Supplementary material: Controlling the familywise error rate when performing multixple comparisons in a Linear Latent Variable Model

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

[1]Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark

[2]Neurobiology Research Unit, Neurobiology Research Unit, University Hospital of Copenhagen, Copenhagen, Denmark

Appendix A Likelihood and derivatives in LVMs

The log-likelihood of a LVM $\mathcal{M}(\boldsymbol{\Theta})$ can be written:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{Y},\boldsymbol{X}) &= \sum_{i=1}^{n} \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{Y}_i,\boldsymbol{X}_i) \\ &= \sum_{i=1}^{n} -\frac{m}{2}\log(2\pi) - \frac{1}{2}log\,|\Omega(\boldsymbol{\Theta})| - \frac{1}{2}(\boldsymbol{Y}_i - \mu(\boldsymbol{\Theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\Theta})^{-1}(\boldsymbol{Y}_i - \mu(\boldsymbol{\Theta},\boldsymbol{X}_i))^{\intercal}\end{aligned}$$

The individual score is obtained by derivating once the log-likelihood, i.e., for $\theta \in \mathbf{\Theta}$:

$$\begin{aligned}\mathcal{S}_i(\theta) = \frac{\partial \mathcal{L}(\mathbf{\Theta}|\boldsymbol{Y}_i, \boldsymbol{X}_i)}{\partial \theta} = &-\frac{1}{2} tr\left(\Omega(\mathbf{\Theta})^{-1}\frac{\partial \Omega(\mathbf{\Theta})}{\partial \theta}\right) \\ &+ \frac{\partial \mu(\mathbf{\Theta}, \boldsymbol{X}_i)}{\partial \theta}\Omega(\mathbf{\Theta})^{-1}(\boldsymbol{Y}_i - \mu(\mathbf{\Theta}, \boldsymbol{X}_i))^\intercal \\ &+ \frac{1}{2}(\boldsymbol{Y}_i - \mu(\mathbf{\Theta}, \boldsymbol{X}_i))\Omega(\mathbf{\Theta})^{-1}\frac{\partial \Omega(\mathbf{\Theta})}{\partial \theta}\Omega(\mathbf{\Theta})^{-1}(\boldsymbol{Y}_i - \mu(\mathbf{\Theta}, \boldsymbol{X}_i))^\intercal.\end{aligned}$$

Derivating a second time the log-likelihood and taking the minus the expectation of the result gives the expected information matrix for a single observation, $\mathcal{I}_1(\mathbf{\Theta})$, with elements:

$$\mathcal{I}_1(\theta, \theta') = \frac{1}{2} tr\left(\Omega(\mathbf{\Theta})^{-1}\frac{\partial \Omega(\mathbf{\Theta})}{\partial \theta'}\Omega(\mathbf{\Theta})^{-1}\frac{\partial \Omega(\mathbf{\Theta})}{\partial \theta}\right) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mu(\mathbf{\Theta}, \boldsymbol{X}_i)}{\partial \theta}\Omega(\mathbf{\Theta})^{-1}\frac{\partial \mu(\mathbf{\Theta}, \boldsymbol{X}_i)}{\partial \theta'}^\intercal.$$

Appendix B Normalized iid decomposition of the score

B.1 Notations

Consider a LVM $\mathcal{M}_p$ with parameters $\boldsymbol{\theta}$ and an extended LVM $\mathcal{M}_{p+1}$ with parameters $\mathbf{\Theta} = (\boldsymbol{\theta}, \beta)$. We denote by $\mathbf{\Theta}_0 = (\boldsymbol{\theta}_0, 0)$ the value of $\mathbf{\Theta}$ used to generate the data, by $\widehat{\mathbf{\Theta}} = (\widehat{\boldsymbol{\theta}}, \widehat{\beta})$ the ML estimator in the extended model and $\widetilde{\mathbf{\Theta}} = (\widetilde{\boldsymbol{\theta}}, 0)$ the ML estimator in $\mathcal{M}_p$. The score statistic relative to the parameter β is:

$$U = \mathcal{S}\left(\widetilde{\mathbf{\Theta}}\right)\mathcal{I}^{-1}\left(\widetilde{\mathbf{\Theta}}\right)\mathcal{S}\left(\widetilde{\mathbf{\Theta}}\right)^\intercal$$

We further denote $\mathcal{S}(\mathbf{\Theta}) = \begin{bmatrix}\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{\Theta}) & \mathcal{S}_{\beta}(\mathbf{\Theta})\end{bmatrix}$ and $\mathcal{I}(\mathbf{\Theta}) = \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\mathbf{\Theta}) & \mathcal{I}_{\boldsymbol{\theta},\beta}(\mathbf{\Theta}) \\ \mathcal{I}_{\beta,\boldsymbol{\theta}}(\mathbf{\Theta}) & \mathcal{I}_{\beta,\beta}(\mathbf{\Theta})\end{bmatrix}$ as the score and the information matrix in the extended LVM evaluated at $\mathbf{\Theta}$. As in the main text, n will denote the sample size and $\mathcal{X}_i$ the observations for individual i. We also denote by

$\mathcal{H}$ the second derivative of the likelihood (i.e. the gradient of the score) and $\mathcal{I}^{-1}(\boldsymbol{\Theta})$ by $\Sigma(\boldsymbol{\Theta}) = \begin{bmatrix} \Sigma_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\Theta}) & \Sigma_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta}) \\ \Sigma_{\beta,\boldsymbol{\theta}}(\boldsymbol{\Theta}) & \Sigma_{\beta,\beta}(\boldsymbol{\Theta}) \end{bmatrix}$.

We will also use the following results on matrix inversion:

- $\Sigma^{-1}_{\beta,\beta}(\boldsymbol{\Theta}) = \mathcal{I}_{\beta,\beta} - \mathcal{I}_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta})\mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\Theta})\mathcal{I}_{\beta,\boldsymbol{\theta}}(\boldsymbol{\Theta})$
- $\Sigma^{-1}_{\beta,\beta}(\boldsymbol{\Theta})\Sigma_{\beta,\boldsymbol{\theta}}(\boldsymbol{\Theta}) = -\mathcal{I}_{\beta,\boldsymbol{\theta}}(\boldsymbol{\Theta})\mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\Theta})$

B.2 Lemma

Lemma 1. *Constrained maximum likelihood estimator under the null hypothesis (adapted from section 4.4. of Wood (2015))*

$$\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}} + \widehat{\beta}\mathcal{I}_{\beta,\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right)\mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(n^{-\frac{1}{2}}\right)$$

Proof: Using a first order Taylor expansion around $\widetilde{\boldsymbol{\Theta}}$:

$$\mathcal{S}\left(\widehat{\boldsymbol{\Theta}}\right) = \mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) + \left(\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right)\mathcal{H}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(\widehat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right)$$

The score evalutated the ML estimate, $\mathcal{S}\left(\widehat{\boldsymbol{\Theta}}\right)$, equals 0. Under the null hypothesis, the model is well specified and $\mathcal{H}(\widetilde{\boldsymbol{\Theta}}) = -\mathcal{I}(\widetilde{\boldsymbol{\Theta}}) + o_p(n^{-\frac{1}{2}})$. Using that $\widetilde{\beta} = 0$, we obtain:

$$\mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) = \begin{bmatrix} \widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}} & \widehat{\beta} \end{bmatrix} \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\widetilde{\boldsymbol{\Theta}}) & \mathcal{I}_{\boldsymbol{\theta},\beta}(\widetilde{\boldsymbol{\Theta}}) \\ \mathcal{I}_{\beta,\boldsymbol{\theta}}(\widetilde{\boldsymbol{\Theta}}) & \mathcal{I}_{\beta,\beta}(\widetilde{\boldsymbol{\Theta}}) \end{bmatrix} + o_p\left(\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\right) + o_p\left(n^{-\frac{1}{2}}\right)$$

In the constrained model, only the first components of the score (i.e. the one relative to $\widetilde{\boldsymbol{\theta}}$)

equals 0. This leads to:

$$
\begin{aligned}
0 &= \left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right) \mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) + \widehat{\beta}\mathcal{I}_{\beta,\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\right) + o_p\left(n^{-\frac{1}{2}}\right) \\
\widetilde{\boldsymbol{\theta}} &= \widehat{\boldsymbol{\theta}} + \widehat{\beta}\mathcal{I}_{\beta,\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) \mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\right) + o_p\left(n^{-\frac{1}{2}}\right)
\end{aligned}
$$

Finally since $\widehat{\beta}$ converges toward 0 (the true value) at rate square root of n, $o_p\left(\widetilde{\boldsymbol{\Theta}} - \widehat{\boldsymbol{\Theta}}\right)$ is $o_p\left(n^{-\frac{1}{2}}\right)$. □

B.3 Normalized iid decomposition of the score

We would like to identify a random variable $\mathcal{U}\left(\widetilde{\boldsymbol{\Theta}}\right)$ such that $\mathcal{U}\left(\widetilde{\boldsymbol{\Theta}}\right)\mathcal{U}\left(\widetilde{\boldsymbol{\Theta}}\right)^\intercal = U + o_p(n^{-\frac{1}{2}})$ while under the null hypothesis $\mathcal{U}\left(\widetilde{\boldsymbol{\Theta}}\right) \overset{d}{\sim} \mathcal{N}\left(0, \Sigma_{\mathcal{U}}\right)$ where $\Sigma_{\mathcal{U}}$ can be consistently estimated using the iid decomposition of $\mathcal{U}$. More precisely, under the null hypothesis,

$$
\mathcal{U}\left(\widetilde{\boldsymbol{\Theta}}\right) = \frac{1}{n}\sum_{i=1}^{n} \psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i) + o_p\left(n^{-\frac{1}{2}}\right)
$$

and $\widehat{\Sigma}_{\mathcal{U}} = \frac{1}{n}\sum_{i=1}^{n} \psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i)\psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i)^\intercal$.

We first relate the score and the estimated parameters using a first order Taylor expansion around $\widetilde{\boldsymbol{\Theta}}$:

$$
\begin{aligned}
\mathcal{S}\left(\hat{\boldsymbol{\Theta}}\right) &= \mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) + \left(\hat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right)\mathcal{H}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(\hat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right) \\
\mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) &= \left(\hat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right)\mathcal{I}\left(\widetilde{\boldsymbol{\Theta}}\right) + o_p\left(n^{-\frac{1}{2}}\right)
\end{aligned}
$$

where we have used (i) that the score evalutated the ML estimate equals 0, (ii) under the null hypothesis both $\widetilde{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\Theta}}$ converges toward the true value at rate square root of n, (iii)

$\mathcal{H}(\boldsymbol{\Theta}) = -\mathcal{I}(\boldsymbol{\Theta}) + o_p(n^{-\frac{1}{2}})$. Applying Lemma 2, we can rewrite the first term as:

$$\begin{aligned}\left(\hat{\boldsymbol{\Theta}} - \widetilde{\boldsymbol{\Theta}}\right)\mathcal{I}\left(\widetilde{\boldsymbol{\Theta}}\right) &= \widehat{\beta}\begin{bmatrix}-\mathcal{I}_{\beta,\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right)\mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}^{-1}\left(\widetilde{\boldsymbol{\Theta}}\right) & 1\end{bmatrix}\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) & \mathcal{I}_{\boldsymbol{\theta},\beta}\left(\widetilde{\boldsymbol{\Theta}}\right)\\ \mathcal{I}_{\beta,\boldsymbol{\theta}}\left(\widetilde{\boldsymbol{\Theta}}\right) & \mathcal{I}_{\beta,\beta}\left(\widetilde{\boldsymbol{\Theta}}\right)\end{bmatrix}\\ &= \widehat{\beta}\begin{bmatrix}0 & \mathcal{I}_{\beta,\beta}(\widetilde{\boldsymbol{\Theta}}) - \mathcal{I}_{\beta,\boldsymbol{\theta}}(\widetilde{\boldsymbol{\Theta}})\mathcal{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}^{-1}(\widetilde{\boldsymbol{\Theta}})\mathcal{I}_{\boldsymbol{\theta},\beta}(\widetilde{\boldsymbol{\Theta}})\end{bmatrix}\end{aligned}$$

Therefore, denoting $M = \begin{bmatrix}0 & \Sigma_{\beta,\beta}^{-1}\left(\widetilde{\boldsymbol{\Theta}}\right)\end{bmatrix}$, we have:

$$\mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) = \widehat{\beta}M + o_p\left(n^{-\frac{1}{2}}\right) \tag{A}$$

The iid decomposition of $\widehat{\beta}$ under the global null hypothesis is:

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\psi_\beta(\boldsymbol{\Theta}_0, \mathcal{X}_i) + o_p\left(n^{-\frac{1}{2}}\right)$$

where $\psi_\beta(\boldsymbol{\Theta}_0, \mathcal{X}_i) = \mathcal{S}_{\beta,i}(\boldsymbol{\Theta}_0)\Sigma_{\beta,\beta}(\boldsymbol{\Theta}_0) + \mathcal{S}_{\boldsymbol{\theta},i}(\boldsymbol{\Theta}_0)\Sigma_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta}_0)$. So

$$\mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right) = \frac{1}{n}\sum_{i=1}^{n}\psi_\beta(\boldsymbol{\Theta}_0, \mathcal{X}_i)M + o_p(n^{-\frac{1}{2}})$$

We now introduce the normalized iid decomposition of the score:

$$\psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i) = \psi_\beta(\boldsymbol{\Theta}_0, \mathcal{X}_i)M\mathcal{I}^{-\frac{1}{2}}\left(\widetilde{\boldsymbol{\Theta}}\right)$$

which satisfies

$$\left(\frac{1}{n}\sum_{i=1}^{n}\psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n}\psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i)\right)^{\intercal} = \mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right)\mathcal{I}^{-1}\left(\widetilde{\boldsymbol{\Theta}}\right)\mathcal{S}\left(\widetilde{\boldsymbol{\Theta}}\right)^{\intercal} + o_p(n^{-\frac{1}{2}}) = U + o_p(n^{-\frac{1}{2}})$$

Here $M^{\frac{1}{2}}$ is the symmetric square root, i.e., the matrix satisfying $M^{\frac{1}{2}}M^{\frac{1}{2}\intercal} = M$. We further

note that an asymptotically equivalent expression of the iid terms is:

$$\psi_{\mathcal{U}}(\boldsymbol{\Theta}_0, \mathcal{X}_i) = \left(\mathcal{S}_{\beta,i}(\boldsymbol{\Theta}_0) - \mathcal{S}_{\boldsymbol{\theta},i}(\boldsymbol{\Theta}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\Theta}_0)\mathcal{I}_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta}_0)\right)\left[\mathcal{I}^{-\frac{1}{2}}\left(\widetilde{\boldsymbol{\Theta}}\right)\right]_{\beta,\cdot}$$

where $[M]_{\beta,\cdot}$ indicates the restriction of the matrix M to the rows corresponding to the β parameter. Indeed:

$$\begin{aligned}\psi_{\beta}(\boldsymbol{\Theta}_0, \mathcal{X}_i)M &= \left[0 \quad \left(\mathcal{S}_{\beta,i}(\boldsymbol{\Theta}_0)\Sigma_{\beta,\beta}(\boldsymbol{\Theta}_0) + \mathcal{S}_{\boldsymbol{\theta},i}(\boldsymbol{\Theta}_0)\Sigma_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta}_0)\right)\Sigma^{-1}_{\beta,\beta}(\widetilde{\boldsymbol{\Theta}})\right] \\ &= \left[0 \quad \mathcal{S}_{\beta,i}(\boldsymbol{\Theta}_0) - \mathcal{S}_{\boldsymbol{\theta},i}(\boldsymbol{\Theta}_0)\mathcal{I}^{-1}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\Theta}_0)\mathcal{I}_{\boldsymbol{\theta},\beta}(\boldsymbol{\Theta}_0)\right] + o_p(n^{-\frac{1}{2}})\end{aligned}$$

NOTE: we can retrieve the fact that $U \sim \chi^2_1$ plugging equation (A) into equation 3. Indeed:

$$\begin{aligned}U &= \widehat{\beta}M\mathcal{I}^{-1}(\widetilde{\boldsymbol{\theta}})M^{\intercal}\widehat{\beta}^{\intercal} + o_p(n^{-\frac{1}{2}}) \\ &= \widehat{\beta}\Sigma^{-1}_{\beta,\beta}(\widetilde{\boldsymbol{\theta}})\widehat{\beta}^{\intercal} + o_p(n^{-\frac{1}{2}}) \\ &= \widehat{\beta}\Sigma^{-1}_{\beta,\beta}(\widehat{\boldsymbol{\theta}})\widehat{\beta}^{\intercal} + o_p(n^{-\frac{1}{2}})\end{aligned}$$

Appendix C Comparison with step-up procedures

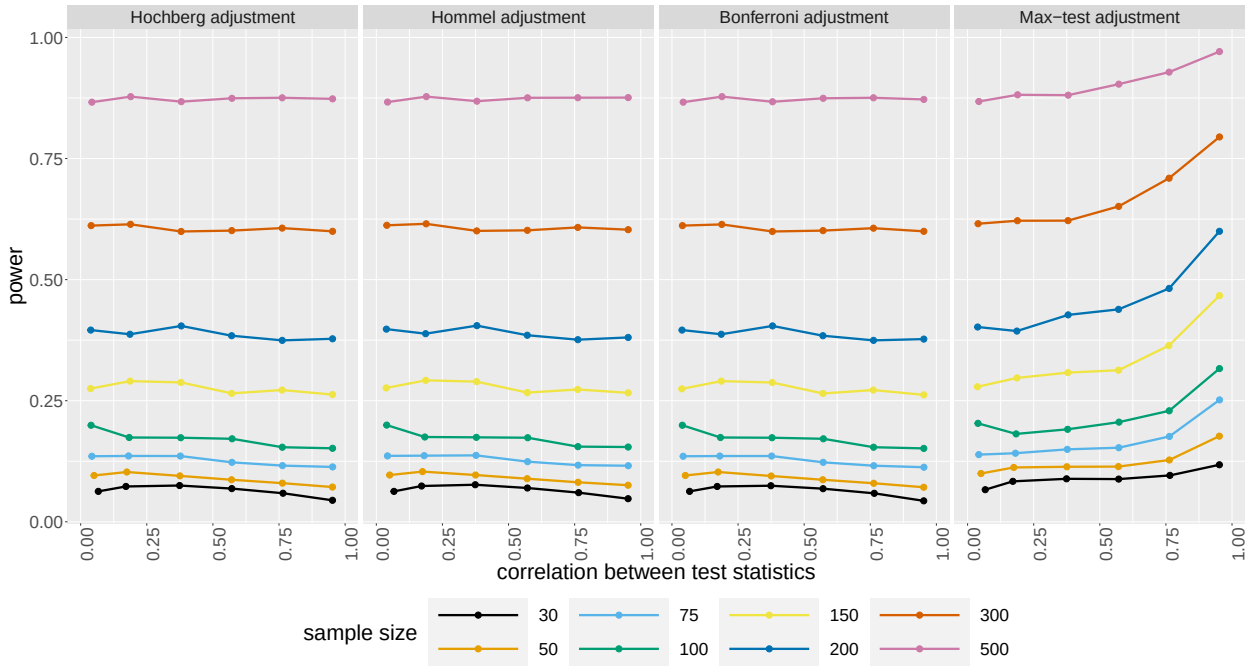

Figure A: Power when testing 9 null hypotheses with Wald tests using Hochberg, Hommel, Bonferroni or the max-test procedure to adjust the p-values for multiple comparisons. The first two are step-up procedure and the last two single-step procedures. Only the last one takes into account the correlation between the test statistics.

Appendix D Conditional inference in the FSS

D.1 Definition of the type 1 error

Consider, to simplify, that we rejected the first null hypothesis at the first stage of the FSS. At the second stage of the FSS, the null and alternative hypotheses are:

$$\left\{\begin{array}{lcl} \widehat{\mathcal{H}}_0 & : & \forall j \in \{2,\dots,c\}, \beta_j = 0 \\ & vs. & \\ \widehat{\mathcal{H}}_1 & : & \exists j \in \{2,\dots,c\}, \beta_j \neq 0 \end{array}\right\}$$

(the hat in the notation of the null hypothesis refers to the fact that the hypotheses are chosen in a data-driven way). Controlling the type 1 error is then a post-selection inference problem so we need to first explicit what type of error rate we want to control. As discussed in the litterature (e.g. section 2 of Lee et al. (2016) or in Fithian et al. (2014)) controlling the selective type 1 error rate:

$$\mathbb{P}\left[\text{reject } \widehat{\mathcal{H}}_0 \middle| \widehat{\mathcal{M}}_{p+1} = \mathcal{M}_{p+1}\right] \leq \alpha$$

is a reasonable approach.

D.2 Distribution of the max statistic

Conditioning on having selected the model $\widehat{\mathcal{M}}_{p+1}$ is equivalent to conditioning on the score statistics of the first constraint, $U^{\mathcal{M}^{(1)}_{p+1}}$, being strictly greater than the score statistics of the remaining constraints, $\left(U^{\mathcal{M}^{(j)}_{p+1}}\right)_{j\in\{2,\dots,p\}}$, and greater than the critical quantile q_α^{p+1}. To control the selective type 1 error rate, we are then interested in the conditional distribution

of the max statistic:

$$U_{\max}^{\mathcal{M}_{p+2}} = \max_{j\in\{2,\dots,p\}}(|U^{\mathcal{M}_{p+2}^{(j)}}|) \;\Bigg|\; |U^{\mathcal{M}_{p+1}^{(1)}}| \geq q_\alpha^{p+1}, |U^{\mathcal{M}_{p+1}^{(1)}}| > |U^{\mathcal{M}_{p+1}^{(2)}}|, \dots, |U^{\mathcal{M}_{p+1}^{(1)}}| > |U^{\mathcal{M}_{p+1}^{(q)}}| \tag{B}$$

In theory, we could adapt our max-step procedure to accound for the conditioning:

(i) sampling in the joint distribution of $\left(\mathcal{U}^{\mathcal{M}_{p+1}^{(1)}}, \dots, \mathcal{U}^{\mathcal{M}_{p+1}^{(c)}}, \mathcal{U}^{\mathcal{M}_{p+2}^{(2)}}, \dots, \mathcal{U}^{\mathcal{M}_{p+2}^{(c)}}\right)$ under the null hypothesis $\widehat{\mathcal{M}}_{p+2}$.

(ii) discard the samples that would not lead to retain model $\mathcal{M}_{p+1}^{(1)}$ at the first step.

(iii) compute the max statistic of the second step.

We note that whenever there was a clear winner at the first step, i.e. $|U^{\mathcal{M}_{p+1}^{(1)}}| >> |q_\alpha^{p+1}|$ and $\forall j \in \{2, \dots, q\}$ $|U^{\mathcal{M}_{p+1}^{(1)}}| >> |U^{\mathcal{M}_{p+1}^{(j)}}|$, then very few samples will be discarded so ignoring step (ii) will lead to essentially to the same results. Otherwise, by ignoring step (ii), one can expect to obtain conservative p-values since we ignore the conditioning on events like $|U^{\mathcal{M}_{p+1}^{(2)}}| < q_\alpha^{p+1}$. This is examplified in the next section.

D.3 Impact of the FSS on the control of the FWER on the subsequent analysis

When at least one of the null hypotheses is rejected, it is a common practice to use the extended LVM with the largest score statistic instead of the initial model. If this new LVM is not saturated (i.e. additional parameters could be added), the score tests are once again performed on this new LVM, and the LVM is updated until no additional parameter is found to significantly improve the model fit. Then, standard statistical tests (e.g. Wald tests) are used on the parameter of interest (which should not be one of the parameter added based on

the score tests). This, however, ignores that the extended model has been chosen based on the data, i.e., that we condition on specific bounds for the score statistics. This may create a bias in the p-value (see Tibshirani et al. (2016)). For linear models with independent and homoschedastic errors, Tibshirani et al. (2016) derived a method to estimate p-values in FSS with known number of steps that appropriately control the so called selective type one error rate. However, we found that generalizing these results to our setting is difficult and we only aim to assess the impact of the score tests on the subsequent Wald tests.

We re-use the generative model defined by equations (7), (8), (9), and (10). We added 2 covariance links to the generative model (link between regions 1-2 and region 6-8) and set the group effect to 0. The values of the remaining parameters were obtained by fitting the unconstrained model to the data used for the illustration (see section 6). The investigator model is defined by equations (7), (8), (9), and (10) under the constraint $\lambda_1 = 1$, i.e. has no covariance links. The investigator is interested in testing the group effect as in section 5.1 but first run a FSS to add any missing covariance links. This procedure was repeated many times to obtain 10 000 repetitions where the FSS identified the correct model for a sample size of 50, 100, and 500. This was the case in, respectively to the sample size, 9.4%, 53.1%, and 100% of the simulations. In these simulations the FWER for the group effect was, respectively to the sample size, 5.59%, 5.37%, 4.93%. This is compatible with the statistical fluctuations around 0.05 in a Monte Carlo simulation. The score statistics at the first step of the FSS where only very weakly correlated with the max Wald statistic used to test the group effect at the end of the FSS (absolute value of the Pearson correlation coefficient <0.02). Looking at the bivariate density plot did not reveal any clear association. This suggests that, in this specific setting, the distribution of the max Wald statistic may be independent of the distribution of the statistics used in the FSS, explaining why the selection mechanism does not have a noticeable impact on the control of the FWER.

Appendix E Estimated parameters in the initial latent variable model

The latent variable model defined by the investigator in the application (equations 7-10 under the constraint $\lambda_1 = 1$) contained 45 parameters:

- 8 parameters from ν, one for each region except for thalamus where the intercept was fixed to 0. The estimated value can be read from the software output below, section `Intercepts` from `log.pallidostriatum` to `log.corpusCallosum`.

- 8 parameters from Λ, one for each region except for thalamus where the loading was fixed to 1. The estimated value can be read from the software output below, section `Measurements` from `log.pallidostriatum~eta` to `log.corpusCallosum~eta`.

- 18 parameters from K, one for each region and each covariate (genotype and group). The estimated value can be read from the software output below, section `Regressions`.

- 9 parameters from Σ_ε, one for each region. The estimated value can be read from the software output below, section `Residual Variances` from `log.thalamus` to `log.corpusCallosum`.

- 1 parameter from α corresponding to the intercept of the latent variable. Its value can be read on the line `eta` section `Intercepts` in the software output below.

- 1 parameter from Σ_ζ corresponding to the variance of the latent variable. Its value can be read on the line `eta` section `Residual Variance` in the software output below.

Latent variables: eta

	Estimate	Std. Error	Z-value	P-value
Measurements:				
log.thalamus~eta	1.00000			
log.pallidostriatum~eta	0.83452	0.09235	9.03608	<1e-12
log.neocortex~eta	0.85006	0.08348	10.18336	<1e-12
log.midbrain~eta	0.84739	0.07774	10.90004	<1e-12
log.pons~eta	0.86516	0.09556	9.05311	<1e-12
log.cingulateGyrus~eta	0.76682	0.07412	10.34541	<1e-12
log.hippocampus~eta	0.76147	0.09489	8.02438	<1e-12
log.supramarginalGyrus~eta	0.87999	0.08552	10.29020	<1e-12
log.corpusCallosum~eta	0.67779	0.09941	6.81842	9.205e-12
Regressions:				
log.thalamus~genotypeHAB	0.60113	0.09229	6.51368	7.333e-11
log.thalamus~groupconcussion	0.11999	0.09451	1.26960	0.2042
log.pallidostriatum~genotypeHAB	0.57197	0.07757	7.37391	<1e-12
log.pallidostriatum~groupconcussion	0.11358	0.07943	1.42992	0.1527
log.neocortex~genotypeHAB	0.57948	0.07598	7.62683	<1e-12
log.neocortex~groupconcussion	0.04283	0.07781	0.55051	0.582
log.midbrain~genotypeHAB	0.61591	0.07427	8.29282	<1e-12
log.midbrain~groupconcussion	0.09895	0.07606	1.30105	0.1932
log.pons~genotypeHAB	0.53958	0.08036	6.71447	1.888e-11
log.pons~groupconcussion	0.01545	0.08229	0.18771	0.8511
log.cingulateGyrus~genotypeHAB	0.65551	0.06822	9.60927	<1e-12

log.cingulateGyrus~groupconcussion	0.15936	0.06986	2.28119	0.02254
log.hippocampus~genotypeHAB	0.57525	0.07406	7.76779	<1e-12
log.hippocampus~groupconcussion	0.11901	0.07584	1.56920	0.1166
log.supramarginalGyrus~genotypeHAB	0.57436	0.07841	7.32519	<1e-12
log.supramarginalGyrus~groupconcussion	0.05089	0.08030	0.63378	0.5262
log.corpusCallosum~genotypeHAB	0.57192	0.07092	8.06380	<1e-12
log.corpusCallosum~groupconcussion	0.17416	0.07263	2.39780	0.01649
Intercepts:				
log.thalamus	0.00000			
log.pallidostriatum	0.07255	0.13764	0.52709	0.5981
log.neocortex	-0.08699	0.12440	-0.69927	0.4844
log.midbrain	0.25676	0.11585	2.21625	0.02667
log.pons	0.28991	0.14243	2.03549	0.0418
log.cingulateGyrus	0.09924	0.11046	0.89840	0.369
log.hippocampus	0.09823	0.14143	0.69451	0.4874
log.supramarginalGyrus	-0.12540	0.12745	-0.98393	0.3252
log.corpusCallosum	-0.00549	0.14816	-0.03705	0.9704
eta	1.43044	0.07746	18.46597	<1e-12
Residual Variances:				
log.thalamus	0.01308	0.00339	3.85729	
log.pallidostriatum	0.00987	0.00254	3.88728	
log.neocortex	0.00603	0.00166	3.63644	
log.midbrain	0.00402	0.00120	3.33602	
log.pons	0.01053	0.00271	3.88466	
log.cingulateGyrus	0.00451	0.00126	3.58286	

log.hippocampus	0.01247	0.00311	4.00907
log.supramarginalGyrus	0.00612	0.00170	3.60185
log.corpusCallosum	0.01602	0.00391	4.09882
eta	0.06319	0.01777	3.55679

F Reference

Fithian, W., Sun, D., and Taylor, J. (2014). Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J. E., et al. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927.

Tibshirani, R. J., Taylor, J., Lockhart, R., and Tibshirani, R. (2016). Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association*, 111(514):600–620.

Wood, S. N. (2015). *Core statistics*, volume 6. Cambridge University Press.