Joint modeling of biomarkers dynamics and survival with competing risks to predict the prognosis of patients hospitalized with severe infectious diseases

Alexandra Lavalley-Morelle[*1], France Mentré[1,2], Emmanuelle Comets[1,3], Jimmy Mullaert[1,2]

*alexandra,lavalley-morelle@inserm.fr
1 Université Paris Cité, UMR 1137 IAME, INSERM, F-75018 Paris, France
2 Department of Epidemiology, Biostatistics and Clinical Research, AP-HP, Bichat-Claude Bernard University Hospital, F-75018 Paris, France
3 Université de Rennes, Inserm, EHESP, Irset - UMRS 1085, F-35000 Rennes, France

29/06/2023

Provide every patient with a self and adapted medical treatment

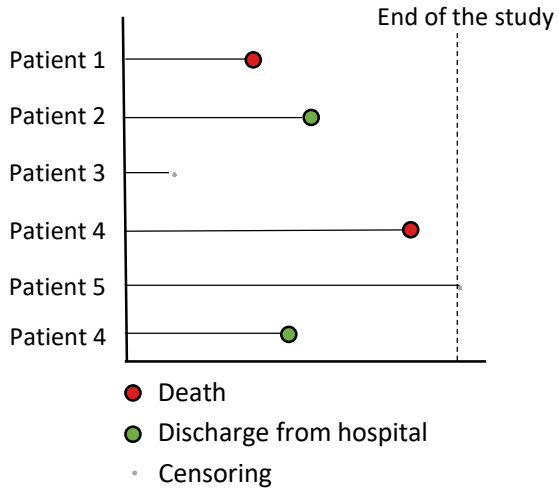
Personalized medicine

1- Rizopoulos. *Biometrics*, 2011

Provide every patient with a self and adapted medical treatment

Personalized medicine

Predict the death of hospitalized patients for severe infectious diseases



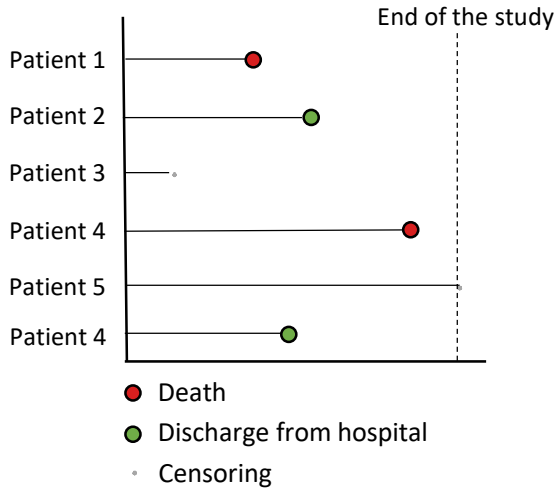
Competing risks

1- Rizopoulos. *Biometrics*, 2011

Provide every patient with a self and adapted medical treatment

Personalized medicine

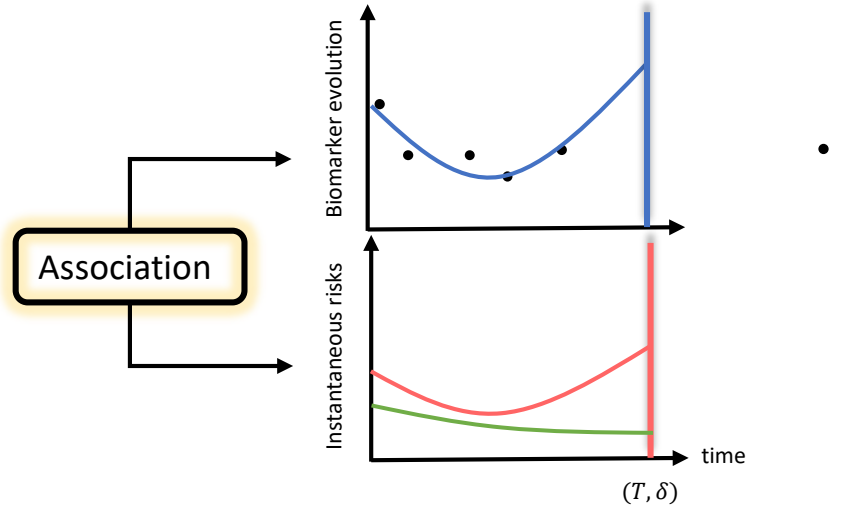
Predict the death of hospitalized patients for severe infectious diseases



Competing risks

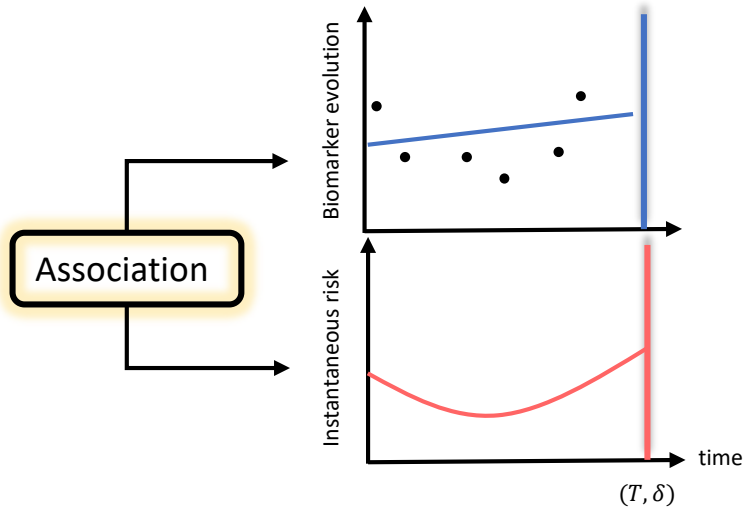
Hospitals equipped with laboratory information systems that routinely gather results of biological analyses

Consecutive biological observations can be used in a joint model to provide individual dynamic predictions[1] of patient prognosis

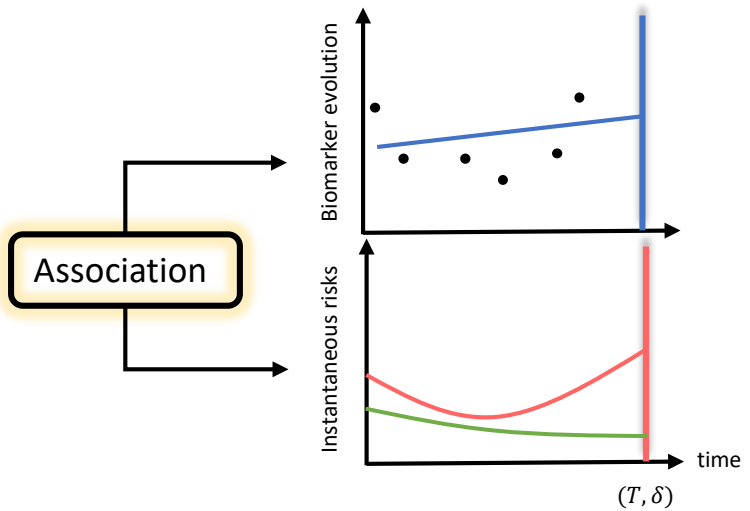


1- Rizopoulos. *Biometrics*, 2011

Joint modeling of **LMEM** and **single event**



Joint modeling of **LMEM** and **competing risks**

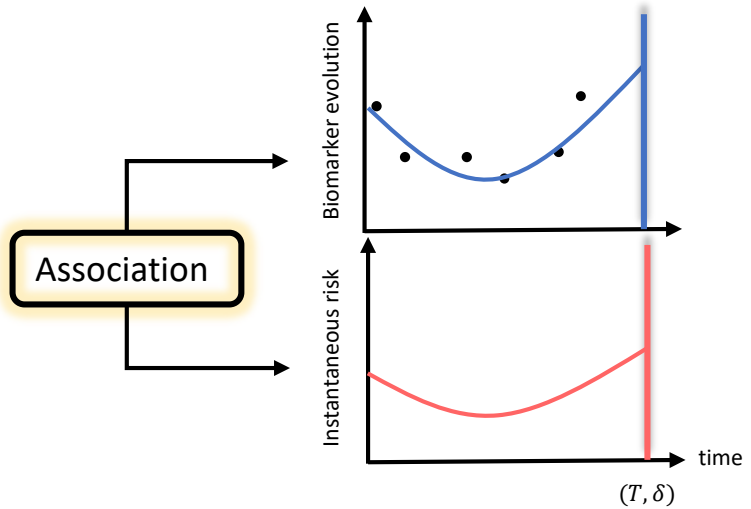


Widely developed in literature (single event[1,2,3],competing risks[4,5,6])
Estimation available in various software: (R, SAS, Monolix, NONMEM,…)

LMEM = linear mixed-effects model

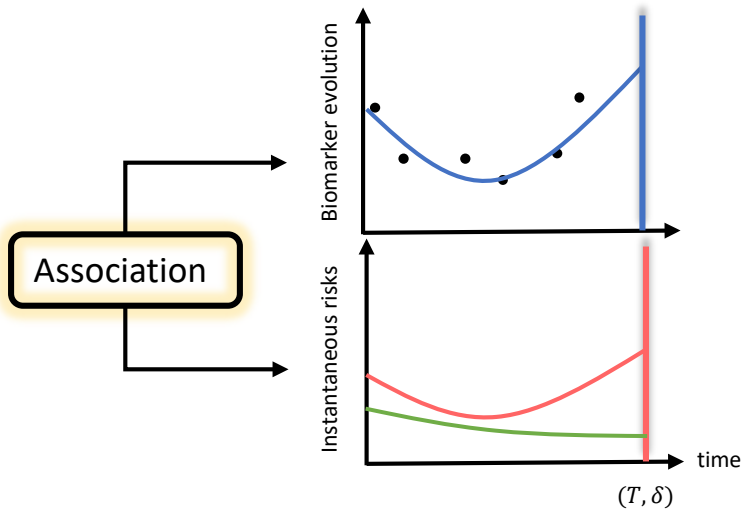
1- Rizopoulos. *Biometrics*, 2011
2- Angeli et al. *The AAPS Journal,* 2016
3- Elashoff et al. *Biometrics*, 2008
4- Deslandes and Chevret. *BMC Medical Research and Methodology*, 2010
5- Musoro et al. *Statistica Neerlandica*, 2018
6- Alvares and Rubio. *Statistics in Medicine*, 2021

Joint modeling of **NLMEM** and **single event**



Widely developed in literature[7,8,9]
Some softwares available: Monolix, NONMEM

Joint modeling of **NLMEM** and **competing risks**

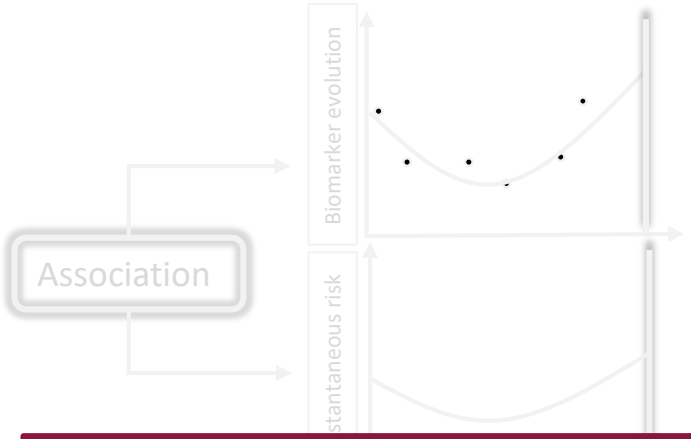


Very few developed in literature (1 published work[10])
Software used: NONMEM

NLMEM = nonlinear mixed-effects model

7- Desmée et al. *The AAPS Journal*, 2015
8- Tardivon et al. *Clinical Pharmacology & Therapeutics,* 2019
9- Kerioui et al. *Statistics in Medicine*, 2020
10- Krishnan et al. *CPT:Pharmacometrics & Systems Pharmacology,* 2021

Joint modeling of **NLMEM** and **single event**

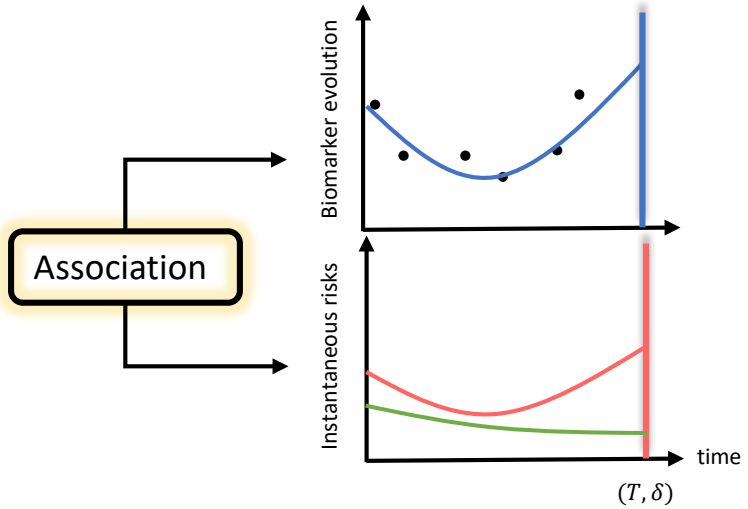


1st PhD work:

- Show that estimation well performs in **Monolix** software
- Application on sepsis patients

- Lavalley-Morelle et al. *CPT:Pharmacometrics & Systems Pharmacology*, 2022
- 29th PAGE conference (2021), "Individual dynamic predictions and their evaluation in joint analysis with nonlinear longitudinal model and parametric competing risks model: application on sepsis patients"

Joint modeling of **NLMEM** and **competing risks**



Very few developed in literature (1 published work[10])
Software used: NONMEM

7- Desmée et al. *The AAPS Journal*, 2015
8- Tardivon et al. *Clinical Pharmacology & Therapeutics,* 2019
9- Kerioui et al. *Statistics in Medicine*, 2020

10- Krishnan et al. *CPT:Pharmacometrics & Systems Pharmacology,* 2021

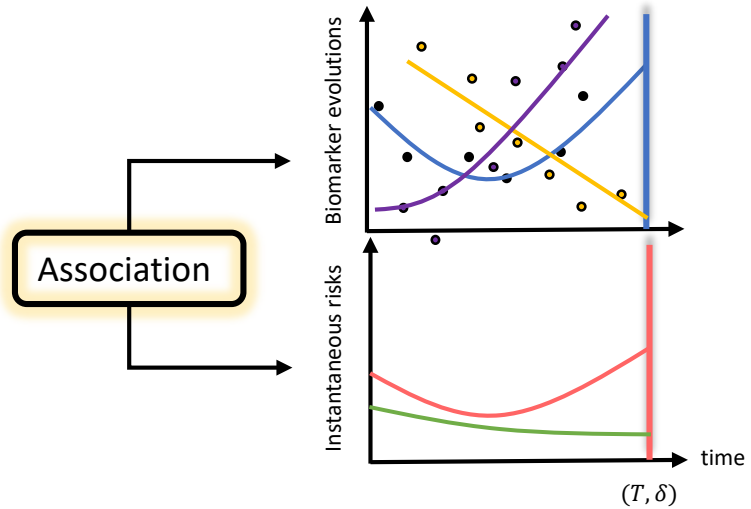
Comes with computational and identifiability issues due to the high number of random effects[11]

Published models mostly limited to longitudinal models with at most two biomarkers[12,13]

Certainly improve prediction accuracy

Shifting the modeling from one to several biomarkers

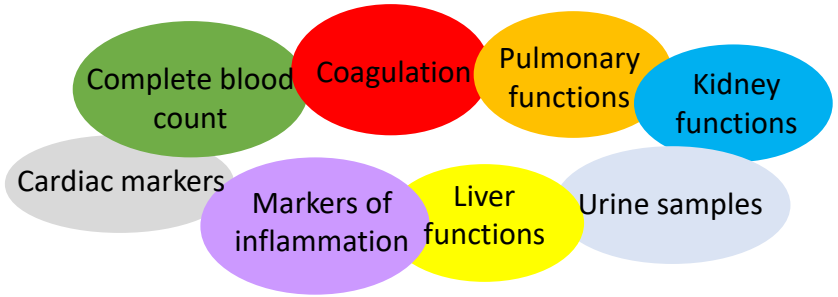
Need of methods for variable selection in this context



11- Shen et al. *Statistics in Medicine*, 2021
12- Long and Mills. *BMC Medical Research Methodology*, 2018
13- Rajeswaran et al. *Statistics in Biosciences*,2018

RisCoV database

- 327 patients hospitalized during the first wave of the COVID-19 pandemic
- 59 biomarkers (classified in 8 categories) variably followed until death or discharge
- Prognosis (death/discharge from hospital)



How to select biomarkers most associated with prognosis ?

Objectives of the work

COVID-19 case study

- Develop a multivariate joint model and a strategy to select a subset of biomarkers to predict the death of patients hospitalized for SARS-CoV-2 infection

Methodological assessment

- Evaluate the SAEM algorithm implemented in Monolix for multivariate joint models under competing risks
- Assess the validity of the proposed selection strategy

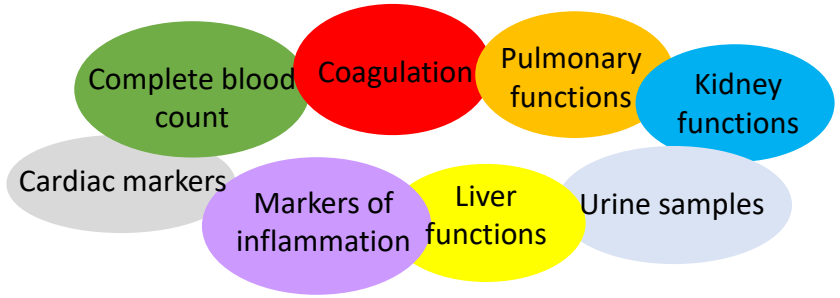
saemix extension

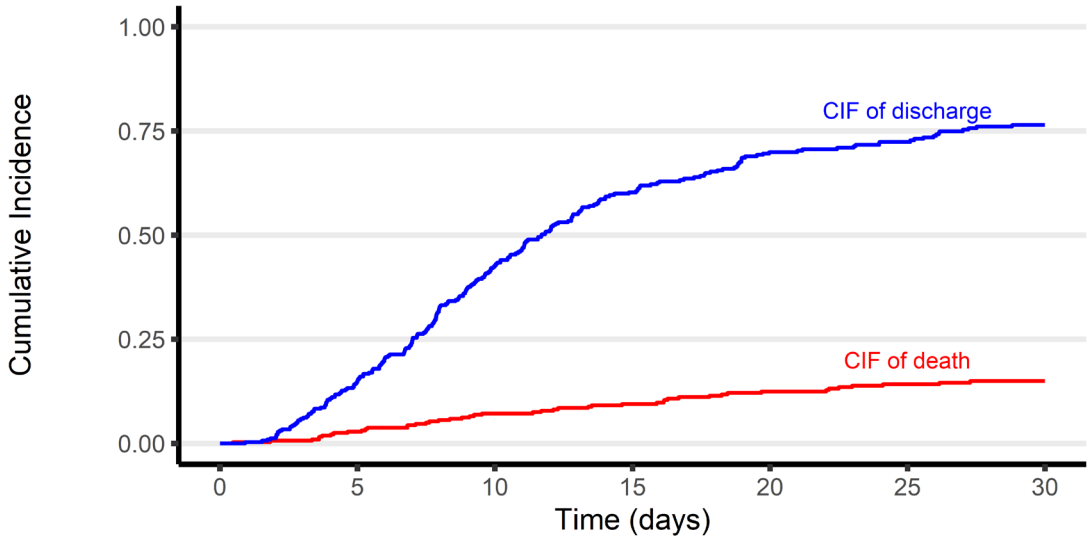
- Extend the R *saemix*[14] package to the case of multi-response and joint models

14- Comets et al. *Journal of Statistical Software*, 2017

RisCoV database

- 327 patients hospitalized in France during the first wave of the COVID-19 pandemic (January to July 2020)
- 59 biomarkers (classified in 8 categories) followed until death or discharge
- 4C Score[15] available at admission
- At D30: 14% deaths
  73% discharges





Cumulative incidence functions for both events

Components of the 4C-Score

Age, years	**< 50**	0				
	50 – 59	+2				
	60 – 69	+4				
	70 – 79	+6				
	≥ 80	+7				
Sex at birth	**Female**	0	Male	+1		
Number of commorbidities	**0**	0	**1**	+1	**≥ 2**	+2
Respiratory rate, breaths/min	**< 20**	0	**20 – 29**	+1	**≥ 30**	+2
Peripheral oxygen saturation on room air	**≥ 92%**	0	**< 92%**	+2		
Glasgow Coma Scale	**15**	0	**< 15**	+2		
Urea (mmoL/L) at admission	**< 7**	0	**7 – 14**	+1	**> 14**	+3
C-reactive protein (mg/L) at admisson	**< 50**	0	**50 – 100**	+1	**≥ 100**	+2

15- Knight et al. *British Medical Journal*, 2020

General notations

y_{ijk}: obs of marker k in patient i at time t_{ijk}
$Score_i$: baseline 4C-Score for patient i

$$y_{ijk} = \boldsymbol{m_k(\psi_{ik}, t_{ijk})} + g\left[m_k(\psi_{ik}, t_{ijk}), \sigma_k\right]\varepsilon_{ij} \longrightarrow \text{Mixed-effects model}$$

Subdistribution parametrization

$$h_{1ik}(t) = h_{01k} \times \exp(\boldsymbol{\alpha_{1k}} \times \boldsymbol{m_k(\psi_{ik}, t)} + \beta_{1k} \times Score_i) \longrightarrow \text{instantaneous risk of in-hospital death}$$

$$h_{2ik}(t) = h_{02k} \times \exp(\boldsymbol{\alpha_{2k}} \times \boldsymbol{m_k(\psi_{ik}, t)} + \beta_{2k} \times Score_i) \longrightarrow \text{instantaneous risk of discharge from hospital}$$

Linear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{1ik} \times t_{ijk}$

Nonlinear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{aik} \times \left[\exp(\psi_{1ik} \times t_{ijk}) - \exp(\psi_{2ik} \times t_{ijk})\right]$

$\psi_{.ik} = \mu_{.k} + \eta_{.ik}$

$\psi_{aik} = \mu_{ak} \times \exp(\eta_{aik})$

$\eta_{.ik} \sim \mathcal{N}(0, \Omega_k)$

$\varepsilon_{ij} \sim \mathcal{N}(0,1)$

General notations

y_{ijk}: obs of marker k in patient i at time t_{ijk}
$Score_i$: baseline 4C-Score for patient i
K: number of biomarkers involved

$$y_{ij1} = \boldsymbol{m_1(\psi_{i1}, t_{ij1})} + g\left[m_1(\psi_{i1}, t_{ij1}), \sigma_1\right]\varepsilon_{ij}$$
$$\dots$$
$$y_{ijK} = \boldsymbol{m_K(\psi_{iK}, t_{ijK})} + g\left[m_K(\psi_{iK}, t_{ijK}), \sigma_K\right]\varepsilon_{ij}$$

$$h_{1i}(t) = h_{01} \times \exp(\boldsymbol{\alpha_{11}} \times \boldsymbol{m_1(\psi_{i1}, t)} + \dots + \boldsymbol{\alpha_{1K}} \times \boldsymbol{m_K(\psi_{iK}, t)} + \beta_1 \times Score_i)$$
$$h_{2i}(t) = h_{02} \times \exp(\boldsymbol{\alpha_{21}} \times \boldsymbol{m_1(\psi_{i1}, t)} + \dots + \boldsymbol{\alpha_{2K}} \times \boldsymbol{m_K(\psi_{iK}, t)} + \beta_2 \times Score_i)$$

Linear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{1ik} \times t_{ijk}$

Nonlinear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{aik} \times \left[\exp(\psi_{1ik} \times t_{ijk}) - \exp(\psi_{2ik} \times t_{ijk})\right]$

$\psi_{.ik} = \mu_{.k} + \eta_{.ik}$

$\psi_{aik} = \mu_{ak} \times \exp(\eta_{aik})$

$\eta_{.ik} \sim \mathcal{N}(0, \Omega_k)$

$\varepsilon_{ij} \sim \mathcal{N}(0,1)$

General notations

y_{ijk}: obs of marker k in patient i at time t_{ijk}
$Score_i$: baseline 4C-Score for patient i
K: number of biomarkers involved

$$y_{ij1} = \boldsymbol{m_1(\psi_{i1}, t_{ij1})} + g\left[m_1(\psi_{i1}, t_{ij1}), \sigma_1\right]\varepsilon_{ij}$$
$$\dots$$
$$y_{ijK} = \boldsymbol{m_K(\psi_{iK}, t_{ijK})} + g\left[m_K(\psi_{iK}, t_{ijK}), \sigma_K\right]\varepsilon_{ij}$$

$$h_{1i}(t) = h_{01} \times \exp(\boldsymbol{\alpha_{11}} \times \boldsymbol{m_1(\psi_{i1}, t)} + \dots + \boldsymbol{\alpha_{1K}} \times \boldsymbol{m_K(\psi_{iK}, t)} + \beta_1 \times Score_i)$$
$$h_{2i}(t) = h_{02} \times \exp(\boldsymbol{\alpha_{21}} \times \boldsymbol{m_1(\psi_{i1}, t)} + \dots + \boldsymbol{\alpha_{2K}} \times \boldsymbol{m_K(\psi_{iK}, t)} + \beta_2 \times Score_i)$$

Linear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{1ik} \times t_{ijk}$

Nonlinear model: $\boldsymbol{m_k(\psi_{ik}, t_{ijk})} = \psi_{0ik} + \psi_{aik} \times \left[\exp(\psi_{1ik} \times t_{ijk}) - \exp(\psi_{2ik} \times t_{ijk})\right]$

$$\psi_{.ik} = \mu_{.k} + \eta_{.ik}$$
$$\psi_{aik} = \mu_{ak} \times \exp(\eta_{aik})$$
$$\eta_{.ik} \sim \mathcal{N}(0, \Omega_k)$$
$$\varepsilon_{ij} \sim \mathcal{N}(0,1)$$

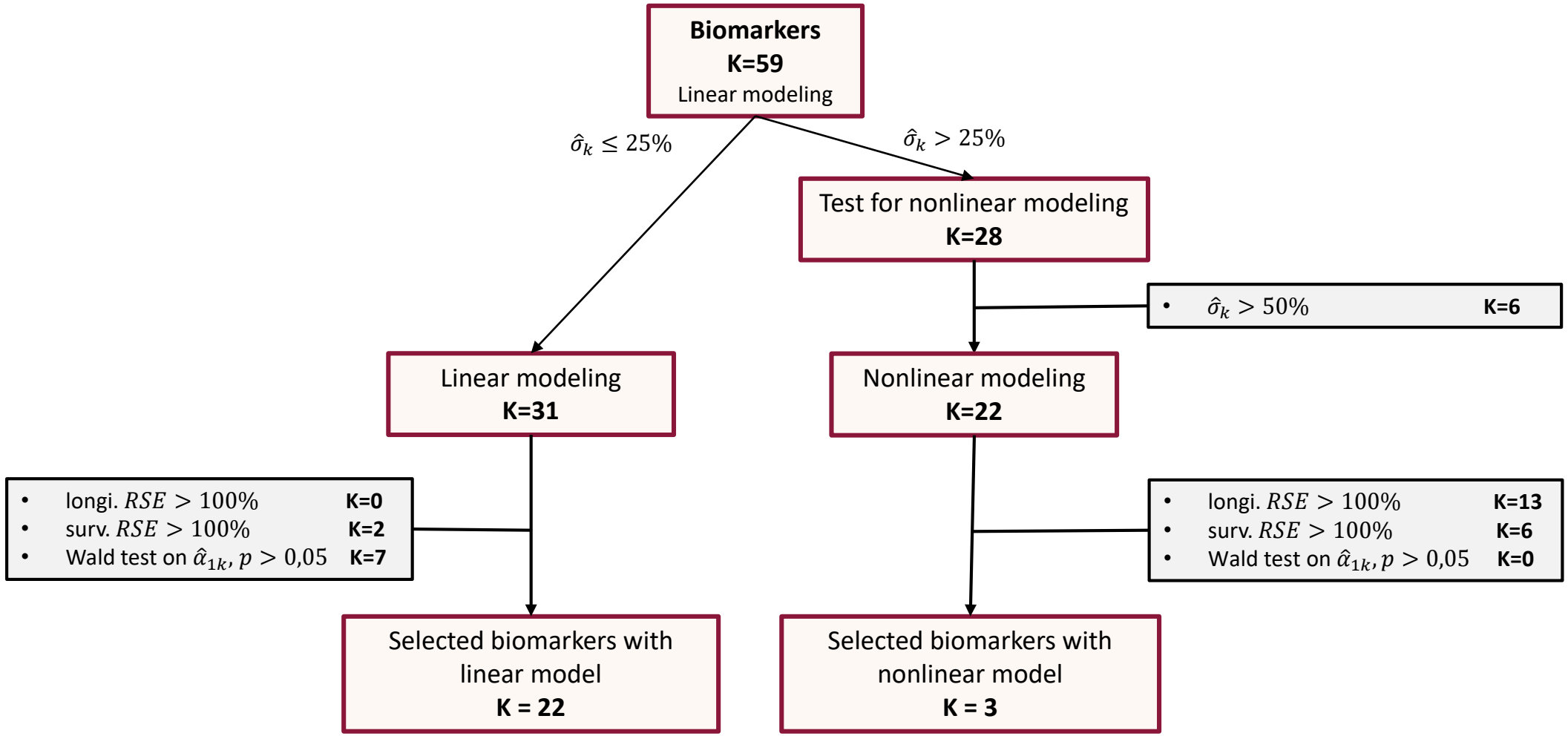
Estimation

$\theta = (\mu, \Omega, \sigma, h_{01}, h_{02}, \alpha_1, \alpha_2, \beta_1, \beta_2)$

Monolix software version 2018R2

$\mu = (\mu_1, \dots, \mu_K)$
$\Omega = diag(\omega_1, \dots, \omega_K)$
$\sigma = (\sigma_1, \dots, \sigma_K)$
$\alpha_1 = (\alpha_{11}, \dots, \alpha_{1K})$
$\alpha_2 = (\alpha_{21}, \dots, \alpha_{2K})$

Univariate joint models



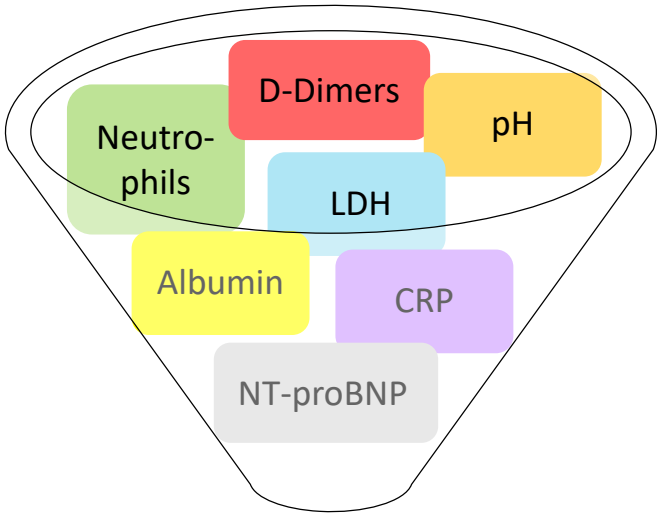
Univariate joint models

Biomarker	α_1	RSE (α_1)	-log10(p)	model
Complete blood count				
Neutrophil polynuclear cells	**0.24**	**16.17**	**9.20**	**nonlin**
Platelets	-0.004	27.66	3.52	lin
Erythrocytes	-0.44	45.28	1.57	lin
Hemoglobin	-0.14	49.06	1.38	lin
Coagulation				
D-Dimers	**1.08**	**14.86**	**10.78**	**lin**
Activated facteur V	0.04	18.40	7.26	lin
aPTT	1.50	20.00	6.24	lin
Fibrinogen	0.70	22.10	5.22	lin
Activated facteur II	-0.02	45.62	1.55	lin
Pulmonary functions				
pHa	**-20.61**	**11.18**	**18.42**	**lin**
pCO2a	0.19	12.48	14.95	lin
Oxyhemoglobin ratio	-2.04	45.23	1.57	lin
Markers of inflammation				
CRP	**1.25**	**17.63**	**7.85**	**lin**
Haptoglobin	0.42	19.03	6.83	lin
Orosomucoid	1.85	19.87	6.32	lin

Biomarker	α_1	RSE (α_1)	-log10(p)	model
Blood kidney functions/cellular lysis				
Lactate deshydrogenase (LDH)	**0.01**	**12.97**	**13.90**	**lin**
Uremia	0.07	18.11	7.48	nonlin
Kaliuresis	0.10	19.19	6.73	nonlin
Magnesium	6.65	23.96	4.52	lin
Calcemia	-6.00	25.27	4.12	lin
Creatininemia	0.003	32.27	2.71	lin
Phosphates	1.92	39.10	1.98	lin
Kalemia	0.99	44.01	1.64	lin
Urine kidney functions				
Liver/pancreatic functions				
Albuminemia	**-0.11**	**27.43**	**3.57**	**lin**
Lipasemia	0.88	17.22	8.19	lin
Cardiac markers				
NT-proBNP	**0.48**	**23.01**	**4.86**	**lin**

Multivariate joint models

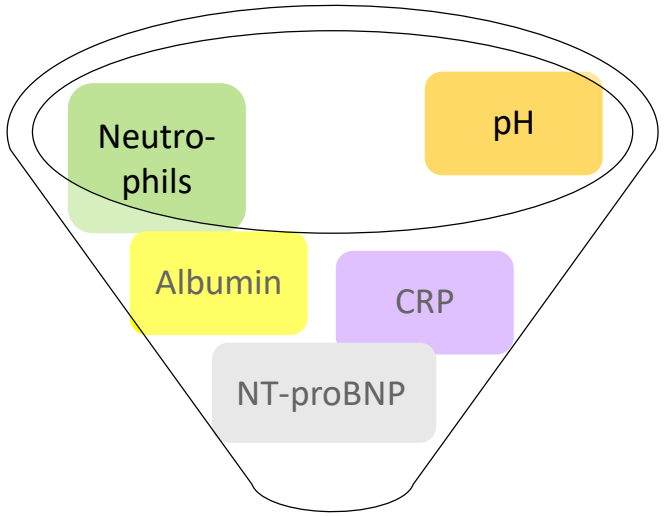
Initial multivariate joint model



death

$\hat{\alpha}_{1k}$

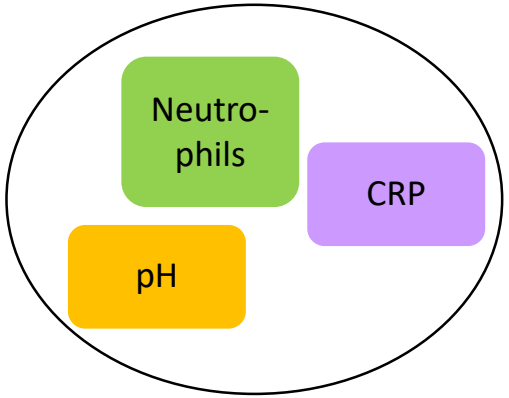
Intermediate multivariate joint model



discharge

$\hat{\alpha}_{2k}$

Final multivariate joint model

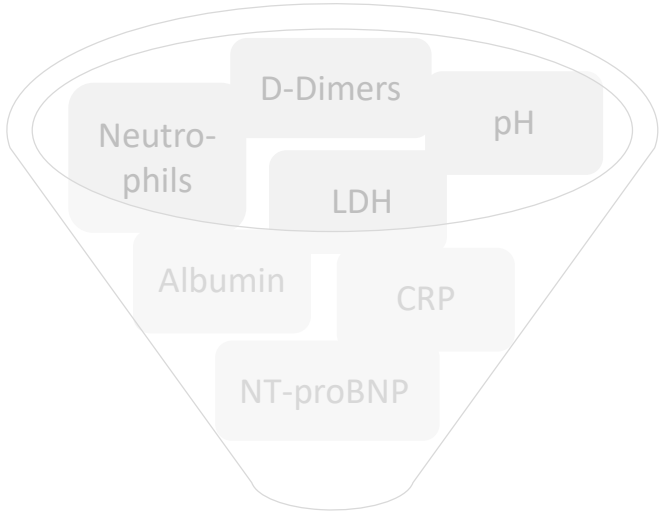


Backward selections

- Removed the highest p-value Wald test for $\hat{\alpha}_{.k}$
- Stop when all p-values for $< 5\%$

Multivariate joint models

Initial multivariate joint model



$\hat{\alpha}_{1k}$

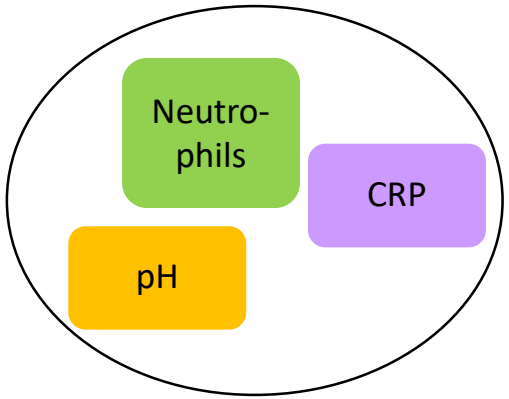
Intermediate multivariate joint model

Parameter	Value	R.S.E.(%)	p-value
Death			
α_{1neut} (L×10^{-9})	0.14	24.2	$< 10^{-4}$
α_{1pH}	-11.4	24.9	$< 10^{-4}$
α_{1CRP} (L.mg^{-1})	0.63	34.7	0.004
β_1	0.33	18.1	$< 10^{-4}$
Discharge			
α_{2neut} (L×10^{-9})	-0.14	39.1	0.01
α_{2pH}	25.2	25.2	$< 10^{-4}$
α_{2CRP} (L.mg^{-1})	-1.09	16.2	$< 10^{-4}$
β_2	-0.12	27.3	0.0002

Link and covariate parameter estimates (final multivariate joint model)

- Stop when all p-values for $< 5\%$

Final multivariate joint model

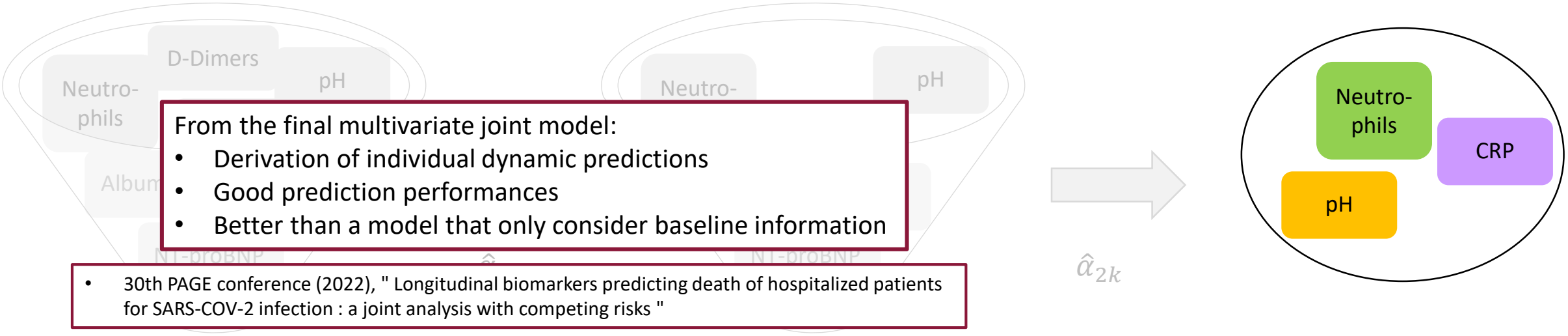


Multivariate joint models

Initial multivariate joint model

Intermediate multivariate joint model

Final multivariate joint model



From the final multivariate joint model:

- Derivation of individual dynamic predictions
- Good prediction performances
- Better than a model that only consider baseline information

- 30th PAGE conference (2022), " Longitudinal biomarkers predicting death of hospitalized patients for SARS-COV-2 infection : a joint analysis with competing risks "

Backward selections

- Removed the highest p-value Wald test for $\hat{\alpha}_{.k}$
- Stop when all p-values for $< 5\%$

Objectives

- Evaluate the performances of the estimation
- Assess the validity of the proposed selection strategy

Objectives

- Evaluate the performances of the estimation
- Assess the validity of the proposed selection strategy

Data generating mechanism

$M = 100$ datasets of $N = 300$ patients

$K = 7$ biomarkers (bm_1 to bm_7) simulated according to the design of the application (multivariate stage)

2 failure causes: death (event 1) and discharge (event 2)

Objectives

- Evaluate the performances of the estimation
- Assess the validity of the proposed selection strategy

Data generating mechanism

$M = 100$ datasets of $N = 300$ patients

$K = 7$ biomarkers (bm_1 to bm_7) simulated according to the design of the application (multivariate stage)

2 failure causes: death (event 1) and discharge (event 2)

Biomarker	Longitudinal submodel	Error model	Measurement frequency	Association with event 1	Correlation on slopes
bm_1	Nonlinear	Proportional	Every 2 days	✓	
bm_2	Linear	Additive	Every 1.5 days	✓	
bm_3	Linear	Additive	Every 2 days	✓	
bm_4	Linear	Additive	Every 3 days		$\rho(\eta_{.4}, \eta_{.2}) = 0.8$
bm_5	Linear	Proportional	Every 3 days		$\rho(\eta_{.5}, \eta_{.3}) = 0.8$
bm_6	Linear	Proportional	Every 3 days		
bm_7	Linear	Proportional	Every 3 days		

$$h_{i1}(t, \psi_i; \theta) = \frac{p_1 g_1 \exp(-g_1 \times t)}{1 - p_1(1 - \exp(-g_1 \times t))} \exp\left(\alpha_{11} \times m_1(\psi_{i1}, t) + \cdots + \alpha_{13} \times m_3(\psi_{i3}, t)\right)$$

$$h_{i2}(t, \psi_i; \theta) = \frac{1}{b} \times \frac{(1 - F_1(\infty)) \exp(-t/b)}{1 - (1 - F_1(\infty))(1 - \exp(-t/b))}$$

Estimands and performances measures

Objective 1: assess the performances of the estimation

For a simulation $m \in \{1, \dots, M\}$:

- Estimation of θ (true model parameters)
- Performances assessed with relative estimation errors:

$$\mathrm{REE}^m(\hat{\theta}) = \frac{\hat{\theta}^m - \theta}{\theta} \times 100$$

Estimands and performances measures

Objective 1: assess the performances of the estimation

For a simulation $m \in \{1, \dots, M\}$:

- Estimation of θ (true model parameters)
- Performances assessed with relative estimation errors:

$$\mathrm{REE}^m(\hat{\theta}) = \frac{\hat{\theta}^m - \theta}{\theta} \times 100$$

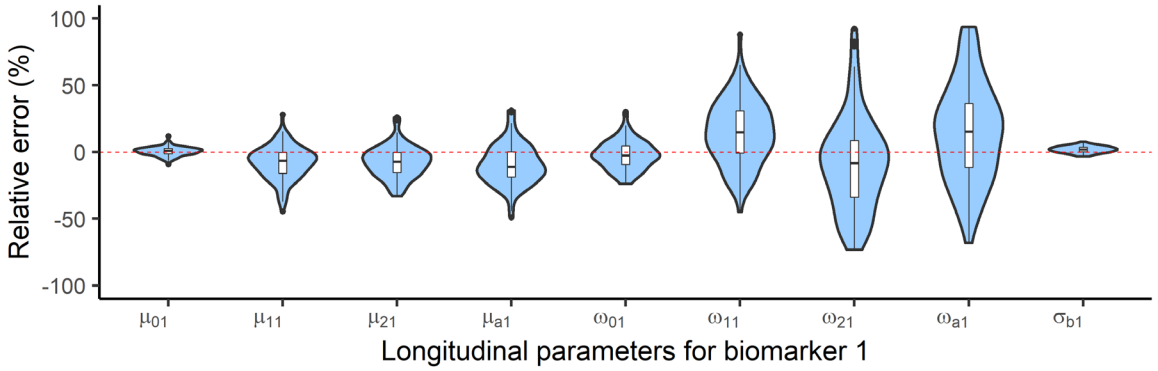
Objective 2: assess the ability of the backward strategy to find the "true" model

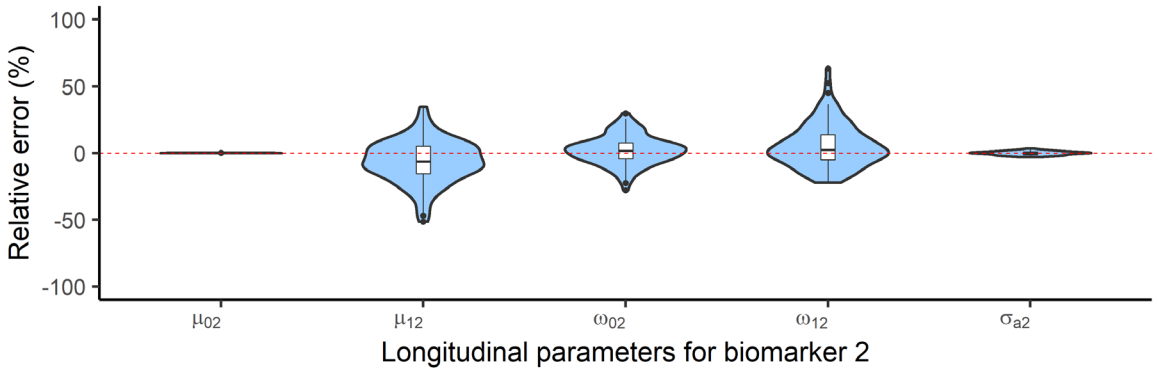
For a simulation $m \in \{1, \dots, M\}$:

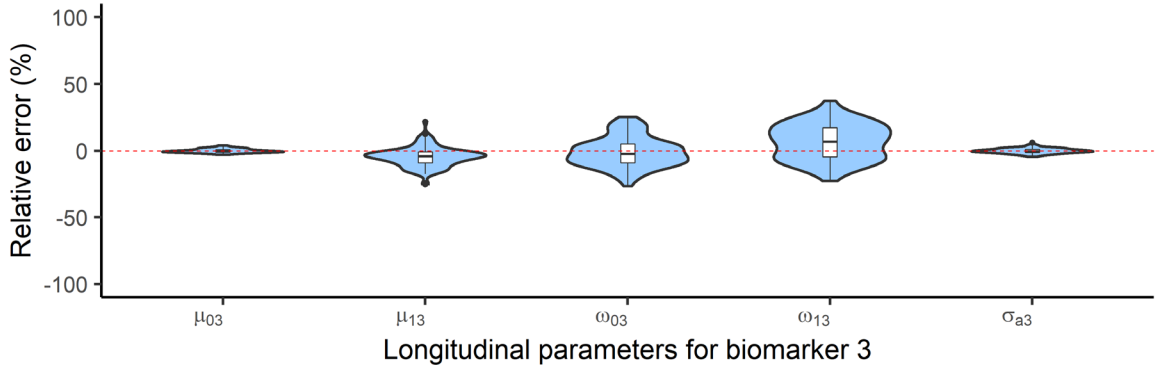
- Start with the full multivariate model (7 biomarkers)
- Backward process on $\hat{\alpha}_{1k}$ (stop when all p-values are $< 5\%$)
- Performances assessed by reporting the final set of biomarkers

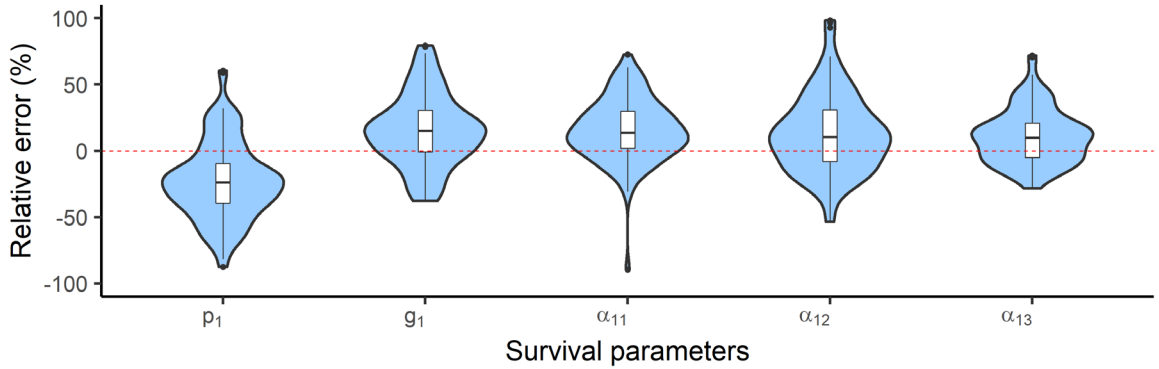
Results

Objective 1: assess the performances of the estimation





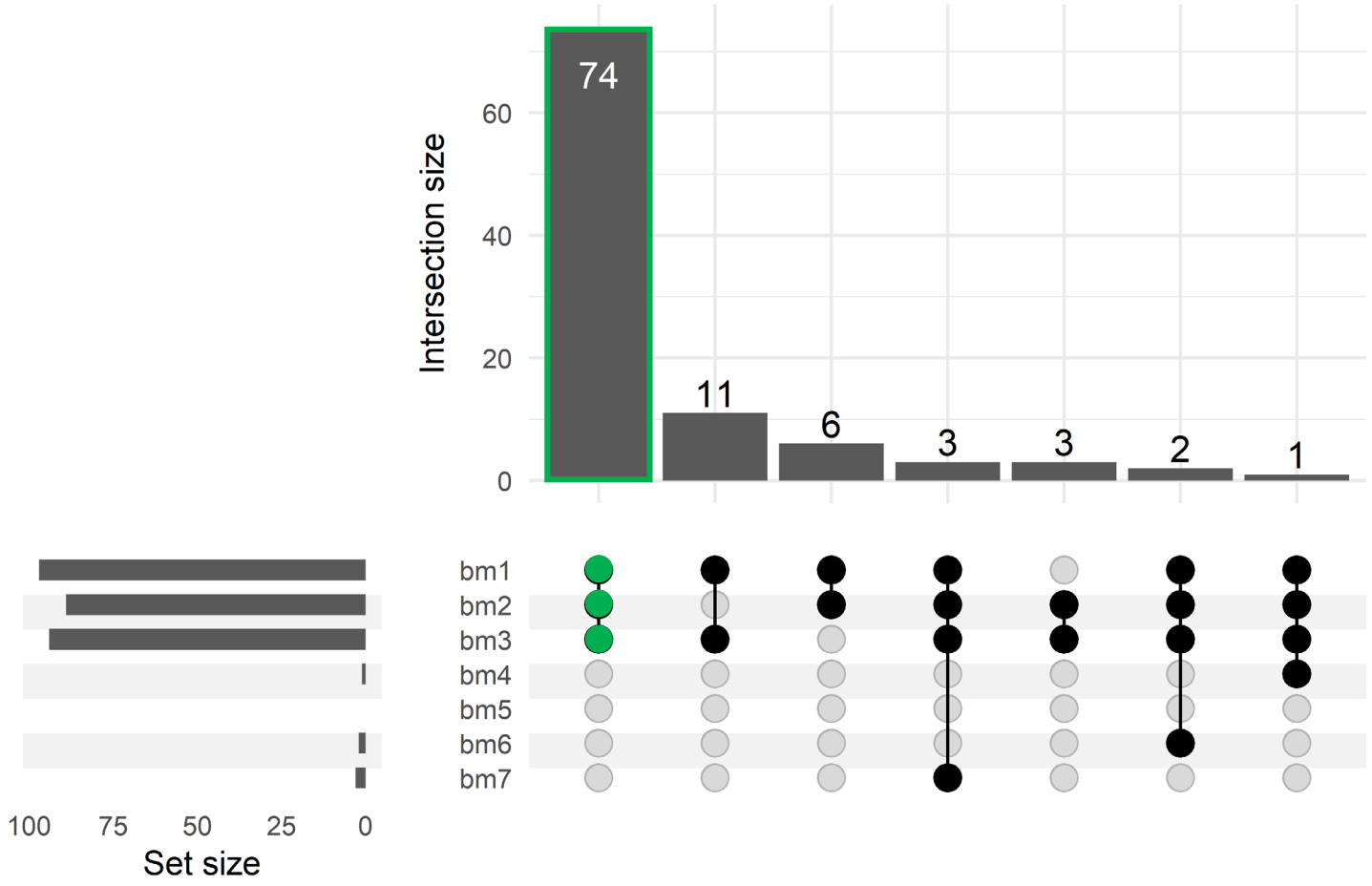




Distribution of the Relative Estimation Errors for each model parameter

Results

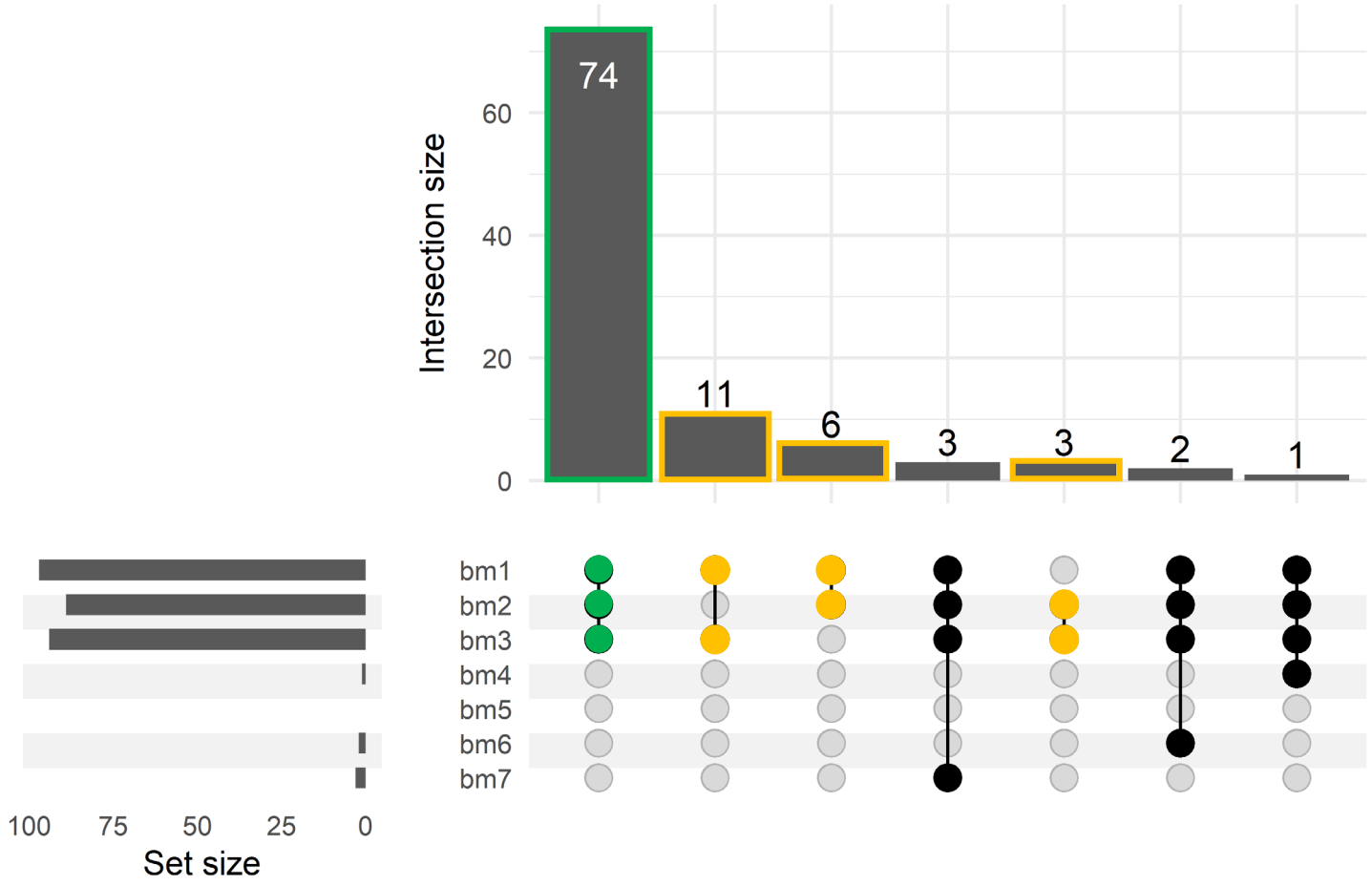
Objective 2: assess the ability of the backward strategy to find the "true" model



Final set of biomarkers selected after the backward process for each simulation

Results

Objective 2: assess the ability of the backward strategy to find the "true" model



Final set of biomarkers selected after the backward process for each simulation

Objectives: extend the *saemix* package to the case of multi-responses and joint models

- Re-defining the likelihood expression (to allow for multiple outcomes)
- Implementing an algorithm for Fisher Information Matrix (FIM) computation, noted $I(\hat{\theta})$

16- Louis, *Journal of the Royal Statistical Society*, 1982
17- Delattre and Kuhn, *hal-02285712v2,* 2023

Objectives: extend the *saemix* package to the case of multi-responses and joint models

- Re-defining the likelihood expression (to allow for multiple outcomes)
- Implementing an algorithm for Fisher Information Matrix (FIM) computation, noted $I(\hat{\theta})$
  - Variance-covariance matrix of $\hat{\theta}$:

$$\hat{\Sigma} = I(\hat{\theta})^{-1}$$

  - In joint model context: $I(\hat{\theta})$ computed by stochastic approximations
  - Monolix software: Louis's method[16] (time consuming…)

16- Louis, *Journal of the Royal Statistical Society*, 1982
17- Delattre and Kuhn, *hal-02285712v2,* 2023

Objectives: extend the *saemix* package to the case of multi-responses and joint models

- Re-defining the likelihood expression (to allow for multiple outcomes)
- Implementing an algorithm for Fisher Information Matrix (FIM) computation, noted $I(\hat{\theta})$
  - Variance-covariance matrix of $\hat{\theta}$:

$$\hat{\Sigma} = I(\hat{\theta})^{-1}$$

  - In joint model context: $I(\hat{\theta})$ computed by stochastic approximations
  - Monolix software: Louis's method[16] (time consuming…)
  - Alternative stochastic algorithm developed by Delattre and Kuhn[17]

→ interesting properties because it is integrated in the SAEM algoritm (time saving)
→ but not currently evaluated for joint models

16- Louis, *Journal of the Royal Statistical Society*, 1982
17- Delattre and Kuhn, *hal-02285712v2,* 2023

Evaluation by simulations

Aims

1. Evaluate the SAEM algorithm extended in R *saemix* package[14] for joint models:

	Single event	2 competing risks
Linear mixed-effects model	LMEM – TTE	LMEM – CR
Nonlinear mixed-effects model	NLMEM – TTE	NLMEM – CR

2. Evaluate the algorithm developed by Delattre and Kuhn[17] for the 4 previous models (standard errors estimation)

14- Comets et al. *Journal of Statistical Software*, 2017
17- Delattre and Kuhn, *hal-02285712v2,* 2023

Evaluation by simulations

Data generating mechanism

For each of the 4 models presented, we simulate M = 100 datasets of N = 100 patients.
Biomarker measurements available each day until time-to-event for at most 30 days

LMEM – TTE

$$y_{ij} = m_l(\psi_i, t_{ij}) + \sigma\epsilon_{ij}$$
$$h_i(t, \psi_i; \theta) = h_0 \times \exp\left(\boldsymbol{\alpha} \times m_l(\psi_i, t)\right)$$

NLMEM-TTE

$$y_{ij} = m_{nl}(\psi_i, t_{ij}) + \sigma\epsilon_{ij}$$
$$h_i(t, \psi_i; \theta) = h_0 \times \exp\left(\boldsymbol{\alpha} \times m_{nl}(\psi_i, t)\right)$$

LMEM-CR

$$y_{ij} = m_l(\psi_i, t_{ij}) + \sigma\epsilon_{ij}$$
$$h_{i1}(t, \psi_i; \theta) = \frac{p_1 g_1 \exp(-g_1 \times t)}{1 - p_1(1 - \exp(-g_1 \times t))} \exp\left(\boldsymbol{\alpha_1} \times m_l(\psi_i, t)\right)$$
$$h_{i2}(t, \psi_i; \theta) = \frac{1}{b} \times \frac{(1 - F_1(\infty)) \exp(-t/b)}{1 - (1 - F_1(\infty))(1 - \exp(-t/b))}$$

NLMEM-CR

$$y_{ij} = m_{nl}(\psi_i, t_{ij}) + \sigma\epsilon_{ij}$$
$$h_{i1}(t, \psi_i; \theta) = \frac{p_1 g_1 \exp(-g_1 \times t)}{1 - p_1(1 - \exp(-g_1 \times t))} \exp\left(\boldsymbol{\alpha_1} \times m_{nl}(\psi_i, t)\right)$$
$$h_{i2}(t, \psi_i; \theta) = \frac{1}{b} \times \frac{(1 - F_1(\infty)) \exp(-t/b)}{1 - (1 - F_1(\infty))(1 - \exp(-t/b))}$$

Evaluation by simulations

Estimands

Objective 1: assess parameter estimation

JM LMEM/NLMEM – TTE

$\theta = (\mu, \Omega, \sigma, h_0, \alpha)$

JM LMEM/NLMEM – CR

$\theta = (\mu, \Omega, \sigma, p_1, g_1, \alpha, b)$

Performance measures

Relative estimation errors:

$$\mathrm{REE}^m(\hat{\theta}) = \frac{\hat{\theta}^m - \theta}{\theta} \times 100$$

Evaluation by simulations

Estimands

Objective 1: assess parameter estimation

JM LMEM/NLMEM – TTE

$\theta = (\mu, \Omega, \sigma, h_0, \alpha)$

JM LMEM/NLMEM – CR

$\theta = (\mu, \Omega, \sigma, p_1, g_1, \alpha, b)$

Objective 2: assess standard error estimation

JM LMEM/NLMEM – TTE/CR

$\hat{\Sigma}$ = variance-covariance matrix of $\hat{\theta}$

Performance measures

Relative estimation errors:

$$\mathrm{REE}^m(\hat{\theta}) = \frac{\hat{\theta}^m - \theta}{\theta} \times 100$$

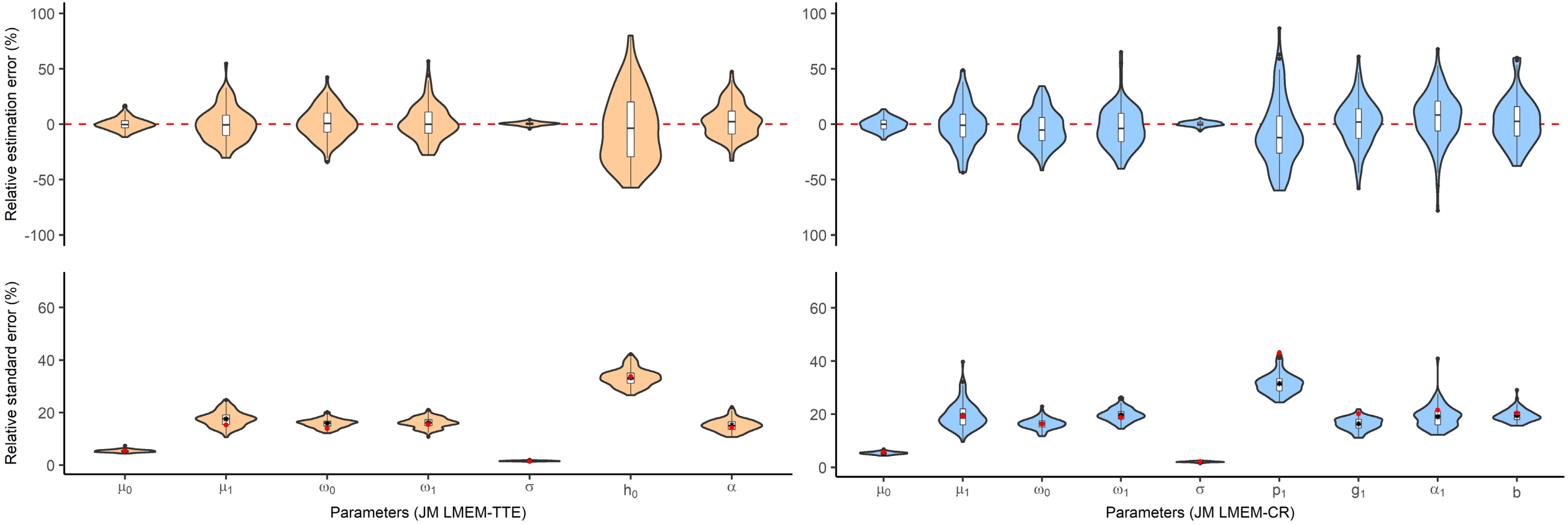
Relative standard errors:

$$\mathrm{RSE}^m(\hat{\theta}) = \frac{\sqrt{diag(\hat{\Sigma}^m)}}{\hat{\theta}^m} \times 100$$

Relative empirical error:

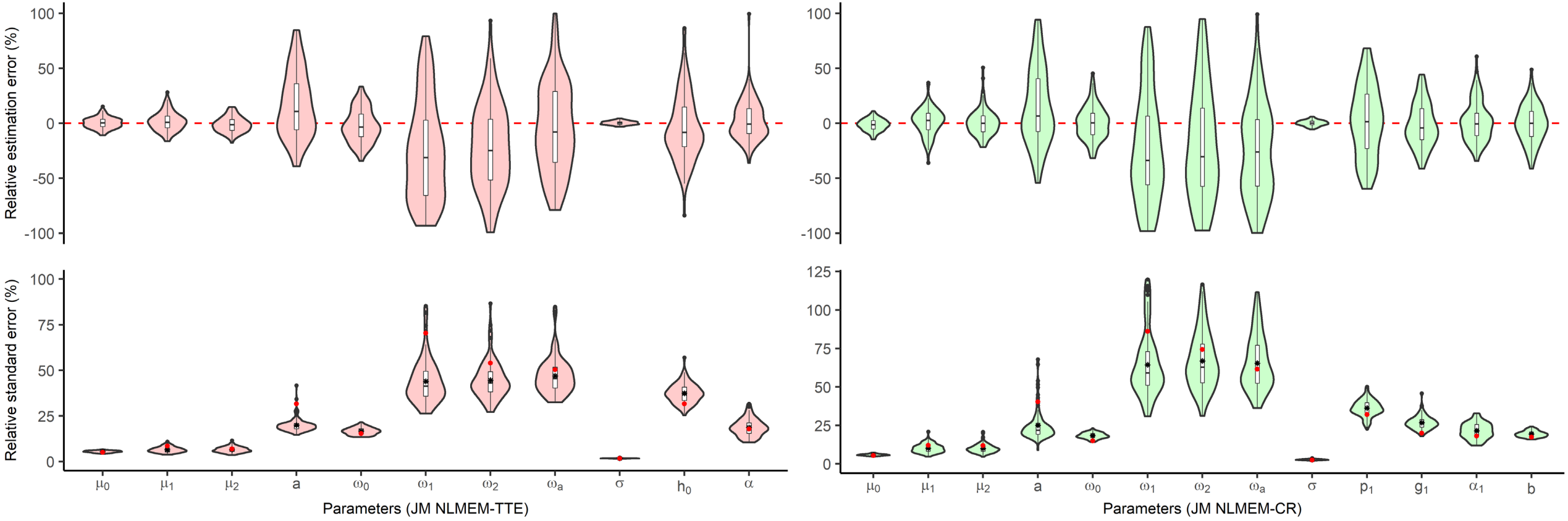
$$\mathrm{RSE}^{emp}(\hat{\theta}) = \sqrt{\frac{1}{m-1}\sum_{m=1}^{M}\left(\hat{\theta}^m - \bar{\hat{\theta}}\right)^2} \times \frac{100}{\bar{\hat{\theta}}}$$

Distribution of REE (top) and stochastic RSE versus empirical RSE (bottom) - JM LMEM-TTE and JM LMEM-CR



- Empirical RSE
- ★ Mean distribution of the stochastic RSE

Distribution of REE (top) and stochastic RSE versus empirical RSE (bottom) - JM NLMEM-TTE and JM NLMEM-CR



- Empirical RSE
- Mean distribution of the stochastic RSE

COVID-19 case study

- Developments of a multivariate joint model to predict the death of patients hospitalized for SARS-CoV-2 infection and a strategy to select among various biomarkers
- Evolution of neutrophils, pH and CRP are predictive of the death/discharge of patients
- Identify biomarkers in other emergent diseases

COVID-19 case study

- Developments of a multivariate joint model to predict the death of patients hospitalized for SARS-CoV-2 infection and a strategy to select among various biomarkers
- Evolution of neutrophils, pH and CRP are predictive of the death/discharge of patients
- Identify biomarkers in other emergent diseases

Methodological assessment

- Monolix software (version 2018R2) provides unbiased and accurate estimates of such complex model parameters
- The backward strategy yields good performances

COVID-19 case study

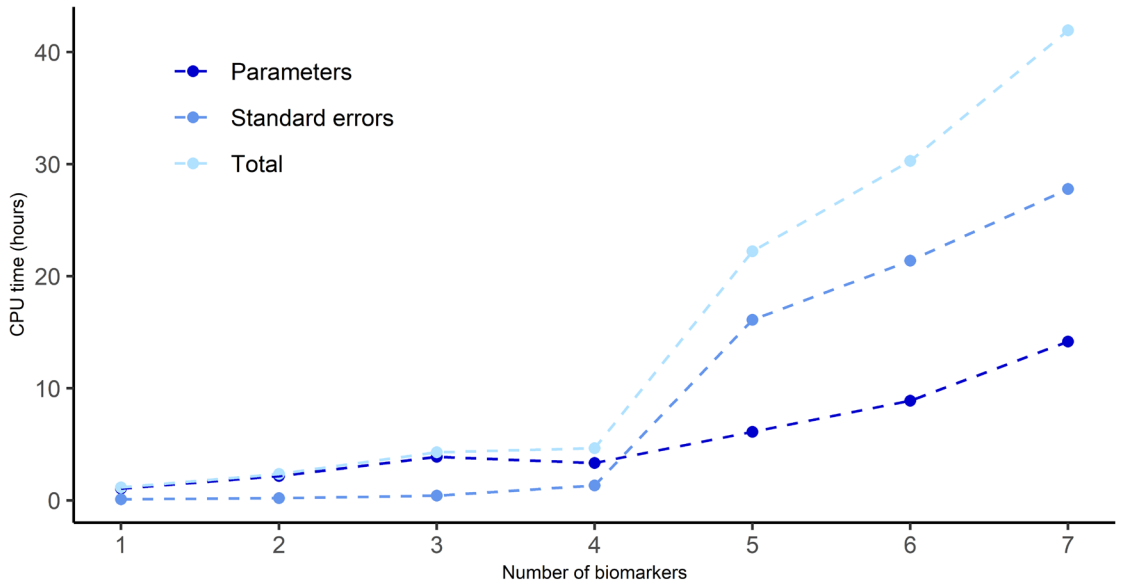
- Developments of a multivariate joint model to predict the death of patients hospitalized for SARS-CoV-2 infection and a strategy to select among various biomarkers
- Evolution of neutrophils, pH and CRP are predictive of the death/discharge of patients
- Identify biomarkers in other emergent diseases

Methodological assessment

- Monolix software (version 2018R2) provides unbiased and accurate estimates of such complex model parameters
- The backward strategy yields good performances

Limitations

- Limited number of biomarkers included in the multivariate analysis (computational limit)
- Backward strategy usually outperformed by penalized regression methods[18] (LASSO penalization)



CPU time spent to estimate a joint model involving a given number of biomarkers

18- Tibshirani, *Journal of the Royal Statistical Society,* 1996

saemix extension

- Extension of the R package *saemix* to the case of joint models
- Good properties for parameter and standard errors estimation
- Flexible tool in parametric joint model framework
- Users define the likelihood of the model (very specific joint models can be considered)

Perspectives

- Need to evaluate for multiple longitudinal biomarkers
- Need to develop goodness of fit tools in the package

Functions and examples available on Github:
https://github.com/saemixdevelopment/saemixextension/tree/master/joint

Acknowledgements

Pr Jean-François Timsit
The Outcomerea network

Pr Xavier Lescure
Dr Nathan Peiffer-Smadja
Dr Simon Gressens
Dr Alexandre Lahens
Dr Agathe Bounhiol
Dr Bérénice Souhail

Thank you for your attention !