

Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment

Loïc Ferrer^{1,†}, Hein Putter² and Cécile Proust-Lima¹

⁽¹⁾Biostatistics Team, INSERM U1219, University of Bordeaux, Bordeaux, France

⁽²⁾Department of Medical Statistics and Bioinformatics, Leiden University Medical Center,
Leiden, The Netherlands

^(†) *email*: loic.ferrer@inserm.fr

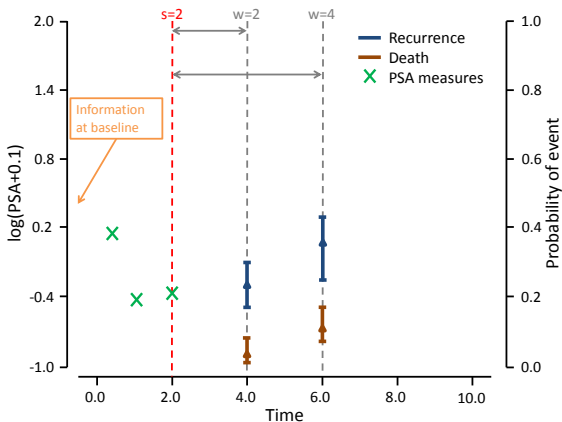
Journées GDR/SFB

ISPED, Université de Bordeaux, France

October 05, 2017

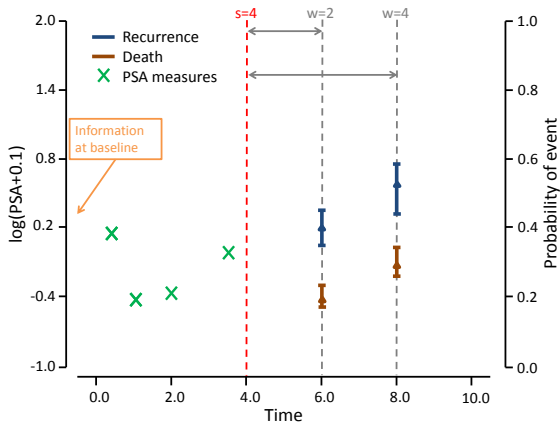
Dynamic predictions in prostate cancer

- Cumulative incidences of two causes of event computed from the landmark time $s = 2$ for two horizons $w = 2$ and $w = 4$



Dynamic predictions in prostate cancer

- Cumulative incidences of two causes of event computed from the landmark time $s = 4$ for two horizons $w = 2$ and $w = 4$



Quantity of interest

- ▶ Landmark specific cumulative incidence of the event of cause k between the times s and $s + w$ for the new subject \star

$$\pi_{\star}^k(s, w; \theta) = \Pr(s < T_{\star} \leq s + w, \delta_{\star} = k | T_{\star} > s, \mathcal{Y}_{\star}(s), X_{\star}; \theta)$$

- ▶ $T_{\star} = \min(\{T_{\star,k}\}_k)$ the earliest event time
- ▶ $T_{\star}^{\dagger} = \min(T_{\star}, C_{\star})$ the observed event time, C_{\star} the censoring time
- ▶ $\Delta_{\star} = \delta_{\star} \cdot \mathbb{1}\{T_{\star} \leq C_{\star}\}$, where $\delta_{\star} = k$ is the cause of event
- ▶ $\mathcal{Y}_{\star}(s)$ and X_{\star} the individual information collected until s

Objectives

- ▶ When $\mathcal{Y}_*(s)$ includes internal time-dependent covariates \Rightarrow needs to handle the dependency between \mathcal{Y}_* and (T_*, δ_*)
 - ▶ Joint modelling
 - ▶ Landmarking
 - Cause-specific PH models
 - Pseudo-observations approach
- \Rightarrow Validation of the estimators from the joint and landmark models and their uncertainty
- \Rightarrow Comparison of the joint and landmark models in terms of prediction accuracy and robustness to the model hypotheses

Joint model

- ▶ Information \mathcal{O} used in the learning step: Whole information

$$\mathcal{O} = \{(T_i^\dagger, \Delta_i, \mathcal{Y}_i(T_i^\dagger), X_i)\}_{i=1, \dots, N}$$

- ▶ Model formulation

$$\begin{cases} Y_i(t) &= Y_i^*(t) + \epsilon_i(t) \\ &= X_i^L(t)^\top \beta + Z_i(t)^\top \mathbf{b}_i + \epsilon_i(t) \\ \lambda_i^k(t) &= \lambda_{k,0}(t) \exp \{X_{k,i}^E{}^\top \gamma_k + \mathbf{W}_{k,i}(t|\mathbf{b}_i; \beta, \mathbf{D})^\top \eta_k\} \end{cases}$$

- ▶ $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$; $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I_{n_i})$; $\lambda_{0,k}(\cdot)$ the param. baseline hazard
- ▶ Cumulative incidence estimator for new subject \star

$$\hat{\pi}_\star^k(s, w; \hat{\theta}) = \int_{\mathbb{R}^q} \hat{\pi}_\star^k(s, w | \mathbf{b}_\star; \hat{\theta}) f(\mathbf{b}_\star | T_\star > s, \mathcal{Y}_\star(s), X_\star; \hat{\theta}) d\mathbf{b}_\star$$

- ▶ $f(\cdot)$ a probability density function

Landmark cause-specific PH models

- Information \mathcal{O} used in the learning step: Reduced information

$$\mathcal{O} = \{(\min(T_i^\dagger, s + w), \Delta_i \cdot \mathbb{1}\{s < T_i \leq s + w\}, \mathcal{Y}_i(s), X_i) | T_i^\dagger > s\}_{i=1, \dots, N_s^\dagger}$$

- Model formulation

$$\lambda_i^k(t) = \lambda_{k,0}(t) \exp \{X_{k,i}^E \top \gamma_k + W_{k,i}(s) \top \eta_k\}$$

Naive model

- $\lambda_{k,0}(\cdot)$ unspecified baseline hazard
- $W_{k,i}(s) = Y_i(t_{in_i}(s))$
 - last observed marker value
 - with $t_{in_i}(s) = \max(t_{ij} | t_{ij} < s)$

Two-stage model

- $\lambda_{k,0}(\cdot)$ unspecified baseline hazard
- $W_{k,i}(s) = \widehat{W}_{k,i}(s; \widehat{\beta}, \widehat{D}, \widehat{b}_i)$
 - marker dynamics predicted at s
 - with $\widehat{\beta}, \widehat{D}$ from a LMM
 - $\widehat{b}_i = \mathbb{E}(b_i | \mathcal{Y}_i(s))$

- Cumulative incidence estimator for new subject \star

$$\widehat{\pi}_\star^k(s, w; \widehat{\theta})$$

$$\widehat{\pi}_\star^k(s, w; \widehat{\theta}, \widehat{b}_\star)$$

with $\widehat{b}_\star = \mathbb{E}(b_\star | \mathcal{Y}_\star(s))$

Two-stage landmark pseudo-values approach

- Information \mathcal{O} used in the learning step: Reduced information

$$\mathcal{O} = \{(T_i^\dagger, \Delta_i, \mathcal{Y}_i(s), X_i) | T_i^\dagger > s\}_{i=1, \dots, N_s^\dagger}$$

- Model formulation

$$g \left[E\{\widehat{\mu}_{i,s,w}^k | T_i^\dagger > s\} \right] = \gamma_0 + X_{k,i}^E \top \gamma_1 + \widehat{W}_{k,i}(s; \widehat{\beta}, \widehat{D}, \widehat{b}_i) \top \eta$$

- where the dynamic pseudo-observation is

$$\widehat{\mu}_{i,s,w}^k = N_s \widehat{F}^k(s, w) - (N_s - 1) \widehat{F}_{(-i)}^k(s, w)$$

- $N_s = \sum_i \mathbb{1}\{T_i^* > s\}$, $\widehat{F}^k(s, w)$ non-par. estimator of $\pi^k(s, w)$

- Cumulative incidence estimator for new subject \star ($g(x) = \text{cloglog}(x)$)

$$\widehat{\pi}_\star^k(s, w; \widehat{\theta}, \widehat{b}_\star) = 1 - \exp \left[- \exp \{ \widehat{\gamma}_0 + X_{k,\star}^E \top \widehat{\gamma}_1 + \widehat{W}_{k,\star}(s; \widehat{\beta}, \widehat{D}, \widehat{b}_\star) \top \widehat{\eta} \} \right]$$

Prediction accuracy / Robustness

- ▶ Comparison of the joint and landmark models in terms of
 - ▶ Prediction error
 - ▶ Discriminatory power
 - ▶ Robustness to the model hypotheses
- ▶ 4 scenarii
 - ▶ Well-specification of the joint model
 - ▶ Misspecification of the dependence function
 - ▶ Violation of the PH assumption
 - ▶ Misspecification of the longitudinal trend of the marker

Tools for comparison

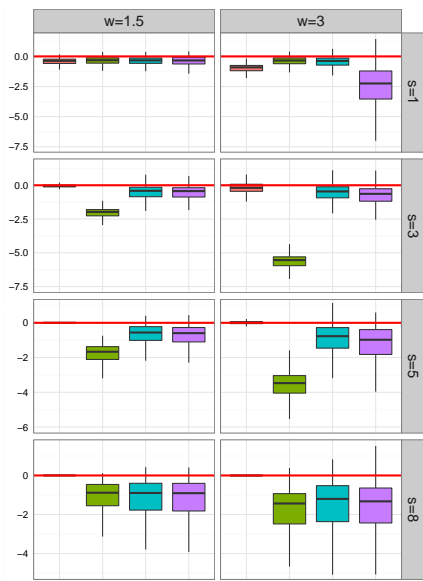
- ▶ For each scenario, data generated from JM
 - For each new subject $\star = 1, \dots, N_s^{\text{new}}$ ($N_0^{\text{new}} = 500$)

$$\pi_{\star}^k(s, w; \theta) = \int_{\mathbb{R}^q} \pi_{\star}^k(s, w | b_{\star}; \theta) f(b_{\star} | T_{\star} > s, \mathcal{Y}_{\star}(s), X_{\star}; \theta) db_{\star}$$

- ▶ For each replicate $r = 1, \dots, 500$ and each $\star = 1, \dots, N_s^{\text{new}}$
 - $\widehat{\pi}_{\star,r}^k(s, w; \widehat{\theta})$ computed using JM, Naive-LM-PH, 2s-LM-PH and 2s-LM-PV approaches
- ▶ Comparison between the prediction models using boxplots of the differences of Mean Square Error of Prediction (MSEP) over the 500 replicates, where

$$\text{MSEP}_r^k(s, w) = \frac{1}{N_s^{\text{new}}} \sum_{\star=1}^{N_s^{\text{new}}} \left(\pi_{\star}^k(s, w; \theta) - \widehat{\pi}_{\star,r}^k(s, w; \widehat{\theta}) \right)^2 \times 1000$$

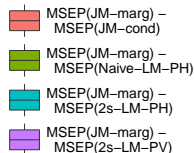
Well-specification of the joint model : Recurrence



- ▶ Generation: JM with $W_{k,i}(t|b_i; \beta, D) = (Y_i^*(t), \partial Y_i^*(t)/\partial t)^\top$
- ▶ Estimation: Based on the model of generation

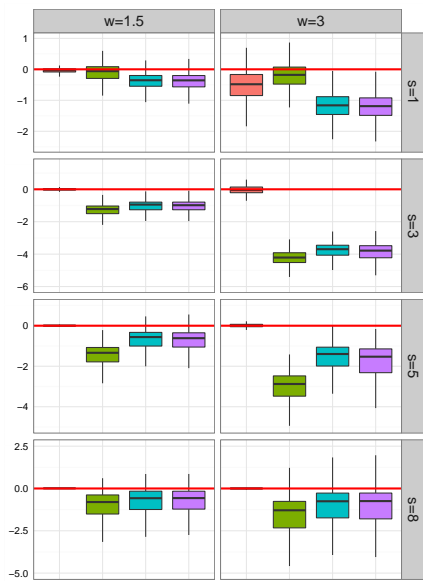
▶ $\bar{N} = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 947 \\ 699 \\ 482 \\ 253 \end{pmatrix}$; $\bar{D}_1 = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 1.5 & 3 \\ 66 & 123 \\ 52 & 86 \\ 29 & 43 \\ 6 & 9 \end{pmatrix}$

- ▶ Legend:



better than JM
 similar to JM
 less than JM

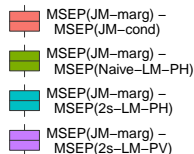
Misspecification of the dependence function: Rec.



- ▶ Generation: JM with $W_{k,i}(t|b_i; \beta, D) = (Y_i^*(t), \partial Y_i^*(t)/\partial t)^\top$
- ▶ Estimation: JM, 2s-LM-PH and 2s-LM-PV using $W_{k,i}(t|b_i; \beta, D) = Y_i^*(t)$

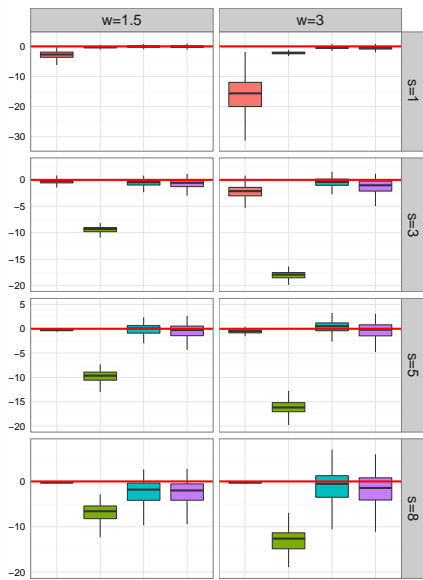
▶ $\bar{N} = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 947 \\ 699 \\ 482 \\ 253 \end{pmatrix}; \bar{D}_1 = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 1.5 & 3 \\ 66 & 123 \\ 52 & 86 \\ 29 & 43 \\ 6 & 9 \end{pmatrix}$

- ▶ Legend:



better than JM
~~similar to JM~~
 less than JM

Violation of the PH assumption: Rec.



► Generation: JM with interaction of η_k with $\log(1+t)$

► Estimation: By neglecting $\log(1+t)$

►
$$\bar{N} = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 965 \\ 689 \\ 422 \\ 199 \end{pmatrix}; \quad \bar{D}_1 = \begin{matrix} 1 & 1.5 & 3 \\ 3 & 76 & 183 \\ 5 & 101 & 162 \\ 8 & 48 & 71 \\ & 9 & 14 \end{matrix}$$

► Legend:

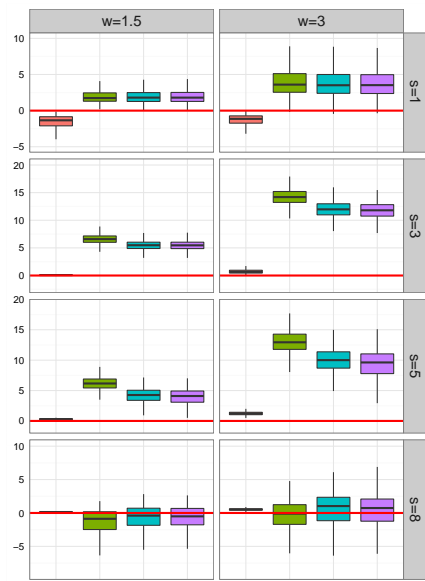
- MSEP(JM-marg) – MSEP(JM-cond)
- MSEP(JM-marg) – MSEP(Naive-LM-PH)
- MSEP(JM-marg) – MSEP(2s-LM-PH)
- MSEP(JM-marg) – MSEP(2s-LM-PV)

better than JM

similar to JM

less than JM

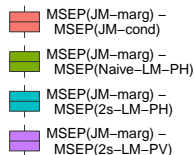
Misspecification of the longit. marker trajectory: Rec.



- ▶ Generation: JM with biphasic trajectory of the marker
- ▶ Estimation: Linear trajectory of the marker for JM, 2s-LM-PH and 2s-LM-PV

$$\bar{N} = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 972 \\ 730 \\ 484 \\ 238 \end{pmatrix}; \quad \bar{D}_1 = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 1.5 & 3 \\ 46 & 118 \\ 72 & 125 \\ 45 & 67 \\ 10 & 14 \end{pmatrix}$$

- ▶ Legend:



better than JM

~~similar to JM~~

less than JM

Discussion (1/2)

- ▶ Accuracy/Robustness
 - ▶ JM better than LM when well-specified
 - ▶ JM better than LM when misspecification of the association function but much worse performances for both in terms of PE
 - ▶ JM & LM comparable when **strong** violation of the PH assumption
 - ▶ LM better than JM when **strong** misspecification of the longitudinal marker trajectory
- ▶ Discriminatory power
 - ▶ AUC examinations
 - ▶ yet sometimes to a lesser extent probably explained by the lesser sensitivity of AUC and the possibly preserved discriminatory power in the presence of worse calibration

Discussion (2/2)

- ▶ Validation of the estimators and their variability
 - ▶ Good individual coverage rates of cumulative incidences for two-stage landmark CS PH model and joint model
 - ▶ Individual relative bias → JM unbiased, LM biased in earlier landmark times
 - ▶ Length of the individual 95% confidence intervals → estimator more efficient using JM than LM
- ▶ Implementation in R
 - ▶ JM with extensions for joint modelling
 - ▶ `mstate` for cause-specific modelling
 - ▶ `pseudo` and `geepack` for dynamic pseudo-values approach

References

- ▶ Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F., and Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1):102–113.
- ▶ Ferrer, L., Putter, H., & Proust-Lima, C. (available on request). Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment.
- ▶ Maziarz, M., Heagerty, P., Cai, T., & Zheng, Y. (2016). On longitudinal prediction with time-to-event outcome: Comparison of modeling options. *Biometrics*.
- ▶ Nicolaie, M. A., van Houwelingen, J. C., de Witte, T. M., and Putter, H. (2013). Dynamic Pseudo-Observations: A Robust Approach to Dynamic Prediction in Competing Risks. *Biometrics*, 69(4):1043–1052.
- ▶ Proust-Lima, C., & Blanche, P. (2015). Dynamic Predictions. In *WileyStatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd.
- ▶ Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–29.
- ▶ van Houwelingen, H. and Putter, H. (2011). *Dynamic Prediction in Clinical Survival Analysis*. Monographs on Statistics & Applied Probab #123. Chapman & Hall/CRC, London.

Landmark cause-specific PH models

- Linear mixed model based on the longitudinal data collected until the landmark time s , $\mathcal{Y}_i(s)$

$$\begin{aligned} Y_i(u|s, \mathbf{b}_i) &= Y_i^{*(s)}(u|\mathbf{b}_i) + \epsilon_i^{(s)}(u) \\ &= \mathbf{X}_i^L(u)^\top \beta^{(s)} + \mathbf{Z}_i(u)^\top \mathbf{b}_i^{(s)} + \epsilon_i^{(s)}(u), \quad u \leq s \end{aligned}$$

- One uses the marker dynamics predicted at the landmark time s

$$\lambda_i^k(t|s, w, \widehat{\mathbf{b}}_i^{(s)}) = \lambda_{0,k}(t|s, w) \exp \left\{ \mathbf{X}_{k,i}^{E\top} \gamma_k^{(s,w)} + m_i(s|\widehat{\mathbf{b}}_i^{(s)}, \widehat{\boldsymbol{\theta}}_Y^{(s)})^\top \boldsymbol{\eta}_k^{(s,w)} \right\},$$

$$s \leq t \leq s + w$$

- Predicted conditional cumulative incidence of cause k

$$\widehat{\pi}_i^k(s, w|\widehat{\mathbf{b}}_i^{(s)}; \widehat{\boldsymbol{\theta}}^{(s,w)}) = \Pr(s < T_i^* \leq s + w, \delta_i = k | T_i^* > s, \mathbf{X}_i, \widehat{\mathbf{b}}_i^{(s)}; \widehat{\boldsymbol{\theta}}^{(s,w)})$$

Comparison through several cases: simulation study

- ▶ For each subject $i = 1, \dots, 1000$ of the 500 replicates
- ▶ Data generated according to the joint model

$$\begin{cases} Y_i(t) &= Y_i^*(t) + \epsilon_i(t) \\ &= (\beta_0 + \beta_1 X_{1,i} + b_{i0}) + (\beta_2 + \beta_3 X_{1,i} + b_{i1})t + \epsilon_i(t) \\ \lambda_i^k(t) &= \lambda_{0,k}(t) \exp \left\{ \gamma_k X_{2,i} + \eta_{1,k} Y_i^*(t) + \eta_{2,k} \frac{\delta Y_i^*(t)}{\delta t} \right\} \end{cases}$$

- ▶ $\log(\lambda_{0,k}(t))$ combination of cubic B-splines with one internal knot
 - ▶ k the cause of event (Recurrence ; Death)
 - ▶ Predictions on the same subjects $j = 1, \dots, N_s$, where N_s is the number of subjects at risk at time s ($N(0) = 500$)
- ⇒ Comparison of the generated and estimated cumulative incidences using differences of MSE of prediction

$$\text{MSE}_{k,r}(s, w) = \sum_{i=1}^{N_s} \frac{1}{N_s} \left\{ \pi_i^k(s, w; \theta) - \hat{\pi}_i^{k,r}(s, w; \hat{\theta}) \right\}^2 \times 1000 \quad \text{for the } r^{\text{th}} \text{ replicate}$$

Misspecification of the dependence function

- ▶ Data generated using the joint model

$$\left\{ \begin{array}{l} Y_i(t) = Y_i^*(t) + \epsilon_i(t) \\ \quad = (\beta_0 + \beta_1 X_{1,i} + b_{i0}) + (\beta_2 + \beta_3 X_{1,i} + b_{i1}) t + \epsilon_i(t) \\ \lambda_i^k(t) = \lambda_{0,k}(t) \exp \left\{ \gamma_k X_{2,i} + \eta_{1,k} Y_i^*(t) + \eta_{2,k} \frac{\delta Y_i^*(t)}{\delta t} \right\} \end{array} \right.$$

- ▶ Models estimated by neglecting $\frac{\delta Y_i^*(t)}{\delta t}$

Violation of the PH assumption

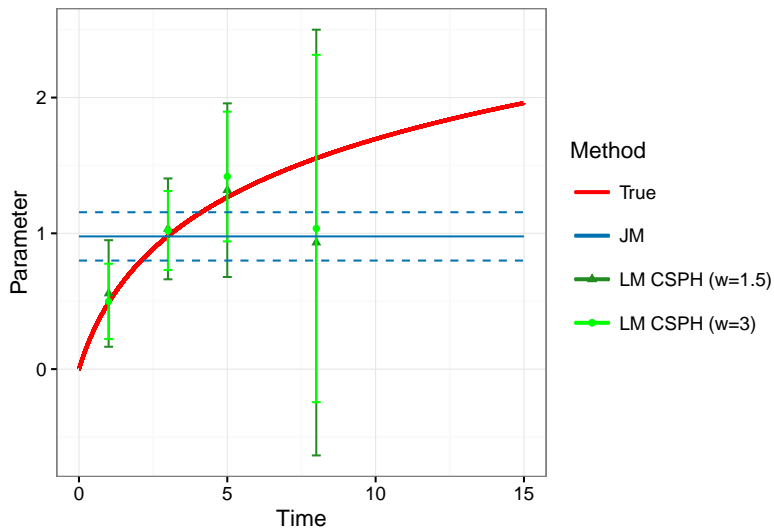
- ▶ Data generated using the joint model

$$\left\{ \begin{array}{l} Y_i(t) = Y_i^*(t) + \epsilon_i(t) \\ \quad = (\beta_0 + \beta_1 X_{1,i} + b_{i0}) + (\beta_2 + \beta_3 X_{1,i} + b_{i1}) t + \epsilon_i(t) \\ \lambda_i^k(t) = \lambda_{0,k}(t) \exp \left\{ \gamma_k X_{2,i} + \eta_{1,k} \log(1+t) Y_i^*(t) + \eta_{2,k} \log(1+t) \frac{\delta Y_i^*(t)}{\delta t} \right\} \end{array} \right.$$

- ▶ Models estimated by neglecting $\log(1+t)$

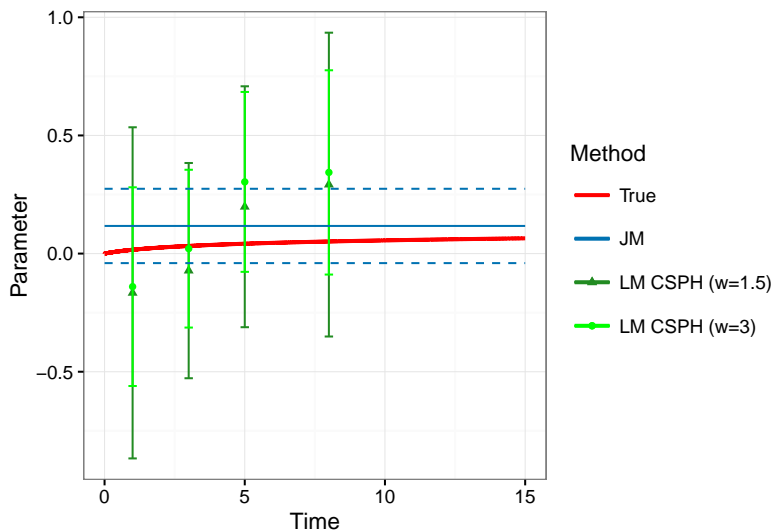
Comparison between $\eta_{1, \text{Rec.}} \times \log(1 + t)$ and $\hat{\eta}_{1, \text{Rec.}}$

1st replicate



Comparison between $\eta_{1,\text{Death}} \times \log(1 + t)$ and $\hat{\eta}_{1,\text{Death}}$

1st replicate



Misspecification of the longitudinal trajectory

▶ Data generated using the joint model

▶ Longitudinal sub-model

- **Case A (biphasic shape on all data)**

$$\begin{aligned} Y_i(t) &= Y_i^*(t) + \epsilon_i(t) \\ &= (\beta_0 + \beta_{0,X}X_{1,i} + b_{i0}) + (\beta_1 + \beta_{1,X}X_{1,i} + b_{i1})f_1(t) + \\ &\quad (\beta_2 + \beta_{2,X}X_{1,i} + b_{i2})t + \epsilon_i(t) \end{aligned}$$

- **Case B (splines on post-nadir data)**

$$\begin{aligned} Y_i(t) &= (\beta_0 + \beta_{0,X}X_{1,i} + b_{i0}) + \\ &\quad \sum_{p=1}^3 (\beta_p + \beta_{p,X}X_{1,i} + b_{ip})B(t, \kappa_p) + \epsilon_i(t) \end{aligned}$$

▶ Survival sub-model

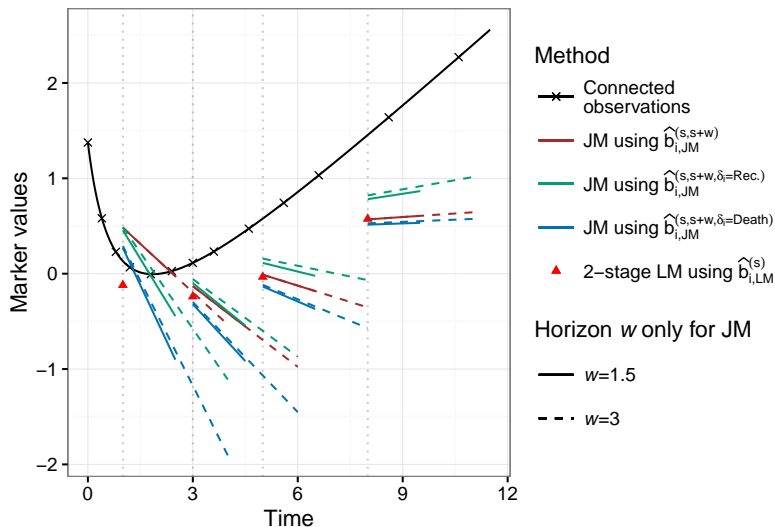
$$\lambda_i^k(t) = \lambda_{0,k}(t) \exp \left\{ \gamma_k X_{2,i} + \eta_{1,k} Y_i^*(t) + \eta_{2,k} \frac{\delta Y_i^*(t)}{\delta t} \right\}$$

▶ Models estimated by considering

$$Y_i(t) = (\beta_0 + \beta_1 X_{1,i} + b_{i0}) + (\beta_2 + \beta_3 X_{1,i} + b_{i1})t + \epsilon_i(t)$$

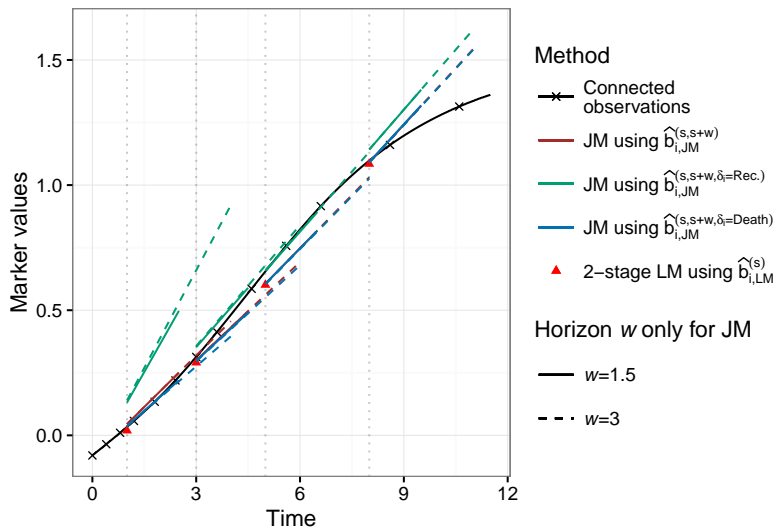
Comparison between used predictions – Exple case A

1st simulation, ID with $b_i = 0, X_i = \bar{X}$

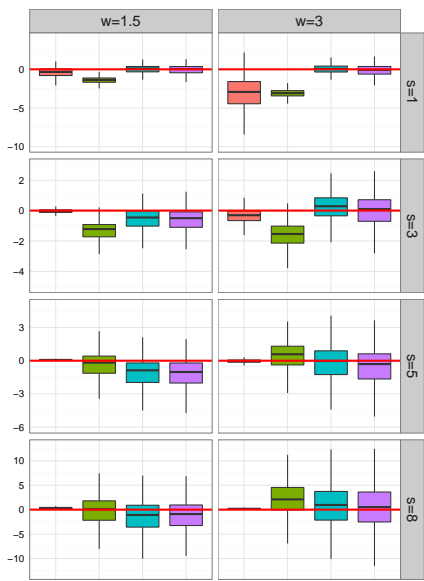


Comparison between used predictions – Exple case B

1st simulation, ID with $b_i = 0, X_i = \bar{X}$



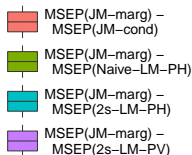
Misspecification of the longit. marker trajectory: Rec.



- ▶ Generation: JM with trajectory of the marker modelled by splines
- ▶ Estimation: Linear trajectory of the marker for JM, 2s-LM-PH and 2s-LM-PV

$$\bar{N} = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 870 \\ 568 \\ 365 \\ 172 \end{pmatrix}; \quad \bar{D}_1 = \begin{matrix} 1 \\ 3 \\ 5 \\ 8 \end{matrix} \begin{pmatrix} 1.5 & 3 \\ 125 & 201 \\ 64 & 99 \\ 29 & 42 \\ 6 & 10 \end{pmatrix}$$

- ▶ Legend:



better than JM

similar to JM

less than JM

First objective

- ▶ Validation of the proposed estimators and their uncertainty
- ⇒ Simulation study with focus on the bias and coverage rates
 - ▶ Data generated according to a joint model
 - ⇒ For each new subject $\star = 1, \dots, N_s^{\text{new}}$ ($N_0^{\text{new}} = 500$)

$$\pi_{\star}^k(s, w; \theta) = \int_{\mathbb{R}^q} \pi_{\star}^k(s, w | b_{\star}; \theta) f(b_{\star} | T_{\star} > s, \mathcal{Y}_{\star}(s), X_{\star}; \theta) db_{\star}$$

- ▶ For each replicate $r = 1, \dots, 500$ and each $\star = 1, \dots, N_s^{\text{new}}$
 - $\widehat{\pi}_{\star, r}^k(s, w; \widehat{\theta})$ computed using JM, Naive-LM-PH, 2s-LM-PH and 2s-LM-PV approaches

Validation of the estimators and their uncertainty

- ▶ Interest in the distribution over the subjects of the relative bias (RB) of $\widehat{\pi}_*^k(s, w; \widehat{\theta})$

$$\text{RB}_*^k(s, w) = \frac{1}{R} \sum_{r=1}^R \frac{\widehat{\pi}_{*,r}^k(s, w; \widehat{\theta}) - \pi_*^k(s, w; \theta)}{\pi_*^k(s, w; \theta)}$$

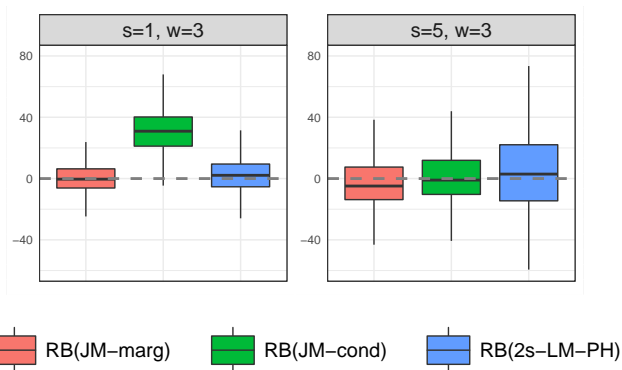
- ▶ Interest in the distribution over the subjects of the coverage rates (COV)

$$\text{COV}_*^k(s, w) = \frac{1}{R} \sum_{r=1}^R \left(\pi_*^k(s, w; \theta) \in \left[q_{0.025} \{ \widetilde{\pi}_{*,r}^{k,(l)}(s, w; \widetilde{\theta}^{(l)}); l = 1, \dots, L \}; \right. \right. \\ \left. \left. q_{0.975} \{ \widetilde{\pi}_{*,r}^{k,(l)}(s, w; \widetilde{\theta}^{(l)}); l = 1, \dots, L \} \right] \right)$$

- ▶ Results obtained for $R = 500$ replicates, $N_0^{\text{new}} = 500$

Relative bias

- Distribution over the subjects of the relative bias (RB) of $\hat{\pi}_*^k(s, w; \hat{\theta})$



Variability of the predictions

- ▶ Essential to compute valid 95% confidence intervals
 - ▶ Joint model
 - ▶ Existing method proposed by D. Rizopoulos (JM package)
 - ▶ Naive landmark model
 - ▶ Existing analytic formula
 - ▶ Two-stage landmark models
 - ▶ No available method
- Validation of the estimators using coverage rates of $\pi_{\star}^k(s, w; \theta)$

Confidence intervals of the predictions from JM

Parametric bootstrap: for each $l = 1, \dots, L$

1. Generate $\tilde{\theta}^{(l)} \sim \mathcal{N}(\hat{\theta}, \widehat{V}(\hat{\theta}))$
2. 2 solutions: compute $\tilde{\pi}_*^{k,(l)}(s, w; \tilde{\theta}^{(l)})$ using
 - a. either the conditional probability

$$\tilde{\pi}_*^{k,(l)}(s, w; \tilde{\theta}^{(l)}, \widehat{b}_*^{(l)})$$

- where $\widehat{b}_*^{(l)} = \arg \max_b f(b|T_* > s, \mathcal{Y}_*(s), X_*; \tilde{\theta}^{(l)})$

- b. or the marginal probability

$$\int_{\mathbb{R}^q} \tilde{\pi}_*^{k,(l)}(s, w|b_*; \tilde{\theta}^{(l)}) db_*$$

\implies Deduce 95% confidence interval

$$\left[q_{0.025} \{ \tilde{\pi}_*^{k,(l)}(s, w; \tilde{\theta}^{(l)}); l = 1, \dots, L \} ; q_{0.975} \{ \tilde{\pi}_*^{k,(l)}(s, w; \tilde{\theta}^{(l)}); l = 1, \dots, L \} \right]$$

Confidence intervals of the predictions from 2s-LM-PH

Parametric bootstrap [1.] and perturbation-resampling [2.]

For each $l = 1, \dots, L$

1. Generate $\tilde{\theta}^{(l)} \sim \mathcal{N}(\hat{\theta}, \hat{V}(\hat{\theta}))$
 - ▶ $\hat{b}_i^{(l)} = \mathbb{E}(b_i | \mathcal{Y}_i(s); \tilde{\theta}^{(l)})$ and $\hat{b}_\star^{(l)} = \mathbb{E}(b_\star | \mathcal{Y}_\star(s); \tilde{\theta}^{(l)})$
2. a. Perturb the observed time-to-event data: for $i = 1, \dots, N_s$
 - $\nu_i \sim 4 \cdot \text{Beta}(\frac{1}{2}, \frac{3}{2})$
 - $\tilde{T}_i^* = \nu_i \cdot T_i^*$; $\tilde{\Delta}_i = \nu_i \cdot \Delta_i$

b. Update the Breslow-estimator $\tilde{\Lambda}_{k,0}^{(l)}(\cdot) = f(\tilde{\theta}^{(l)}, \{\hat{b}_i^{(l)}, \tilde{T}_i^*, \tilde{\Delta}_i\}_{i=1, \dots, N_s})$
3. Compute $\tilde{\pi}_\star^{k,(l)}(s, w; \tilde{\theta}^{(l)}) = f(\tilde{\Lambda}_{k,0}^{(l)}(\cdot), \tilde{\theta}^{(l)}, \hat{b}_\star^{(l)})$

⇒ Deduce the 95% confidence interval

$$\left[q_{0.025} \{ \tilde{\pi}_\star^{k,(l)}(s, w; \tilde{\theta}^{(l)}); l = 1, \dots, L \} ; q_{0.975} \{ \tilde{\pi}_\star^{k,(l)}(s, w; \tilde{\theta}^{(l)}); l = 1, \dots, L \} \right]$$

Coverage rates

- Distribution over the subjects of the coverage rates (COV)

