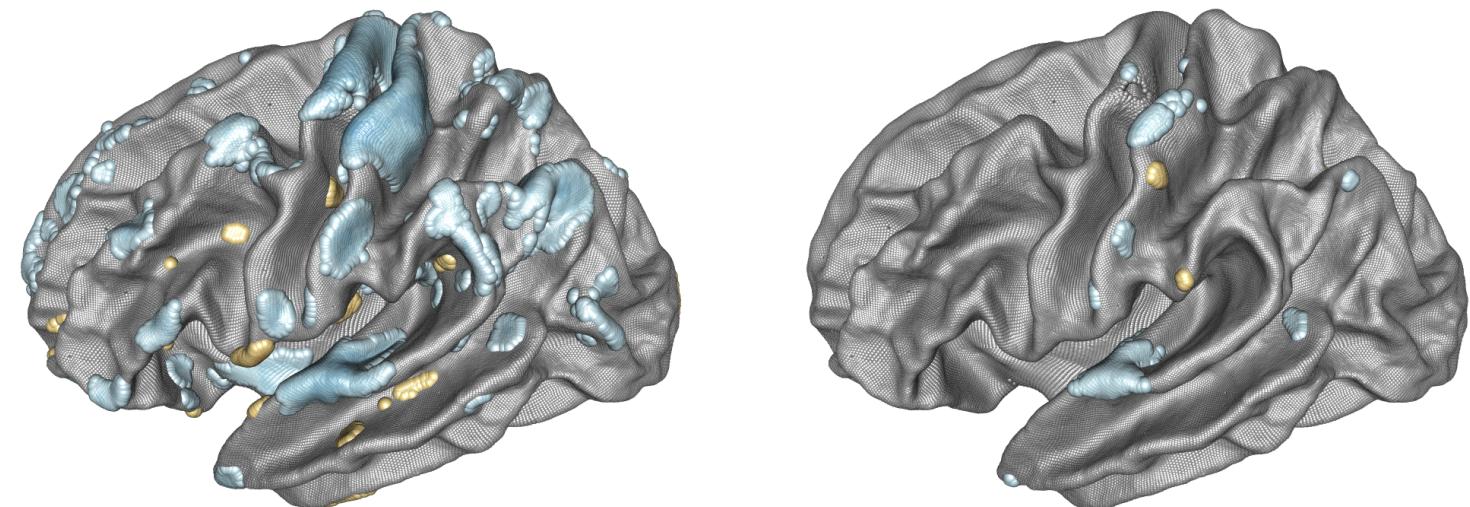
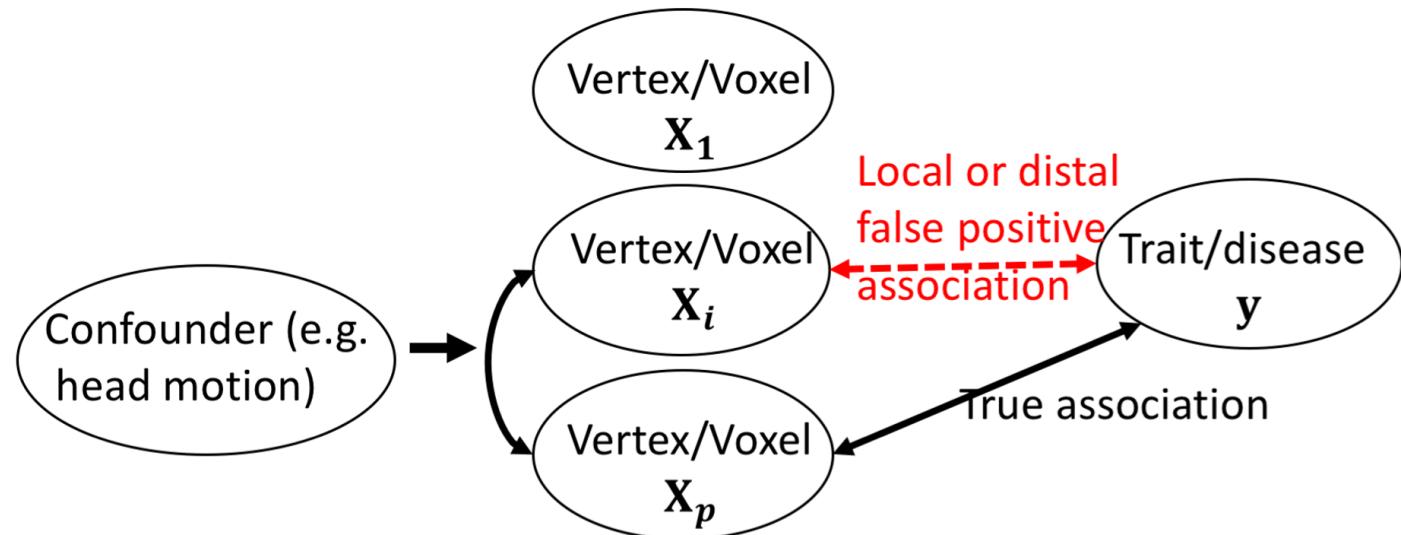
5

**Rationale:** control for other brain feature in the model

- ⇒ remove redundant associations caused by confounded covariance structure
- ⇒ does not require knowledge or direct measure of confounder

All brain feature fitted as a random effect to accommodate dimension ( $p$  feature  $\gg N$  observations)

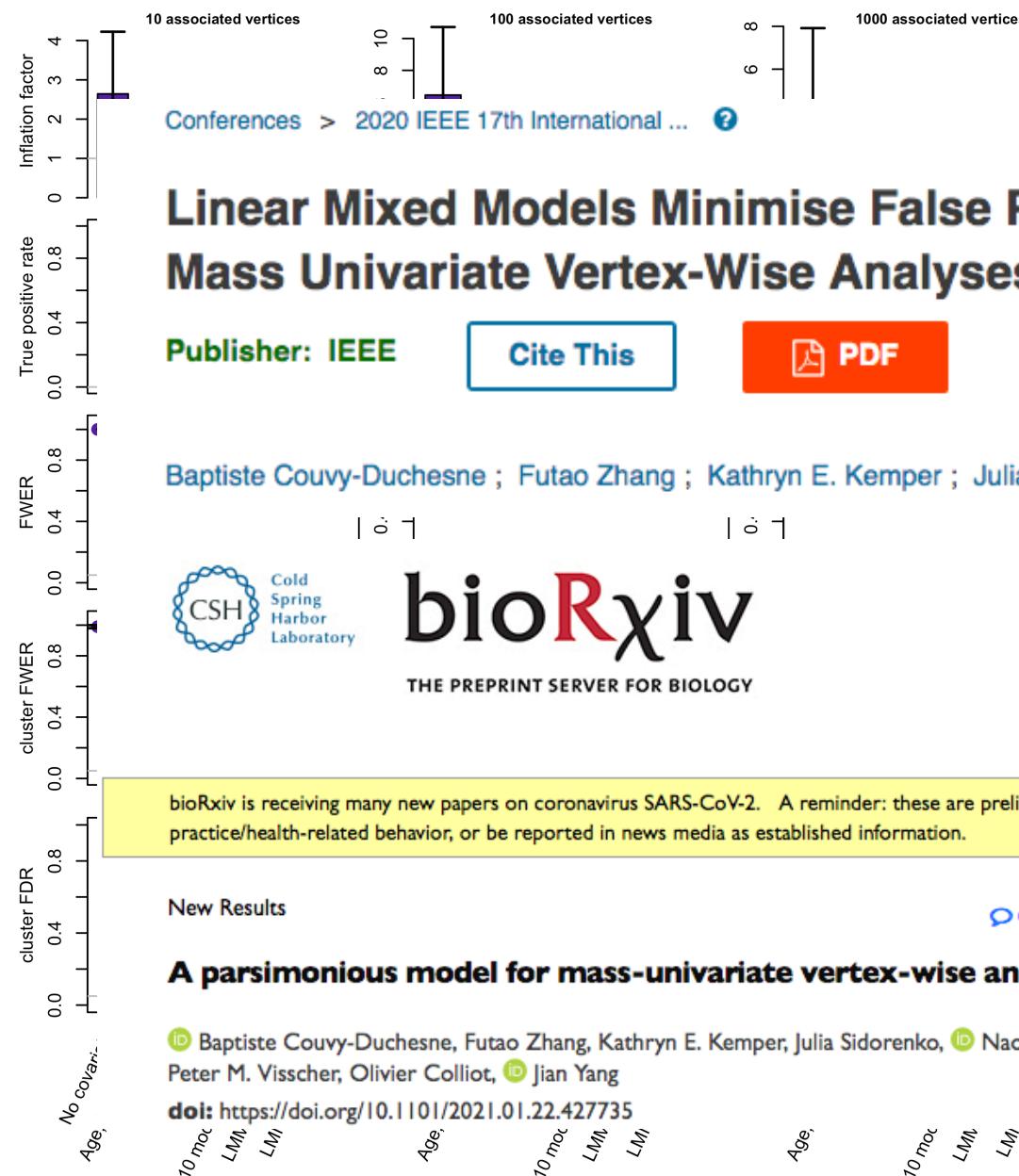
LMM now routinely used in Genome-Wide Association Studies.  
*Yang et al. Advantages and pitfalls in the application of mixed model association methods, Nat Genetics, 2014*



$$\mathbf{y} = \mathbf{X}_i b_i + \boldsymbol{\varepsilon} \quad (\text{GLM})$$

$$\mathbf{y} = \mathbf{X}_i b_i + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\text{LMM})$$
$$\boldsymbol{\beta} \sim \mathcal{N}(0, \mathbf{I}\sigma_{\boldsymbol{\beta}}^2)$$

# LMM vs. GLM



We compared different models:

Linear Mixed Models Minimise False Positive Rate and Enhance Precision of Mass Univariate Vertex-Wise Analyses of Grey-Matter

Baptiste Couvy-Duchesne ; Futao Zhang ; Kathryn E. Kemper ; Julia Sidorenko ; Naomi R. Wray ; Peter M. Visscher ; Olivier Coll... All Authors

HOME | ABOUT | SUBMIT | NEWS & NOTES | ALERTS / RSS | CHANNELS

Search



Advanced Search



**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not been peer-reviewed. They should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information.

New Results

Comment on this paper

Previous

Next

## A parsimonious model for mass-univariate vertex-wise analysis

Posted January 22, 2021.

Baptiste Couvy-Duchesne, Futao Zhang, Kathryn E. Kemper, Julia Sidorenko, Naomi R. Wray,

Peter M. Visscher, Olivier Colliot, Jian Yang

doi: <https://doi.org/10.1101/2021.01.22.427735>

Download PDF

Supplementary Material

Email

Citation Tools

**MOMENT:** Multi-cOmponent Mixed model ExcludiNg the Target

**Objectives:** solve limitations of LMM

- Reduced power
- False positive rate still > 5%

Model first proposed for DNA-Methylation-Wide Association Studies

Zhang *et al.* *Genome Biology* (2019) 20:107  
<https://doi.org/10.1186/s13059-019-1718-z>

Genome Biology

METHOD

Open Access

**OSCA:** a tool for omic-data-based complex trait analysis



Futao Zhang<sup>1</sup>, Wenhan Chen<sup>1</sup>, Zhihong Zhu<sup>1</sup>, Qian Zhang<sup>1</sup>, Marta F. Nabais<sup>1,2</sup>, Ting Qi<sup>1</sup>, Ian J. Deary<sup>3</sup>, Naomi R. Wray<sup>1,4</sup>, Peter M. Visscher<sup>1,4</sup>, Allan F. McRae<sup>1</sup> and Jian Yang<sup>1,4,5\*</sup>

# MOMENT formulation

$$\mathbf{y} = \mathbf{X}_i b_i + \boldsymbol{\varepsilon} \quad (\text{GLM})$$

$$\begin{aligned}\mathbf{y} &= \mathbf{X}_i b_i + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\text{LMM}) \\ \boldsymbol{\beta} &\sim \mathcal{N}(0, \mathbf{I} \sigma_{\boldsymbol{\beta}}^2)\end{aligned}$$

$$\mathbf{y} = \mathbf{X}_i b_i + \mathbf{X}_0 \boldsymbol{\beta}_0 + \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \quad (\text{MOMENT})$$

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(0, \mathbf{I} \sigma_{\boldsymbol{\beta}_0}^2)$$

$$\boldsymbol{\beta}_1 \sim \mathcal{N}(0, \mathbf{I} \sigma_{\boldsymbol{\beta}_1}^2)$$

MOMENT: extension of LMM with two random effects - in practice bins brain feature into two groups

- $\mathbf{X}_0$  all “non-associated vertices” - set of feature with small effects
- $\mathbf{X}_1$  the “associated vertices” i.e. feature with large(r) effects
- $\mathbf{X}_1$  and  $\mathbf{X}_0$  exclude target vertex  $\mathbf{X}_i$  (and flanking region) - prevents from “double fitting”

In practice: we select  $\mathbf{X}_0$  and  $\mathbf{X}_1$  using results from GLM

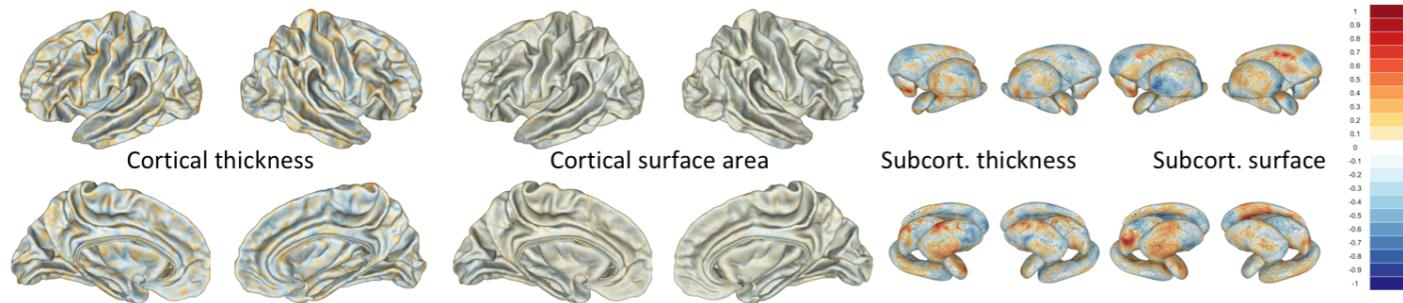
Definition of “flanking region” - correlation level with  $\mathbf{X}_i$

=> fast, efficient, data driven

=> hyper-parameter

# Realistic simulation framework

- Simulation of synthetic traits from real processed images



**Vertex-wise measurements (standardised)**

$m \sim 650,000$  measurements per individuals

Processing of T1w +T2w MRI images  
(FreeSurfer 6.0 “fsaverage” mesh + ENIGMA-shape)

- Sample of  $N=8,662$  participants from the UK Biobank
- Associated vertices selected at random
- 100 synthetic traits simulated for each scenario



**Scenario 1:**

10 associated vertices  
Total  $R^2=0.2$

⇒ On average, each vertex association  $r^2=0.02$

**Scenario 2:**

100 associated vertices  
Total  $R^2=0.5$

⇒ On average, each vertex association  $r^2=0.005$

**Scenario 3:**

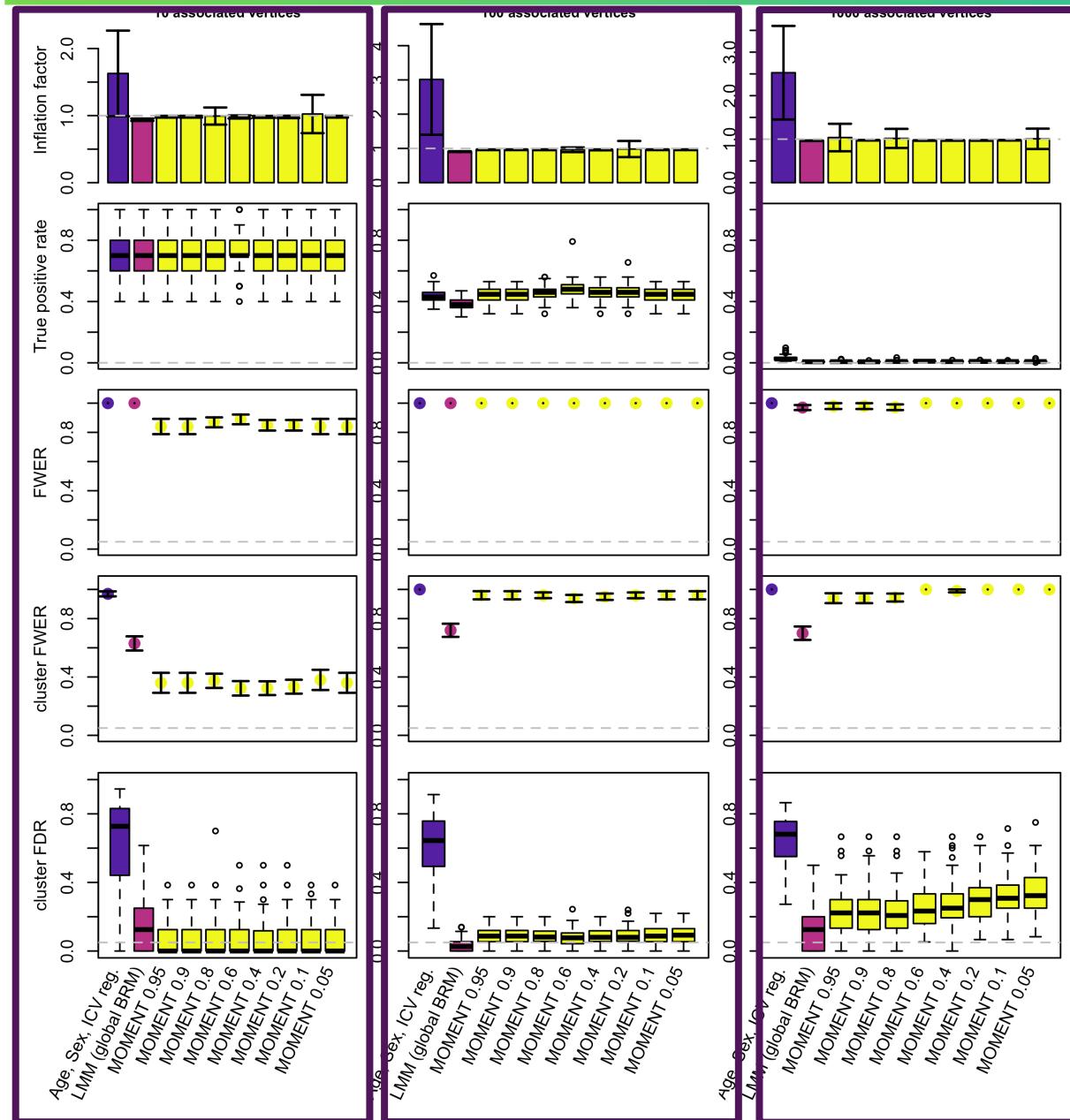
1000 associated vertices  
Total  $R^2=0.4$

⇒ On average, each vertex association  $r^2=0.0004$

- For each synthetic trait and model, we corrected for multiple testing using Bonferroni correction

# MOMENT vs. GLM or LMM

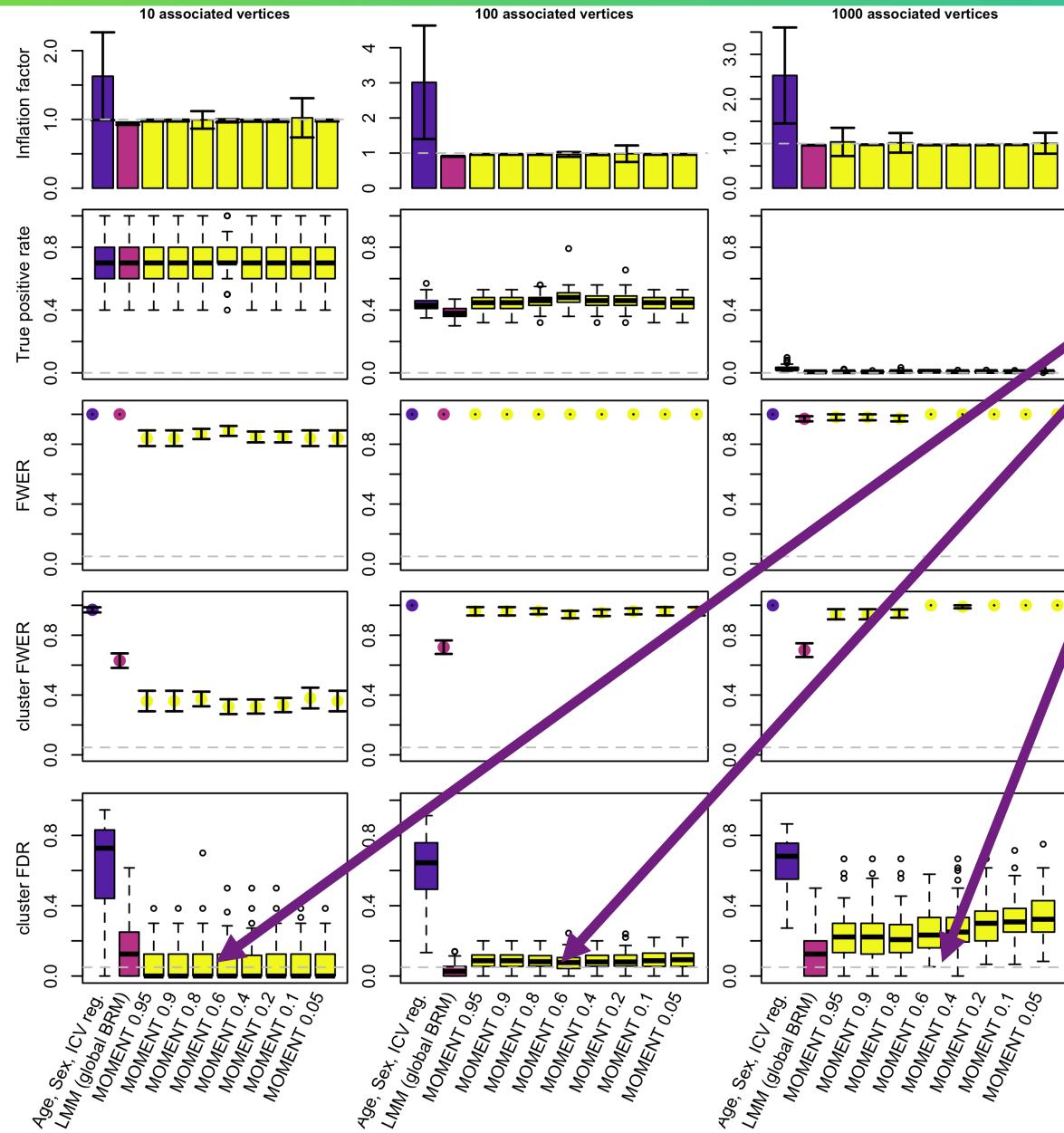
10



The 3 simulation scenarios

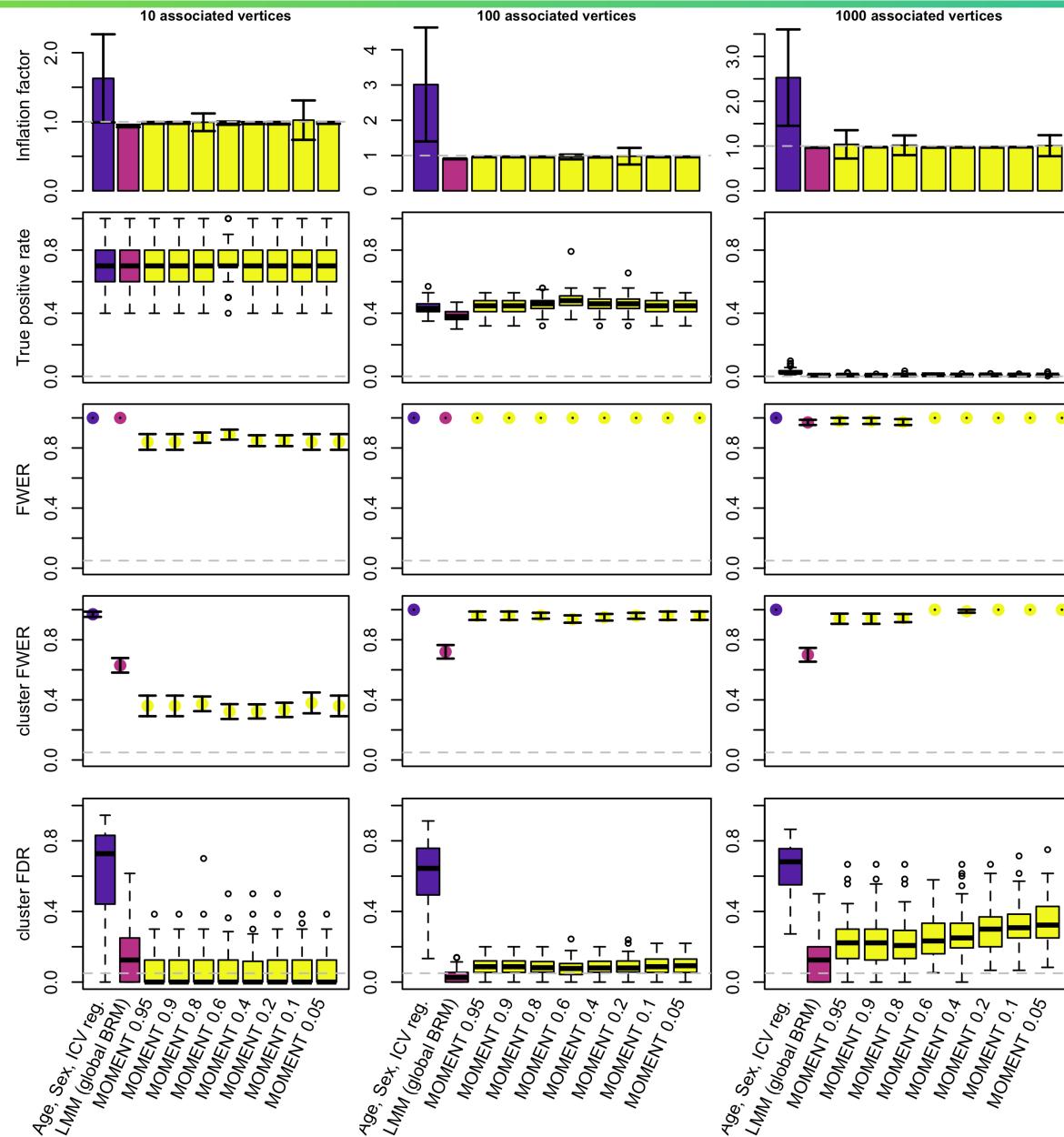
# MOMENT vs. GLM or LMM

11



The different models, incl. several definition of the flanking region (correlation cut-offs) for MOMENT

# MOMENT vs. GLM or LMM



Inflation factor - expected value of 1  
(median of “null” test-statistic)

True Positive Rate i.e. statistical power  
**MOMENT on par with GLM**

FWER: proportion of replicates with false positive vertex

**Expected to be high in presence of association clusters**

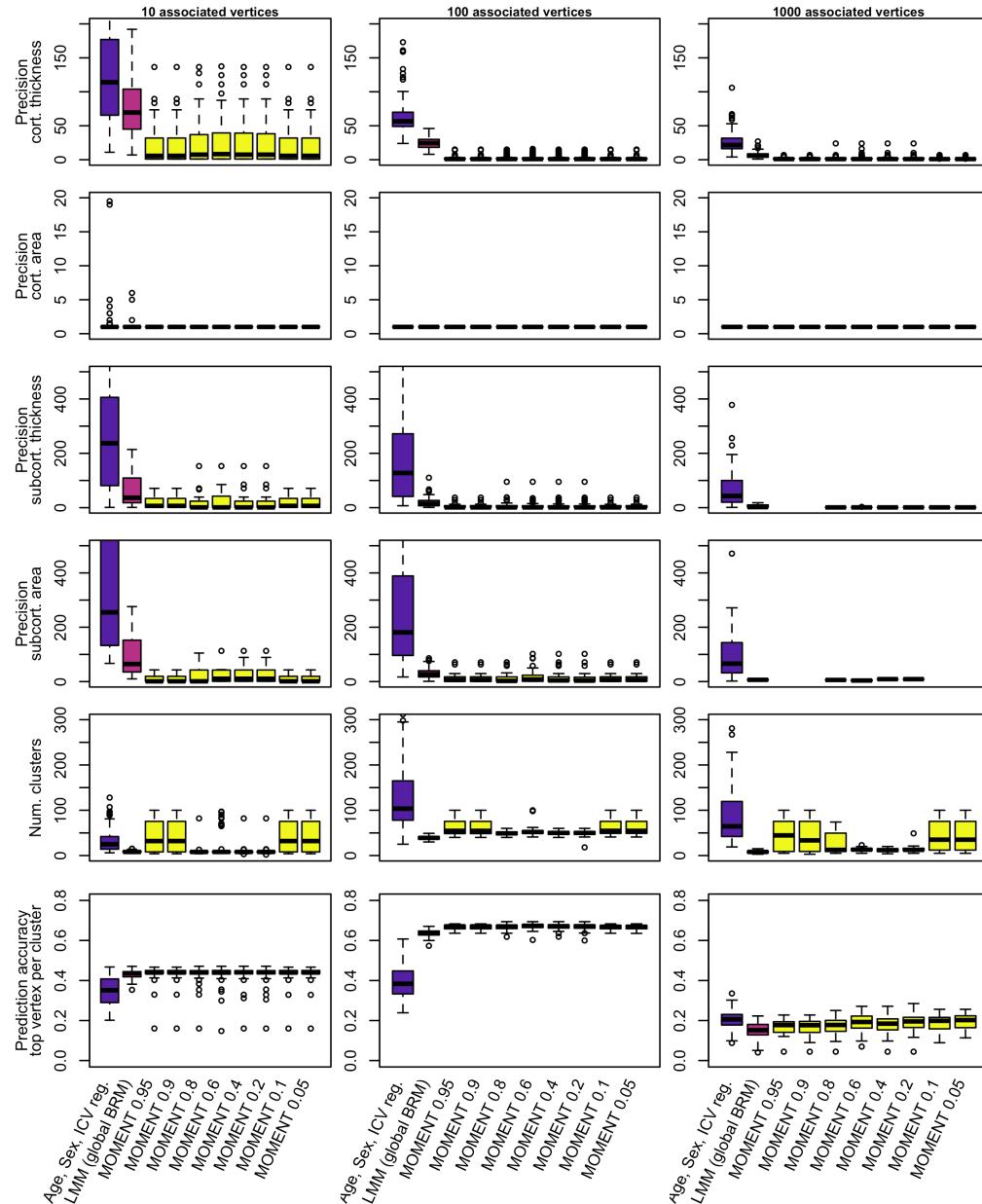
Cluster FWER: proportion of replicates with false positive cluster

**For complex traits - MOMENT always return at least one FP cluster**

Cluster FDR: proportion of false positive clusters

**MOMENT still yields fewer FP clusters than GLM**

# MOMENT vs. GLM or LMM (II)



**Mapping precision:** size of the true positive clusters

- Broken down by type of measurement (e.g. cortical / subcortical, thickness / surface area)

**MOMENT maximises precision**

**Prediction accuracy achieved from the significant clusters (top vertex in each cluster)**  
**Prediction evaluated in independent UKB sample**

**MOMENT maximises prediction accuracy**

Overall the use of mixed model (LMM or MOMENT) vs. GLM

- Reduce the rate of redundant/false positive associations
- Improve mapping precision
- Improve prediction accuracy

## Advantages of MOMENT:

- Statistical power
- False positive rate, in presence of large effects / simple(r) traits
- Overall limited number of FP cluster (vs. GLM)
- Maximal prediction accuracy

## Limitations:

- Presence of a hyper-parameter / definition of “flanking region”
  - Cor>0.6 best one within the ones we tested
  - May be trait/brain region dependent
  - May be improved to include spatial information
- Computational requirements

<https://cnsgenomics.com/software/osca/#Overview>

Efficient implementation of GLM, LMM and MOMENT in C++, data management and more!

# Thank you

15

Big thank you to all (UKB) participants who contributed their time and data



Algorithmes, modèles et méthodes pour les images et les signaux du cerveau humain  
**(ARAMIS lab)**  
Paris Brain Institute  
UKB application: 53185



**Program for Complex Trait Genomics (PCTG)**  
Institute for Molecular Bioscience, the University of Queensland  
UKB application: 12505

<https://github.com/baptisteCD>  
@BaptisteCouvy   
baptiste.couvy@icm-institute.org