RESEARCH ARTICLE

Statistics in Medicine WILEY

Landmarking 2.0: Bridging the gap between joint models and landmarking

Hein Putter | Hans C. van Houwelingen

Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

Correspondence
Hein Putter, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
Email: h.putter@lumc.nl

The problem of dynamic prediction with time-dependent covariates, given by biomarkers, repeatedly measured over time, has received much attention over the last decades. Two contrasting approaches have become in widespread use. The first is joint modeling, which attempts to jointly model the longitudinal markers and the event time. The second is landmarking, a more pragmatic approach that avoids modeling the marker process. Landmarking has been shown to be less efficient than correctly specified joint models in simulation studies, when data are generated from the joint model. When the mean model is misspecified, however, simulation has shown that joint models may be inferior to landmarking. The objective of this article is to develop methods that improve the predictive accuracy of landmarking, while retaining its relative simplicity and robustness. We start by fitting a working longitudinal model for the biomarker, including a temporal correlation structure. Based on that model, we derive a predictable time-dependent process representing the expected value of the biomarker after the landmark time, and we fit a time-dependent Cox model based on the predictable time-dependent covariate. Dynamic predictions based on this approach for new patients can be obtained by first deriving the expected values of the biomarker, given the measured values before the landmark time point, and then calculating the predicted probabilities based on the time-dependent Cox model. We illustrate the approach in predicting overall survival in liver cirrhosis patients based on prothrombin index.

KEYWORDS
joint models, landmarking, longitudinal biomarkers, survival prediction

1 | INTRODUCTION

Biomarkers are commonly used in clinical research and treatment to monitor progression of patients. Prominent examples are PSA in prostate cancer,[1,2] CD4+ T-cell count and HIV-RNA in HIV-infected individuals,[3,4] and eGFR in patients with end-stage renal disease.[5,6] They are used to study the impact of (changes) in the biomarker on disease progression and survival, and to obtain updated prognosis for patients, based on observed marker values, that is, for dynamic prediction of survival.

Broadly speaking, two approaches are in widespread use for dynamic prediction based on longitudinally measured biomarkers. The first is the use of models that jointly characterize the development of the longitudinal biomarkers and the time to event.[7,8] The advantage of the joint modeling approach is that predictions based on joint models are quite efficient when the model is well specified, and there is software available that can fit these models and produce dynamic predictions in standard situations.[9,10]

The second approach is landmarking,[11,12] which is a pragmatic approach that avoids specifying a model for the longitudinal markers. The advantage of landmarking is that it is easy to implement; no specialized software is needed to obtain dynamic predictions from landmarking. The disadvantage is that it is less efficient than joint modeling, and can yield small bias when last observation carried forward is used and the biomarkers are coarsely observed.

The objective of this article is to bridge the gap between joint modeling and landmarking, and develop a method that improves on standard landmarking while avoiding complex integration over random effects, which makes joint modeling computationally demanding.

2 | NOTATION AND COMMON APPROACHES

We assume that we follow patients, indexed by $i = 1, \ldots, n$, from time $t = 0$ until an event (called death here) occurs. Let T_i be the time of death, and C_i an independent censoring time; define $\tilde{T}_i = \min(T_i, C_i)$ and the status indicator $D_i = I(T_i \leq C_i)$. There is a continuous biomarker process $X_i(t)$, defined as long as individual i is alive. This process is observed at observation times t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$. The observation times may be irregular, the number of observations may differ across subjects, but it is assumed that they are uninformative, that is, that they do not depend on unobserved marker values or unobserved characteristics. The observations have measurement error or day-to-day variation (white noise), and the actual observed measurements of $X_i(t)$ at t_{ij} are denoted by x_{ij}. Other covariates might be present, but will be ignored for the sake of simplicity; they can be included into each of the models we describe below in a straightforward way. The observations are $(\tilde{t}_i, d_i, \mathbf{x}_i)$, with $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,n_i})^\top$.

The objective is to use part of the information of the biomarkers of the patient to estimate the conditional probability that the patient is still alive after a predefined time window. More specifically, at a prediction time point s we want to estimate the conditional probability that the patient is still alive at time $s + w$, conditionally on being alive at time s and conditional on the history of the biomarkers up to time s, that is,

$$\pi_i(s + w \,|\, s) = P(T_i > s + w \,|\, T_i \geq s, \overline{x}_i(s)),$$

with $\overline{x}_i(s)$ denoting the history of all biomarker measurements up to s.

A Cox model with a time-dependent covariate $X(t)$

$$\lambda(t \,|\, X(t)) = \lambda_0(t) \exp(\beta X(t))$$

is helpful in understanding biology, but useless in predicting the future. The reason for this is that to obtain $\pi_i(s + w \,|\, s)$, based on information at the prediction time s one would need the future values of $X(t)$ after time s. Unless the time-dependent covariate is exogenous one would not know these future values.

2.1 | Joint modeling

One way to be able to derive dynamic predictions is to make a model for how $X(t)$ might change over time, given knowledge at the prediction time. For this it is assumed that $X_i(t)$ follows a Gaussian process with mean $\mu_i(t)$, possibly depending on covariates, and covariance function $C(t_1, t_2) = \text{cov}(X_i(t_1), X_i(t_2))$. A popular choice is a linear mixed model like $X_i(t) = \beta_0 + b_{i0} + (\beta_1 + b_{i1})t$ with fixed effects β_0 and β_1 (possibly depending on covariates) and random effects (b_{i0}, b_{i1}) assumed to be bivariate normal with mean zero. $X_i(t)$ is observed at t_{ij} with independent measurement errors e_{ij}. The standard joint model assumes that the hazard of dying at time t depends on the current value of the biomarker, for instance given by the proportional hazards model

$$\lambda(t \,|\, \overline{x}_i(t)) = \lambda_0(t) \exp(\beta \mu_i(t)).$$

Other options, where the hazard depends on the random effects directly, or on the slope or the area under the curve are also possible. Rizopoulos[13] discusses how to obtain dynamic prediction from such joint models, and software is available in the **JM** and **JMbayes** packages.[9,10]

Simulation studies have shown that the joint model efficiently estimates the underlying parameters, when the model is correctly specified,[14] and that it is reasonably robust against modest misspecification of the dependence function and against modest deviations of proportional hazards, but that it is quite sensitive to misspecification of the longitudinal trajectory.[15]

2.2 | Landmarking

Landmarking[11,12] avoids modeling the marker process. The idea behind landmarking is to select, for a given landmark time point $t_{\text{LM}} = s$, all subjects alive and under follow-up at time s. The time-dependent information until time s is summarized in some way. Possibilities to summarize the history are the last observed measurement (last observation carried forward, LOCF), or the last observed measurement and the slope. An extension is to use the "age" of the last observation (difference between s and last observed time before s) as additional covariate. This summary of the time-dependent covariate is subsequently used in a Cox model in the landmark data set. When interest is in estimating the dynamic prediction probability at $s + w$, it is common to apply administrative censoring at $s + w$. This administrative censoring, or "stopped Cox"[16] is introduced to make the procedure robust against violations of proportional hazards, although for long term prediction (large w) time-varying effects might lead to some bias. A concern is that the staleness ("aging") of the predictor when using LOCF leads to a mismatch between the true underlying value of the biomarker and the last observation. This measurement error leads to violation of proportional hazards,[12] which would call for modeling time-varying effects, but it is challenging to find adequate models while at the same time avoiding the threat of overfitting.

A number of approaches have been proposed to improve the simple LOCF landmarking approach. One of them is a two-stage approach,[17-19] where the data of the time-dependent covariate(s) before the landmark prediction time-point s are used and a mixed model is fit to those data (or all data). Then the empirical Bayes best linear unbiased predictor (BLUP) is used as a predictor at s. It is called "error free" but that could be too optimistic. It partly solves the staleness problem of the predictor at t_{LM}, but does not take care of the problem that the effect of the last measured biomarker before time s typically becomes smaller as t gets more removed from s, leading to a decay of the $\beta(t)$.

3 | LANDMARKING 2.0: GETTING CLOSER TO THE JOINT MODEL

The joint model approach leads to the following model for the conditional survival:[20]

$$\pi_i(s + w \mid s) = \mathbf{E}\left[\exp\left(-\int_s^{s+w} \lambda_0(t) \exp(\beta X_i(t))\mathrm{d}t\right) \mid T_i \geq s, \overline{x}_i(s)\right]. \tag{1}$$

Following Tsiatis et al[3] in their treatment of measurement errors in survival analysis, see also Andersen and Liestøl,[21] the conditional survival can by approximated by

$$\pi_i(s + w \mid s) \approx \exp\left(-\int_s^{s+w} \tilde{\lambda}_0(t) \exp\left(\tilde{\beta}\mathbf{E}[X_i(t) \mid T_i \geq t, \overline{x}_i(s)]\right)\mathrm{d}t\right). \tag{2}$$

Note that the regression coefficient $\tilde{\beta}$ and baseline hazard $\tilde{\lambda}_0(t)$ in the approximation (2) differ from the original ones in Equation (1). Also note the conditioning on $T_i \geq t$ in (2), rather than on $T_i > s$, by definition of the hazard. We expect the approximation in Equation (2) to be accurate when β and the baseline hazard are not too large and when $X(t)$ is not too variable. The approximation in (2) leads to the following proposal for what we call *landmarking 2.0*:

1. Define and fit a working Gaussian process with trend $\mu(t)$ and covariance matrix $C(t_1, t_2)$ of the observed X_{ij}. Note that it is not assumed that $X_i(t)$ follows a Gaussian process, although a transformation of the longitudinal measurements in order to make it approximately normal before fitting the model is probably wise anyway.

2. Use the fitted Gaussian process to estimate $E\{X_i(t) \mid T_i \geq t, \overline{x}_i(s)\}$ for $t \geq s$ by least squares yielding the predictable time-dependent covariate $\hat{x}_i(t \mid s)$, given by

$$\hat{x}_i(t \mid s) = \mu_2 + \Sigma_{12}^{\top}\Sigma_{11}^{-1}(\overline{x}_i(s) - \mu_1),$$

at each of the event time points in the data, where μ_1 and μ_2 denote the means of the fitted Gaussian process, evaluated at the observed measurement time points before s and the event time points after s, respectively, and Σ_{11}, Σ_{12}, and Σ_{22} the relevant sub-matrices of the variance-covariance matrix at the collection of those time points.

3. Fit a landmark Cox model with a fixed effect of the time-dependent covariate $\hat{x}_i(t \mid s)$, yielding estimates $\hat{\beta}$ and $\hat{\lambda}_0(t)$ of $\tilde{\beta}$ and $\tilde{\lambda}_0(t)$ in Equation (2).

4. Use the resulting landmark Cox model to obtain dynamic predictions for a new patient with observed $\overline{x}^*(s)$ by
   – Using the Gaussian process again to estimate $E\{X^*(t) \mid T \geq t, \overline{x}^*(s)\}$ for $t \geq s$ by least squares yielding the predictable time-dependent covariate $\hat{x}^*(t \mid s)$. Let μ_1 and μ_2 denote the means of the fitted Gaussian process, evaluated at the observed measurement time points before s and the event time points after s, respectively, and Σ_{11}, Σ_{12}, and Σ_{22} the relevant sub-matrices of the variance-covariance matrix at the collection of those time points, then

$$\hat{x}^*(t \mid s) = E\{X^*(t) \mid T \geq t, \overline{x}^*(s)\} = \mu_2 + \Sigma_{12}^{\top}\Sigma_{11}^{-1}(\overline{x}^*(s) - \mu_1).$$

   – Calculating the predicted hazard increments $\hat{\lambda}_0(u)\exp\{\hat{\beta}\hat{x}^*(u \mid s)\}$ for each event time point u between s and $s + w$ in the data.
   – The estimated conditional survival probability is given by

$$\hat{\pi}(s + w \mid s) = \exp\left[-\sum_{s < u \leq s+w} \hat{\lambda}_0(u)\exp\{\hat{\beta}\hat{x}^*(u \mid s)\}\right].$$

The approach is obviously more complex than naïve landmarking, but it is computationally considerably less challenging than joint modeling, because it avoids latent variables and integration over random effects. It gives a robust estimate of the survival given the predictable $\hat{x}(t \mid s)$. Note that the first two steps above could be replaced by any other approach that would give estimates of $\hat{x}_i(t \mid s)$ for $t \geq s$. Later we will use revival modeling[22] for this purpose. Landmarking 2.0 might be less efficient than the joint model, but it allows closer inspection and direct modeling of the trajectories of the survivors before estimating the regression parameters of the survival model. Note that the BLUP approach[17-19] is similar in spirit, but uses $\hat{x}_i(s \mid s)$ rather than $\hat{x}_i(t \mid s)$ in the landmark model.

As a working longitudinal model, we propose to take a variance components approach related to an autoregressive model, as also used in Dempsey and McCullagh.[22] This involves specifying a model for the trend $\mu(t)$ and one for the temporal covariance $C(t_1, t_2) = \mathrm{cov}(X(t_1), X(t_2))$. For the temporal covariance we follow[22] by taking as variance components a between individuals variance σ_1^2, a within individuals variance σ_2^2 with a temporal correlation $\exp(-\lambda|t_1 - t_2|)$, and a white noise error component σ_3^2, leading to

$$C(t_1, t_2) = \sigma_1^2 + \sigma_2^2 \exp(-\lambda|t_1 - t_2|) + \sigma_3^2 \mathbf{1}\{t_1 = t_2\}.$$

Other options, depending on the fit of the model, are of course possible. The mode can be fitted with standard software for linear mixed effects models, such as the R package **nlme**.

3.1 | Revival

The working Gaussian process is not the only way to obtain estimates $\hat{x}(t \mid s)$ to be used in the landmark Cox model. Another interesting way to achieve the same goal is to base $\hat{x}(t \mid s)$ on the revival approach.[22] In this approach, a longitudinal model is used for the biomarker, backwards in time from the time of death of the individual. Bayes' formula can be used to obtain dynamic predictions of $\pi_i(s + w \mid s)$. A problem that has to be dealt with is the possibility of censoring, that is, of not observing the time of death. We follow here the approach suggested by the commentary on the paper by Dempsey and McCullagh by van Houwelingen.[23] We start by giving some details on the model, then show how the model

can be used to obtain dynamic predictions, and continue to illustrate how the same model can be used to obtain estimates of the predictable time-dependent covariate $\hat{x}(t \mid s)$, to be used in landmarking 2.0.

3.1.1 | The revival model

Following van Houwelingen,[23] we start by defining an observation limit τ, and denote the subset of "dead" subjects by those that died before τ, and the subset of "survivors" by those that were alive at time τ (including those that are observed to die after time τ). We define and fit separate models for the longitudinal markers of the dead subjects and the survivors. Subjects that were censored before τ are not included in either model. They are used later on when obtaining and assessing dynamic prediction probabilities. For subject i belonging to the subset of dead people, let t_i be the time of death, denote $u = t_i - t$ the reverse time to death of subject i, and define the time-reversed process of subject i as $Z_i(u) = X_i(t_i - u)$. The distribution of this time-reversed process may depend on subject-specific factors like age, sex, and treatment. For all subjects belonging to the subset of survivors, we denote $u = \tau - t$ the reverse time to the observation limit, and define the time-reversed process as $Z_i(u) = X_i(\tau - u)$.

3.1.2 | Dynamic prediction using revival

When models have been defined for the time-reversed marker processes, backwards in time from the time of death t_i of the dead subjects, and from the horizon τ for the survivors, conditional probabilities $\pi_i(s + w \mid s)$ can be obtained by Bayes' rule, after having obtained an estimate of the conditional survival probabilities $P(T > t \mid T > s)$. The latter may be obtained from a Kaplan-Meier estimate, possibly stratified by covariates, or a simple baseline Cox model. Since these all yield estimates that concentrate their probability mass on the observed event time points, let $t > s$ be such an event time point. Then Bayes' rule gives

$$P(T = t \mid T > s, \overline{X}(s)) = \frac{P(\overline{X}(s) \mid T = t, T > s) \cdot P(T = t \mid T > s)}{\sum_{u > s} P(\overline{X}(s) \mid T = u, T > s) \cdot P(T = u \mid T > s)}. \tag{3}$$

Note that in the above, the sum in the denominator is over all event time points $u > s, u < \tau$, plus the predefined horizon τ. Note also that $P(\overline{X}(s) \mid T = t, T > s)$ can be simplified to $P(\overline{X}(s) \mid T = t)$. With slight abuse of notation, we denote by P either a discrete probability or a (joint) density. For each event time point, $P(\overline{X}(s) \mid T = t)$ is the joint density of the observed marker values before the landmark time point s, which can be obtained from the distribution of the time-reversed marker process $Z(t - u)$. For $t = \tau$, the joint density of the observed marker values before the landmark time point s can be obtained from the distribution of the process $Z(\tau - u)$.

3.1.3 | Landmarking 2.0 using revival

Equation (3) describes how to obtain dynamic prediction probabilities of survival, given observed marker values. We shall refer to this as *direct* dynamic prediction using revival. The time-reversed marker processes also imply conditional distributions of the marker at time points $t > s$, given survival until time t and the observations of the marker process at time points before time s. Here we want to extract the conditional expectations of $X(t)$, given the observed history until time $\overline{x}(s)$, and given $T \geq t$. Using the same abuse of notation (P denoting either density or probability), we have

$$P(X(t) = x \mid T \geq t, \overline{x}(s)) = \int_t^\tau P(X(t) = x, T = u \mid T \geq t, \overline{x}(s)) \, du$$

$$= \int_t^\tau P(X(t) = x \mid T = u, T \geq t, \overline{x}(s)) \cdot P(T = u \mid T \geq t, \overline{x}(s)) \, du$$

$$= \int_t^\tau P(X(t) = x \mid T = u, \overline{x}(s)) \cdot P(T = u \mid T \geq t, \overline{x}(s)) \, du.$$

When working with Cox models or nonparametric models for the time-to-event distribution, the integral is in fact a sum over event time points u, including the separate time point τ, representing the survivors. This implies that the

conditional expectation $\hat{x}(t \mid s)$ is given by

$$
E\left(X(t) \mid T \geq t, \overline{x}(s)\right) = \int xP(X(t) = x \mid T \geq t, \overline{x}(s))\, dx
$$

$$
= \sum_{t \leq u \leq \tau} E\left(X(t) \mid T = u, \overline{x}(s)\right) \cdot P(T = u \mid T \geq t, \overline{x}(s)).
$$

The last term can be written as

$$
P(T = u \mid T \geq t, \overline{x}(s)) = \frac{P(\overline{x}(s) \mid T = u) \cdot P(T = u \mid T \geq t)}{\sum_{u' \geq t} P(\overline{x}(s) \mid T = u') \cdot P(T = u' \mid T \geq t)},
$$

similar to (3), the direct revival dynamic prediction probability of dying at time u. Both u and u' include τ. Furthermore, the first term, $E\left(X(t) \mid T = u, \overline{x}(s)\right)$ can be obtained from the joint distribution of $(\overline{X}(s), \overline{X}(s, u])$, given $T = u$. Here $\overline{X}(s, u]$ refers to the vector of $X(t)$'s for the event times t in $(s, u]$. If we denote this distribution as multivariate normal with mean vector $\begin{pmatrix} \mu_s \\ \mu_{su} \end{pmatrix}$, and covariance matrix $\begin{pmatrix} \Sigma_{ss} & \Sigma_{s,su} \\ \Sigma_{su,s} & \Sigma_{su,su} \end{pmatrix}$, then we obtain

$$
E\left(\overline{X}(s, u] \mid T = u, \overline{x}(s)\right) = \mu_{su} + \Sigma_{su,s}\Sigma_{ss}^{-1}\left(\overline{x}(s) - \mu_s\right).
$$

All this implies the following procedure to calculate $\hat{x}(t \mid s) = E\left(X(t) \mid T \geq t, \overline{x}(s)\right)$ based on revival, for use in landmarking 2.0: first calculate the direct dynamic prediction probabilities $P(T = u \mid T \geq t, \overline{x}(s))$, then loop over event time points $u > s$, including $u = \tau$, and

- Calculate conditional expectations and variances, given $T = u$, of $(\overline{X}(s), \overline{X}(s, u])$, yielding expectation vector $\begin{pmatrix} \mu_s \\ \mu_{su} \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \Sigma_{ss} & \Sigma_{s,su} \\ \Sigma_{su,s} & \Sigma_{su,su} \end{pmatrix}$.

- Calculate $E\left(\overline{X}(s, u] \mid T = u, \overline{x}(s)\right) = \mu_{su} + \Sigma_{su,s}\Sigma_{ss}^{-1}\left(\overline{x}(s) - \mu_s\right)$.

- Combine elements $E\left(X(t) \mid T = u, \overline{x}(s)\right)$ of these with $P(T = u \mid T \geq t, \overline{x}(s))$ and sum over $u \in (t, \tau]$ to obtain $E\left(X(t) \mid T \geq t, \overline{x}(s)\right)$.

4 | ILLUSTRATION

We will illustrate our methods using data from the CSL-1 trial, conducted in Copenhagen in 1962 to 1969, randomizing patients with histologically verified liver cirrhosis to placebo or prednisone. The subset used in this article consists of 488 patients, 251 in the prednisone and 237 in the placebo arm. Figure 1 shows the Kaplan-Meier estimate of overall survival for all subjects in the trial, and the reverse Kaplan-Meier estimate of the censoring distribution in the trial, both by randomized treatment.

The longitudinal marker of interest is the prothrombin index, a composite blood coagulation index related to liver function, measured initially at 3-month intervals and subsequently at roughly 12-month intervals. The prothrombin measurements over time for all patients in the trial by randomized treatment are shown in Figure 2, along with a loess smoothed average.

A Gaussian process was fitted on the prothrombin measurements excluding $t = 0$, where the mean $\mu(t)$ was fitted using different linear trends for the two treatment arms, but a common covariance function was used for both treatments. The estimated linear trend was $69.03(\pm 1.69) + 2.19(\pm 0.37)t$ for the placebo arm (estimate $\pm$ standard error) and $80.57(\pm 1.64) + 1.03(\pm 0.34)t$ for the prednisone arm, with covariance parameters as shown in Table 1.

4.1 | Revival

Figure 3 shows spaghetti plots of the prothrombin measurements in reverse time, separately for the placebo and prednisone patients, and separately for subjects that died within the observation limit of $\tau = 9$ years and subjects that were alive at τ. A total of 36 patients were censored before time τ.
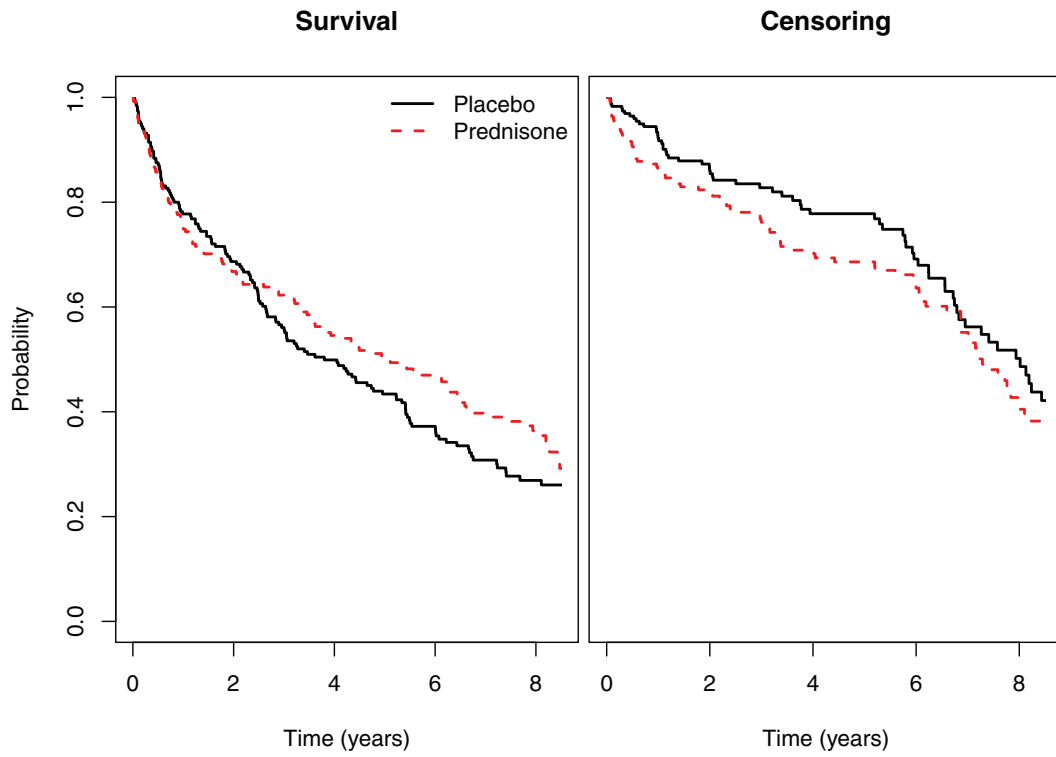
Survival **Censoring**

FIGURE 1 Kaplan-Meier estimates of overall survival (left) and censoring distribution (right)

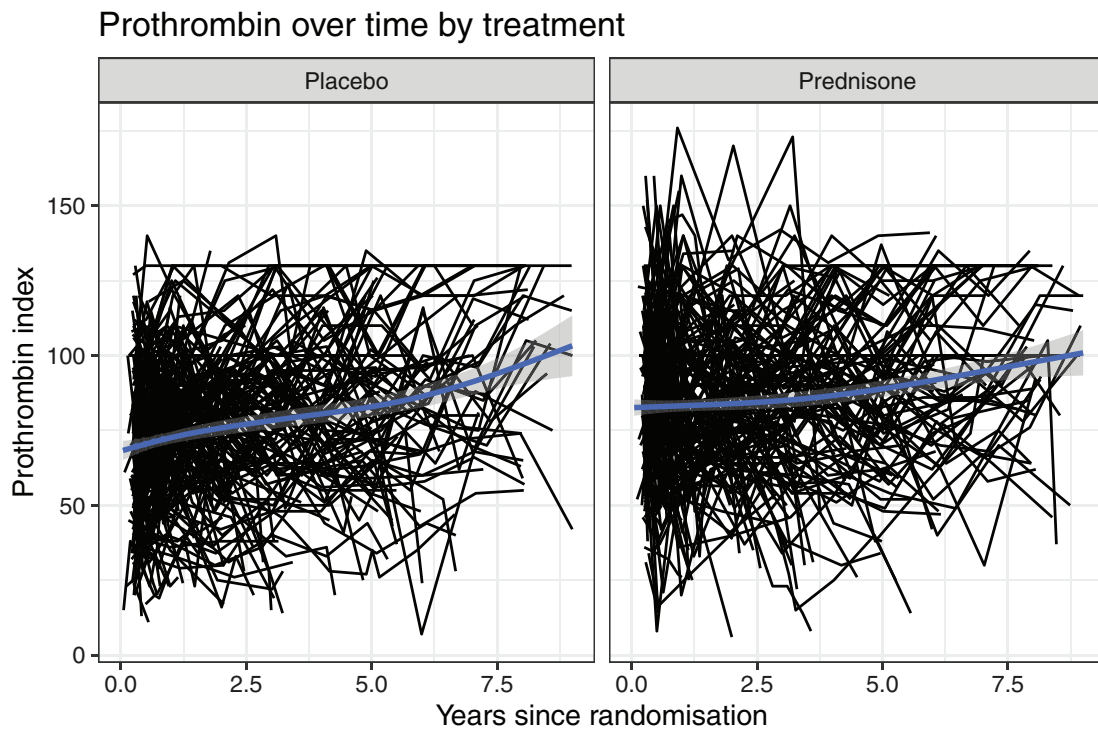## Prothrombin over time by treatment

FIGURE 2 Spaghetti plot of the prothrombin values over time

TABLE 1 Estimated covariance parameters of the Gaussian process

Component	Variance	Estimate	Standard error
Between individuals	σ_1^2	308.4	39.1
Within individuals	σ_2^2	240.8	55.8
Temporal decay parameter	λ	0.52	0.49
White noise	σ_3^2	184.3	50.0

FIGURE 3 Spaghetti plots of the prothrombin measurements in backward time, separately for the placebo and prednisone patients, and separately for subjects that died within $\tau = 9$ years ("dead") and subjects that were alive at 9 years ("survivor")

Denote A_i as the treatment indicator (0 = placebo, 1 = prednisone). A Gaussian process for the time-reversed marker process was used, with

$$\mathbf{E}(Z_i(u) \mid T = t_i, A_i) = \beta_0 + \beta_1 A_i + \beta_2 u + \beta_3 \log(u + \varepsilon) + \beta_4 t_i,$$
$$\text{cov}(Z_i(u), Z_i(u') \mid T = t_i) = \sigma_1^2 + \sigma_2^2 \exp(-\lambda|u - u'|) + \sigma_3^2 \mathbf{1}\{u = u'\},$$

for subjects that died before τ (t_i being the time of death of subject i), and with

$$\mathbf{E}(Z_i(u) \mid T > \tau, A_i) = \tilde{\beta}_0 + \tilde{\beta}_1 A_i + \tilde{\beta}_2 u + \tilde{\beta}_3 \log(u + \varepsilon),$$
$$\text{cov}(Z_i(u), Z_i(u') \mid T > \tau) = \tilde{\sigma}_1^2 + \tilde{\sigma}_2^2 \exp(-\tilde{\lambda}|u - u'|) + \tilde{\sigma}_3^2 \mathbf{1}\{u = u'\},$$

for subjects that were alive and under follow-up at time τ. For ε we took 1 day. The results are shown in Table 2.

Figure 4 illustrates the mean model, for patients in both treatment arms, dying at 3, 6, and 9 years, and surviving until τ.

TABLE 2 Estimates of the longitudinal time-reversed models

Parameter	Died		Censored	
	Estimate	SE	Estimate	SE
Intercept	66.39	2.57	95.85	4.51
t_i	1.73	0.56		
Revival (u)	−1.79	0.56	−1.39	0.68
$\ln(u + \delta)$	4.58	0.45	−1.65	1.51
Prednisone	8.37	2.55	9.53	5.34
Between individuals (σ_1^2)	221.5	53.8	202.4	98.5
Within individuals (σ_2^2)	243.6	70.7	191.1	282.2
Temporal decay parameter (λ)	0.62	0.62	0.35	2.28
White noise (σ_3^2)	161.9	52.6	161.9	209.4
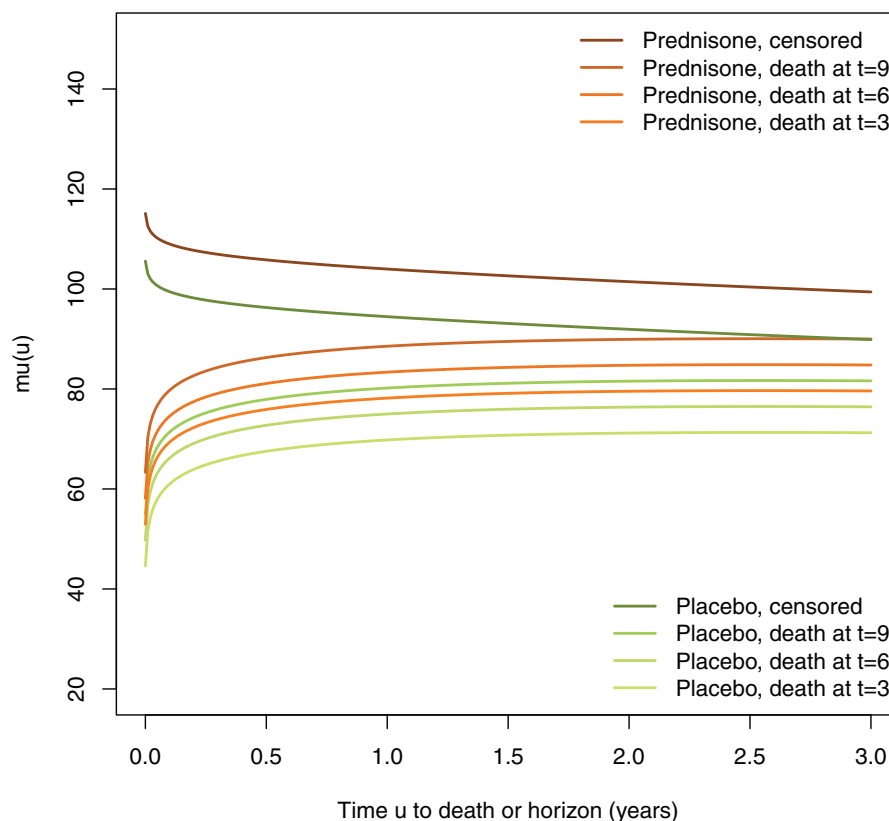
FIGURE 4 Model-based means for patients in both treatment arms, dying at 3, 6, and 9 years, and surviving until τ

4.2 | Dynamic prediction

Our aim is to illustrate our new proposed method in obtaining dynamic prediction probabilities, and to compare these dynamic prediction probabilities with those obtained by other methods. For this purpose we fix the prediction time point at $s = 3$ years and the prediction window to $w = 2$ years. Using the marker values up to s, the following methods are considered for estimating $\pi_i(s + w \mid s)$.

- Joint model (JM): Joint model, where the linear mixed effects model used fixed and random intercepts (unstructured), separately for the two treatment arms, and a proportional hazards model with treatment for the survival part and piecewise constant baseline hazard, using the JM package.[9]

- Revival: Direct revival, using Equation (3).
- LOCF: Last observation carried forward, this is the naïve landmark method, where at time s the last observed marker value before time s is used in a Cox model.
- $\hat{x}(s \mid s)$ (Xhats): This is the BLUP method,[17-19] based on the fitted working Gaussian process.
- $\hat{x}(t \mid s)$ (Xhat): The newly proposed landmark method, with $\hat{x}(t \mid s)$ based on the fitted working Gaussian process.
- $\hat{x}(t \mid s)$ based on revival (Xhatrevival): The newly proposed landmark method, with $\hat{x}(t \mid s)$ based on revival model.

Dynamic prediction probabilities were obtained by leave-one-out cross-validation; for each subject, the above models were fitted on data with the subject left out and subsequently used to obtain the predicted probability for that subject.

Figure 5 shows the evolution of $\hat{x}(t \mid s)$ for each of the patients in the CSL-1 trial, based on their observed marker values until the landmark time $s = 3$ years, by randomized treatment. The $\hat{x}(t \mid s)$ in the top row of Figure 5 is based on the fitted working Gaussian process, while those in the bottom row are based on the revival model.
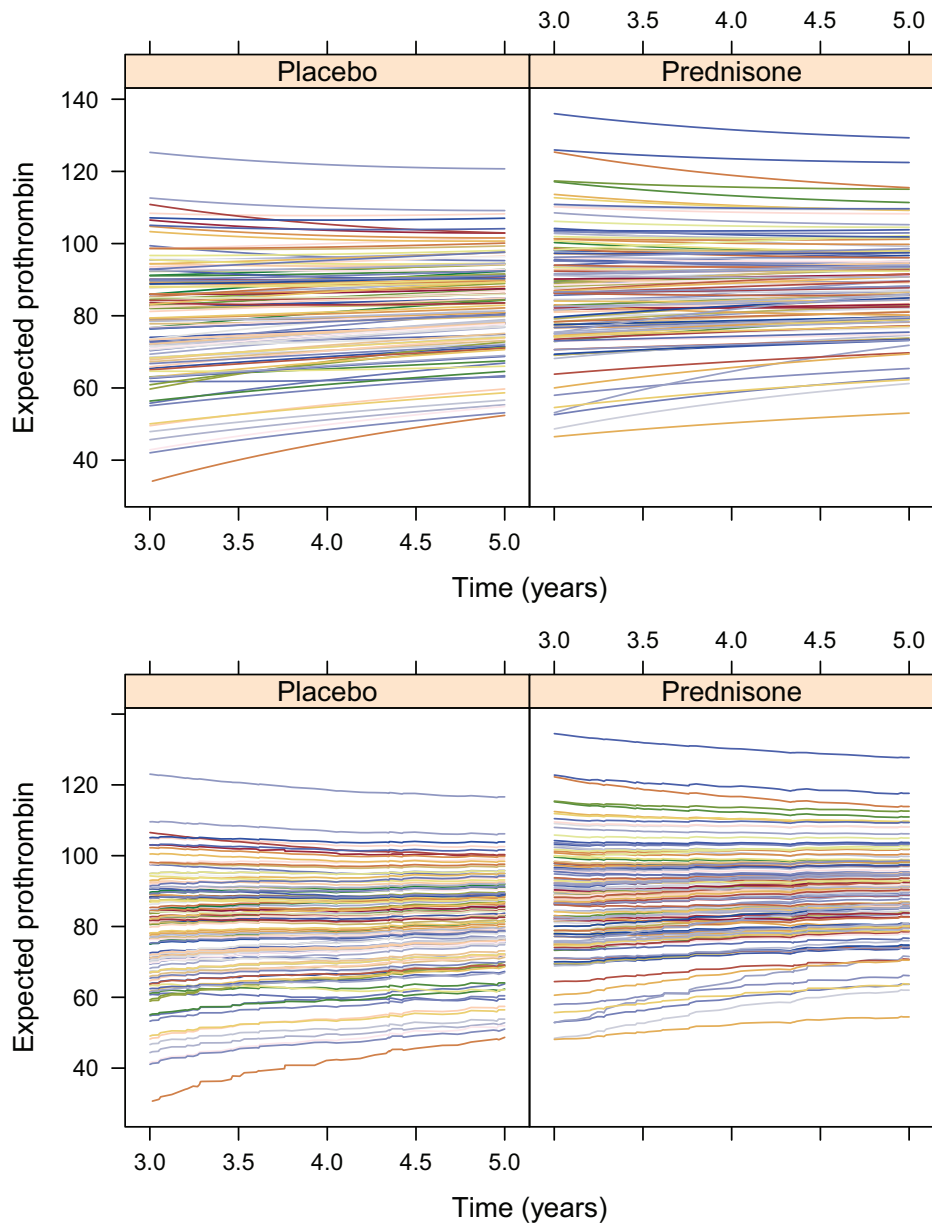

FIGURE 5 Evolution of $\hat{x}(t \mid s)$ for $s < t \leq s + w$ for all patients by treatment, based on the Gaussian process (top row) and the revival model (bottom row)

Figure 6 shows a matrix plot of the cross-validated dynamic predictions obtained from the different approaches.

Figure 6 reveals an aspect already noted,[23] namely that the revival models are not well calibrated. The cross-validated prediction probabilities of the direct revival model are much more narrowly distributed around its mean than the other methods, while those of the landmarking based on revival (Xhatrevival) seem to have a somewhat lower average than the other methods. The miscalibration of the direct revival model could be due to a misspecification of the revival models (the longitudinal models in reverse time), leading to incorrect prediction probabilities after using Bayes' rule, as in Equation (3).

In order to compare the predictive information of the different methods, we transformed the original cross-validated predicted probabilities using the complementary log-log transformation. We then entered each of the transformed cross-validated dynamic prediction probabilities in a univariate proportional hazards model in the landmark data, using administrative censoring at the horizon. The results are shown in Table 3.

It can be seen that landmarking 2.0 with $\hat{x}(t\,|\,s)$ based on revival as predictable time-dependent covariate has the highest univariate χ^2 value (χ^2 column). Landmarking 2.0 ($\hat{x}(t\,|\,s)$) and the BLUP method ($\hat{x}(s\,|\,s)$) are very close with
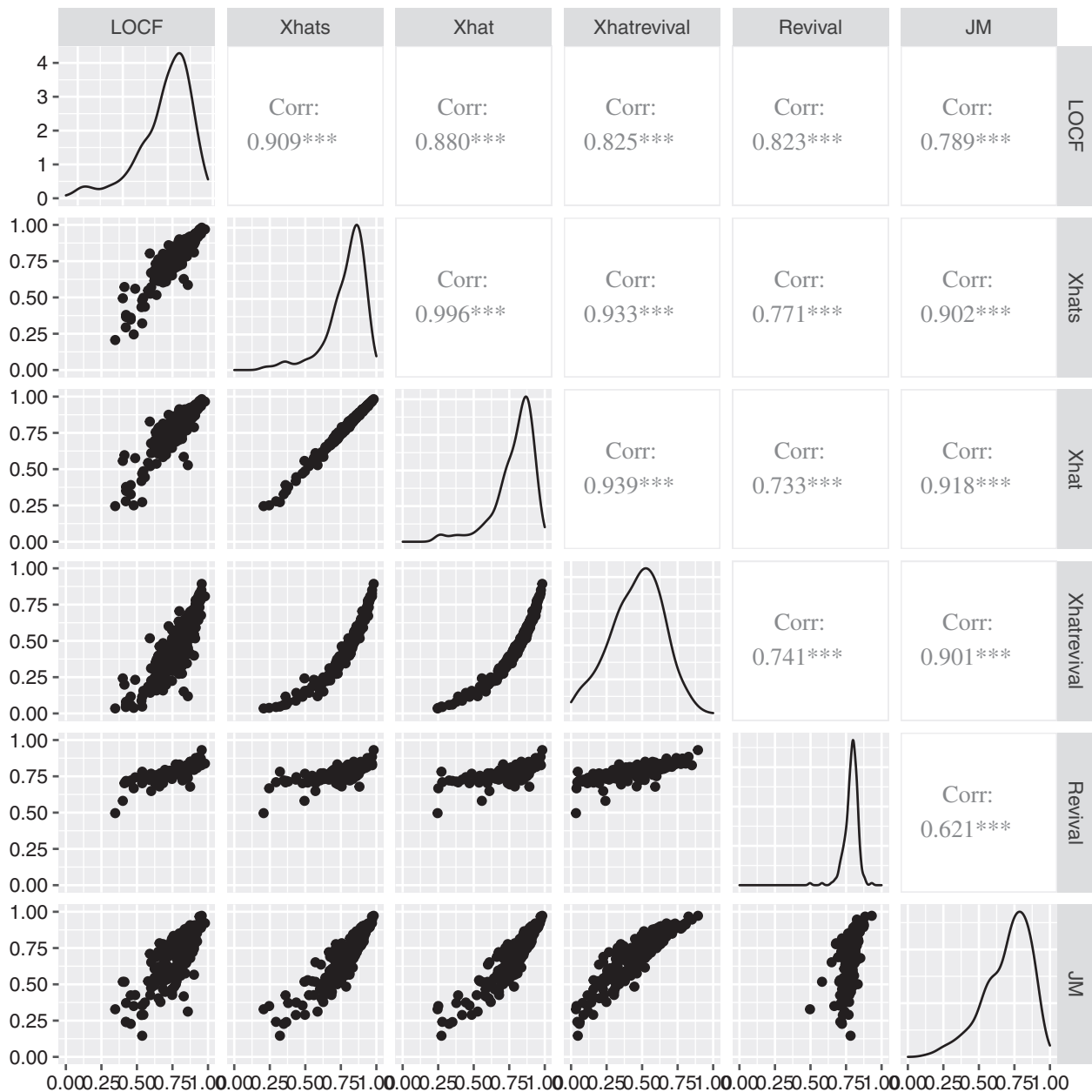

FIGURE 6 Matrix plot of the dynamic predictions obtained from the different approaches for the prothrombin data

TABLE 3 Estimated regression coefficients, standard errors, and chi-squared statistics for the univariate (χ^2) and bivariate vs univariate model with $\hat{x}(t\,|\,s)$ revival (LRT)

Model	Beta	SE	χ^2	LRT
Joint model	0.908	0.221	17.29	0.56
Direct revival	2.717	0.622	17.69	1.76
Last observation	0.858	0.238	12.99	0.17
$\hat{x}(s\,|\,s)$	0.886	0.198	19.30	0.10
$\hat{x}(t\,|\,s)$	0.874	0.194	19.45	0.02
$\hat{x}(t\,|\,s)$ revival	1.168	0.256	20.30	—

TABLE 4 Cross-validated Brier and Kullback-Leibler (KL) prediction errors of different prediction methods; in brackets after the prediction errors are the percentage reduction of prediction error, compared to the null model

Model	Prediction error	
	Brier	KL
Null model	0.1683	0.5206
Joint model	0.1649 (2.1%)	0.5048 (3.0%)
Direct revival	0.1565 (7.0%)	0.4858 (6.7%)
Last observation	0.1585 (5.8%)	0.4932 (5.3%)
$\hat{x}(s\,|\,s)$	0.1549 (8.0%)	0.4797 (7.9%)
$\hat{x}(t\,|\,s)$	0.1549 (8.0%)	0.4791 (8.0%)
$\hat{x}(t\,|\,s)$ revival	0.1536 (8.7%)	0.4751 (8.7%)

respect to their univariate χ^2 value. The fact that direct revival is not well calibrated is also evident from this table, with an estimated regression coefficient of 2.72. In the other methods the calibration slope is acceptable, but it must be noted that calibration in the large was not assessed here. For that, a parametric model like Poisson or Weibull[12,24,25] could be used. After having selected landmarking 2.0 with $\hat{x}(t\,|\,s)$, we fitted bivariate proportional hazards models with the cloglog-transformed landmarking 2.0 with $\hat{x}(t\,|\,s)$ cross-validated dynamic prediction probabilities along with each of the other transformed cross-validated dynamic prediction probabilities. The column LRT reports the likelihood ratio test statistic of each of the bivariate Cox models, compared with the univariate Cox model with only $\hat{x}(t\,|\,s)$ revival. The direct revival dynamic prediction probabilities gives the highest LRT, but its value is not dramatic; adding it would not yield statistical significance at the 5% level.

Finally, Table 4 reports the cross-validated prediction errors (both Brier and Kullback-Leibler, KL) and the percentage of prediction error reduction with respect to the null model, containing no covariates. For the revival models, the calibrated dynamic prediction probabilities (the model-based prediction probabilities based on the univariate Cox models described above) were used.

5 | SIMULATION STUDY

We conducted a simulation to study more closely the differences in terms of predictive accuracy of different landmark models and joint models.

Following our previous paper,[20] we base the biomarker process on a mean zero Ornstein-Uhlenbeck (OU) process $X^*(t)$, starting at $X^*(0) = 0$, and further defined by

$$dX^*(t) = -\theta X^*(t)dt + \sigma \, dW(t), \tag{4}$$

with $W(t)$ a Wiener process and θ and σ are parameters describing the degree of mean reversal (to zero) and influence

of the random fluctuations of the Wiener process, respectively. The solution of (4) is a stationary Wiener process with covariance function

$$\text{cov}(X^*(s), X^*(t)) = \frac{\sigma^2}{2\theta} \exp(-\theta|t - s|).$$

The biomarker process of subject i at time t is obtained by adding a common mean $\mu(t)$, taken to be equal to δt, and a random person effect $b_i \sim N(0, \omega^2)$ to the OU process, leading to

$$X_i(t) = \mu(t) + b_i + X_i^*(t). \tag{5}$$

After adding the random person effect b, the result is a Wiener process[20] with

$$\text{cov}(X_i(s), X_i(t)) = \omega^2 + \frac{\sigma^2}{2\theta} \exp(-\theta|t - s|) = \sigma_{\text{tot}}^2 (\rho + (1 - \rho) \exp(-\theta|t - s|)),$$

where $\sigma_{\text{tot}}^2 = \omega^2 + \sigma^2/(2\theta)$ is the total variance of $X(t)$ and $\rho = \omega^2/\sigma_{\text{tot}}^2$ is the intra-class correlation, the proportion of the total variance represented by the random person effect variance. For the base scenario, the following values were taken: $\sigma_{\text{tot}}^2 = 0.75$, $\rho = 0.75$, $\theta = 1$, $\delta = 0.1$, from which it follows that $\omega^2 = 0.5625$ and $\sigma^2 = 0.375$.

The baseline hazard is taken to be Weibull with rate a and shape b, and hazard $h_0(t; a, b) = abt^{b-1}$. The values for a and b are 0.1 and 1.5, respectively. The resulting hazard, given the value of the biomarker, is given by $h(t \mid X(t)) = h_0(t; a, b) \exp(\beta X(t))$. The value for β was set to 0.5 in the base scenario.

For a given scenario, defined by specific choices of the parameters involved, we started out by generating a single pool of validation data for a large number ($N = 6250$) of individuals. Data for each individual consists of a full set of biomarker values $X_i(t)$ at a fine grid ($\Delta t = 0.01$) from $t = 0$ until a fixed horizon $t = \tau = 10$. The generated biomarker values define the hazard of the individual through $h(t \mid X_i(t)) = h_0(t; a, b) \exp(\beta X_i(t))$, from which "true" conditional survival probabilities can be calculated through $\pi_i(s, t) = \exp(-\int_s^t h(u \mid X_i(u)) \mathrm{d}u)$. Observed data in the validation set are defined as follows. First, the generated biomarkers also define the overall survival probabilities of subject i as $S_i(t) = \exp(-\int_0^t h(u \mid X_i(u)) \mathrm{d}u)$, from which, through the inverse method, a single event time $T_i = S_i^{-1}(U)$ is drawn, with U uniform on $(0, 1)$, and with administrative censoring at τ. The biomarker process is observed at more or less regular intervals. For this we used an observation frequency Δ, chosen so that one would expect on average five measurements before the median marginal time-to-event. The biomarker process then is observed at time 0, and at $k\Delta$ with a disturbance uniform on the interval $[-\Delta/3, \Delta/3]$, all observation times of subject i occurring before the event time T_i and before τ. At these observation times, the observed value of the biomarker was taken to be the true biomarker value at that time with an independent measurement error from a mean zero normal distribution with standard deviation τ, which was taken to be 0.2.

For each of $M = 250$ replications, we then independently generated a set of training data, consisting of $n = 500$ individuals, with data generated in the same way as the observed validation data. We fitted three landmark models, one (LM1.0) using the last observed biomarker value, one (LM1.5) based on the BLUP approach,[17] and one (LM2.0) on the newly proposed landmark approach. We also fit a joint model with linear trend, random intercept and slope for the longitudinal part, and with the hazard based on the current value of the biomarker (JM), using the JM package.[9] Revival models were not considered in the simulation study because they were computationally too expensive. After having fit the landmark models and joint model, the models were used to estimate conditional probabilities $\pi_i(s, t) = P(T_i > t \mid T_i \geq s, \bar{x}_i(s))$ for $i = 1, \ldots, N$ in the large pool of validation data, where s is the landmark time point, t is the prediction horizon, and $\bar{x}_i(s)$ is the set of observed data of subject i before time s. We also calculated conditional probabilities based on Kaplan-Meier estimates (NULL). For s and t, we used the quintiles of the marginal distribution of the generated event times in the validation set (which could differ from scenario to scenario). In the results, we will simply denote s and t with the numbers of these quintiles, so "24" for instance will stand for s being the second quintile, the 40% quantile, and t being the fourth quintile, the 80% quantile of the marginal time to event distribution. After having obtained estimated conditional probabilities $\hat{\pi}_i(s, t)$ for a given method, these were compared with the true $\pi_i(s, t)$ values obtained from the full true biomarker trajectories as explained above, and we obtained estimates of bias $N^{-1} \sum_{i=1}^N [\hat{\pi}_i(s, t) - \pi_i(s, t)]$ and mean squared error (MSE) $N^{-1} \sum_{i=1}^N [\hat{\pi}_i(s, t) - \pi_i(s, t)]^2$.

Figure S1 shows the marginal survival function of the event times of the validation set for the base scenario, while Figure S2 shows the full trajectories and observed data of the first four subjects of the validation set. Figure S3 shows the individual survival functions of these first four subjects. The mean squared errors of the base scenario are shown in

Figure 7 (bias was seen to be virtually zero for the landmarking methods and in the order of 0.01 to 0.04 for the joint model, data not shown).

Clearly, all landmark methods and the joint model perform better than the null model. For all scenarios and all combinations of s and t, the landmarking BLUP approach (LM1.5) improved on the naive landmarking method (LM 1.0), and landmarking 2.0 (LM2.0) further improved on LM1.5. Interestingly, all landmarking methods outperformed the joint model when the prediction horizon is close to the landmark time point ("12," "23," and "34"), and the joint model was only competitive for LM2.0 when the prediction horizon is further away ("14").

The Supplementary Material also shows the reduction in MSE (($\text{MSE}_{\text{NULL}} - \text{MSE}_{\text{method}}$)/$\text{MSE}_{\text{NULL}}$ (Figure S4) and the runtimes for the base scenario (Table S1). The Supplementary Material also shows MSE and reductions in MSE for the other scenarios. The other scenarios are based on the base scenario with one of the parameters altered. In general, results are quite consistent over different scenarios. The relative positions of the different landmark methods is very stable.
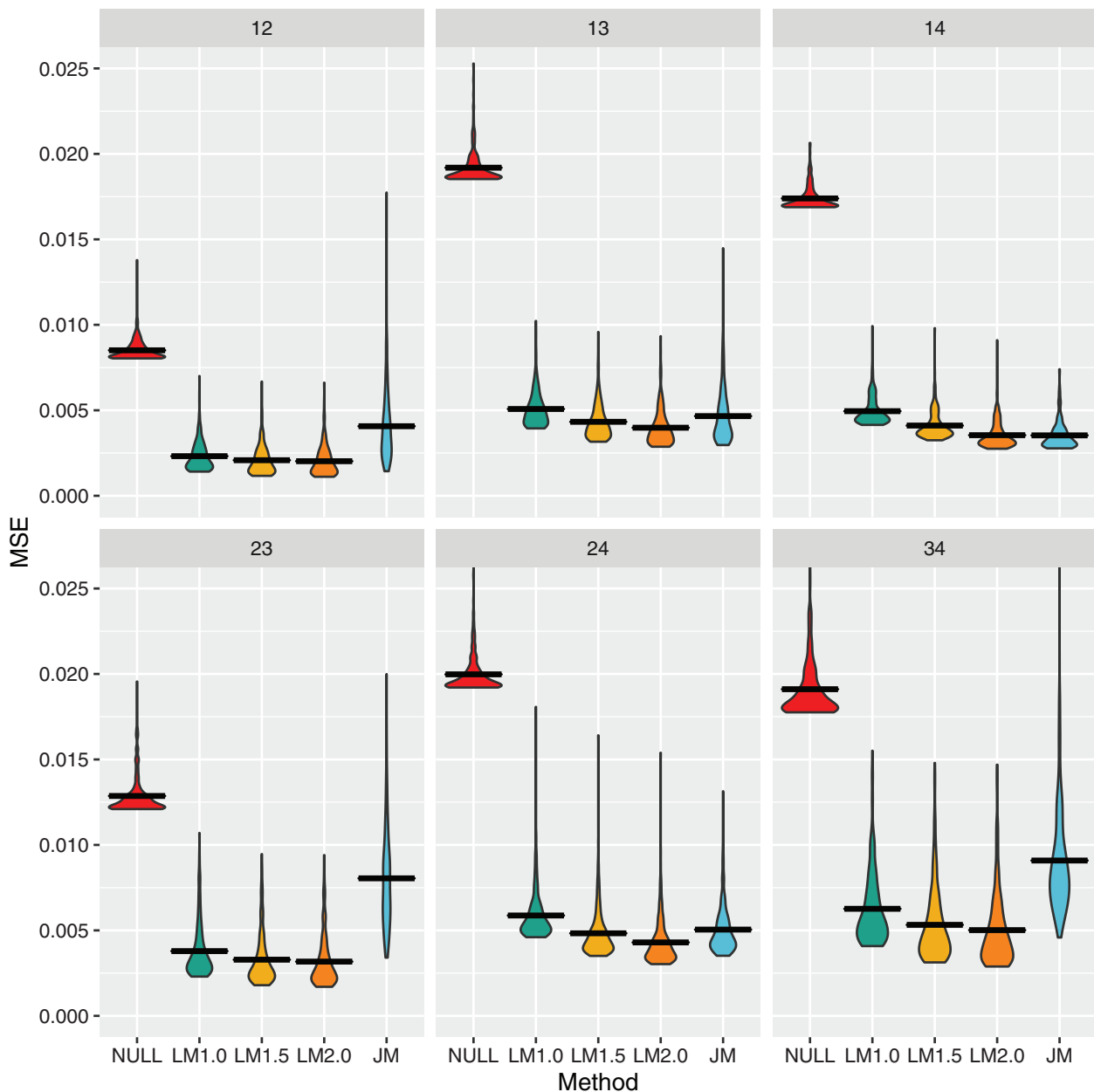

FIGURE 7 Mean squared error of predicted probabilities in the base scenario; the numbers 12 through 34 refer to the combinations st of landmark time s (1 through 3 representing the first through third quintile of the marginal time-to-event distribution) and prediction horizon t (2 through 4 representing the second through fourth quintile of the marginal time-to-event distribution)

Compared to the base scenario, the joint model performs relatively worse for the scenario with lower ρ, and relatively better for the scenarios with higher β and higher τ.

A possible reason for the worse performance of the joint model approach is that the data generating process for the longitudinal data is a stochastic process exhibiting random deviations from a random intercept and deterministic slope, whereas the joint model assumed a random intercept and slope joint model with measurement error. So far most simulation studies comparing joint models with landmarking have not used data generating processes for which the joint models were misspecified.

6 | DISCUSSION

The landmarking principle that "prediction should depend only on the past and nothing but the past in a transparent way" firmly stands. Nevertheless there is a lot to learn from the "future of the past." We have incorporated this in landmarking 2.0 by defining a predictable time-dependent covariate to be used in a time-dependent Cox model, from the landmark time point onwards until the prediction horizon. This predictable time-dependent covariate at time t is defined as the conditional expectation of the biomarker, given alive at time t and given the observed biomarker values before the landmark time point, and could be determined on the basis of an underlying Gaussian process or on a reverse-time model ("revival"). The proposed procedure is more computer intensive than landmarking 1.0, but still considerably less so than joint models, because integration over random effects is avoided, see also Table S1 in the Supplementary Material. In the application we considered, we found that landmarking 2.0, especially when combined with revival, showed the best predictive performance, but it should be emphasized that this was just one application. We do not claim that landmarking 2.0 with revival is always the best performing procedure; more study and experience is needed to better understand the relative advantages and disadvantages of different approaches.

It is well known that landmarking is not consistent in the sense of Jewell and Nielsen,[26] as already discussed in Chapter 7 of our book.[12] In a sense, considering the longitudinal process in the landmarking procedure might bring landmarking closer to Jewell and Nielsen's consistency. On the other hand, the pragmatic aim of using landmarking for dynamic prediction at time s is to provide a simple and robust model that is optimal at time s. There could be unexpected changes over time in the relation between biomarkers and hazard that are difficult to detect and/or to capture in a model; not relying on a single coherent overarching model makes one less vulnerable to misspecification of the model, leading to more robust estimates. In that sense, consistency is both a curse and a blessing. As a reviewer pointed out, more robustness could be gained by fitting separate longitudinal models at each event time. This could be an interesting angle for future study.

The good performance of revival is interesting, but it should be used with caution. First, revival methods seem to work best when the biomarker shows a marked increase or decrease in value toward the event time point. Figure 3 and the model-based version, Figure 4, show that this marked decrease of prothrombin values takes place only 2 months before death, which means that this decrease is hard to foresee after more than 2 months, and therefore its use in long-term prediction (also 2 years as used in our application) may be more limited than it seems on first instance. The revival models also need calibration. This point was already raised in van Houwelingen's commentary;[23] through Bayes' rule in Equation (3) any misspecification in the longitudinal revival models translate in possible miscalibration of the conditional event probabilities. The effect of the misspecification of the longitudinal revival models in landmarking 2.0 combined with revival is more obscured and hard to judge. We advise to always calibrate both direct and indirect revival models in practice.

Extension to higher-dimensional biomarkers is possible in principle, by simultaneously fitting Gaussian processes to the biomarkers. This requires more thinking how to handle the correlation between the biomarker components. The extension to higher-dimensional biomarkers is easier for landmarking 2.0 than for joint models, because for the latter approach dealing with the (typically higher-dimensional) random effects becomes comparatively much more difficult.

In this article, we covered the situation of dynamic prediction based on a biomarker, repeatedly measured over time. Analysis and prediction with this type of time-dependent covariates is often performed using joint models. Another common situation of dynamic prediction with time-dependent covariates concerns the case where the time-dependent covariate is a binary covariate, most commonly changing from 0 to 1 over the course of time. Examples include prediction of survival based on the occurrence of some intermediate event like relapse, progression or response to treatment. This setting has been studied by Suresh et al.[27] In that context the time-dependent covariate always starts as $X(t) = 0$ at $t = 0$ and cannot revert from the value 1 back to 0. Then dynamic prediction at prediction time s is of most interest

when $X(s) = 0$ (which implies $\bar{x}(s) \equiv 0$). One approach, the equivalent of the joint model approach for longitudinally measured biomarkers, is multi-state models, in the present case with states 0 (alive and $X(t) = 0$), 1 (alive and $X(t) = 1$), and 2 (dead). In that case, Equation (2) can also be used, and $\mathbf{E}[X(t) \mid T \geq t, \bar{x}(s) \equiv 0]$ is given by the prevalence probability $\pi_{01}(s, t) = \frac{P_{01}(s,t)}{P_{00}(s,t)+P_{01}(s,t)}$, where $P_{gh}(s, t)$ are the *transition probabilities* in the multi-state model.[20] The conditional survival probability can also be expressed directly as one minus the transition probability $P_{02}(s, t)$, which can also be calculated in the multi-state model. Simulations in Suresh et al[27] showed that predictions based on a correctly specified joint model (Markov illness-death model) provided more accurate predictions than landmark models. They also derived extensions of the simple landmarking approaches that provided improvements in simulations. The added value of landmarking 2.0 could be that the result does not depend on the Markov assumption.[28] It would be of interest to study this case further.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The code used to perform the analyses in this article is available on https://github.com/survival-lumc/Landmarking2.0. Data are publicly available in the **joineR** package.[29]

ORCID

Hein Putter https://orcid.org/0000-0001-5395-1422

REFERENCES

1. Pauler DK, Finkelstein DM. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Stat Med.* 2002;21(24):3897-3911.
2. Taylor JM, Park Y, Ankerst DP, et al. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics.* 2013;69(1):206-213.
3. Tsiatis A, Degruttola V, Wulfsohn M. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc.* 1995;90(429):27-37.
4. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics.* 1997;53(1):330-339.
5. Asar Ö, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol.* 2015;44(1):334-344.
6. Hu B, Li L, Greene T. Joint multiple imputation for longitudinal outcomes and clinical events that truncate longitudinal follow-up. *Stat Med.* 2016;35(17):2991-3006.
7. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin.* 2004;14(3):809-834.
8. Rizopoulos D. *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R.* Boca Raton, FL: Chapman & Hall/CRC Press; 2012.
9. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw.* 2010;35(9):1-33.
10. Rizopoulos D, The R. Package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J Stat Softw.* 2016;72(7):1-45.
11. van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand J Stat.* 2007;34(1):70-85.
12. van Houwelingen H, Putter H. *Dynamic Prediction in Clinical Survival Analysis.* Boca Raton, FL: Chapman & Hall/CRC Press; 2011.
13. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics.* 2011;67(3):819-829.
14. Rizopoulos D, Molenberghs G, Lesaffre EM. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical J.* 2017;59(6):1261-1276.
15. 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):3649-3666.
16. van Houwelingen HC, Putter H. Comparison of stopped Cox regression with direct methods such as pseudo-values and binomial regression. *Lifetime Data Anal.* 2015;21(2):180-196.
17. 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):4514-4528.
18. Paige E, Barrett J, Pennells L, et al. Use of repeated blood pressure and cholesterol measurements to improve cardiovascular disease risk prediction: an individual-participant-data meta-analysis. *Am J Epidemiol.* 2017;186(8):899-907.
19. Paige E, Barrett J, Stevens D, et al. 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):1530-1538.
20. Putter H, van Houwelingen HC. Understanding landmarking and its relation with time-dependent Cox regression. *Stat Biosc.* 2017;9(2):489-503.
21. Andersen PK, Liestøl K. Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics.* 2003;4(4):633-649.

22. Dempsey W, McCullagh P. Survival models and health sequences. *Lifetime Data Anal*. 2018;24(4):550-584.

23. van Houwelingen HC. Commentary to the paper by Walter Dempsey and Peter McCullagh. *Lifetime Data Anal*. 2018;24(4):595-600.

24. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401-3415.

25. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res*. 2016;25(4):1692-1706.

26. Jewell NP, Nielsen JP. A framework for consistent prediction rules based on markers. *Biometrika*. 1993;80(1):153-164.

27. Suresh K, Taylor JM, Spratt DE, Daignault S, Tsodikov A. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometr J*. 2017;59(6):1277-1300.

28. van Houwelingen HC, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Anal*. 2008;14:447-463.

29. Philipson P, Sousa I, Diggle PJ, et al. joineR: joint modelling of repeated measurements and time-to-event data; 2018. R package version 1.2.5.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Putter H, van Houwelingen HC. Landmarking 2.0: Bridging the gap between joint models and landmarking. *Statistics in Medicine*. 2022;41(11):1901-1917. doi: 10.1002/sim.9336