

Morphometricity is Biased by Image Smoothness

Nicolas Salvy¹, Thomas Nichols²

¹Université Paris-Saclay, Inria, CentraleSupélec, CEA, Palaiseau & Gif-sur-Yvette, France

²University of Oxford, Oxford, United Kingdom



Morphometricity

The proportion of **phenotypic** variation that can be explained by **macroscopic brain morphology**.

BMI

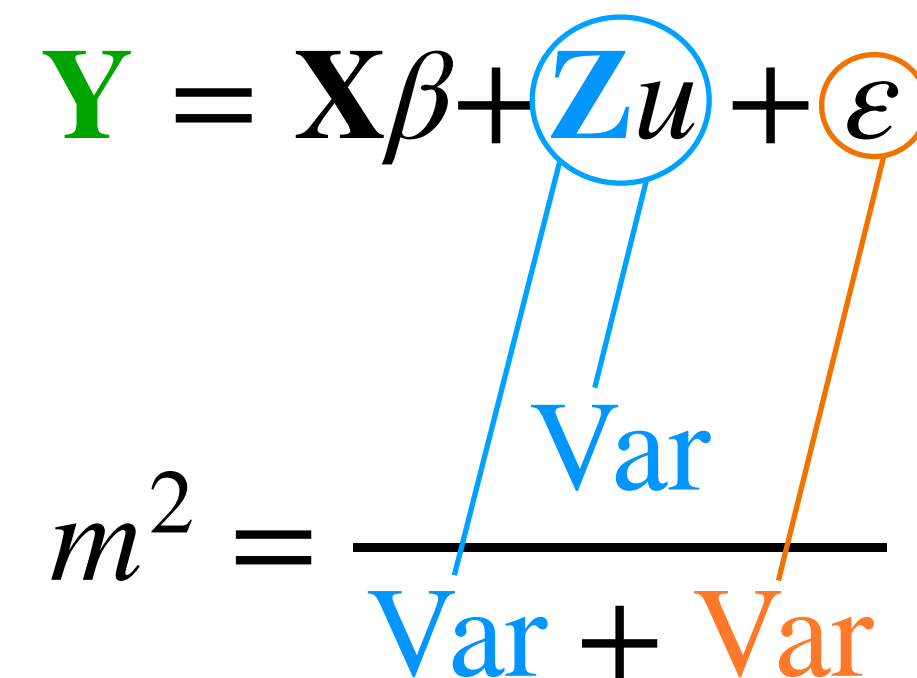
The proportion of **Fluid IQ** variation that can be explained by **MRI images of the brain**.

...

→ Morphometricity summarises the evidence.

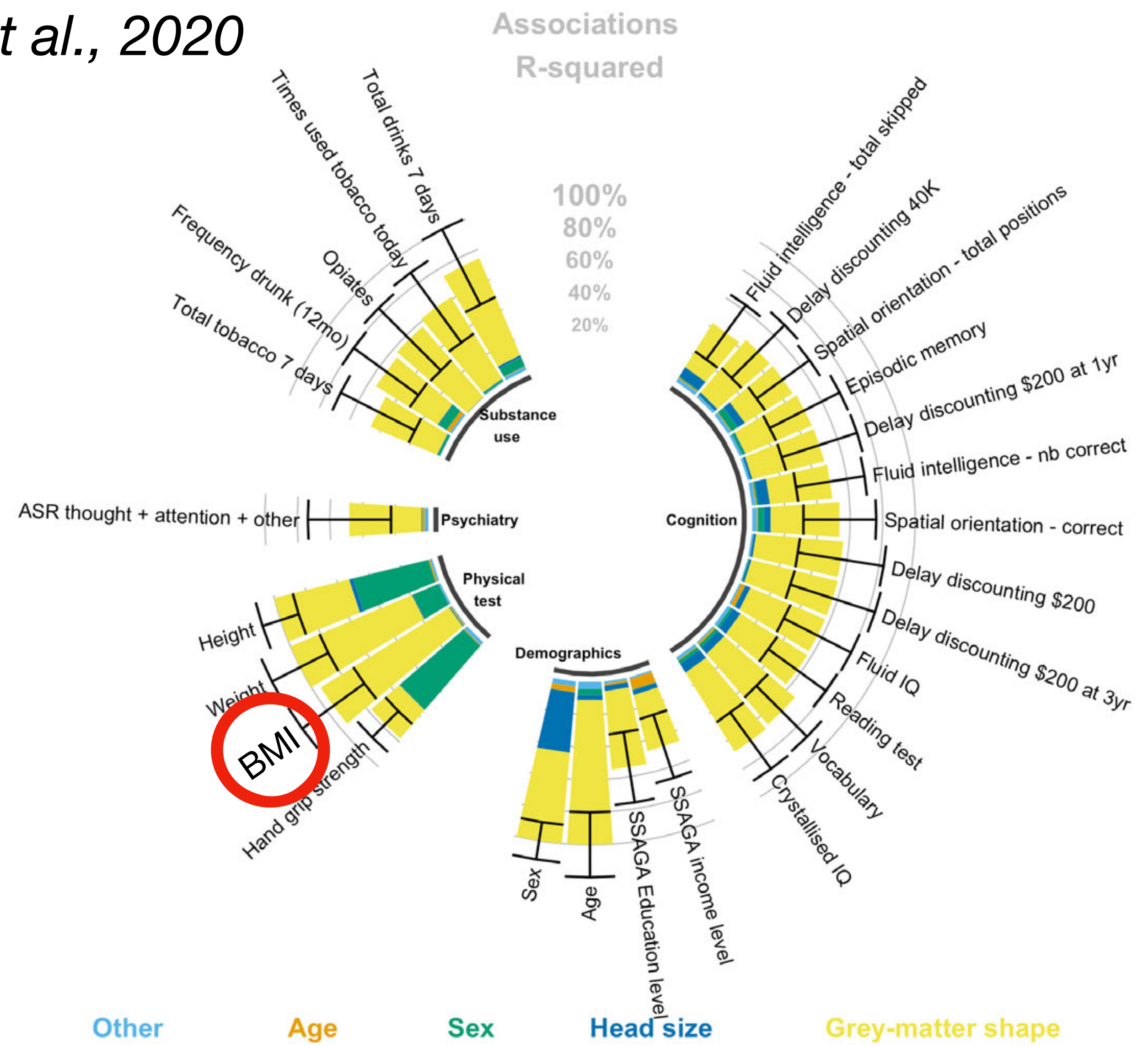
X covariates (age, sex, ...)

Z is standardized

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$
$$m^2 = \frac{\text{Var}}{\text{Var} + \text{Var}}$$


Unexpected results

Couvy-Duchesne et al., 2020



Generative vs Fitted model

$$\mathbf{Y}_{N,1} = \mathbf{X}_{N,c}\beta_{c,1} + \mathbf{Z}_{N,M}u_{M,1} + \varepsilon_{N,1}$$

N number of subjects

M number of MRI data points ≈ 150000

$$\begin{aligned} \varepsilon &\sim \mathcal{N}(0, \mathbf{I}\sigma_e^2) \\ u &\sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M) \end{aligned} \perp\!\!\!\perp$$

Generative vs Fitted model

$$\mathbf{Y}_{N,1} = \mathbf{X}_{N,c}\beta_{c,1} + \mathbf{Z}_{N,M}u_{M,1} + \varepsilon_{N,1}$$

N number of subjects

M number of MRI data points ≈ 150000

$$\begin{aligned} \varepsilon &\sim \mathcal{N}(0, \mathbf{I}\sigma_e^2) \\ u &\sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M) \quad \perp\!\!\!\perp \end{aligned}$$

$$\text{Var}(\mathbf{Z}u) = \mathbf{Z} \text{Var}(u) \mathbf{Z}^\top = \mathbf{Z} \frac{\sigma_b^2}{M} \mathbf{Z}^\top = \mathbf{B} \sigma_b^2$$

$$\mathbf{B}_{N,N} = \frac{\mathbf{Z}\mathbf{Z}^\top}{M}$$

Brain Relatedness Matrix: inter-subject correlation of the brain data

Generative vs Fitted model

$$\mathbf{Y}_{N,1} = \mathbf{X}_{N,c}\beta_{c,1} + \mathbf{Z}_{N,M}u_{M,1} + \varepsilon_{N,1}$$

N number of subjects

M number of MRI data points ≈ 150000

$$\varepsilon \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$$

$$u \sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M) \perp\!\!\!\perp$$

$$\text{Var}(\mathbf{Z}u) = \mathbf{Z} \text{Var}(u) \mathbf{Z}^\top = \mathbf{Z} \frac{\sigma_b^2}{M} \mathbf{Z}^\top = \mathbf{B} \sigma_b^2$$

$$\mathbf{B}_{N,N} = \frac{\mathbf{Z}\mathbf{Z}^\top}{M}$$

Brain Relatedness Matrix: inter-subject correlation of the brain data

Generative model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$

$$u \sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M)$$

Linear Mixed Models

$$\mathbf{Y} = \mathbf{X}\beta + b + \varepsilon$$

$$b \sim \mathcal{N}(0, \mathbf{B}\sigma_b^2)$$

Fitted model

Morphometricity: $m^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$

$\mathbf{Z}$ is standardized

The key independence hypothesis

Generative model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$

Linear Mixed
Models

Fitted model

$$\mathbf{Y} = \mathbf{X}\beta + b + \varepsilon$$

Standard assumption: independence

$$u \sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M)$$

$$\begin{aligned}\text{Var}(\mathbf{Z}u) &= \mathbf{Z} \text{Var}(u) \mathbf{Z}^\top \\ &= \mathbf{Z} \frac{\sigma_b^2}{M} \mathbf{Z}^\top = \mathbf{B}\sigma_b^2\end{aligned}$$

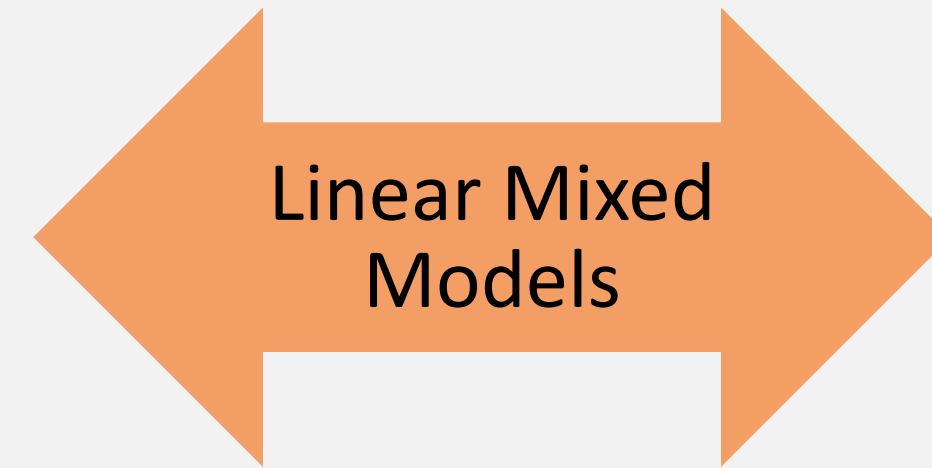
$$\begin{aligned}b &\sim \mathcal{N}(0, \mathbf{B}\sigma_b^2) \\ \mathbf{B}_{N,N} &= \frac{\mathbf{Z}\mathbf{Z}^\top}{M}\end{aligned}$$

Assumes **independence** of the voxel-specific contributions:
the contribution from one voxel is **independent** from that of the next

The key independence hypothesis

Generative model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$



Fitted model

$$\mathbf{Y} = \mathbf{X}\beta + b + \varepsilon$$

Standard assumption: independence

$$u \sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M)$$

$$\begin{aligned} \text{Var}(\mathbf{Z}u) &= \mathbf{Z} \text{Var}(u) \mathbf{Z}^\top \\ &= \mathbf{Z} \frac{\sigma_b^2}{M} \mathbf{Z}^\top = \mathbf{B}\sigma_b^2 \end{aligned}$$

$$\begin{aligned} b &\sim \mathcal{N}(0, \mathbf{B}\sigma_b^2) \\ \mathbf{B}_{N,N} &= \frac{\mathbf{Z}\mathbf{Z}^\top}{M} \end{aligned}$$

More general assumption on u

$$u \sim \mathcal{N}(0, \Sigma_u \sigma_b^2/M)$$

$$\text{Var}(\mathbf{Z}u) = \mathbf{Z} \Sigma_u \frac{\sigma_b^2}{M} \mathbf{Z}^\top$$

$$b \sim \mathcal{N}(0, \mathbf{Z} \Sigma_u \mathbf{Z}^\top \sigma_b^2/M)$$

→ incorporate **smoothness** in the voxel-specific contributions

The key independence hypothesis

Generative model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$

Linear Mixed
Models

Fitted model

$$\mathbf{Y} = \mathbf{X}\beta + b + \varepsilon$$

Standard assumption: independence

$$u \sim \mathcal{N}(0, \mathbf{I}\sigma_b^2/M)$$

$$\begin{aligned}\text{Var}(\mathbf{Z}u) &= \mathbf{Z} \text{Var}(u) \mathbf{Z}^\top \\ &= \mathbf{Z} \frac{\sigma_b^2}{M} \mathbf{Z}^\top = \mathbf{B}\sigma_b^2\end{aligned}$$

$$\begin{aligned}b &\sim \mathcal{N}(0, \mathbf{B}\sigma_b^2) \\ \mathbf{B}_{N,N} &= \frac{\mathbf{Z}\mathbf{Z}^\top}{M}\end{aligned}$$

More general assumption on u

$$u \sim \mathcal{N}(0, \Sigma_u \sigma_b^2/M)$$

$$\text{Var}(\mathbf{Z}u) = \mathbf{Z} \Sigma_u \frac{\sigma_b^2}{M} \mathbf{Z}^\top$$

$$b \sim \mathcal{N}(0, \mathbf{Z} \Sigma_u \mathbf{Z}^\top \sigma_b^2/M)$$

Stationary smooth noise process assumption

$$\Sigma_u = c_f \mathbf{K}_f \mathbf{K}_f^\top, \quad c_f = \frac{M}{\text{Tr}(\mathbf{K}_f \mathbf{K}_f^\top)}$$

$M \times M$ smoothing matrix, FWHM f

$$b \sim \mathcal{N}(0, \mathbf{B}_f \sigma_b^2 c_f)$$

$$\mathbf{B}_f = \frac{(\mathbf{Z}\mathbf{K}_f)(\mathbf{Z}\mathbf{K}_f)^\top}{M}$$

Simulations

Generative model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}u + \varepsilon$$

Linear Mixed
Models

Fitted model

$$\mathbf{Y} = \mathbf{X}\beta + b + \epsilon$$

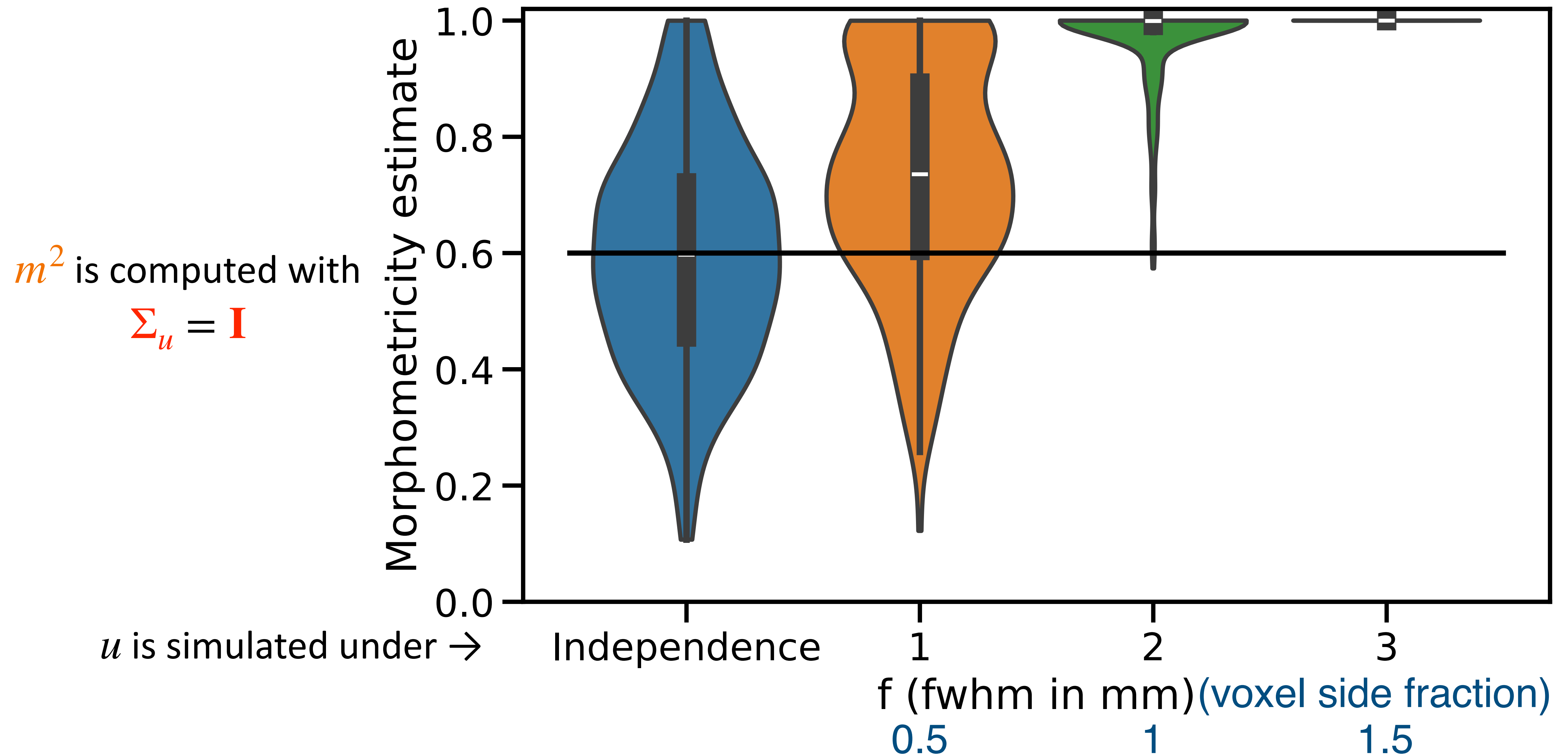
real $\mathbf{Z}$ N : 500 subjects from the UK Biobank (*Miller et al., 2016*) $\rightarrow$ standardize
 M : 154055 voxel values from gray matter VBM images

$\rightarrow$ compute $\text{Var}(b)$ under the different assumptions

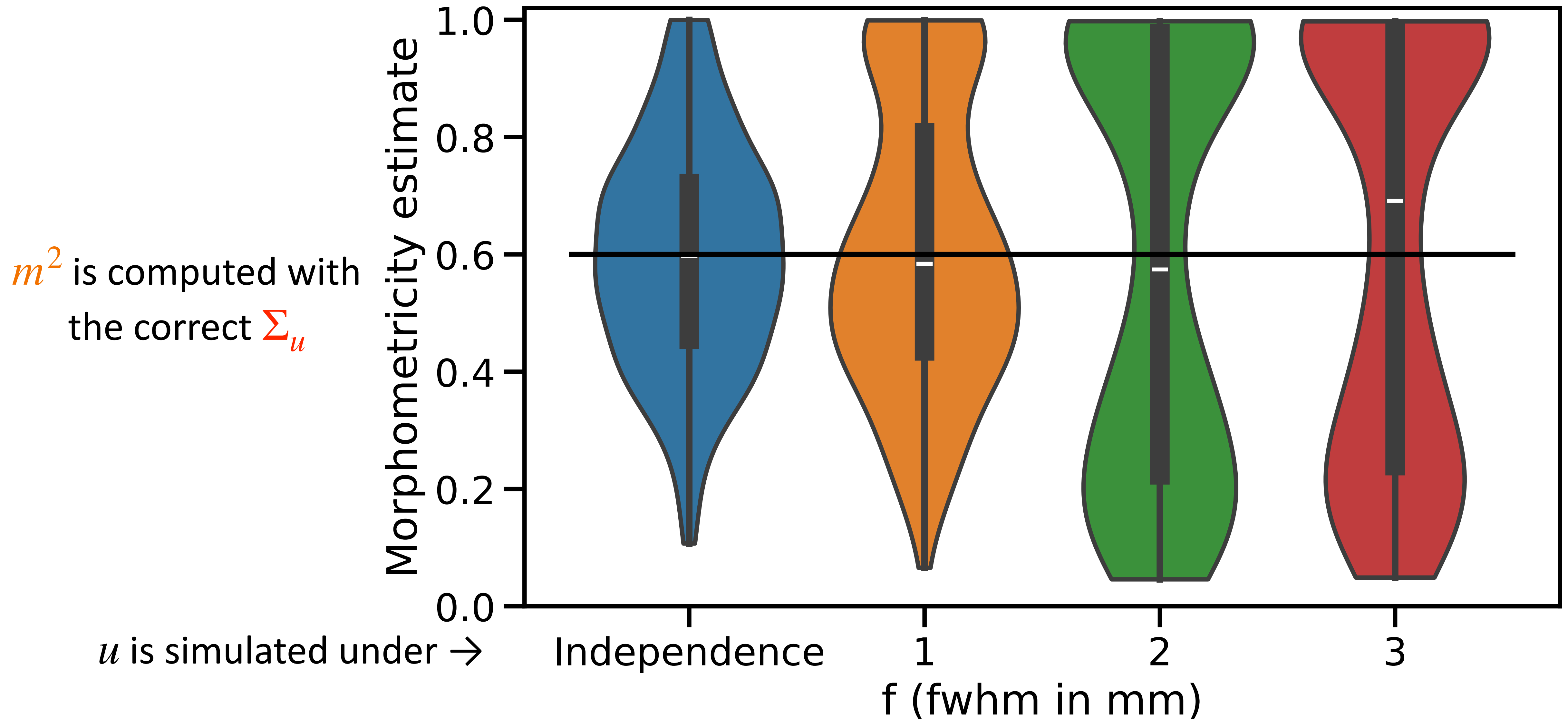
simulate ε 250 times
simulate u 250 times for each scenario $\rightarrow$ compute $\mathbf{Y}$

Estimate the fitted model with R's lme4 (*Therneau, 2012*) and compute m^2

Simulation: Impact of Model Misspecification



Simulation: Accounting for Smoothness



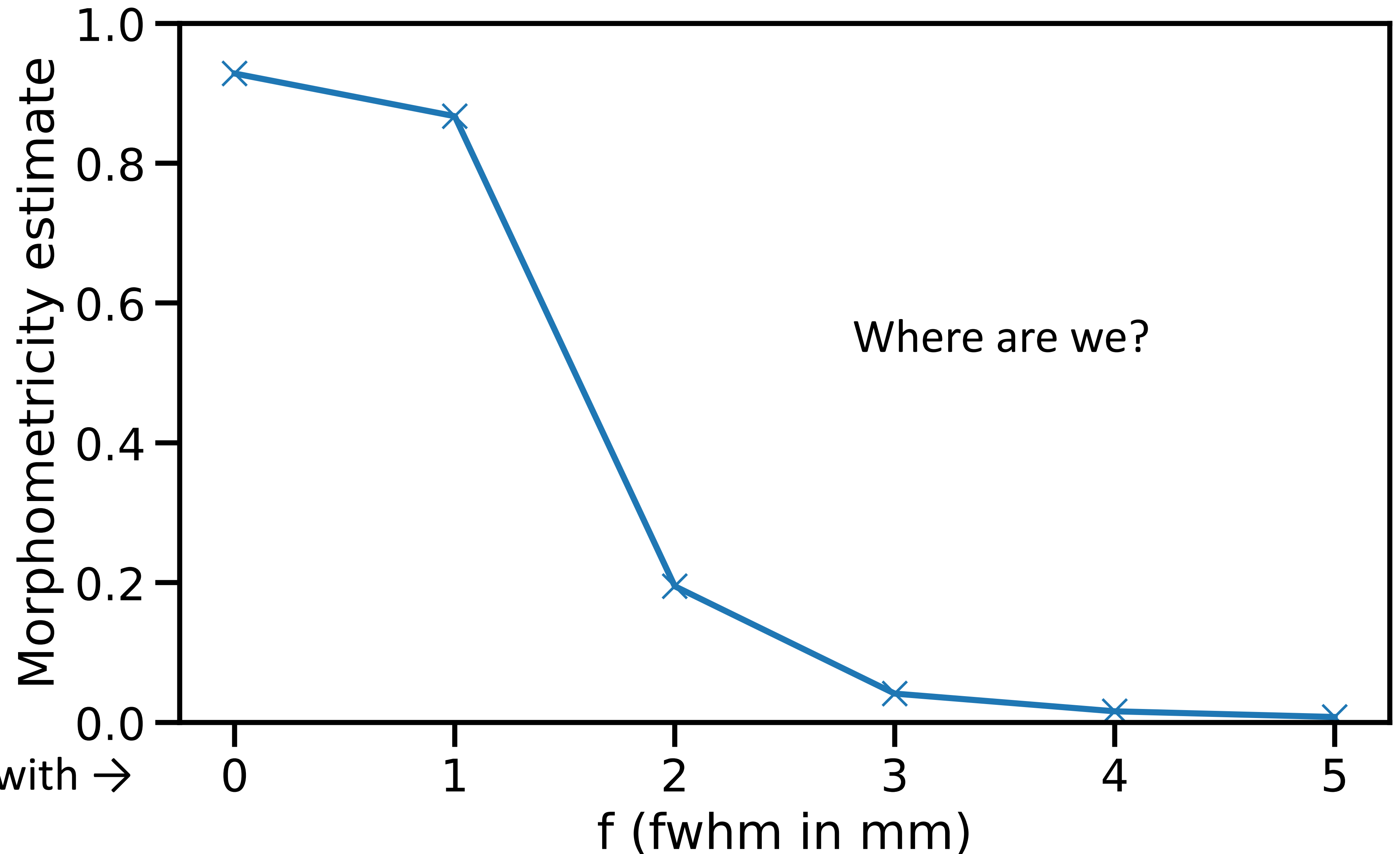
Real data

Z: UKB-VBM

Y: BMI

X: Sex, Age, ICV,
Date, Center, Head
motion

m^2 is computed with $\rightarrow$



Conclusions

- When latent effects of brain measures u are assumed independent, **morphometricity is positively biased.**
- If the true correlation structure is known, an unbiased morphometricity could be recovered but it appears very unstable, and learning the true correlation on real data may not be practical.
- These considerations suggest that morphometricity should not be used until new methods are developed that account for the impact of smoothness.

Thank you!

References

Couvy-Duchesne, B. et al. (2020), 'A unified framework for association and prediction from vertex-wise grey- matter structure', Human brain mapping, vol. 41, no 14, p. 4062-4076.

Miller, K. L. et al. (2016), 'Multimodal population brain imaging in the UK Biobank prospective epidemiological study', Nature neuroscience, vol. 19, no 11, p. 1523-1536.

Sabuncu, M. R. et al. (2016), 'Morphometricity as a measure of the neuroanatomical signature of a trait', Proceedings of the National Academy of Sciences, vol. 113, no. 39, p. E5749-E5756.

Therneau, T. (2012), 'The lme4 function', Rochester, MN: Mayo Clinic.