Fast and accurate modelling of longitudinal neuroimaging data

B. Guillaume[1 2 4] L. Waldorp[3] T. Nichols[1]

[1]Department of Statistics
University of Warwick

[2]Cyclotron Research center
University of Liège

[3]Department of Psychological Methods
University of Amsterdam

[4]GlaxoSmithKline

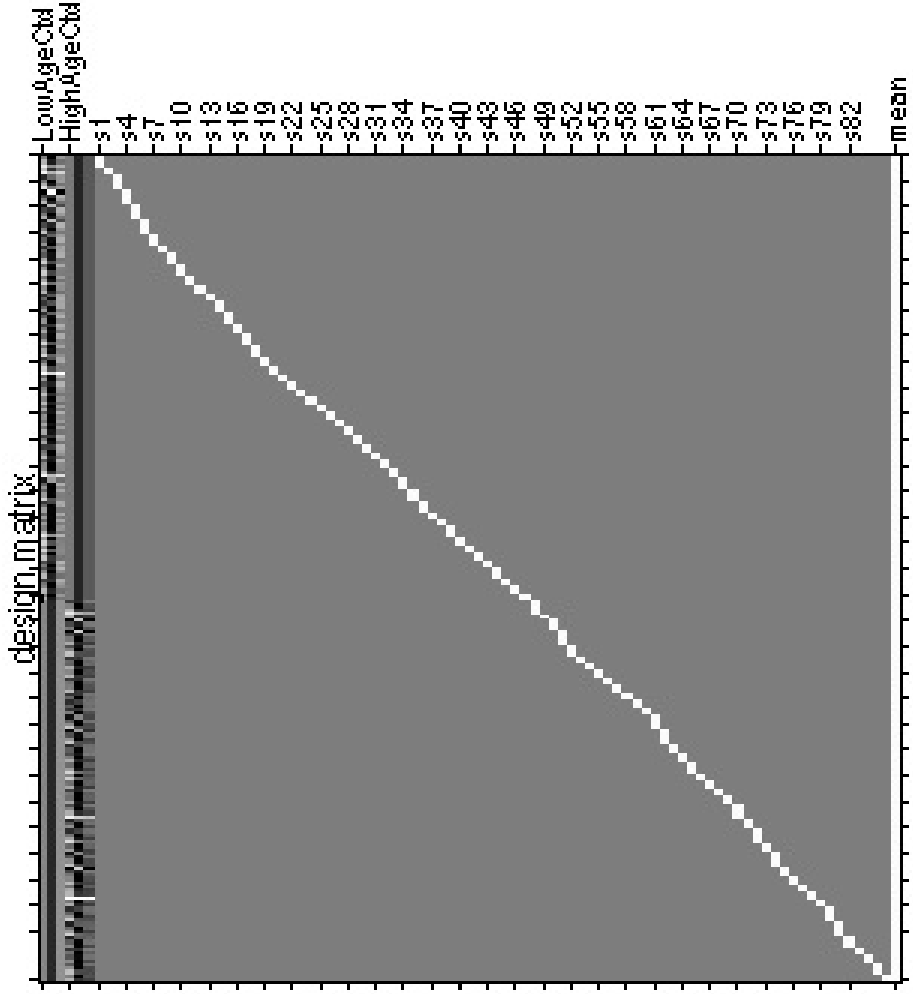
OHBM, 2012

An example of a longitudinal study in brainimaging

fMRI study of longitudinal changes in a population of adolescents at risk for alcohol abuse linked to Heitzeg et al. (2010)

- 86 subjects
- 2 groups (Low risk: 47 subj. and High risk: 39 subj.)
- Missing data (1, 2, 3 or 4 scans/subject)
- Go/No-go task
- No common time points

2nd-level standard modelling in SPM

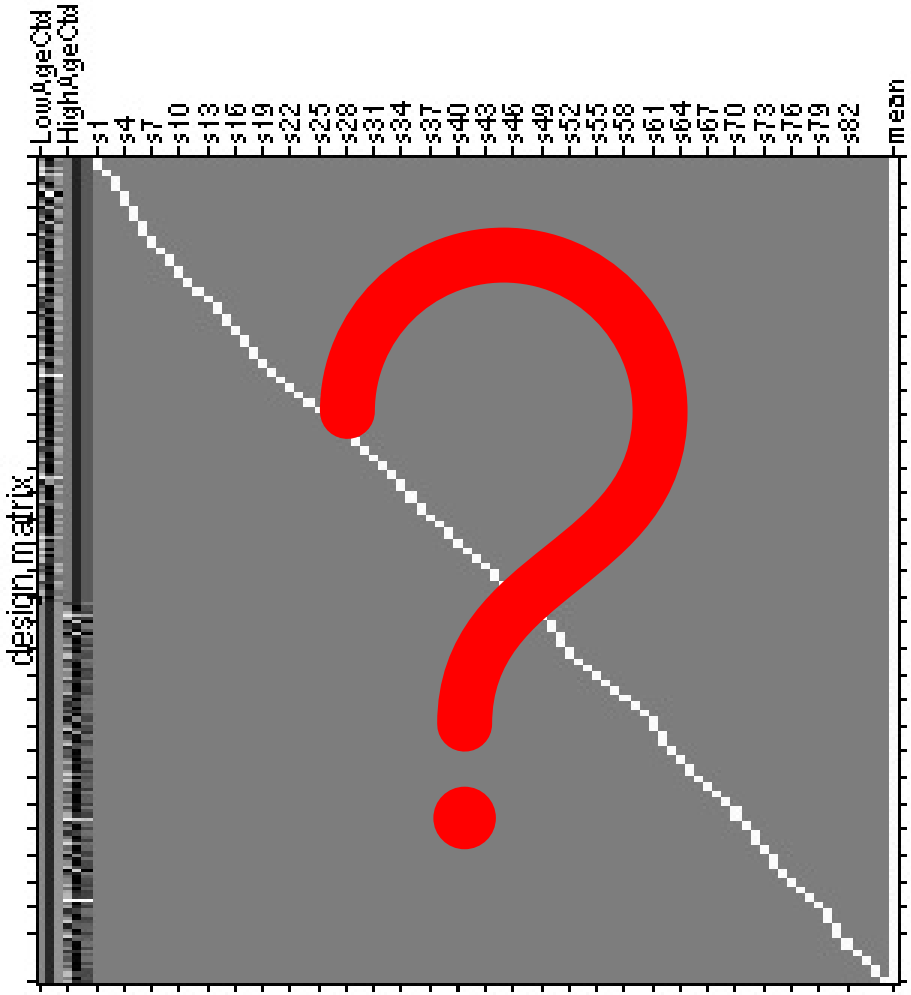


An example of a longitudinal study in brainimaging

fMRI study of longitudinal changes in a population of adolescents at risk for alcohol abuse linked to Heitzeg et al. (2010)

- 86 subjects
- 2 groups (Low risk: 47 subj. and High risk: 39 subj.)
- Missing data (1, 2, 3 or 4 scans/subject)
- Go/No-go task
- No common time points

2nd-level standard modelling in SPM

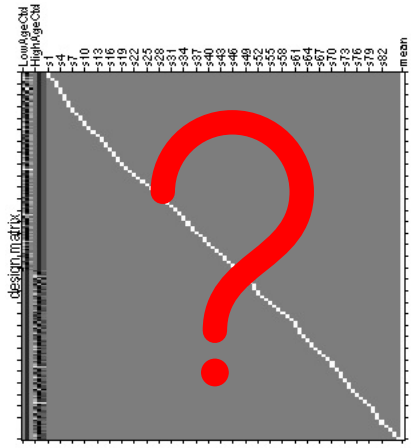


An example of a longitudinal study in brainimaging

fMRI study of longitudinal changes in a population of adolescents at risk for alcohol abuse linked to Heitzeg et al. (2010)

- 86 subjects
- 2 groups (Low risk: 47 subj. and High risk: 39 subj.)
- Missing data (1, 2, 3 or 4 scans/subject)
- Go/No-go task
- No common time points

2nd-level standard modelling in SPM



Only valid under **Compound Symmetry**:

$$\text{var}(y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}$$

Is there an alternative method?

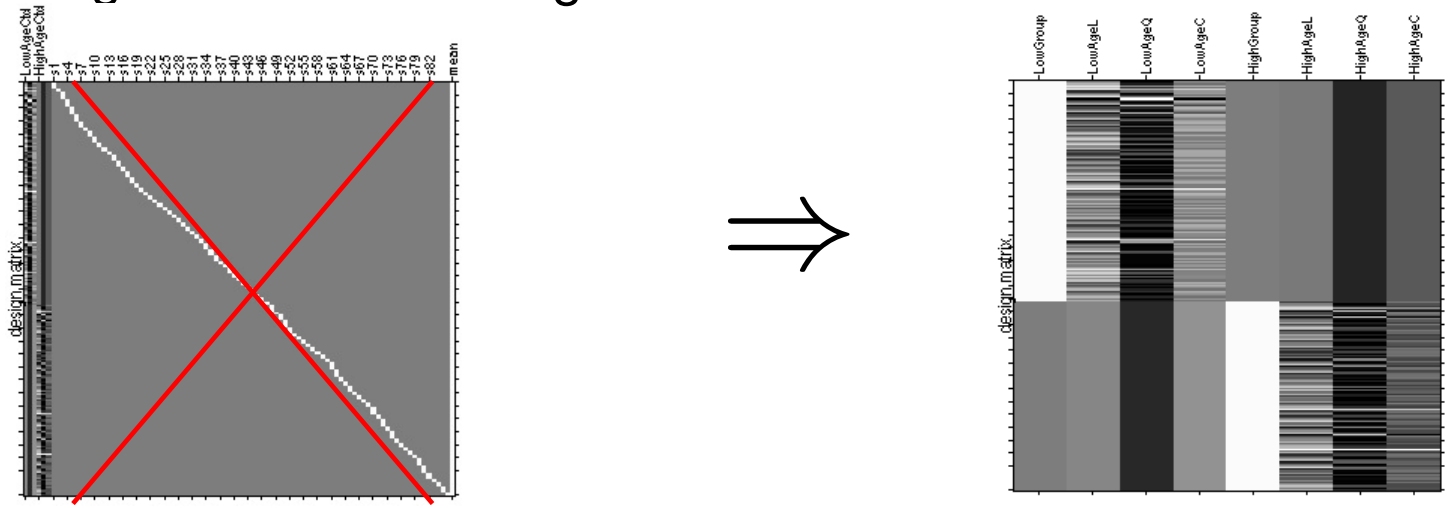
- Gold standard method for longitudinal data in Biostatistics: Linear Mixed Effects (LME) modelling (Laird and Ware, 1982) :

$$y_i = \underbrace{X_i\beta}_{\text{Fixed effects}} + \underbrace{Z_i b_i}_{\text{Random effects}} + e_i$$

- Unfortunatly, LME has drawbacks:
  - Random effects not easy to specify
  - Use of iterative algorithms
    - generally slow
    - may fail to converge to a solution:
      E.g., 12 subjects, 8 visits, Toeplitz (linearly decaying) correlation struct., LME with unstructured intra-visit correlation fails to converge 95 % of the time.
      E.g., 12 subjects, 8 visits, Compound Symmetry , LME with random int. and random slope fails to converge 2 % of the time.

The Sandwich Estimator (SwE) method

- Marginal OLS modelling used:



$$y_i = \underbrace{X_i\beta}_{\text{Fixed effects}} + \underbrace{\cancel{\beta_{0i}}}_{\text{Subject indicator covariates}} + e_i$$

- β estimated by OLS estimate $\hat{\beta}_{OLS}$
- var($\hat{\beta}_{OLS}$) estimated by the Sandwich Estimator (Eicker,1963):

$$\text{SwE} = \underbrace{\left(\sum_{i=1}^{M} X_i' X_i\right)^{-1}}_{\textit{Bread}} \underbrace{\left(\sum_{i=1}^{M} X_i' \hat{V}_i X_i\right)}_{\textit{Meat}} \underbrace{\left(\sum_{i=1}^{M} X_i' X_i\right)^{-1}}_{\textit{Bread}}$$

$$\text{with} \quad \hat{V}_i = r_i r_i' \quad \text{and} \quad r_i = y_i - X_i\hat{\beta}$$

The adjusted Sandwich Estimator (SwE) method

- SwE property:

$$\lim_{M \to +\infty} \text{SwE} = \text{var}(\hat{\beta}_{OLS})$$

  $\rightarrow$ **Large sample assumption**

- In order to enhance the accuracy of the SwE method in **small samples**, we propose to use:
  - Small sample bias adjustment (MacKinnon and White, 1985)
  - Small sample distributional adjustment of the statistical test Null distribution (Waldorp, 2009)
  - Assumption of a common covariance matrix among subjects

Assessment method

- Methods accuracy assessed by Null Gaussian Monte Carlo simulations (10,000 realizations)
- Metric used:

$$\text{Relative False Positive Rate (rel. FPR)} = \frac{\text{E}(\text{FPR}_{Method})}{\text{Nominal FPR}}$$

- 2 correlation structures tested:

Compound Symmetry

E.g., for subject 1:

$$\begin{pmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{pmatrix}$$

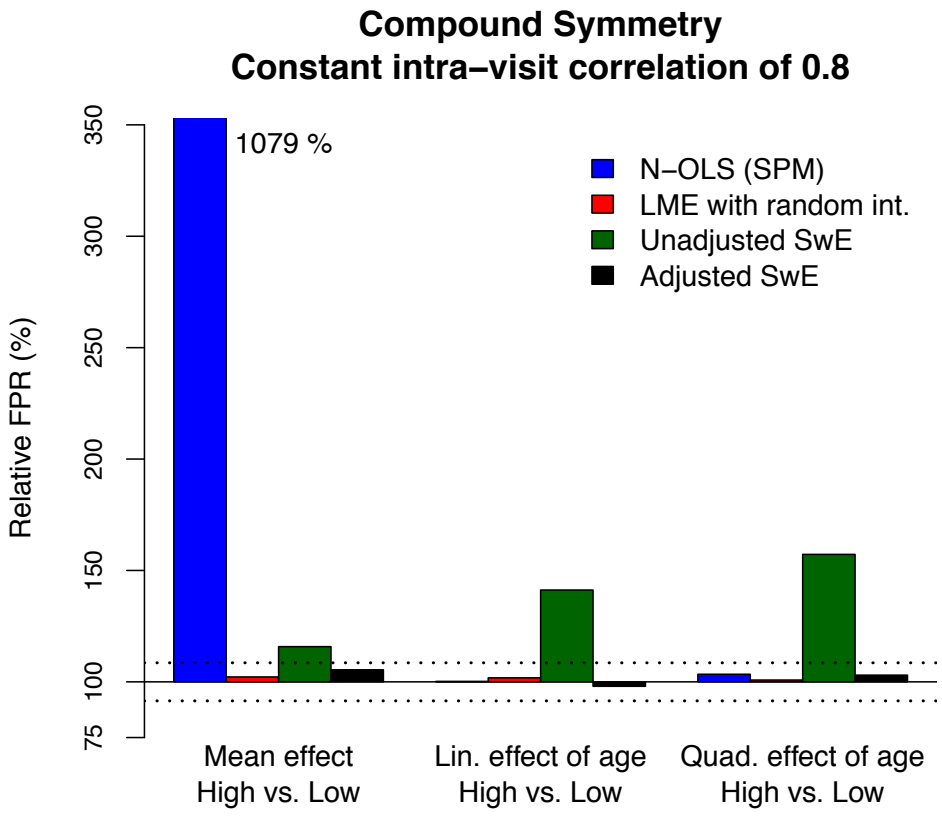
Non Compound Symmetry

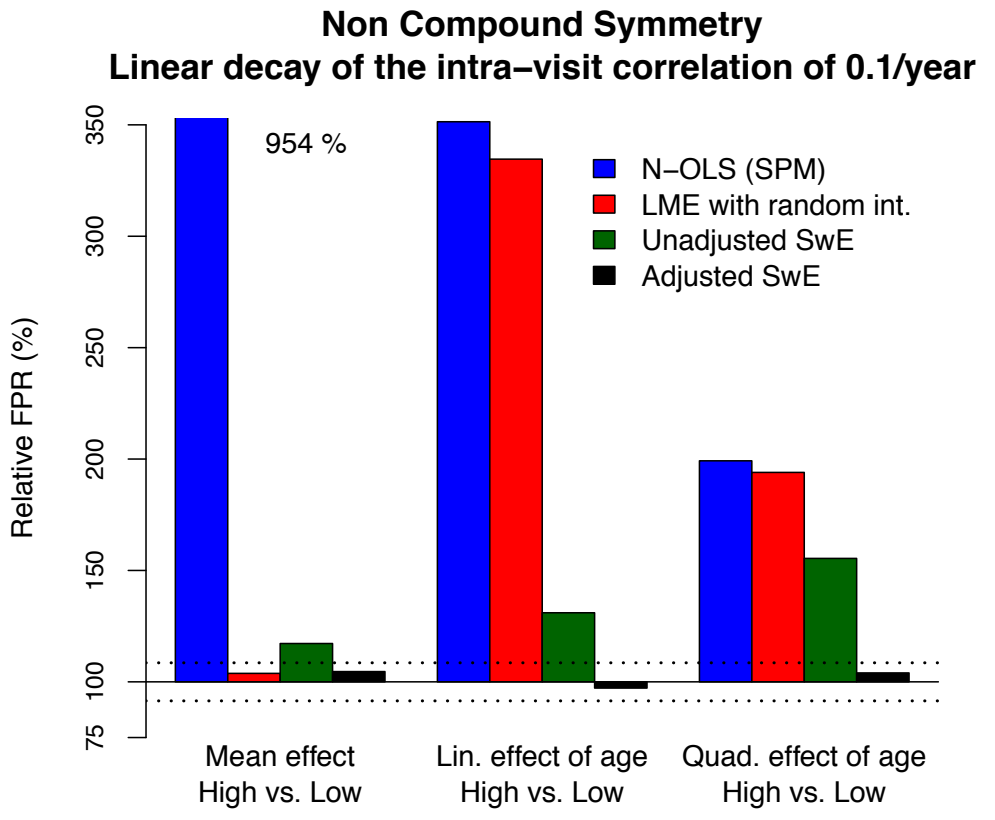
E.g., for subject 1:

$$\begin{pmatrix} 1 & 0.7 & 0.51 & 0.31 \\ 0.7 & 1 & 0.81 & 0.61 \\ 0.51 & 0.81 & 1 & 0.8 \\ 0.31 & 0.61 & 0.8 & 1 \end{pmatrix}$$

Simulation results

- F-test at 0.05 as level of significance





Summary

- Longitudinal standard methods not really appropriate to neuroimaging data:
  - LME
    - Difficult to specify
    - generally slow
    - Convergence issues
  - N-OLS
    - Issues when Compound Symmetry does not hold
    - Cannot accommodate pure between covariates (e.g., gender)
- The SwE method, particularly with small samples adjustments,
  - Accurate in a large range of settings
  - Easy to specify
  - No iteration needed
    - Quite fast
    - No convergence issues
  - Can accommodate pure between covariates (e.g., gender)

Acknowledgment

- Data
  - M. Heitzeg

- Funding
  - GlaxoSmithKline through the Marie Curie Initial Training Network "Neurophysics"