Investigating the impact on dynamic predictions and effect sizes when misspecifying the associations between outcomes

BMS-ANed spring meeting: Modern Mixed Models

Eleni-Rosalina Andrinopoulou, PhD

13 May, 2022

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Introduction

Introduction: Motivation

A lot of information is available

→ Electronic medical records

Introduction: Motivation

A lot of information is available

→ Electronic medical records

Different types of information

→ Baseline characteristics

→ Longitudinal outcomes

→ Time-to-event outcomes

Introduction: Examples

Applications

- ➔ Stroke
- ➔ Cystic Fibrosis

Introduction: Examples

Applications

→ Stroke

- ◇ Action Research Arm Test
- ◇ Fugl-Meyer Upper Extremity
- ◇ Barthel Index

→ Cystic Fibrosis

Introduction: Examples

Applications

- ➔ Stroke
- ➔ Cystic Fibrosis

- ◇ FEV_1
- ◇ BMI
- ◇ Time-to death/exacerbation

Introduction: Common practice

Separate analysis

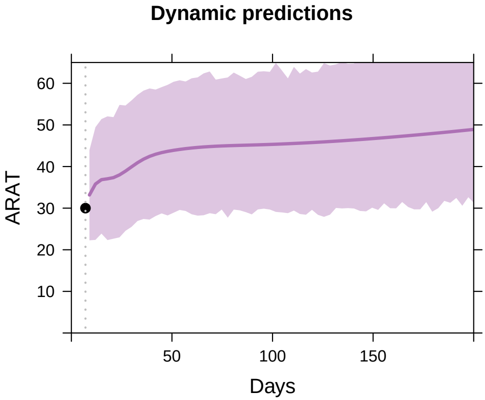
- ➔ Each longitudinal outcome
- ➔ Survival outcomes

Introduction: Common practice

Separate analysis - Stroke data

- 450 patients
- Outcome:
  Action Research Arm Test

Selles, R. W., Andrinopoulou, E. R., et al (2021). Computerised patient-specific prediction of the recovery profile of upper limb capacity within stroke services: the next step. Journal of Neurology, Neurosurgery & Psychiatry.

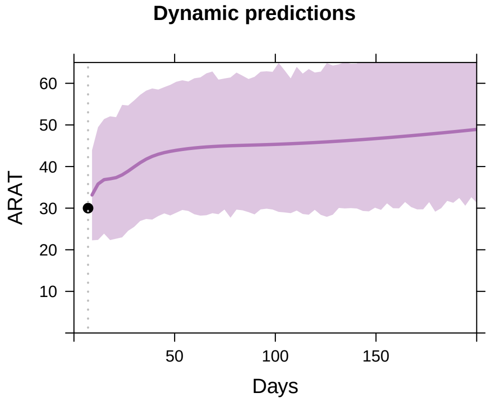


Introduction: Common practice

Separate analysis - Stroke data

- ◇ 450 patients
- ◇ Outcome:
  Action Research Arm Test

Selles, R. W., Andrinopoulou, E. R., et al (2021). Computerised patient-specific prediction of the recovery profile of upper limb capacity within stroke services: the next step. Journal of Neurology, Neurosurgery & Psychiatry.



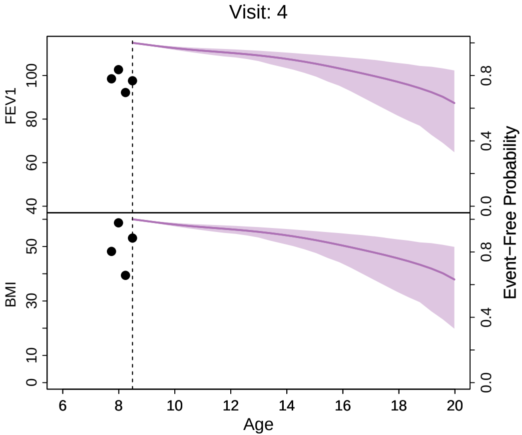
The SAFE model is used in clinical practice!

Introduction: Extensions

Combined analysis - Cystic Fibrosis data

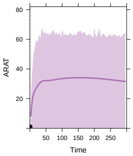
- ◇ 17,100 patients
- ◇ Outcomes:
  - **FEV_1**
  - **BMI**
  - **Exacerbation**

Andrinopoulou, E. R., Clancy, J. P., & Szczesniak, R. D. (2020). Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. BMC pulmonary medicine, 20, 1-11.



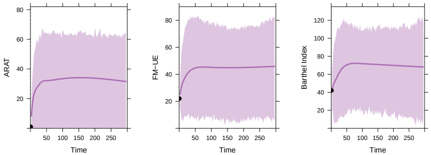
Introduction: Challenges and Opportunities

Separate prediction models



Introduction: Challenges and Opportunities

Single prediction model



Statistical Models

Let's assume that we have a longitudinal outcome

Statistical Models: Mixed Models

$$
\begin{aligned}
g_1[E\{y_{i1}(t) \mid \boldsymbol{b}_{i1}\}] &= m_{i1}(t) \\
&= x_{i1}^\top(t)\boldsymbol{\beta_1} + z_{i1}^\top(t)\boldsymbol{b}_{i1}
\end{aligned}
$$

where

- $g_1(.)$ is the link function
- $\boldsymbol{b}_{i1} \sim N(\mathbf{0}, \mathbf{\Sigma}_b^2)$

Statistical Models

Let's assume that we have multiple longitudinal outcomes

Statistical Models: Multivariate Mixed Models

$$\begin{aligned} g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] &= m_{ik}(t) \\ &= x_{ik}^\top(t)\boldsymbol{\beta}_1 + z_{ik}(t)^\top \boldsymbol{b}_{ik} \end{aligned}$$

where

- $\boldsymbol{b}_i^\top = (\boldsymbol{b}_{ik}^\top, \ldots, \boldsymbol{b}_{iK}^\top) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b^2)$

Statistical Models: Multivariate Mixed Models

$$
\begin{aligned}
g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] &= m_{ik}(t) \\
&= x_{ik}^\top(t)\boldsymbol{\beta}_1 + z_{ik}(t)^\top \boldsymbol{b}_{ik}
\end{aligned}
$$

where

- $\boldsymbol{b}_i^\top = (\boldsymbol{b}_{ik}^\top, \ldots, \boldsymbol{b}_{iK}^\top) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b^2)$

Challenge: Quantify the association between the outcomes

Statistical Models: Multivariate Mixed Models

Erasmus MC University Medical Center Rotterdam

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = m_{ik}(t) + \alpha \; f\{\mathcal{M}_{ip}(t)\}$$

where

- α denotes the association
- $\mathcal{M}_{ip}(t)$ denotes the history of the true unobserved longitudinal process up to time point t
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Statistical Models: Multivariate Mixed Models

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = m_{ik}(t) + \alpha m_{ip}(t)$$

where

- α denotes the association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

Statistical Models: Multivariate Mixed Models

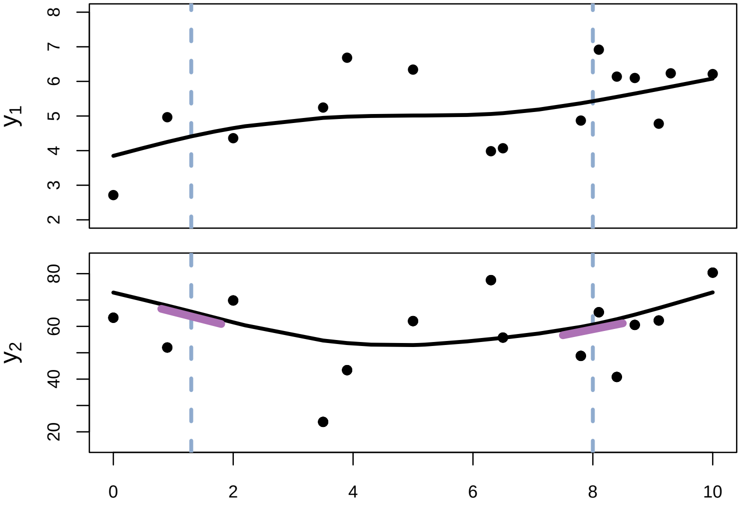
$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = m_{ik}(t) + \alpha m_{ip}(t)$$

where

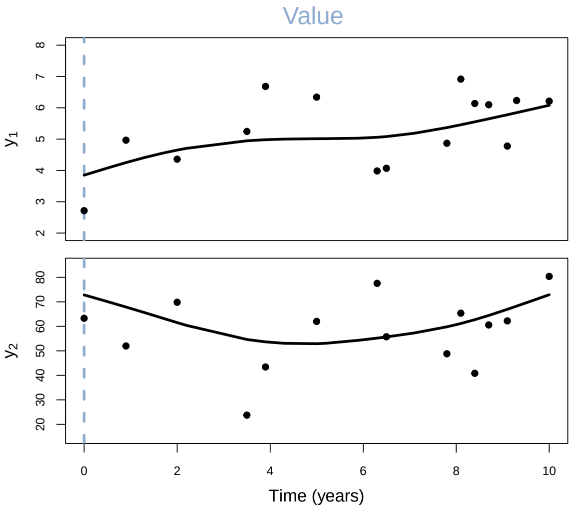
- α denotes the association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

Challenge: Is that our only option?

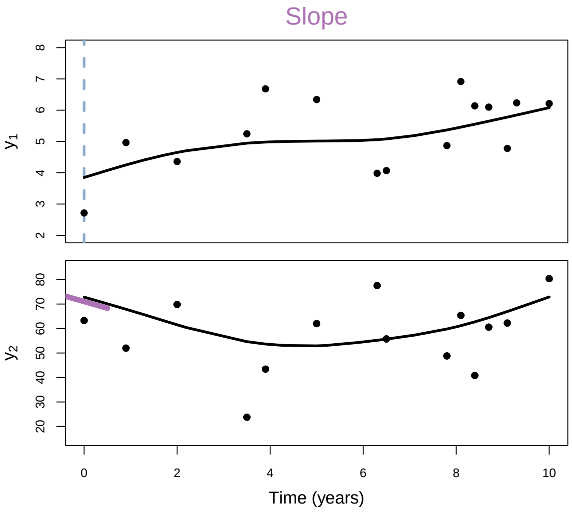
Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models



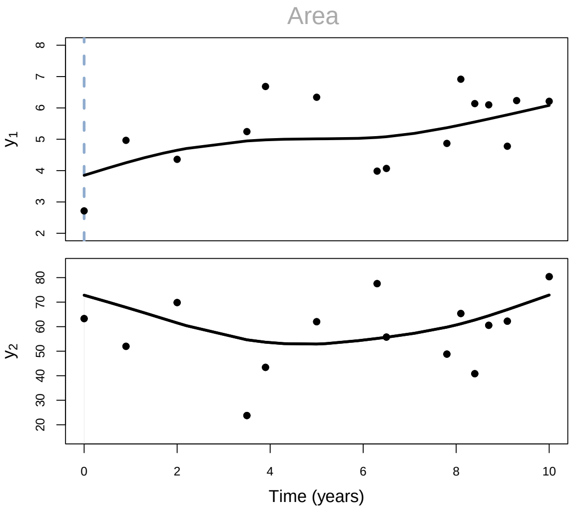
Statistical Models: Multivariate Mixed Models

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = m_{ik}(t) + \alpha \frac{d}{dt} m_{ip}(t),$$

where

- α denotes the association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = m_{ik}(t) + \alpha \int_0^t m_{ip}(s)dt,$$

where

- α denotes the association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

Statistical Models

Let's assume that we have multiple longitudinal and a survival outcome

Statistical Models: Multivariate Joint Models

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = x_{ik}^\top(t)\boldsymbol{\beta}_1 + z_{ik}(t)^\top \boldsymbol{b}_{ik}$$

$$h_i(t) = h_0(t)[\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \sum_{k=1}^{K} \boldsymbol{\alpha}_{Sk} f\{\mathcal{M}_{ik}(t)\}],$$

where

- $\boldsymbol{\alpha}_{Sk}$ denote the associations

Statistical Models

What about the association between the longitudinal outcomes?

Statistical Models: Multivariate Joint Models

Erasmus MC University Medical Center Rotterdam

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = x_{ik}^\top(t)\boldsymbol{\beta}_1 + z_{ik}(t)^\top \boldsymbol{b}_{ik} + \alpha_L f\{\mathcal{M}_{ip}(t)\}$$

$$h_i(t) = h_0(t)[\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \sum_{k=1}^{K} \boldsymbol{\alpha}_{Sk} f\{\mathcal{M}_{ik}(t)\}]$$

where

- $\boldsymbol{\alpha}_{Sk}$ denotes the survival association
- α_L denotes the longitudinal association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

Statistical Models: Multivariate Joint Models

$$g_k[E\{y_{ik}(t) \mid \boldsymbol{b}_{ik}\}] = x_{ik}^{\top}(t)\boldsymbol{\beta}_1 + z_{ik}(t)^{\top}\boldsymbol{b}_{ik} + \alpha_L f\{\mathcal{M}_{ip}(t)\}$$

$$h_i(t) = h_0(t)[\boldsymbol{\gamma}^{\top}\boldsymbol{w}_i + \sum_{k=1}^{K} \boldsymbol{\alpha}_{Sk} f\{\mathcal{M}_{ik}(t)\}]$$

where

- $\boldsymbol{\alpha}_{Sk}$ denotes the survival association
- α_L denotes the longitudinal association
- p indicates a longitudinal outcome that is not the main outcome k $(p \neq k)$

- **Multiple functional forms and outcomes**: Shrinkage

Andrinopoulou, E. R.,& Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. Statistics in medicine, 35(26), 4813-4823.

Statistical Models: Prognostic models

Personalized dynamic predictions

- We assume the following setting for a new patient l
  - all baseline information
  - available longitudinal outcomes (K) up to time t, $\tilde{Y}_{lk}(t) = y_{lk}(s), 0 \leq s < t$
- We are interested in **future longitudinal outcomes / events** in the medically relevant interval $(t, t + \Delta t]$

Based on the models we can get

- $E\{y_{lk}(t + \Delta t) \mid \tilde{Y}_{lk}(t), \boldsymbol{D}_n\}$
- $Pr\{T_l^* \geq t + \Delta t \mid T_l^* > t, \tilde{Y}_{lk}(t), \boldsymbol{D}_n\}$

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Statistical Models: Prognostic models

Measuring Predictive Performance

- **Longitudinal and survival outcomes**: the distance between the predicted outcome and the actual outcome (PE)
- **Survival outcomes**: how well can the longitudinal biomarker(s) discriminate between subject of low and high risk for the event (AUC)

Andrinopoulou, E. R., Eilers, P. H., Takkenberg, J. J., & Rizopoulos, D. (2018). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. Biometrics, 74(2), 685-693.

Andrinopoulou, E. R., Harhay, M. O., Ratcliffe, S. J., & Rizopoulos, D. (2021). Reflection on modern methods: Dynamic prediction using joint models of longitudinal and time-to-event data. International Journal of Epidemiology, 50(5), 1731-1743.

Simulations

Fit Joint Models

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Value of longitudinal outcome

Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time

Treatment

→ **Survival outcome**

Treatment

Value of longitudinal outcome

Fit

→ **Longitudinal outcome**

Non linear time

Treatment

→ **Survival outcome**

Treatment

Slope/Area of longitudinal outcome

Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Value of longitudinal outcome

Fit

→ **Longitudinal outcome**

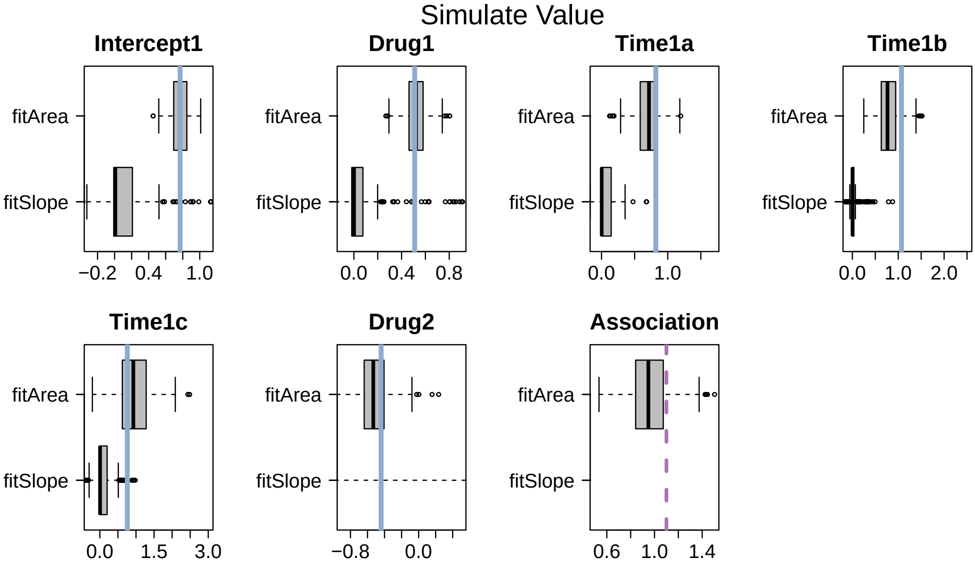
Non linear time
Treatment

→ **Survival outcome**

Treatment
Slope/Area of longitudinal outcome

All models were fitted under the Bayesian framework

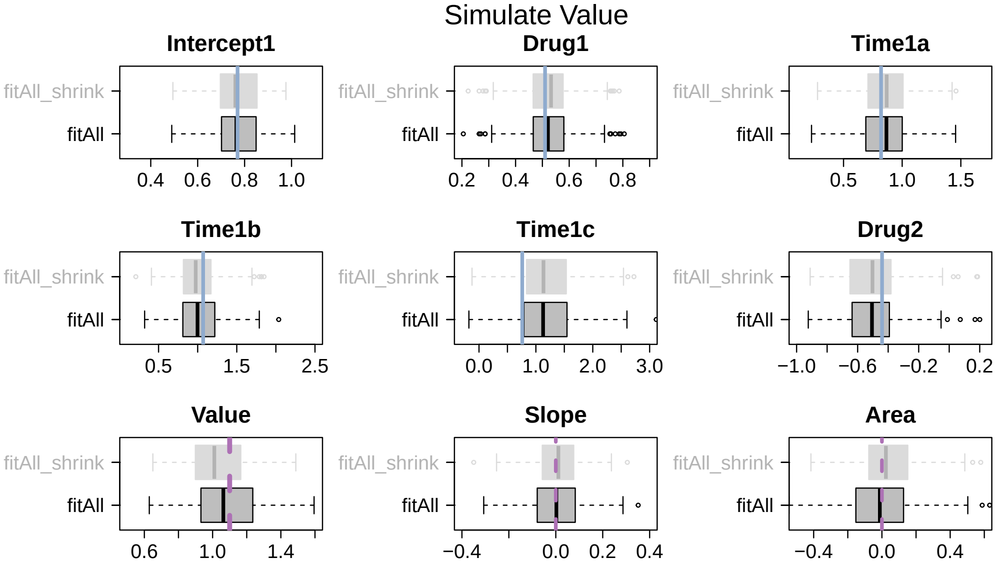
Simulations: Results



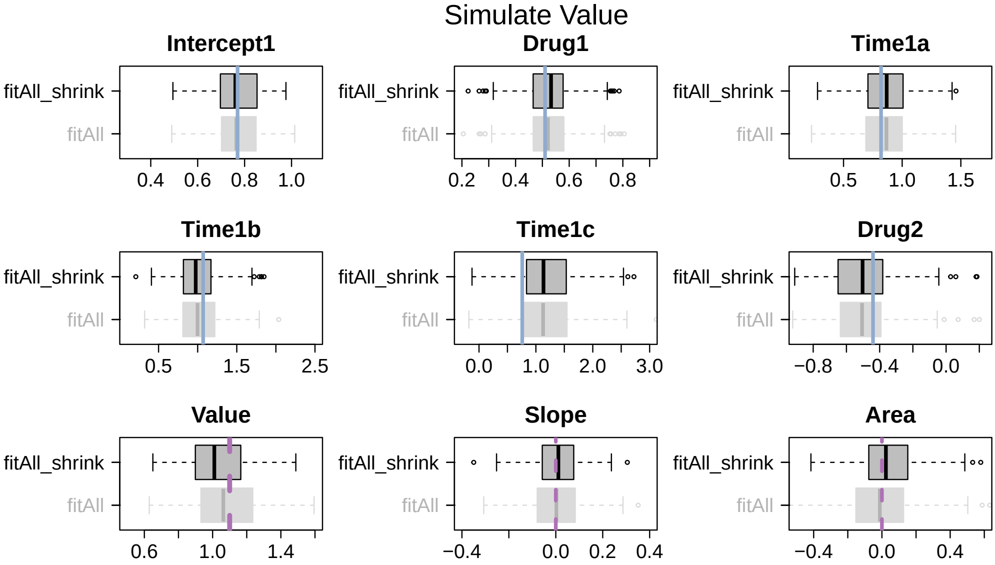
Simulations

Fit Joint Models with all Functional Forms

Simulations: Results



Simulations: Results



Simulations

Fit Multivariate Joint Models

Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Non linear time

Treatment

Value of longitudinal outcome 2

➔ **Longitudinal outcome 2**

Linear time

➔ **Survival outcome**

Treatment

Value of longitudinal outcome 1

Simulations: Scenario

Simulate

→ **Longitudinal outcome 1**

Non linear time
Treatment
Value of longitudinal outcome 2

→ **Longitudinal outcome 2**

Linear time

→ **Survival outcome**

Treatment
Value of longitudinal outcome 1

Fit

→ **Longitudinal outcome 1**

Non linear time
Treatment
~~Value of longitudinal outcome 2~~

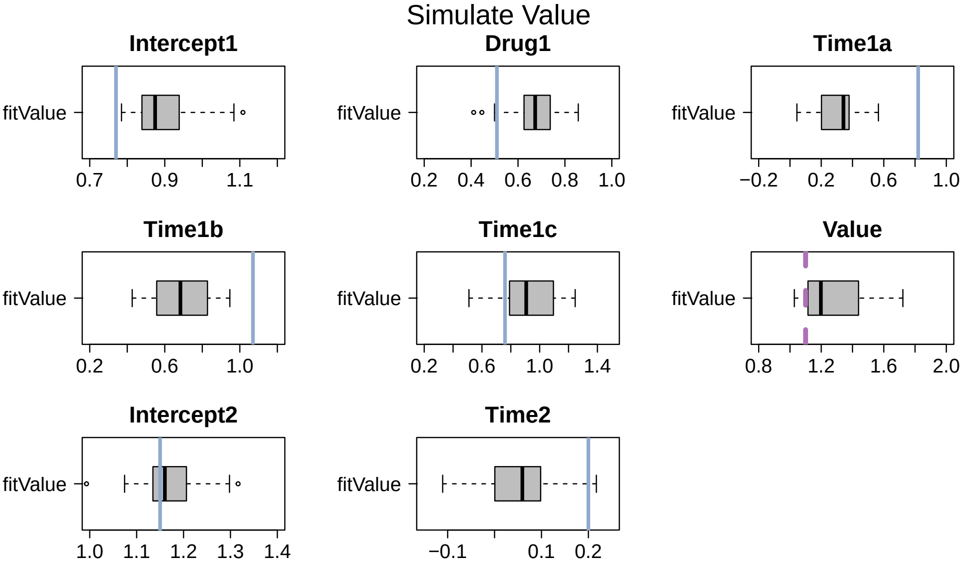
→ **Longitudinal outcome 2**

Linear time

→ **Survival outcome**

Treatment
Value of longitudinal outcome 1

Simulations: Results



Simulations

Prediction in Joint Models

Simulations: Scenario

- ➔ Split simulated data in 5 subsets
- ➔ Use 4 subsets to fit the model
- ➔ Obtain predictions at $(t, t + \Delta t]$ for the patients left out (1 subset)
- ➔ Calculate AUC and PE
- ➔ Repeat the above steps 100 times

Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time

Treatment

→ **Survival outcome**

Treatment

Value of longitudinal outcome

Simulations: Scenario

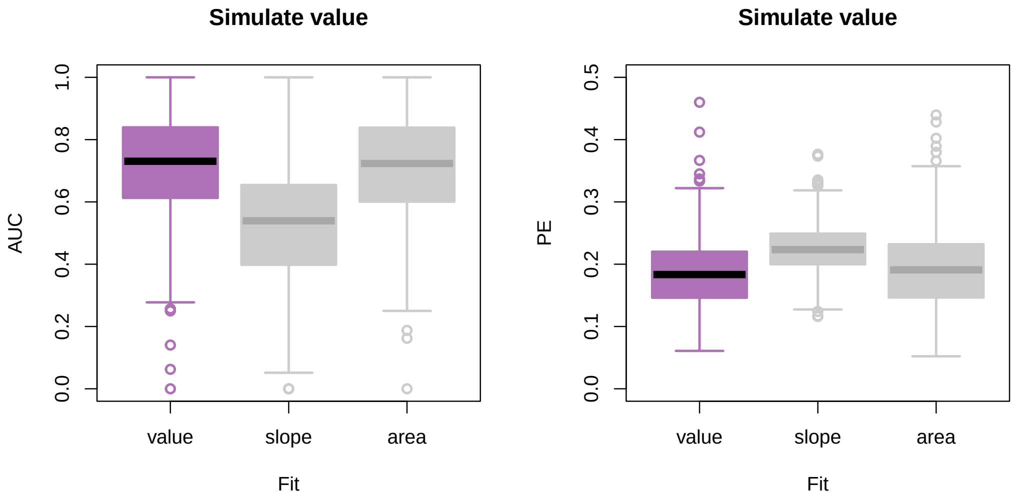
Simulate

- → **Longitudinal outcome**
  - Non linear time
  - Treatment
- → **Survival outcome**
  - Treatment
  - Value of longitudinal outcome

Predict

- → **Longitudinal outcome**
  - Non linear time
  - Treatment
- → **Survival outcome**
  - Treatment
  - Value/slope/Area of longitudinal outcome

Simulations: Results



Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Slope of longitudinal outcome

Simulations: Scenario

Erasmus MC
University Medical Center Rotterdam

Simulate

→ **Longitudinal outcome**

Non linear time

Treatment

→ **Survival outcome**

Treatment

Slope of longitudinal outcome

Predict

→ **Longitudinal outcome**

Non linear time

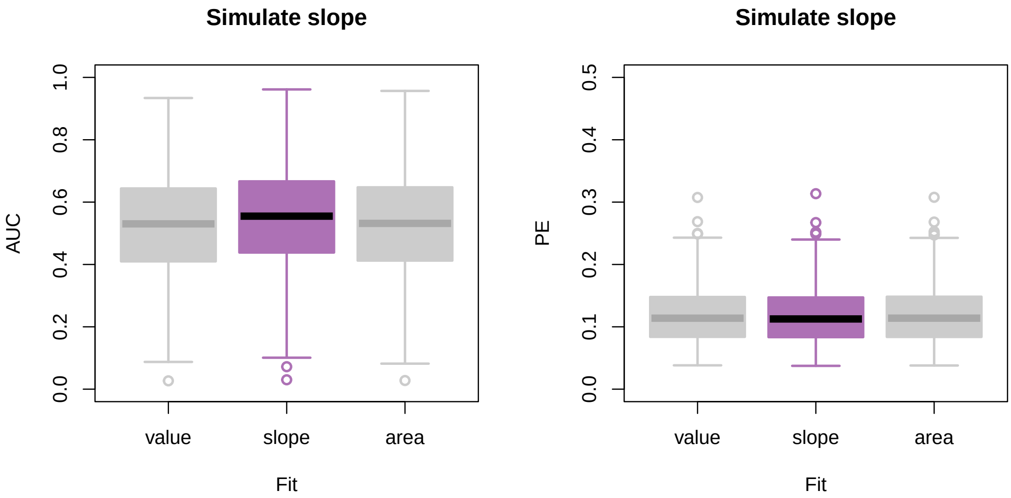
Treatment

→ **Survival outcome**

Treatment

Value/Slope/Area of longitudinal outcome

Simulations: Results



Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Area of longitudinal outcome

Simulations: Scenario

Simulate

➔ **Longitudinal outcome**

Non linear time
Treatment

➔ **Survival outcome**

Treatment
Area of longitudinal outcome

Predict

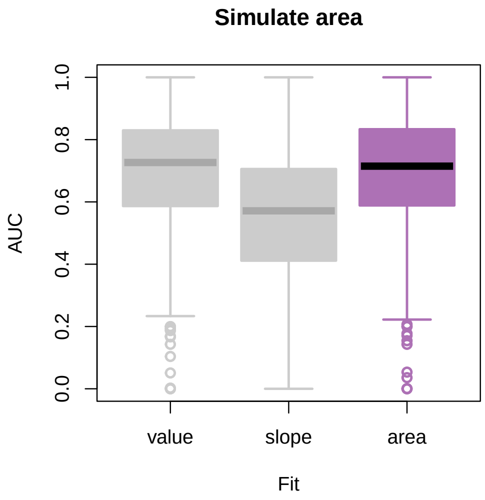
➔ **Longitudinal outcome**

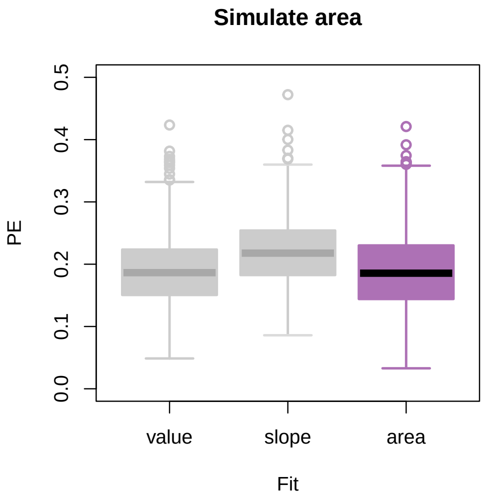
Non linear time
Treatment

➔ **Survival outcome**

Treatment
Value/Slope/Area of longitudinal outcome

Simulations: Results





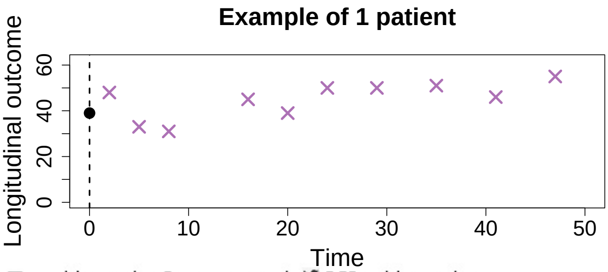
Simulations

Prediction in Multivariate Mixed Models

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou

Simulations: Scenario

- ➔ Split simulated data in 5 subsets
- ➔ Use 4 subsets to fit the model
- ➔ Obtain predictions at every future t for the patients left out (1 subset)
- ➔ Calculate PE
- ➔ Repeat the above steps 100 times



Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Value of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Value of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Predict

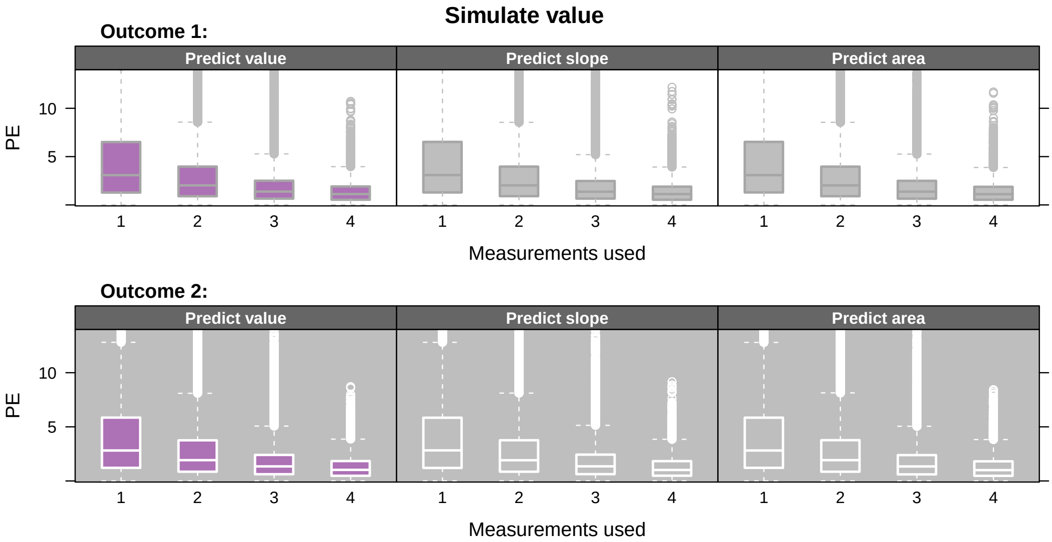
➔ **Longitudinal outcome 1**

Linear time
Treatment
Value/Slope/Area of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Simulations: Results



Simulations: Scenario

Simulate

→ **Longitudinal outcome 1**

Linear time
Treatment
Slope of outcome 2

→ **Longitudinal outcome 2**

Linear time

Simulations: Scenario

Simulate

→ **Longitudinal outcome 1**

Linear time
Treatment
Slope of outcome 2

→ **Longitudinal outcome 2**

Linear time

Predict

→ **Longitudinal outcome 1**

Linear time
Treatment
Value/Slope/Area of outcome 2

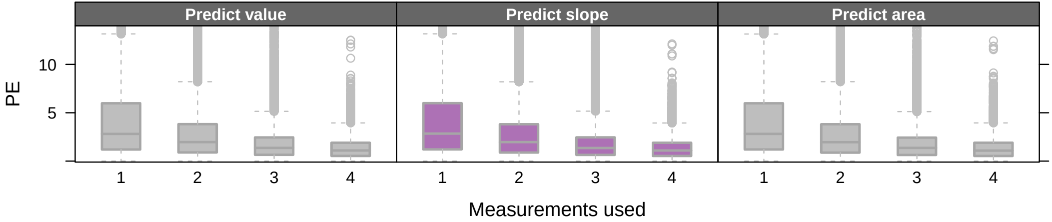
→ **Longitudinal outcome 2**

Linear time

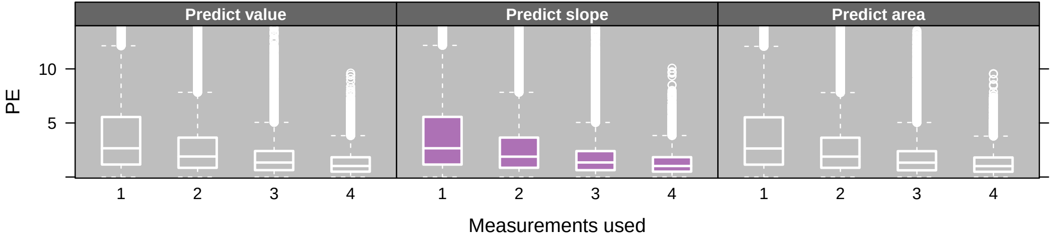
Simulations: Results

Simulate slope

Outcome 1:



Outcome 2:



Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Area of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Area of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Predict

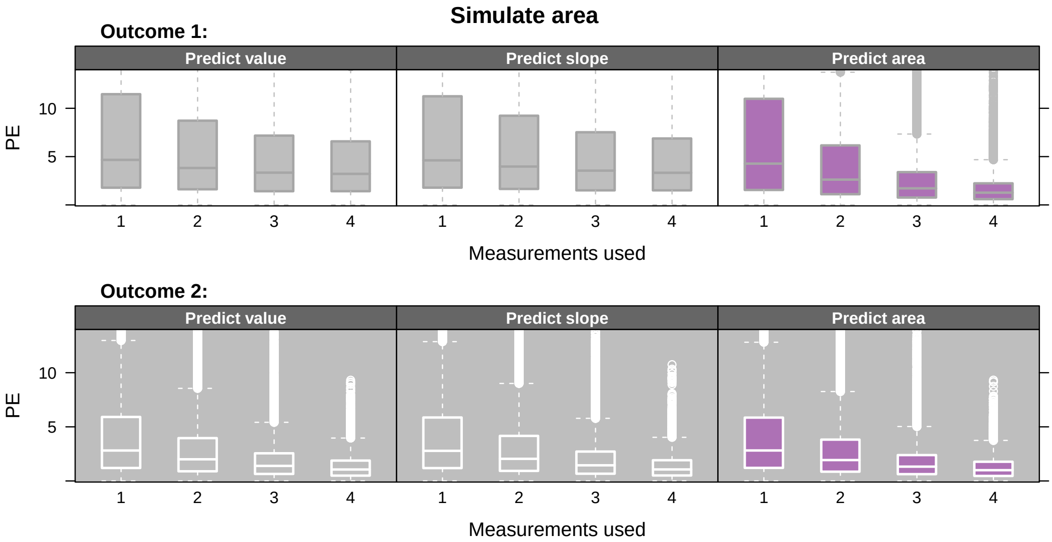
➔ **Longitudinal outcome 1**

Linear time
Treatment
Value/Slope/Area of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Simulations: Results



Summary and Discussion

Summary and Discussion

→ A lot of information is available

→ Correlation between outcomes

Summary and Discussion

- ➔ A lot of information is available
- ➔ Correlation between outcomes

- ➔ Challenges and opportunities
  - ◇ Functional forms

Thank you for your attention!

www.erandrinopoulou.com eandrinopoulou@erasmusmc.nl @ERandrinopoulou