Quantifying and comparing dynamic predictive accuracy of joint models

for longitudinal marker and time-to-event with competing risks

P. Blanche, C. Proust-Lima, L. Loubère, H. Jacqmin-Gadda

UNIVERSITÉ BORDEAUX SEGALEN

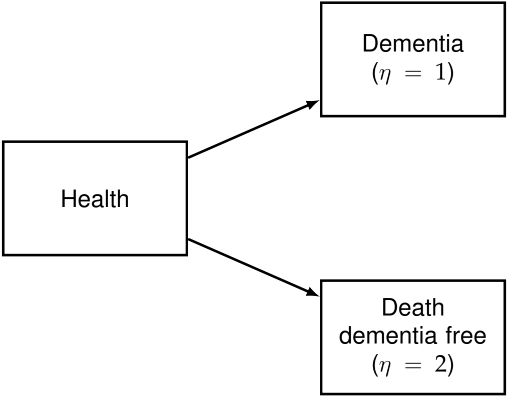
OBJECTIVE

- Question : How to evaluate and compare dynamic predictive accuracy of joint-models?

- Data: Cohorts of elderly people Paquid (training, $n = 2970$) and 3-City (validation, $n = 3880$)
  - Dynamic prediction of dementia
  - Using repeated measurements of cognitive tests

- Statistical Goal : making inference with dynamic accuracy measures
  - Estimating dynamic predictive accuracy curves
  - Testing whether or not 2 curves of predictive accuracy differ

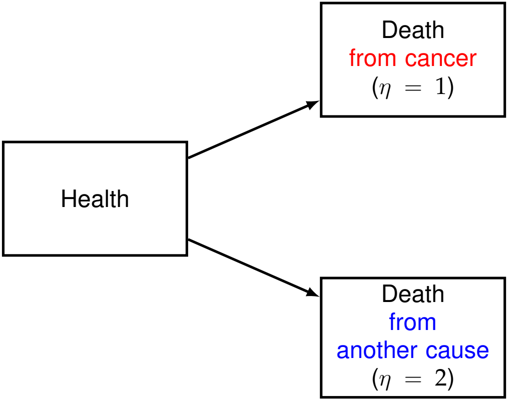
Competing risks : Motivation example



Notations:

- T : time-to-event
- η : type of event

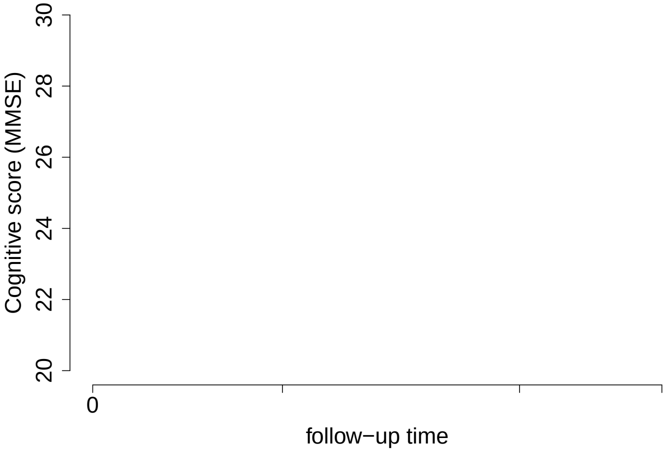
COMPETING RISKS IN CANCER



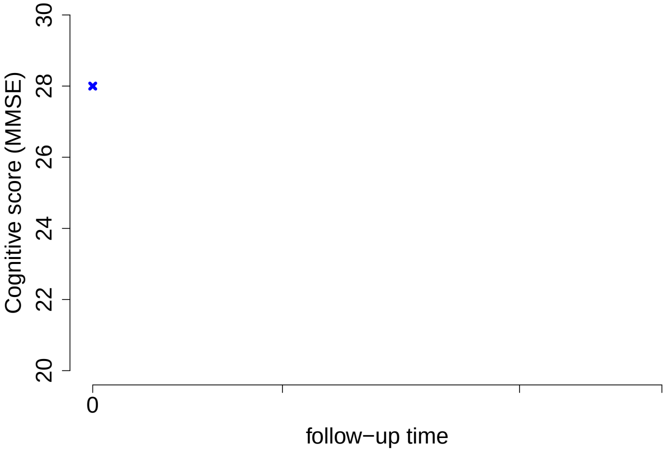
Notations:

- T : time-to-event
- η : type of event

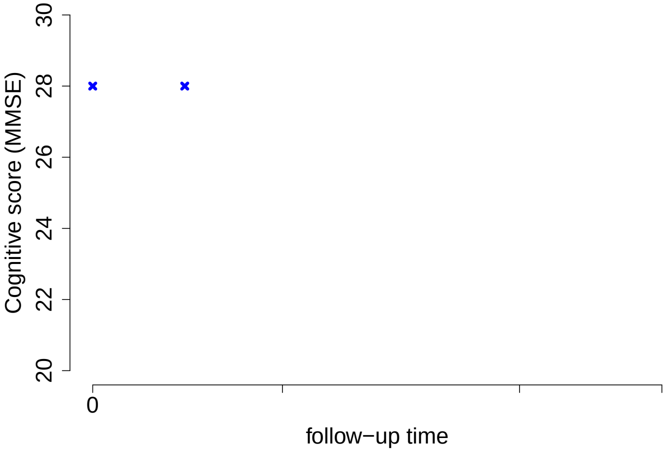
DYNAMIC PREDICTION



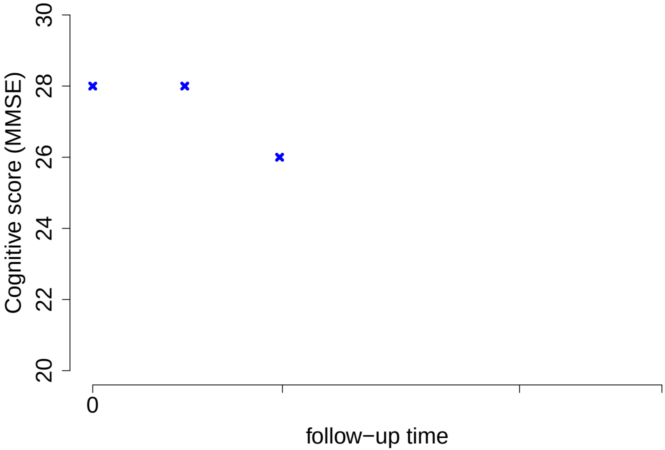
Dynamic prediction



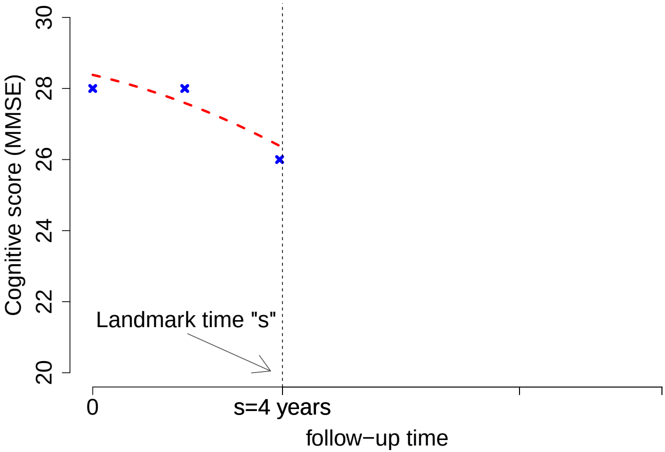
DYNAMIC PREDICTION



DYNAMIC PREDICTION

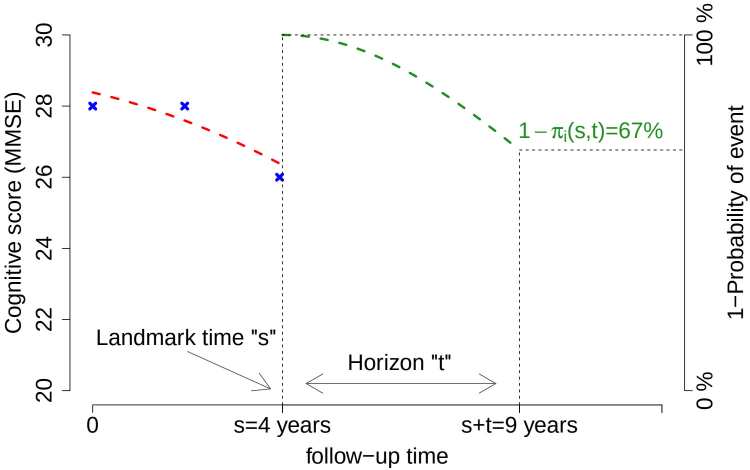


DYNAMIC PREDICTION



Dynamic prediction

Landmark time "*s*" at which predictions are made varies, horizon "*t*" is fixed.



Notations for population parameters

- Event-time and event-type : (T_i, η_i)
- Indicator of disease occurrence in $(s, s+t]$:

$$D_i(s,t) = \mathbb{1}\{s < T_i \le s+t, \eta_i = 1\}$$

- Dynamic predictions:

$$\pi_i(s,t) = \mathbb{P}_{\widehat{\xi}}\Big(D_i(s,t) = 1 \Big| T_i > s, \mathcal{Y}_i(s), \mathbf{X}_i\Big)$$

$$= \mathbb{P}_{\widehat{\xi}}(s < T_i \le s+t, \eta_i = 1 | T_i > s, \mathcal{Y}_i(s), \mathbf{X}_i)$$

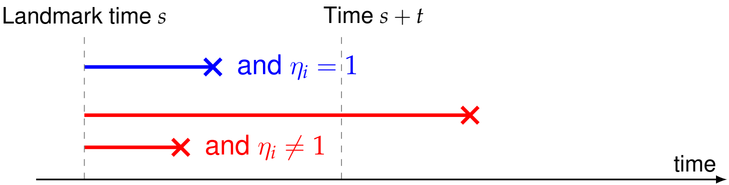
  - $\mathcal{Y}_i(s)$: set of marker measurements measured before time s
  - $\mathbf{X}_i$: baseline covariates
  - $\widehat{\xi}$: estimated model parameters (from independent training data)

PREDICTIVE ACCURACY : DISCRIMINATION

$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

- Does a higher predicted risk really mean more likely to experience the event ?

- How often $\pi_i(s,t) > \pi_j(s,t)$ and $D_i(s,t) = 1$, $D_j(s,t) = 0$?



DEFINITIONS OF ACCURACY: AUC(s,t)

$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

AUC (Area under ROC curve):

$$\boxed{\text{AUC}(s,t) = \mathbb{P}\Big(\pi_i(s,t) > \pi_j(s,t) \Big| D_i(s,t) = 1, D_j(s,t) = 0, T_i > s, T_j > s\Big)}$$

with i and j two independent subjects.

- the higher the better
- Discrimination measure
- Does NOT depend on incidence in $(s, s+t]$

PREDICTIVE ACCURACY : PREDICTION ERROR

$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

- How close are the predicted risks $\pi_i(s,t)$ from the "true underlying" risk of event given the available information ?

- Is it true that :

$$\begin{aligned} \pi_i(s,t) &\approx \mathbb{E}\Big[D_i(s,t)\Big| T_i > s, \mathcal{Y}_i(s), \mathbf{X}_i\Big] \\ &\approx \mathbb{P}\big(s < T_i \leq s+t, \eta_i = 1 \big| T_i > s, \mathcal{Y}_i(s), \mathbf{X}_i\big) \quad ? \end{aligned}$$

DEFINITIONS OF ACCURACY: BS(s,t)

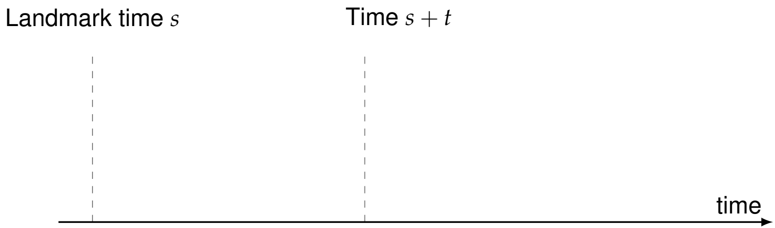
$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

Expected Brier Score:

$$\boxed{\text{BS}(s,t) = \mathbb{E}\left[\left\{D(s,t) - \pi(s,t)\right\}^2 \middle| T > s\right]}$$

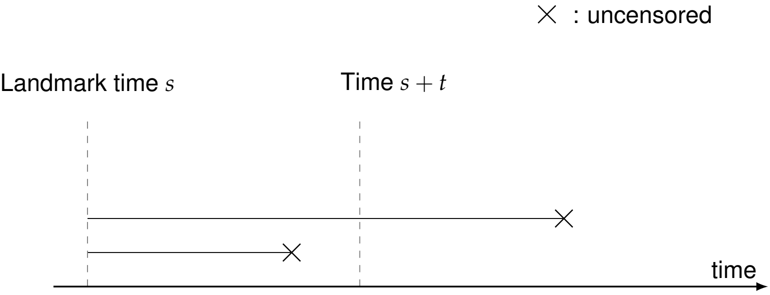
- the lower the better
- BS $\approx$ Bias^2 + Variance
- Calibration and Discrimination
- Depends on incidence in $(s, s+t]$

RIGHT CENSORING ISSUE



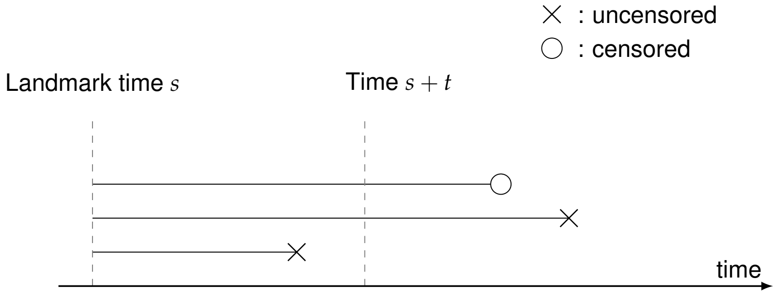
$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

Right censoring issue



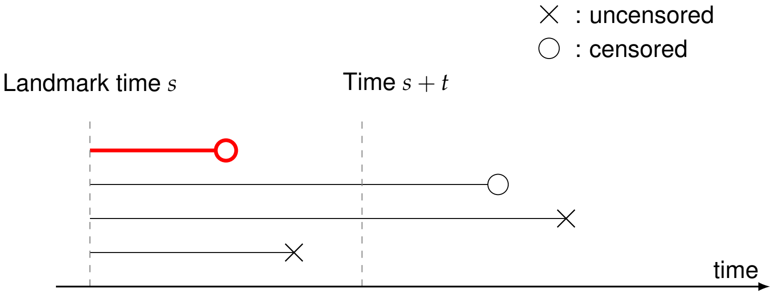
$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

RIGHT CENSORING ISSUE



$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

Right censoring issue



For subject i censored within $[s, s+t)$ the status

$$D_i(s,t) = \mathbb{1}\{s < T_i \leq s+t, \eta_i = 1\}$$

is unknown.

Notations for right censored observation

Observed iid sample :

$$\Big\{ \big(\widetilde{T}_i, \Delta_i, \widetilde{\eta}_i, \pi_i(\cdot,\cdot) \big), i = 1, \ldots, n \Big\}$$

with

$$\widetilde{T}_i = \min(T_i, C_i) \quad \text{and} \quad \widetilde{\eta}_i = \Delta_i \eta_i$$

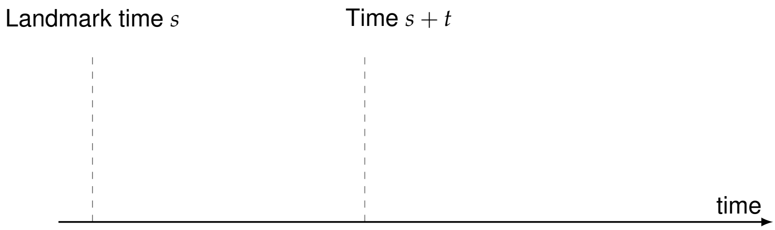
where

- C_i: censoring
- $\Delta_i = \mathbb{1}\{T_i \leq C_i\}$: censoring indicator.

Inverse Probability of Censoring Weighting (IPCW) estimators (1/2)

$$\boxed{\widehat{W}_i(s,t) = \qquad\qquad\qquad + \qquad\qquad +}$$

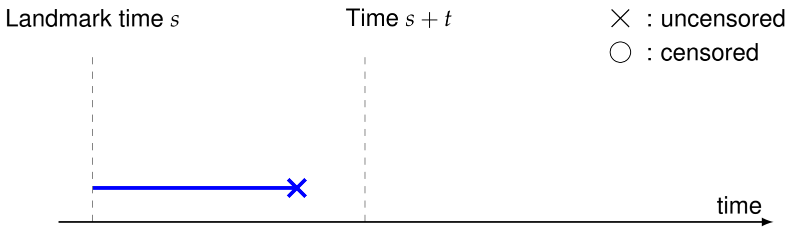
with $\widehat{G}(u)$ the Kaplan-Meier estimator of $\mathbb{P}(C > u)$.



INVERSE PROBABILITY OF CENSORING WEIGHTING (IPCW) ESTIMATORS (1/2)

$$\boxed{\widehat{W}_i(s,t) = \frac{\mathbb{1}\{s < \widetilde{T}_i \le s+t\}\Delta_i}{\widehat{G}(\widetilde{T}_i|s)} + \qquad\qquad +}$$

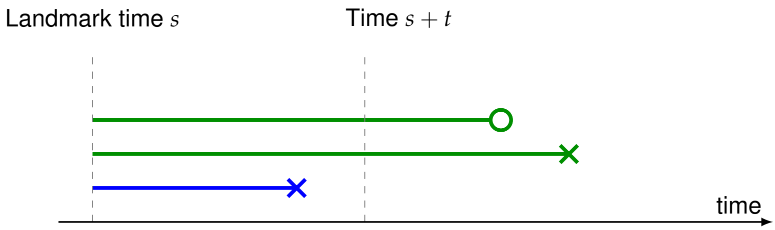
with $\widehat{G}(u)$ the Kaplan-Meier estimator of $\mathbb{P}(C > u)$.



Inverse Probability of Censoring Weighting (IPCW) estimators (1/2)

$$\boxed{\widehat{W}_i(s,t) = \frac{\mathbb{1}\{s < \widetilde{T}_i \le s+t\}\Delta_i}{\widehat{G}(\widetilde{T}_i|s)} + \frac{\mathbb{1}\{\widetilde{T}_i > s+t\}}{\widehat{G}(s+t|s)} +}$$

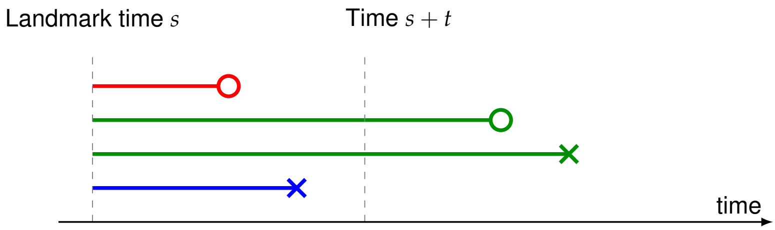
with $\widehat{G}(u)$ the Kaplan-Meier estimator of $\mathbb{P}(C > u)$.



INVERSE PROBABILITY OF CENSORING WEIGHTING (IPCW) ESTIMATORS (1/2)

$$\boxed{\widehat{W}_i(s,t) = \frac{\mathbb{1}\{s < \widetilde{T}_i \le s+t\}\Delta_i}{\widehat{G}(\widetilde{T}_i|s)} + \frac{\mathbb{1}\{\widetilde{T}_i > s+t\}}{\widehat{G}(s+t|s)} + 0}$$

with $\widehat{G}(u)$ the Kaplan-Meier estimator of $\mathbb{P}(C > u)$.



Inverse Probability of Censoring Weighting (IPCW) estimators (2/2)

- Indicator of "observed disease occurrence" in $(s, s+t]$:

$$\widetilde{D}_i(s,t) = \mathbb{1}\{s < \widetilde{T}_i \leq s+t, \widetilde{\eta}_i = 1\}$$

(instead of $D_i(s,t)$).

INVERSE PROBABILITY OF CENSORING WEIGHTING (IPCW) ESTIMATORS (2/2)

- Indicator of "observed disease occurrence" in $(s, s+t]$:

$$\widetilde{D}_i(s,t) = \mathbb{1}\{s < \widetilde{T}_i \leq s+t, \widetilde{\eta}_i = 1\}$$

(instead of $D_i(s,t)$).

- Expected Brier score estimator:

$$\widehat{BS}(s,t) = \frac{1}{n}\sum_{i=1}^{n} \widehat{W}_i(s,t) \left\{\widetilde{D}_i(s,t) - \pi_i(s,t)\right\}^2$$

$\widehat{AUC}(s,t)$ similarly defined...

ASYMPTOTIC IID REPRESENTATION

Let θ denote either AUC or BS.

LEMMA: Assume that the censoring time C is independent of $(T, \eta, \pi(\cdot,\cdot))$, then

$$\boxed{\sqrt{n}\left(\widehat{\theta}(s,t) - \theta(s,t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathsf{IF}_\theta(\widetilde{T}_i, \widetilde{\eta}_i, \pi_i(s,t), s, t) + o_p(1)}$$

where $\mathsf{IF}_\theta(\widetilde{T}_i, \widetilde{\eta}_i, \pi_i(s,t), s, t)$ being :

- zero-mean iid terms
- easy to estimate (plugging in Nelson-Aalen & Kaplan-Meier)

PROOF OF ASYMPTOTIC IID REPRESENTATION

The proof consists in 3 steps:

(i) Martingale theory to account for Kaplan-Meier estimator variability

(ii) Taylor expansions to connect variability of estimated weights to variability of the weighted sum.
$\rightarrow$ sum of non-iid terms

(iii) Hájek projection to rewrite the sum of non-iid terms as an equivalent sum of iid-terms (U-statistic theory)

POINTWISE CONFIDENCE INTERVAL (FIXED s)

- Asymptotic normality:

$$\sqrt{n}\Big(\widehat{\theta}(s,t) - \theta(s,t)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2_{s,t}\right)$$

- 95% confidence interval:

$$\left\{\widehat{\theta}(s,t) \pm z_{1-\alpha/2}\frac{\widehat{\sigma}_{s,t}}{\sqrt{n}}\right\}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\mathcal{N}(0,1)$.

- Variance estimator:

$$\widehat{\sigma}^2_{s,t} = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{\mathsf{IF}}_\theta(\widetilde{T}_i, \widetilde{\eta}_i, \pi_i(s,t), s, t)\right\}^2$$

SIMULTANEOUS CONFIDENCE BAND OVER A SET OF LANDMARK TIMES $s \in \mathcal{S}$

$$\left\{ \widehat{\theta}(s,t) \pm \widehat{q}_{1-\alpha}^{(\mathcal{S},t)} \frac{\widehat{\sigma}_{s,t}}{\sqrt{n}} \right\}, \quad s \in \mathcal{S}$$

Mimicking Lin, et al. (Biometrika, 1994)

Simultaneous confidence band over a set of landmark times $s \in \mathcal{S}$

$$\left\{ \widehat{\theta}(s,t) \pm \widehat{q}_{1-\alpha}^{(\mathcal{S},t)} \frac{\widehat{\sigma}_{s,t}}{\sqrt{n}} \right\}, \quad s \in \mathcal{S}$$

Computation of $\widehat{q}_{1-\alpha}^{(\mathcal{S},t)}$ by the simulation algorithm:

1. For $b = 1, \ldots, B$, say $B = 4000$, do:
   1.1 Generate $\{\omega_1^b, \ldots, \omega_n^b\}$ from n iid $\mathcal{N}(0,1)$.
   1.2 Using the plug-in estimator $\widehat{\mathrm{IF}}_\theta(\cdot)$, compute :

$$\Upsilon^b = \sup_{s \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i^b \frac{\widehat{\mathrm{IF}}_\theta(\widetilde{T}_i, \widetilde{\eta}_i, \pi_i(s,t), s, t)}{\widehat{\sigma}_{,s,t}} \right|$$

2. Compute $\widehat{q}_{1-\alpha}^{(\mathcal{S},t)}$ as the $100(1-\alpha)$th percentile of $\left\{\Upsilon^1, \ldots, \Upsilon^B\right\}$

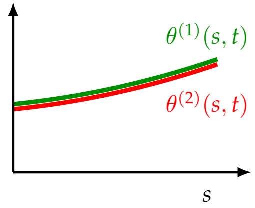
Mimicking Lin, et al. (Biometrika, 1994)

COMPARING DYNAMIC PREDICTIVE ACCURACY CURVES (1/2)

Doing similarly with a difference in predictive accuracy of 2 dynamic predictions $\pi^{(l)}(\cdot, t)$, $l = 1, 2$, we are able

- to test

$$\boxed{\mathcal{H}_0 : \forall s \in \mathcal{S} \quad \theta^{(1)}(s,t) - \theta^{(2)}(s,t) = 0}$$



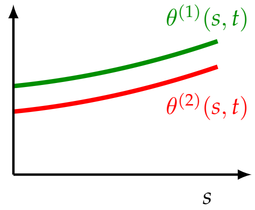
by observing whether or not the zero function is contained within the confidence band of $\theta^{(1)}(s,t) - \theta^{(2)}(s,t)$ versus s

Comparing dynamic predictive accuracy curves (2/2)

Doing similarly with a difference in predictive accuracy of 2 dynamic predictions $\pi^{(l)}(\cdot, t)$, $l = 1, 2$, we are able

- to assert

$$\boxed{\forall s \in \mathcal{S} \quad \theta^{(1)}(s,t) > \theta^{(2)}(s,t)}$$



by observing whether or not the confidence band $\theta^{(1)}(s,t) - \theta^{(2)}(s,t)$ versus s overlaps the zero line.

Data from 2 cohorts of elderly subjects

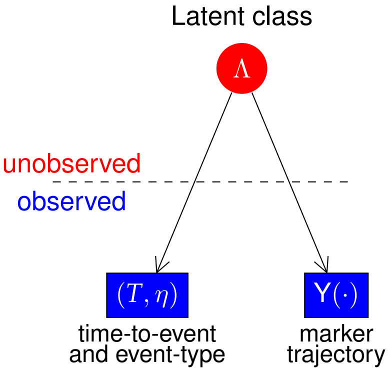
- Population based studies of elderly subjects:

	No. of subjects	follow-up
training cohort: Paquid	2970	20 years
validation cohort: 3-City	3880	9 years

- Repeated measurements of 2 cognitive tests:
  - Mini Mental State Examination (MMSE):
    $\rightarrow$ global index of cognition
  - Isaac Score Test (IST):
    $\rightarrow$ evaluates speed of verbal production

Joint Latent Class Model

(T, η) and $\mathrm{Y}(\cdot)$ are joint by the latent class Λ



Baseline covariates: Age, Education level and Sex

JOINT LATENT CLASS MODELING ($K = 3$ CLASSES)

JOINT LATENT CLASS MODELING ($K = 3$ CLASSES)

- MMSE (transformed) or IST decline given class $\Lambda_i = g$:

$$\begin{aligned} Y_i(t_{ij})|_{\Lambda_i=g} =& \beta_0 + \beta_{0,age}\mathbf{AGE}_i + \beta_{0,educ}\mathbf{EDUC}_i + \beta_{0,learn}\mathbb{1}\{t_{ij}=0\} + b_{i0|\Lambda_i=g} \\ &+ \left(\beta_{1g} + \beta_{1,age}\mathbf{AGE}_i + b_{i1|\Lambda_i=g}\right) \times t_{ij} \\ &+ \left(\beta_{2g} + \beta_{2,age}\mathbf{AGE}_i + b_{i2|\Lambda_i=g}\right) \times t_{ij}^2 + \varepsilon_i(t_{ij}), \end{aligned}$$

with $(b_{i0|\Lambda_i=g}, b_{i1|\Lambda_i=g}, b_{i2|\Lambda_i=g}) \sim \mathcal{N}(\mathbf{0}, \sigma_g^2\mathbf{B})$

Joint Latent Class Modeling ($K = 3$ classes)

- MMSE (transformed) or IST decline given class $\Lambda_i = g$:

$$\begin{aligned} Y_i(t_{ij})|_{\Lambda_i=g} = & \beta_0 + \beta_{0,age}\mathbf{AGE}_i + \beta_{0,educ}\mathbf{EDUC}_i + \beta_{0,learn}\mathbb{1}\{t_{ij}=0\} + b_{i0|\Lambda_i=g} \\ & + \left(\beta_{1g} + \beta_{1,age}\mathbf{AGE}_i + b_{i1|\Lambda_i=g}\right) \times t_{ij} \\ & + \left(\beta_{2g} + \beta_{2,age}\mathbf{AGE}_i + b_{i2|\Lambda_i=g}\right) \times t_{ij}^2 + \varepsilon_i(t_{ij}), \end{aligned}$$

with $(b_{i0|\Lambda_i=g}, b_{i1|\Lambda_i=g}, b_{i2|\Lambda_i=g}) \sim \mathcal{N}(\mathbf{0}, \sigma_g^2\mathbf{B})$

- Risk of events given class $\Lambda_i = g$:
  - dementia

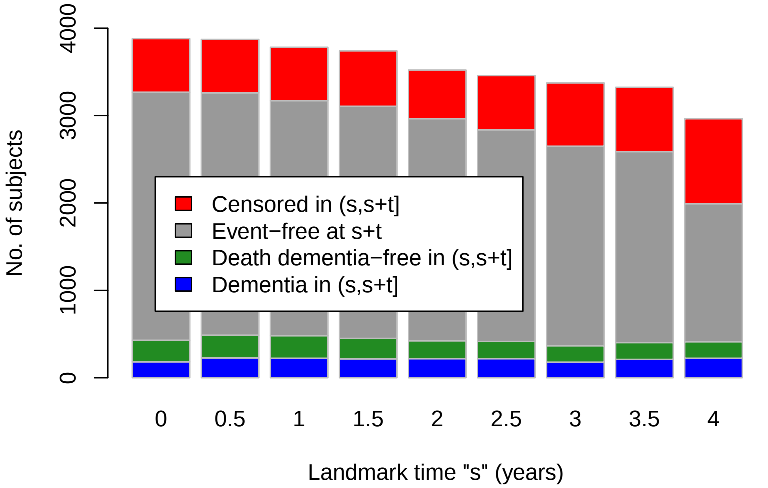
$$\lambda_{i,1}(t|\Lambda_i = g) = \lambda_{01,g}(t)\exp\left(\alpha_{11,g}\mathbf{AGE}_i + \alpha_{21,g}\mathbf{EDUC}_i\right)$$

  - death dementia-free

$$\lambda_{i,2}(t|\Lambda_i = g) = \lambda_{02,g}(t)\exp\left(\alpha_{12,g}\mathbf{AGE}_i + \alpha_{22,g}\mathbf{EDUC}_i + \alpha_{32,g}\mathbf{SEX}_i\right).$$

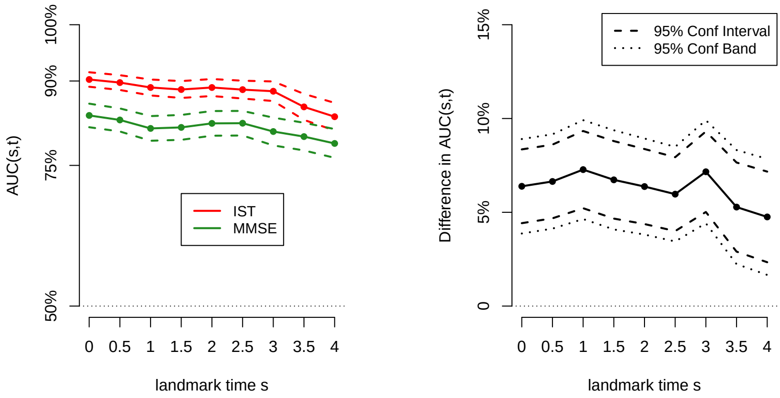
Descriptive statistics & Right censoring issue

$t = 5$ years, $s \in \mathcal{S} = \{0, 0.5, \ldots, 4\}$ years



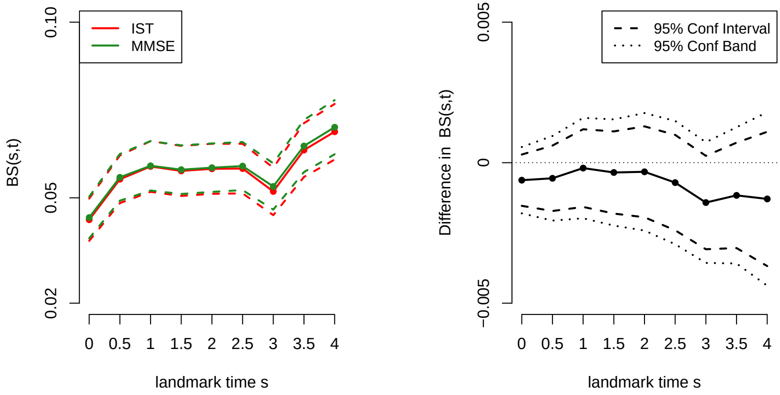
Dynamic prediction accuracy curves: AUC

$t = 5$ years, $s \in \mathcal{S} = \{0, 0.5, \ldots, 4\}$ years



COMPARING PREDICTION ACCURACY CURVES: BS

$t = 5$ years, $s \in \mathcal{S} = \{0, 0.5, \ldots, 4\}$ years



PERSPECTIVE: R^2-LIKE CRITERIA

- Interpretation difficulties for $s \mapsto BS(s,t)$:
  - Scaling meaning ?
  - BS value depends on cumulative incidence in $(s, s+t]$
  - Increase/decrease when s varies not explainable

PERSPECTIVE: R^2-LIKE CRITERIA

- Interpretation difficulties for $s \mapsto BS(s,t)$:
  - Scaling meaning ?
  - BS value depends on cumulative incidence in $(s, s+t]$
  - Increase/decrease when s varies not explainable

- "Explained variation" criteria :

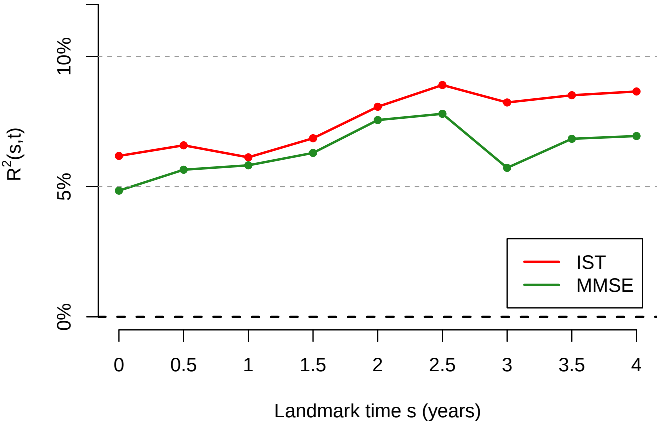
$$R^2(s,t) = 1 - \frac{BS(s,t)}{BS_{NULL}(s,t)}$$

where $BS_{NULL}(s,t)$ is BS of the null model predicting the same risk for all subjects (=cumulative incidence in $(s, s+t]$).

  - the higher the better & easier scaling
  - cumulative incidence free

Perspective: inference for R^2-like criteria

$t = 5$ years, $s \in \mathcal{S} = \{0, 0.5, \ldots, 4\}$



Computation of confidence regions (easy): ongoing work ...

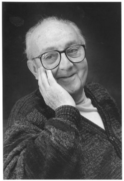
Conclusion (1/2)

- New testing approach to simultaneously compare dynamic predictions over all times at which predictions are made
- Nonparametric methodology provides a model-free comparison.

CONCLUSION (1/2)

- New testing approach to simultaneously compare dynamic predictions over all times at which predictions are made
- Nonparametric methodology provides a model-free comparison.

"Essentially, all models are wrong, but some are useful.", G. Box



$\Rightarrow$ We do not assume any correct model specification.

Conclusion (2/2)

- Asymptotic results established
- Good simulation results with finite sample size (not shown)

CONCLUSION (2/2)

- Asymptotic results established
- Good simulation results with finite sample size (not shown)
- Beyond the joint modeling framework ?

  $\approx$ provide inference procedures for comparing any kind of dynamic prediction tools

e.g : Joint modeling vs Landmarking ?

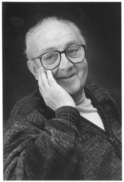
CONCLUSION (2/2)

- Asymptotic results established
- Good simulation results with finite sample size (not shown)
- Beyond the joint modeling framework ?

  $\approx$ provide inference procedures for comparing any kind of dynamic prediction tools

e.g : Joint modeling vs Landmarking ?

"Statisticians, like artists, have the bad habit of falling in love with their models.",
G. Box



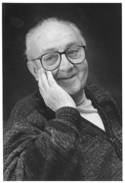
CONCLUSION (2/2)

- Asymptotic results established
- Good simulation results with finite sample size (not shown)
- Beyond the joint modeling framework ?

  $\approx$ provide inference procedures for comparing any kind of dynamic prediction tools

e.g : Joint modeling vs Landmarking ?

"Statisticians, like artists, have the bad habit of falling in love with their models.", G. Box



THANK YOU FOR YOUR ATTENTION!