# Multiple testing in latent variable models

Brice Ozenne[1,2], Esben Budtz-Jørgensen[2], Sebastian Elgaard Ebert[1]

[1] Neurobiology Reasearch Unit, Rigshospitalet, Copenhagen, Denmark.
[2] University of Copenhagen, Section of Biostatistics, Copenhagen, Denmark.

ISCB40 - July 16th 2019

# Typical neuroimaging study (Ebert et al., 2019)

Association study:

- After a mild traumatic brain injury,
  is there a neuroinflammatory response in the brain?

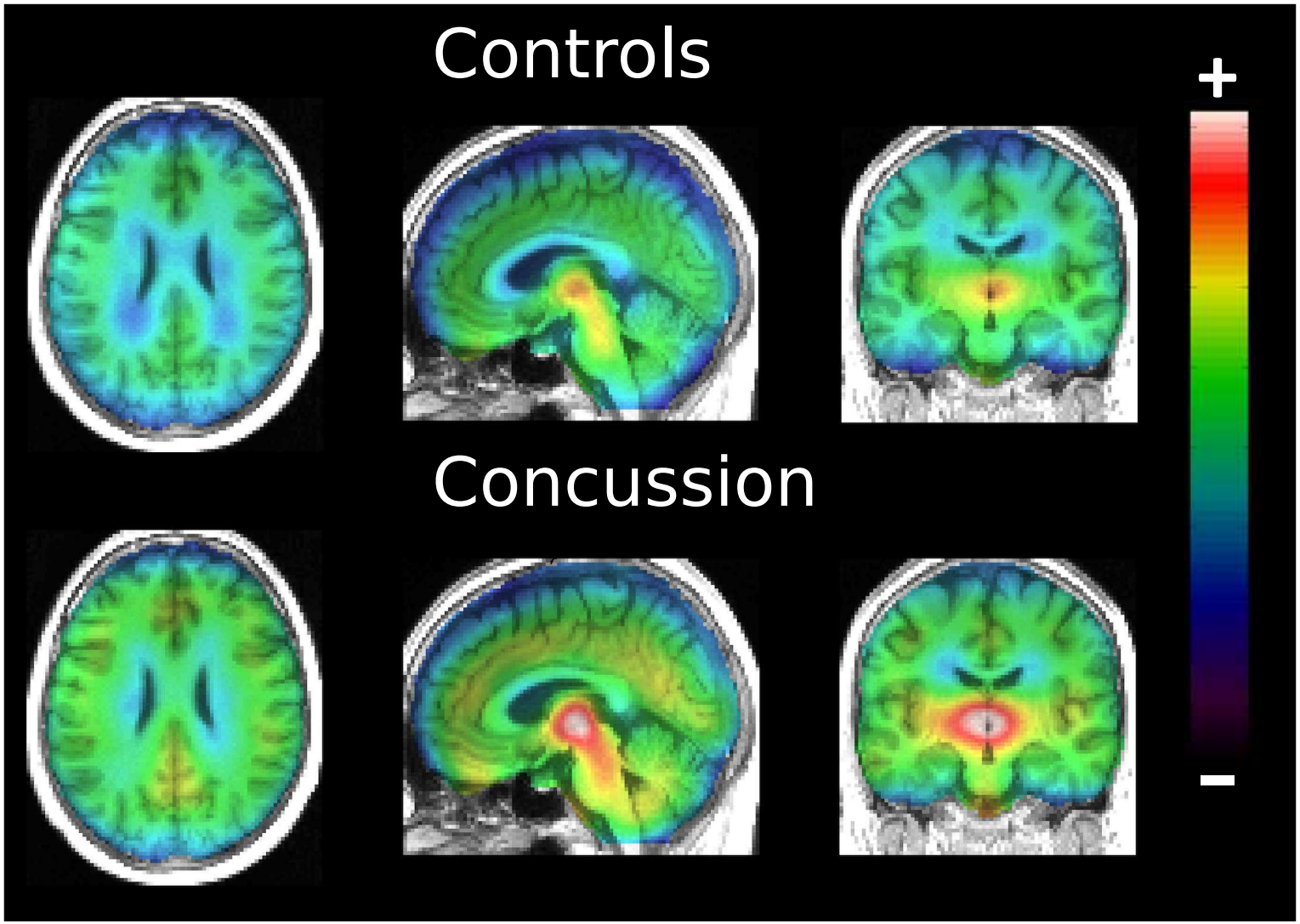
Many measurements, few subjects:

- SPECT scan: TSPO measurements over the whole brain
- 22 healthy controls and 14 patients

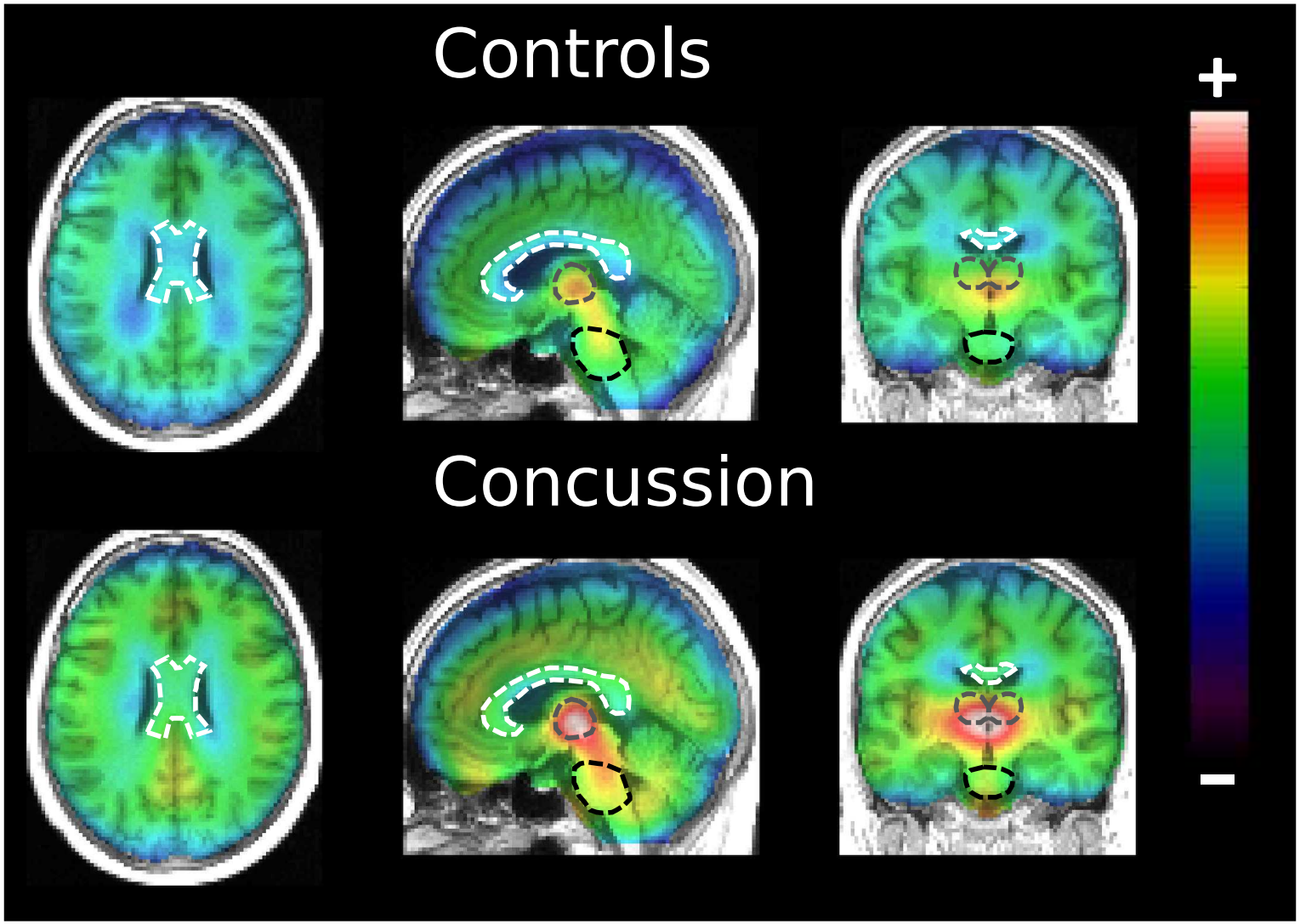
Often, some complications:

- genetic factors can influence the TSPO measurements

# Processed data "averaged" over individuals



# Dimension reduction: 9 ROIs




Corpus Callosum (white), thalamus (grey), and pons (black).

# Latent variable models (LVM)

We have:

- measurements in 9 brain regions $\boldsymbol{Y} = (Y_1, \ldots, Y_9)$
- 2 covariates $\boldsymbol{X} = (X_1 = group, X_2 = gene)$
- 1 latent variables $\eta$ modeling the individual TSPO level

# Latent variable models (LVM)

We have:

- measurements in 9 brain regions $\boldsymbol{Y} = (Y_1, \ldots, Y_9)$
- 2 covariates $\boldsymbol{X} = (X_1 = group, X_2 = gene)$
- 1 latent variables $\eta$ modeling the individual TSPO level

We define our LVM using

- a measurement model:

$$\forall r \in \{1, \ldots, 9\}, \qquad Y_{ir} = \nu_r + \eta_i \lambda_r + \boldsymbol{X}_i K_r + \varepsilon_{ir}$$

where $(\varepsilon_{i1}, \ldots, \varepsilon_{i9}) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_9} \end{bmatrix}\right)$

- a structural model:

$$\eta_i = \alpha + \zeta_i, \text{ where } \zeta_i \sim \mathcal{N}\left(0, \sigma^2_\zeta\right)$$

# Latent variable models (LVM)

We have:

- measurements in 9 brain regions $\boldsymbol{Y} = (Y_1, \ldots, Y_9)$
- 2 covariates $\boldsymbol{X} = (X_1 = group, X_2 = gene)$
- 1 latent variables $\eta$ modeling the individual TSPO level

We define our LVM using

- a measurement model:

$$\forall r \in \{1, \ldots, 9\}, \qquad Y_{ir} = \nu_r + \eta_i \lambda_r + \boldsymbol{X}_i K_r + \varepsilon_{ir}$$

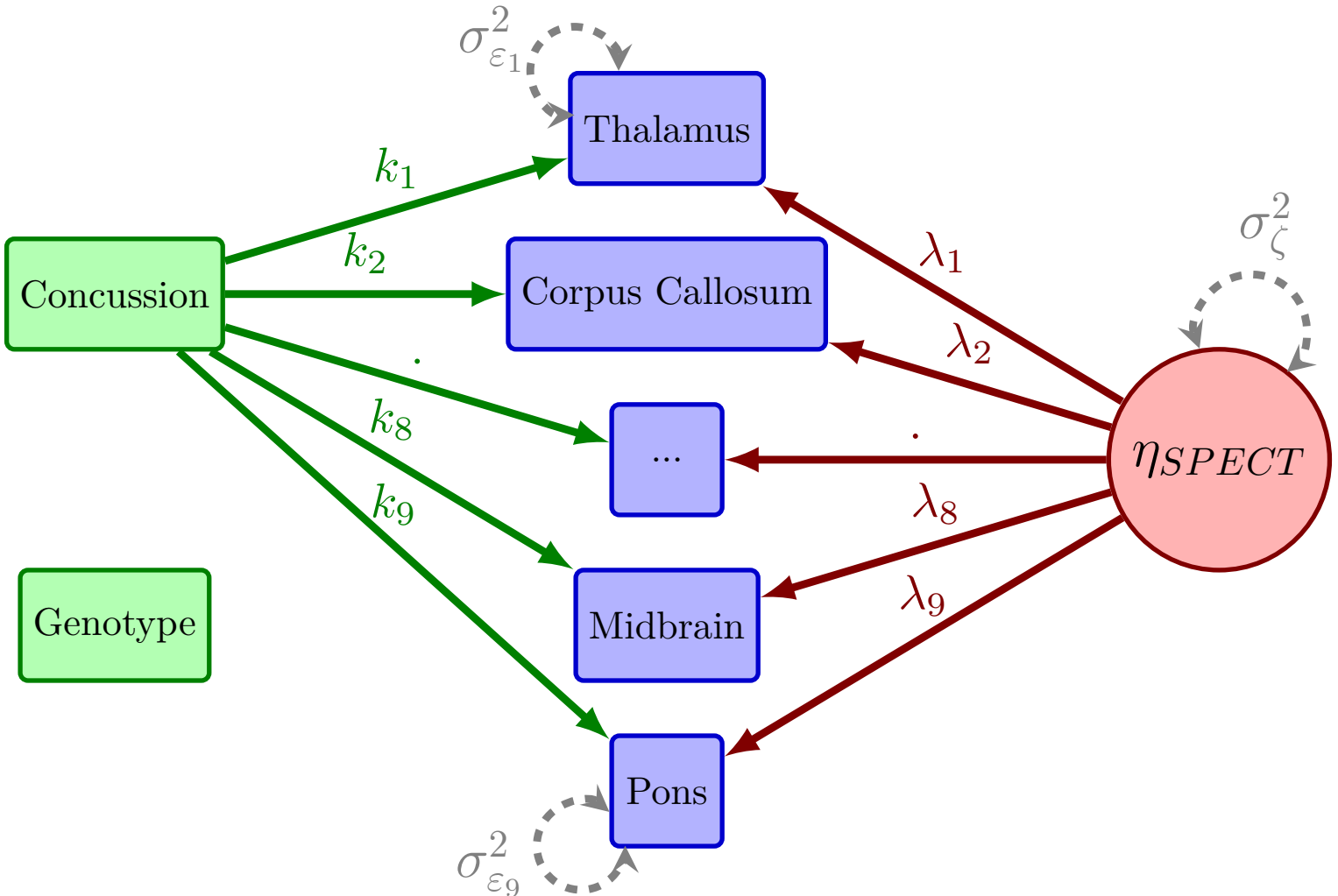
where $(\varepsilon_{i1}, \ldots, \varepsilon_{i9}) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_9} \end{bmatrix}\right)$

- a structural model:

$$\eta_i = \alpha + \zeta_i, \text{ where } \zeta_i \sim \mathcal{N}\left(0, \sigma^2_\zeta\right)$$

- identifiability constrains, e.g. $\nu_1 = 0$, $\lambda_1 = 1$.

# Path diagram




$$\mathbb{C}ov\left[Y_{ir}, Y_{ir*}\right] = \begin{cases} \lambda_r^2\sigma_\zeta^2 + \sigma_r^2 \text{ if } r = r* \\ \lambda_r\lambda_{r*}\sigma_\zeta^2 \text{ if } r \neq r* \end{cases}$$

# Clinical hypotheses

Is there any effect at all of concussion?

$$(\mathcal{H}_0^a) : k_1 = k_2 = \ldots = k_9 = 0$$

In which region is there an effect of concussion?

$$(\mathcal{H}_0^{b1}) : k_1 = 0$$
$$\ldots$$
$$(\mathcal{H}_0^{b9}) : k_9 = 0$$

# Test 1 clinical hypothesis

Easy. Estimate the LVM by maximum likelihood (ML):

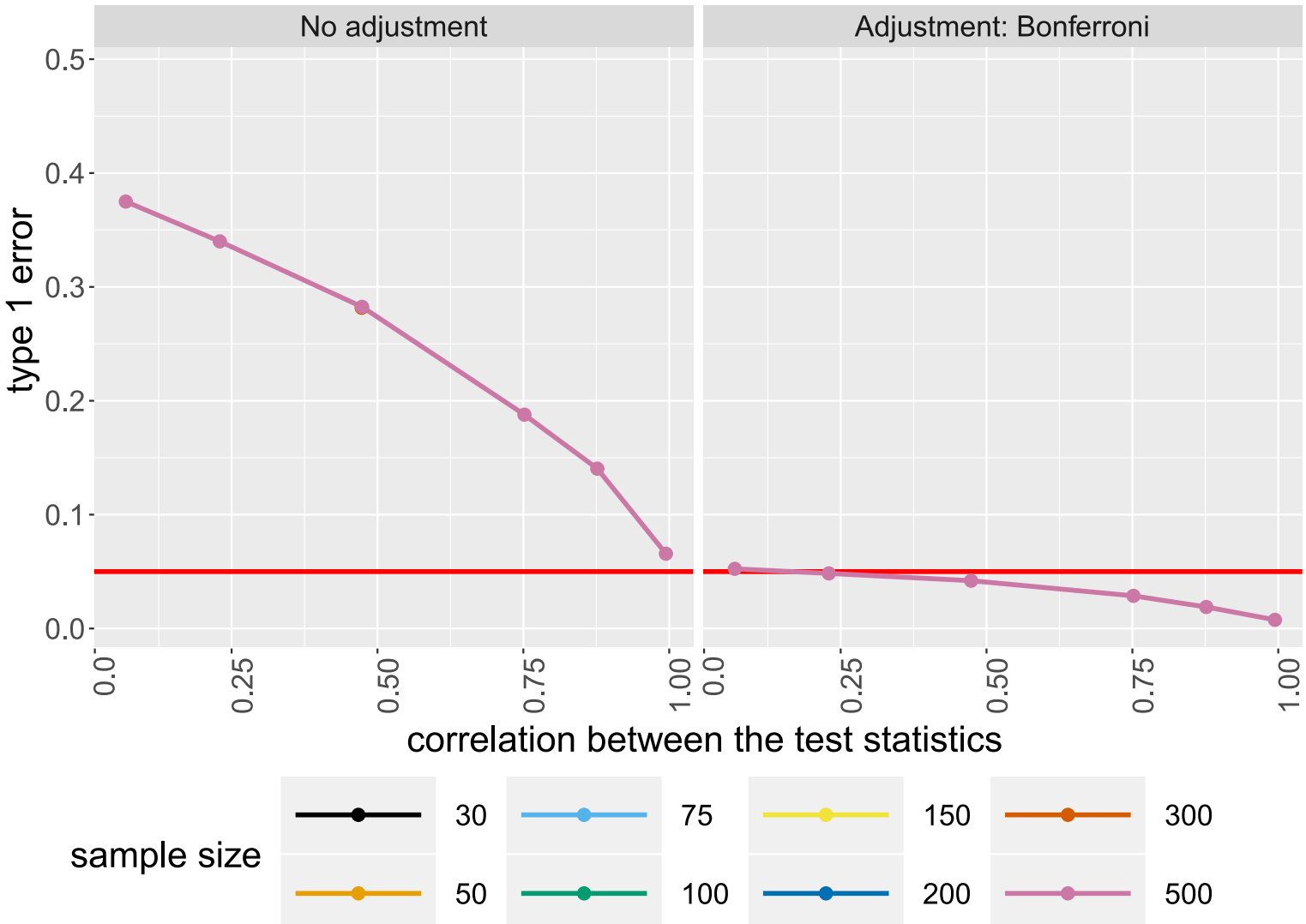
- $\hat{\Theta}$: ML estimate of the model parameters
- $\hat{\Sigma}_{\hat{\Theta}}$: Estimate of the variance-covariance of the model parameters
- $C$: Contrast matrix

Wald test:

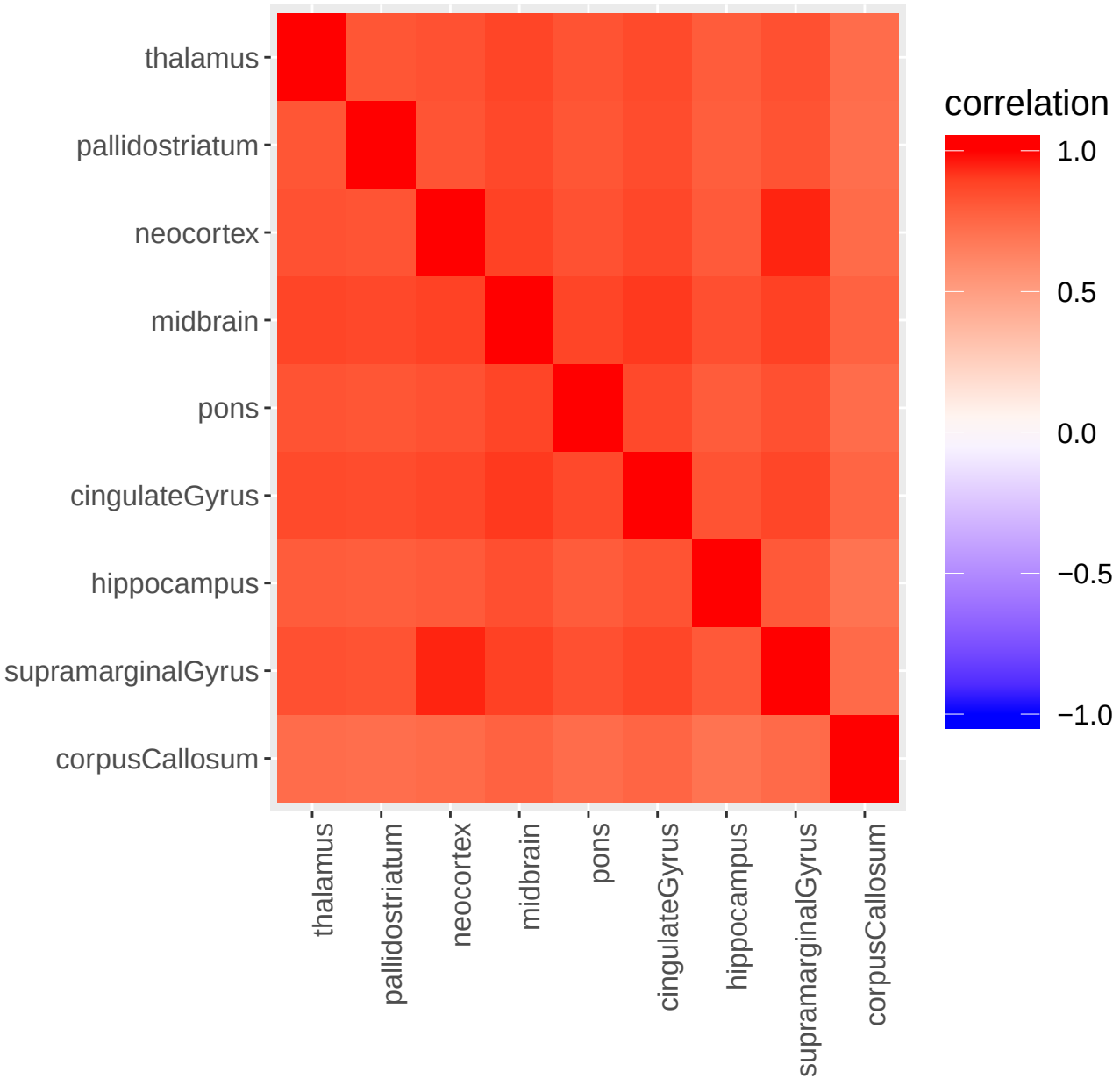
$$\left(C\hat{\Theta}\right)^{\mathsf{T}}\left(C\hat{\Sigma}_{\hat{\Theta}}C^{\mathsf{T}}\right)^{-1}\left(C\hat{\Theta}\right) \xrightarrow[n\to\infty]{} \chi^2_r$$

```
data: chisq = 24.193, df = 9, p-value = 0.004006
```

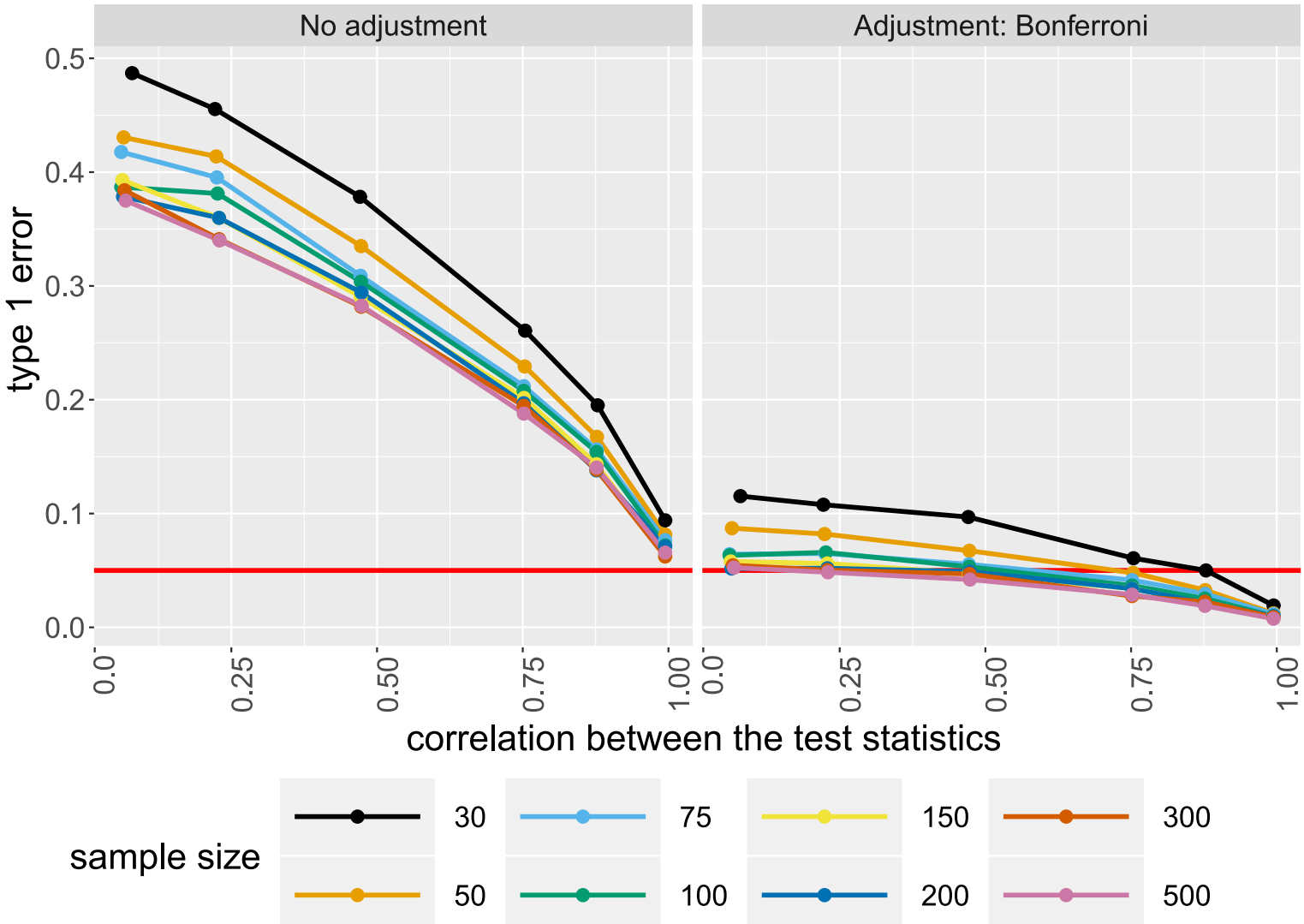
# Test several clinical hypotheses



# Test several clinical hypotheses



# Test several clinical hypotheses



# Handling multiple comparisons

# Notations

We consider the contrast matrix $C$ such that we want to test:

$$C\Theta = \mathbf{0} \qquad = (k_1, \ldots, k_9)$$

and denote the vector of Wald statistics by:

$$\mathbf{T} = diag(C\Sigma_{\hat{\Theta}}C^\intercal)^{-\frac{1}{2}}C\Theta \qquad = \left(\frac{k_1}{\sigma_{k_1}}, \ldots, \frac{k_9}{\sigma_{k_9}}\right)$$

From maximum likelihood theory, we know that:

$$\sqrt{n}\left(\hat{\Theta} - \Theta\right) \overset{d}{\sim} \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

where $\left(n\mathcal{I}_1(\Theta)\right)^{-1} = \Sigma_{\hat{\Theta}}$ in correctly specified models.

# Max test procedure (Hothorn et al., 2008)

The vector of Wald statistics is asymptotically normally distributed:

$$\sqrt{n}\,\mathbf{T} \underset{\mathcal{H}_0}{\overset{d}{\sim}} \mathcal{N}\left(0, \Sigma_{\mathbf{T}}\right)$$

with $\Sigma_{\mathbf{T}} = f(C, \mathcal{I}_1(\Theta))$

# Max test procedure (Hothorn et al., 2008)

The vector of Wald statistics is asymptotically normally distributed:

$$\sqrt{n}\,\mathbf{T} \underset{\mathcal{H}_0}{\overset{d}{\sim}} \mathcal{N}(0, \Sigma_{\mathbf{T}})$$

with $\Sigma_{\mathbf{T}} = f(C, \mathcal{I}_1(\Theta))$

We define

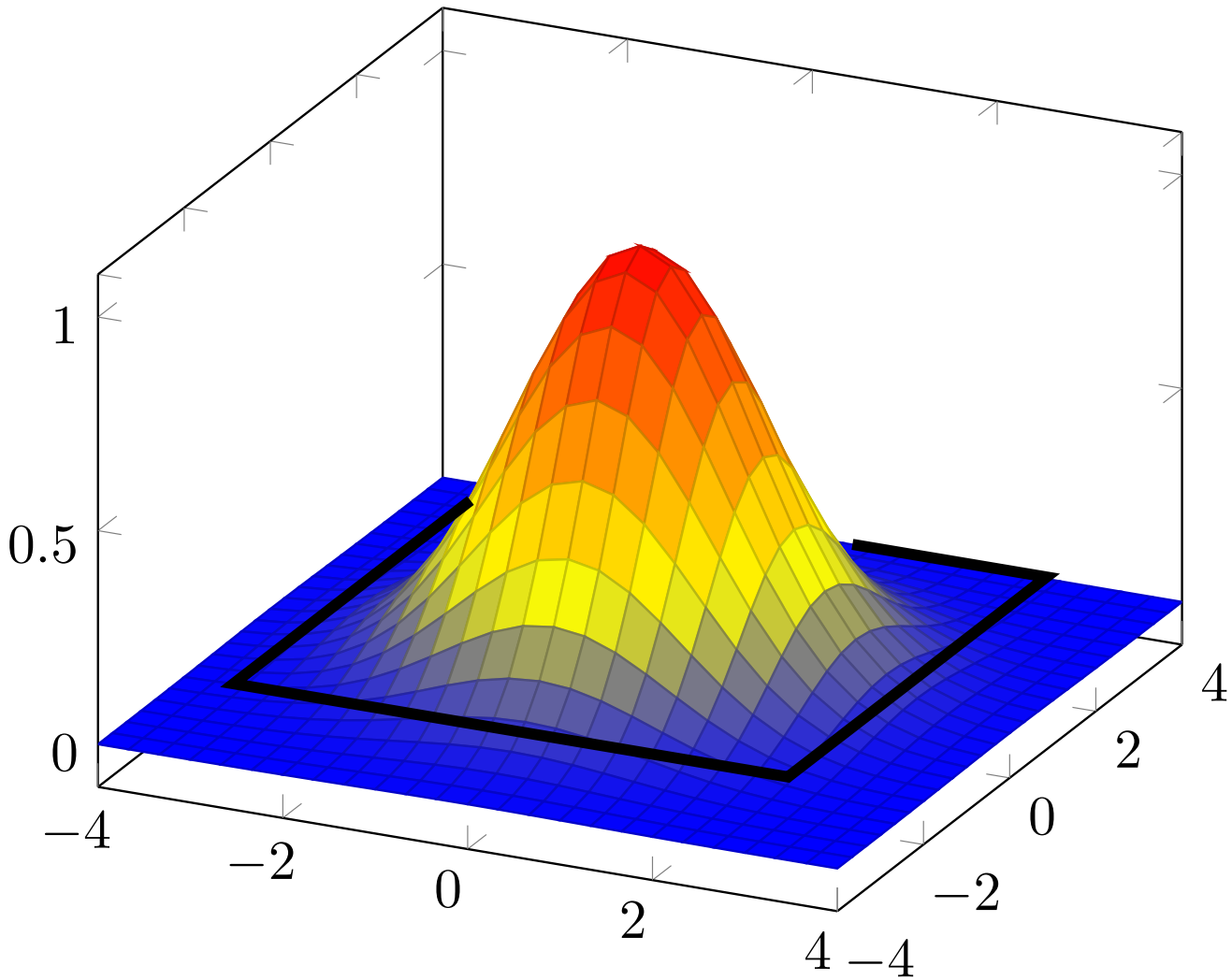
- the max statistic: $|T|_{max} = max(\mathbf{T})$
- the observed max statistic: $|t|_{max}$ (e.g. $t_1$)

We obtain an adjusted p-value for the largest observed $T$-statistic by computing (under the null):

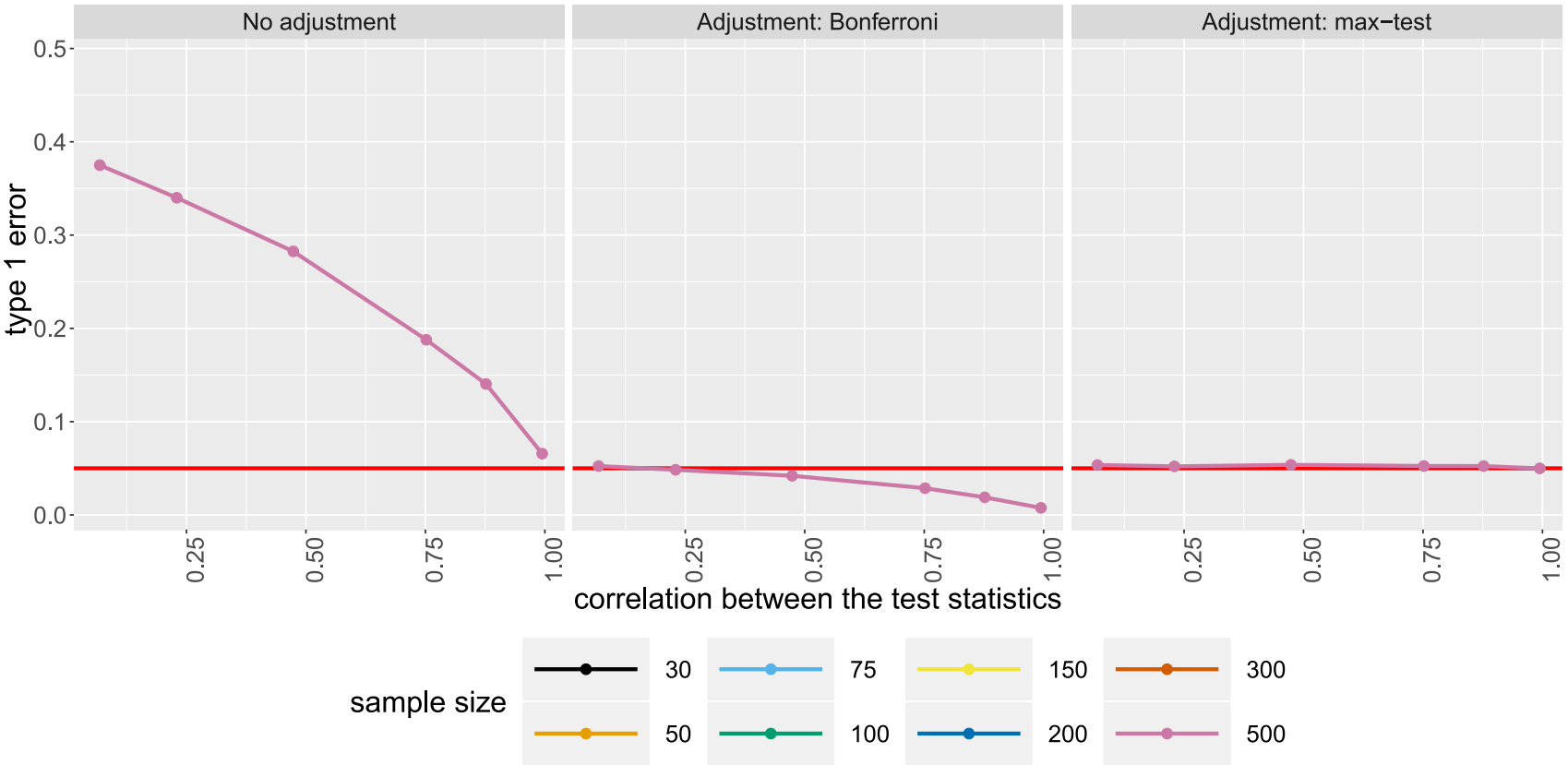
$$1 - \mathbb{P}\left[|T|_{max} < |t|_{max}\right]$$

# Computation of $\mathbb{P}\left[|T|_{max} < |t|_{max}\right]$ using `mvtnorm`

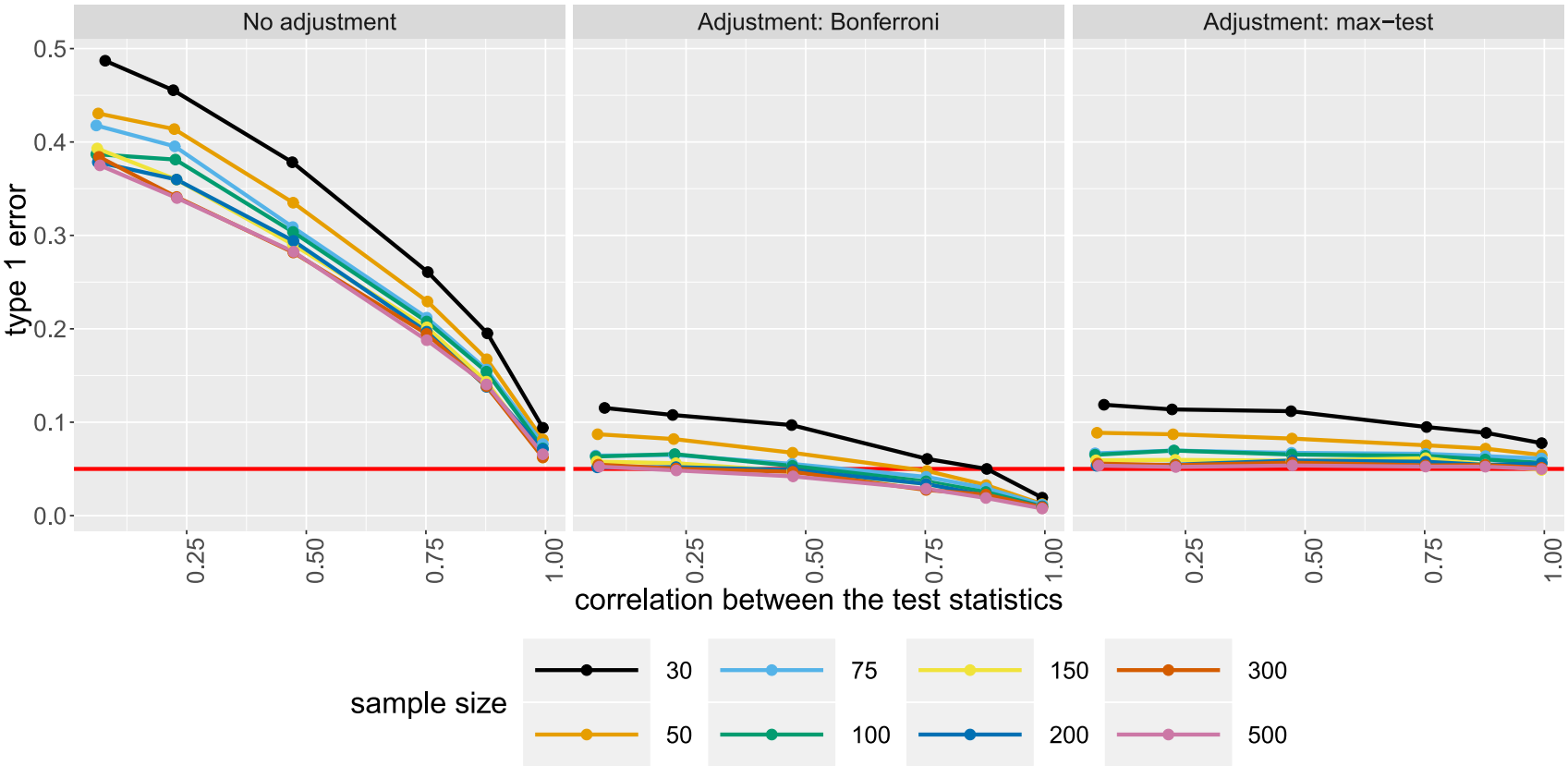
Black thick line: $|t|_{max}$



# Simulation study



# Simulation study



# Handling small sample size

# Back to univariate linear regression

For the Wald test, asymptotically:

$$\frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}} = \frac{\hat{\theta}}{\sqrt{(X^\intercal X)^{-1}\hat{\sigma}^2}} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0,1)$$

where $\hat{}$ stands for the maximum likelihood estimate (MLE).

# Back to univariate linear regression

For the Wald test, asymptotically:

$$\frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}} = \frac{\hat{\theta}}{\sqrt{(X^\intercal X)^{-1}\hat{\sigma}^2}} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0,1)$$

where $\hat{}$ stands for the maximum likelihood estimate (MLE).

Standard corrections:

(A) correct finite sample bias of MLE:

$$\hat{\sigma}^{2,c} = \frac{n}{n-p}\hat{\sigma}^2$$

(B) use a t-distribution.

(C) correct the degrees of freedom: $n - p$ instead of $n$.

# (A) bias correction in LVM (1/2)

Denoting the observed residuals:

$$\xi_i(\hat{\Theta}) = \boldsymbol{Y}_i - \mu(\hat{\Theta})$$

their variance is smaller than the (true) conditional variance of $Y$:

$$\mathbb{E}\left[\xi_i(\hat{\Theta})^\intercal \xi_i(\hat{\Theta})\right] = \Omega(\Theta) - \Psi_i + o_p(n^{-1})$$

where $\Psi_i$ is the first order bias:

$$\Psi_i = \frac{\partial \mu(\Theta)}{\partial \Theta}^\intercal \Sigma_{\hat{\Theta}} \frac{\partial \mu(\Theta)}{\partial \Theta}$$

and $\mu(\Theta) = \mathbb{E}\left[\boldsymbol{Y}|\boldsymbol{X}\right]$, $\Omega(\Theta) = \mathbb{V}ar\left[\boldsymbol{Y}|\boldsymbol{X}\right]$.

# (A) bias correction in LVM (1/2)

Denoting the observed residuals:

$$\xi_i(\hat{\Theta}) = \boldsymbol{Y}_i - \mu(\hat{\Theta})$$

their variance is smaller than the (true) conditional variance of $Y$:

$$\mathbb{E}\left[\xi_i(\hat{\Theta})^\intercal \xi_i(\hat{\Theta})\right] = \Omega(\Theta) - \Psi_i + o_p(n^{-1})$$

where $\Psi_i$ is the first order bias:

$$\Psi_i = \frac{\partial \mu(\Theta)}{\partial \Theta}^\intercal \Sigma_{\hat{\Theta}} \frac{\partial \mu(\Theta)}{\partial \Theta}$$

and $\mu(\Theta) = \mathbb{E}\left[\boldsymbol{Y}|\boldsymbol{X}\right]$, $\Omega(\Theta) = \mathbb{V}ar\left[\boldsymbol{Y}|\boldsymbol{X}\right]$.
**Example**: $Y_i = \beta X_i + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$

$$\hat{\Psi}_i = \hat{\sigma}^2 \frac{p}{n}$$

# (A) bias correction in LVM (2/2)

Assuming that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\hat{\Theta})^{\intercal}\xi_i(\hat{\Theta})\right]$ and $\Omega(\Theta)$ are subject to the same type of bias

- we can use $\hat{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_i$ to correct $\hat{\Omega}(\Theta)$

# (A) bias correction in LVM (2/2)

Assuming that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\hat{\Theta})^{\mathsf{T}}\xi_i(\hat{\Theta})\right]$ and $\Omega(\Theta)$ are subject to the same type of bias

- we can use $\hat{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_i$ to correct $\hat{\Omega}(\Theta)$
- and the new $\hat{\Omega}(\Theta)$ to better estimate $\Sigma_{\hat{\Theta}}$ and $\Psi$

$$\Psi_i(\Omega(\Theta)) = \frac{\partial\mu(\Theta)}{\partial\Theta}^{\mathsf{T}}\Sigma_{\hat{\Theta}}(\Omega(\Theta))\frac{\partial\mu(\Theta)}{\partial\Theta}$$

# (A) bias correction in LVM (2/2)

Assuming that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i(\hat{\Theta})^\intercal\xi_i(\hat{\Theta})\right]$ and $\Omega(\Theta)$ are subject to the same type of bias

- we can use $\hat{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Psi}_i$ to correct $\hat{\Omega}(\Theta)$
- and the new $\hat{\Omega}(\Theta)$ to better estimate $\Sigma_{\hat{\Theta}}$ and $\Psi$

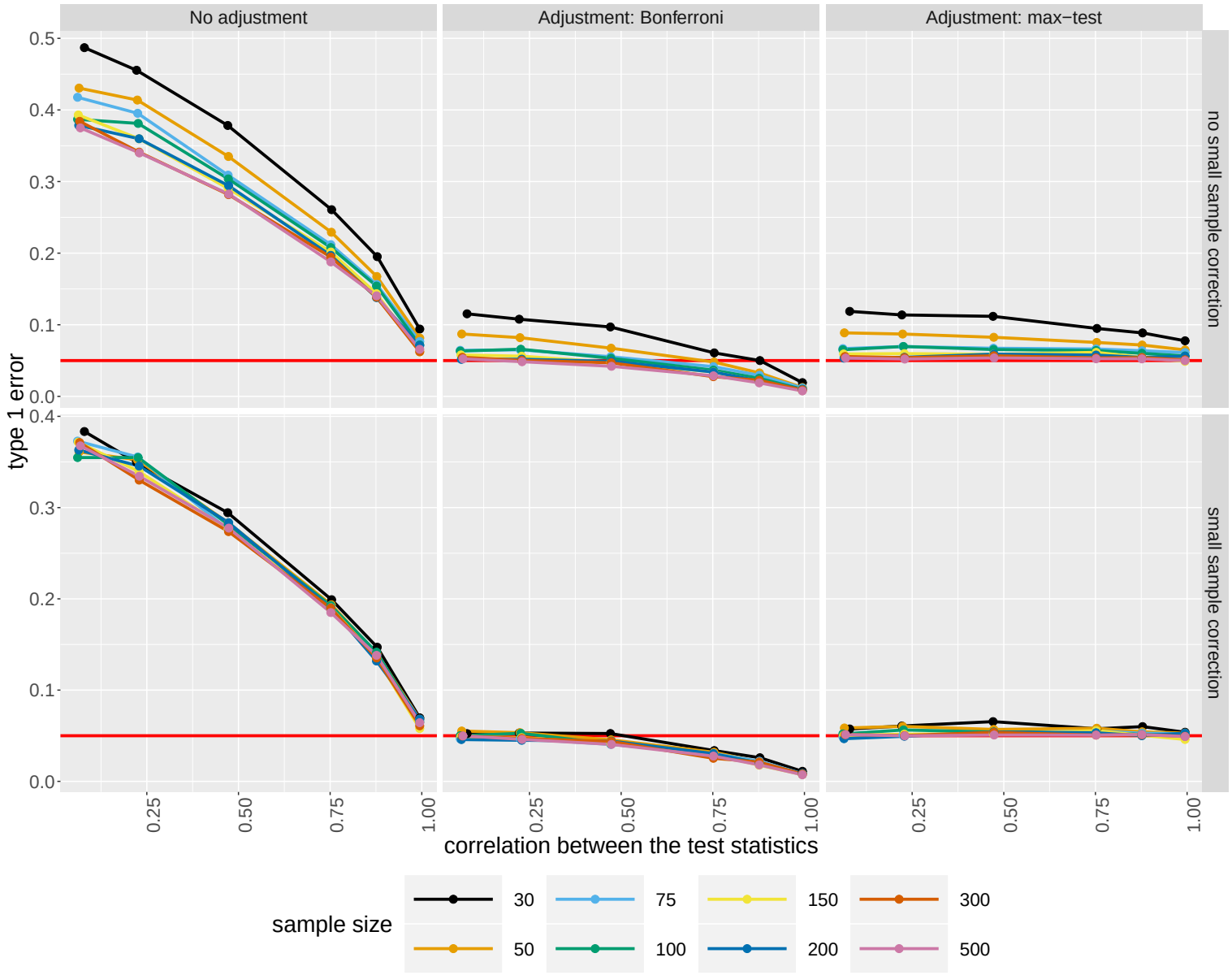
**Example**: $Y_i = \beta X_i + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

$$\hat{\Psi}_i = \hat{\sigma}^2\frac{p}{n} \text{ and } \tilde{\sigma}^2 = \hat{\sigma}^2 + \hat{\Psi} = \left(1 + \frac{p}{n}\right)\hat{\sigma}^2$$

Iterating the procedure gives:

$$\tilde{\sigma}^{\infty} = \sum_{k=0}^{\infty}\left(\frac{p}{n}\right)^k\hat{\sigma}^2 = \frac{n}{n-p}\hat{\sigma}^2$$

# Simulation study



# Application

| | no correction | permutations [1] | small sample correction |
|---|---|---|---|
| $(\mathcal{H}_0^a)$: p-value | 0.004 | 0.0166 | 0.011 |

| | | p-value (with small sample correction) | | |
|---|---|---|---|---|
| region | effect (%) | unadjusted | Bonferroni | max-test |
| thalamus | 12.75 | 0.23 | 1.00 | 0.53 |
| pallidostriatum | 12.03 | 0.18 | 1.00 | 0.43 |
| neocortex | 4.38 | 0.60 | 1.00 | 0.97 |
| midbrain | 10.40 | 0.22 | 1.00 | 0.51 |
| pons | 1.56 | 0.86 | 1.00 | 1.00 |
| cingulate gyrus | 17.28 | 0.03 | 0.30 | 0.11 |
| hippocampus | 12.64 | 0.14 | 1.00 | 0.36 |
| supramarginal gyrus | 5.22 | 0.55 | 1.00 | 0.94 |
| corpus callosum | 19.02 | 0.03 | 0.23 | 0.09 |

[1] 10 000 samples, false cv 0.13%, CPU time $\approx$ 1h

# Conclusion

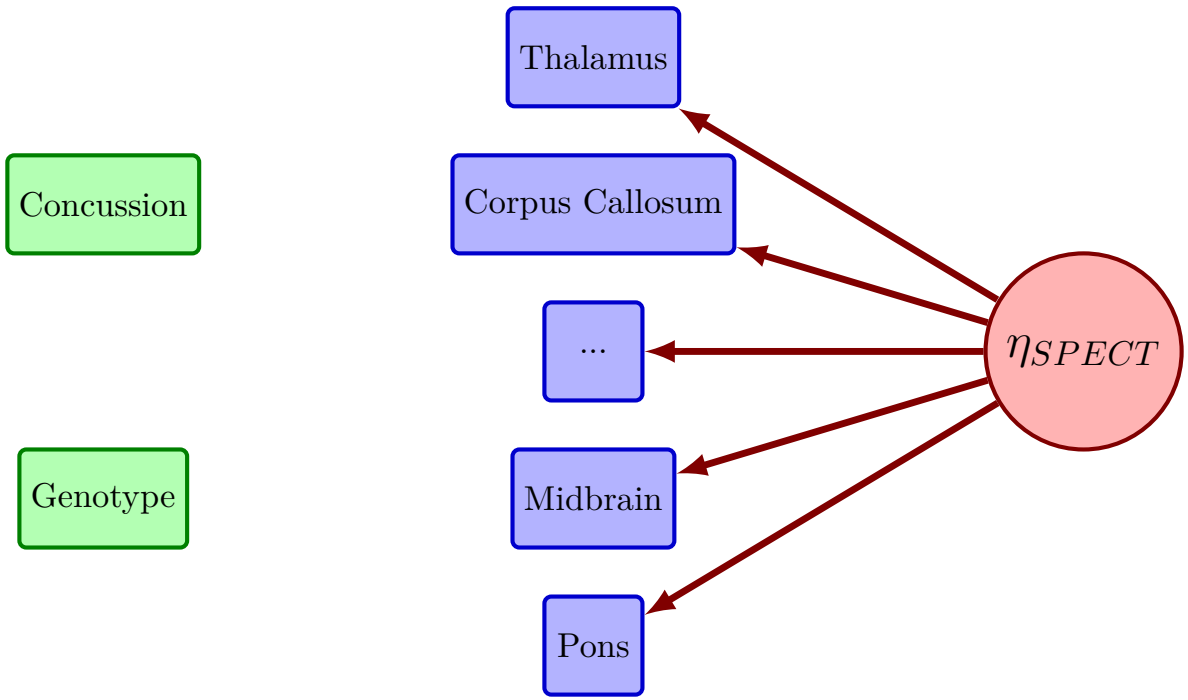
LVM are a flexible framework to analyze regional imaging data:

- R package `lava`

We propose inference tools in the R package `lavaSearch2`:

- inference in small samples (not perfect but better than `lava`)
- adjustment for multiple comparisons (via `mvtnorm`)

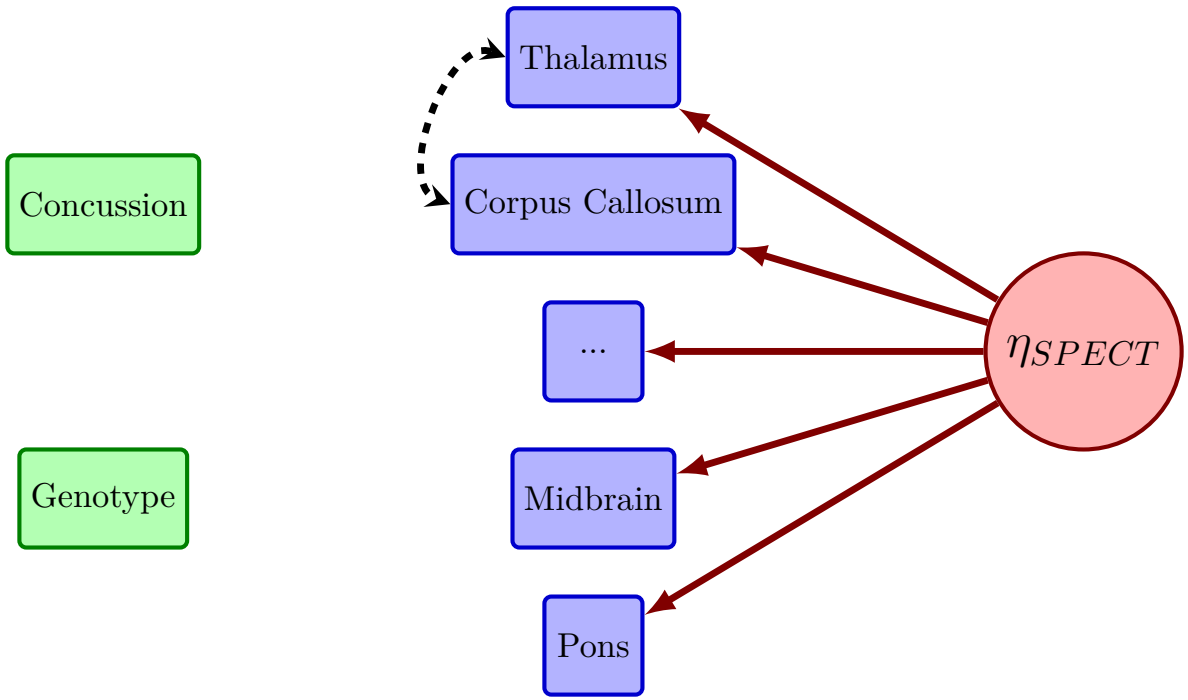
# Extention




Investigate model misspecification using score tests

- adjustment for multiple comparisons
- max-test procedure for score statistics

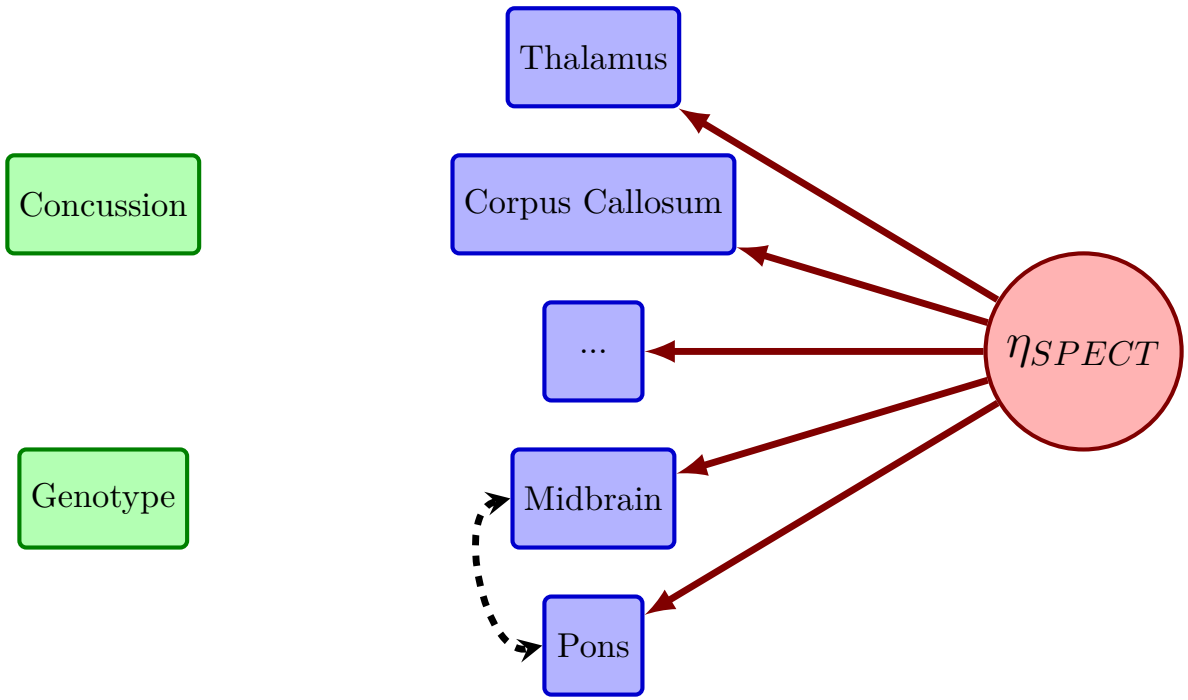
# Extention




Investigate model misspecification using score tests

- adjustment for multiple comparisons
- max-test procedure for score statistics

# Extention




Investigate model misspecification using score tests

- adjustment for multiple comparisons
- max-test procedure for score statistics

# Reference I

Ebert, S. E., Jensen, P., Ozenne, B., Armand, S., Svarer, C., Stenbaek, D. S., Moeller, K., Dyssegaard, A., Thomsen, G., Steinmetz, J., Forchhammer, B. H., Knudsen, G. M., and Pinborg, L. H. (2019). Molecular imaging of neuroinflammation in patients after mild traumatic brain injury; a longitudinal 123i-clinde spect study. *European Journal of Neurology*.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical journal*, 50(3):346–363.

Wei, B.-C., Hu, Y.-Q., and Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of statistics*, 25(1):25–37.

# Generic LVM

- a measurement model:

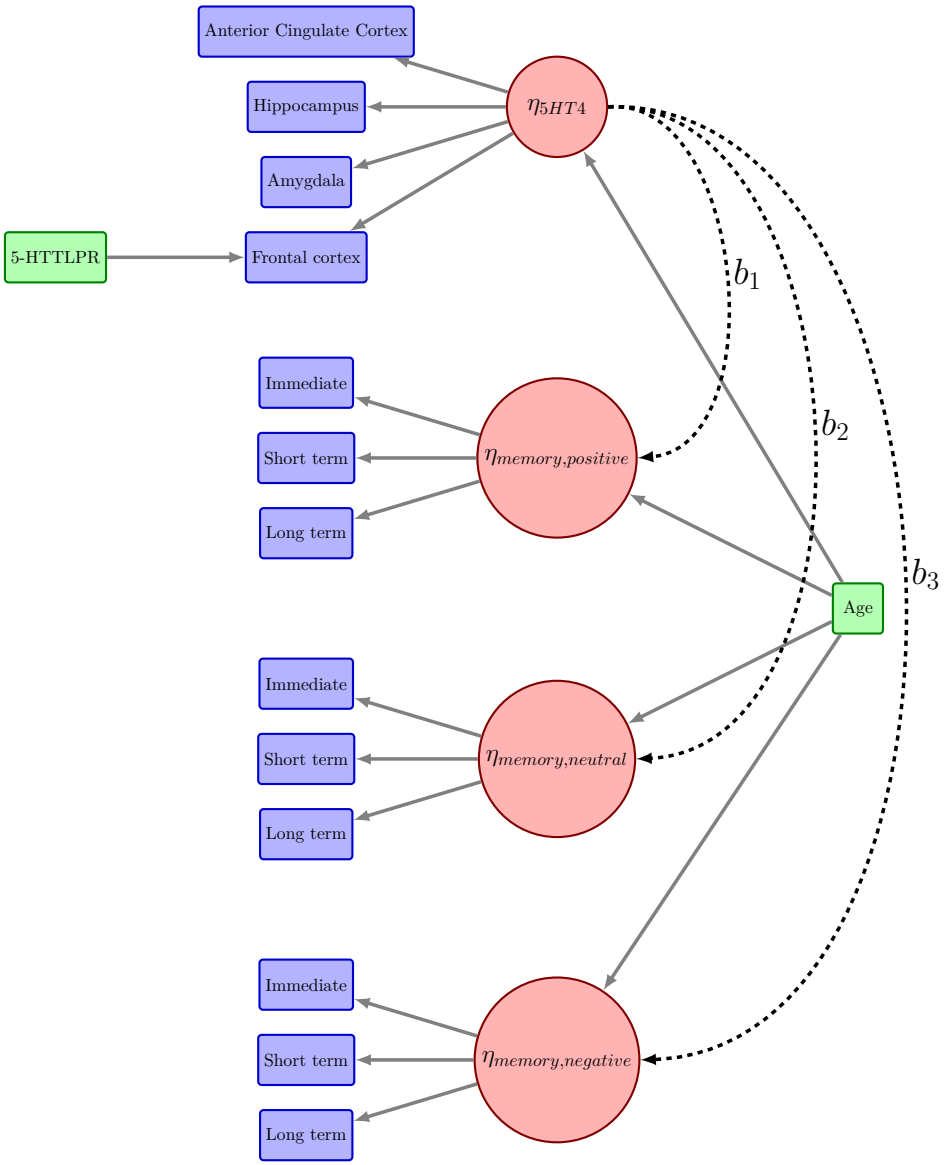
$$\boldsymbol{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \boldsymbol{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

- a structural model:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\eta}_i B + \boldsymbol{X}_i \Gamma + \boldsymbol{\zeta}_i, \text{ where } \boldsymbol{\zeta}_i \sim \mathcal{N}(0, \Sigma_\zeta)$$

- identifiability constrains, e.g. $\nu_1 = 0$, $\lambda_1 = 1$, $diag(B) = \mathbf{0}$

# Generic LVM - illustration



# LVM as a Gaussian model

$$\boldsymbol{Y}_i|\boldsymbol{X}_i \sim \mathcal{N}\left(\mu(\Theta, \boldsymbol{X}_i), \Omega(\Theta)\right)$$

with a specific structure for the conditional mean:

$$\mu(\Theta, \boldsymbol{X}_i) = \boldsymbol{\nu} + \boldsymbol{\alpha}(1-B)^{-1}\Lambda + \boldsymbol{X}_i\Gamma(1-B)^{-1}\Lambda + \boldsymbol{X}_i K$$

and the conditional variance:

$$\Omega(\Theta) = \Lambda^\intercal(1-B)^{-\intercal}\Sigma_\zeta(1-B)^{-1}\Lambda + \Sigma_\varepsilon$$

Note:

$$\begin{aligned}\mathcal{I}_1(\theta, \theta') =& \frac{1}{2} tr\left(\Omega(\Theta)^{-1}\frac{\partial\Omega(\Theta)}{\partial\theta}\Omega(\Theta)^{-1}\frac{\partial\Omega(\Theta)}{\partial\theta'}\right) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mu(\Theta)}{\partial\theta}\Omega(\Theta)^{-1}\frac{\partial\mu(\Theta)^\intercal}{\partial\theta'}\end{aligned}$$

# (B) Satterthwaite approximation

We model the distribution of the variance of our estimator:

$$k\hat{\sigma}^2_{\hat{\theta}} \sim \chi^2(df)$$

We identify $k$ and $df$ using the method of moments:

$$\mathbb{E}\left[k\hat{\sigma}^2_{\hat{\theta}}\right] = \mathbb{E}\left[\chi^2(df)\right] = df$$
$$\mathbb{V}ar\left[k\hat{\sigma}^2_{\hat{\theta}}\right] = \mathbb{V}ar\left[\chi^2(df)\right] = 2df$$

i.e.

$$df = 2\frac{\mathbb{E}\left[\hat{\sigma}^2_{\hat{\theta}}\right]^2}{\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]}$$

# (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

# (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

Estimates from maximum likelihood estimator satisfies:

$$n^{1/2}(\hat{\Theta} - \Theta) \sim \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

# (B) Estimating $\mathbb{V}ar\left[\hat{\sigma}^2_{\hat{\theta}}\right]$

We can relate the estimated variance of our estimator to the model parameters:

$$\hat{\sigma}_{\hat{\theta}} = f(\hat{\Theta})$$

Estimates from maximum likelihood estimator satisfies:

$$n^{1/2}(\hat{\Theta} - \Theta) \sim \mathcal{N}\left(0, \mathcal{I}_1(\Theta)^{-1}\right)$$

Then, the delta method gives:

$$n^{1/2}(\hat{\sigma}_{\hat{\Theta}} - \sigma_{\hat{\Theta}}) \sim \mathcal{N}\left(0, \nabla_\Theta f(\Theta)\mathcal{I}_1(\Theta)^{-1}\nabla_\Theta f(\Theta)\right)$$

# (C) Effective sample size

$$\mathbf{n}^c = \sum_{i=1}^{n} \frac{\partial \xi_i(\hat{\Theta})}{\partial \boldsymbol{Y}_i} = n - \sum_{i=1}^{n} \frac{\partial \mu(\hat{\Theta}, \boldsymbol{X}_i)}{\partial \boldsymbol{Y}_i}$$

where $\frac{\partial \mu(\hat{\Theta}, \boldsymbol{X}_i)}{\partial \boldsymbol{Y}_i}$ are the generalized leverage defined by (Wei et al., 1998).

**Example**: univariate linear regression

$$\hat{n}^c = n - \sum_{i=1}^{n} \boldsymbol{X}_i \frac{\partial \beta}{\partial Y_i} = n - \sum_{i=1}^{n} \boldsymbol{X}_i \left(\boldsymbol{X}^\intercal \boldsymbol{X}\right)^{-1} \boldsymbol{X}_i{}^\intercal = n - p$$