Joint model with heterogeneous variance for studying the risk of stroke associated with blood pressure variability

Léonie Courcoul, Antoine Barbieri, Christophe Tzourio, Hélène Jacqmin-Gadda

Univ. Bordeaux, ISPED, Inserm BPH U1219, Biostatistic Team

Journées des Biostatistiques, 17 & 18th November 2022, Rennes

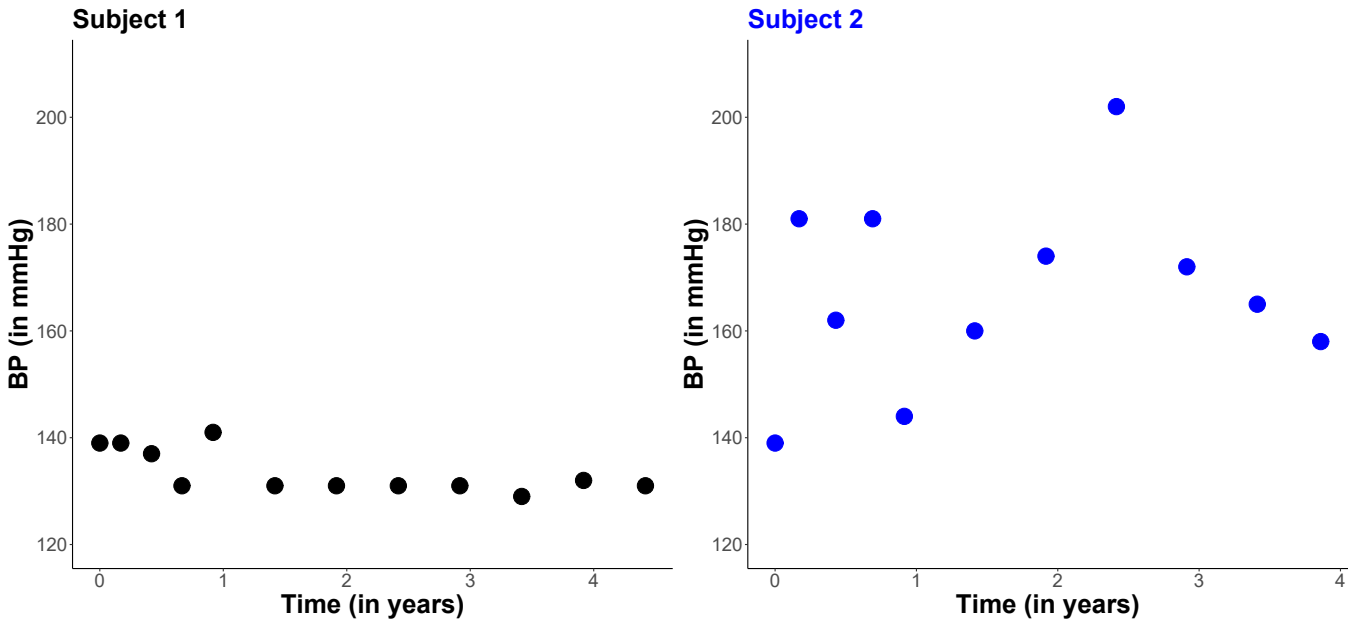
Clinical motivation - Stroke

- 15 million stroke victims per year
- The first cause of acquired physical disability
- The second leading cause of dementia after Alzheimer's disease
- The second leading cause of death
- High risk of recidivism

$\Longrightarrow$ **Need to identify the risk factors**

Clinical motivation - Blood pressure

- High level of blood pressure : a known risk factor
- Blood pressure variability : supposed to be a risk factor by some studies



Methodological motivation

- Methodological weaknesses of existing studies: [1]
  - conditioning on the future => risk of bias
  - selection bias
  - measurement error on the standard deviation not taken into account
  - competing deaths from other causes not considered

[1] H de Courson et al. European heart journal. 39(47):4220–4220, 2018.

Methodological motivation - Objective

Objective

To propose a joint model with a longitudinal marker and competing risks which depend on the current value, the slope and the intra-subject variability of the marker

- **Main features:**
  - Flexible dependance structure (current value and slope instead of random effects only)
  - Flexible baseline risk function
  - Competing risks and left truncation

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variability: $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variability: $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$

Random effects: $\begin{pmatrix} b_i \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_b & 0 \\ 0 & \tau_\sigma^2 \end{pmatrix}\right),$

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variability: $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$

Random effects: $\begin{pmatrix} b_i \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_b & 0 \\ 0 & \tau_\sigma^2 \end{pmatrix}\right),$

Survival submodel for competing events: For $k \in \{1, 2\}$:

$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp\left(W_i^\top \gamma_k \right)$$

$\lambda_{0k}(t)$: exponential, Weibull or B-splines

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variability: $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$

Random effects: $\begin{pmatrix} b_i \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_b & 0 \\ 0 & \tau_\sigma^2 \end{pmatrix}\right),$

Survival submodel for competing events: For $k \in \{1, 2\}$:

$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp\left(W_i^\top \gamma_k + \alpha_{1k}\tilde{y}_i(t) + \alpha_{2k}\tilde{y}_i'(t)\right)$$

$\lambda_{0k}(t)$: exponential, Weibull or B-splines

Variance-dependent joint model

Longitudinal submodel:

$$Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \epsilon_{ij} = X_{ij}^\top \beta + Z_{ij}^\top b_i + \epsilon_{ij}$$

Variability: $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$

Random effects: $\begin{pmatrix} b_i \\ \omega_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_b & 0 \\ 0 & \tau_\sigma^2 \end{pmatrix}\right),$

Survival submodel for competing events: For $k \in \{1, 2\}$:

$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp\left(W_i^\top \gamma_k + \alpha_{1k}\tilde{y}_i(t) + \alpha_{2k}\tilde{y}_i'(t) + \alpha_{\sigma k}\sigma_i\right)$$

$\lambda_{0k}(t)$: exponential, Weibull or B-splines

Joint Likelihood

- T_i : observed time
- δ_i : indicator of event (0 : censoring, 1 : first event, 2 : competing event)
- θ: the set of parameter

Joint Likelihood

- T_i : observed time
- δ_i : indicator of event (0 : censoring, 1 : first event, 2 : competing event)
- θ: the set of parameter

The individual contribution to the likelihood is :

$$L_i(\theta; Y_i, T_i, \delta_i) = \int \prod_{k=1}^{2} \exp(-\Lambda_k(T_i|b_i, \omega_i; \theta))\lambda_k(T_i|b_i, \omega_i; \theta)^{\mathbb{1}_{\delta_i=k}} f(Y_i|b_i, \omega_i; \theta) f(b_i; \theta) f(\omega_i; \theta) db_i d\omega_i$$

Joint Likelihood

- T_i : observed time
- δ_i : indicator of event (0 : censoring, 1 : first event, 2 : competing event)
- θ: the set of parameter

The individual contribution to the likelihood is :

$$L_i(\theta; Y_i, T_i, \delta_i) = \int \prod_{k=1}^{2} \exp(-\Lambda_k(T_i|b_i, \omega_i; \theta))\lambda_k(T_i|b_i, \omega_i; \theta)^{\mathbb{1}_{\delta_i=k}} f(Y_i|b_i, \omega_i; \theta) f(b_i; \theta) f(\omega_i; \theta) db_i d\omega_i$$

In case of delayed entry, we denote T_{0i} the time of entry and :

$$L_i^{DE}(\theta; Y_i, T_i, \delta_i) = \frac{L_i(\theta; Y_i, T_i, \delta_i)}{\int \exp(-\Lambda_1(T_{0i}|b_i, \omega_i; \theta)) - \Lambda_2(T_{0i}|b_i, \omega_i; \theta)) db_i d\omega_i}$$

Estimation

- Quasi-Monte-Carlo for integrating over the random effects using deterministic quasi random sequences

Estimation

- Quasi-Monte-Carlo for integrating over the random effects using deterministic quasi random sequences
- Gauss-Kronrod approximation (15 points) to estimate survival functions

Estimation

- Quasi-Monte-Carlo for integrating over the random effects using deterministic quasi random sequences
- Gauss-Kronrod approximation (15 points) to estimate survival functions
- Maximisation of the likelihood using the Marquardt-Levenberg algorithm with stringent convergence criteria, in two steps :
  - **Step 1 :** with S1 draws of Quasi-Monte-Carlo
  - **Step 2 :** with results of the step 1 as initial parameters and $S2 > S1$ draws

Design

- Random intercept and slope, intra-subject variability and competing risks :

$$\begin{cases} y_i(t_{ij}) = \tilde{y}_i(t_{ij}) + \epsilon_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij} \\ \end{cases}$$

where :

- $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$ and $\omega_i \sim \mathcal{N}(0, \tau_\sigma^2)$
- $b_i \sim \mathcal{N}(0, \Sigma_b)$ with $\begin{pmatrix} s_0 & 0 \\ s_{01} & s_1 \end{pmatrix}$ the Cholesky decomposition of Σ_b

Design

- Random intercept and slope, intra-subject variability and competing risks :

$$\begin{cases} y_i(t_{ij}) = \tilde{y}_i(t_{ij}) + \epsilon_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij} \\ \lambda_{ik}(t) = \lambda_{0k}(t)\exp(\alpha_{1k}\tilde{y}_i(t) + \alpha_{\sigma k}\sigma_i) \end{cases}$$

where :

  - $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$ and $\omega_i \sim \mathcal{N}(0, \tau_\sigma^2)$
  - $b_i \sim \mathcal{N}(0, \Sigma_b)$ with $\begin{pmatrix} s_0 & 0 \\ s_{01} & s_1 \end{pmatrix}$ the Cholesky decomposition of Σ_b
  - $\lambda_{0k}(t) = \kappa_k t^{\kappa_k - 1} e^{\zeta_{0k}}$: Weibull function

Design

- 2 scenario
- 200 simulations on 200 subjects
- $S1 = 1000$ and $S2 = 8000$
- 13 measures by subjects between 0 and 5 years

Results - Scenario 1 - 200 subjects - 200 simulations

Parameter		True value	Mean estimate	Empirical SE	Mean asymptotic SE	Coverage rate(%)
Longitudinal submodel						
Intercept	β_0	142	142.14	1.07	1.05	95.5
Slope	β_1	0.06	0.046	0.26	0.27	93.0
Variability	μ_σ	2.3	2.30	0.03	0.03	94.0
	τ_σ	0.3	0.30	0.02	0.03	96.5
Cholesky	s_0	14.5	14.36	0.81	0.77	95.0
	s_{01}	-1.1	-1.07	0.32	0.28	91.0
	s_1	2.8	2.77	0.22	0.20	91.5
Survival submodel 1						
Variability	$\alpha_{\sigma 1}$	-0.03	-0.043	0.092	0.086	94.0
Current value	α_{11}	0.03	0.031	0.013	0.013	92.5
Weibull	$\sqrt{\kappa_1}$	0.95	0.962	0.085	0.074	93.0
	ζ_{01}	-7.0	-7.09	2.12	2.06	94.0
Survival submodel 2						
Variability	$\alpha_{\sigma 2}$	0.1	0.09	0.14	0.11	97.5
Current value	α_{12}	0.007	0.0083	0.0197	0.0198	97.0
Weibull	$\sqrt{\kappa_2}$	1.07	1.083	0.128	0.128	96.0
	ζ_{02}	-6.0	-6.45	3.0	3.14	97.0

Results - Scenario 2 - 200 subjects - 200 simulations

Parameter		True value	Mean estimate	Empirical SE	Mean asymptotic SE	Coverage rate(%)
Longitudinal submodel						
Intercept	β_0	142	142.0	1.27	1.05	86.5
Slope	β_1	1.5	1.50	0.31	0.28	94.0
Variability	μ_σ	2.3	2.30	0.03	0.03	93.5
	τ_σ	0.3	0.30	0.03	0.03	95.0
Cholesky	s_0	14.5	14.45	0.89	0.78	90.5
	s_{01}	-1.1	-1.08	0.33	0.27	89.0
	s_1	2.8	2.74	0.24	0.20	87.0
Survival submodel 1						
Variability	$\alpha_{\sigma 1}$	0.3	0.32	0.13	0.12	93.0
Current value	α_{11}	0.03	0.031	0.014	0.014	96.0
Weibull	$\sqrt{\kappa_1}$	0.95	0.976	0.095	0.097	96.5
	ζ_{01}	-10.5	-10.86	2.97	2.97	96.5
Survival submodel 2						
Variability	$\alpha_{\sigma 2}$	0.3	0.32	0.13	0.13	97.0
Current value	α_{12}	0.007	0.0056	0.0165	0.0159	95.0
Weibull	$\sqrt{\kappa_2}$	1.07	1.109	0.110	0.112	96.5
	ζ_{02}	-7.5	-7.90	3.20	3.13	96.5

Clinical trial PROGRESS [2]

- Multicentric, randomised, double-blind, controlled clinical trial
- **Population:** individuals with a history of stroke or transient ischaemic attack within 5 years of inclusion
- **Follow-up:** 5 visits the first year then 2 visits per year during 4.5 years
- **Placebo arm :** 3032 patients, 406 strokes and 213 deaths without stroke

[2] PROGRESS Management Committee. J Hypertens Suppl. 1996

Model

We denote

$$\begin{cases} x_{1i} \text{ is the standardised age of subject i} \\ x_{2i} = 1 \text{ if subject } i \text{ is a man, 0 otherwise} \\ x_{3i} = 1 \text{ if subject } i \text{ is not asian, 0 otherwise} \end{cases}$$

Model

We denote

$$\begin{cases} x_{1i} \text{ is the standardised age of subject i} \\ x_{2i} = 1 \text{ if subject } i \text{ is a man, 0 otherwise} \\ x_{3i} = 1 \text{ if subject } i \text{ is not asian, 0 otherwise} \end{cases}$$

The model is then formally defined by :

$$y(t_{ij}) = \tilde{y}(t_{ij}) + \epsilon_{ij} \quad = \quad (\beta_0 + b_{0i}) + \sum_{k=1}^{3}(\beta_k + b_{ki})ns(t_{ij}) + \beta_4 x_{1i} + \beta_5 x_{2i} + \beta_6 x_{3i} + \epsilon_{ij}$$

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$ and $\omega_i \sim \mathcal{N}(0, \tau_\sigma^2)$.

Model

We denote

$$\begin{cases} x_{1i} \text{ is the standardised age of subject i} \\ x_{2i} = 1 \text{ if subject } i \text{ is a man, 0 otherwise} \\ x_{3i} = 1 \text{ if subject } i \text{ is not asian, 0 otherwise} \end{cases}$$

The model is then formally defined by :

$$y(t_{ij}) = \tilde{y}(t_{ij}) + \epsilon_{ij} \quad = \quad (\beta_0 + b_{0i}) + \sum_{k=1}^{3}(\beta_k + b_{ki})ns(t_{ij}) + \beta_4 x_{1i} + \beta_5 x_{2i} + \beta_6 x_{3i} + \epsilon_{ij}$$

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$ with $\log(\sigma_i) = \mu_\sigma + \omega_i$ and $\omega_i \sim \mathcal{N}(0, \tau_\sigma^2)$.
For $k \in \{1, 2\}$:

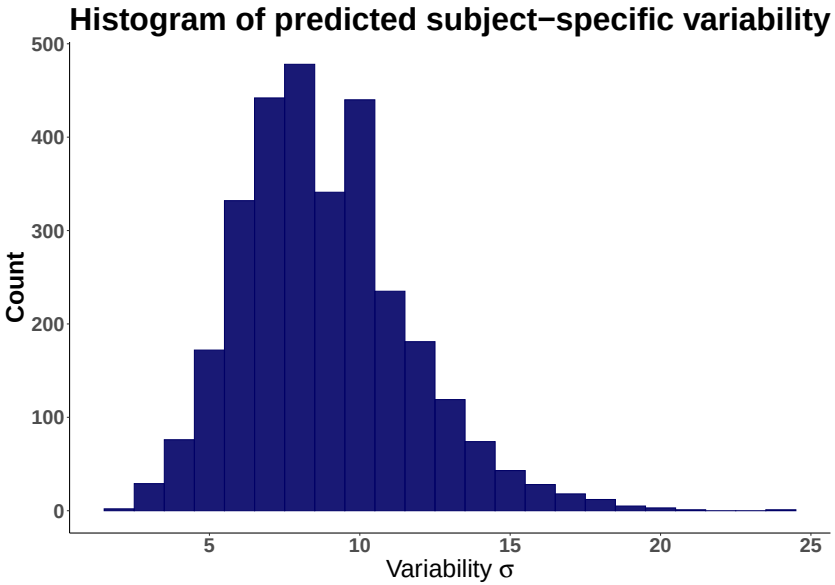
$$\lambda_{ik}(t) = \lambda_{0k}(t) \exp(\gamma_{1k} x_{1i} + \gamma_{2k} x_{2i} + \gamma_{3k} x_{3i} + \alpha_{1k} \tilde{y}_i(t) + \alpha_{\sigma_k} \sigma_i)$$

with $\lambda_{0k}(t)$ the baseline risk function defined on a B-splines bases.

Results - Placebo arm (3032 patients - 406 strokes - 213 deaths)

- Adjusted on sex, ethnicity and age:

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	p-value
Longitudinal submodel - BP			
μ_σ	2.29	0.009	< 0.001
τ_σ	0.36	0.008	< 0.001



Variable	$\hat{\theta}$	SE($\hat{\theta}$)	HR	p-val
Survival submodel 1 - Stroke				
$\alpha_{\sigma 1}$	-0.023	0.023	0.98	0.317
α_{11} *(cv)*	0.024	0.004	1.03	< 0.001
Survival submodel 2 - Death				
$\alpha_{\sigma 2}$	0.071	0.028	1.07	0.011
α_{12} *(cv)*	0.006	0.006	1.01	0.317

Goodness of fit: Longitudinal submodel

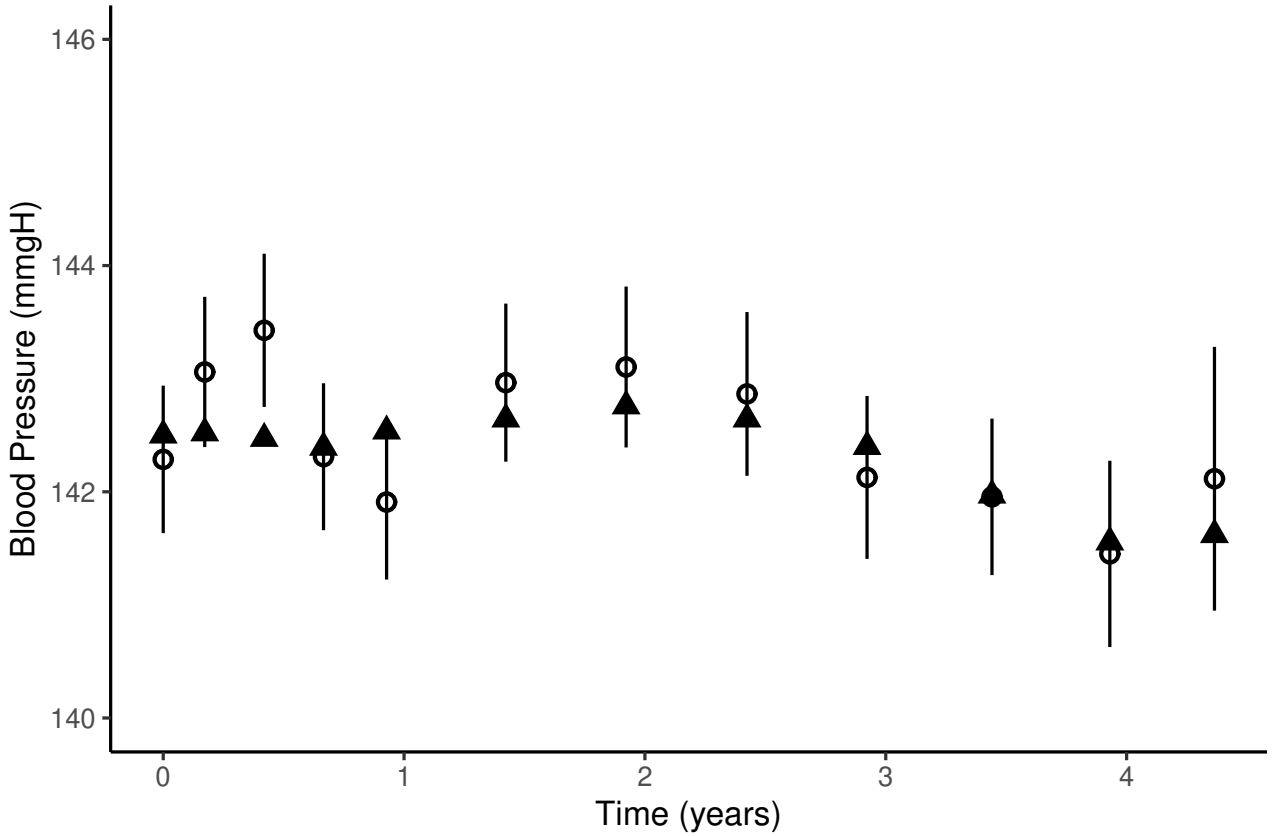


Figure: Longitudinal submodel fit assessment: the means of the BP predictions at each visit (black triangles) and the mean of observed BP values (white dots) with its 95% confidence interval (vertical black lines).

Goodness of fit: Survival submodels

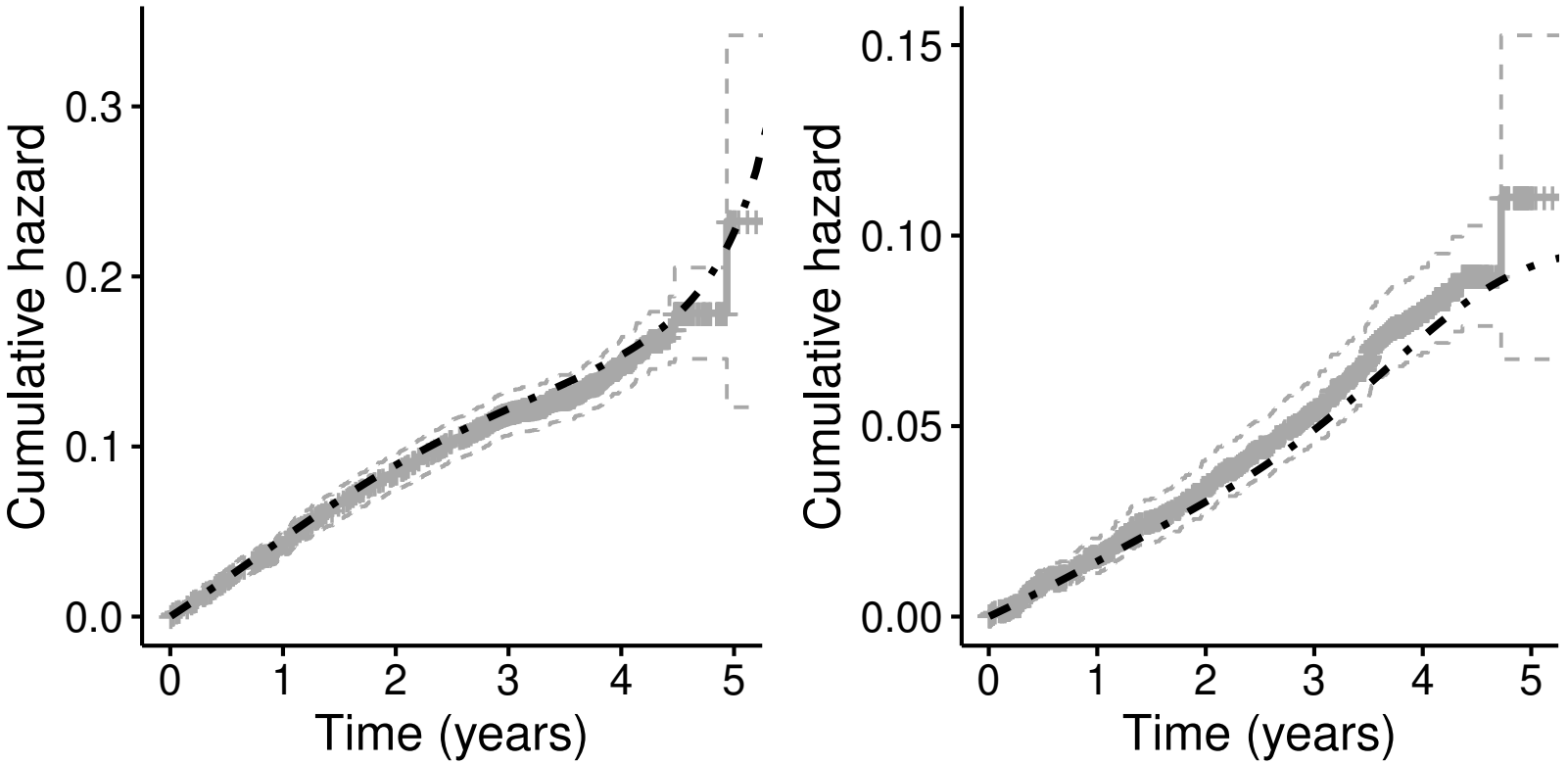


Figure: Survival submodel for stroke (left) and death (right) fit assessment: comparison between predicted cumulative hazard function (in black) and Nelson Aalen estimator (in grey).

Individual Predictions

$$\pi_i(s,t;\widehat{\theta}) = P(s < T_i < s+t, \delta_i = k | T_i > s, \mathcal{Y}_i(s), \widehat{\theta})$$

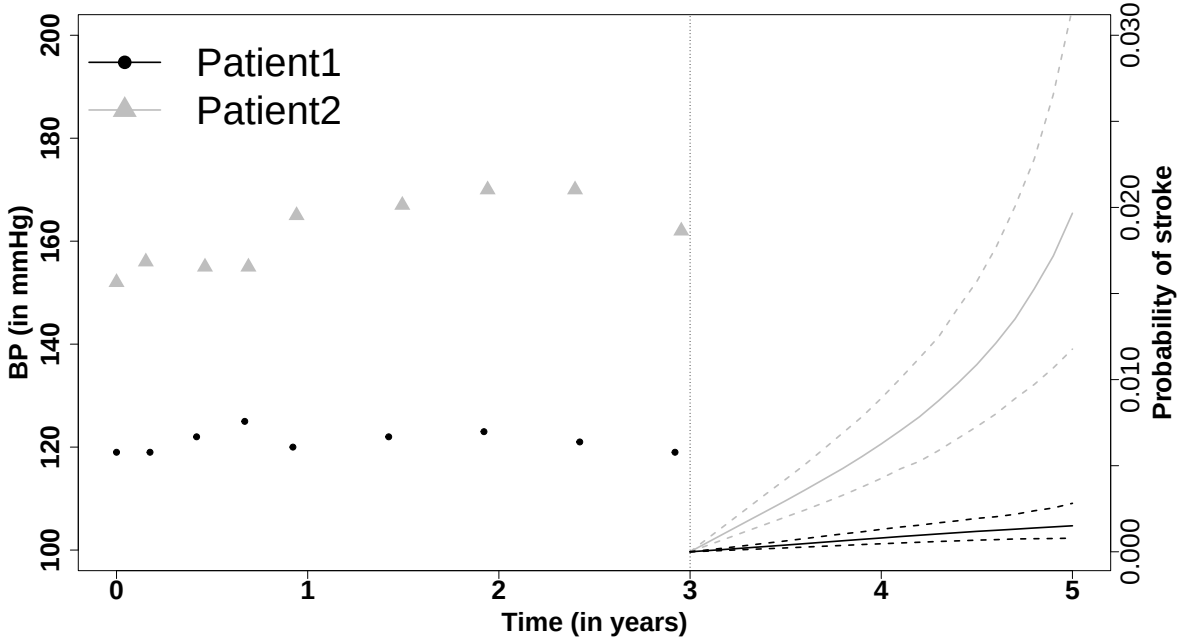
$$= \frac{\int \left[\int_s^{s+t} \exp\left(-\sum_{c=1}^{2} \Lambda_{ic}(u|b_i,\omega_i)\right) \lambda_{ik}(u|b_i,\omega_i) du \right] f(\mathcal{Y}_i(s)|b_i,\omega_i) f(b_i) f(\omega_i) db_i d\omega_i}{\int \exp\left(-\sum_{c=1}^{2} \Lambda_{ic}(s|b_i,\omega_i)\right) f(\mathcal{Y}_i(s)|b_i,\omega_i) f(b_i) f(\omega_i) db_i d\omega_i}$$

- $\mathcal{Y}_i(s)$: marker's trajectory until time s

For brievety $\widehat{\theta}$ is removed from formula

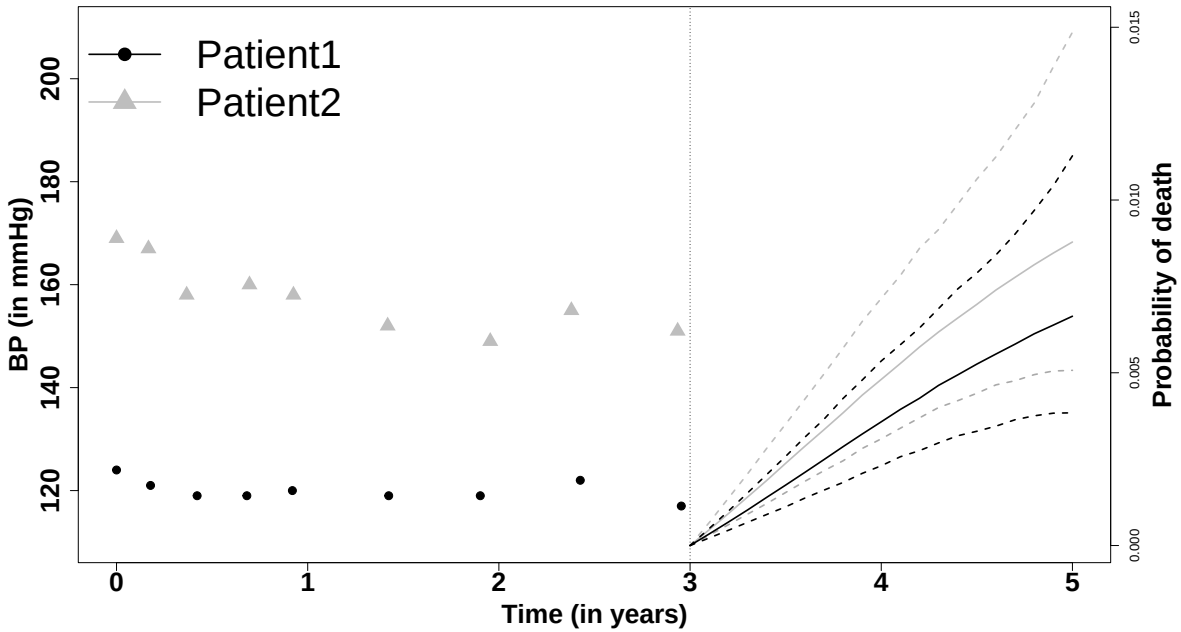
Individual Predictions: BP's current value effect on stroke risk

- **Patient 1**: a 47-year-old Asian man with a low level of blood pressure, censored at 4.40 years,
- **Patient 2**: a 47-year-old Asian man with a higher level of blood pressure with a recurrent stroke at 3.30 years



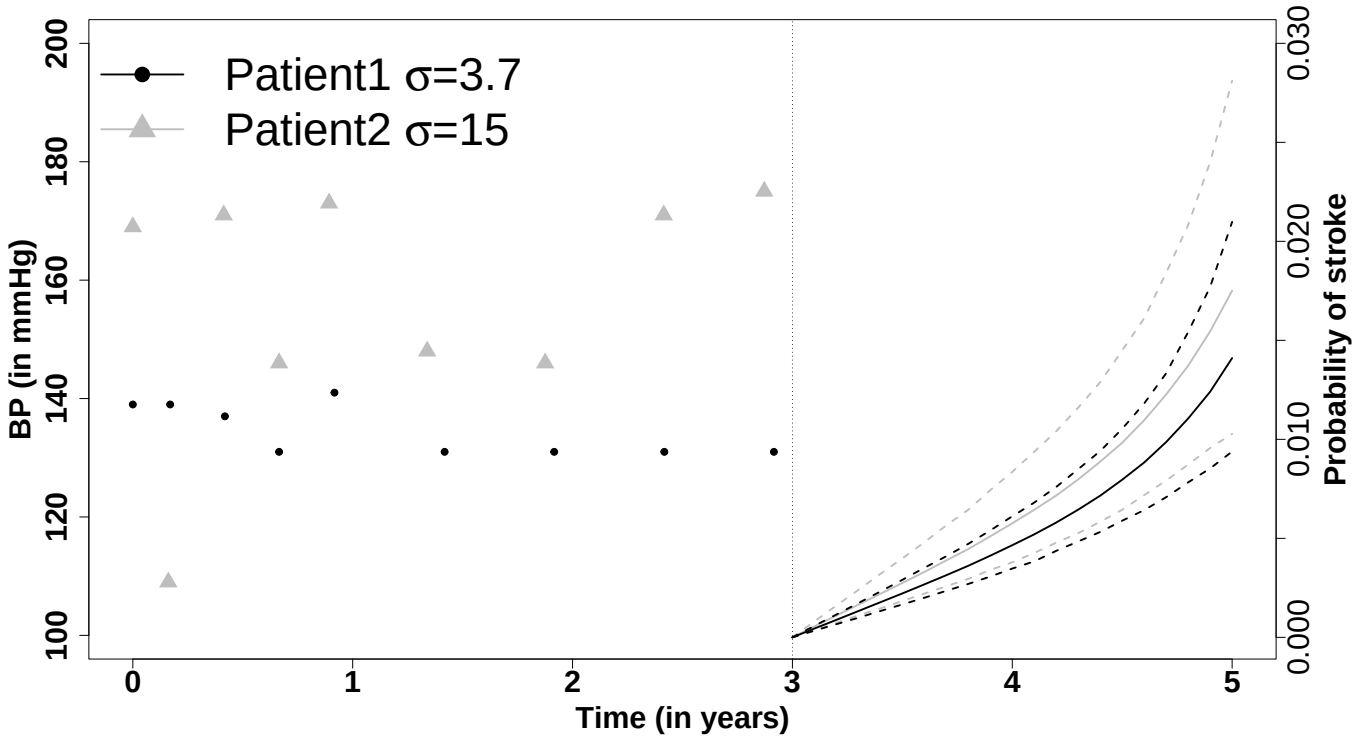
Individual Predictions: BP's current value effect on death risk

- **Patient 1**: a 66-year-old non Asian man with a low level of blood pressure censored at 3.37 years,
- **Patient 2**: a 67-year-old non-Asian man with a higher level of blood pressure who died at 3.56 years.



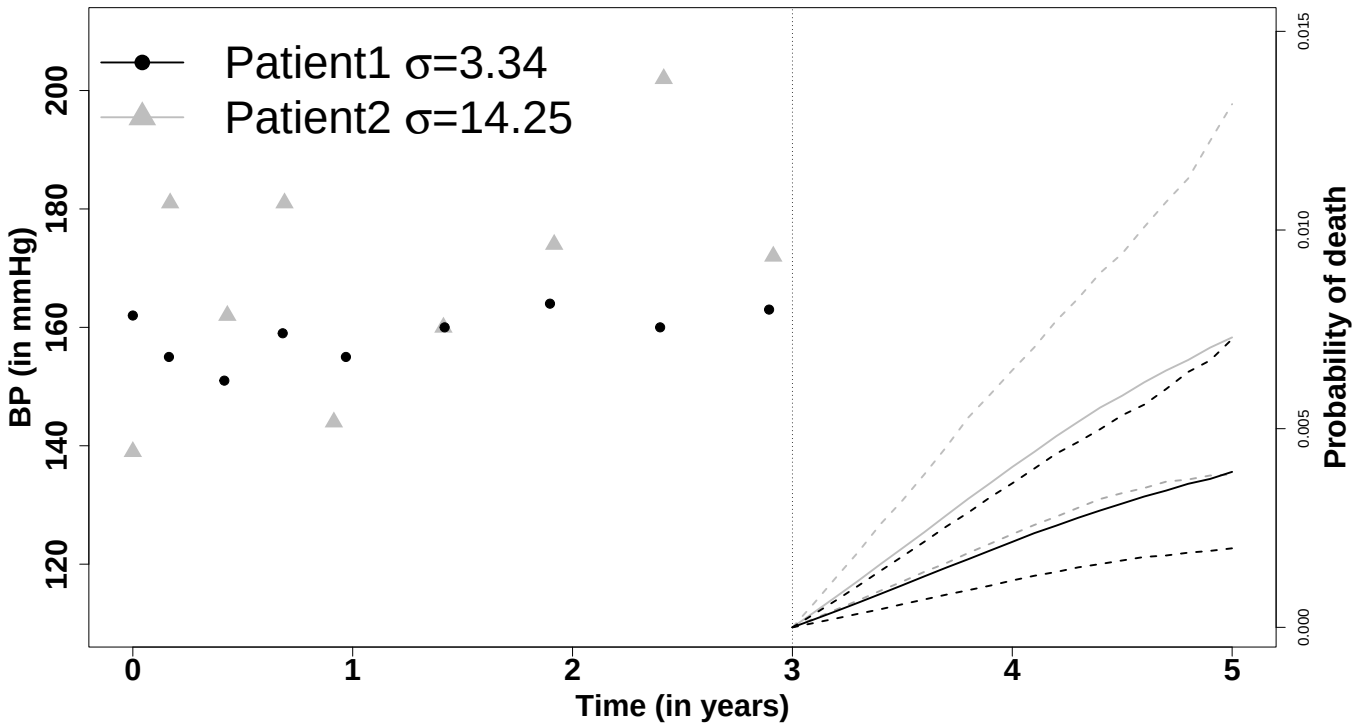
Individual Predictions: BP's variability effect on stroke risk

- **Patient 1**: a 61-year-old Asian man with a low variability, censored at 4.42 years,
- **Patient 2**:a 61-year-old Asian man with a higher variability with a recurrent stroke at 4.24 years,



Individual Predictions: BP's variability effect on death risk

- **Patient 1**: a 70-year-old non-Asian woman with a low variability censored at 4.41 years,
- **Patient 2**: a 68-year-old non-Asian woman with a higher variability who died at 4.27 years.



Additional results

- Treatment arm (3025 patients - 292 strokes - 219 deaths)

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	p-value
Longitudinal submodel - BP			
μ_σ	2.26	0.009	< 0.001
τ_σ	0.37	0.008	< 0.001

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	HR	p-val
Survival submodel 1 - Stroke				
$\alpha_{\sigma 1}$	0.014	0.024	1.01	0.560
α_{11} *(cv)*	0.012	0.005	1.01	0.016
Survival submodel 2 - Death				
$\alpha_{\sigma 2}$	0.076	0.026	1.08	0.003
α_{12} *(cv)*	-0.007	0.006	0.99	0.243

Additional results

- Treatment arm (3025 patients - 292 strokes - 219 deaths)

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	p-value
Longitudinal submodel - BP			
μ_σ	2.26	0.009	< 0.001
τ_σ	0.37	0.008	< 0.001

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	HR	p-val
Survival submodel 1 - Stroke				
$\alpha_{\sigma 1}$	0.014	0.024	1.01	0.560
α_{11} *(cv)*	0.012	0.005	1.01	0.016
Survival submodel 2 - Death				
$\alpha_{\sigma 2}$	0.076	0.026	1.08	0.003
α_{12} *(cv)*	-0.007	0.006	0.99	0.243

- Complete data (6057 patients - 698 strokes - 432 deaths)

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	p-value
Longitudinal submodel - BP			
μ_σ	2.26	0.006	< 0.001
τ_σ	0.37	0.006	< 0.001

Variable	$\hat{\theta}$	SE($\hat{\theta}$)	HR	p-val
Survival submodel 1 - Stroke				
$\alpha_{\sigma 1}$	-0.008	0.017	0.99	0.653
α_{11} *(cv)*	0.019	0.003	1.02	< 0.001
Survival submodel 2 - Death				
$\alpha_{\sigma 2}$	0.071	0.021	1.07	< 0.001
α_{12} *(cv)*	0.001	0.009	1.001	0.876

Conclusion

- Epidemiological results:
  - Current value of BP: risk factor of stroke
  - BP variability: risk factor of death
  - **Perspectives:**
    - Study others events
- Methodological results:
  - Extension of Barrett et al's model[1]
  - Estimation under frequentist paradigm
  - Implementation of goodness-of-fit criteria
  - R-Package on development available: `https://github.com/LeonieCourcoul/FlexVarJM`
  - **Perspectives:**
    - Time-dependent variance

[1] Barrett et al, Statistics in medicine, 38(10):1855–1868, 2019

References

[1] H de Courson, K Leffondre, and C. Tzourio. Blood pressure variability and risk of cardiovascular event : is it appropriate to use the future for pverticting the present ? European heart journal. 39(47):4220–4220, 2018.

[2] J.K. Barrett, R. Huille, R. Parker, Y. Yano, and M. Griswold. Estimating the association between blood pressure variability and cardiovascular disease : An application using the aric study. Statistics in medicine, 38(10):1855–1868, 2019.

[3] PROGRESS Management Committee. Blood pressure lowering for the secondary prevention of stroke: rationale and design for PROGRESS. Journal of Hypertension Supplement. 1996 Sep;14(2):S41-5.

Acknowledgements

This work was funded by the French National Research Agency (grant ANR-21-CE36-0013-01 for the project "Joint Models for Epidemiology and Clinical research").

This PhD program is supported within the framework of the PIA3 (Investment for the Future).
Project reference: 17-EURE-0019.

Thank you for your attention !
Any questions?