Toward a unified framework for analyzing repeated measurements with a continuous outcome

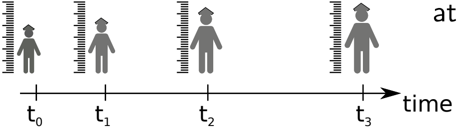
Brice Ozenne[1,2], Julie Lyng Forman[2]
brice.ozenne@nru.dk

[1] Section of Biostatistics, Department of Public Health, University of Copenhagen.
[2] Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet.

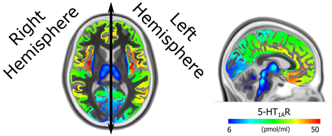
June 7th 2022, Method week at Karolinska Institutet

Repeated measurements in medical research

- **Clinical trial**: outcome measured on the **same patient** at **different timepoints**.

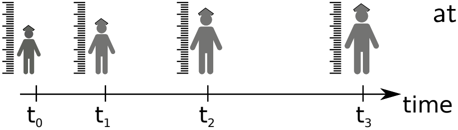


- **Observational**: outcome measured at **different locations**.

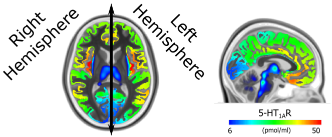


Repeated measurements in medical research

- **Clinical trial**: outcome measured on the **same patient** at **different timepoints**.



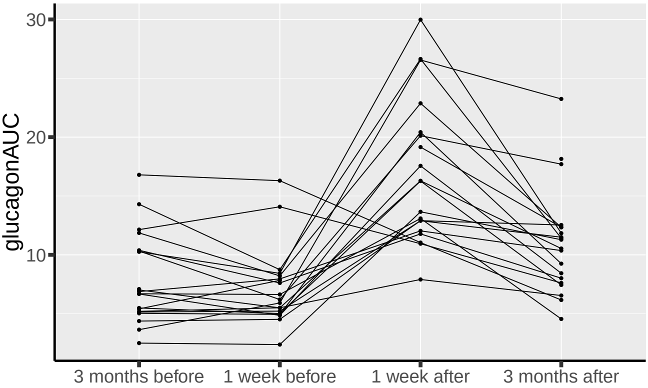
- **Observational**: outcome measured at **different locations**.



⚠ Occasions (i.e. timepoints, brain regions) are fixed by design

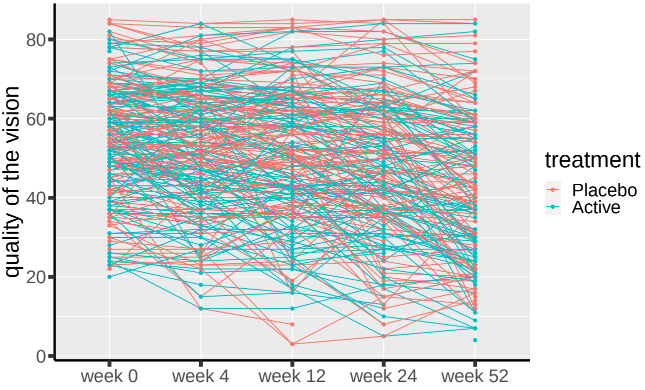
Example 1: gastric bypass study (Jorsal et al., 2020)

- single group of obese subjects
- outcome: gut hormone prior and after surgery



Example 2: ARMD Trial (int, 1997)

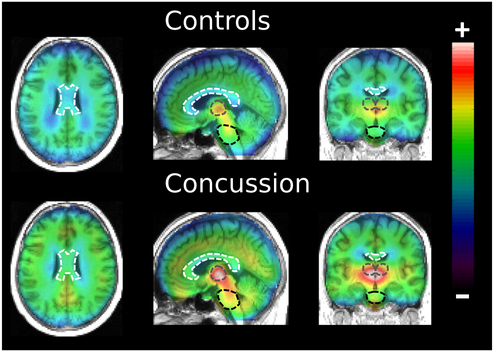
- comparing interferon-α and placebo
- outcome: change in vision over time



Example 3: brain trauma (Ebert et al., 2019)

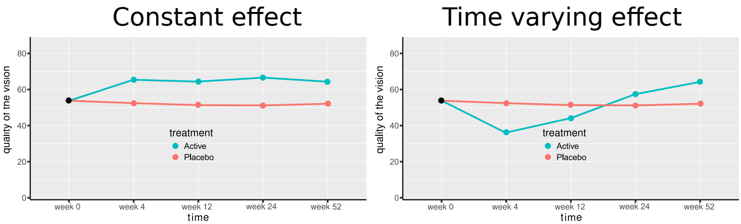
After a mild traumatic brain injury,
is there a neuroinflammatory response in the brain?

- 22 healthy controls and 14 patients
- genetic factors influencing brain measurements



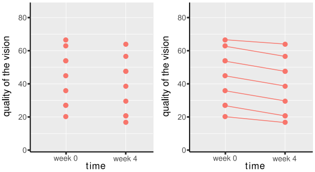
Why repeated repeated measurements? (1/2)

To capture **the time-dynamic** of the treatment effect:



To **get a better estimates**

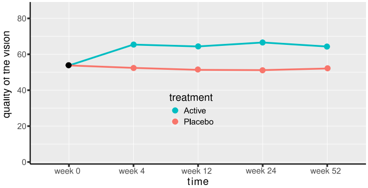
- use each patient
  as his own control



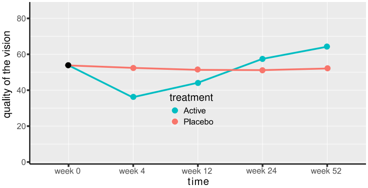
Why repeated repeated measurements? (1/2)

To capture **the time-dynamic** of the treatment effect:

Constant effect

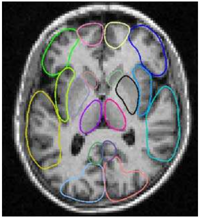


Time varying effect



To **get a better estimates**

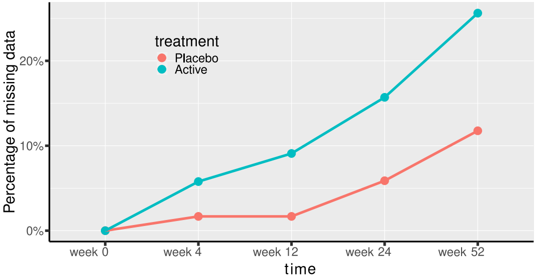
- use each patient as his own control
- extract the signal from noisy measurements



Why repeated repeated measurements? (2/2)

To **better handle missing values**:

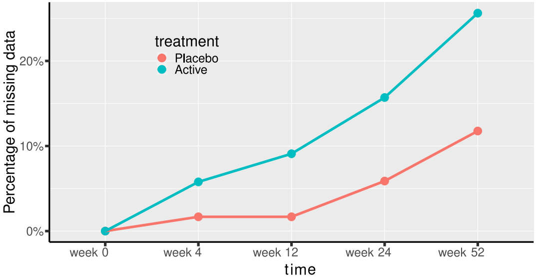
- as the follow-up time increases, patient are more likely to drop-out



Why repeated repeated measurements? (2/2)

To **better handle missing values**:

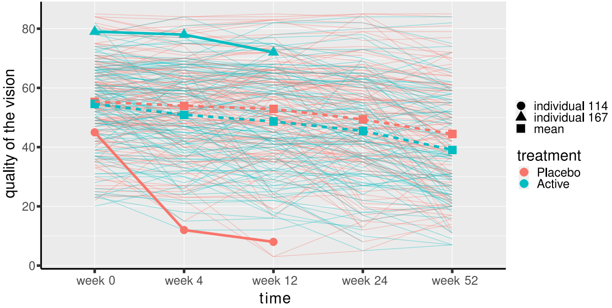
- as the follow-up time increases, patient are more likely to drop-out
- regular follow-up can help:
  - to understand the reason(s) for drop-out
  - to limit the bias/loss in statistical power due to drop-out



Why repeated repeated measurements? (2/2)

To **better handle missing values**:

- as the follow-up time increases, patient are more likely to drop-out
- regular follow-up can help:
  - to understand the reason(s) for drop-out
  - to limit the bias/loss in statistical power due to drop-out



Statistical challenges

Clinical trial:

- incorporate restrictions induced by randomization
- transparent and optimal handling of missing values
- valid statistical inference in small samples

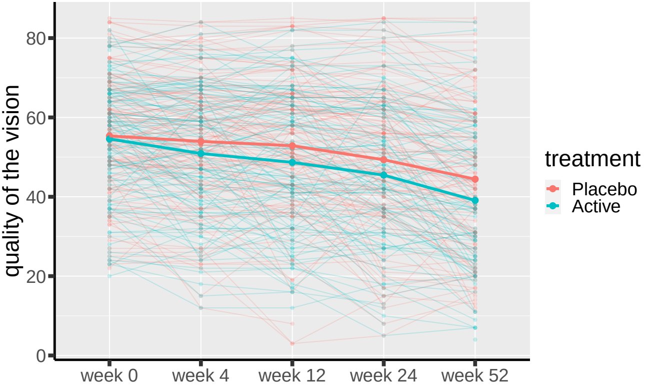
Observational:

- modeling a system of variables
- flexible but interpretable model

$\Longrightarrow$ multivariate approach

But what about a good old t-test?

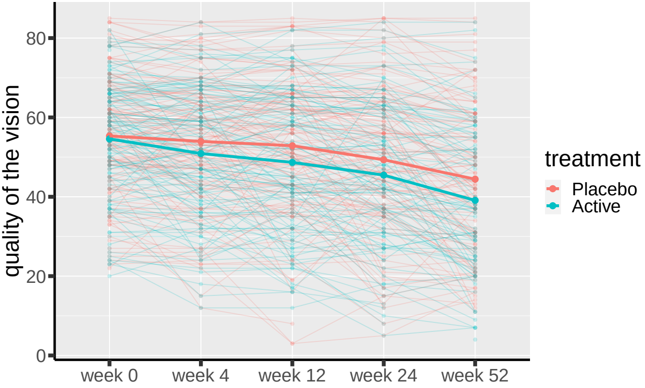
Two sample t-test at the last timepoint



But what about a good old t-test?

Two sample t-test at the last timepoint

- not optimal: 30% of the variance is explained by baseline
- not feasable: missing values



What can be done then?

Complete case analysis (2 timepoints 0-52):

Full information (2 timepoints):

Full information (5 timepoints):

What can be done then?

Complete case analysis (2 timepoints 0-52):

- two sample t-test on the change from baseline: CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.29$, $p = 0.061$
- linear regression `final~baseline+group`: ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.61$, $p = 0.038$

Full information (2 timepoints):
Full information (5 timepoints):

What can be done then?

Complete case analysis (2 timepoints 0-52):

- two sample t-test on the change from baseline: CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.29$, $p = 0.061$
- linear regression `final~baseline+group`: ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.61$, $p = 0.038$

Full information (2 timepoints):

- random intercept model CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.43$, $p = 0.0484$
- unstructured gaussian model CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.38$, $p = 0.055$

Full information (5 timepoints):

What can be done then?

Complete case analysis (2 timepoints 0-52):

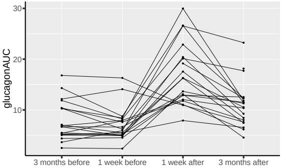
- two sample t-test on the change from baseline: CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.29$, $p = 0.061$
- linear regression `final~baseline+group`: ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.61$, $p = 0.038$

Full information (2 timepoints):
Full information (5 timepoints):

- random intercept model CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.90$, $p = 0.0046$
- unstructured gaussian model CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.87$, $p = 0.037$
- unstructured gaussian model ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.916$, $p = 0.0308$

A more dramatic example

Using only two timepoints
(+3 months vs. -3 months)
where there is **no missing values**:



- t-test on the change from baseline:
  $\hat{\mu}_A - \hat{\mu}_P = 3.20258$, $p = 0.04132$
- random intercept model:
  `boundary (singular) fit:  see ?isSingular=`
- unstructured gaussian model:
  $\hat{\mu}_A - \hat{\mu}_P = 3.20258$, $p = 0.04132$

Missing data

Two-timepoint study

Consider a study with two timepoints and $n = n_1 + n_2$ participants:

- n_1 with only baseline measurement — missing data
- n_2 with baseline and follow-up measurements — full data

Denote:

- $Y = (Y_1, Y_2)$ the outcome over time
- $\mu = (\mu_1, \mu_2)$ the mean over time
- $\Omega = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ the variance-covariance over time
- $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ the model parameters

Likelihood in a two-timepoints study

Under a Gaussian model

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_2} \frac{1}{2\pi \left|\Omega\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{bmatrix} Y_{1i}-\mu_1 & Y_{2i}-\mu_2 \end{bmatrix} \Omega^{-1} \begin{bmatrix} Y_{1i}-\mu_1 \\ Y_{2i}-\mu_2 \end{bmatrix}\right)$$
$$\prod_{i=n_2+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(Y_{1i}-\mu_1)^2\right)$$

Likelihood in a two-timepoints study

Under a Gaussian model

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_2} \frac{1}{2\pi |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{bmatrix} Y_{1i}-\mu_1 & Y_{2i}-\mu_2 \end{bmatrix} \Omega^{-1} \begin{bmatrix} Y_{1i}-\mu_1 \\ Y_{2i}-\mu_2 \end{bmatrix}\right)$$
$$\prod_{i=n_2+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(Y_{1i}-\mu_1)^2\right)$$

After some calculation, the score regarding μ_2 can be expressed:

$$\mathcal{S}_{\mu_2} = \frac{\partial \log \mathcal{L}(\theta)}{\partial \mu_2} = \sum_{i=1}^{n_2} \frac{Y_{2i}-\mu_2}{\sigma_2^2(1-\rho^2)} - \rho \frac{Y_{1i}-\mu_1}{\sigma_1\sigma_2(1-\rho^2)}$$

Likelihood in a two-timepoints study

Under a Gaussian model

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_2} \frac{1}{2\pi |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{bmatrix} Y_{1i}-\mu_1 & Y_{2i}-\mu_2 \end{bmatrix} \Omega^{-1} \begin{bmatrix} Y_{1i}-\mu_1 \\ Y_{2i}-\mu_2 \end{bmatrix}\right)$$
$$\prod_{i=n_2+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(Y_{1i}-\mu_1)^2\right)$$

After some calculation, the score regarding μ_2 can be expressed:

$$\mathcal{S}_{\mu_2} = \frac{\partial \log \mathcal{L}(\theta)}{\partial \mu_2} = \sum_{i=1}^{n_2} \frac{Y_{2i}-\mu_2}{\sigma_2^2(1-\rho^2)} - \rho \frac{Y_{1i}-\mu_1}{\sigma_1\sigma_2(1-\rho^2)}$$

$$\mathcal{S}_{\mu_2} = 0 \implies \mu_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i} - \rho\sigma_2 \frac{Y_{1i}-\mu_1}{\sigma_1} = [...]$$

Likelihood in a two-timepoints study

Under a Gaussian model

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_2} \frac{1}{2\pi |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\begin{bmatrix} Y_{1i}-\mu_1 & Y_{2i}-\mu_2 \end{bmatrix} \Omega^{-1} \begin{bmatrix} Y_{1i}-\mu_1 \\ Y_{2i}-\mu_2 \end{bmatrix}\right)$$

$$\prod_{i=n_2+1}^{n} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(Y_{1i}-\mu_1)^2\right)$$

After some calculation, the score regarding μ_2 can be expressed:

$$\mathcal{S}_{\mu_2} = \frac{\partial \log \mathcal{L}(\theta)}{\partial \mu_2} = \sum_{i=1}^{n_2} \frac{Y_{2i}-\mu_2}{\sigma_2^2(1-\rho^2)} - \rho \frac{Y_{1i}-\mu_1}{\sigma_1\sigma_2(1-\rho^2)}$$

$$\mathcal{S}_{\mu_2} = 0 \implies \mu_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i} - \rho\sigma_2 \frac{Y_{1i}-\mu_1}{\sigma_1} = [...]$$

$$\implies \mu_2 = \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mu_2 + \rho\sigma_2 \frac{Y_{1i}-\mu_1}{\sigma_1}\right)$$

A classical result

Gaussian conditional distribution

If $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$ then:

$$Y_2|y_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2\right)$$

A classical result

Gaussian conditional distribution

If $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$ then:

$$Y_2|y_1 \; \sim \; \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2\right)$$

- If $\rho = 0$, observing $Y_1 = y_1$ is useless:
$$Y_2|y_1 \; \sim \; \mathcal{N}\left(\mu_2, \sigma_2^2\right)$$
- if $\rho = 1$, μ_2 is corrected by how much y_1 deviates from μ_1
$$Y_2|y_1 \; = \; \mu_2 + \sigma_2\frac{(y_1 - \mu_1)}{\sigma_1}$$

Estimate in a two-timepoints study

$$\mu_2 = \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mu_2 + \rho\sigma_2 \frac{Y_{1i} - \mu_1}{\sigma_1}\right)$$
$$= \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mathbb{E}\left[Y_2 | Y_1\right]\right)$$

$\rightarrow$ model parameter identified as an average of observed values (when available) or conditional mean (when missing).

The mixed model can be seen as an **imputation model** for the missing values!

Estimate in a two-timepoints study

$$
\begin{aligned}
\mu_2 &= \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mu_2 + \rho\sigma_2 \frac{Y_{1i} - \mu_1}{\sigma_1}\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mathbb{E}\left[Y_2 | Y_1\right]\right)
\end{aligned}
$$

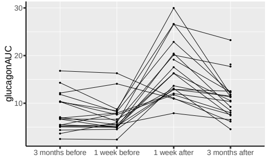
$\rightarrow$ model parameter identified as an average of observed values (when available) or conditional mean (when missing).

The mixed model can be seen as an **imputation model** for the missing values!

- not a good idea if missingness is due to death

Back to example 1

Using only two timepoints
(+3 months vs. -3 months)
where there is **no missing values**:



- t-test on the change from baseline: $\hat{\mu}_A - \hat{\mu}_P = 3.20258,\ p = 0.04132$
- unstructured gaussian model:
  $\hat{\mu}_A - \hat{\mu}_P = 3.20258,\ p = 0.04132$

No missing value = no need for imputation

Estimate in a multiple-timepoints study

The formula generalizes to multiple timepoints:
(i.e. $\mathbf{Y}_1$ and $\boldsymbol{\mu}_1$ are vectors)

$$\mu_2 = \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mu_2 + \mathbb{C}ov(Y_2, \mathbf{Y}_1)\mathbb{V}ar\left[\mathbf{Y}_1\right]^{-1}(\mathbf{Y}_{1i} - \boldsymbol{\mu}_1)\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n_2} Y_{2i} + \sum_{i=n_2+1}^{n} \mathbb{E}\left[Y_2|\mathbf{Y}_1\right]\right)$$

The quality of the imputation and therefore of the estimate depends on:

- the **mean**, **variance**, **correlation** structure,
  e.g. may be group dependent
- bias / variance trade-off

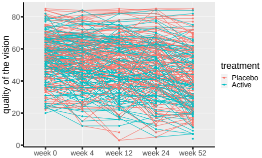
Back to example 2

Full information (2 timepoints):

- Random intercept model:
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.43,\ p = 0.0484$
- Unstructured gaussian model:
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.38,\ p = 0.055$

Full information (5 timepoints):

- Random intercept model
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.90,\ p = 0.0046$
- Unstructured gaussian model
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.87,\ p = 0.037$



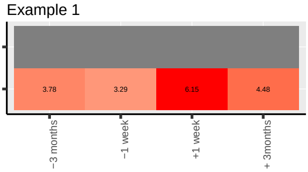

simplified imputation model

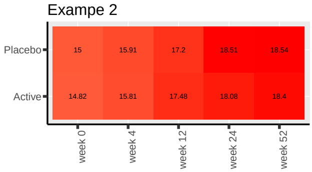
- 2 vs. 5 timepoints: $\mathbb{E}[Y_5 | Y_1, X] = \mathbb{E}[Y_5 | Y_1, Y_2, Y_3, Y_4, X]$?

Parametrization

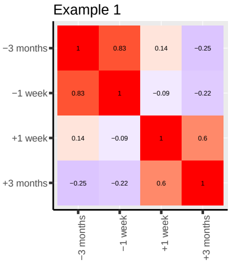
Empirical covariance pattern

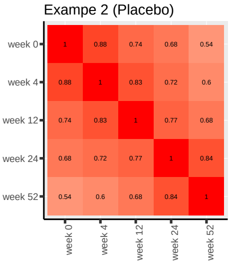
- dispertion over time (standard deviation)

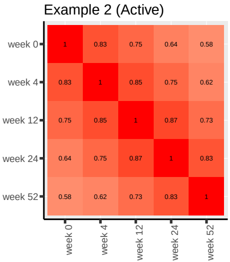




- dependency over time (Pearson correlation)







Typical random effects models (1/2)

Random intercept:

$$Y_{it} = X_{it}\beta + u_i + \varepsilon_{it} \text{ where } u_i \sim \mathcal{N}(0, \tau) \perp\!\!\!\perp \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

Typical random effects models (1/2)

Random intercept:

$$Y_{it} = X_{it}\beta + u_i + \varepsilon_{it} \text{ where } u_i \sim \mathcal{N}(0, \tau) \perp\!\!\!\perp \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

$$\mathbb{V}ar\left[\mathbf{Y}_i\right] = \mathbb{V}ar\left[u_i + \varepsilon_{it}\right] = \begin{bmatrix} \sigma^2 & \tau + \sigma^2 & \tau + \sigma^2 & \dots \\ \tau + \sigma^2 & \sigma^2 & \tau + \sigma^2 & \dots \\ \tau + \sigma^2 & \tau + \sigma^2 & \sigma^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \Omega$$

Typical random effects models (1/2)

Random intercept:

$$Y_{it} = X_{it}\beta + u_i + \varepsilon_{it} \text{ where } u_i \sim \mathcal{N}(0,\tau) \perp\!\!\!\perp \varepsilon_{it} \sim \mathcal{N}(0,\sigma^2)$$

$$\mathbb{V}ar\left[\mathbf{Y}_i\right] = \mathbb{V}ar\left[u_i + \varepsilon_{it}\right] = \begin{bmatrix} \sigma^2 & \tau+\sigma^2 & \tau+\sigma^2 & \dots \\ \tau+\sigma^2 & \sigma^2 & \tau+\sigma^2 & \dots \\ \tau+\sigma^2 & \tau+\sigma^2 & \sigma^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \Omega$$

Implied marginal model

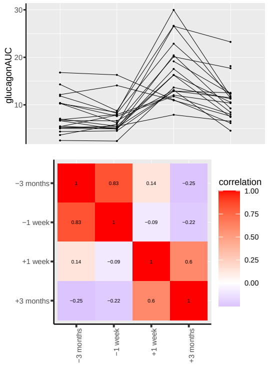
$$\mathbf{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{\nu}_i \text{ where } \boldsymbol{\nu}_i \sim \mathcal{N}(0,\Omega)$$

- variance and correlation constant over time and group
- positive correlation

Back to example 1

Using only two timepoints
(+3 months vs. -3 months)
where there is **no missing values**:

- random intercept model:
  `boundary (singular) fit: see ?isSingular=`



Random intercept cannot handle negative correlation

Typical random effects models (2/2)

Random slope

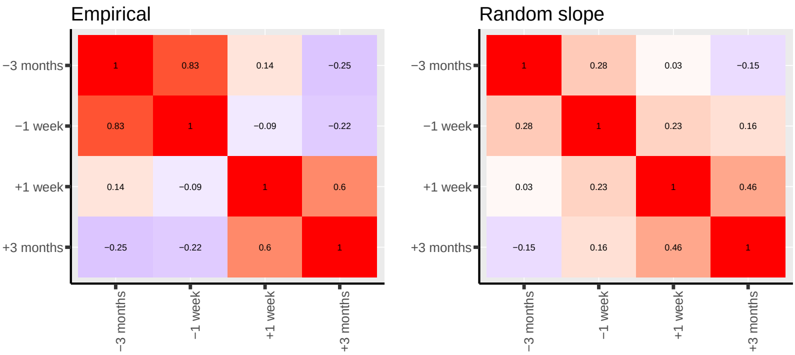
$$Y_{it} = X\beta + u_i + tv_i + \varepsilon_{it}$$
$$\text{where } \begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{12} & \tau_{22} \end{bmatrix}\right) \perp\!\!\!\perp \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2)$$

Implied marginal model

$$\mathbf{Y}_i = \boldsymbol{X}_i\beta + \boldsymbol{\nu}_i \text{ where } \boldsymbol{\nu}_i \sim \mathcal{N}(0, \Omega)$$

- here Ω has a complex expression $\mathbb{C}ov(Y_{is}, Y_{it}) = \tau_{11} + st\tau_{22} + (s+t)\tau_{12}$
- τ_{11}, τ_{22}, τ_{12} variance and correlation parameters

Back to example 1



- 4 parameters for variance and correlation do not provide a good fit (in this example)

Multivariate linear model - general expression

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}(0, \Omega)$$

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \cdots \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \cdots \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} & \cdots \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Multivariate linear model - general expression

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}(0, \Omega)$$

$$\begin{aligned}\Omega = &\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} & \sigma_1\sigma_3\rho_{1,3} & \sigma_1\sigma_4\rho_{1,4} \\ \sigma_2\sigma_1\rho_{1,2} & \sigma_2^2 & \sigma_2\sigma_3\rho_{2,3} & \sigma_2\sigma_4\rho_{2,4} \\ \sigma_3\sigma_1\rho_{1,3} & \sigma_3\sigma_2\rho_{2,3} & \sigma_3^2 & \sigma_3\sigma_4\rho_{3,4} \\ \sigma_4\sigma_1\rho_{1,4} & \sigma_4\sigma_2\rho_{2,4} & \sigma_4\sigma_3\rho_{3,4} & \sigma_4^2 \end{bmatrix} \\ = &diag(\boldsymbol{\sigma}) \times R \times diag(\boldsymbol{\sigma})\end{aligned}$$

Multivariate linear model - general expression

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}(0,\Omega)$$

$$\begin{aligned}\Omega =& diag(\boldsymbol{\sigma}) \times R \times diag(\boldsymbol{\sigma}) \\ =& \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}\end{aligned}$$

"Direct" parametrisation in term of:

- variance ($\sigma \in]0;+\infty[$, $k \in]0;+\infty[$)
- correlation ($\rho \in]-1,1[$)

Multivariate linear model - general expression

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}(0, \Omega)$$

$$\begin{aligned}\Omega =& diag(\boldsymbol{\sigma}) \times R \times diag(\boldsymbol{\sigma}) \\ =& \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}\end{aligned}$$

"Direct" parametrisation in term of:

- variance ($\sigma \in]0; +\infty[$, $k \in]0; +\infty[$)
- correlation ($\rho \in] -1, 1[$)

Assumption:
distinct parameters for the mean, variance, correlation

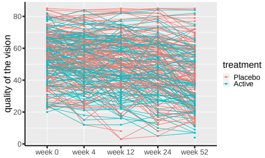
Back to example 2

Full information (2 timepoints):

- Random intercept model:
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.43, \; p = 0.0484$
- Unstructured gaussian model:
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.38, \; p = 0.055$

Full information (5 timepoints):

- Random intercept model
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.90, \; p = 0.0046$
- Unstructured gaussian model
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.87, \; p = 0.037$



simplified imputation model

- Random intercept vs. Unstructured:
  variance constant over time? correlation time independent?

Covariance pattern model in **R**

Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

Covariance pattern model in **R**

Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

LMMstar covariance patterns:
- `CS`: compound symmetry

$$\begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho \\ \sigma^2\rho & \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2\rho & \sigma^2\rho & \sigma^2 \end{bmatrix}$$

-

-

-

Covariance pattern model in **R**

Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

LMMstar covariance patterns:

- `CS`: compound symmetry
  `CS(~ group)` stratification

$$\begin{bmatrix} \sigma_k^2 & \sigma_k^2\rho_k & \sigma_k^2\rho_k & \sigma_k^2\rho_k \\ \sigma_k^2\rho_k & \sigma_k^2 & \sigma_k^2\rho_k & \sigma_k^2\rho_k \\ \sigma_k^2\rho_k & \sigma_k^2\rho_k & \sigma_k^2 & \sigma_k^2\rho_k \\ \sigma_k^2\rho_k & \sigma_k^2\rho_k & \sigma_k^2\rho_k & \sigma_k^2 \end{bmatrix}$$

-
-
-

Covariance pattern model in R

Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

LMMstar covariance patterns:

- `CS`: compound symmetry
  `CS(~ group)` stratification
- `UN` unstructured
  `UN(~ group)` stratification
-
-

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} & \sigma_1\sigma_3\rho_{1,3} & \sigma_1\sigma_4\rho_{1,4} \\ \sigma_2\sigma_1\rho_{1,2} & \sigma_2^2 & \sigma_2\sigma_3\rho_{2,3} & \sigma_2\sigma_4\rho_{2,4} \\ \sigma_3\sigma_1\rho_{1,3} & \sigma_3\sigma_2\rho_{2,3} & \sigma_3^2 & \sigma_3\sigma_4\rho_{3,4} \\ \sigma_4\sigma_1\rho_{1,4} & \sigma_4\sigma_2\rho_{2,4} & \sigma_4\sigma_3\rho_{3,4} & \sigma_4^2 \end{bmatrix}$$

Covariance pattern model in **R**

Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

LMMstar covariance patterns:
- CS: compound symmetry
  CS($\sim$ group) stratification
- UN unstructured
  UN($\sim$ group) stratification
- ...
-

`CS(~baseline)`
where baseline indicates time 1 and 2

$$\begin{bmatrix} \sigma_1^2 & \sigma_1^2\rho_1 & \sigma_1\sigma_2\rho_3 & \sigma_1\sigma_2\rho_3 \\ \sigma_1^2\rho_1 & \sigma_1^2 & \sigma_1\sigma_2\rho_3 & \sigma_1\sigma_2\rho_3 \\ \sigma_1\sigma_2\rho_3 & \sigma_1\sigma_2\rho_3 & \sigma_2^2 & \sigma_2^2\rho_2 \\ \sigma_1\sigma_2\rho_3 & \sigma_1\sigma_2\rho_3 & \sigma_2^2\rho_2 & \sigma_2^2 \end{bmatrix}$$

Covariance pattern model in **R**

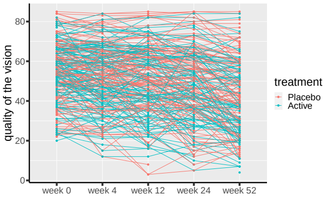
Gold standard: `gls` function

- we developed our own software solution!
  Include several covariance patterns for balanced design

LMMstar covariance patterns:

- `CS`: compound symmetry
  `CS(~ group)` stratification
- `UN` unstructured
  `UN(~ group)` stratification
- `...`
- user-specific pattern

Back to example 2




```
$group

        week 0 week 4 week 12 week 24 week 52
  Placebo   119    119     119     119     119
  Active    121    121     121     121     121

$treat

        week 0 week 4 week 12 week 24 week 52
  Placebo   240    119     119     119     119
  Active      0    121     121     121     121
```

Mixed model with baseline adjustment

At baseline, the groups are comparable. We constrain:

- same baseline mean
- same baseline variance

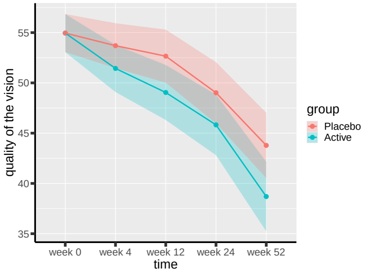
```
lmm(visual ~ time:treat,
    repetition = ~time:treat|subject,
    structure = UN,
    control = list(optimizer = "FS"), data = armd.long)
```

Mixed model with baseline adjustment

At baseline, the groups are comparable. We constrain:

- same baseline mean
- same baseline variance

```
lmm(visual ~ time:treat,
    repetition = ~time:treat|subject,
    structure = UN,
    control = list(optimizer = "FS"), data = armd.long)
```

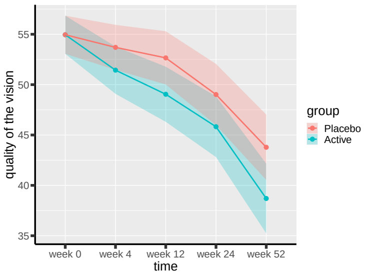


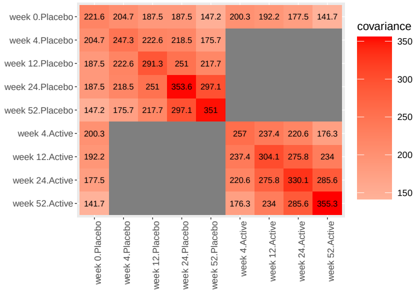
Mixed model with baseline adjustment

At baseline, the groups are comparable. We constrain:

- same baseline mean
- same baseline variance

```
lmm(visual ~ time:treat,
    repetition = ~time:treat|subject,
    structure = UN,
    control = list(optimizer = "FS"), data = armd.long)
```





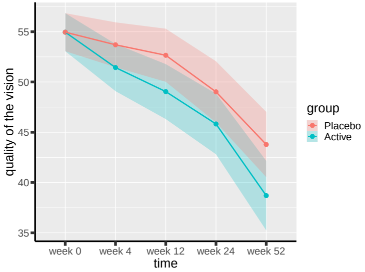
	week 0.Placebo	week 4.Placebo	week 12.Placebo	week 24.Placebo	week 52.Placebo	week 4.Active	week 12.Active	week 24.Active	week 52.Active
week 0.Placebo	221.6	204.7	187.5	187.5	147.2	200.3	192.2	177.5	141.7
week 4.Placebo	204.7	247.3	222.6	218.5	175.7				
week 12.Placebo	187.5	222.6	291.3	251	217.7				
week 24.Placebo	187.5	218.5	251	353.6	297.1				
week 52.Placebo	147.2	175.7	217.7	297.1	351				
week 4.Active	200.3					257	237.4	220.6	176.3
week 12.Active	192.2					237.4	304.1	275.8	234
week 24.Active	177.5					220.6	275.8	330.1	285.6
week 52.Active	141.7					176.3	234	285.6	355.3

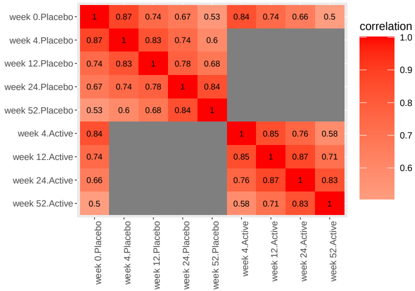
Mixed model with baseline adjustment

At baseline, the groups are comparable. We constrain:

- same baseline mean
- same baseline variance

```
lmm(visual ~ time:treat,
    repetition = ~time:treat|subject,
    structure = UN,
    control = list(optimizer = "FS"), data = armd.long)
```





But what about the ANCOVA?

Complete case analysis (2 timepoints 0-52):

- two sample t-test on the change from baseline: CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.29$, $p = 0.061$
- linear regression `final~baseline+group`: ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.61$, $p = 0.038$

Full information (5 timepoints):

- unstructured gaussian model CHANGE
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.87$, $p = 0.037$
- unstructured gaussian model ANCOVA
  $\Delta\hat{\mu}_A - \Delta\hat{\mu}_P = -4.916$, $p = 0.0308$

Parameter of interest:

CHANGE: $\psi_1 = \mathbb{E}[Y_5 - Y_1 | X = 1] - \mathbb{E}[Y_5 - Y_1 | X = 0]$

ANCOVA: $\psi_2 = \mathbb{E}[Y_5 | X = 1, Y_1 = y_1] - \mathbb{E}[Y_5 | X = 0, Y_1 = y_1]$

Using the same classical result

Gaussian conditional distribution

If $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$ then:

$$Y_2|y_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2\right)$$

$$\mathbb{E}\left[Y_5|X=1, Y_1 = y_1\right] = \mu_5(X=1) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=1)\right)$$

$$\mathbb{E}\left[Y_5|X=0, Y_1 = y_1\right] = \mu_5(X=0) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=0)\right)$$

Using the same classical result

Gaussian conditional distribution

If $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$ then:

$$Y_2|y_1 \ \sim \ \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2 \right)$$

$$\mathbb{E}\left[Y_5 | X = 1, Y_1 = y_1 \right] = \mu_5(X=1) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=1) \right)$$

$$\mathbb{E}\left[Y_5 | X = 0, Y_1 = y_1 \right] = \mu_5(X=0) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=0) \right)$$

So $\psi_2 = \mu_5(X=1) - \mu_5(X=0) - \rho_{5,1}\dfrac{\sigma_5}{\sigma_1}\left(\mu_1(X=1) - \mu_1(X=0) \right)$

Using the same classical result

Gaussian conditional distribution

If $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$ then:

$$Y_2|y_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), (1-\rho^2)\sigma_2^2\right)$$

$$\mathbb{E}\left[Y_5|X=1, Y_1 = y_1\right] = \mu_5(X=1) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=1)\right)$$

$$\mathbb{E}\left[Y_5|X=0, Y_1 = y_1\right] = \mu_5(X=0) + \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(Y_{i1} - \mu_1(X=0)\right)$$

So $\psi_2 = \mu_5(X=1) - \mu_5(X=0) - \rho_{5,1}\frac{\sigma_5}{\sigma_1}\left(\mu_1(X=1) - \mu_1(X=0)\right)$

$\rightarrow$ specific to each covariance pattern

$\rightarrow$ UN, no missing value: same estimate as the usual ANCOVA

Conclusion and perspectives

Take home messages

Mixed models is a transparent way to handle missing values:

- implicit imputation of the conditional mean

Requires correct modeling of the joint covariance structure Ω:

- indirect parametrisation (random effects)
- direct parametrisation (covariance patterns)
  $\rightarrow$ intuitive and very flexible
  $\rightarrow$ well suited for studing balanced designs

Conditional effects can be deduced from the joint model:

- as in the ANCOVA example
- typically involves elements from Ω

Perspective: choice of the covariance pattern

With few repetitions:

- UN is safe pattern
- stratified UN can handle a binary covariate at the cluster level (e.g. treatment)

Otherwise:

- assumptions have to be made ...
- ...which can often be tested,
  e.g. by considering a simple or more complex model
- does selection of the covariance pattern affects statistical inference on the mean parameters?

Perpsective: optimization

Optimization **under constraints**

- R must be positive definite
- may create complex constraints on the support of $\boldsymbol{\rho}$

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}\left(0, \Omega = diag(\boldsymbol{\sigma}) \times R \times diag(\boldsymbol{\sigma})\right)$$

$$\text{where } R(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \dots \\ \rho_{1,2} & 1 & \rho_{2,3} & \dots \\ \rho_{1,3} & \rho_{2,3} & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Perpsective: optimization

Optimization **under constraints**

- R must be positive definite
- may create complex constraints on the support of $\boldsymbol{\rho}$

Transform R or $\boldsymbol{\rho}$ to an unconstrained space

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \varepsilon_i \text{ where } \varepsilon_i \sim \mathcal{N}\left(0, \Omega = diag(\boldsymbol{\sigma}) \times R \times diag(\boldsymbol{\sigma})\right)$$

$$\text{where } R(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \dots \\ \rho_{1,2} & 1 & \rho_{2,3} & \dots \\ \rho_{1,3} & \rho_{2,3} & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Perpsective: optimization

Optimization **under constraints**

- R must be positive definite
- may create complex constraints on the support of ρ

Univariate transformations help but are not enough.

Multivariate transformations
(Pinheiro and Bates, 1996; Zhang et al., 2015):

- Cholesky decomposition $R = r r^\intercal$
- Reparametrization of r to obtain unit column vectors

✔ uniquely defined and guarantee positive define R

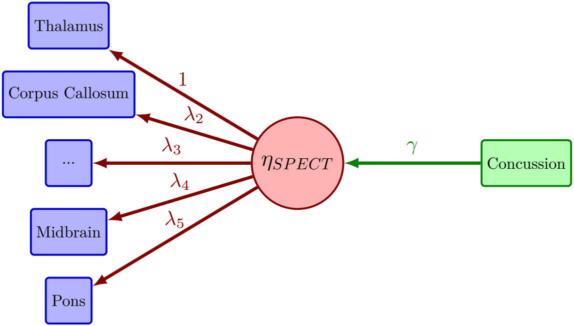
$\rightarrow$ routinely used for UN patterns

✘ difficult to apply to any covariance pattern

Perpsective: dimension reduction

Latent Variable Models can be seen as an extension of mixed models:

- parameters can influence both mean, variance, and correlation!



Comments or questions?

https://www.goodvibeblog.com/got-mixed-feelings/

Reference I

(1997). Interferon Alfa-2a Is Ineffective for Patients With Choroidal Neovascularization Secondary to Age-Related Macular Degeneration: Results of a Prospective Randomized Placebo-Controlled Clinical Trial. *Archives of Ophthalmology*, 115(7):865–872.

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*.

Reference II

Jorsal, T., Christensen, M. M., Mortensen, B., Nygaard, E. B., Zhang, C., Rigbolt, K., Wandall, E., Langholz, E., Friis, S., Worm, D., et al. (2020). Gut mucosal gene expression and metabolic changes after roux-en-y gastric bypass surgery. *Obesity*, 28(11):2163–2174.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and computing*, 6(3):289–296.

Zhang, W., Leng, C., and Tang, C. Y. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):219–238.

Latent Variable Models (LVMs)

A LVM is defined by

- a measurement model:

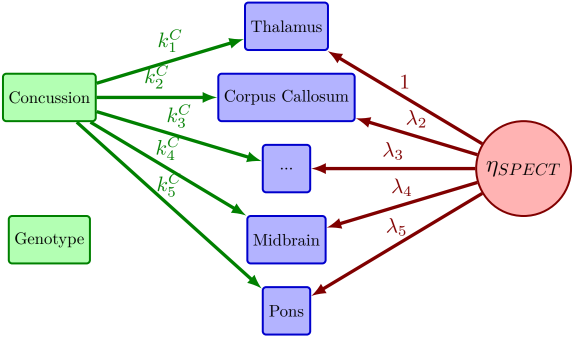
$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\eta}_i \Lambda + \mathbf{X}_i K + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \Omega_\varepsilon\right)$$

- a structural model:

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

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

Illustration on example 3

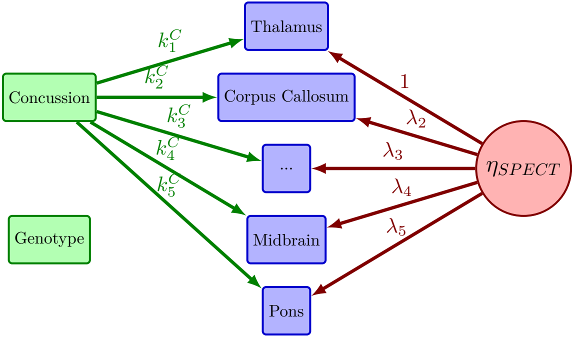


$$Y_{ir} = \nu_r + \eta_i \lambda_r + \text{Concussion}_i k_r^C + \text{Genotype}_i k_r^G + \varepsilon_{ir}$$

$$\text{where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, diag(\sigma_1^2, \ldots, \sigma_9^2)\right)$$

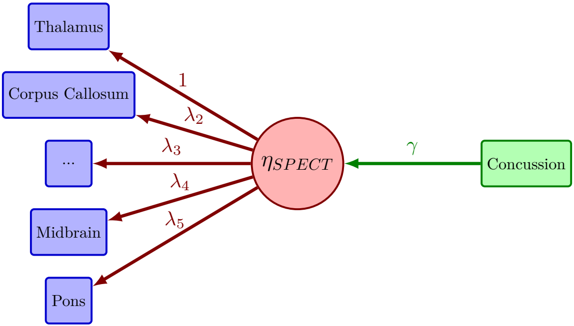
$$\eta_i = \alpha + \zeta_i, \text{ where } \zeta_i \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$$

Illustration on example 3



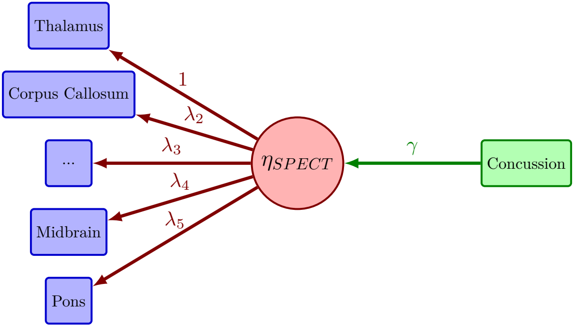
$$\Omega = \begin{bmatrix} \sigma_1^2 + \tau & \lambda_2\tau & \lambda_3\tau & \dots \\ \lambda_2\tau & \sigma_2^2 + \lambda_2^2\tau & \lambda_2\lambda_3\tau & \dots \\ \lambda_3\tau & \lambda_2\lambda_3\tau & \sigma_3^2 + \lambda_3^2\tau & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

LVM for dimension reduction



$$Y_{ir} = \nu_r + \eta_i \lambda_r + \text{Genotype}_i k_r^G + \varepsilon_{ir}$$
$$\text{where } \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, diag(\sigma_1^2, \dots, \sigma_9^2)\right)$$
$$\eta_i = \alpha + \text{Concussion}_i \gamma + \zeta_i, \text{ where } \zeta_i \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$$

LVM for dimension reduction



Same covariance structure but different mean structure:

- γ: global concussion effect
- $(\gamma, \gamma\lambda_2, \ldots, \gamma\lambda_5)$: region specific concussion effect

$\lambda_.$ parameters influence mean, variance, and correlation!