

RESEARCH

Open Access



# Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker

Kaci L Pickett<sup>1</sup>, Krithika Suresh<sup>2,3\*</sup>, Kristen R Campbell<sup>1</sup>, Scott Davis<sup>4</sup> and Elizabeth Juarez-Colunga<sup>1,2</sup>

## Abstract

**Background:** Risk prediction models for time-to-event outcomes play a vital role in personalized decision-making. A patient's biomarker values, such as medical lab results, are often measured over time but traditional prediction models ignore their longitudinal nature, using only baseline information. Dynamic prediction incorporates longitudinal information to produce updated survival predictions during follow-up. Existing methods for dynamic prediction include joint modeling, which often suffers from computational complexity and poor performance under misspecification, and landmarking, which has a straightforward implementation but typically relies on a proportional hazards model. Random survival forests (RSF), a machine learning algorithm for time-to-event outcomes, can capture complex relationships between the predictors and survival without requiring prior specification and has been shown to have superior predictive performance.

**Methods:** We propose an alternative approach for dynamic prediction using random survival forests in a landmarking framework. With a simulation study, we compared the predictive performance of our proposed method with Cox landmarking and joint modeling in situations where the proportional hazards assumption does not hold and the longitudinal marker(s) have a complex relationship with the survival outcome. We illustrated the use of the RSF landmark approach in two clinical applications to assess the performance of various RSF model building decisions and to demonstrate its use in obtaining dynamic predictions.

**Results:** In simulation studies, RSF landmarking outperformed joint modeling and Cox landmarking when a complex relationship between the survival and longitudinal marker processes was present. It was also useful in application when there were several predictors for which the clinical relevance was unknown and multiple longitudinal biomarkers were present. Individualized dynamic predictions can be obtained from this method and the variable importance metric is useful for examining the changing predictive power of variables over time. In addition, RSF landmarking is easily implementable in standard software and using suggested specifications requires less computation time than joint modeling. (Continued on next page)

\*Correspondence: [krithika.suresh@cuanschutz.edu](mailto:krithika.suresh@cuanschutz.edu)

<sup>2</sup>Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA

<sup>3</sup>Adult and Child Consortium for Health Outcomes and Delivery Science, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA

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



© The Author(s). 2021 **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.

(Continued from previous page)

**Conclusions:** RSF landmarking is a nonparametric, machine learning alternative to current methods for obtaining dynamic predictions when there are complex or unknown relationships present. It requires little upfront decision-making and has comparable predictive performance and has preferable computational speed.

**Keywords:** Area under the curve, Joint modeling, Landmarking, Prediction accuracy, Variable importance

## Background

Risk prediction models assist physicians in making personalized clinical care decisions for their patients. In predicting a survival outcome, prognostic models are traditionally developed using variables collected on patients at a baseline time. When the variables are measured multiple times over a patient's follow-up, such as medical lab results, many models ignore the longitudinal trajectory of these markers and utilize only baseline values. Thus, these models fail to take into account how changes in the marker over time may affect risk. Ideally, a patient's survival prediction should be updated dynamically as the values of the longitudinal marker change. Dynamic prediction incorporates time-dependent marker information collected during a patient's follow-up to produce updated, more accurate estimates of their survival probability.

Two commonly used statistical approaches for dynamic prediction are joint modeling and landmarking. Joint modeling involves specifying a model for the longitudinal marker process, a model for the survival outcome, and uses a function to link the two [1, 2]. Estimation of this model and computation of the conditional survival probabilities involve numerical integration and can require substantial computation time [2, 3]. It also requires assumptions about the relationship between the marker process and event time that may be unknown, or even when known, often lead to convergence issues and pose a computational burden as the complexity increases [2]. When there is interest in incorporating multiple longitudinal markers, approximation techniques must be used to evaluate parameter estimates but extensions are limited as is the software to perform the computations [4, 5].

Landmarking requires specifying a sequence of survival models for the subsample of individuals still at risk at prediction times of interest during follow-up, referred to as landmark times. At each landmark time, the model incorporates the value of the longitudinal marker at that time, or with extension, the longitudinal marker history up to that time [6, 7]. In classic landmarking, a Cox proportional hazards model is used. This approach avoids specifying the distribution of the stochastic marker process in time, making it appealing compared to the distributional assumptions required by joint modeling. It has been shown that landmarking prediction accuracy is affected by misspecification of the dependence structure

between the longitudinal process and the survival process, but is less sensitive to misspecification of the longitudinal marker trajectory and violation of the proportional hazards assumption [7–9].

Both joint modeling and landmarking have limitations when used to capture complex relationships between the marker and survival. These methods have yet to be explored for dynamic prediction in scenarios with a multitude of available patient and clinical information, where the quantity and complex relationships between these variables can present challenges for these classic techniques. If many variables are plausible candidates for affecting the survival outcome, and these relationships are largely unknown or complex, correctly specifying a model to describe their relationship with the survival outcome can be difficult. This complexity increases with multiple longitudinally-measured variables, for which the longitudinal marker trajectories, their dependence with the survival process, and their dependence with each other must be specified. Both joint modeling and landmarking also rely on a survival model with a proportional hazards assumption [6, 8], which can be restrictive and unrealistic in practice. Thus, we hypothesize that a nonparametric approach for dynamic prediction that does not require explicit specification of the covariate relationships can improve upon prediction accuracy and ease-of-implementation.

Machine learning algorithms are nonparametric methods that assume no prior knowledge about the data and have become popular in the prediction of survival outcomes [10–14]. These methods have been shown to have superior predictive performance compared to traditional regression methods when there are a large number of predictors relative to the sample size (high-dimensional data) and predictors have nonlinear, complex relationships in the hazard for the survival outcome [15, 16]. Dynamic prediction with machine learning has been explored with the use of a machine learning ensemble for binary outcomes adapted to survival data [17]. The predictive performance of this approach was similar to joint modeling and Cox landmarking in a data application. However, the method is complex and does not provide an interpretation of the covariate effects on survival. We propose a machine learning approach for dynamic prediction that uses the tree-based method of random survival forests and is

easily implementable in standard software. This approach will provide an alternative way of obtaining accurate dynamic predictions in the presence of multiple covariates without need for prior specification of covariate relationships.

A random survival forest (RSF) is a nonparametric ensemble method for the analysis of right censored survival data, built as a time-to-event extension of random forests for classification [12, 18]. The method can handle multiple covariates, noise covariates, as well as complex, nonlinear relationships between covariates without need for prior specification [19]. As such, RSFs often fill a role in replacing classic Cox regression when the proportional hazards assumption is in question as it is assumption-free [20]. There is recent work that explores the use of RSF in the context of dynamic prediction. One such work includes an investigation of using a landmark-type model to predict a time-to-event based on a discrete-time survival model [21]. Another approach uses counting processes to dynamically predict risk based on a piecewise constant hazard model [22]. A new definition of the receiver operating characteristic curve has also been developed for evaluating the performance of RSF [23]. Approaches to handle multiple longitudinal covariates have been proposed that reduce the dimensionality of the covariates and subsequently apply RSF to dynamically predict the event [24, 25]. In our proposal, RSFs fit at each landmark time can incorporate updated longitudinal information in survival prediction while requiring no assumptions about the relationship between the longitudinal trajectory and the survival process.

We aim to assess the utility of using RSF in a landmarking framework as an approach for dynamic prediction of time-to-event outcomes using longitudinal biomarkers. A simulation study is performed comparing the predictive performance of a joint model, a Cox landmarking model and the proposed RSF landmark approach under scenarios of non-proportional hazards and model misspecification as well as inclusion of multiple longitudinal markers. In an application to dynamically predicting death after heart valve transplant, we compare the predictive performance of different model-building decisions for RSF landmarking. Additionally, we illustrate the use of RSF landmarking to obtain dynamic predictions for the development of de novo Donor Specific Antibodies (dnDSA) using longitudinal measurements of the immunosuppression drug Tacrolimus (TAC) in kidney transplant patients.

The structure of the paper is as follows. In Section “Methods,” we describe the extension of RSF to be used in a landmarking framework and give details about how to assess the prediction accuracy. Sections “Simulation study” and “Application” assess the methods via simulation and data application, respectively. The “Discussion” section includes concluding remarks as well as future

directions, followed by “Conclusions” which includes brief recommendations.

## Methods

### Dynamic predictions

The observed data is given by  $D_n = \{T_i, \delta_i, x_i, y_i; i = 1, \dots, n\}$  where  $T_i = \min(T_i^*, C_i)$  is the observed event time for the  $i$ -th subject ( $i = 1, \dots, n$ ), with  $T_i^*$  denoting the true event time,  $C_i$  is the censoring time, and  $\delta_i = I(T_i^* \leq C_i)$  the event indicator. We observe  $x_i$ , the baseline covariate vector, in addition to  $y_i$ , a continuous vector of the longitudinal measurements for the marker, with  $y_{ij} = y_i(t_{ij})$  denoting the value of the marker observed at time  $t_{ij}$  for  $j = 1, \dots, n_i$ , where  $n_i$  is the number of marker measurements for individual  $i$ .

We are interested in obtaining a predicted probability of survival for a new subject  $m$  from the same population given their history of longitudinal marker measurements and baseline covariate data. Specifically, the aim is to obtain the predicted probability of surviving to a prediction horizon  $\tau + s, s > 0$ , given that subject  $m$  has survived up to landmark time  $\tau$ , where  $s$  is a specified prediction window of interest. That is, our dynamic predictions are defined as:

$$\begin{aligned} \pi_m(\tau + s \mid \tau, x_m, \bar{y}_m(\tau)) \\ = Pr(T_m^* \geq \tau + s \mid T_m^* > \tau, D_n, x_m, \bar{y}_m(\tau)). \end{aligned} \quad (1)$$

This prediction is conditional on a summary value of the history of the marker up to time  $\tau$ ,  $\bar{y}_m$ . For the purposes of this manuscript, we assume that this summary measure is the scalar value of the last observation carried forward (LOCF),  $y_m(\tau)$ . This formulation can be extended to include multiple longitudinally-collected markers.

### Landmarking

Landmarking involves selecting only the subset of subjects still event-free at landmark time  $\tau$  and using a survival model to estimate the probability of surviving to the prediction horizon  $\tau + s$ , where the prediction window  $s$  is pre-specified. There is often interest in many landmark times  $\tau_1, \tau_2, \dots, \tau_L$ , and thus landmarking involves developing a prediction model at each time. We construct a prediction data set for each landmark time  $\tau_l$  for the risk set  $R(\tau_l) = \{i : T_i > \tau_l\}$ , consisting of individuals still event-free and uncensored at time  $\tau_l$ . To reduce the potential for bias that can occur from violation of the proportional hazards assumption within a prediction interval, administrative censoring is applied to the individuals' event times at the pre-specified prediction horizon,  $\tau_l + s$  [6, 7]. The selection of multiple landmark times allows the same subject to contribute to the estimation of many predicted residual time distributions. For a model to be fit at each landmark time, every subject still in the risk set must have a value for the longitudinal marker

at that landmark time. If the marker is not continuously observed, this value may not be available in practice and is imputed, most commonly using LOCF [6]. A separate survival model is then fit at each  $\tau_l$  to obtain an estimate of the landmark-specific effect of the marker for predicting survival between  $\tau$  and the fixed horizon  $\tau + s$  for the risk set  $R(\tau)$ . For Cox landmarking, the proportional hazards model at each landmark time  $\tau$  is given by  $h(t|\tau, x, y(\tau)) = h_{0,\tau}(t) \exp\{x'\xi_{1,\tau} + y(\tau)\xi_{2,\tau}\}$  for  $\tau \leq t \leq \tau + s$ , where  $h_{0,\tau}(t)$  is the landmark-specific baseline hazard function and  $\xi_\tau = (\xi_{1,\tau}, \xi_{2,\tau})'$  is the vector of landmark-specific coefficient estimates corresponding to that landmark time.

**Random survival forests**

Random Survival Forests is an ensemble tree-based method for the analysis of right-censored survival data and is an extension of the random forest method [18, 26]. Survival trees are built by recursively partitioning the covariate space using binary splits to form groups of subjects who are similar according to the survival outcome [27]. The RSF predictor ensemble is formed by aggregating the results of many survival trees. Here, we present the RSF algorithm for creating a tree ensemble for predicting survival for a given set of covariates, as described in [28].

- 1 Draw  $B$  ( $n_{tree}$ ) subsamples of a specified sample size from the original dataset (with or without replacement).
- 2 Grow a survival tree for each subsample  $b = 1, \dots, B$ .
  - (a) At each tree node select a subset of the predictor variables available to try as candidates for splitting ( $m_{try}$ ).
  - (b) Select the node split from the candidate variables that maximizes the survival difference between the daughter nodes based on selected split criterion ( $splitrule$ ) up to the number of predefined split points ( $n_{split}$ ).
  - (c) Repeat (a)-(b) recursively on each daughter node until stopping criterion is met, often the number of unique cases in each terminal node ( $nodesize$ ).
- 3 Calculate the cumulative hazard function for each tree using the terminal nodes. Aggregate information from the  $B$  survival trees to obtain a risk prediction ensemble.

This method is implemented in the R package **randomForestSRC** with the function `rfsrc()` and can be tailored based on different ensemble parameters that affect tree building [29]. Selection of these parameters can change the results of the predictions [29], and in Table 1 we show the inputs used in this manuscript [26]. The parameters

**Table 1** Parameters of a Random Survival Forest with suggested values

Parameter	Description	Suggested value
<i>n<sub>tree</sub></i>	how many subsamples should be performed	500-1000
<i>sampsiz</i> e	proportion of the data should be selected for each subsample	63% of available data
<i>samptype</i>	Sampling Done With Replacement (swr) or without replacement (swor)	swr
<i>nodesize</i>	number of unique cases in each terminal node	15
<i>m<sub>try</sub></i>	how many variables should be included as candidates for splitting	$\sqrt{\# \text{ variables}}$
<i>n<sub>split</sub></i>	how many splits should be considered for each candidate variable	10
<i>splitrule</i>	impurity measure which aims to create the two most dissimilar daughter nodes at each split	"logrank"

that affect each node split include  $m_{try}$  and  $n_{split}$ , which allow selection of how many variables should be included as candidates for splitting and how many splits should be considered for each candidate variable respectively. Specifying the number of splits considered for each candidate variable can reduce computation time compared to testing all possible split points for each covariate. The parameter  $nodesize$  denotes the number of unique cases in each terminal node and is mentioned in the existing software package documentation as something that should be experimented with to improve predictive performance.

**RSF ensemble for dynamic prediction**

At each landmark time, the general RSF algorithm is applied to the corresponding landmark data set to build a risk prediction ensemble. Specifically, for landmark time  $\tau$  we apply the RSF algorithm to the risk set  $R(\tau) = \{i : T_i > \tau\}$  using predictor variables  $K_i = (x_i, \bar{y}_i(\tau))$  for each subject  $i$  in the risk set. Once each tree has reached the stopping criteria of the selected split type, the ensemble construction consists of aggregating tree-based Nelson-Aalen estimators. We extend the Morgensen et. al [28] formulation to our landmark framework. For each subsample  $b$  drawn from risk set  $R(\tau)$ , we obtain a survival tree, denoted  $Q_b^\tau$  and a unique terminal node of subjects with the covariate vector  $K = (x, \bar{y}(\tau))$ , denoted  $Q_b^\tau(K)$ . Let  $c_{ib}^\tau$  be the number of times that subject  $i$  is selected in the  $b$ th subsample for landmark time  $\tau$ . For subject  $i$ , we define the counting process as  $N_i(t) = I(T_i \leq t)$  and the at-risk process as  $Y_i(t) = I(T_i > t)$  [30]. For landmark time  $\tau$ , we then define the landmark counting process and at-risk processes as  $N_i^\tau(t) = I(T_i \geq \tau)N_i(t)$  and  $Y_i^\tau(t) = I(T_i \geq \tau)Y_i(t)$ , respectively. Then, using

counting process notation as in [28], we define the aggregated counting and at-risk processes for each subsample  $b$  based on the landmark population, respectively, as

$$N_b^\tau(t, K) = \sum_{i=1}^n c_{ib}^\tau I(K_i \in Q_b^\tau(K)) N_i^\tau(t);$$

$$Y_b^\tau(t, K) = \sum_{i=1}^n c_{ib}^\tau I(K_i \in Q_b^\tau(K)) Y_i^\tau(t).$$

where in the terminal node corresponding to covariate vector  $K$ ,  $Y_b^\tau(t, K)$  is the number at risk at time  $t$  and  $N_b^\tau(t, K)$  is the number of uncensored events up to time  $t$ . To estimate the conditional cumulative hazard  $H_b^\tau(t|\tau, K)$  for the terminal node corresponding to covariate vector  $K$  in the  $b$ th subsample, we use a nonparametric conditional Nelson-Aalen estimator [31] that is based on subjects at risk at time  $\tau$  and is given by

$$\hat{H}_b^\tau(t | K) = \int_\tau^t \frac{N_b^\tau(du, K)}{Y_b^\tau(u, K)}.$$

The dynamic prediction for a new patient with covariate vector  $K_m = (x_m, \bar{y}_m)$  at prediction time  $\tau$  as defined in Eq. (1) is then given by

$$\hat{\pi}_m(\tau + s | \tau, K_m) = \exp \left\{ -\frac{1}{B} \sum_{b=1}^B \hat{H}_b^\tau(\tau + s | K_m) \right\}.$$

**Additional considerations for building RSF models**

Since the landmark formulation involves fitting a new ensemble at each landmark time  $\tau$ , special consideration of the data and model formulation can occur at each prediction time. Instead of excluding individuals with missing baseline predictors, since we are using a tree-based algorithm, a forest can be grown in RSF to impute missing values to improve predictions [26]. This process can be done both when building the predictive risk ensembles, or when obtaining predictions for new individuals. In addition, the user-specified tuning parameters for the landmark-specific RSF can be adjusted separately at each landmark time to achieve optimal predictive performance. The function `tune.rfsrc()` is available in the **randomForestSRC** package and uses grid search to determine the optimal terminal node size (`nodesize`) and number of variables to try as candidates for splitting (`mtry`) [29]. The optimal values for these parameters are selected to minimize the out-of-bag (OOB) prediction error, which is estimated using Harrell’s concordance index [26]. Tuning is performed using fast trees, which builds survival forests using the specified number of subsamples and sampling type. This reduces computational speed but may not be as accurate as using the full forest [29].

**Variable importance**

RSFs can capture nonlinear relationships between multiple covariates for predicting survival, and the landmarking RSF allows for these relationships to change over time. RSF gives an interpretation of variables that play a key role in predicting the survival outcome by providing a Variable Importance (VIMP) measure for each predictor. VIMP is defined as the change in prediction error on a new test case if variable  $x$  were not available, given  $x$  was used to grow the original forest. This is calculated via permutation of out-of-sample (OOB) observations that were not involved in building the trees [12, 32]. VIMP is a relative measure compared to the other variables included in building the forest ensemble. A zero or negative VIMP value at a particular landmark time reflects that the variable does not contribute predictive power for surviving the prediction window at that prediction time during a patient’s follow-up. VIMP can change during follow-up for the different covariates and can be used by clinicians to identify how the importance of the covariates change as the patient survives to time points beyond baseline.

**Assessing predictive performance**

We assess the predictive performance of each method based on calibration and discrimination. Area under the receiver operating characteristic curve (AUC) was used to assess discrimination, and Brier score (BS) as an overall measure for both calibration and discrimination, and root mean square prediction error (RMSPE) was used to compare predicted probabilities to the true survival probabilities. To account for the censoring present in a survival outcome as well as the changing incidence at each landmark time, we use time-dependent versions of these measures [33]. The time-dependent AUC,  $AUC(\tau, s)$ , is used to determine how well the predicted probabilities discriminate between those likely and unlikely to experience an event in the prediction window  $s$  given that they are event-free at landmark time  $\tau$ . This value ranges from 0 to 1, with higher values indicating better discrimination. The dynamic Brier score,  $BS(\tau, s)$ , is defined as the mean square error between the true event indicator and the predicted probability of an event in the prediction window  $s$  for those event-free at landmark time  $\tau$  [34], with lower values indicating better predictive ability.

**Simulation study**

A simulation study was performed to compare the predictive performance of the RSF landmarking approach to a joint model and Cox landmarking method. Specifically, the four main aims were to compare performance under the scenarios of (1) a simple relationship between the longitudinal and survival processes, (2) a violation of the proportional hazard assumption and a complex dependence of the longitudinal marker process, (3) multiple

longitudinal outcomes, and (4), multiple noise variables not associated with the survival process.

**Data generation**

The data were generated using a joint model. The longitudinal processes for  $k$  longitudinal outcomes were modeled using a linear mixed effects model that included only main effects for three binary baseline covariates and a random intercept and slope.

$$y_{ik}(t) = m_{ik}(t) + \epsilon_{ik}(t) = x'_{ik}\beta_k + z'_{ik}b_{ik} + \epsilon_{ik}(t)$$

where the longitudinal measurements  $y_{ijk} = \{y_{ik}(t_j), j = 1, \dots, n_{ik}\}$  are simulated for each subject  $i$  for measurement time  $t_{ijk}$ , and  $x_{ik}$  and  $z_{ik}$  are covariate vectors associated, respectively, with the vector of fixed effects  $\beta_k$  and vector of random effects  $b_{ik}$ ,  $b_{ik} \sim N(0, G_k)$ , where  $G_k$  is the variance matrix, and is independent of  $\epsilon_{ik}(t) \sim N(0, \sigma_k^2)$ . The estimate of true unobserved value of each underlying longitudinal covariate is  $m_{ik}(t)$ . In this simulation study, we simulate two longitudinal markers,  $m_{i1}(t)$  and  $m_{i2}(t)$ , where  $x_{i1} = x_{i2} = (1, t, X_1, X_2, X_3)'$  and  $z_{i1} = z_{i2} = (1, t)'$ .

For Aims 1-2 including only one longitudinal outcome, two different scenarios were assumed for the survival process, each with varying complexity in the association between the survival and longitudinal processes. In Scenario I, we simulated from a standard joint model where the hazard of the event depended only on the true current marker level. Scenario II included an interaction between  $\log(1 + t)$  and the association parameter to induce a violation of proportional hazards, and also included a quadratic trend in the relationship between the longitudinal marker and the survival outcome:

Scenario I:  $h_i(t) = h_0(t) \exp\{\gamma'w_i + \alpha_1 m_{i1}(t)\}$ .

Scenario II:  $h_i(t) = h_0(t) \exp\{\gamma'w_i + \alpha_1 m_{i1}(t)^2 \log(1 + t)\}$

where  $w_i = (X_{i1}, X_{i2}, X_{i3})'$  is the vector of baseline covariates associated with the vector of coefficients  $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$ . The scalar  $\alpha_1$  quantifies the association between the error free current measurement,  $m_{i1}(t)$ , and the time-to-event outcome. We use a Weibull baseline hazard,  $h_0(t) = \exp(\sigma_1)\sigma_2 t^{\sigma_2 - 1}$ .

For Aim 3, we included the second longitudinal marker using a quadratic relationship with the marker and the survival process:

Scenario III:  $h_i(t) = h_0(t) \exp\{\gamma'w_i + \alpha_1 m_{i1}(t) + \alpha_2 m_{i2}(t)^2\}$ .

A total of 500 simulated data sets of 1000 subjects each were generated. The subjects were assumed to have been followed for a period of up to 15 years, with longitudinal measurements at baseline and at up to 30 follow-up times generated from an exponential distribution with

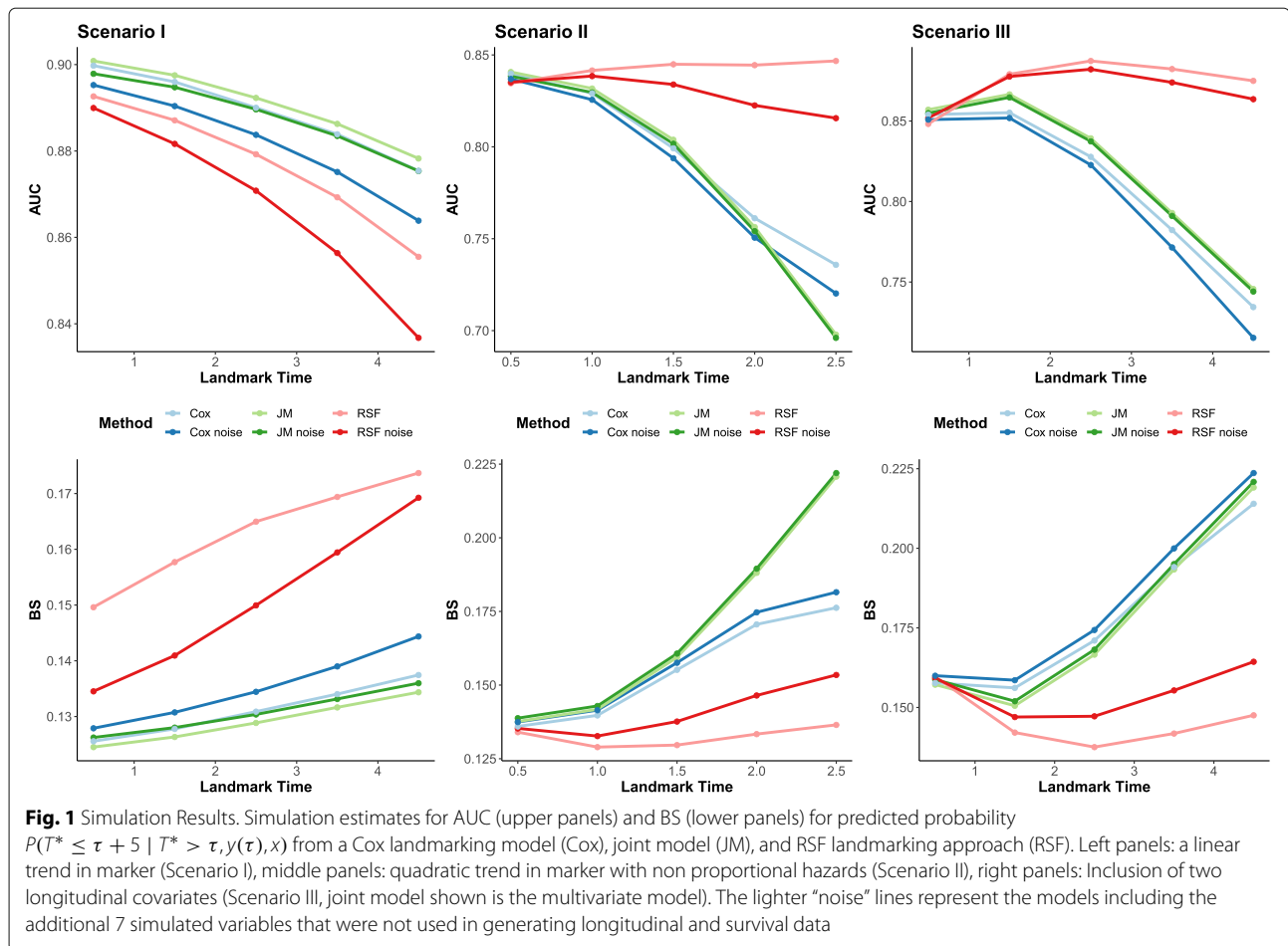
a rate of 2 per year for both longitudinal outcomes. Censoring was simulated under a uniform distribution. Each of the scenarios resulted in approximately 50%, 43%, and 45% censoring, respectively. To address the fourth aim of the simulation study, an additional set of 7 binary “noise” variables,  $X_{i4}, \dots, X_{i10}$ , were simulated with varying prevalence, but not included in generating the survival and longitudinal data under Scenarios I, II, or III. Additional details and the parameter values for the simulation are given in Supplementary Section 1 in Additional file 1.

**Model fitting**

An identical procedure was repeated on data sets from Scenario I, II and III. Each dataset was randomly split into 50% training and 50% testing data sets. Each prediction method was then fit to the training sets and used to compute dynamic survival probabilities on the 500 test individuals, which we used to assess predictive performance. Evaluation of predictive ability via AUC, Brier score and RMSPE was performed at five follow-up times for each scenario, and values were averaged across simulated data sets. A prediction window of  $s = 5$  years was used for all three scenarios.

We fit a joint model as specified in data generation with a linear mixed effects model for the longitudinal process and the true current value of the marker included in the survival process. This model was correctly specified for Scenario I, but misspecified for Scenario II. For Scenario III, we fit two misspecified joint models. One in which fit the longitudinal marker  $y_1$  as in Scenarios I and II, and the longitudinal marker  $y_2$  is treated as a baseline covariate, and a second in which we include both longitudinal markers using a multivariate shared parameter joint model for longitudinal and survival outcomes under a Bayesian estimation approach (implemented using the **JMBayes** R package), ignoring the quadratic relationship with the hazard. We used Cox landmarking by fitting a simple Cox model with LOCF imputation of the longitudinal marker at each of the landmark times. Details of the Cox model and joint model that were fit and how dynamic predictions were obtained can be found in Supplementary Section 2 (Additional file 1).

For the RSF landmarking approach we fit a RSF at each landmark time, using LOCF for the longitudinal marker as in the Cox landmarking. Each RSF was run with 1000 trees and all other suggested values in Table 1. We also fit the RSF models with parameter tuning at each landmark time to select the number of covariates to try as candidates ( $mtry$ ) and final node size. We fit all the modeling approaches twice: (1) with  $w$  including only the 3 baseline variables that were used in the data generation, as well as (2) including the 7 simulated noise variables.



## Results

In Scenario I (Fig. 1, left panel), when there is a simple relationship between the survival and marker processes where the survival depends only on the true current level of the marker, the RSF landmarking approach does not perform better than the Cox landmarking or correctly specified joint model based on both AUC and Brier score. The RMSPE followed a similar pattern to BS (Additional file 1 Table S3) The AUC of all models decrease over time, which is likely due to a selection process that induces increasing homogeneity in the at-risk population at later prediction time points [33]. The landmark Cox and RSF models experience a greater decrease in AUC over time possibly since these models are trained on smaller data sets at later prediction times.

In Scenario II (Fig. 1, middle panel), when we violate the proportional hazards assumption and introduce a more complex relationship between the marker and the survival process, RSF landmarking performed better in terms of both AUC and BS at every time point except the earliest landmark where it performed similarly to the Cox landmarking approach and the misspecified joint model.

The performance of RSF landmarking was similar across all landmark times. In contrast, both the misspecified joint model and Cox landmarking had worsening predictive performance at later landmark times, with the largest decrease over time seen in the misspecified joint model. The RMSPE of the models can be seen in Additional File 1 Table S6, and follows a similar pattern to BS.

In Scenario III (Additional File Tables S7-9), the naïve joint model that uses only longitudinal information from one marker performed similarly to RSF at baseline, but at later landmarks had worse performance. The multivariate joint model that uses both longitudinal markers (Fig. 1, right panel) performed slightly better than RSF at early landmark times but followed the same pattern of reduction over time as the naïve joint model. The Cox landmarking approach had performance similar to the joint models at later landmark times as the longitudinal quadratic relationship between  $y_2$  and the hazard changed. The RSF method performed consistently across time points and was closest to the true probability at every landmark time, with AUC, BS, and RMSPE showing similar patterns.

In all scenarios, all methods (Cox noise, JM noise, RSF noise) had decreased performance when variables were included that were not associated with the survival process (Fig. 1). In Scenario I (simple relationship), the performance of the RSF decreased more with the inclusion of the noise variables compared to the other two methods. In Scenario II (complex relationship), the reduction in the RSF landmark and joint modeling approach was less severe than the Cox landmark approach. All predictions were similar with and without noise in Scenario III. In Figures S1 in Additional file 1, we demonstrate that for Scenario II the VIMP of the noise variables under the RSF landmarking approach is close to 0 across the landmark times.

Estimation and prediction using the Cox landmarking model, the joint model, and RSF landmarking took on average about 25 s (s), 80s, and 19s, respectively, for the 5 landmark times for all scenarios. The multivariate joint model took much longer at about 15 min on average. When noise covariates were included, estimation and prediction of the RSF landmarking increased to around 30s. RSF landmarking with tuning increased time by about 10s for the correct set of covariates and about 30s with the noise variables included. No substantial improvements in predictive performance were seen from tuning in Scenario I and across all scenarios when no noise is present. But, with the more complex relationships in Scenarios II and III, slight improvements in AUC, BS and RMSPE can be seen in the tuned compared to author-selected parameter models when noise variables are included (Additional File 1, Tables S1-S9).

Overall, from the results of our simulation study we found that (1) RSF landmarking did not have good predictive performance when there was a simple relationship between the survival and longitudinal marker processes and few baseline covariates; (2) When the relationship between the marker and survival process was complex, and the proportional hazards assumption violated, the RSF landmark approach performed better than Cox landmarking and a misspecified, simple joint model; (3) The RSF method can accommodate more than one longitudinal covariate with good predictive performance relative to misspecified joint modeling and Cox landmarking; (4) When extra noise variables are included and there are complex relationships present, including these covariates in the RSF landmarking reduces predictive performance only slightly; (5) Parameter tuning when predicting with a small number of covariates did not substantially improve performance and may not be necessary.

### Application

We use two data applications to demonstrate (1) the predictive performance of various RSF model specifications, and (2) the computation and use of variable importance

and individualized dynamic predictions from a RSF landmarking approach.

### Heart valve transplant

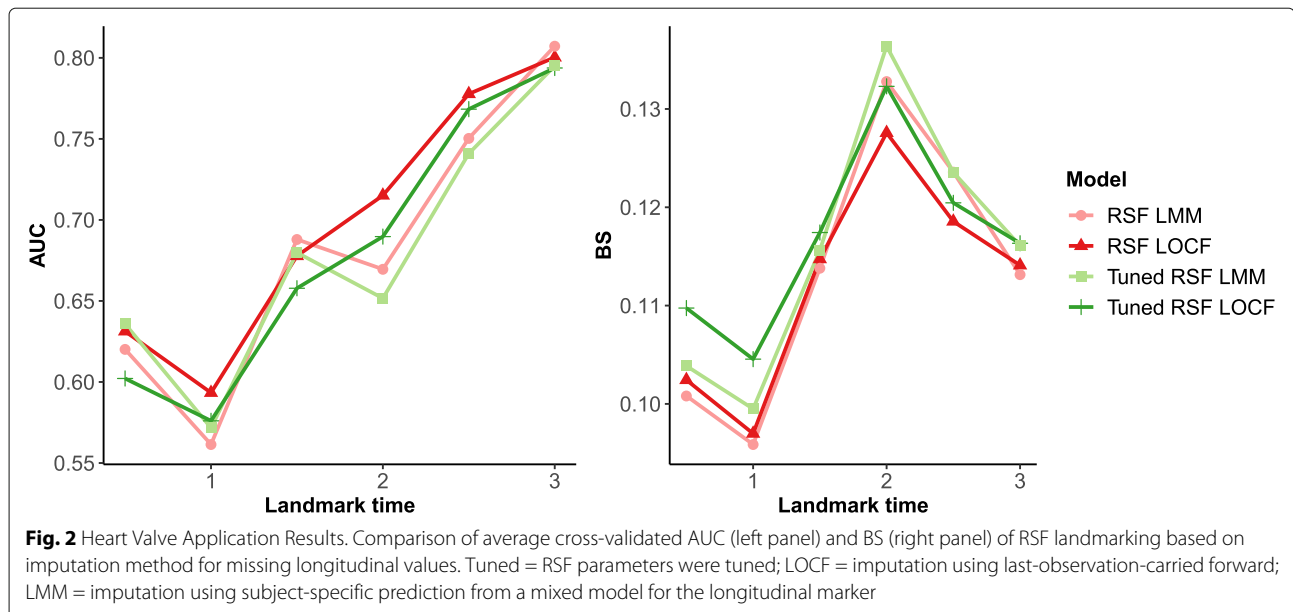
We use an observational study of patients who received an aortic valve replacement surgery [35] to assess the predictive performance of RSF landmarking and evaluate various model specifications in a clinical application. Data were available for 256 patients who were followed post surgery for whom longitudinal measurements were collected for two different heart function measures, left ventricular mass index (LVMI) and ejection fraction. Data are publicly available in the **joiner** package in R software [36]. The aim was to use these longitudinal marker measurements and baseline patient characteristics to predict the probability of surviving a prediction window of 3 years at landmark times 0.5, 1, 1.5, 2, 2.5, 3 years post surgery.

We fit several variations of the RSF landmarking approach to evaluate various RSF modeling specifications: imputation, administrative censoring, tuning, and variable selection (Additional file 1, Table S10). We compare the predictive performance of these RSF landmark models to each other, as well as to Cox landmarking, and a joint model for survival and LVMI that assumes a longitudinal marker trajectory for LVMI, and a dependence structure based on the true current LVMI level. All models included the following clinically relevant variables: sex, age, preoperative left ventricular ejection fraction, presence of concomitant coronary artery bypass graft, and implanted aortic prosthesis type. We also fit RSF models with a larger set of 16 baseline predictors. The longitudinal marker of ejection fraction was included in the RSF and Cox models, but excluded from the joint model due to the associated increase in modelling complexity. For both the RSF and Cox landmarking, the longitudinal measures were imputed at landmark times using two different approaches, LOCF and subject-specific predictions from a linear mixed model (LMM). Predictive performance measures of BS and AUC, as described previously, were computed at each landmark time using 5-fold cross-validation, repeated and averaged across 100 iterations.

First, we compare the RSF landmark approach with Cox landmarking and joint modeling (Additional file 1, Figure S2). The RSF approach has similar predictive performance to Cox landmarking in terms of AUC and BS, with the RSF models having slightly higher AUC and lower BS at earlier landmark times. The RSF model has similar BS to the joint model at earlier landmark times but worse AUC, indicating that although the model has worse discrimination it is possibly better calibrated at earlier times.

Second, comparing within the various RSF specifications, we assess RSF models that use LOCF versus LMM imputation (Fig. 2). Based on AUC and BS, both the tuned





and selected parameter RSF models that used the LOCF imputation performed better than their LMM imputed counterparts at later landmark times, but similarly or worse at earlier landmark times.

Third, as a nonparametric method, RSF does not rely heavily on the proportional hazards assumption, and thus the administrative censoring of the event time that is used in Cox landmarking may not be necessary. RSF landmarking without administrative censoring had similar, but slightly worse, predictive performance compared to the method using administrative censoring (Additional file 1, Figure S3).

Fourth, in high-dimensional settings, Cox and joint models can be restrictive in the number of variables that they can include for model stability. We compare a RSF model with a clinically-relevant set of variables to RSF landmark models that uses all of the 15 available baseline variables. AUC and BS were very similar for both methods, with the RSF models using all variables having slightly higher AUC at some landmark times (Additional file 1, Figure S4).

Fifth, we assess the performance of RSF for dynamic prediction using tuning to optimize the inputs for the number of variables to try at each node (*mtry*) and terminal node size. We found that tuned RSFs have worse performance in terms of AUC and BS compared to the selected settings given in Table 1 (Fig. 2, Additional file 1 Figure S4). This was seen in both the full and reduced models, as well as with both types of imputation.

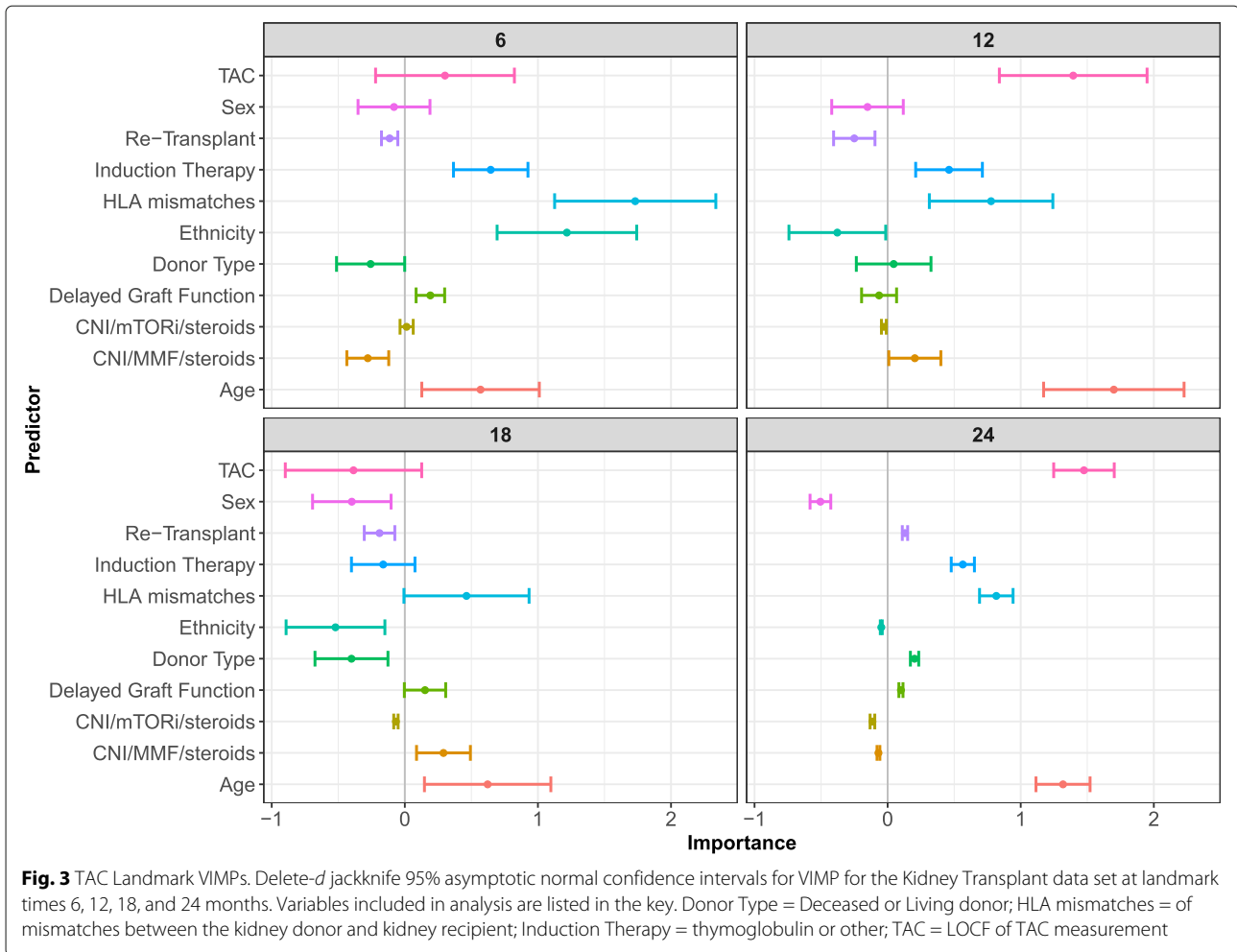
Overall, we found that RSF shows good performance using the suggested parameter inputs and can allow for increased prediction abilities by imputing missing data

within the RSF procedure without significantly affecting performance. In this particular data application, we find that the RSF models perform similarly to Cox landmarking, but not as well as the simple joint model. This is possibly due to the lack of complexity in the relationships between variables, as was explored in the simulation, or due to the limited sample size on which to build the RSF models.

#### Kidney transplant study

To demonstrate RSF landmarking for computing individualized dynamic predictions, we use a retrospective study conducted at the University of Colorado Hospital which followed patients post-kidney transplant for up to 7 years. Data collected included basic baseline patient characteristics, as well as longitudinal immunosuppression drug levels (Tacrolimus, TAC), and the timing of the development of de novo Donor Specific Antibodies (dnDSA), an early warning signs of adverse outcomes. Additional study details can be found elsewhere [37, 38]. The objective was to determine an individual patient's risk of developing dnDSA within 1 year using baseline factors and a patient's up-to-date longitudinal history of TAC. With this application, we demonstrate the assessment of RSF variable importance (VIMP), and illustrate computation of patient-specific dynamic predictions of dnDSA using baseline and TAC data.

There were 535 individuals that met the final inclusion criteria, of which 178 (33%) experienced dnDSA during follow-up. Post transplant, patients were monitored for up to 7 years with up to 90 measures (median 22 [IQR: 15-32]) of TAC ranging from 0 to 30. Ten baseline covariates were included, the full list is given in Fig. 3.

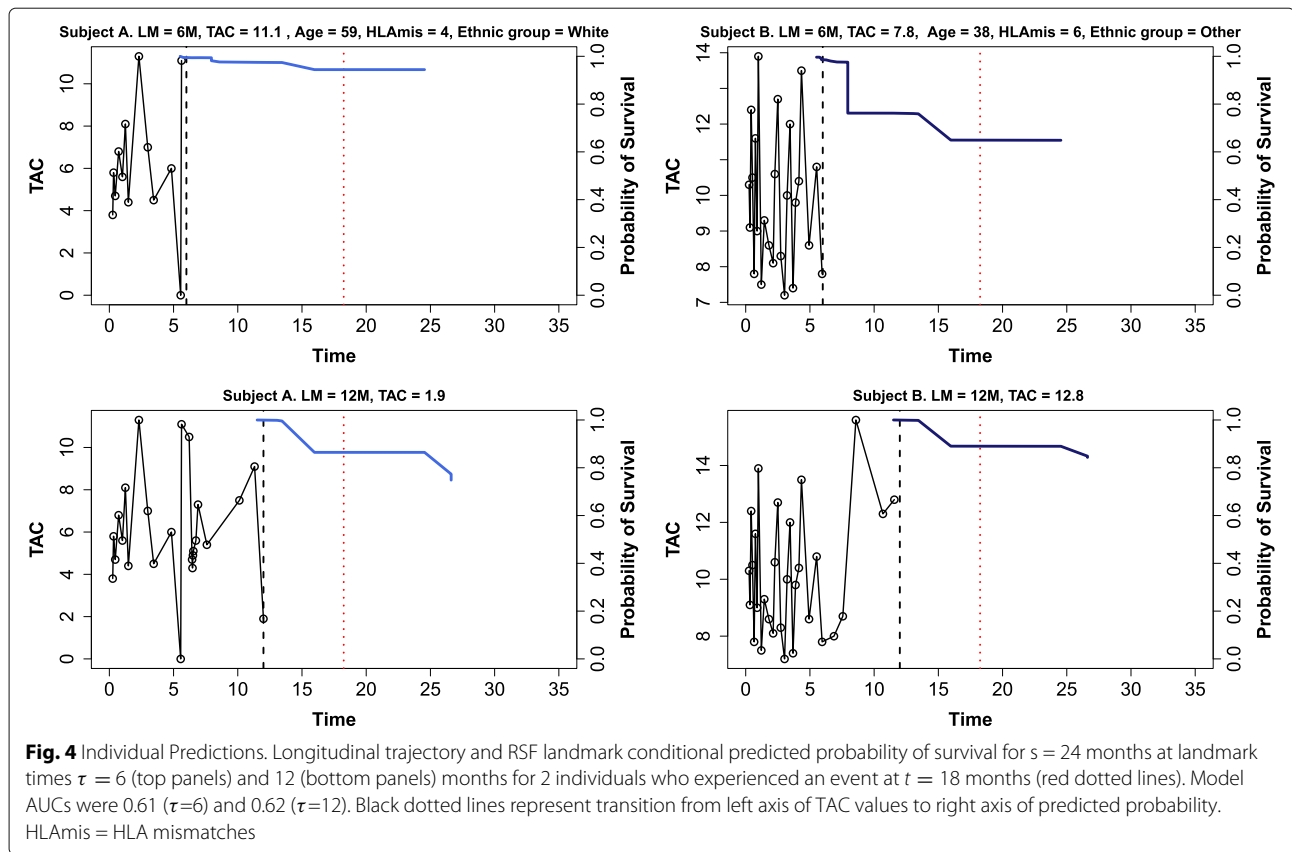


We fit a RSF landmark model at each of the landmark times including all available covariates. The history of longitudinal TAC was summarized at each landmark time as a scalar value using LOCF. The repeated 5-fold average cross-validated AUCs at landmark times 6, 12, 18, 24 months were 0.58, 0.57, 0.52, and 0.64, respectively.

Variable importance assessments were performed at landmark times  $\tau = 6, 12, 18, 24$  months with a prediction window of  $s = 12$  months (Fig. 3). VIMP confidence intervals were obtained using a delete-*d* jackknife estimator, that was calculated using 1000 subsamples [32]. The VIMP of TAC was variable across landmark times, but was generally becoming more important in predicting dnDSA as time from transplant increased, though reductions at landmark time 18 require additional investigation. While ethnicity was important early on, it had VIMP values below 0 for later times indicating that it may not be of much use predicting dnDSA when considerable time has lapsed since transplant. Age and the number of Human Leukocyte Antigens (HLA) mismatches proved to

be relatively important at all timepoints and additional covariates to investigate include the type of induction and maintenance immunotherapy.

As an example of dynamic prediction with RSF landmarking, we fit a reduced RSF landmark model including the covariates age, ethnic group, HLA mismatch, and last-observed TAC, that were shown to be important from the variable importance plot. We present the predicted conditional dnDSA-free survival curves for two individuals in the data set at landmark times of  $\tau = 6, 12$  months (Fig. 4). These two individuals had similar event times but differing baseline covariates. Subject A had a high TAC value at the first landmark time, while Subject B had a low TAC value, and Subject B's survival curve is noticeably lower. Given that both individuals survive to 12 months, we update the survival predictions using their up-to-date TAC values 12 months post-transplant, and find that Patient B's 12-month survival probability has improved based on their increased TAC value, indicating that changes in drug exposure can impact the development of dnDSA.



### Discussion

Dynamic prediction of survival outcomes is a powerful tool in making important clinical decisions. In this paper, we proposed a dynamic prediction method using a RSF landmarking approach. The use of a machine learning approach for dynamic prediction has practical and performance advantages over the existing methods of Cox landmarking and joint modeling.

In our simulation study, we assessed the performance of the RSF landmarking approach. When there are a small number of baseline variables and complex relationships do not exist between the marker and survival process, the RSF landmark approach does not have good predictive performance compared to Cox landmarking and a correctly specified joint model. However, when there was a complex relationship for the dependence between the marker process and the hazard and the proportional hazards assumption was violated, RSF landmarking outperformed a simple, misspecified joint model and Cox landmarking. In the setting with two longitudinal markers, the RSF landmarking approach, which does not require specification of the marker relationships in the survival process, outperformed the joint model that included both longitudinal markers but

misspecified their relationship in the hazard function. As well, when there were noise variables included in the modeling that were not used in the data generation process, the predictive performance of all methods worsened, but the decrease in performance was less for the RSF landmark models than Cox landmarking. Thus, RSF landmarking is beneficial in situations in which there are complex relationships present between the marker and survival outcomes and when there are multiple predictors present, but their clinical relevance and predictive power is unknown.

In the heart valve transplant application, we assess the predictive performance of RSF landmarking and its various specifications in a clinical setting. We compared the imputation of the longitudinal markers in RSF landmarking using a LMM or LOCF, and found that with increasing landmark time the LOCF imputation had better predictive performance. This is possibly due to a misspecified LMM for the longitudinal markers that produces biased predictions at later landmark times. RSF with administrative censoring, which is used in Cox landmarking due to its reliance on the proportional hazards assumption, had slightly better predictive performance for the nonparametric RSF landmark models. Tuning of the RSF parameters increased computation time substantially

but often did not improve predictive performance compared to using suggested parameters. This would indicate that the application of the default *tune.rfsrc*, which is based on Harrell's concordance index, may not be sufficient in this context. While this measure is related to the time-dependent AUC [39], it may not be appropriate for assessing predictive performance for a fixed prediction horizon [40]. Instead, we can consider coding the selection of optimal tuning parameters that maximizes the time-dependent AUC metric at each landmark time. Selecting a small subset of clinically relevant variables compared to including all available variables did not improve performance consistently across landmark times. Overall, using RSF landmarking and LOCF imputation with suggested specifications was found to be sufficient for maintaining predictive performance and did not pose a computational burden.

In the kidney transplant application, we demonstrate the use of RSF landmarking for obtaining and understanding dynamic predictions to answer clinical hypotheses. Previous studies have found that age, ethnicity, HLA mismatches, and Tacrolimus were associated with development of dnDSA [37]. This was confirmed with the use of VIMP plots that also identified the novel predictor of type of maintenance immunosuppression. Ethnicity was important in avoiding dnDSA immediately following transplant, but less so as time progressed post-transplant. Thus, the use of VIMP plots demonstrates how relationships between the survival outcome and predictors can change over time. This application also demonstrated how predictions of dnDSA can be updated in real time based on the TAC value at the landmark time. This provides a tool for personalized medicine, as it may help to inform the clinician whether to change the dosing of TAC based on the patient-specific risk of dnDSA.

Thus, an advantage of RSF landmarking is that this nonparametric, machine learning approach allows for us to capture complex, nonlinear relationships between the longitudinal and baseline variables in the hazard of the survival outcome. It does not rely on a proportional hazards assumption as in Cox landmarking, and does not require prior correct specification of the longitudinal trajectory of the marker or the dependence relationship between the marker and survival process as in joint modeling. As well, it has good predictive performance in the presence of noise variables, indicating that it is a useful approach when there are multiple predictors for which the clinical relevance is unknown. RSF landmarking is also beneficial in situations where multiple longitudinal biomarkers are present, as joint models require the specification of the longitudinal trajectory and their dependence in the hazard for each marker process. VIMP can also be useful in this case to see the changing predic-

tive power of the markers over time. In addition, RSF landmarking is easily implementable in standard software and using specifications suggested in this manuscript requires less computation time than joint modeling, especially when there is more than one longitudinal marker or complex covariate relationships exist.

A limitation of RSF landmarking is that the predictions are not linked over time due to the use of independent RSF models at each landmark time results. Cox landmarking with landmark-specific baseline hazards and covariate effects suffers from the same issue, and thus both approaches do not provide a consistent prediction model [41]. Extensions to Cox landmarking include specifying the coefficient estimates and baseline hazard as a smooth function of landmark time to link the landmark models. To apply a similar smoothing technique in an RSF landmarking frameworks, we can compute the cumulative hazard used in the survival prediction at time  $\tau$  as an average of the cumulative hazards for a sequence of landmark times surrounding  $\tau$ . A similar approach was used for dynamic prediction with large scale data, where a sliding window in combination with RSF was used to put more emphasis on the recent versus older past [42].

In this work, we present a comparison of RSF landmarking and its various specifications to two existing methods and demonstrate situations in which it can be expected to have greater utility and predictive performance. Future work will require exploring the use of RSF landmarking in various situations for which extensions to RSFs exist but should be assessed before used in a landmarking framework, such as competing risks [13] and interval censored survival outcomes [43]. Additionally, a comparison of our RSF approach to recently developed conditional inference and relative risk forests [44] for time-varying covariates is of interest. We have demonstrated the use of RSF landmarking using models built using the log-rank split rule that is based on a proportional hazards assumption [27]. Although, we did not find that this method had poor performance in situations where the proportional hazards assumption was violated, the use of other survival ensemble methods that use other splitting rules, such as conditional inference forests can be explored [45]. Additionally, as in Cox landmarking, we use a LOCF imputation for the longitudinal marker at landmark times at which it is not observed. We extend this in the heart valve data application where we explore the use of a model for the longitudinal marker to impute the marker value, as in [46]. However, alternative specifications can be considered using other summary measures such as the marker slope, or using a function of multiple time-varying covariates to grow trees [24, 25]. The ability of RSFs to deal with correlation among its predictors allows for us to include and evaluate

multiple marker specifications when building the predictive model. Additional studies are required to demonstrate the use of RSF landmarking in high-dimensional settings.

## Conclusions

RSF landmarking is a nonparametric, machine learning alternative for dynamic prediction that can capture complex, nonlinear relationships between the longitudinal and baseline variables in the hazard of the survival outcome. It is a useful approach when there are multiple candidate predictors for which the clinical relevance is unknown or in situations where multiple longitudinal biomarkers are present. Examining the VIMP over time can provide insight into the changing predictive power of the markers over time. In addition, RSF landmarking is easily implementable in standard software and using specifications suggested in this manuscript requires less computation time than joint modeling.

## Abbreviations

TAC: Tacrolimus; dnDSA: de novo Donor Specific Antibody; RSF: Random survival forest; HLA: Human leucocyte antigen; AUC: Area under the curve; BS: Brier score; LOCF: Last observation carried forward; JM: Joint model

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-021-01375-x>.

**Additional file 1:** Supplemental materials. This is a PDF of supplemental information including equations and figures referenced in text.

## Acknowledgments

This paper is based on Kaci Pickett's master's thesis supervised by Dr. Elizabeth Juarez-Colunga, Dr. Krithika Suresh and Kristen Campbell.

## Authors' contributions

KP, KC, EJC, and KS performed all statistical analysis mentioned in the paper. SD provided clinical context and guidance. KC served as the liaison between the statistical and clinical co-authors. All authors read and approved the final manuscript.

## Funding

Supported by NIH/NCATS Colorado CTSa Grant Number UL1 TR002535. The design of the study and the collection, analysis, and interpretation of data are the authors' sole responsibility and do not necessarily represent official NIH views.

## Availability of data and materials

The heart valve data set is publicly available in R package *joiner*. The TAC dataset is not publicly available but may be made available upon reasonable request. Code for simulation datasets as well as models for simulation and heart valve data are publicly available at [https://github.com/picketka/rsf\\_landmarking](https://github.com/picketka/rsf_landmarking).

## Declarations

### Ethics approval and consent to participate

Study data collection was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee at the University of Colorado (COMIRB 13-3137). All data collected was retrospective in nature and de-identified, so individual consent forms were not needed and administrative data privileges were not needed.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Pediatrics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. <sup>2</sup>Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. <sup>3</sup>Adult and Child Consortium for Health Outcomes and Delivery Science, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. <sup>4</sup>Division of Renal Diseases and Hypertension, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA.

Received: 17 February 2021 Accepted: 21 August 2021

Published online: 17 October 2021

## References

- Tsai AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin*. 2004;14(3):809–34. <https://doi.org/10.1201/b17622-15>.
- Rizopoulos D. Joint Models for Longitudinal and Time-to-event Data: With Applications In R. Boca Raton: Chapman & Hall/CRC; 2012.
- Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011;67(3): 819–29.
- Hickey GL, Philipson P, Jorgensen A, Kolumunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol*. 2016;16(1):117. <https://doi.org/10.1186/s12874-016-0212-5>.
- Hickey GL, Philipson P, Jorgensen A, Kolumunnage-Dona R. *joinerml*: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Med Res Methodol*. 2018;18(1):50.
- van Houwelingen H, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. Boca Raton: CRC Press; 2011.
- 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>.
- Rizopoulos D, Molenberghs G, Lesaffre EM. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom J*. 2017;59(6):1261–76.
- 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>.
- Wang CKR, Li PY. Machine learning for survival analysis: A survey. *ACM Comput Surv (CSUR)*. 2019;51.6:1–36.
- Segal MR. Regression trees for censored data. *Biometrics*. 1988;44:35–47.
- Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS, et al. Random survival forests. *Ann Appl Stat*. 2008;2(3):841–60.
- Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics*. 2014;15(4):757–73.
- Van Belle V, Pelckmans K, Van Huffel S, Suykens JA. Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artif Intell Med*. 2011;53(2):107–18.
- Gong X, Hu M, Zhao L. Big data toolsets to pharmacometrics: application of machine learning for time-to-event analysis. *Clin Transl Sci*. 2018;11(3): 305–11.
- 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.
- 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*. 2021;184(1):3–30. <https://doi.org/10.1111/rssa.12611>.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
- Bou-Hamad I, Larocque D, Ben-Ameur H. A review of survival trees. *Stat Surv*. 2011;5:44–71.

20. Omurlu IK, Ture M, Tokatli F. The comparisons of random survival forests and cox regression analysis with simulation and an application related to breast cancer. *Expert Syst Appl.* 2009;36(4):8582–8.
21. Moradian H, Yao W, Larocque D, Simonoff JS, Frydman H. Dynamic estimation with random forests for discrete-time survival data. 2021. arXiv preprint arXiv:2103.01355.
22. Wongvibulsin S, Wu KC, Zeger SL. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (rf-slam) data analysis. *BMC Med Res Methodol.* 2020;20(1):1–14.
23. Sun Y, Chiou SH, Wang M-C. Roc-guided survival trees and ensembles. *Biometrics.* 2020;76(4):1177–89.
24. Lin J, Li K, Luo S. Functional survival forests for multivariate longitudinal outcomes: Dynamic prediction of alzheimer's disease progression. *Stat Methods Med Res.* 2021;30(1):99–111. <https://doi.org/10.1177/0962280220941532>.
25. 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;70(1):66–79.
26. 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>.
27. Nasejje JB, Mwambi H, Dheda K, Lesosky M. A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. *BMC Med Res Methodol.* 2017;17(1):115.
28. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *J Stat Softw.* 2012;50(11):1–23. <https://doi.org/10.18637/jss.v050.i11>.
29. Ishwaran H, Kogalur U. Fast unified random forests for survival, regression, and classification (rf-src), r package version 2.11.0. 2021.
30. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*: Springer; 2012. <https://doi.org/10.1007/978-1-4612-4348-9>.
31. Beran R. Nonparametric regression with randomly censored survival data. Technical report. Berkeley: University of California; 1981.
32. Ishwaran H, Lu M. Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat Med.* 2019;38(4):558–82. <https://doi.org/10.1002/sim.7803>.
33. 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.
34. Schoop R, Graf E, Schumacher M. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics.* 2008;64(2):603–10.
35. Lim E, Ali A, Theodorou P, Sousa I, Ashrafian H, Chamageorgakis T, Duncan A, Henein M, Diggle P, Pepper J. Longitudinal Study of the Profile and Predictors of Left Ventricular Mass Regression After Stentless Aortic Valve Replacement. *Ann Thorac Surg.* 2008;85(6):2026–9. <https://doi.org/10.1016/j.athoracsur.2008.02.023>.
36. Philipson P, Sousa I, Diggle PJ, Williamson P, Kolamunnage-Dona R, Henderson R, Hickey GL. *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*. 2018. R package version 1.2.5. <https://github.com/graemeleehickey/joineR/>.
37. Campbell KR, Juarez-Colunga E, Grunwald GK, Cooper J, Davis S, Gralla J. Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: application to kidney transplantation. *BMC Med Res Methodol.* 2019;19(1):130.
38. Davis S, Gralla J, Klem P, Tong S, Wedermyer G, Freed B, Wiseman A, Cooper JE. Lower tacrolimus exposure and time in therapeutic range increase the risk of de novo donor-specific antibodies in the first year of kidney transplantation. *Am J Transplant.* 2018;18(4):907–15.
39. Heagerty PJ, Zheng Y. Survival model predictive accuracy and roc curves. *Biometrics.* 2005;61(1):92–105.
40. Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of year predicted risks. *Biostatistics.* 2019;20(2):347–57.
41. Jewell NP, Nielsen JP. A framework for consistent prediction rules based on markers. *Biometrika.* 1993;80(1):153–64.
42. Yang T, Yang Y, Jia Y, Li X. Dynamic prediction of hospital admission with medical claim data. *BMC Med Inform Dec Making.* 2019;19(1):18.
43. Yao W, Frydman H, Simonoff JS. An ensemble method for interval-censored time-to-event data. *Biostatistics.* 2019;22(1):198–213. <https://doi.org/10.1093/biostatistics/kxz025>.
44. Yao W, Frydman H, Larocque D, Simonoff JS. Ensemble methods for survival data with time-varying covariates. 2020. arXiv preprint arXiv:2006.00567.
45. Hothorn T, Lausen B, Benner A, Radespiel-Tröger M. Bagging survival trees. *Stat Med.* 2004;23(1):77–91.
46. Maziarz M, Heagerty P, Cai T, Zheng Y. On longitudinal prediction with time-to-event outcome: comparison of modeling options. *Biometrics.* 2017;73(1):83–93.

## 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)

