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

Research Seminars, Athens University of Economics and Business, Department of Statistics

Eleni-Rosalina Andrinopoulou, PhD

11 March, 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

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

Introduction: Examples

Applications

- ➔ Stroke
- ➔ Head & Neck Cancer
- ➔ Cystic Fibrosis

Introduction: Examples

Applications

- → Stroke
  - ◇ Action Research Arm Test
  - ◇ Fugl-Meyer Upper Extremity
  - ◇ Barthel Index
- → Head & Neck Cancer
- → Cystic Fibrosis

Introduction: Examples

Applications

- ➔ Stroke
- ➔ Head & Neck Cancer
  - ◇ Quality of life
  - ◇ Time-to recurrence
- ➔ Cystic Fibrosis

Introduction: Examples

Applications

- ➔ Stroke
- ➔ Head & Neck Cancer
- ➔ Cystic Fibrosis

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

Introduction: Common practice

Separate analysis

- ➔ Each longitudinal outcome
- ➔ Survival outcomes

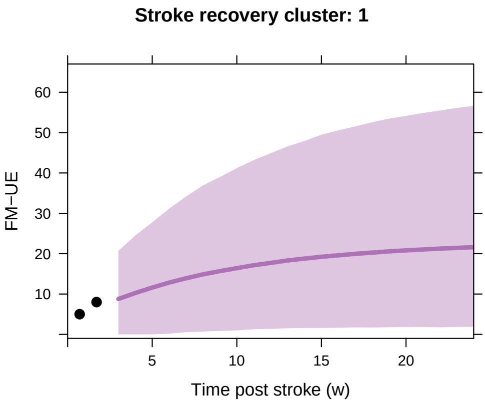
Introduction: Common practice

Separate analysis - Stroke data

- 412 patients
- Outcome:

  Fugl–Meyer

van der Vliet, R., Selles, R. W., Andrinopoulou, E. R., et al (2020). Predicting upper limb motor impairment recovery after stroke: a mixture model. Annals of Neurology.



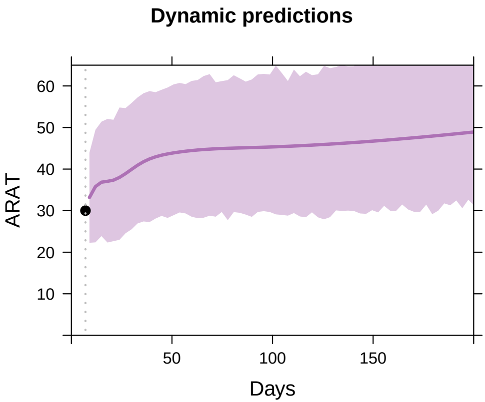
Introduction: Common practice

Separate analysis - Stroke data

- 450 patients
- Outcome:

  Action Research Arm Test

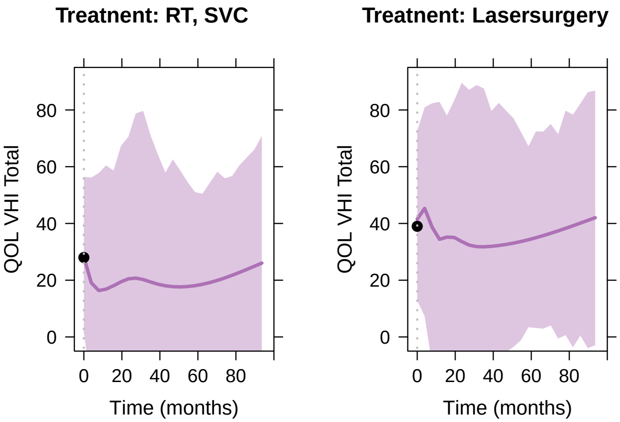
Selles, R. W., Andrinopoulou, E. R., et al (2021). Computerised patient-specific prediction of the recovery profile of upper limb capacity within stroke services: the next step. Journal of Neurology, Neurosurgery & Psychiatry.



Introduction: Common practice

Separate analysis - Head & neck cancer data

- 293 patients
- Outcome:
  Quality of life



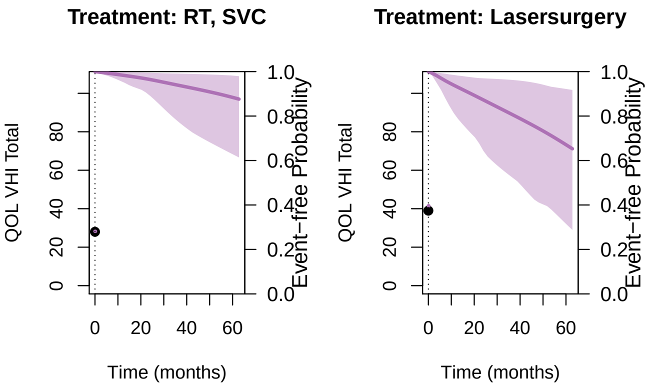
Introduction: Extensions

Combined analysis - Head & neck cancer data

- 293 patients
- Outcome:

  Quality of life

  Recurrence

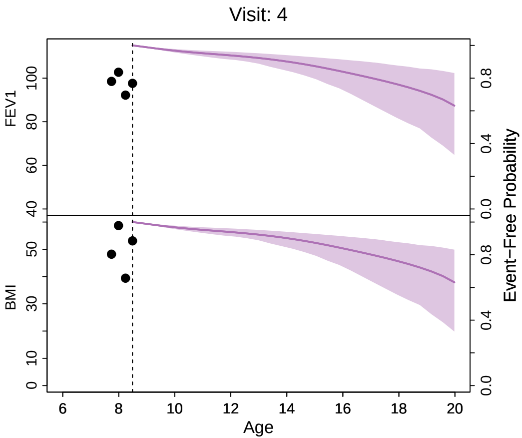


Introduction: Extensions

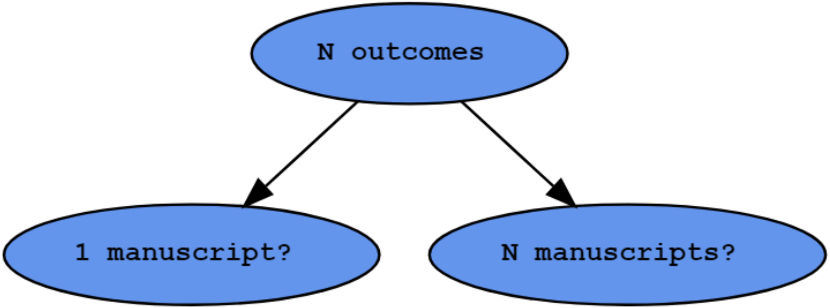
Combined analysis - Cystic Fibrosis data

- 17,100 patients
- Outcomes:
  - **$\mathbf{FEV}_1$**
  - **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



Statistical Models

Let's assume that we have a longitudinal outcome

Statistical Models: Mixed Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^{\top}(t)\beta_1 + z_{1i}^{\top}(t)b_{1i} + \epsilon_{1i}(t)$$

where

- $b_{1i} \sim N(0, D)$
- $\epsilon_{1i}(t) \sim N(0, \Sigma_{1i})$

Statistical Models

Let's assume that we have two longitudinal outcomes

Statistical Models: Multivariate Mixed Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

where

- $b_i^\top = (b_{1i}^\top, b_{2i}^\top) \sim N(0, D)$

Statistical Models: Multivariate Mixed Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top} b_{1i} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top} b_{2i} + \epsilon_{2i}(t)$$

where

- $b_i^{\top} = (b_{1i}^{\top}, b_{2i}^{\top}) \sim N(0, D)$

Challenge: Quantify the association between y_1 and y_2

Statistical Models: Multivariate Mixed Models

Erasmus MC University Medical Center Rotterdam

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha \, \boxed{f\{\mathcal{M}_{2i}(t)\}} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

where

- α denotes the association
- $\mathcal{M}_{2i}(t)$ denotes the history of the true unobserved longitudinal process up to time point t

Statistical Models: Multivariate Mixed Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^{\top}(t)\beta_1 + z_{1i}(t)^{\top}b_{1i} + \alpha m_{2i}(t) + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^{\top}(t)\beta_1 + z_{2i}(t)^{\top}b_{2i} + \epsilon_{2i}(t)$$

where

- α denotes the association

Statistical Models: Multivariate Mixed Models

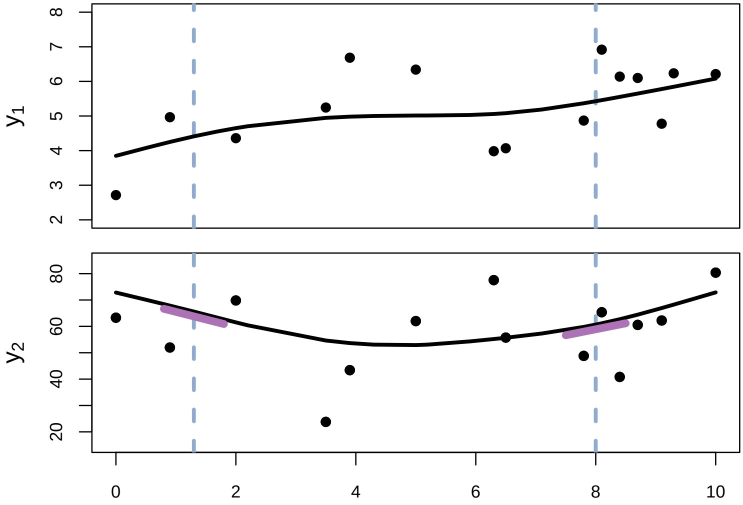
$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha m_{2i}(t) + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

where

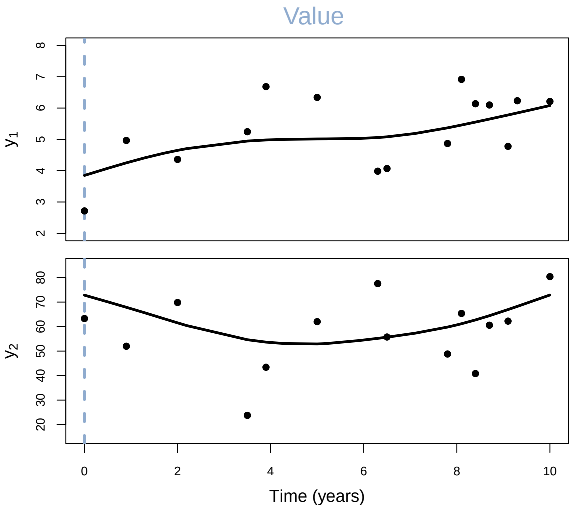
- α denotes the association

Challenge: Is that our only option?

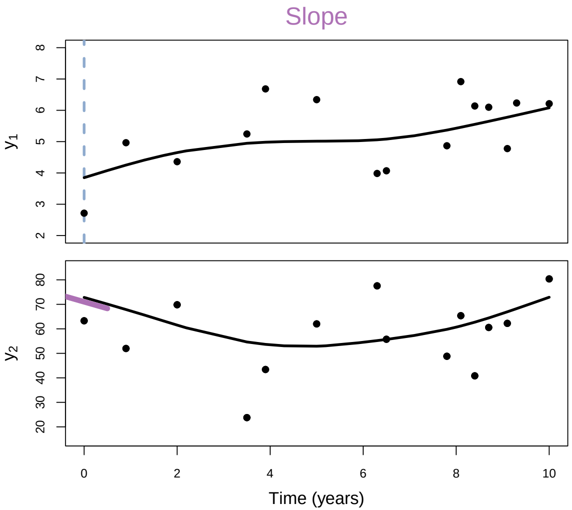
Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models

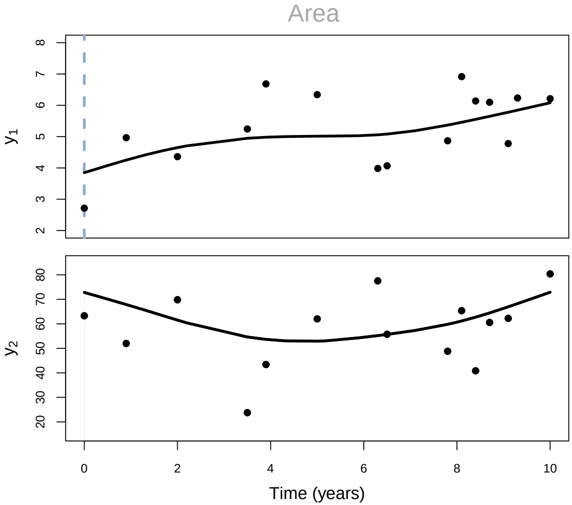
$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha \boxed{\frac{d}{dt} m_{2i}(t)} + \epsilon_{1i}(t),$$

$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t),$$

where

- α denotes the association

Statistical Models: Multivariate Mixed Models



Statistical Models: Multivariate Mixed Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha \int_0^t m_{2i}(s)dt + \epsilon_{1i}(t),$$

$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t),$$

where

- α denotes the association

Statistical Models

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

Statistical Models: Multivariate Joint Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

$$h_i(t) = h_0(t)[\gamma^\top w_i + \alpha_{S1} f\{\mathcal{M}_{1i}(t)\} + \alpha_{S2} f\{\mathcal{M}_{2i}(t)\}],$$

where

- α_{S1} and α_{S2} denote the associations

Statistical Models

What about the association between the longitudinal outcomes?

Statistical Models: Multivariate Joint Models

Erasmus MC University Medical Center Rotterdam

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha_L f\{\mathcal{M}_{2i}(t)\} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

$$h_i(t) = h_0(t)[\gamma^\top w_i + \alpha_S f\{\mathcal{M}_{1i}(t)\}]$$

where

- α_S denotes the survival association
- α_L denotes the longitudinal association

Statistical Models: Multivariate Joint Models

$$y_{1i}(t) = m_{1i}(t) + \epsilon_{1i} = x_{1i}^\top(t)\beta_1 + z_{1i}(t)^\top b_{1i} + \alpha_L f\{\mathcal{M}_{2i}(t)\} + \epsilon_{1i}(t)$$
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

$$h_i(t) = h_0(t)[\gamma^\top w_i + \alpha_S f\{\mathcal{M}_{1i}(t)\}]$$

where

- α_S denotes the survival association
- α_L denotes the longitudinal association

- **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

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

Statistical Models: Prognostic models

Measuring Predictive Performance

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

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

Fitting 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

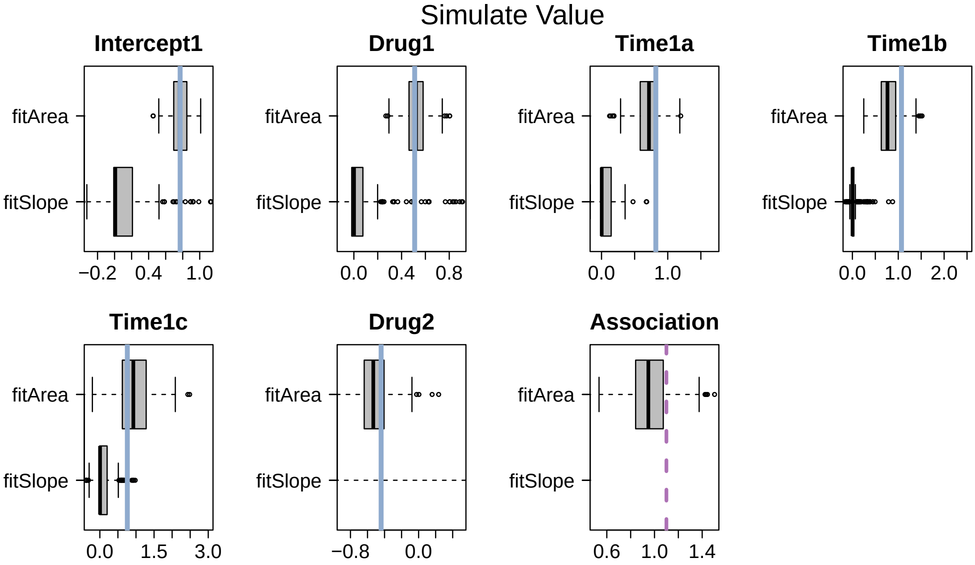
→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Slope/Area of longitudinal outcome

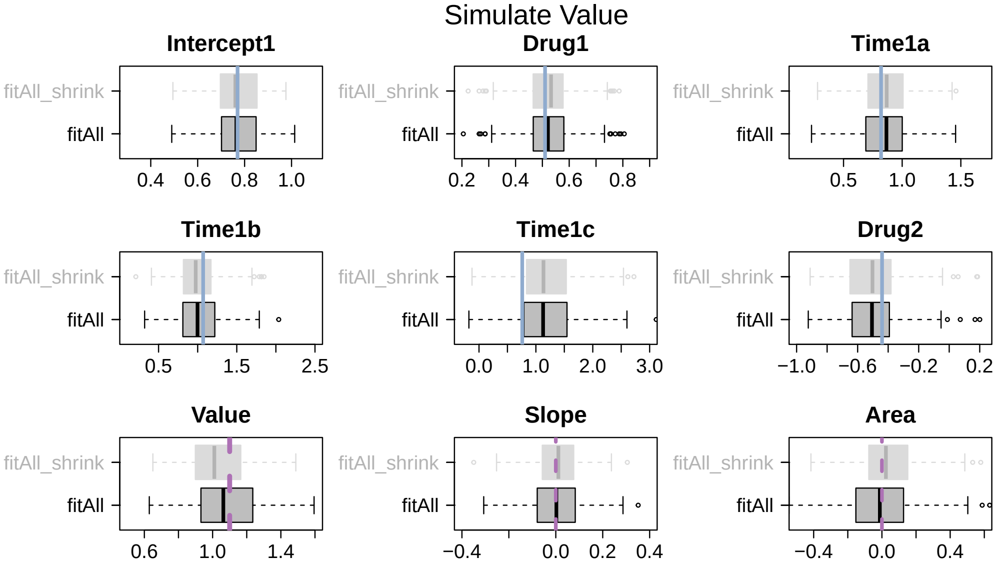
Simulations: Results



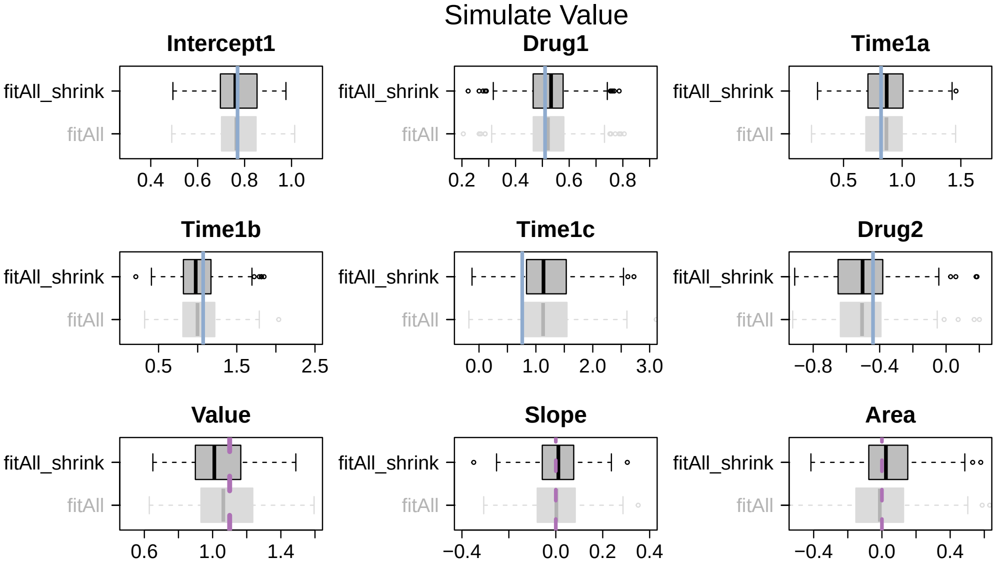
Simulations

Fitting Joint Models with all Functional Forms

Simulations: Results



Simulations: Results



Simulations

Fitting Multivariate Joint Models

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

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~~

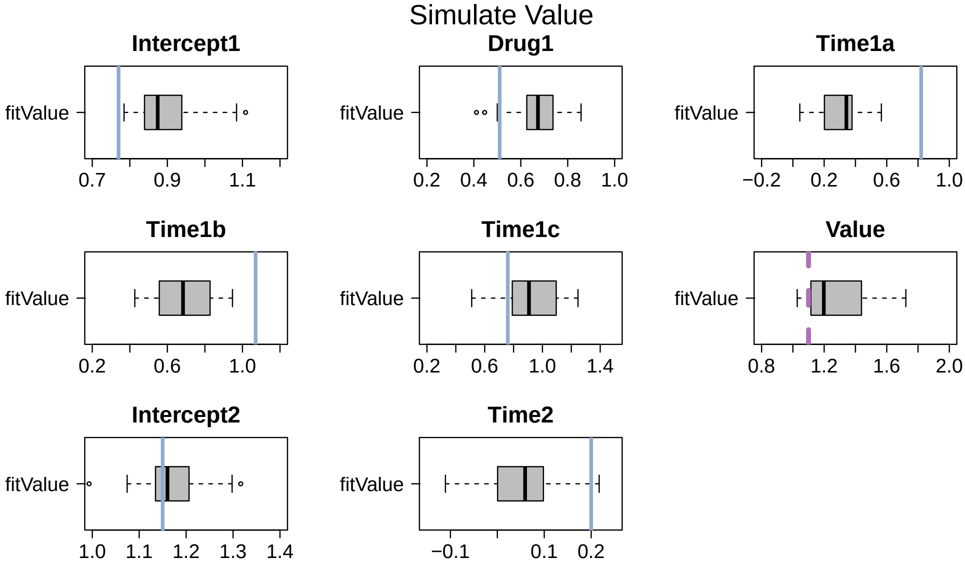
➔ **Longitudinal outcome 2**

Linear time

➔ **Survival outcome**

Treatment
Value of longitudinal outcome 1

Simulations: Results



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

Simulations

Prediction in Joint 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 for the patients left out (1 subset)
- → Calculate AUC and PE

Simulations: Scenario

Simulate

→ **Longitudinal outcome**

Non linear time
Treatment

→ **Survival outcome**

Treatment
Value of longitudinal outcome

Simulations: Scenario

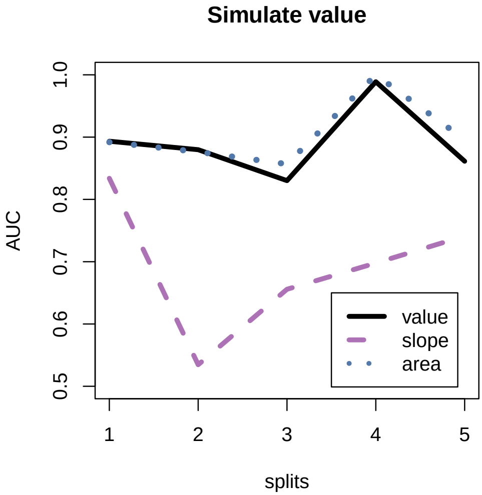
Simulate

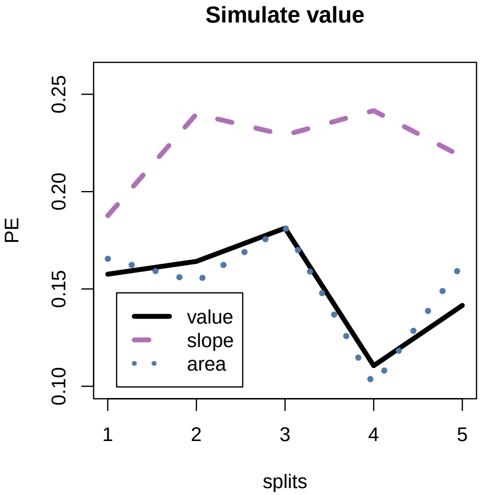
- ➔ **Longitudinal outcome**
  - Non linear time
  - Treatment
- ➔ **Survival outcome**
  - Treatment
  - Value of longitudinal outcome

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

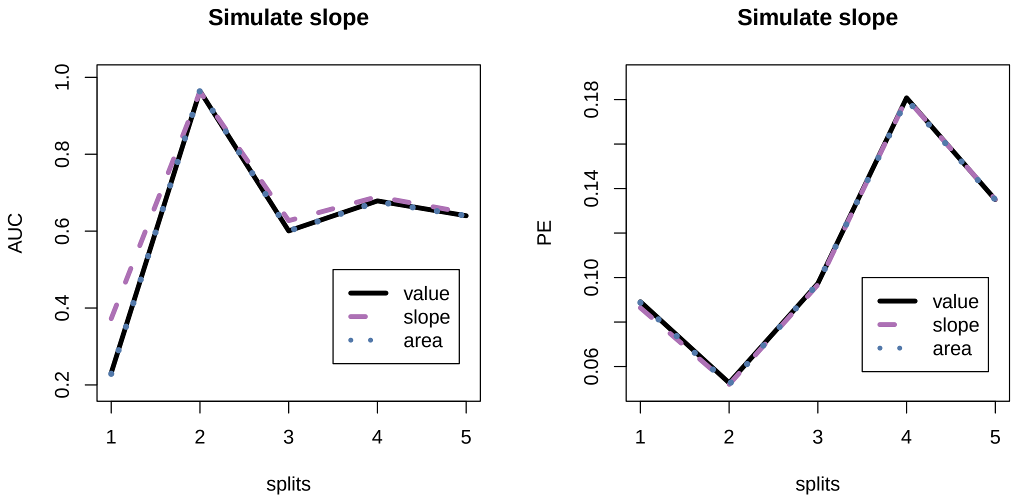
Simulate

- ➔ **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**

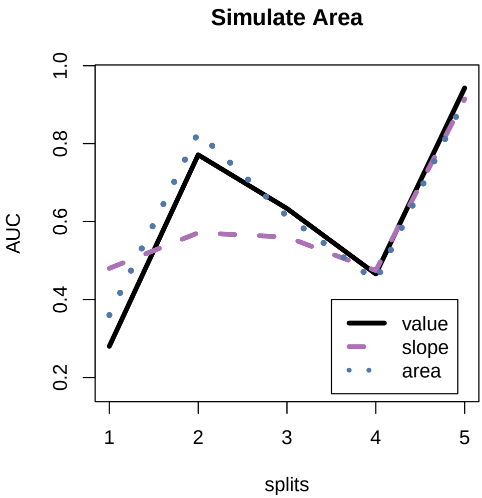
Non linear time
Treatment

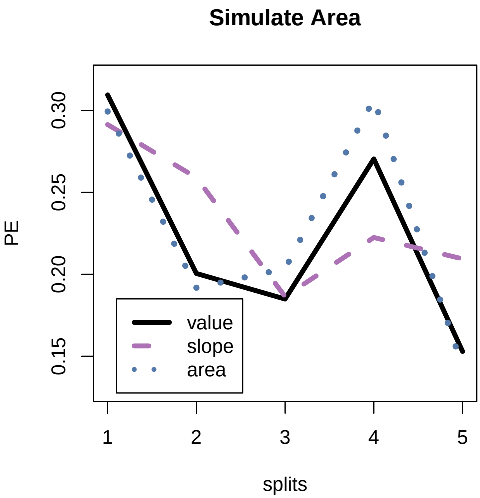
➔ **Survival outcome**

Treatment
Value/Slope/Area of longitudinal outcome

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

Simulations: Results





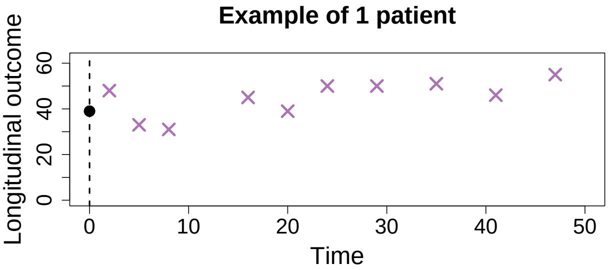
Simulations

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 for the patients left out (1 subset)
- → Calculate PE



Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Value/Slope of outcome 2

➔ **Longitudinal outcome 2**

Linear time

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

Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Value/Slope of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Predict

➔ **Longitudinal outcome 1**

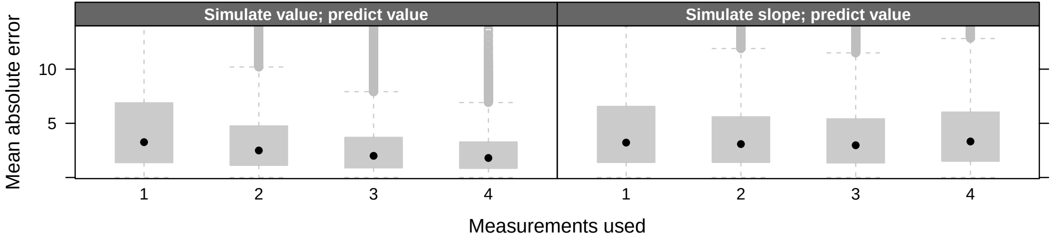
Linear time
Treatment
Value of outcome 2

➔ **Longitudinal outcome 2**

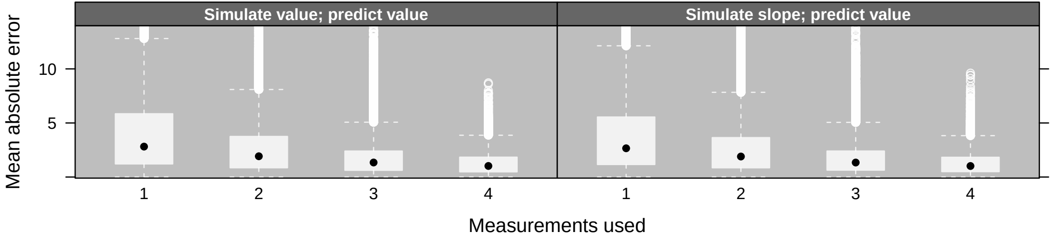
Linear time

Simulations: Results

Outcome 1:



Outcome 2:



Simulations: Scenario

Simulate

➔ **Longitudinal outcome 1**

Linear time
Treatment
Value/Area of outcome 2

➔ **Longitudinal outcome 2**

Linear time

Simulations: Scenario

Simulate

→ **Longitudinal outcome 1**

Linear time
Treatment
Value/Area of outcome 2

→ **Longitudinal outcome 2**

Linear time

Predict

→ **Longitudinal outcome 1**

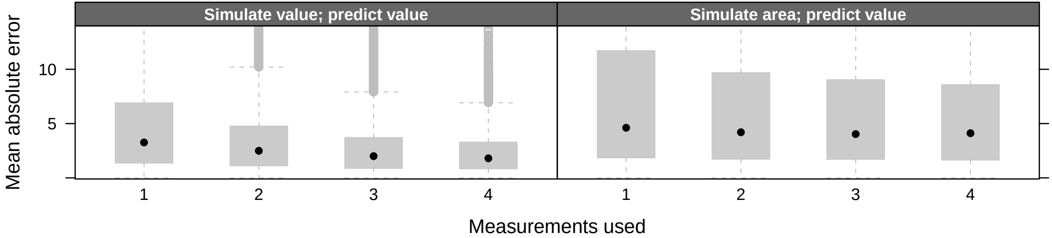
Linear time
Treatment
Value of outcome 2

→ **Longitudinal outcome 2**

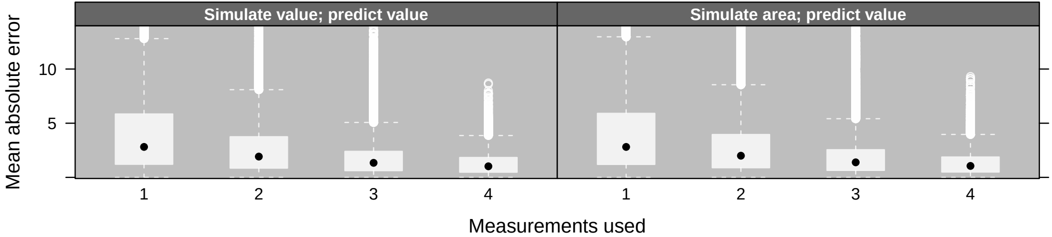
Linear time

Simulations: Results

Outcome 1:



Outcome 2:



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