


TECHNICAL ADVANCE

Open Access



# Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history: a landmark approach

Anthony Devaux<sup>1\*</sup> , Robin Genuer<sup>1,2</sup>, Karine Peres<sup>1</sup> and Cécile Proust-Lima<sup>1</sup>

## Abstract

**Background:** The individual data collected throughout patient follow-up constitute crucial information for assessing the risk of a clinical event, and eventually for adapting a therapeutic strategy. Joint models and landmark models have been proposed to compute individual dynamic predictions from repeated measures to one or two markers. However, they hardly extend to the case where the patient history includes much more repeated markers. Our objective was thus to propose a solution for the dynamic prediction of a health event that may exploit repeated measures of a possibly large number of markers.

**Methods:** We combined a landmark approach extended to endogenous markers history with machine learning methods adapted to survival data. Each marker trajectory is modeled using the information collected up to the landmark time, and summary variables that best capture the individual trajectories are derived. These summaries and additional covariates are then included in different prediction methods adapted to survival data, namely regularized regressions and random survival forests, to predict the event from the landmark time. We also show how predictive tools can be combined into a superlearner. The performances are evaluated by cross-validation using estimators of Brier Score and the area under the Receiver Operating Characteristic curve adapted to censored data.

**Results:** We demonstrate in a simulation study the benefits of machine learning survival methods over standard survival models, especially in the case of numerous and/or nonlinear relationships between the predictors and the event. We then applied the methodology in two prediction contexts: a clinical context with the prediction of death in primary biliary cholangitis, and a public health context with age-specific prediction of death in the general elderly population.

**Conclusions:** Our methodology, implemented in R, enables the prediction of an event using the entire longitudinal patient history, even when the number of repeated markers is large. Although introduced with mixed models for the repeated markers and methods for a single right censored time-to-event, the technique can be used with any other appropriate modeling technique for the markers and can be easily extended to competing risks setting.

**Keywords:** Individual prediction, Landmark, Longitudinal data, Survival data, Machine learning methods

\*Correspondence: [anthony.devaux@u-bordeaux.fr](mailto:anthony.devaux@u-bordeaux.fr)

<sup>1</sup>INSERM, BPH, U1219, Univ. Bordeaux, Bordeaux, France

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

A central issue in health care is to quantify the risk of disease, disease progression or death at the individual level, for instance to initiate or adapt a treatment strategy as soon as possible. To achieve this goal, the information collected at a given time (at diagnosis or at the first visit) is often not sufficient and repeated measurements of markers are essential. For example, repeated prostate specific antigen (PSA) data are highly predictive of the risk of prostate cancer recurrence [1–3], and markers such as diabetic status or blood pressure level over time are crucial in predicting the risk of cardiovascular disease [4, 5]. Including longitudinal information into the prediction of a clinical event defines the framework for individual dynamic predictions [1, 6, 7]. In some contexts, a single marker may be sufficient to predict the occurrence of the event (e.g., in prostate cancer with PSA) but often the complete patient history with possibly many repeated markers should be exploited (see Fig. 1). Yet, statistical developments for individual prediction of event have so far either focused on the repeated nature of the information or on its large dimension.

When using repeated information to develop dynamic prediction tools, two approaches are commonly used: joint models [1, 6] and landmark models [8]. Joint models simultaneously analyze the longitudinal and event time processes by assuming a structure of association built on summary variables of the marker dynamics [9]. This model which uses all the information on the longitudinal and time-to-event processes to derive the prediction tool is widely used in the case of a single repeated marker but becomes intractable in the presence of more than a few repeated markers due to high computational complexity [7].

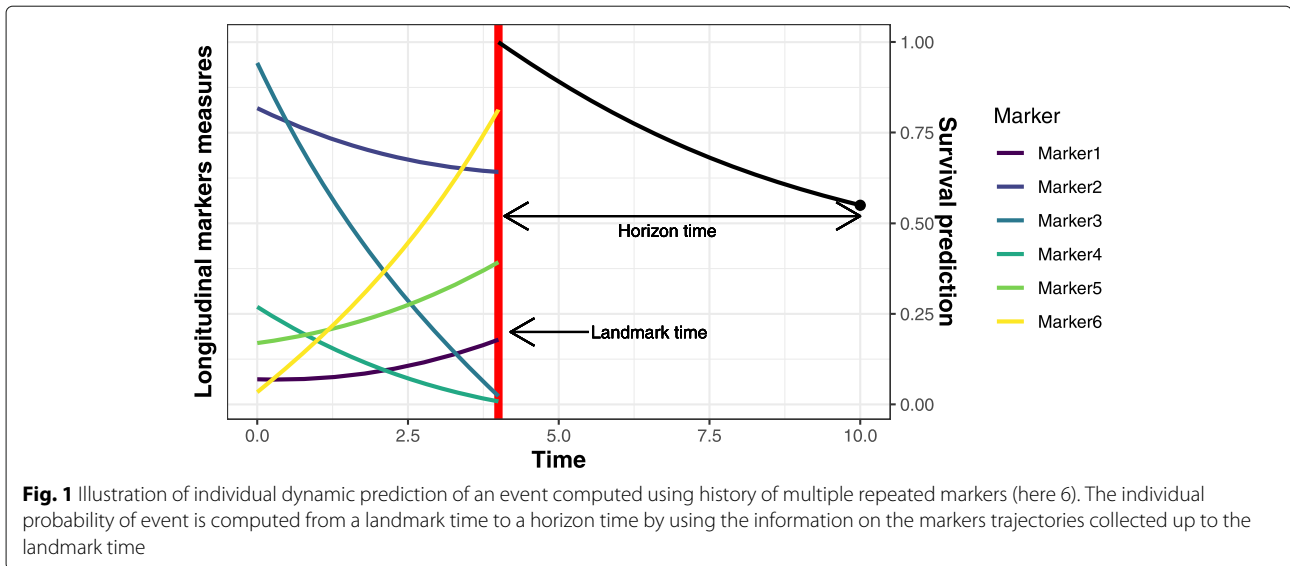
An alternative is to use partly conditional survival model [10] or landmark models [8] which consist in directly focusing on the individuals still at risk at the landmark time and consider their history up to the landmark time (see Fig. 1). When individual history includes repeated measures of an endogenous marker, summaries of the marker derived from preliminary mixed models can be included in the survival model, instead of only the last observed value [1, 5]. Although the landmark models do not use as much information as the joint model (only information from the at-risk individuals at the landmark time is exploited) and thus may lack of efficiency, they have shown competitive predictive performances, easier implementation (much less numerical problems) and better robustness to misspecification than joint models [7]. However, as joint models, they necessitate to consider the actual nature of the relationship between the marker and the event.

Although the landmark approach is *per se* very general, in practice its definition is based on standard sur-

vival models, namely the Cox model, which prevents the methodology to be applied in large dimensional contexts usually encountered in applications. Indeed the Cox model becomes rapidly limited in the presence of: 1) a large number of predictors, 2) highly correlated predictors, and 3) complex relationships between the predictors and the event [11]. Yet, in the context of dynamic prediction from multiple repeated markers, these three limits are rapidly reached. Indeed, the large dimension of the predictors does not only come from the number of markers but also from the number of (potentially correlated with each other) marker-specific summaries that are necessary to approximate the actual nature of the relationship between the marker and the event.

Machine learning methods, including regularized regressions or decision trees and random forests, have been specifically developed to predict outcomes while tackling the aforementioned issues [12]. Their good predictive performances have been largely demonstrated in the literature [13]. Initially proposed for continuous or binary outcomes, they have been recently extended to handle right censored time-to-event data. For instance, Simon et al. [14] developed penalized Cox models either using Ridge, Lasso or Elastic-Net penalty, Bastien et al. [15] developed a Cox model based on deviance residuals-based sparse-Partial Least Square, as an extension of sparse-Partial Least Square [16] for survival data, and Ishwaran et al. [17] extended random forests to survival data. However, they were mostly applied to predict time-to-event from time-independent marker information. Our purpose is thus to show how these machine learning methods can also be leveraged to provide dynamic individual predictions from large dimensional longitudinal biomarker data.

Computing dynamic predictions in the context of a large number of repeated markers is a very new topic in statistics, and only a few proposals have been made very recently. Zhao et al. [18] and Jiang et al. [19] focused on random forests. Using a landmark approach, Zhao et al. transformed the survival data into pseudo-observations and incorporated in each tree the marker information at a randomly selected time. Although handling repeated markers, this method neither accounts for measurement errors of the biomarkers nor their trajectory shapes. By considering a functional ensemble survival tree, Jiang et al. overcame this issue. They characterized the changing patterns of continuous time-varying biomarkers using functional data analysis, and incorporated those characteristics directly into random survival forests. By concomitantly analyzing the markers and the event, this approach belongs to the two-stage calibration approaches [20] and may suffer from the same biases [21]. Finally Tanner et al. [22] proposed to extend the landmark approach to incorporate multiple repeated markers with measure-



ments errors. For the survival prediction method, they chose to discretize the time and use an ensemble of classical binary classifiers to predict the event.

In comparison with this emerging literature, our proposal goes one step forward. As in Tanner et al., we chose to rely on a landmark approach and consider various prediction methods rather than only random forests. However, we also chose to directly exploit the survival data in continuous time. In addition, our methodology handles markers of different nature, accounts for their measurement error and intermittent missing data, and for a possibly large number of summary characteristics of each marker.

In the following sections, we first describe the proposed method. We then demonstrate in a simulation study the performances of the methodology and the benefit of using machine learning methods to handle the large dimensional aspect. We then illustrate the methodology in two very different contexts: a clinical context with the prediction of death in primary biliary cholangitis, and a public health context with the prediction of 5-year death at different ages in the general elderly population. The paper ends with the discussion of the strengths and weaknesses of the proposed method.

## Methods

### Framework, notations and general principle

Let us consider a landmark time  $t_{LM}$  of interest and a population of  $N_{t_{LM}}$  individuals that are still at risk of the event at  $t_{LM}$ . For an individual  $i \in \{1, \dots, N_{t_{LM}}\}$ , we denote  $T_i$  the true event time,  $C_i$  the independent censoring time. We define  $T_i^* = \min(T_i, C_i)$  the observed time event and  $\delta_i = \mathbb{1}(T_i < \min(C_i, t_{LM} + t_{Hor}))$  the event indicator with  $t_{Hor}$  the horizon time. We consider a single event for simplicity.

At the landmark time,  $P$  time-independent covariates  $X_i$  are available, and the history of  $K$  time-dependent markers  $Y_{ijk}$  ( $k \in \{1, \dots, K\}$ ) measured at time  $t_{ij}$  ( $j \in \{1, \dots, n_i\}$ ) and  $t_{ij} \leq t_{LM}$ .

The target individual probability of event from the landmark time  $t_{LM}$  to the horizon time  $t_{Hor}$  of a subject  $\star$  is defined as:

$$\pi_\star(t_{LM}, t_{Hor}) = P(T_\star \leq t_{LM} + t_{Hor} \mid T_\star > t_{LM}, \{Y_{\star jk}; k = 1, \dots, K, t_{\star jk} \leq t_{LM}\}, X_\star) \tag{1}$$

By assuming that the history of the  $K$  marker trajectories up to  $t_{LM}$  can be summarized into a vector  $\Gamma_\star$ , we define the following probability:

$$\tilde{\pi}_\star(t_{LM}, t_{Hor}) = P(T_\star \leq t_{LM} + t_{Hor} \mid T_\star > t_{LM}, \Gamma_\star(t_{LM}), X_\star) \tag{2}$$

This probability is estimated by  $\hat{\pi}_\star^{(m)}(t_{LM}, t_{Hor})$  in 4 steps on a learning sample for each survival prediction method  $m$ :

- 1 Each marker trajectory is modeled using the information collected up to  $t_{LM}$
- 2 The vector of summary variables  $\Gamma_i(t_{LM})$  is computed for each individual  $i$
- 3  $\Gamma_i(t_{LM})$  and additional baseline covariates  $X_i$  are entered into survival prediction method  $m$
- 4 The predicted probability of event  $\hat{\pi}_\star^{(m)}(t_{LM}, t_{Hor})$  is computed from survival method  $m$

Once the estimator defined (i.e., the survival prediction method trained) on the learning sample, the summary variables  $\Gamma_\star(t_{LM})$  can be computed for any new external individual  $\star$  at risk of event at  $t_{LM}$ , and the corresponding individual predicted probability of event can be deduced.

**Step 1. longitudinal model for markers history**

Longitudinal markers are usually measured at intermittent times with error. The first step consists to estimate the error-free trajectory of the marker of each individual over the history period. We propose to use generalized mixed models [23] defined as:

$$g(E(Y_{ijk}|b_{ik})) = Y_{ik}^*(t_{ijk}) = X_{ik}^\top(t_{ijk})\beta_k + Z_{ik}^\top(t_{ijk})b_{ik} \quad (3)$$

where  $X_{ik}^\top(t_{ijk})$  and  $Z_{ik}^\top(t_{ijk})$  are the  $p_k$ - and  $q_k$ -vectors associated with the fixed effects  $\beta_k$  and random effects  $b_{ik}$  (with  $b_{ik} \sim \mathcal{N}(0, B_k)$ ), respectively. The link function  $g(\cdot)$  is chosen according to the nature of  $Y_{ijk}$  (e.g. identity function for Gaussian continuous markers or logit function for binary markers).

**Step 2. summary characteristics of the marker trajectories**

Once the parameters of the model have been estimated (indicated by  $\hat{\cdot}$  below), any summary that captures the marker behavior up to the time  $t_{LM}$  can be computed. We give here a non-exhaustive list for individual  $i$ :

- Predicted individual deviations to the mean trajectory:  $\hat{b}_{ik} = \hat{B}_k Z_{ik}^\top \hat{V}_{ik}^{-1} (Y_{ik} - X_{ik} \hat{\beta}_k)$  where  $\hat{V}_{ik} = Z_{ik} \hat{B}_k Z_{ik}^\top + \hat{\sigma}_{\epsilon k} I_{n_i}$ , if the marker is continuous. Otherwise,  $\hat{b}_{ik} = \underset{b_{ik}}{\operatorname{argmax}} f(b_{ik}|Y_{ik}^*) = \underset{b_{ik}}{\operatorname{argmax}} f(Y_{ik}^*|b_{ik})f(b_{ik})$  with  $f(\cdot)$  the density function;
- Error-free level at time  $u \leq t_{LM}$ :  $\hat{Y}_{ik}^*(u) = X_{ik}^\top(u)\hat{\beta}_k + Z_{ik}^\top(u)\hat{b}_{ik}$  ;
- Error-free slope at time  $u \leq t_{LM}$ :  $\hat{Y}_{ik}^{*'}(u) = \frac{\partial \hat{Y}_{ik}^*(t)}{\partial t} |_{t=u}$  ;
- Cumulative error-free level during period  $\mathcal{T}$ :  $\hat{h}_{ik}(t_{LM}) = \int_{t_{LM}-\mathcal{T}}^{t_{LM}} \hat{Y}_{ik}^*(u) du$ .

Any additional summary that is relevant for a specific disease can be considered as soon as it is a function of the error-free marker trajectory (e.g., time spent above/below a given threshold). All the individual summary characteristics across the  $K$  markers are stored into a vector  $\Gamma_i$ . Using the list above and  $u = t_{LM}$ ,  $\Gamma_i(t_{LM}) = \{\hat{\Gamma}_{ik}(t_{LM}), k = 1, \dots, K\}$  with  $\hat{\Gamma}_{ik}(t_{LM}) = (b_{ik}, \hat{Y}_{ik}^*(t_{LM}), \hat{Y}_{ik}^{*'}(t_{LM}), \hat{h}_{ik}(t_{LM}))^\top$  is of length  $\sum_{k=1}^K (q_k + 3)$ . This vector may have a large amount of summaries which can also be highly correlated with each other. These particularities have to be taken into account in survival prediction methods.

**Step 3. prediction methods for survival data in a large dimensional context**

To predict the risk of event from  $t_{LM}$  to a horizon time  $t_{Hor}$  using the vector  $\mathcal{X}_i = (\Gamma_i, X_i)$  of summaries  $\Gamma_i$  and time-independent variables  $X_i$  of length  $P$ , we can use any technique that handles 1) right-censored time-to-event

data, 2) the possibly high dimension, 3) and the correlation between the predictors. We focused in this work on Cox model, Penalized-Cox model, Deviance residuals-based sparse-Partial Least Square and Random Survival Forests, although other techniques could also be applied. For each technique, several sub-methods were considered that differ according to the type of variable selection and/or the hyperparameters choices. We briefly describe the different techniques and sub-methods below, and refer to Section 1 in supplementary material for further details.

**Cox models**

The Cox model is a semi-parametric regression which models the instantaneous risk of event according to a log-linear combination of the independent covariates:

$$\lambda_i(t|\Gamma_i, X_i) = \lambda_0(t) \exp(X_i \gamma + \Gamma_i \eta) \quad (4)$$

with  $\lambda_0$  the baseline hazard function,  $\gamma$  and  $\eta$  the coefficients estimated by partial likelihood. We defined two sub-models whether variable selection was performed according to backward selection procedure using `step()` R function (called *Cox-SelectVar*) or not (*Cox-AllVar*).

**Penalized-Cox models**

Penalized-Cox models extend the Cox model defined in (4) to handle a high number of possibly correlated predictors. The partial log-likelihood is penalized with norm  $\ell_2$  (Ridge penalty), norm  $\ell_1$  (Lasso penalty [24]) which enables covariate selection, or a mixture of both (Elastic-Net [14]). These methods require the tuning of the norms mixing parameter (0 for Lasso, 1 for Ridge, ]0; 1[ for Elastic-Net) and the penalty parameter. We used `cv.glmnet()` function (from the `glmnet` R package) with internal cross-validation to tune the penalty parameter, and we defined three sub-models according to the norms mixing parameter (i.e. Lasso, Ridge or Elastic-Net). There are called *Penal-Cox-Lasso*, *Penal-Cox-Ridge* and *Penal-Cox-Elastic*, respectively.

**Deviance residuals-based sparse-Partial least square (sPLS-DR)**

Partial Least Square (PLS) is a method of dimension reduction where components (or latent variables) are built to maximize the covariance with the outcome. Sparse-PLS (sPLS) [16] adds a variable selection within each component using Lasso penalty. First developed in the framework of linear regression, this method was extended to survival data [15] (sPLS-DR). The principle is to apply a sPLS regression on the deviance residuals which are a normalized transformation of the martingale residuals  $\widehat{M}_i = \delta_i - \widehat{\Lambda}_i(t)$ , with  $\widehat{\Lambda}_i(t)$  the Nelson-Aalen cumulative hazard function estimate. Then, a Cox model is applied using the  $C$  identified components  $f_c(\Gamma_i, X_i)$  as covariates. In sPLS, the number of components  $C$  and the Lasso penalty

parameter on each component (which controls the sparsity on each component) have to be properly tuned. We used `cv.coxsplsDR()` function (from `splsRcox` R package) with internal cross-validation to tune the number of components, and considered three variants for the penalty: no penalty (called *sPLS-NoSparse*), maximum penalty (called *sPLS-MaxSparse*), or an optimized penalty from a grid of values (called *sPLS-Optimize*).

**Random survival forests**

Random forests [12] are a non-parametric machine learning tool that can handle high-dimensional data with possibly complex input-output relationships. Random forests, originally developed in a context of regression or classification, were later adapted to right-censored survival data [17] and called random survival forests (RSF). A RSF aggregates  $B$  survival trees, each one built on a different bootstrap sample from the original data (subjects not included in one bootstrap sample are called out-of-bag (OOB)). As any tree-based predictor, a survival tree recursively splits the sample into subgroups until the subgroups reach a certain minimal size  $S$ . To deal with time-to-event data, the splitting rule is usually based on the log-rank statistics although other splitting rules have also been proposed (e.g. gradient-based brier score splitting [17]). In RSF, at each node of each tree, a subset of  $M$  predictors is randomly drawn and the split is optimized among splits candidates only involving those predictors. The size of the predictors subset  $M$  and the minimal size  $S$  have to be tuned.

The interpretation of the link between the predictors and the event is not as easy in RSF as in (penalized) regression methods. To address this issue, RSF provide a quantification of this association, also known as variable importance (VIMP). For a given predictor  $p$ ,  $VIMP^{(p)}$  measures the mean (over all trees in the forest) increase of a tree error on its associated OOB sample, after randomly permuting the values of the  $p^{th}$  predictor in the OOB sample. Large VIMP values indicate variables with prediction ability while null (or even negative) VIMP values indicate variables that could be removed from the prediction tool.

Using `rfsrc()` function (from `randomForestSRC` R package), three RSF sub-methods were considered that differed according to  $M$  and  $S$  parameter tuning: default software parameters  $M = \text{square root of the number of predictors}$ ,  $S = 15$  (called *RSF-Default*),  $M$  and  $S$  that minimize the OOB error (called *RSF-Optimize*) or  $M$  and  $S$  optimized plus a variable selection using the VIMP statistic (called *RSF-SelectVar*).

**Step 4. predicted individual probability of event**

The estimator of individual probability of event  $\hat{\pi}_\star^{(m)}(t_{LM}, t_{Hor})$  for a new patient  $\star$  becomes:

- For Cox, penalized-Cox and sPLS-DR models:

$$\hat{\pi}_\star^{(m)}(t_{LM}, t_{Hor}) = 1 - \exp\left(-\hat{\Lambda}_0(t_{Hor}) \exp(\hat{P}_\star)\right) \quad (5)$$

with  $\hat{\Lambda}_0(\cdot)$  the Nelson-Aalen estimator, and  $\hat{P}_\star$  the predicted linear predictor directly obtained from  $\Gamma_\star$  and  $X_\star$  for Cox and Penalized-Cox models, or from the  $C$  components  $f_c(\Gamma_\star, X_\star)$  ( $c = 1, \dots, C$ ) for sPLS-DR.

- For RSF:

$$\hat{\pi}_\star^{(m)}(t_{LM}, t_{Hor}) = 1 - \exp\left(-\frac{1}{B} \sum_{b=1}^B \hat{\Lambda}_\star^b(t_{Hor})\right) \quad (6)$$

with  $\hat{\Lambda}_\star^b(t_{Hor})$  the Nelson-Aalen estimator in the leaf of tree  $b$  containing individual  $\star$ .

**Predictive accuracy assessment**

We assessed the predictive performances of the models using the time-dependent Area Under the ROC Curve (AUC) [25] defined as:

$$AUC(t_{LM}, t_{Hor}) = P\left(\pi_i(t_{LM}, t_{Hor}) > \pi_j(t_{LM}, t_{Hor}) \mid D_i(t_{LM}, t_{Hor}) = 1, D_j(t_{LM}, t_{Hor}) = 0, T_i > t_{LM}, T_j > t_{LM}\right) \quad (7)$$

and time-dependent Brier score [26] defined as:

$$BS(t_{LM}, t_{Hor}) = E\left[(D_i(t_{LM}, t_{Hor}) - \pi(t_{LM}, t_{Hor}))^2 \mid T > t_{LM}\right] \quad (8)$$

where  $D_i(t_{LM}, t_{Hor})$  is the survival status at time  $t_{LM} + t_{Hor}$ . We used estimators of these quantities that specifically handle the censored nature of  $D_i(t_{LM}, t_{Hor})$  using inverse censoring probability weighting (see [26, 27] for details).

In the applications, predictive accuracy assessment was done using a cross-validation approach to ensure independence between the samples on which each predictive tool was learnt and the samples on which their predictive accuracy was assessed (Fig. 2A). This induced a two-layer cross-validation since a cross-validation (or a bootstrap) was also performed within each training set to determine the method-specific hyperparameters.

**Combining the predictions into a single super learner**

Each survival prediction method  $m$  ( $m = 1, \dots, M$ ) provides a different individual predicted probability  $\hat{\pi}_\star^{(m)}$  (Eq. 2). In some cases, one will prefer to select the best predictive tool and rely on it. In other cases, one can also choose to combine the predictive tools into a Super-Learner predictive tool [28, 29]. It consists in defining the



final predicted probability as a weighted mean over the survival method-specific predictions:

$$\hat{\Pi}_* = \sum_{m=1}^M \omega_m \hat{\pi}_*^{(m)} \tag{9}$$

where the weights  $\omega_m$  (defined in  $[0, 1]$  with  $\sum_{m=1}^M \omega_m = 1$ ) are determined so that the Super-Learner predictive tool  $\hat{\Pi}$  minimizes a loss function. In our work, we chose to minimize the BS function defined in Eq. 8 by internal cross-validation. This lead to a three-layer cross-validation for the superlearner building and validation (see Fig. 2B).

## Results

### Performances of the methodology through a simulation study

We contrasted the performances of the different survival prediction methods according to different scenarios, based on Ishwaran et al. [30], in an extensive simulation study. Prediction tools were trained on  $R = 250$  learning datasets and their predictive performances were compared on a unique external validation dataset.

### Design

The  $R$  learning datasets and the validation dataset were generated according to the same design. They included  $N = 500$  individuals at risk of the event at a landmark time  $t_{LM}$  of 4 years. Up to landmark time, repeated information on 17 continuous biomarkers was generated according to linear mixed models, as described in Eq. 3 with identity link.

For each biomarker, measurement times were randomly generated according to a  $\mathcal{N}(0, 0.15)$  around 5 theoretical visit times at -4, -3, -2, -1 and 0 years prior to  $t_{LM}$ . Different shapes of individual trajectory were considering depending on the biomarker, although all followed an individual polynomial function of time (see Fig. S1 in supplementary material). Summary characteristics of each error-free marker trajectory were computed (as defined in “Methods” Section) leading to a total of 92 summaries statistics, stored in a vector  $\Gamma_i^0$ . An additional vector  $X_i^0$  of 10 time-independent covariates was generated at the landmark time: 5 according to a standard normal distribution and 5 according to binomial distribution with success probability of 0.5.

The risk of event after the landmark time was defined according to a proportional hazard model with  $\Gamma_i^0$  and

$X_i^0$ , and a Weibull distribution for the base hazard function, in order to not disadvantage the methods based on the Cox model. Five different scenarios were built according to the number of summaries actually associated to the event (18 or 4 summaries) and the form of the dependence function: biomarkers summaries were entered into the linear predictor either linearly, linearly with interactions across biomarkers, or non-linearly with polynomial functions and binarization of summaries. Details on the generation model and scenarios are respectively given in Section 2 and Table S1 of supplementary material.

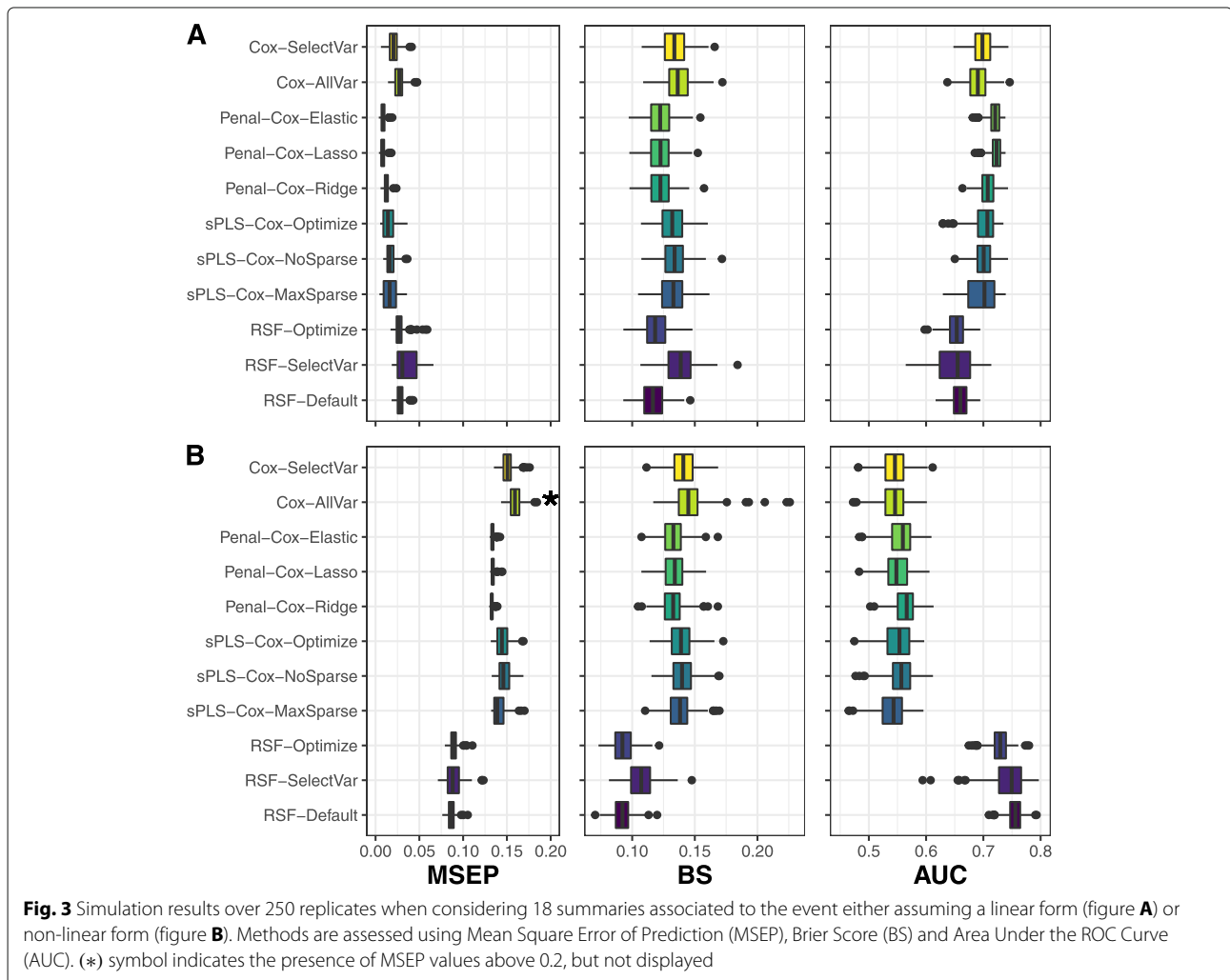
The target prediction was the probability of event up to a horizon of 3 years. The predictive performances of all the survival methods were compared on the external dataset using the BS and AUC previously introduced, as well as the Mean Square Error of Prediction (MSEP),  $MSEP = \frac{1}{N} \sum_{i=1}^N (\hat{\pi}_i - \pi_i^0)^2$ , which measures the average squared difference between the estimated probability  $\hat{\pi}_i$  and the true generated probability  $\pi_i^0$  over all individuals.

**Results**

Predictive performances for scenarios with 18 summaries are summarized in Fig. 3. The same figure for scenarios with 4 summaries is given in Fig. S2 of supplementary material.

When considering summaries entered linearly, the penalized-Cox provided the smallest BS and MSEP, and the highest AUC in both scenarios with 4 or 18 summaries associated with the event. When the relationships became increasingly complex (linear with interactions and non-linear), RSF provided better predictive performance than the other methods for both AUC, BS and MSEP regardless of the number of summaries considered.

This simulation study highlights that the penalized-Cox model provides more accurate predictions in the case of simple relationships between the predictors and the event while RSF outperforms the others in the case of complex relationships (no matter how many summaries are considered). In contrast, classical Cox model was systematically



outperformed by the other methods which points out the potential benefit of using advanced methods to predict the event in landmark approaches.

#### Individual prediction of death in primary biliary cholangitis

We first illustrated our method for predicting death among patients with primary biliary cholangitis (PBC). PBC is a chronic liver disease possibly leading to liver failure. For these patients, the only useful treatment is a liver transplantation [31], and prediction of the risk of death can be useful in that context for patient stratification. We focused on the widely known PBC data from a clinical trial [32] including repeated measures of 11 biomarkers (7 continuous and 4 binary), such as bilirubin value, albumin value or presence of hepatomegaly, and 3 additional demographic variables collected at the enrollment in the study (see Table S2 in supplementary material for the complete list). We aimed to predict the occurrence of death at horizon time  $t_{Hor} = 3$  using information collected up to landmark time  $t_{LM} = 4$  years on the  $N = 225$  patients still at risk at  $t_{LM}$  (see the flow chart Fig. S3 in supplementary material).

After a normalization for continuous markers which did not follow a gaussian distribution using splines [33], we modeled independently the markers according to generalized mixed models (see Eq. 3) with natural splines on time measurements to capture potentially complex behavior over time [34] (see Section 3.1 in supplementary material for details on the models).

We used a 10-fold cross-validation to compute the predictive performances of the individual predicted probabilities. The distribution of the event times did not differ across folds (Fig. S4 in supplementary material). For the superlearner, the optimal weights were determined in a second-layer 9-fold cross-validation. We repeated this process  $R = 50$  times for all methods to assess the variability of the results across different cross-validation partitions. RSF hyperparameters tuning (according to OOB error) is reported in supplementary material Fig. S5.

Predictive performances are displayed in Fig. 4A. All the prediction tools provided satisfying predictive performance for both BS (from 0.076 to 0.089 in mean) and AUC (from 0.73 to 0.87 in mean). Nevertheless, we found that Cox models gave much worst indicators, especially for AUC (the only ones below 0.80 in mean), illustrating the limits of classical methods compared to machine learning methods that handle high dimension and correlation. In this application, the most discriminating and accurate predictions were given by the Cox model with Lasso penalty according to BS (0.076 in mean) and AUC (0.87 in mean). Results from the superlearner did not show substantial improvement in predictive performance. The weights of the superlearner indicated that it was mostly driven by penalized Cox models and RSF (Fig. 4B).

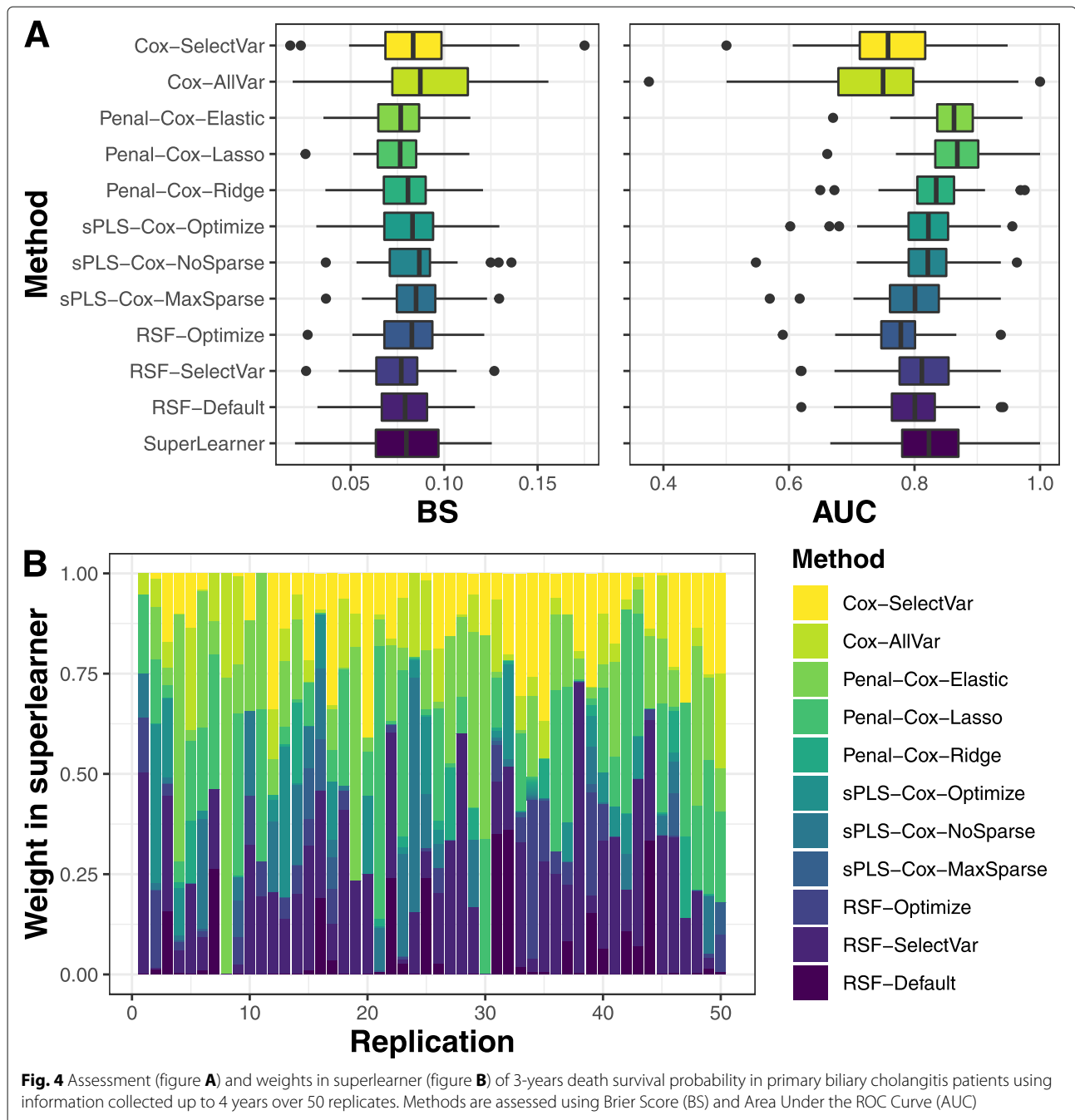
For comparison, we also developed predictive tools based on (1) only baseline information for the 11 biomarkers and 3 covariates, (2) information on the 3 covariates and the trajectory of one biomarker over time (either serum bilirubin, albumin or platelets). The predictive tools based only on baseline information provided poorer cross-validation BS (32% higher in mean over the methods) and AUC (8% lower in mean over the methods) nicely illustrating the gain in updating the biomarker information over follow-up (Fig. 5). The predictive performances were also worse when considering only repeated albumin or platelets with in mean 22% and 37% higher BS (1% and 11% lower AUC), respectively. In contrast, the predictive tools based on serum bilirubin (the main biomarker in PBC) provided similar performances as the multivariate predictive tool.

#### Individual prediction of 5-years death at 80 and 85 years old in the general population

In this second application, we aimed to predict the 5-year risk of death from any cause in the general older population at two different ages: 80 and 85 years old. We relied on the French prospective population-based aging cohort Paquid [35] which included 3777 individuals aged 65 years and older, and followed them up to more than 30 years with general health assessment every two to three years and continuous reporting of death. Beyond the individual quantification of the risk of death, our aim was to identify the main predictors of death and assess whether they differed according to age. The use of landmark models was perfectly adapted to this context with the definition of an age-specific prediction model. We chose to predict the 5-year risk of death from information on 9 markers of aging: depressive symptoms, 3 cognitive functions (general cognition, verbal fluency and executive function), functional dependency, incontinence, dyspnea, the live alone status, and polymedication as a global and easily collected marker of multimorbidity [36]. For each one, we focused on the trajectory over the last 5 years prior to the landmark age. In addition, we considered 18 other predictors including socio-demographic information (such as generation or sex) and medical history at the last visit prior to the landmark age (such as cardiovascular disease). Complete information on the markers and covariate definitions is given in Section 3.2 and Tables S3/S4 of supplementary material. The analysis was done on the samples of individuals still alive at  $t_{LM} = 80$  and  $t_{LM} = 85$ , and with at least one measure for each of the predictors resulting in  $N = 1561$  and  $N = 1240$  subjects for  $t_{LM} = 80$  and  $t_{LM} = 85$ , respectively (see flowchart Fig. S6 in supplementary material).

We used the exact same strategy as explained in the previous application for (i) modeling the trajectories of each marker except that time was the backward time

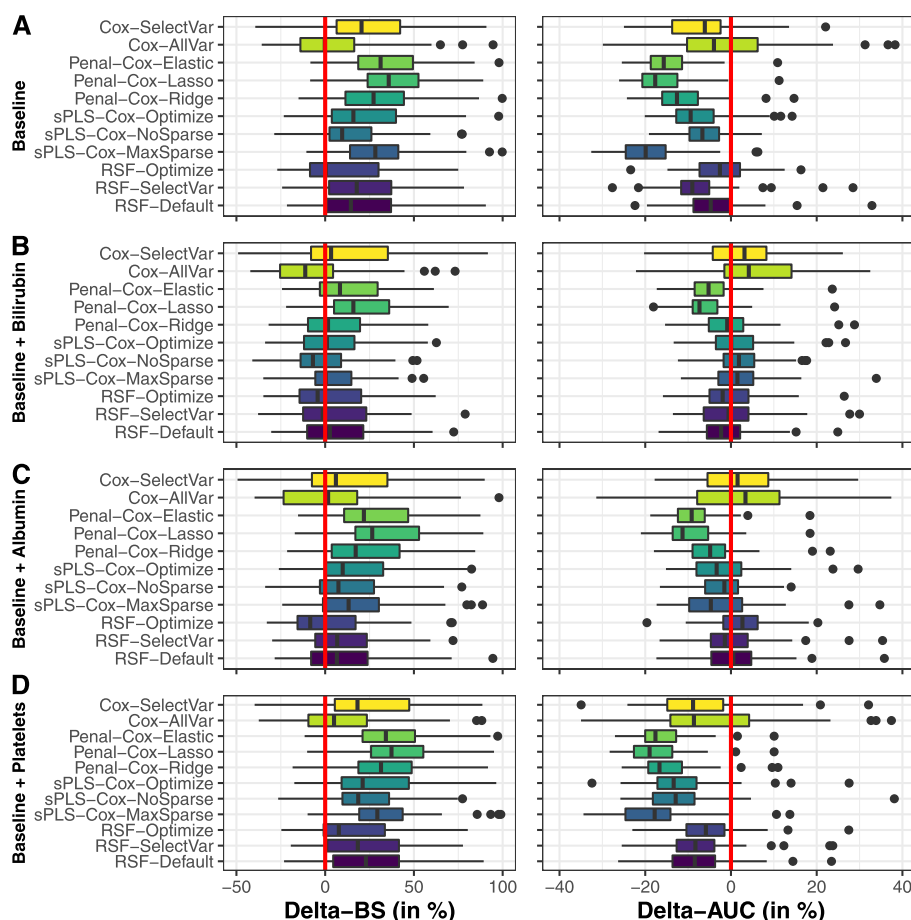




(from -5 to 0 years) from landmark; (ii) computing the external probabilities with a 10-fold cross-validation and computing the superlearner with an internal 5-fold cross-validation. The event time distribution did not differ across folds (see Figs. S7 and S8 in supplementary material). Note that due to the impossibility of using predictors with zero or near zero variance in sPLS-DR models, we removed from these models the following predictors: level of education, hearing, dementia, housing and dependency

(ADL). RSF hyperparameters tuning (according to OOB error) is reported in supplementary material Figs. S9 and S10.

Overall, the predictive performances of all the prediction models were very low with AUC ranging from 0.55 to 0.64 in mean and BS ranging from 0.123 to 0.135 in mean (see Figs. S11A and S12A in supplementary material) showing the difficulty to accurately predict the age-specific risk of all-cause death in the general



**Fig. 5** Assessment of 3-year survival probability in primary biliary cholangitis patients using baseline information on the 11 biomarkers and 3 covariates (figure **A**), baseline information and repeated measures collected up to 4 years of either serum bilirubin (figure **B**), albumin (figure **C**) or presence of platelets (figure **D**). The 10-fold cross-validation was replicated 50 times. The figure displays the difference (in percentage) of Brier Score (BS) and Area Under the ROC Curve (AUC) compared to the method using all the information with positive values for BS and negative values for AUC indicating a lower predictive accuracy

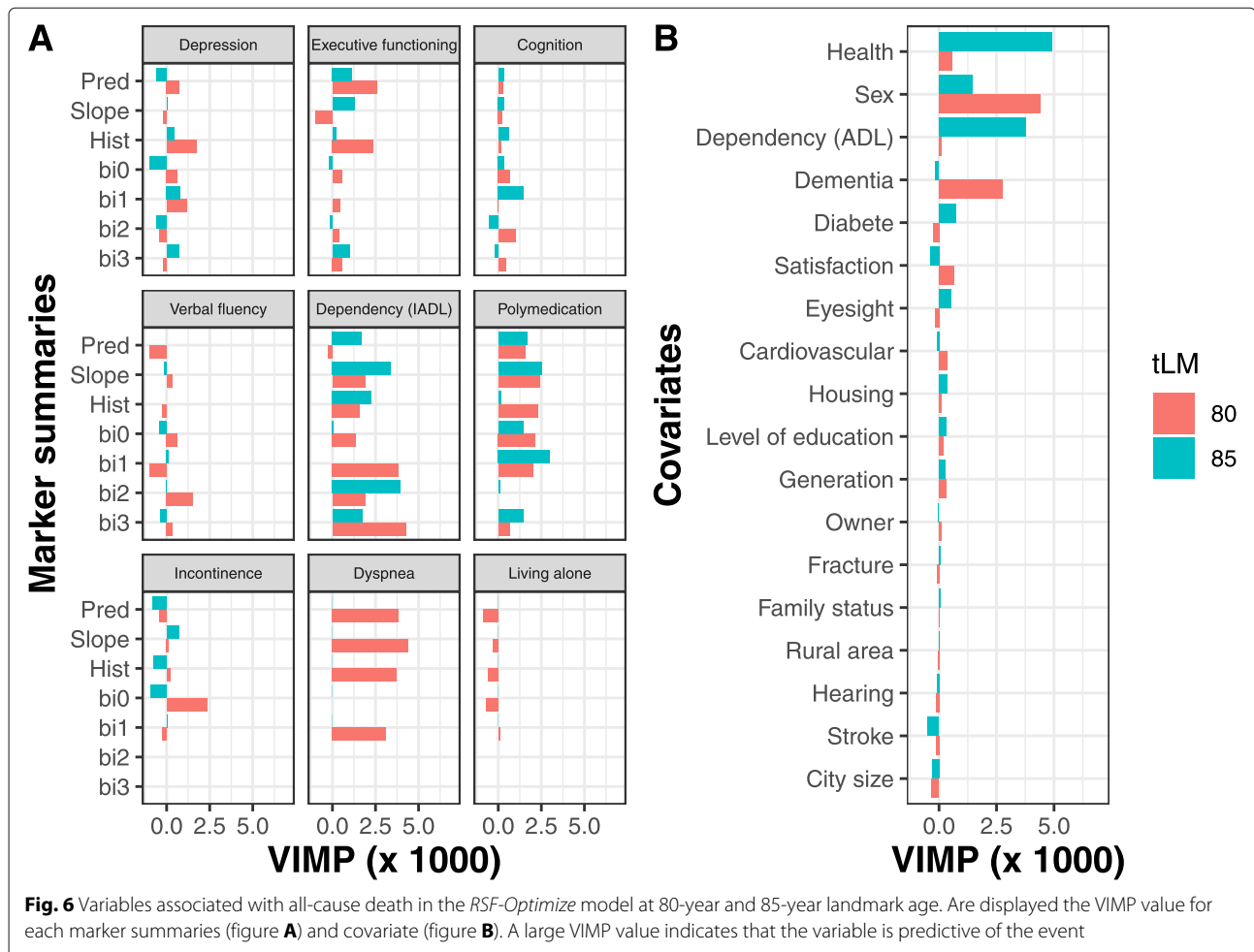
population. For both  $t_{LM} = 80$  and  $t_{LM} = 85$ , RSF and the superlearner (which was mostly driven by the RSF (see Figs. S11B and S12B in supplementary material) provided the lowest BS, whereas Cox with variable selection and penalized Cox models gave the highest AUC (0.66 in mean). Comparison to baseline information is also available in Fig. S13 in supplementary material.

This application mainly aimed at identifying and contrasting the main age-specific predictors of death at 80 and 85 years old. Fig. 6 reports the VIMP from the optimized RSF (variables selected by the Lasso are shown in supplementary material Fig. S14). The main predictors of 5-year death were mainly the trajectory of moderate functional dependency and polymedication both at 80 and 85 years old, dyspnea, sex and dementia at 80 years old as well as general self-assessment of health and severe dependency status at 85 years old. The predictors of 5-year death did not substantially differ between the two landmark times

for RSF, except for dyspnea, general self-assessment of health and sex.

## Discussion

We introduced in this paper an original methodology to compute individual dynamic predictions from a large number of time-dependent markers. We proposed to compute this prediction using a landmark approach combined with machine learning methods adapted to survival data. The idea was to incorporate a set of individual summaries of each marker trajectory (obtained in a preliminary longitudinal analysis) as well as other covariates in various prediction methods that could handle a large number of possibly correlated predictors, and complex associations. In addition to each prediction tool, we also proposed a superlearner adapted to time-to-event data, as a weighted mean of tool-specific predictions where weights were determined in an internal cross-validation



to provide a minimal Brier Score. Through an extensive simulation study, we showed that regularized Cox models and RSF provided better cross-validated predictive performance over standard Cox model in different scenarios where there was a large number of markers and/or complex associations with the event. This was also observed in two real case applications: a clinical setting where death was predicted from monitored markers in primary biliary cholangitis, and in a setting where all-cause age-specific death was predicted in the general population from main markers of aging. We precise that given the wellknown overfitting when assessing the predictive accuracy on the same sample as used for the training, we systematically assessed the predictive accuracy on an external sample in the simulations. However, in the real data analyses, in the absence of available external data, we used K-fold cross-validations (repeated 50 times to account for the variability due to the cross-validation partitioning).

Providing accurate predictions of health events that can exploit all the available individual information, even measured repeatedly over time, has become a major issue with

the expansion of precise medicine. After the first proposals of dynamic predictions from repeated marker information [1, 6], some authors have recently begun to tackle the problem of large dimension of longitudinal markers [18, 19, 22]. In comparison with this recent literature, our method has the advantage of (i) considering any nature of markers with measurement error while other considered only continuous outcomes [19], (ii) proposing the use of many summaries from the biomarkers as individual posterior computation from the longitudinal model (compared for instance to [22] who only include one or two summaries), (iii) exploiting the time-continuous information from survival data rather than discretized scale as in [22], and (iv) considering a vast variety of machine learning techniques as well as a superlearner rather than focusing only on one specific technique [18]. Our methodology does not limit to the specific model and techniques described in the paper, it allows the use of any relevant method at each step. For example, we suggested to capture individual trajectories using generalized mixed models, but we also used functional principal component analysis

[37] to characterize the individual variation of the trajectories using eigenfunctions leading to similar results (not shown here). We could also estimate the individual probability using other techniques such as deep learning [38] or random forests based on pseudo-observations [18]. Finally, although we considered for simplicity a single cause of event in this paper, our methodology could be extended to take into account several events through competing risks. For example, we could easily replace random survival forests by their extension that takes into account competing risks [30].

In our simulations and applications, we considered only a few dozens of markers repeatedly measured over time since this is already a challenging situation in individual dynamic prediction context where classical techniques are limited to a few markers. Yet, the method would also apply in a much higher dimensional context (e.g., with omics repeated data) or with a much larger amount of subjects. Indeed, our methodology primarily relies on prediction methods (random forests, penalized regressions, dimension reduction regressions) that were shown to scale very well in high-dimensional context [39]. The preliminary step we added to determine the set of summary features is a univariate mixed model performed independently on each marker. Therefore, it isn't affected by the number of markers. However, in high-dimensional contexts (highly large number of subjects and/or highly large number of markers), we anticipate the method to become computationally very intensive. Reducing the computational time in such high-dimensional contexts remains a future direction of research.

Our work presents the same limitations as any landmark approach. First, only the subjects at risk at the landmark time are considered which can induce a loss of efficiency. In addition, predictions from landmark approaches are not consistent since the time-varying covariate with the time-to-event are not linked at all times [40]. However, the landmark method in an extensive simulation study with one time-varying covariate, the landmark approach was shown to provide very good predictive performances compared to the joint modelling technique and better robustness to misspecification [7]. Finally our methodology is limited to the prediction of an event from a landmark time that is common over subjects or for a small number of common landmark times as done in the application. In other settings where any landmark time should be considered, our methodology would need to be adapted as it currently involves as many prediction tools and the number of landmark times which would result in a considerable increase of computational burden. A possible solution might be to define the prediction tools as a continuous function of the landmarks, following the super landmark models idea [41] but we leave such development for future research.

## Conclusions

By extending the landmark approach to the large dimensional and repeated setting, our methodology addresses a current major issue in biomedical studies with a complete methodology that has the assets of being (i) easy to implement in standard software (R code is provided at <https://github.com/anthonydevaux/hdlandmark> and more details are given in Section 4 in supplementary material) and (ii) generic as it can be used with any new machine learning technique adapted to survival data, any methodology to model repeated markers over time, any type of possible summary characteristics for the markers, and any number of markers.

## Abbreviations

AUC: Area under the ROC curve; BS: Brier score; MSEP: Mean square error of prediction; OOB: Out-of-bag; PBC: Primary biliary cholangitis; SPLS: sparse-Partial least square; sPLS-DR: Deviance residuals-based sparse-Partial least square; PSA: Prostate specific antigen; RSF: Random survival forests; VIMP: Variable importance

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-022-01660-3>.

**Additional file 1:** Additional information and results from simulations and applications.

## Acknowledgements

This work was presented at the 42st annual conference of the International Society for Clinical Biostatistics (abstract published at <https://easychair.org/smart-program/ISCB2021/2021-07-20.html>).

## Authors' contributions

All the authors contributed significantly to the work at different stages: conception and design (all the authors), acquisition of data (KP), analysis (AD, RG, CPL), interpretation (all authors), manuscript draft (AD, RG, CPL), critical revision (all authors). The authors read and approved the final manuscript.

## Funding

This work was funded by the French National Research Agency (grant number ANR-18-CE36-0004-01 for project DyMES) and within the framework of the PIA3 (Investment for the future) (project reference 17-EURE-0019). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

Anonymized Paquid data can be shared by reasonable request to the Paquid scientific committee at [celine.meillon@u-bordeaux.fr](mailto:celine.meillon@u-bordeaux.fr). PBC2 data are publicly available in R package `joinerML`. Scripts for replicating the analyses can be sent on demand to the corresponding author. Source R code is provided at <https://github.com/anthonydevaux/hdlandmark>.

## Declarations

### Ethics approval and consent to participate

Written informed consent was obtained for all participants and the ethics committee of the University Hospital in Bordeaux approved the research according to the principles embodied in the Declaration of Helsinki (see Helmer et al. [35]).

### Consent for publication

Not Applicable.

**Competing interests**

The authors have no competing interests.

**Author details**

<sup>1</sup>INSERM, BPH, U1219, Univ. Bordeaux, Bordeaux, France. <sup>2</sup>INRIA Bordeaux Sud-Ouest, Talence, France.

Received: 9 February 2021 Accepted: 15 June 2022

Published online: 11 July 2022

**References**

- Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostat (Oxford)*. 2009;10(3):535–49. <https://doi.org/10.1093/biostatistics/kxp009>.
- Sène M, Bellera CA, Proust-Lima C. Shared random-effect models for the joint analysis of longitudinal and time-to-event data: application to the prediction of prostate cancer recurrence. *J Soc Fr Stat*. 2014;155(1):134–55. Accessed 07 May 2014.
- Taylor JMG, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models: Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models. *Biometrics*. 2013;69(1):206–13. <https://doi.org/10.1111/j.1541-0420.2012.01823.x>.
- Paige E, Barrett J, Stevens D, Keogh RH, Sweeting MJ, Nazareth I, Petersen I, Wood AM. Landmark Models for Optimizing the Use of Repeated Measurements of Risk Factors in Electronic Health Records to Predict Future Disease Risk. *Am J Epidemiol*. 2018;187(7):1530–38. <https://doi.org/10.1093/aje/kwy018>.
- Sweeting MJ, Barrett JK, Thompson SG, Wood AM. The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study. *Stat Med*. 2017;36(28):4514–28. <https://doi.org/10.1002/sim.7144>.
- Rizopoulos D. Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data. *Biometrics*. 2011;67(3):819–29. <https://doi.org/10.1111/j.1541-0420.2010.01546.x>.
- Ferrer L, Putter H, Proust-Lima C. Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. *Stat Methods Med Res*. 2019;28(12):3649–66. <https://doi.org/10.1177/0962280218811837>.
- Van Houwelingen HC. Dynamic Prediction by Landmarking in Event History Analysis. *Scand J Stat*. 2007;34(1):70–85. <https://doi.org/10.1111/j.1467-9469.2006.00529.x>.
- Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin*. 2004;14(3):809–34.
- Maziarz M, Heagerty P, Cai T, Zheng Y. On longitudinal prediction with time-to-event outcome: Comparison of modeling options: Prediction Based on Longitudinal and Time-to-Event Data. *Biometrics*. 2017;73(1):83–93. <https://doi.org/10.1111/biom.12562>.
- Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J*. 2016;38(23):1805–14. <https://doi.org/10.1093/eurheartj/ehw302>.
- Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
- Lebedev AV, Westman E, Van Westen GJP, Kramberger MG, Lundervold A, Aarsland D, Soyninen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Simons A. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clin*. 2014;6:115–25. <https://doi.org/10.1016/j.nicl.2014.08.023>.
- Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011;39(5):. <https://doi.org/10.18637/jss.v039.i05>.
- Bastien P, Bertrand F, Meyer N, Maumy-Bertrand M. Deviance residuals-based sparse PLS and sparse kernel PLS regression for censored data. *Bioinformatics*. 2015;31(3):397–404. <https://doi.org/10.1093/bioinformatics/btu660>.
- Chun H, Keles S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Ser B Stat Methodol*. 2010;72(1):3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>.
- Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841–60. <https://doi.org/10.1214/08-AOAS169>.
- Zhao L, Murray S, Mariani LH, Ju W. Incorporating longitudinal biomarkers for dynamic risk prediction in the era of big data: A pseudo-observation approach. *Stat Med*. 2020;39(26):3685–99. <https://doi.org/10.1002/sim.8687>.
- Jiang S, Xie Y, Colditz GA. Functional ensemble survival tree: Dynamic prediction of Alzheimer's disease progression accommodating multiple time-varying covariates. *J R Stat Soc: Ser C: Appl Stat*. 2021;2449. <https://doi.org/10.1111/rssc.12449>.
- Ye W, Lin X, Taylor JMG. Semiparametric Modeling of Longitudinal Measurements and Time-to-Event Data-A Two-Stage Regression Calibration Approach. *Biometrics*. 2008;64(4):1238–46. <https://doi.org/10.1111/j.1541-0420.2007.00983.x>.
- Albert PS, Shih JH. On Estimating the Relationship between Longitudinal Measurements and Time-to-Event Data Using a Simple Two-Stage Procedure. *Biometrics*. 2010;66(3):983–87. [https://doi.org/10.1111/j.1541-0420.2009.01324\\_1.x](https://doi.org/10.1111/j.1541-0420.2009.01324_1.x).
- Tanner KT, Sharples LD, Daniel RM, Keogh RH. Dynamic survival prediction combining landmarking with a machine learning ensemble: Methodology and empirical comparison. *J R Stat Soc Ser A Stat Soc*. 2020. <https://doi.org/10.1111/rssa.12611>.
- Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982;38(4):963–74. <https://doi.org/10.2307/2529876>.
- Goeman JJ. L1 Penalized Estimation in the Cox Proportional Hazards Model. *Biom J*. 2009;52(1):70–84. <https://doi.org/10.1002/bimj.200900028>.
- Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381–97. <https://doi.org/10.1002/sim.5958>.
- Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *J Stat Softw*. 2012;50(11):. <https://doi.org/10.18637/jss.v050.i11>.
- Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues J-F, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks: Comparing Dynamic Predictive Accuracy of Joint Models. *Biometrics*. 2015;71(1):102–13. <https://doi.org/10.1111/biom.12232>.
- van der Laan MJ, Polley EC, Hubbard AE. Super Learner. *Stat Appl Genet Mol Biol*. 2007;6(1):. <https://doi.org/10.2202/1544-6115.1309>.
- Golmakani MK, Polley EC. Super Learner for Survival Data Prediction. *Int J Biostat*. 2020;16(2):20190065. <https://doi.org/10.1515/ijb-2019-0065>. Place: Berlin, Boston Publisher: De Gruyter.
- Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics*. 2014;15(4):757–73. <https://doi.org/10.1093/biostatistics/kxu010>.
- Kaplan MM. Primary Biliary Cirrhosis. *N Engl J Med*. 1996;335(21):1570–80. <https://doi.org/10.1056/NEJM199611213352107>.
- Murtaugh PA, Dickson ER, Van Dam GM, Malinchoc M, Grambsch PM, Langworthy AL, Gips CH. Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology*. 1994;20(1):126–34. <https://doi.org/10.1002/hep.1840200120>.
- Proust-Lima C, Philipps V, Liqueur B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lclmm. *J Stat Softw*. 2017;78(2):1–56. <https://doi.org/10.18637/jss.v078.i02>.
- Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol*. 2019;19(1):46. <https://doi.org/10.1186/s12874-019-0666-3>.
- Helmer C, Joly P, Letenneur L, Commenges D, Dartigues J-F. Mortality with Dementia: Results from a French Prospective Community-based Cohort. *Am J Epidemiol*. 2001;154(7):642–48. <https://doi.org/10.1093/aje/154.7.642>.
- Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154(9):854–64. <https://doi.org/10.1093/aje/154.9.854>.
- Yao F, Müller H-G, Wang J-L. Functional Data Analysis for Sparse Longitudinal Data. *J Am Stat Assoc*. 2005;100(470):577–90. <https://doi.org/10.1198/016214504000001745>.

38. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24. <https://doi.org/10.1186/s12874-018-0482-1>.
39. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* vol. 2. New-York: Springer; 2009.
40. Suresh K, Taylor JMG, Spratt DE, Daignault S, Tsodikov A. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biom J*. 2017;59(6):1277–300. <https://doi.org/10.1002/bimj.201600235>. Accessed 14 Apr 2022.
41. Houwelingen JCV, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. Monographs on statistics and applied probability, vol. 123. Boca Raton: CRC Press; 2012.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

