Feature Selection of Longitudinal Biomarkers in Multivariate Joint Models for Longitudinal and Multi-state processes

G. Papageorgiou[1,2], M. M. Mokhles[2], R. Muslem[2], J. J. M. Takkenberg[2], D. Rizopoulos[1]

1. Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands
2. Department of Cardiothoracic Surgery, Erasmus MC, Rotterdam, The Netherlands

July 15th, 2019

ISCB 2019 Leuven

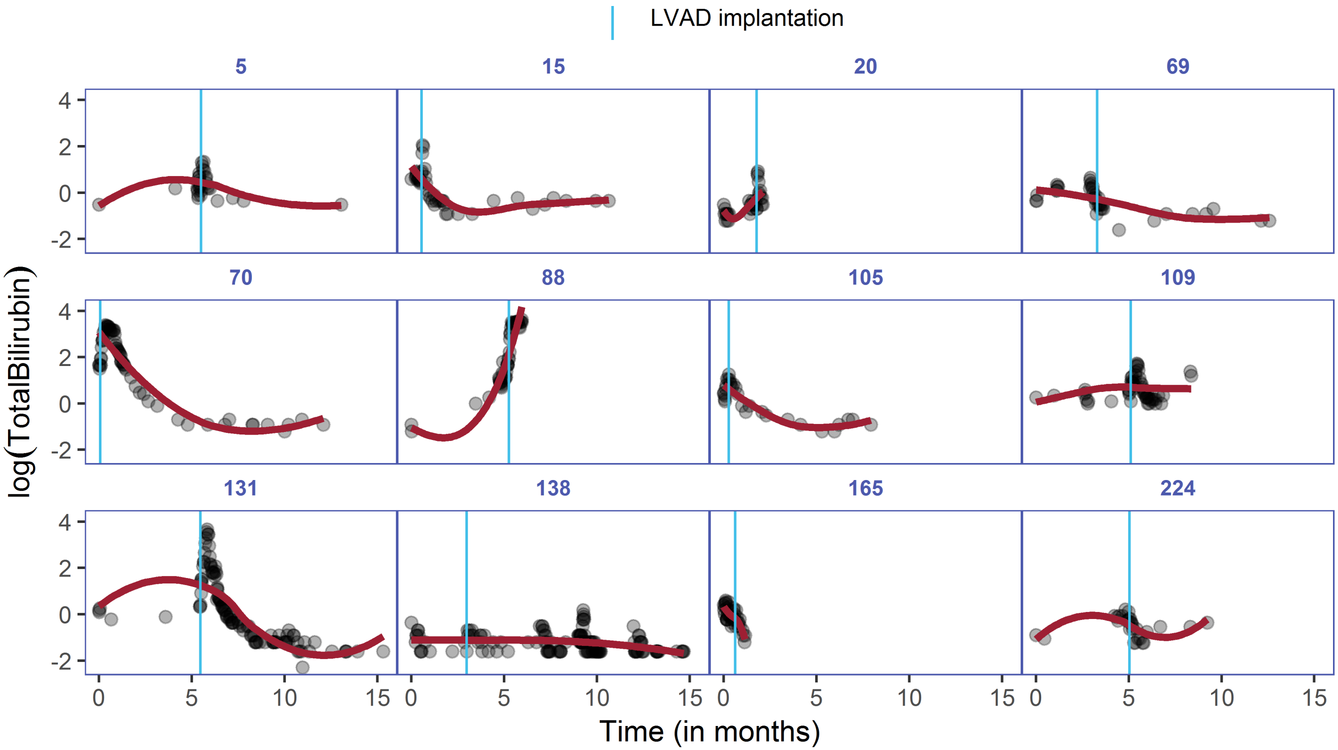
Motivation

- Often in follow-up studies multiple outcomes are collected
- Multiple longitudinal responses:
  - Biomarkers, blood values, etc.
- Times of transitions between states of interest:
  - Transplantation, relapse, clinical complications, death, etc.
  - Possibly successive
- Intermediate events that may alter the course of the disease:
  - Reintervention, etc.

Motivation

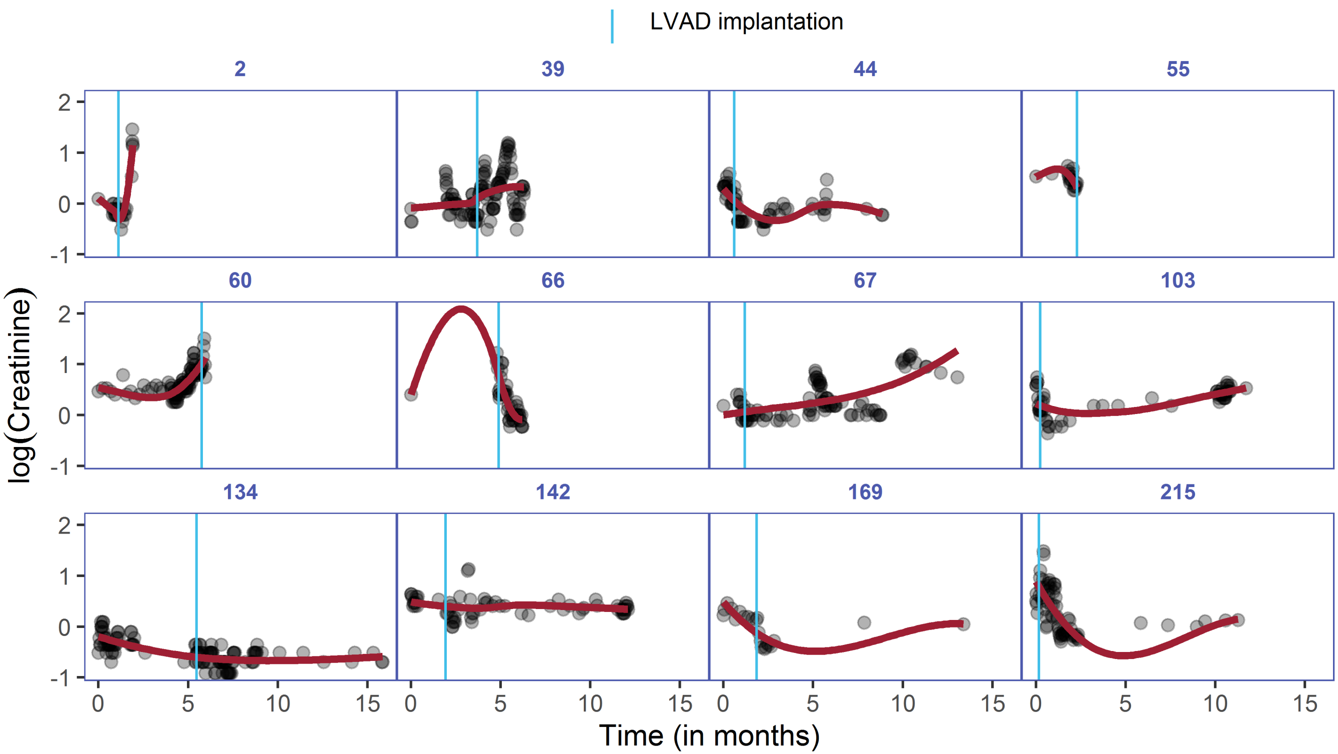
- **232** patients who were followed-up after heart failure
- **Intermediate Event:** Patients received a **L**eft **V**entricular **A**ssist **D**evice during folow-up
- Longitudinal outcomes:
  - Total bilirubin (mg/dl)
  - Creatinine (mg/dl)
- Events of interest:
  - **Complications**: Thrombosis, Embolic Events, Dialysis
  - **Death**

Motivation

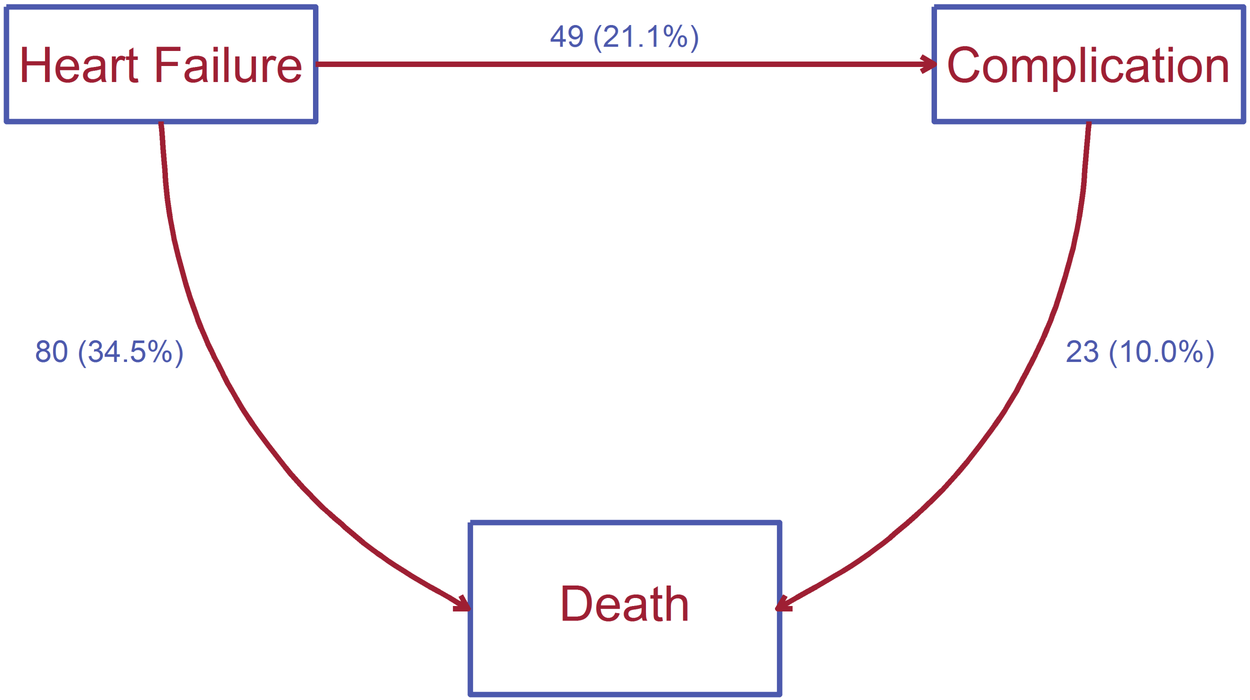


LVAD implantation
5
15
20
69
70
88
105
109
131
138
165
224
log(TotalBilirubin)
4
2
0
-2
0
5
10
15
Time (in months)

Motivation


LVAD implantation
2
39
44
55
60
66
67
103
134
142
169
215
log(Creatinine)
2
1
0
-1
0
5
10
15
Time (in months)

Motivation



Research Goals

- Investigate the association between the longitudinal and multi-state processes
  - Functional form of the association for **each transition**
  - Strength of the association for **each transition**
  - Selection of the association structure for **each transition**
- Investigate the impact of LVAD on the evolution of the markers

Methods

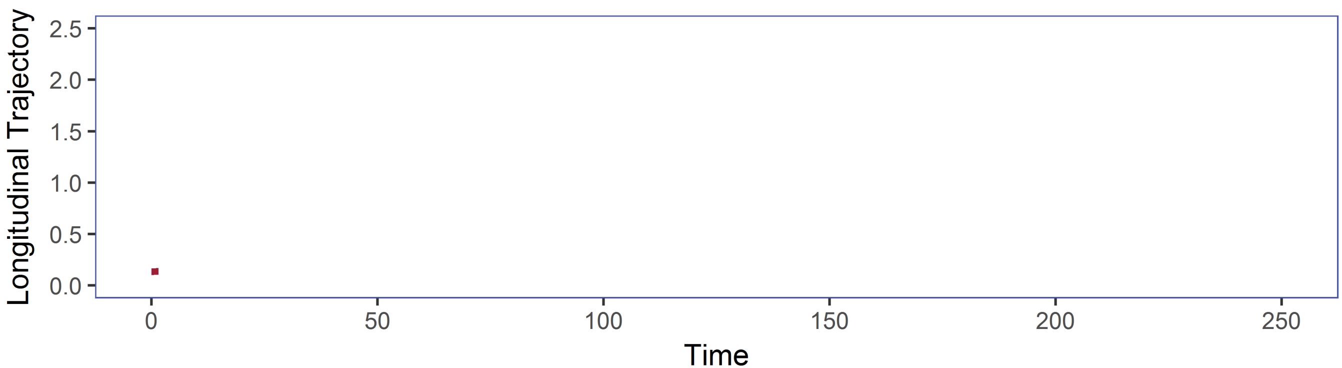
Multivariate generalized linear mixed-effects submodel:

$$g_k\left[E\{y_{ki}(t) \mid b_{ki}\}\right] = \eta_{ki}(t) = \begin{cases} x_{ki}^\top(t)\,\beta_k + z_{ki}^\top(t)\,b_{ki}, & 0 < t < \rho_i \\ x_{ki}^\top(t)\,\beta_k + z_{ki}^\top(t)\,b_{ki} + \tilde{x}_{ki}^\top(t)\,\tilde{\beta}_k + \tilde{z}_{ki}^\top(t)\,\tilde{b}_{ki}, & t \geq \rho_i, \end{cases}$$

- y_{ki}: repeated measurements of the k^{th} outcome for the i^{th} subject, $k = 1, \ldots K$, $i = 1, \ldots, n$
- $b^\top = \left(b_{1i}, \ldots, b_{Ki}, \tilde{b}_{1i}, \ldots, \tilde{b}_{Ki}\right)^\top \sim \mathcal{N}(0, D)$
- ρ_i: time of occurrence of the intermediate event
- $t_{i+} = \max(0, t_{ki} - \rho_i)$: time relative to the occurrence of the intermediate event

$$\eta_{ki}(t) = \begin{cases} x_{ki}^{\top}(t)\,\beta_k + z_{ki}^{\top}(t)\,b_{ki}, & 0 < t < \rho_i \\ x_{ki}^{\top}(t)\,\beta_k + z_{ki}^{\top}(t)\,b_{ki} + \tilde{x}_{ki}^{\top}(t)\,\tilde{\beta}_k + \tilde{z}_{ki}^{\top}(t)\,\tilde{b}_{ki}, & t \geq \rho_i, \end{cases}$$

- $\tilde{x}_{ki}^{\top}(t)\,\tilde{\beta}_k + \tilde{z}_{ki}^{\top}(t)\,\tilde{b}_{ki}$: May be any function of t_{i+}



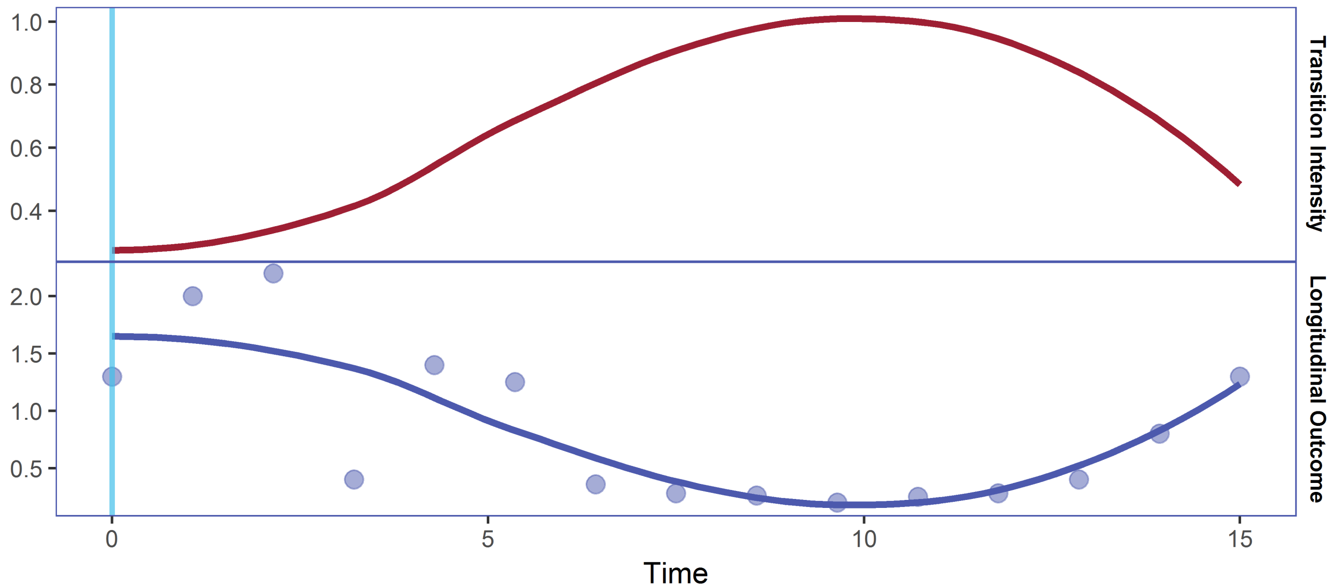
Multi-state submodel:

- $S = \{1, \dots, M\}$: state space
- $T_i = \left(T_{i1}, \dots, T_{im_i}\right)^\top$: vector of observed times for individual i
- $\delta_i = \left(\delta_{i1}, \dots, \delta_{im_i}\right)^\top$: vector of observed transition indicators for individual i

$$\lambda_{hl}^{i}(t) = \begin{cases} \lambda_{hl,0}(t) \exp\left[W_{hl,i}^{S\;\top} \gamma_{hl} + \sum_{k=1}^{K} \sum_{j=1}^{J} f_j \left\{ \eta_{ki}(t), \alpha_{kj} \right\} \right], & 0 < t < \rho_i, \\ \lambda_{hl,0}(t) \exp\left[W_{hl,i}^{S\;\top} \gamma_{hl} + \sum_{k=1}^{K} \sum_{j=1}^{J} f_j \left\{ \eta_{ki}(t), \alpha_{kj} \right\} \right], & t \geq \rho_i. \end{cases}$$

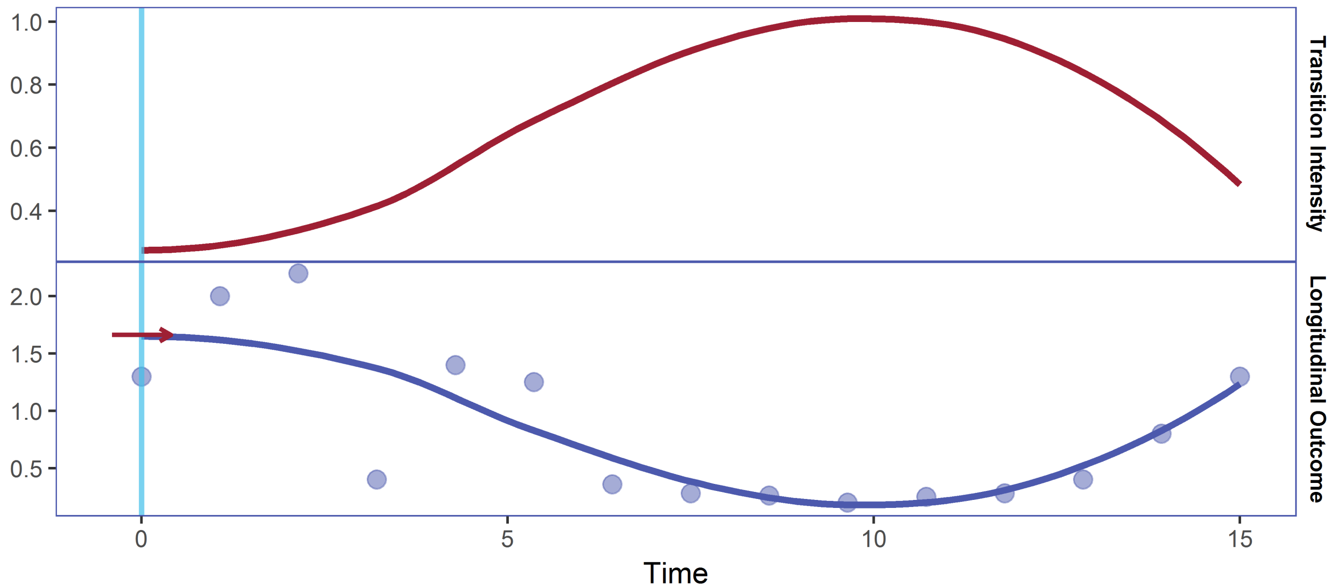
Association Structures

Current value: $f_j\left\{\eta_{ki}\left(t\right),\alpha_{kj}\right\}=\eta_{ki}\left(t\right)\cdot\alpha_{kj}$



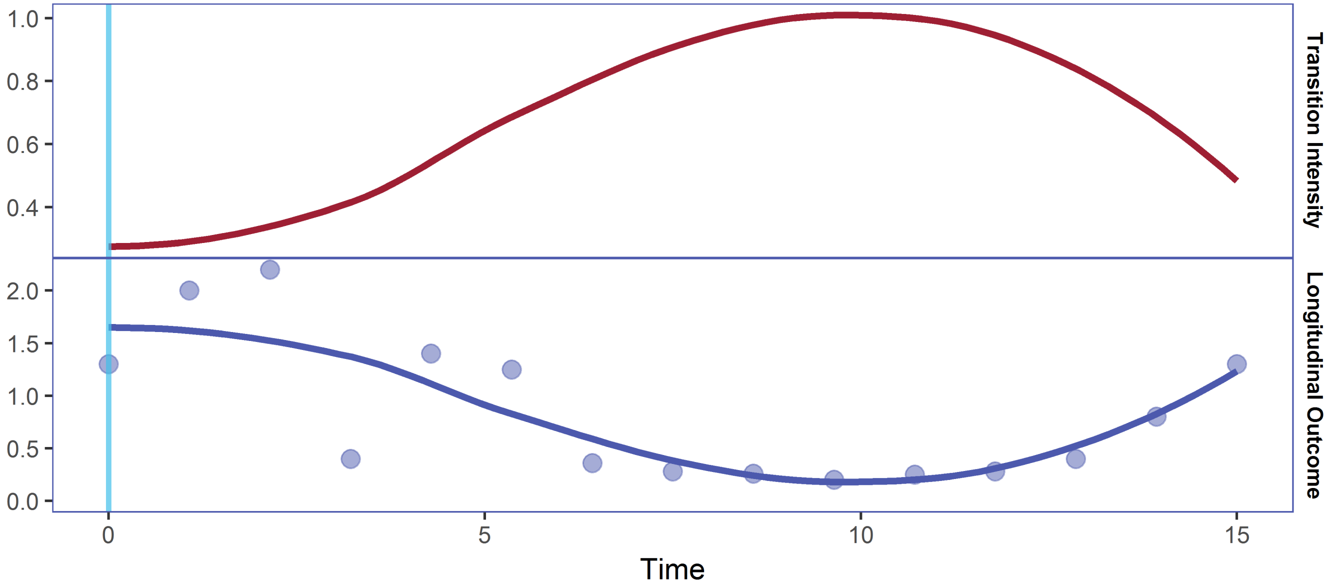
Association Structures

Current slope: $f_j\left\{\eta_{ki}(t), \alpha_{kj}\right\} = \frac{d}{dt}\eta_{ki}(t) \cdot \alpha_{kj}$



Association Structures

Cumulative effect: $f_j\left\{\eta_{ki}(t), \alpha_{kj}\right\} = \int_0^t \eta_{ki}(s)\, ds \cdot \alpha_{kj}$



Feature Selection

- Which features of the longitudinal outcomes are associated with transition intensities?

$$\#\text{Biomarkers} \times \#\text{Features} \times \#\text{Transitions}$$

- High dimensional parameter space
- (Potentially) high correlation among features from the same biomarker
- Feature selection $\rightarrow$ difficult

Bayesian Shrinkage

- Local priors:
  - $\alpha_{jk} \mid \tau_{jk}^2 \sim \mathcal{N}\left(0, \tau_{jk}^2\right)$
  - τ_{jk}^2: Local shrinkage
- Global-local priors:
  - $\alpha_{jk} \mid \tau_{jk}^2, \lambda^2 \sim \mathcal{N}\left(0, \tau_{jk}^2 \lambda^2\right)$
  - τ_{jk}^2: Local shrinkage
  - λ^2: Global shrinkage
- E.R. Andrinopoulou and D. Rizopoulos (2016) investigated the performance of local shrinkage priors

Bayesian Shrinkage

- Global-local horseshoe prior
  - Double inverse-gamma prior leads to $C^{+}(0,1)$

$$
\begin{aligned}
\alpha_{jk} \mid \tau_{jk}^2, \lambda^2 &\sim \mathcal{N}\left(0, \tau_{jk}^2 \lambda^2\right) \\
\tau_{jk}^2 \mid \nu_{jk}^2 &\sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\nu_{jk}}\right) \\
\lambda^2 \mid \xi &\sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\xi}\right) \\
\nu_{1k}, \dots, \nu_{JK}, \xi &\sim \mathcal{IG}\left(\frac{1}{2}, 1\right)
\end{aligned}
$$

- **Properties:**
  - Strong spike that leads to severe shrinkage near 0
  - Narrow tails that allow strong signals to remain strong

Bayesian Shrinkage

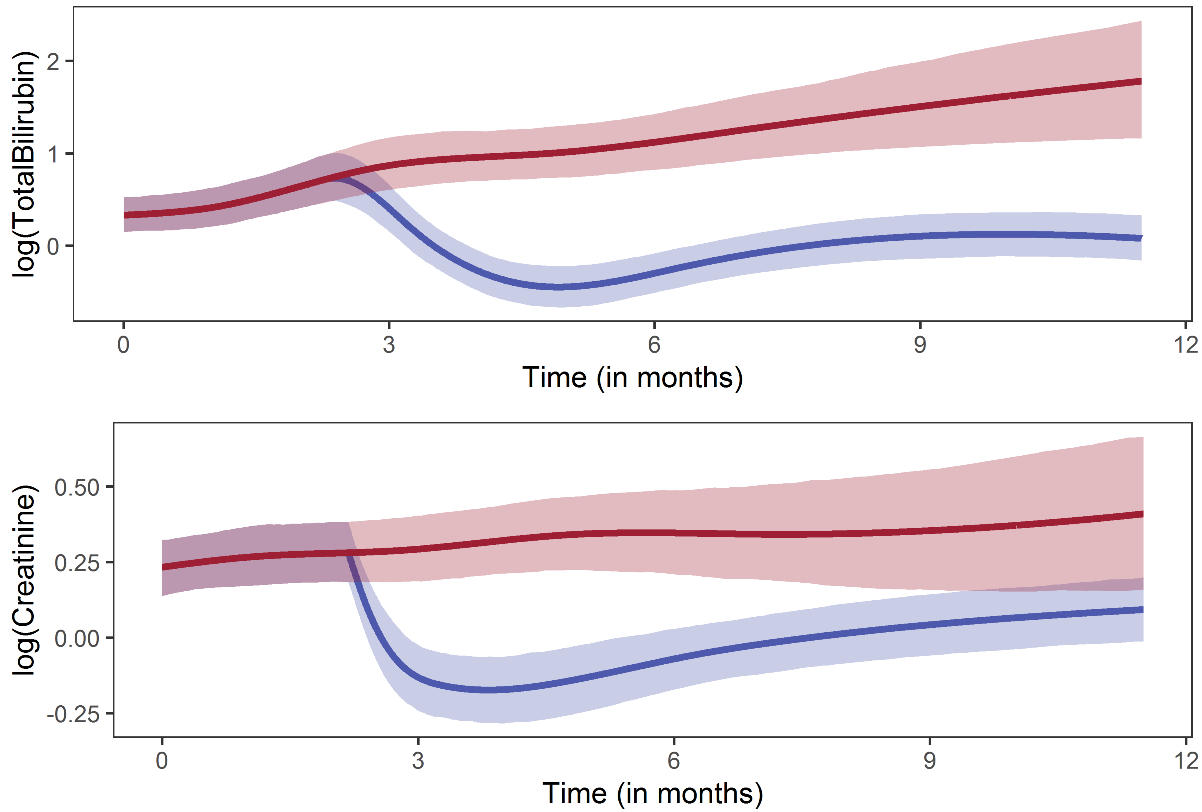
- Global-local ridge prior

$$
\begin{aligned}
\alpha_{jk} \mid \tau_{jk}^2, \lambda^2 &\sim \mathcal{N}\left(0, \tau_{jk}^2 \lambda^2\right) \\
\tau_{jk}^2 &\sim \mathcal{IG}\left(\tfrac{1}{2}, 1\right) \\
\lambda^2 &\sim \mathcal{IG}\left(\tfrac{1}{2}, 1\right)
\end{aligned}
$$

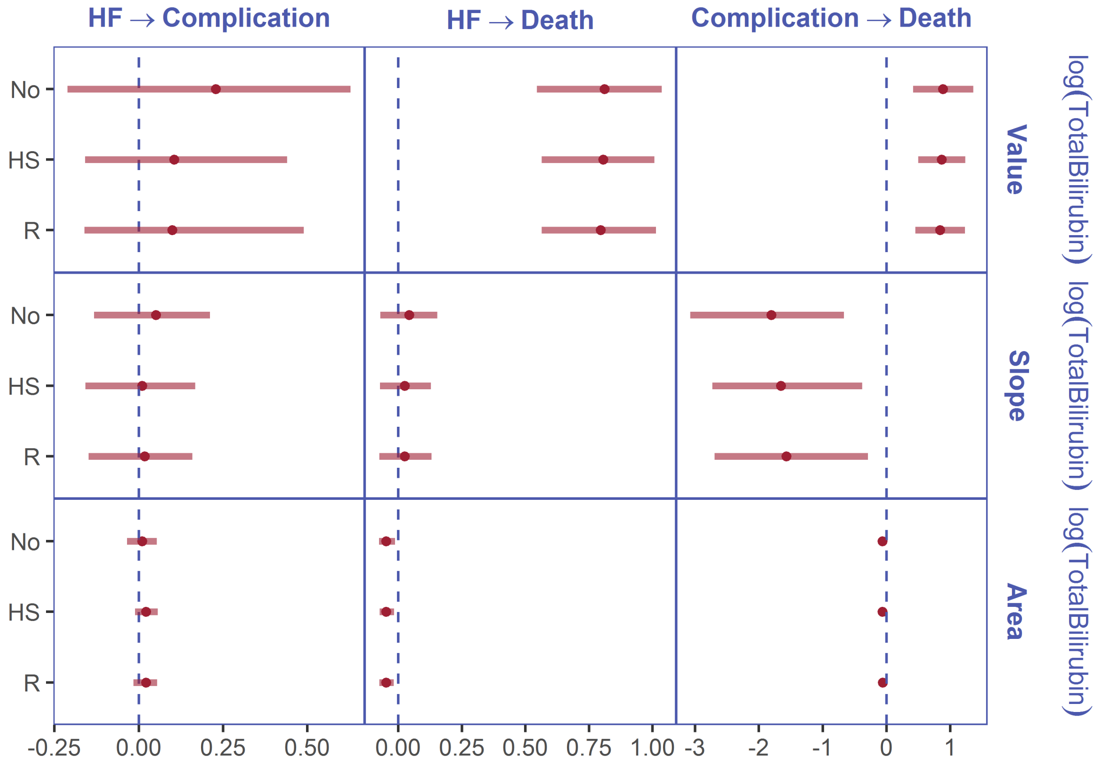
- **Properties:**
  - Less shrinkage near 0
  - Heavier tails than horseshoe

Analysis

- **Longitudinal submodels:**
  - **Fixed-effects**: natural cubic splines for time and time relative to **LVAD** implantation, adjusted for BMI, age, sex and etiology
  - **Random-effects**: natural cubic splines for time and time relative to **LVAD** implantation
- **Multi-state submodel:**
  - **State-specific covariates**: BMI, age, sex and etiology
  - **Markers' features**: value, slope and cumulative effect association with **each transition**

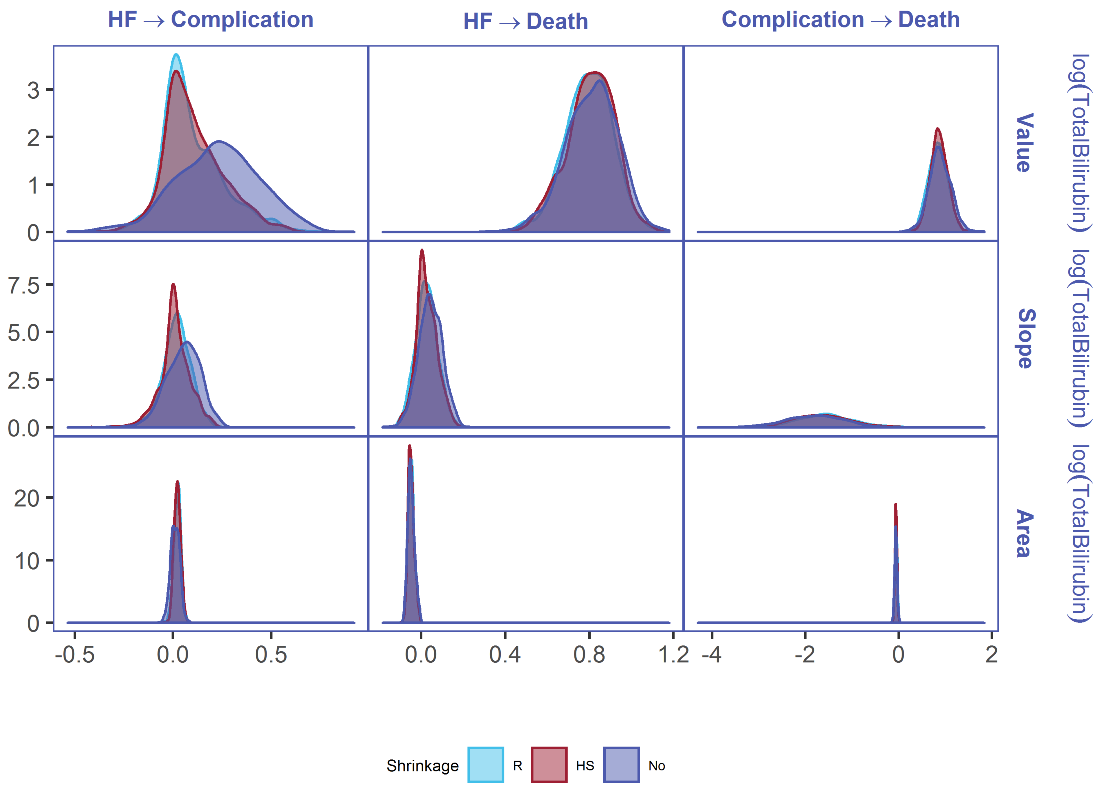

log(TotalBilirubin)
2
1
0
0
3
6
9
12
Time (in months)
log(Creatinine)
0.50
0.25
0.00
-0.25
0
3
6
9
12
Time (in months)

Results

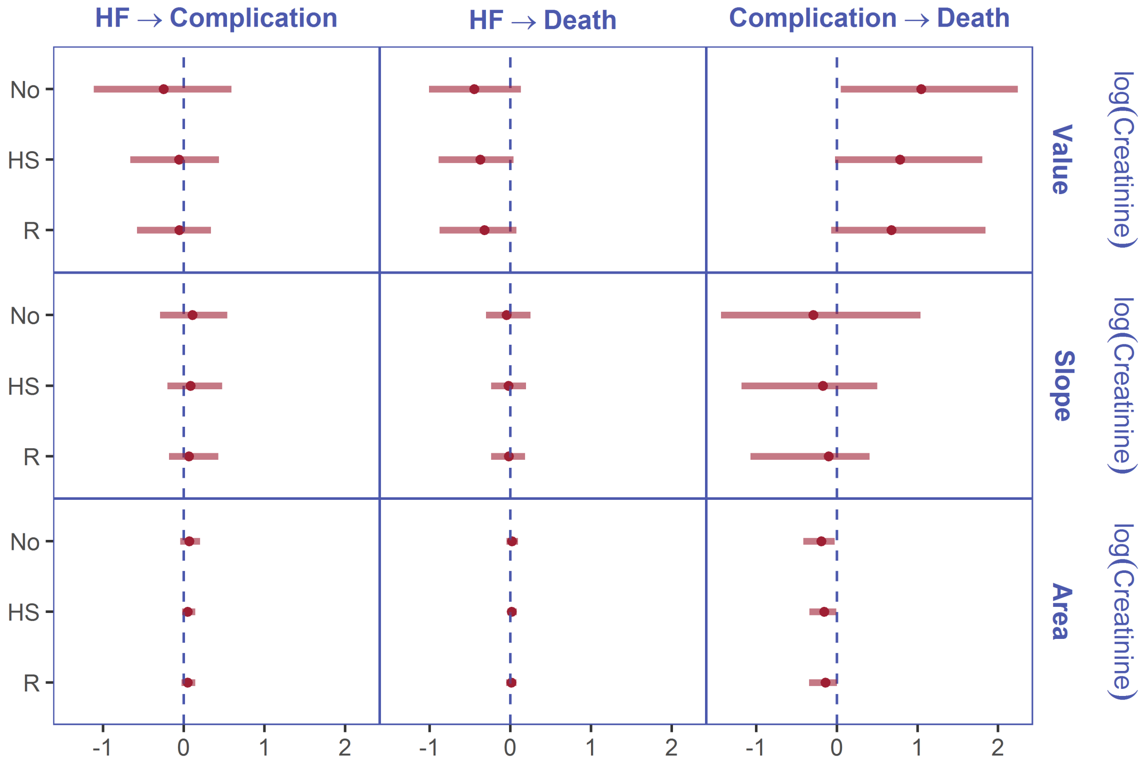


HF → Complication
HF → Death
Complication → Death
No
HS
R
No
HS
R
No
HS
R
-0.25 0.00 0.25 0.50
0.00 0.25 0.50 0.75 1.00
-3 -2 -1 0 1
Value
Slope
Area
log(TotalBilirubin)
log(TotalBilirubin)
log(TotalBilirubin)

Results

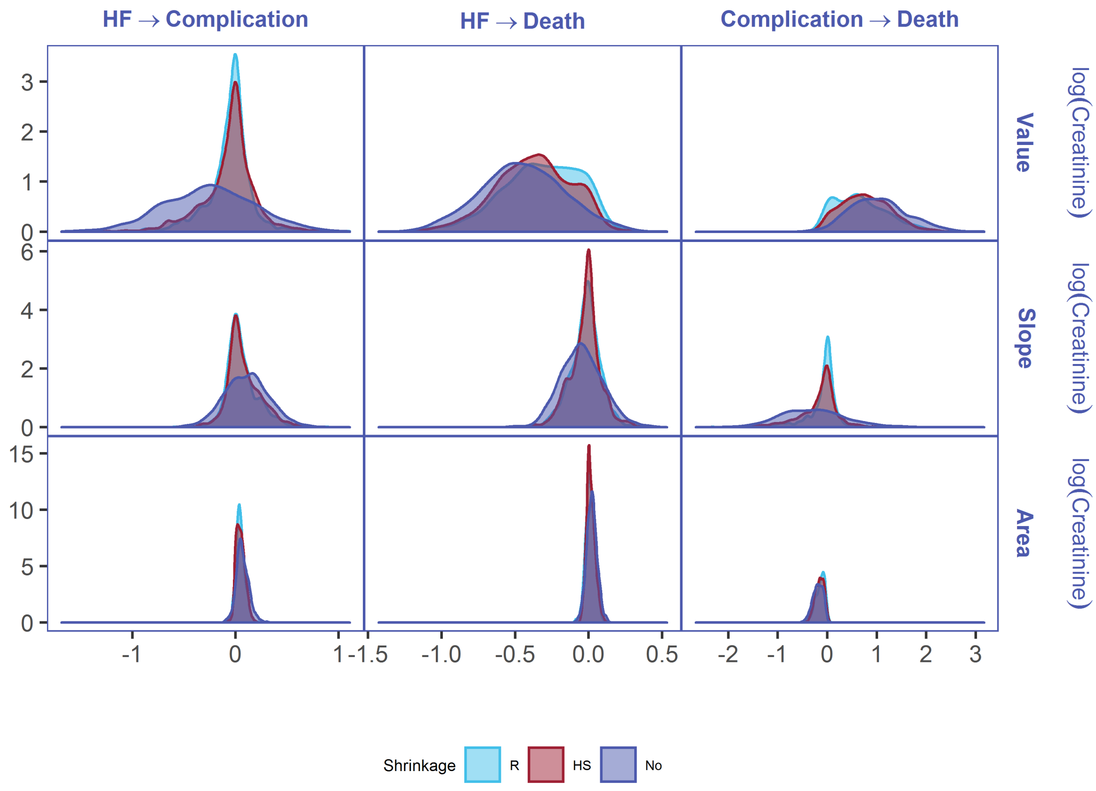


HF → Complication
HF → Death
Complication → Death
Value
Slope
Area
log(TotalBilirubin)
log(TotalBilirubin)
log(TotalBilirubin)
3
2
1
0
7.5
5.0
2.5
0.0
20
10
0
-0.5
0.0
0.5
0.0
0.4
0.8
1.2
-4
-2
0
2
Shrinkage
R
HS
No



HF → Complication
HF → Death
Complication → Death
No
HS
R
No
HS
R
No
HS
R
Value
Slope
Area
log(Creatinine)
log(Creatinine)
log(Creatinine)
-1
0
1
2
-1
0
1
2
-1
0
1
2

Results



HF → Complication
HF → Death
Complication → Death
Value
Slope
Area
log(Creatinine)
log(Creatinine)
log(Creatinine)
3
2
1
0
6
4
2
0
15
10
5
0
-1
0
1
-1.5
-1.0
-0.5
0.0
0.5
-2
-1
0
1
2
3
Shrinkage
R
HS
No

Software

- Software implementation:
  - Already available in **JMbayes**

https://github.com/drizopoulos/JMbayes

Thank you

@gr_papageorgiou

https://github.com/gpapageorgiou

g.papageorgiou@erasmusmc.nl

gregpapageorgiou.com