

RESEARCH ARTICLE

Dynamic risk prediction of survival in liver cirrhosis: A comparison of landmarking approaches

Mitchell Paukner^{1*}, Daniela P. Ladner², Lihui Zhao¹

1 Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States of America, **2** Northwestern University Transplant Outcomes Research (NUTORC), Comprehensive Transplant Center (CTC), Feinberg School of Medicine, Northwestern University, Chicago, IL, United States of America

* mitchell.paukner@northwestern.edu

Abstract

Electronic health records (EHR) data provides the researcher and physician with the opportunity to improve risk prediction by employing newer, more sophisticated modeling techniques. Rather than treating the impact of predictor variables on health trajectories as static, we explore the use of time-dependent variables in dynamically modeling time-to-event data through the use of landmarking (LM) data sets. We compare several different dynamic models presented in the literature that utilize LM data sets as the basis of their approach. These techniques include using pseudo-means, pseudo-survival probabilities, and the traditional Cox model. The models are primarily compared with their static counterparts using appropriate measures of model discrimination and calibration based on what summary measure is employed for the response variable.

OPEN ACCESS

Citation: Paukner M, Ladner DP, Zhao L (2024) Dynamic risk prediction of survival in liver cirrhosis: A comparison of landmarking approaches. PLoS ONE 19(7): e0306328. <https://doi.org/10.1371/journal.pone.0306328>

Editor: Bo Hu, Cleveland Clinic, UNITED STATES

Received: November 16, 2023

Accepted: June 14, 2024

Published: July 5, 2024

Copyright: © 2024 Paukner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from The Chicago Area Patient-Centered Outcomes Research Network (CAPriCORN). <https://www.capricorncdrn.org/>.

Funding: LZ, R01AG070194, National Institute on Aging, <https://www.nia.nih.gov/>; DL, R01DK131164, National Institute of Diabetes, Digestive and Kidney Diseases, <https://www.niddk.nih.gov/> The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1 Introduction

In modern medicine, the collection and storage of patient health data no longer restricts physicians or clinicians to using snapshot measurements in predicting health trajectories. Rather, the availability of comprehensive, longitudinal sequences of risk factors and biomarker changes has served as the impetus for innovation in risk prediction methodology. Leveraging this wealth of electronic health records (EHR) data has become a focal point in the medical and bio-informatics literature [1–4]. Simple statistical tools can be applied to these vast EHR data sets for meaningful analysis, but taking full advantage of the longitudinal nature of these data should lead to improved prediction of health outcomes in personalized settings. The practice of incorporating health histories through time-dependent covariates in modeling is known as dynamic risk prediction (DRP). There is substantial variation in the literature on DRP ranging from methods incorporating multi-linear sparse logistic regression [5], deep neural networks [6], landmarking [7], joint modeling [8], and many other statistical techniques.

The landmarking (LM) technique presents itself as desirable for its methodological simplicity and flexibility. The LM method has emerged as a popular option when attempting to

Competing interests: The authors have declared that no competing interests exist.

dynamically predict outcomes in an EHR setting, updating specified risk prediction models at selected LM times [9, 10]. Some methods create a variable that measures the change in bio-marker measurements from one LM time to the next [10], while others simply incorporate a time variable to interact with selected covariates across the different LM times [11]. While each method can be advocated for based on varying statistical or philosophical advantages, the underlying goal is leveraging longitudinal EHR data to improve the predictive accuracy of relevant health events for patients. This paper takes a simplified approach by incorporating an interaction between the selected covariate levels and a time variable.

Naturally, an equally important component of a DRP model along with covariates is the summary measures for the outcome. Several authors have used landmarking to dynamically model survival probability using a Cox Model [9, 10], while others have explored the pseudo-observation (PO) approach both through survival probability [12] and restricted mean survival time (RMST) [11]. We take a look at both survival probability and RMST and attempt to compare their relative value in an applied setting. With different outcome measures come different measures of prediction accuracy, making any direct comparison potentially misleading. Thus, direct comparison between dynamic methods is avoided, and we focus on comparing the performance of the dynamic models with their static counterparts for each summary measure.

The medical paradigm in which these DRP models will be assessed is the overall survival of patients with liver cirrhosis. Cirrhosis is a leading cause of mortality in the United States, implicated in the death of more than 50,000 patients each year, a number similar to diabetes and pneumonia [13–15] and exceeding most cancers [16, 17]. It remains imperative among hepatologists to identify new ways to maximize the wealth of longitudinal EHR data to improve the prediction of survival outcomes for cirrhotic patients.

The goal of this paper is to compare several different DRP modeling techniques based on the utilization of LM data sets. Specifically, we aim to compare the employment of PO or Cox regression for modeling survival outcomes for subjects diagnosed with liver cirrhosis. These DRP models can all utilize the LM framework but differ in terms of how the survival outcome is summarized (survival probability or RMST) as well as what types of models are employed (Cox model or generalized estimating equations). In Section 2 we will describe the model-building process for each method we intend to compare as well as the data used in training and testing our models. In Section 3 we explore the model performance based on various metrics such as C-statistics and Brier Scores. And, in Section 4 we discuss the various advantages and disadvantages of the modeling procedures.

2 Methods

The first stage in our modeling process is the construction of LM data sets to be used in model fitting. This process is relatively straightforward and begins with selecting a set of LM times ℓ_j . We chose to select a set of 11 landmark times that take place every 0.5 years of follow-up (6 months) for our data analysis. Then, at each LM time ℓ_j we construct a data set L_j that excludes subjects who experienced their event or were censored prior to ℓ_j .

Once the LM data sets are created they can be used in fitting various types of DRP models. Either the LM data sets L_j can be combined (“stacked”) and used in fitting a “super-model” [11], or a new model can be constructed at each ℓ_j [10]. Due to the issue of censoring, many of the standard modeling techniques used in dynamic risk prediction (DRP) are not tenable. In this paper, we will compare the use of PO with the standard use of a Cox proportional hazards model.

2.1 Pseudo observation approach

The basic concept behind the PO approach is to overcome the issue of censoring by creating a measurement of survival for every subject regardless of event or censoring status. This measurement is created through a transformation of the time-to-event data taking the following form:

$$Z_i = nZ - (n - 1)Z^{-i} \quad \text{for } i = 1, \dots, n \tag{1}$$

where Z is the global survival outcome of interest and Z^{-i} is the outcome of interest computed while excluding the i^{th} subject. A PO of this form can be computed for all subjects and can roughly be interpreted as the subject-specific contribution to whatever survival metric is being used. Then, this measurement can be utilized as an outcome variable in a generalized linear model (GLM) [18].

A PO can, in theory, take several different forms, but in application, pseudo-means (PM), and pseudo-probabilities (PP) are intuitive and useful. Due to the challenges of estimating overall means in survival analysis in the presence of censoring, we will be using RMST and computing pseudo-RMST values as outcome variables. The term pseudo-mean will be used to indicate a pseudo-RMST value for the duration of the paper.

RMST, among various other mean-based statistics, has become a regular fixture in survival analysis when considering how to circumvent the issues of modeling and interpreting survival data when hazards are expected to be non-proportional (NPH) [19–23]. RMST can be computed and interpreted as the mean survival time from the beginning of follow-up to some time horizon τ . Thus, it estimates the expected survival duration of a patient until τ . This allows physicians and patients to consider survival outcomes in terms of days, months, and years of life gained or lost based on important covariates. The purpose for using RMST rather than overall mean survival is to account for the limited follow-up time due to censoring. In time-to-event analysis, many subjects won't have their event time recorded if they are either lost to follow-up or the study ends prior to their event occurring. Thus, overall means cannot be directly estimated and may not be identifiable, though various restricted means can.

A restricted mean, $\mu(\tau) = \int_0^\tau S(t)dt$, is graphically represented as the area under the survival curve from randomization ($t = 0$) to some time horizon ($t = \tau$). In practice, $\mu(\tau)$ can easily be estimated by computing the area under the Kaplan-Meier (KM) curve between 0 and τ as follows:

$$\hat{\mu}(\tau) = \sum_{t_i \leq \tau} (t_{i+1}^* - t_i) \hat{S}(t_i) \tag{2}$$

where t_i is the i^{th} event time, $t_{i+1}^* = \min(t_{i+1}, \tau)$, and $\hat{S}(t)$ is the KM estimate of its respective survival function.

The subject-specific PMs to be used as the response variable $Y_i(\tau)$ in a GLM or GEE, where τ is the restriction time, is computed as:

$$\hat{\mu}_i(\tau) = n\hat{\mu}(\tau) - (n - 1)\hat{\mu}^{-i}(\tau) \quad \text{for } i = 1, \dots, n \tag{3}$$

where $\hat{\mu}(\tau)$ is the restricted mean computed based on the KM curve with all subjects and $\hat{\mu}^{-i}(\tau)$ is the restricted mean computed based on the KM curve excluding subject i as in Eq 2. This particular technique was popularized by PK Andersen [24, 25], and has been utilized by several others in DRP modeling in survival analysis [11, 26, 27]. In our case, we compute the PM based on the conditional KM curve, where survival up to time ℓ is the condition. This PM can be interpreted as subject i 's individual contribution to the conditional RMST.

These PMs are then used as the outcome variable in a GEE in order to model the covariate effects on survival. Hence, we can model the effects of covariates on the PMs as:

$$g[\hat{\mu}_i(\ell_j, \tau | \mathbf{X}_i(\ell_j))] = \alpha_j + \mathbf{X}_i(\ell_j)^T \beta(\ell_j) \tag{4}$$

where α_j is the intercept, $\beta(\ell_j) = (\beta_1(\ell_j), \beta_2(\ell_j), \dots, \beta_p(\ell_j))$ are the coefficients of the p predictor variables, $\mathbf{X}_i(\ell_j) = (\mathbf{X}_{i1}(\ell_j), \mathbf{X}_{i2}(\ell_j), \dots, \mathbf{X}_{ip}(\ell_j))$ are the values of the covariates collected until time ℓ_j , and $g()$ is a link function. Ultimately, we used the identity link function.

Interactions between covariates and ℓ were included in the models. For instance, the parameter $\beta(\ell) = (\beta_1(\ell), \beta_2(\ell), \dots, \beta_p(\ell))$ could be a vector of functions that describe changes in the covariate effects according to variations of $\beta_p(\ell) = \beta_{p0} + \beta_{p1}\ell + \beta_{p2}\ell^2$. If one were to apply some model selection criteria to model coefficients, the linear time interaction, or the squared time interaction might be excluded. While others who have used this approach to dynamically modeling EHR data have performed model selection through some type of step-wise technique, we will use a full model for the sake of comparability between our different modeling approaches. Functions from “geepack” in R were used for fitting the GEE.

When using a PP rather than a PM, one must modify Eq 1 to replace the RMST estimates $\hat{\mu}(\tau)$ and $\hat{\mu}^{-i}(\tau)$ with estimated survival probabilities $\hat{p}(\tau)$ and $\hat{p}^{-i}(\tau)$ at time τ based on the KM curve. The remaining modeling techniques will remain the same, though the interpretation of coefficients must be altered appropriately.

2.2 Cox proportional hazards model approach

The LM approach has been applied through the use of the well-known Cox model also. [9, 10] An obvious advantage to the Cox model is familiarity with statistical software packages, though its reliance on the proportional hazards assumption remains to be a relevant downside (especially when incorporating many predictor variables in modeling).

The dynamic Cox model can be expressed as:

$$P(T_i > \tau | T_i > \ell_j, \mathbf{X}_i(\ell_j)) = \exp\{-\Lambda_0(\tau | \ell_j)\exp(\alpha' \mathbf{X}_i(\ell_j))\} \tag{5}$$

where $\Lambda_0(\tau | \ell_j)$ is the baseline cumulative hazard at time τ given survival to ℓ_j and α_j is a vector of regression parameters to be estimated during model fitting. Estimates of α_j can be obtained by maximizing the partial likelihood as would be done in the static case.

Either a new Cox model can be fit at each LM time with its corresponding LM data set, [10], or a “super-model” can be constructed using a stacked LM data set. [9] In this paper, we have chosen to use the method of stacking data sets based on the results of Keogh et al. in which the models fitted from stacked data sets had better performance.

2.3 Static models

In order to assess the value of each dynamic model, we also created several static models for comparison. The first, and most basic static model fits a baseline model with covariates available at time $t = 0$, and uses baseline covariates obtained at time $t = 0$ for survival prediction at time $t = \ell + \tau$, conditional on survival up to ℓ . Hence, if we set $\tau = 3$, then at landmarking time $\ell = 3$ we would be comparing the dynamic models’ 3 year survival predictions (using covariates at $t = 3$) against this static model’s 6-year survival predictions conditional on survival up to year 3 (using covariates at $t = 0$). This was the static model utilized by Yang et al. (2021) and will be referred to as SM1 [11].

The next static model constructed also fits a single model based on subjects’ baseline characteristics at time $t = 0$ but performs survival prediction based on subject covariates collected

at time $t = \ell$. Essentially, a baseline model is constructed, but updated covariates are used with this model in performing prediction at $t = \tau$. Though this model uses updated patient information, it is not dynamic since it doesn't take into account any element of patient history, only the patients' current covariate values. This modeling approach will be referred to as SM2.

The final static model used for comparison fits a new prediction model at each landmark time (rather than only one at baseline) and also performs predictions based on the most recently available covariates at landmark time ℓ . This model may seem like it is encroaching on "dynamic" territory, but it should still be considered static as the model only accounts for covariates measured at a specific time point rather than a patient's health history or progression of health indicators. This final modeling approach will be referred to as SM3.

2.4 Data

We used the Chicago Area Patient-Centered Outcomes Research Network (CAPriCORN) data which capture the diverse population of the greater Chicago metropolitan area. The dataset encompasses 30 hospitals, and 10 health systems, totaling 12.8 million patients, which allows us to model and assess mortality in patients with liver cirrhosis. The longitudinal nature of this EHR data allows us to compare various dynamic and static prediction models. The CAPriCORN data merges institutional databases while protecting patient health information. The CAPriCORN data was pulled for use by the authors on September 8, 2022, and was fully de-identified prior to any author accessing it.

There were numerous possible predictor variables available in the CAPriCORN data. We chose a subset of possible predictors based on relevant hepatological factors. The time-invariant variables utilized were basic demographic factors such as sex and age. The other important baseline variables are known as "cirrhosis etiology" which is a categorization of the cause of liver cirrhosis (e.g. alcohol use disorder or primary biliary cirrhosis). An individual can have multiple etiologies, so the three etiologies considered in this paper were coded as a 1 if a subject ever was diagnosed with that etiology and a 0 otherwise. The time-dependent covariates associated with liver health were the various relevant lab measurements of liver health (meld-sodium score and albumin level), and certain liver-related health events, known as decompensating events. Also, hepatocellular carcinoma was included as a time-dependent covariate.

The need for written informed consent and ethics approval was waived by the Northwestern University institutional review board due to the retrospective nature of the study. The data were de-identified by the National Patient-Centered Clinical Research Network prior to our use. All methods were carried out in accordance with relevant guidelines and regulations.

2.5 Model assessment

In order to assess the accuracy of the prediction models, we evaluate both discrimination and calibration. We measure discrimination in our models by using Harrell's C-index as a form of determining how well each of our dynamic and static models assigns risk based on concordance [28]. This measure is computed as the area under the receiver operating characteristic curve and is thus commonly known as AUC. For the general case, AUC can be computed as:

$$\text{AUC} = \frac{\sum_{i,j} 1_{T_j < T_i} \cdot 1_{\eta_j > \eta_i} \cdot \delta_j}{\sum_{i,j} 1_{T_j < T_i} \cdot \delta_j} \quad (6)$$

where η_i indicates the risk score at the prediction time time $t = \tau$ that can be substituted for $\hat{\mu}_i$ in the RMST case or \hat{p}_i in the survival probability case. The indicator functions, such as $1_{T_j < T_i}$, are 1 when the specified condition is met and 0 otherwise (δ_j is an event indicator). The AUC

can range from 0 to 1 with scores closer to 1 indicating better prediction accuracy. Technically, scores near 0.5 indicate the worst prediction accuracy, as it is akin to flipping a coin.

When measuring calibration (the alignment between observed and predicted survival outcomes), we can take a uniform approach across the different modeling methods, but cannot necessarily make direct comparisons between them. In all cases, we use a Brier score or a slightly modified version of a Brier score for models using PM. Generally, the Brier score is used for measuring the accuracy of probabilistic predictions, such as with Cox models. The standard form for a Brier Score, in the absence of censoring, is as follows:

$$BS = n^{-1} \sum_{i=1}^n (\hat{p}_i - 1_{T_i > t})^2 \tag{7}$$

where $1_{T_i > t}$ is an indicator function that is 1 when the observed time for subject i is greater than the prediction time t . The Brier Score ranges from 0 to 1 with scores closer to 0 indicating better prediction accuracy.

We use a version of a Brier score that incorporates inverse probability censoring weights (IPCW) in order to accommodate for the censoring in our data. Thus, we compute the Brier Score as:

$$BS(t) = \frac{1}{N} \sum_{i=1}^N \left(\frac{(0 - \hat{p}_i)^2 \cdot 1_{T_i \leq t, \delta_i = 1}}{\hat{G}(T_i^-)} + \frac{(1 - \hat{p}_i)^2 \cdot 1_{T_i > t}}{\hat{G}(t)} \right) \tag{8}$$

where $\hat{G}(t)$ is the estimator of the conditional survival function of the censoring times calculated using the KM method, $1_{T_i \leq t, \delta_i = 1}$ is an indicator function that is 1 when the observed time for subject i is less than or equal to the prediction time t but is an event time, and $1_{T_i > t}$ is an indicator function that is 1 when the observed time for subject i is greater than the prediction time t .

In the PM case, since survival probabilities are not being estimated, we have the following form for our “Brier Score”:

$$BS(t) = \frac{1}{N} \sum_{i=1}^N \left(\frac{(T_i - \hat{\mu}_i)^2 \cdot 1_{T_i \leq t, \delta_i = 1}}{\hat{G}(T_i^-)} + \frac{(t - \hat{\mu}_i)^2 \cdot 1_{T_i > t}}{\hat{G}(t)} \right) \tag{9}$$

This resulting “Brier Score” does not have an upper bound of 1, and is not directly comparable to any Brier Score computed based on survival probability, though lower scores do indicate better performance.

For model assessment, we utilized 75% of the available data set for training our model and the remaining 25% for model assessment.

3 Results

The size of the entire CAPriCORN data set available for model building was $n = 7040$. In the overall cohort ($n = 7040$), there is a median follow-up time of 1.12 years ranging from 0.0027 to 11.01 years to death or loss-to-follow-up. During the follow-up period, 1371 subjects (19.47%) died. The overall 3-year survival rate was 28.13%, and the 8-year survival rate was 3.98%. Dynamic and static models were constructed with a $\tau = 3$ years with landmark times of $\ell = 0, 0.5, \dots, 4.5, 5$ years. [Table 1](#) displays the coefficients with their standard errors and p-values included in the dynamic models, and [Table 2](#) displays the results from the SM2 static model fit using baseline covariates for the sake of comparison.

Table 1. The results of the dynamic models with $\tau = 3$ and mortality as the event of interest.

Variable	Time Function	PM			PP			Cox Model		
		$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value
Intercept	1	2.833	0.047	< 0.0001	0.863	0.026	< 0.0001			
Age	1	-0.009	0.001	< 0.0001	-0.004	0.001	< 0.0001	0.037	0.002	< 0.0001
	$\ell/5$	0.025	0.004	< 0.0001	0.009	0.002	0.0001	-0.085	0.015	< 0.0001
	$(\ell/5)^2$	-0.021	0.005	< 0.0001	-0.008	0.003	0.0018	0.056	0.018	0.0398
ETOH	1	0.171	0.025	< 0.0001	0.055	0.012	< 0.0001	-0.317	0.068	0.0021
	$\ell/5$	-0.753	0.127	< 0.0001	-0.278	0.068	< 0.0001	1.677	0.466	0.0169
	$(\ell/5)^2$	0.554	0.131	< 0.0001	0.183	0.073	0.0124	-1.166	0.546	0.1376
Biliary	1	0.066	0.052	0.2025	0.006	0.029	0.8487	-0.223	0.152	0.2686
	$\ell/5$	-0.710	0.252	0.0049	-0.312	0.148	0.0353	3.344	0.927	0.0055
	$(\ell/5)^2$	0.713	0.247	0.0040	0.366	0.150	0.0147	-3.601	1.070	0.0102
Chol	1	0.133	0.065	0.0391	0.069	0.034	0.0403	-0.412	0.214	0.1427
	$\ell/5$	-0.094	0.310	0.7615	0.004	0.176	0.9826	-0.886	1.361	0.5862
	$(\ell/5)^2$	-0.105	0.315	0.7404	-0.162	0.186	0.3837	1.603	1.511	0.3767
MELD-NA	1	-0.026	0.002	< 0.0001	-0.009	0.001	< 0.0001	0.067	0.004	< 0.0001
	$\ell/5$	0.064	0.010	< 0.0001	0.016	0.005	0.0014	-0.113	0.030	0.0113
	$(\ell/5)^2$	-0.049	0.011	< 0.0001	-0.009	0.006	0.1394	0.039	0.037	0.4795
Albumin	1	0.197	0.013	< 0.0001	0.089	0.006	< 0.0001	-0.565	0.043	< 0.0001
	$\ell/5$	-0.452	0.059	< 0.0001	-0.140	0.030	< 0.0001	0.215	0.249	0.5015
	$(\ell/5)^2$	0.367	0.063	< 0.0001	0.117	0.033	0.0004	-0.413	0.286	0.2767
Decomp	1	-0.151	0.024	< 0.0001	-0.053	0.011	< 0.0001	0.230	0.041	0.0005
	$\ell/5$	0.640	0.140	< 0.0001	0.215	0.073	0.0030	-0.828	0.416	0.1613
	$(\ell/5)^2$	-0.671	0.161	< 0.0001	-0.257	0.088	0.0036	1.095	0.519	0.1014
HCC	1	-0.205	0.067	0.0022	-0.087	0.030	0.0044	0.504	0.130	0.0265
	$\ell/5$	-0.376	0.440	0.3924	-0.341	0.230	0.1379	2.956	0.925	0.0497
	$(\ell/5)^2$	0.200	0.507	0.6933	0.272	0.287	0.3424	-2.431	1.106	0.1296

<https://doi.org/10.1371/journal.pone.0306328.t001>

From Table 1 it is clear that the time-varying measures of MELD-Sodium score, Albumin level, and the number of decompensating events are highly statistically significant in the dynamic model, while the baseline covariates of age and etiology of alcoholic cirrhosis (ETOH) are also highly significant. From Table 2 we see that the same predictor variables are significant.

Table 2. The results of the static models fit at baseline with $\tau = 3$ and mortality as the event of interest.

Variable	PM			PP			Cox Model		
	$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value	$\hat{\beta}$	SE	p-value
Intercept	2.898	0.130	< 0.0001	0.896	0.061	< 0.0001			
Age	-0.011	0.001	< 0.0001	-0.005	0.001	< 0.0001	0.030	0.003	< 0.0001
ETOH	0.192	0.033	< 0.0001	0.054	0.016	0.0005	-0.397	0.077	< 0.0001
Biliary	0.085	0.069	0.2188	0.013	0.038	0.7340	-0.229	0.176	0.1936
Chol	0.179	0.088	0.0409	0.079	0.044	0.0768	0.490	0.244	0.0448
MELD-NA	-0.028	0.002	< 0.0001	-0.009	0.001	< 0.0001	0.058	0.004	< 0.0001
Albumin	0.215	0.023	< 0.0001	0.094	0.010	< 0.0001	-0.640	0.053	< 0.0001
Decomp	-0.131	0.028	< 0.0001	-0.048	0.012	0.0001	0.182	0.044	< 0.0001
HCC	-0.210	0.078	0.0073	-0.081	0.034	0.0188	0.483	0.141	0.0006

<https://doi.org/10.1371/journal.pone.0306328.t002>

The time function used in this modeling procedure fits a linear and a quadratic interaction effect between time and the eight covariates of interest. It adds flexibility for modeling the effects of the covariates. A significant p-value with any of the interaction effects indicates a significant relationship between the specific covariate and time (either linear, quadratic, or both). More complicated interactions with time are possible (e.g., splines), though they extend outside the scope of this paper.

The dynamic coefficient of any covariate used in this modeling procedure can be calculated by the following formula:

$$\beta_p(\ell_j) = \beta_{p0} + \beta_{p1} * (\ell_j/5) + \beta_{p2} * (\ell_j/5)^2, \ell_j \in [0, 5] \quad (10)$$

For example, in the PM model at LM time $\ell = 5$, the covariate effect of having a decompensating event is $-0.151 + 0.64 * 1 - 0.671 * 1 = -0.182$. So, holding all other variables constant, patients who experience decompensated at the LM time of 5 years are expected to lose 0.182 mean survival years compared to patients who don't experience a decompensating event.

In practice, model selection criteria should be employed to potentially eliminate certain variables, including certain interaction effects. For example, it may make sense to exclude the linear and quadratic interaction effect between time and HCC status in the PM model as the p-values for those coefficients are high. For our purposes, we wanted each of the dynamic models to be as directly comparable as possible. Hence, we included all covariates and interactions in the final models.

[Fig 1](#) can be used to visualize the difference in 3-year conditional RMST dynamic coefficients over the 5-year LM period with 95% confidence intervals, while the same figures for the PP and Cox models can be found in the [S1 Appendix](#). Essentially, at any specific LM (represented on the x-axis) one can see the covariate-specific effect (on the y-axis) of 3-year RMST of a one-unit increase in the variable of interest (while holding all other variables constant). Due to the inclusion of a quadratic time interaction in our model, the resulting 3-year conditional RMST dynamic coefficients will have a parabolic structure, so long as the quadratic term is not equal to 0. If we had only included a linear time interaction, we would instead see a straight line with some non-zero slope (as long as the interaction was non-zero), and had we included no time interaction they would be horizontal lines. The clinicians and physicians using the dynamic model with time interactions could then interpret the relationship between the landmarking times and the dynamic coefficients and better understand the relationship between time and the covariate-specific effect on survival.

3.1 Individual dynamic prediction

We can also utilize the dynamic model to provide individual 3-year survival predictions for patients with liver cirrhosis. [Fig 2](#) shows the individual predictions with the dynamic and static models for two different patients based on each model type. The solid line represents the dynamic conditional 3-year survival prediction, and the dashed line represents the 3-year survival prediction based on the static model fit with baseline covariates at the beginning of the study (SM2). In a clinical setting, it can be valuable to distinguish the risk of death between patients when evaluating candidates for certain treatment options (such as transplants). In [Fig 2](#), patient A had a relatively high meld score through the duration of follow-up, experienced multiple decompensating events, and was diagnosed with hepatocellular carcinoma (HCC) at year 1, whereas patient B had a relatively low meld score until roughly year 4, experienced no decompensating events, and was never diagnosed with HCC. Notably, patient A experiences a large decrease in their predicted 3-year survival from $\ell = 0.5$ to $\ell = 1$ (dropping from 2.58 years to 2.27 years and from 2.46 years to 1.99 years in the dynamic and static RMST models

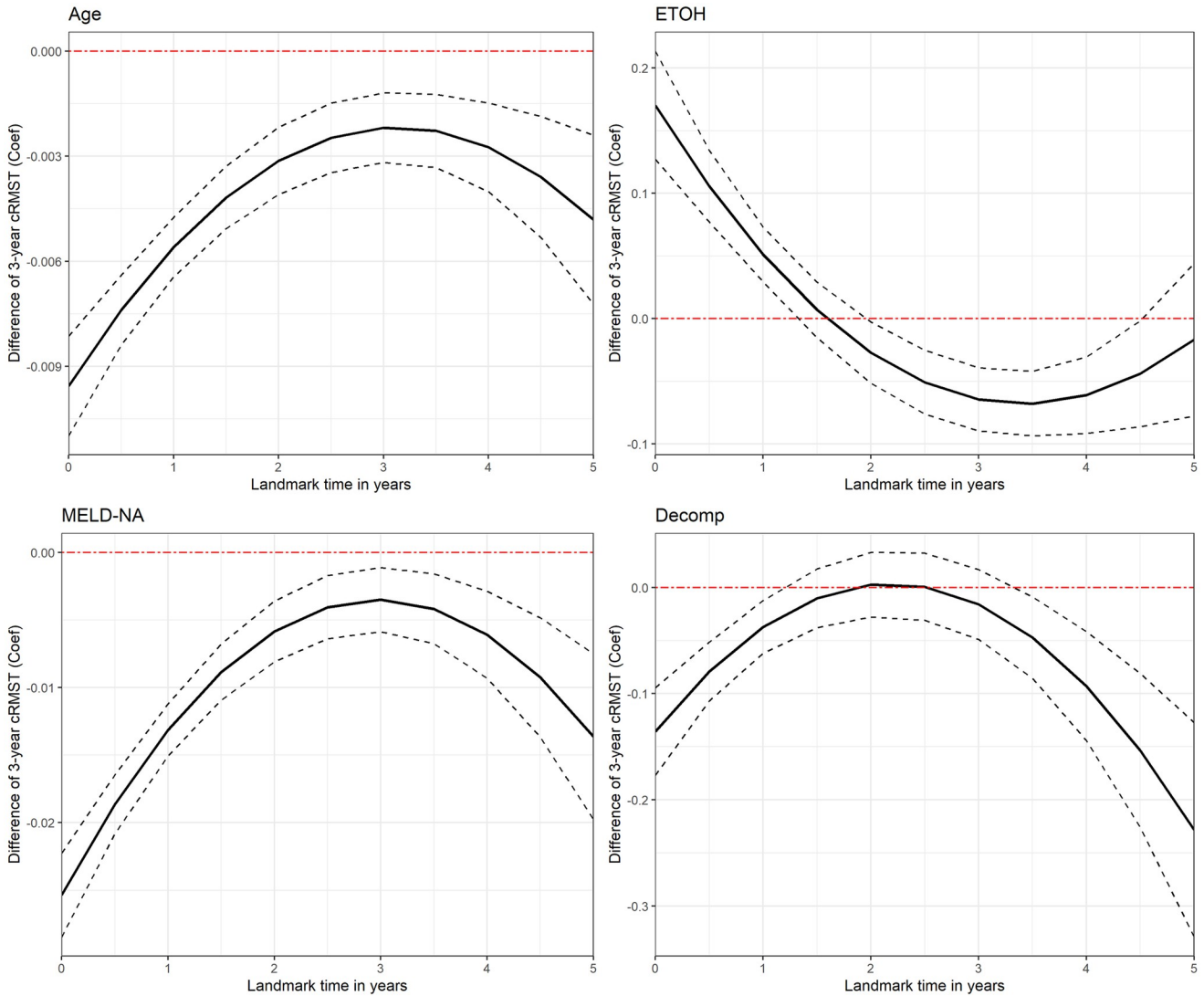


Fig 1. Difference in 3-year conditional RMST in the dynamic RMST model over the 5-year LM period. The solid line represents the dynamic coefficients $\beta_p(\ell)$, or the difference in conditional 3-year RMST resulting from a single unit increase in the p^{th} covariate at ℓ . The black dashed line represents the 95% CI.

<https://doi.org/10.1371/journal.pone.0306328.g001>

respectively) when they were diagnosed with HCC. We also see that the dynamic models are not as strongly influenced by risk factors that take place at a specific time point. For example, when patient B experiences an increase in meld score around year 4, the static model predicts a sharper decrease in survival (2.21 years in the RMST model, 64.80% in the PP model, and 60.95% in the HR model) than does the dynamic model (2.64 years in the RMST model, 77.55% in the PP model, and 97.99% in the HR model) which incorporates the entire history of meld scores. In the event a clinician wishes to evaluate different candidates for transplant at different time points, comparing risk prediction could serve as a useful tool.

3.2 Model performance

As described in Section 2.5, the quality of each dynamic model’s performance was measured in terms of discrimination (AUC) and calibration (prediction error or Brier score) and compared

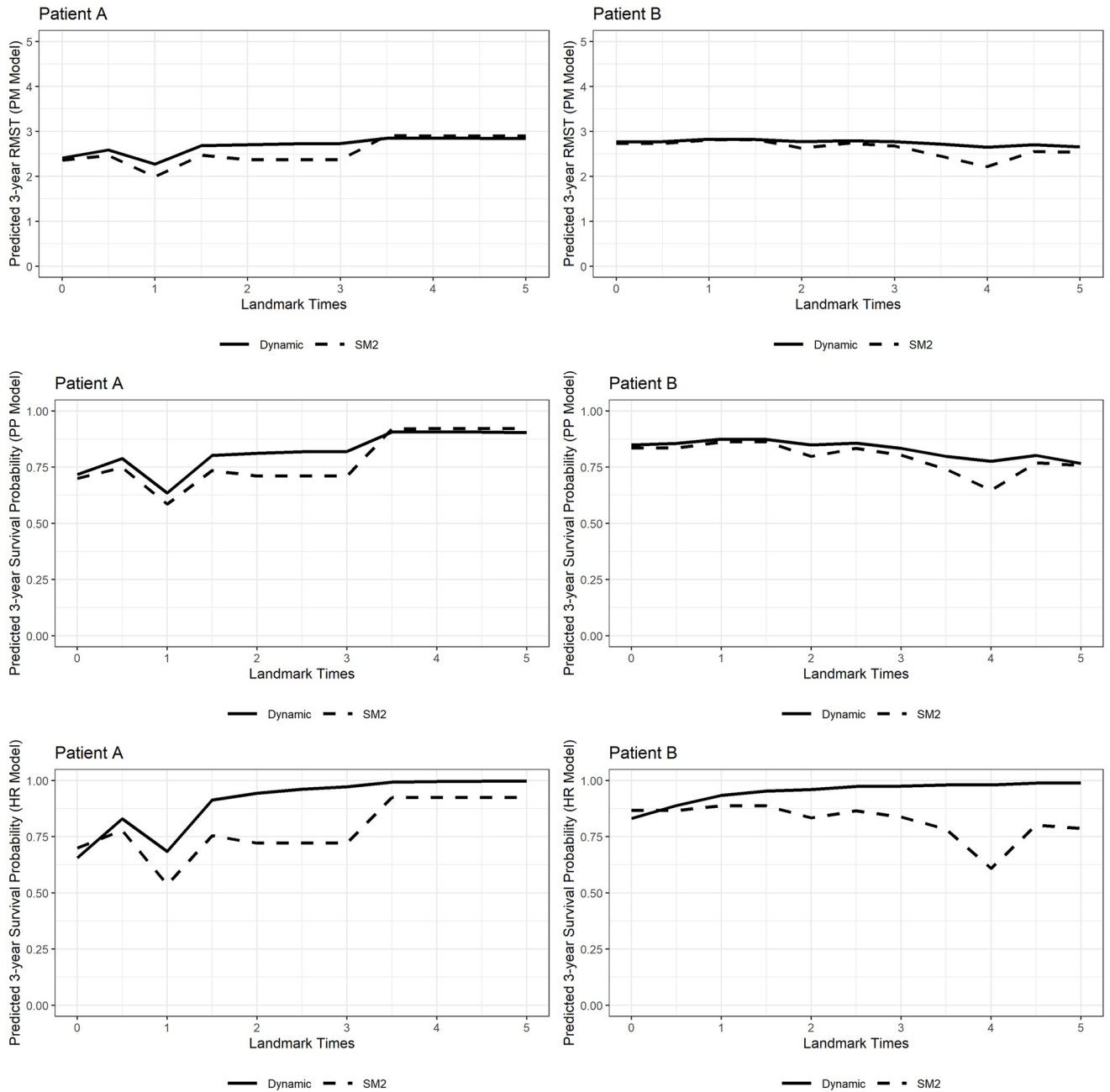


Fig 2. Individual predictions with the dynamic and static models for patients A and B with all-cause mortality as the event of interest. The solid line represents the dynamic 3-year survival prediction, and the dashed line represents the 3-year prediction based on the static model fit with baseline covariates at $t = 0$ (SM2).

<https://doi.org/10.1371/journal.pone.0306328.g002>

with their various static counterparts. A higher AUC score indicates better model performance, while a lower prediction error or Brier score represents better model performance. Fig 3 shows the results of this comparison. It is evident, in terms of AUC, that the dynamic models outperform the crudest static model (SM1), but the static models that merely use updated covariate levels for predictions have similar performance.

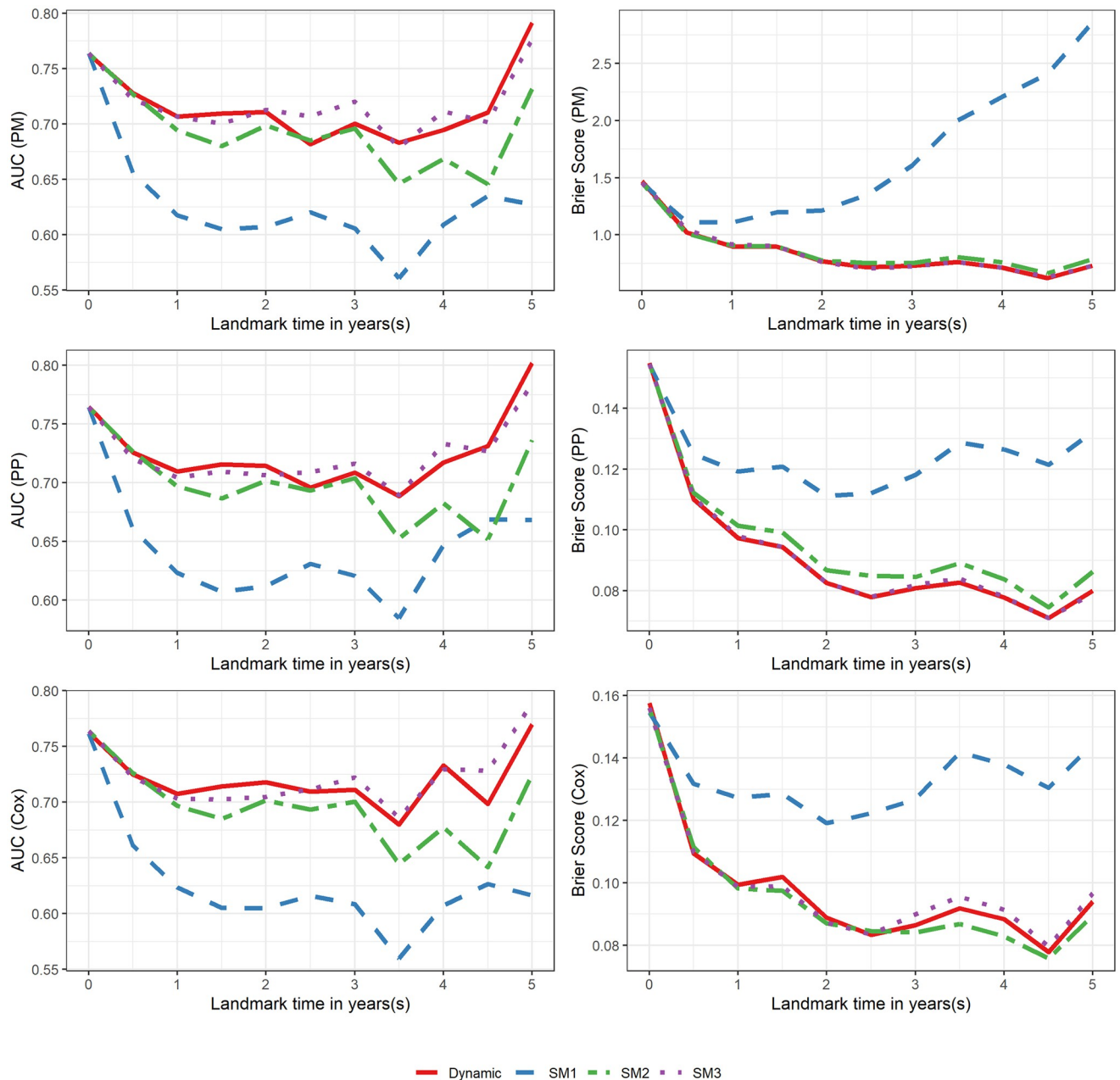


Fig 3. Assessment of model performance in terms of AUC and Brier score. The solid line represents the dynamic model and the dotted/dashed lines represent the 3 different static models.

<https://doi.org/10.1371/journal.pone.0306328.g003>

When considering the calibration of these models, no static model appears to perform better across all 3 modeling procedures than their dynamic counterpart. In the PM model, the crude static model has a better prediction error than either of the other static models but has the worst Brier score in the PP model. Likewise, SM2 has the worst prediction error in the PM model, but the best Brier score in the Cox model. Essentially, even though the dynamic model is not uniformly better than its static counterparts in any given model, it has a very similar AUC and prediction error/Brier score to the best static models.

The strong performance of SM2 and SM3 might be due to their use of the most current covariate information when predicting outcomes. In many health settings, it may be the case that the most current health information does most of the leg work in terms of predicting future outcomes, and that any modeling procedure that incorporates this information will perform relatively well.

The dynamic model seems to be most advantageous when using the PM approach, and least advantageous when applying a Cox modeling procedure. Thus, determining what type of summary measure is desired when modeling should influence a researcher's choice of modeling technique (dynamic or static) in a setting similar to ours. If a summary measure is desired in terms of survival duration, then a dynamic PM model might be best. Whereas, if the prediction is desired in terms of survival probability, and a standard Cox modeling procedure is desired, then a static model fit at baseline (such as SM2) might be a reasonable choice.

From this figure, we can see that SM2 and SM3 perform well in relation to the dynamic models. This may indicate that the most current health information may be of the greatest value when performing prediction.

4 Discussion

Dynamic risk prediction modeling is a vast research-scape and offers medical researchers a new paradigm for dealing with EHR data. In this paper, we have specifically touched on three different modeling techniques that apply the LM approach, but this accounts for a mere fraction of the various complex methods presented in recent literature.

Overall, the LM method proves to be intuitive, straightforward, and useful. Even though the dynamic models built from LM data sets were not an overwhelming favorite in terms of model performance, they were always at least as good as their static counterparts, and better in several specific cases.

One notable downside to the LM approach is the computational time involved in constructing the LM data set. The vast majority of computing time was devoted to this facet of modeling where the `cutLM()` function from the “dynpred” package in R was used. For smaller data sets, this step would not be too intensive, but with EHR data sets containing numerous subjects, this could become a process that requires extremely powerful computing equipment. On the other hand, simple static models could be rapidly built with little resources required.

Alternatively, a modeling procedure based on LM data sets could be employed in a similar fashion to Parast et al. where differences in the time-varying covariates are measured across the LM times and used as a predictor variable in model fitting. This is intuitive, however, Parast et al. found only marginal improvements in model calibration and no improvement in discrimination (as compared to a static model) while applying this methodology to Diabetes Prevention Program (DPP) data [10].

Since the CAPriCORN data is not available to the public, readers could perform our analysis using the “pbc2” data in the “dynpred” package in R. This data similarly measures survival in cirrhotic subjects and has been used to test the quality of DRP methods such as those presented by Yang et al. [11].

5 Conclusion

The ultimate goal of dynamic risk prediction is to improve the accuracy of personalized prediction rules for individual patients. As data collection becomes easier, faster, more expansive, and more accurate, this goal can become more attainable. While the quality of the predictions from the DRP models discussed in this paper may not be overwhelming, their use offers similar performance to static models and encourages the continued advancement in statistical

methodology. Given the prevalent use of risk prediction modeling in medical settings, even simple advances in prediction quality can benefit many patients in the future.

Supporting information

S1 Appendix.
(PDF)

Author Contributions

Conceptualization: Mitchell Paukner, Daniela P. Ladner, Lihui Zhao.

Data curation: Mitchell Paukner.

Formal analysis: Mitchell Paukner.

Funding acquisition: Daniela P. Ladner, Lihui Zhao.

Investigation: Mitchell Paukner.

Methodology: Mitchell Paukner.

Project administration: Lihui Zhao.

Supervision: Daniela P. Ladner.

Visualization: Mitchell Paukner.

Writing – original draft: Mitchell Paukner.

Writing – review & editing: Mitchell Paukner, Daniela P. Ladner, Lihui Zhao.

References

1. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Medical care*. 2010;p. S106–S113. <https://doi.org/10.1097/MLR.0b013e3181de9e17>
2. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*. 2017; 106(1):1–9. <https://doi.org/10.1007/s00392-016-1025-6> PMID: 27557678
3. Kohli R, Tan SSL. Electronic Health Records. *Mis Quarterly*. 2016; 40(3):553–574. <https://doi.org/10.25300/MISQ/2016/40.3.02>
4. Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*. 2017; 24(1):198–208. <https://doi.org/10.1093/jamia/ocw042> PMID: 27189013
5. Wang F, Zhang P, Qian B, Wang X, Davidson I. Clinical risk prediction with multilinear sparse logistic regression. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*; 2014. p. 145–154.
6. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: A deep learning approach. In: *Proceedings of the 2016 SIAM international conference on data mining*. SIAM; 2016. p. 432–440.
7. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*. 2007; 34(1):70–85. <https://doi.org/10.1111/j.1467-9469.2006.00529.x>
8. Andrinopoulou ER, Harhay MO, Ratcliffe SJ, Rizopoulos D. Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data. *International Journal of Epidemiology*. 2021; 50(5):1731–1743. <https://doi.org/10.1093/ije/dyab047> PMID: 33729514
9. Keogh RH, Seaman SR, Barrett JK, Taylor-Robinson D, Szczesniak R. Dynamic prediction of survival in cystic fibrosis: a landmarking analysis using UK patient registry data. *Epidemiology (Cambridge, Mass)*. 2019; 30(1):29. <https://doi.org/10.1097/EDE.0000000000000920> PMID: 30234550

10. Parast L, Mathews M, Friedberg MW. Dynamic risk prediction for diabetes using biomarker change measurements. *BMC medical research methodology*. 2019; 19(1):1–12. <https://doi.org/10.1186/s12874-019-0812-y> PMID: 31412790
11. Yang Z, Wu H, Hou Y, Yuan H, Chen Z. Dynamic prediction and analysis based on restricted mean survival time in survival analysis with nonproportional hazards. *Computer Methods and Programs in Biomedicine*. 2021; 207:106155. <https://doi.org/10.1016/j.cmpb.2021.106155> PMID: 34038865
12. Zhao L, Murray S, Mariani LH, Ju W. Incorporating longitudinal biomarkers for dynamic risk prediction in the era of big data: A pseudo-observation approach. *Statistics in medicine*. 2020; 39(26):3685–3699. <https://doi.org/10.1002/sim.8687> PMID: 32717100
13. Sepanlou SG, Safiri S, Bisignano C, Ikuta KS, Merat S, Saberifiroozi M, et al. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet gastroenterology and hepatology*. 2020; 5(3):245–266. [https://doi.org/10.1016/S2468-1253\(19\)30349-8](https://doi.org/10.1016/S2468-1253(19)30349-8)
14. Xu J, Kochanek KD, Murphy SL, Tejada-Vera B. Deaths: final data for 2014. *National Vital Statistics Reports*. 2016; 65(4). PMID: 27378572
15. Mokdad AA, Lopez AD, Shahrzaz S, Lozano R, Mokdad AH, Stanaway J, et al. Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. *BMC medicine*. 2014; 12(1):1–24. <https://doi.org/10.1186/s12916-014-0145-y> PMID: 25242656
16. Tapper EB, Parikh ND. Mortality due to cirrhosis and liver cancer in the United States, 1999–2016: observational study. *bmj*. 2018; 362. <https://doi.org/10.1136/bmj.k2817> PMID: 30021785
17. Chung W, Jo C, Chung WJ, Kim DJ. Liver cirrhosis and cancer: comparison of mortality. *Hepatology international*. 2018; 12(3):269–276. <https://doi.org/10.1007/s12072-018-9850-5> PMID: 29497986
18. Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*. 2003; 90(1):15–27. <https://doi.org/10.1093/biomet/90.1.15>
19. Magirr D. Non-proportional hazards in immuno-oncology: is an old perspective needed?. *arXiv preprint arXiv:200704767*. 2020;.
20. Pak K, Uno H, Kim DH, Tian L, Kane RC, Takeuchi M, et al. Interpretability of cancer clinical trial results using restricted mean survival time as an alternative to the hazard ratio. *AMA oncology*. 2017; 3(12):1692–1696. <https://doi.org/10.1001/jamaoncol.2017.2797> PMID: 28975263
21. Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of clinical Oncology*. 2014; 32(22):2380. <https://doi.org/10.1200/JCO.2014.55.2208> PMID: 24982461
22. Zhao L, Claggett B, Tian L, Uno H, Pfeffer MA, Solomon SD, et al. On the restricted mean survival time curve in survival analysis. *Biometrics*. 2016; 72(1):215–221. <https://doi.org/10.1111/biom.12384> PMID: 26302239
23. Paukner M, Chappell R. Window mean survival time. *Statistics in Medicine*. 2021; 40(25):5521–5533. <https://doi.org/10.1002/sim.9138> PMID: 34258772
24. Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime data analysis*. 2004; 10(4):335–350. <https://doi.org/10.1007/s10985-004-4771-0> PMID: 15690989
25. Andersen PK, Perme M Pohar. Pseudo-observations in survival analysis. *Statistical methods in medical research*. 2010; 19(1):71–99. <https://doi.org/10.1177/0962280209105020> PMID: 19654170
26. Xiang F, Murray S. Restricted mean models for transplant benefit and urgency. *Statistics in medicine*. 2012; 31(6):561–576. <https://doi.org/10.1002/sim.4450> PMID: 22238131
27. Zhao L, Feng D. Deep neural networks for survival analysis using pseudo values. *IEEE journal of biomedical and health informatics*. 2020; 24(11):3308–3314. <https://doi.org/10.1109/JBHI.2020.2980204> PMID: 32167918
28. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. 1996; 15(4):361–387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4%3C361::AID-SIM168%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4) PMID: 8668867