

Challenges and opportunities in the analysis of clinical data

Statistics seminar, Department of Mathematics, King's College London

Eleni-Rosalina Andrinopoulou, PhD

21 January, 2021

Introduction

A lot of information is available

→ Electronic medical records

A lot of information is available

→ Electronic medical records

Different types of information

→ Baseline characteristics

→ Longitudinal outcomes

→ Time-to-event outcomes

Applications

- Heart valve
- Stroke
- Cystic Fibrosis

Applications

→ Heart valve

- ◇ Aortic gradient
- ◇ Aortic regurgitation
- ◇ Time-to death/reoperation

→ Stroke

→ Cystic Fibrosis

Applications

- Heart valve
- Stroke
 - ◇ Extremity performance
 - ◇ Limb strength
- Cystic Fibrosis

Applications

- Heart valve
- Stroke
- Cystic Fibrosis
 - ◇ FEV_1
 - ◇ BMI
 - ◇ Time-to death/exacerbation

Separate analysis

- Each longitudinal outcome
- Survival outcomes

Separate analysis - Stroke data

- ◇ 412 patients
- ◇ Outcome of interest:

Fugl-Meyer

*van der Vliet, R., Selles, R. W.,
Andrinopoulou, etc (2020). Predicting upper
limb motor impairment recovery after stroke:
a mixture model. Annals of Neurology,
87(3), 383-393.*

Introduction: Extensions

Combined analysis - Cystic Fibrosis data

- ◇ 17,100 patients
- ◇ Outcomes of interest:

FEV_1

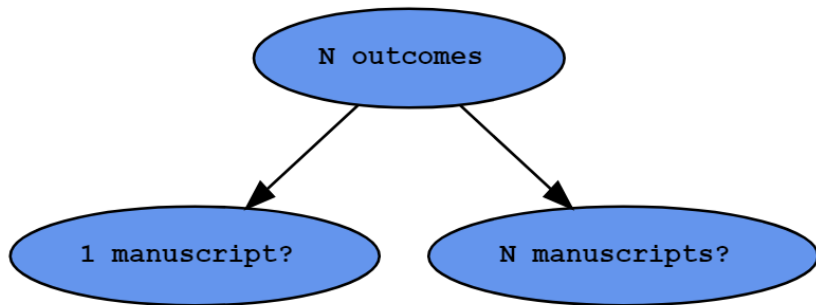
BMI

weight-for-age

height-for-age

time-to first 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.



Combined analysis - Heart valve data

- 296 patients
- Association of **Aortic Gradient** with **time-to-death/reoperation**
 - ◇ Aortic Gradient is measured with error

Combined analysis - Heart valve data

- 296 patients
- Association of **Aortic Gradient** with **time-to-death/reoperation**
 - ◇ Aortic Gradient is measured with error
 - ◇ Different features of Aortic Gradient

Statistical Models

Statistical Models

Let's assume that we have a longitudinal outcome

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

where

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

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

where

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

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

where

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

Let's assume that we have two longitudinal outcomes

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

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

where

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

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

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

Challenge: Quantify the association between y_1 and y_2

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

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

where

◇ α denotes the association

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

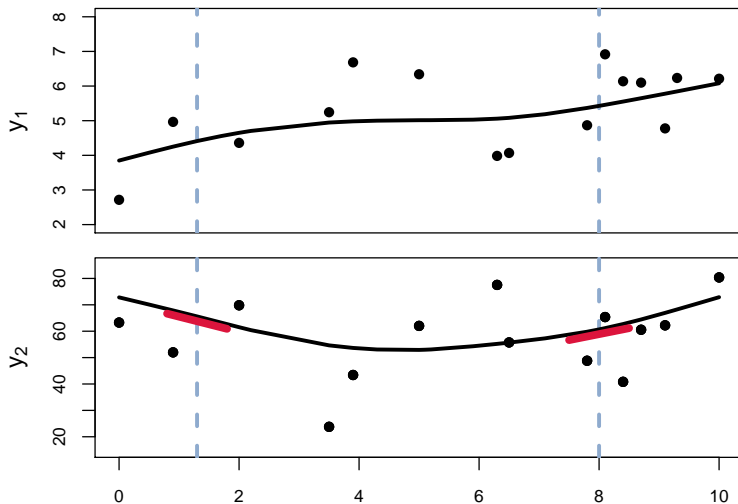
$$y_{2i}(t) = m_{2i}(t) + \epsilon_{2i} = x_{2i}^\top(t)\beta_1 + z_{2i}(t)^\top b_{2i} + \epsilon_{2i}(t)$$

where

◇ α denotes the association

Challenge: Is that our only option?

Statistical Models: Multivariate Mixed Models



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

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

Statistical Models: Multivariate Mixed Models

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

where

◇ α denotes the association

Statistical Models: Multivariate Mixed Models

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

where

◇ α denotes the association

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

Statistical Models: Joint Models

→ Naive joint analysis

- ◇ Cox model using the last observation
- ◇ Time-dependent Cox model

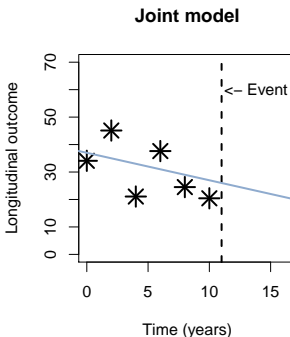
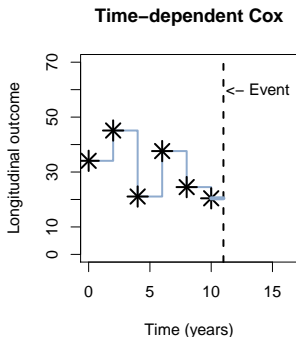
Data is discarded!

Statistical Models: Joint Models

→ Naive joint analysis

- ◇ Cox model using the last observation
- ◇ Time-dependent Cox model

Time-dependent Cox models are suitable only for exogenous covariates!



$$y_i(t) = m_i(t) + \epsilon_i = x_i^\top(t)\beta + z_i^\top(t)b_i + \epsilon_i(t)$$

$$h_i(t) = h_0(t)[\gamma^\top w_i + \alpha m_i(t)]$$

where

◇ α denotes the association

$$y_i(t) = m_i(t) + \epsilon_i = x_i^\top(t)\beta + z_{1i}^\top(t)b_i + \epsilon_i(t)$$

$$h_i(t) = h_0(t) \left[\gamma^\top w_i + \sum_{j=1}^J \alpha_j f_j\{\mathcal{M}_i(t)\} \right]$$

where

- ◇ α_j denotes the association
- ◇ 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.

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

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

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

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

where

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

What about the association between the longitudinal outcomes?

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

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

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

where

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

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

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

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

where

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

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

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

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

where

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

Simulations

Simulations

Multivariate Mixed Models

Simulate

→ Outcome 1

Linear time

Treatment

Value of outcome 2

→ Outcome 2

Linear time

Simulate

→ Outcome 1

Linear time

Treatment

Value of outcome 2

→ Outcome 2

Linear time

Fit

→ Outcome 1

Linear time

Treatment

~~Value of outcome 2~~

→ Outcome 2

Linear time

Simulate

→ Outcome 1

Linear time

Treatment

Value of outcome 2

→ Outcome 2

Linear time

Fit

→ Outcome 1

Linear time

Treatment

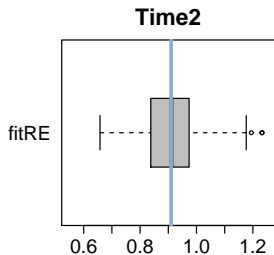
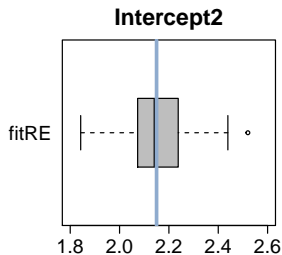
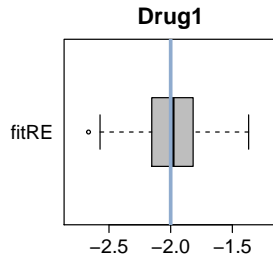
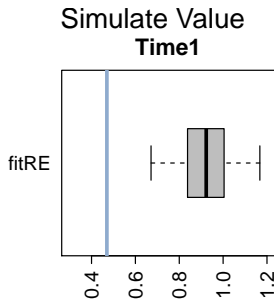
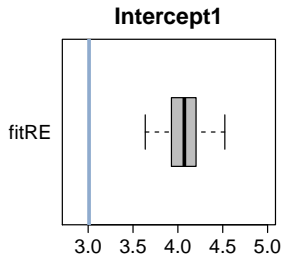
~~Value of outcome 2~~

→ Outcome 2

Linear time

All models were fitted under the Bayesian framework

Simulations: Results



Simulate

→ Outcome 1

Linear time

Treatment

Value of outcome 2

→ Outcome 2

Linear time

Fit

→ Outcome 1

Linear time

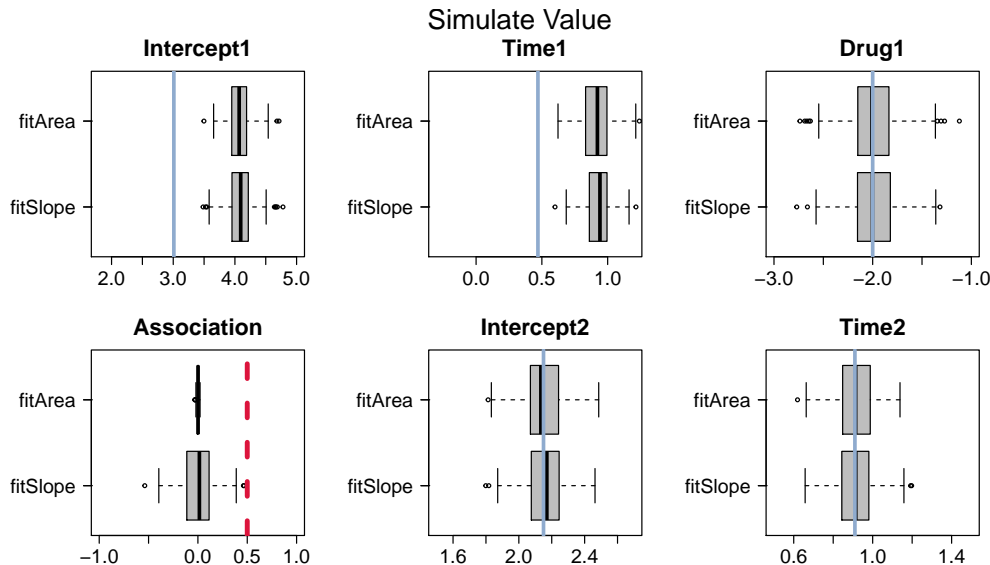
Treatment

Slope/Area of outcome 2

→ Outcome 2

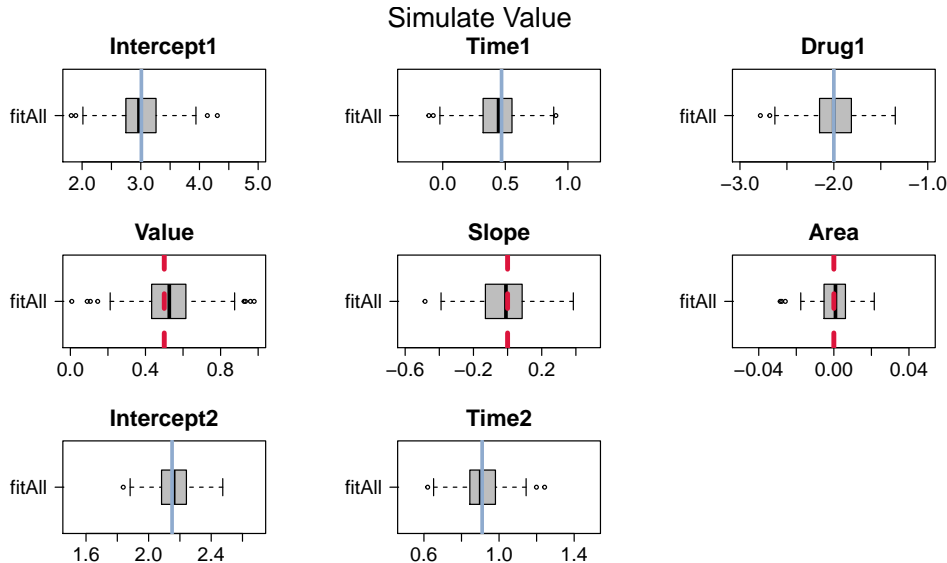
Linear time

Simulations: Results



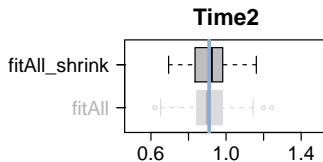
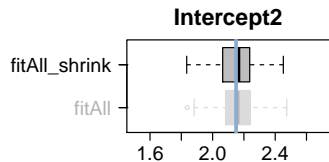
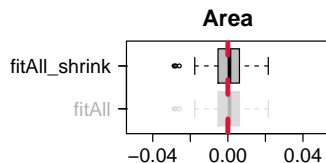
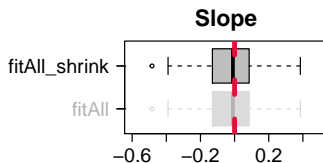
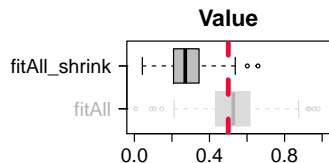
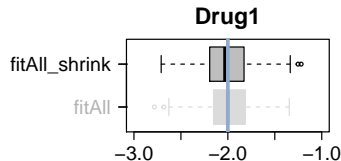
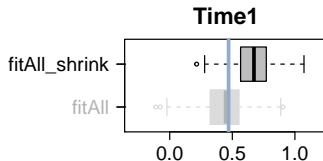
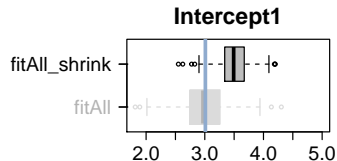
What if we fit all functional forms

Simulations: Results



Simulations: Results

Simulate Value



Let's investigate a more complicated scenario

Simulate

→ Outcome 1

Non linear time

Treatment

Value of outcome 2

Slope of outcome 2

→ Outcome 2

Non linear time

Simulate

→ Outcome 1

Non linear time
Treatment
Value of outcome 2
Slope of outcome 2

→ Outcome 2

Non linear time

Fit

→ Outcome 1

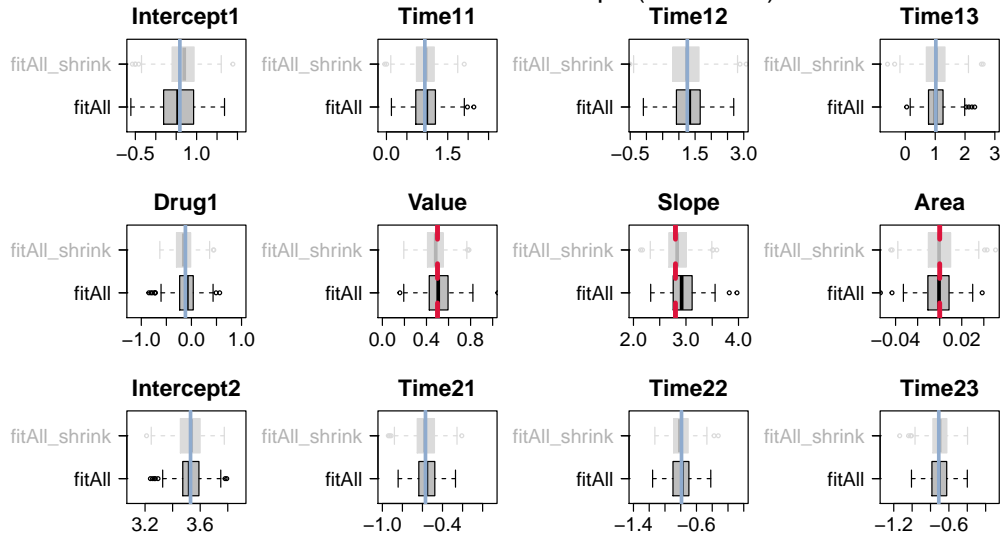
Non linear time
Treatment
Value of outcome 2
Slope of outcome 2
Area of outcome 2

→ Outcome 2

Non linear time

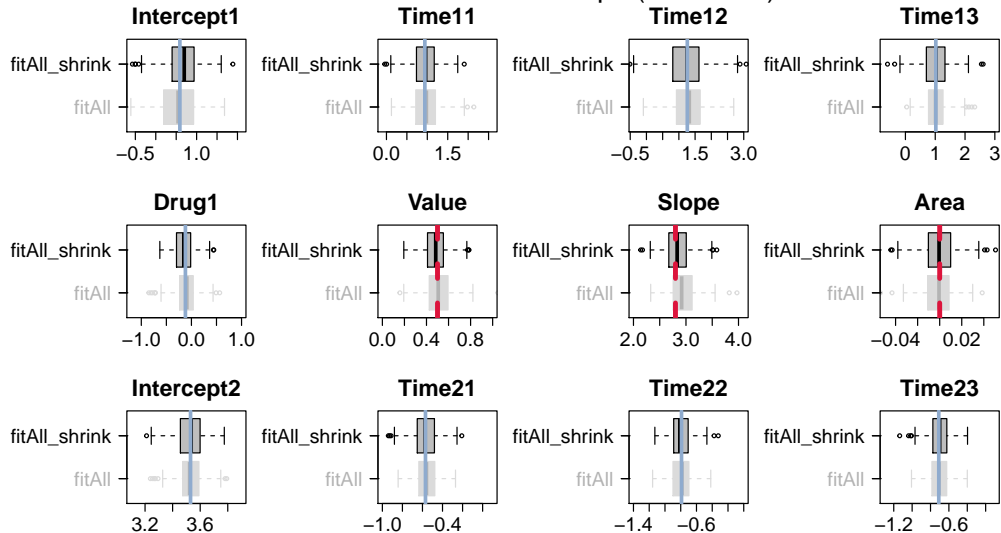
Simulations: Results

Simulate Value and Slope (non linear)



Simulations: Results

Simulate Value and Slope (non linear)



Joint Models

Simulate

→ Longitudinal outcome

Non linear time

Treatment

→ Survival outcome

Treatment

Value of longitudinal
outcome

Simulate

→ Longitudinal outcome

Non linear time
Treatment

→ Survival outcome

Treatment
Value of longitudinal
outcome

Fit

→ Longitudinal outcome

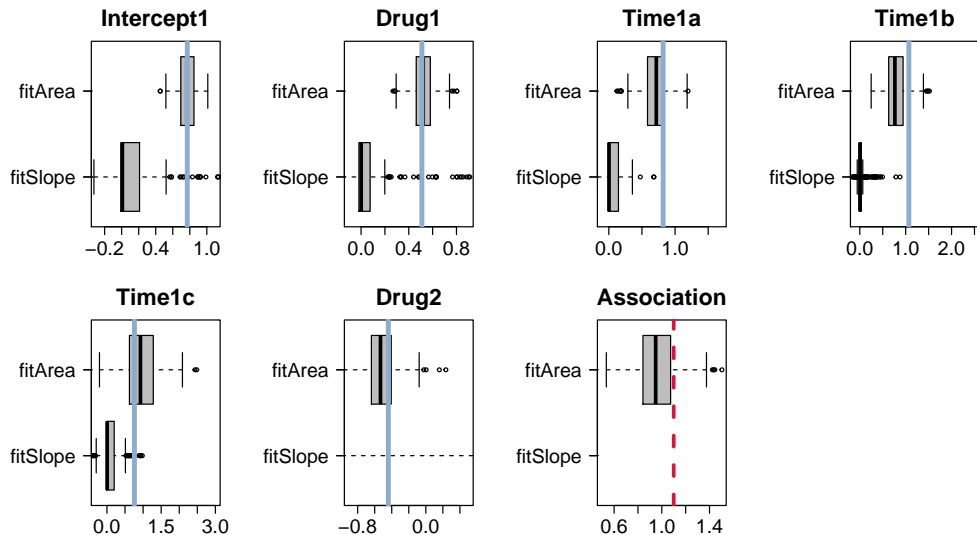
Non linear time
Treatment

→ Survival outcome

Treatment
Slope/Area of longitudinal
outcome

Simulations: Results

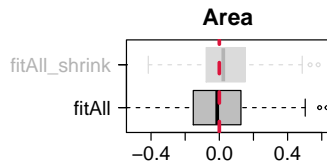
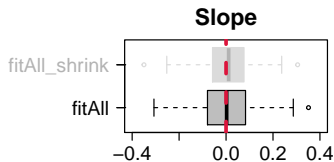
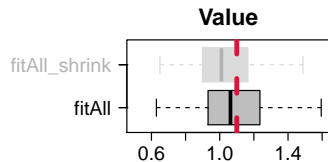
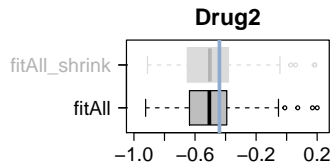
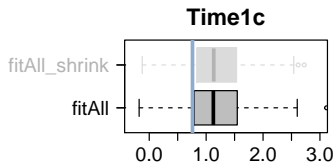
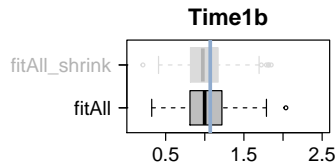
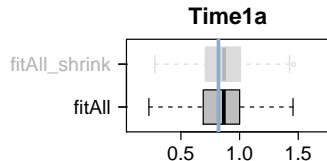
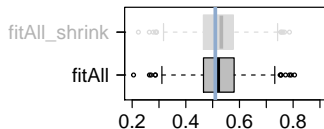
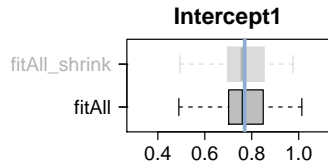
Simulate Value



What if we fit all functional forms

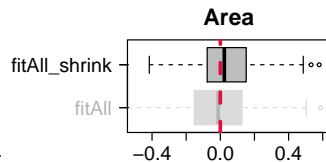
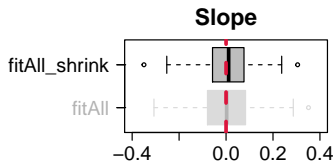
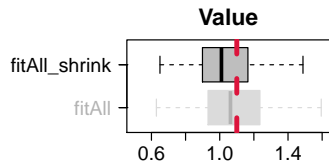
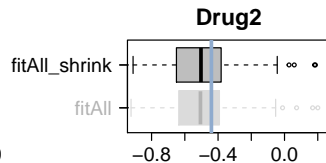
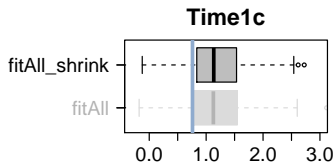
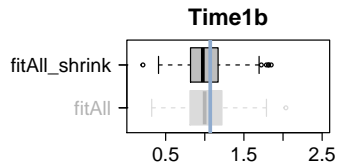
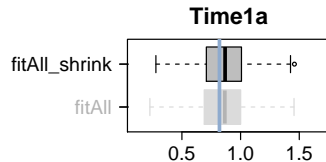
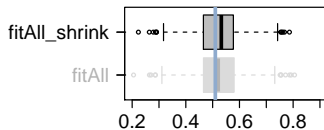
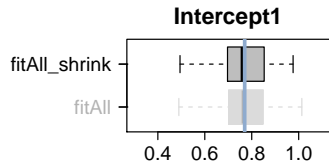
Simulations: Results

Simulate Value Drug1



Simulations: Results

Simulate Value



Let's combine everything

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

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

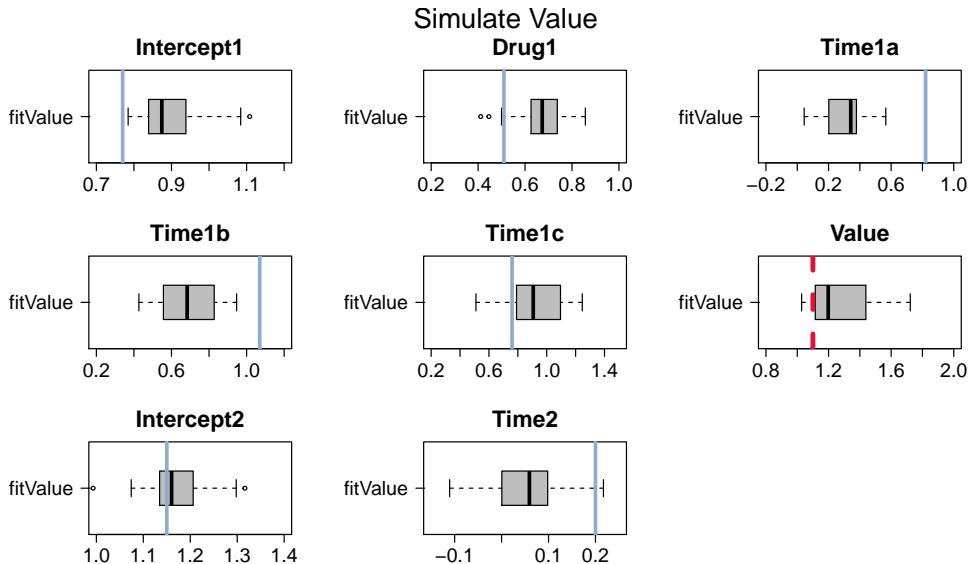
Linear time

→ Survival outcome

Treatment

Value of longitudinal outcome 1

Simulations: Results



Software

→ Joint models

- ◇ JMbayes, JMbayes2, JM
- ◇ joinerR, joinerML
- ◇ frailtypack
- ◇ stan_jm
- ◇ lcmm
- ◇ bamlss
- ◇ JointAI

→ Multivariate mixed models

- ◇ lcmm
- ◇ brms
- ◇ MCMCglmm
- ◇ JointAI

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

- A lot of information is available
- Correlation between outcomes

- Challenges and opportunities
 - ◇ Functional forms

- Future steps
 - ◇ Dynamic predictions

Thank you for your attention!

The slides are available at: <https://www.erandrinopoulou.com>