Dynamic Prediction of Competing Risk Events using Landmark Sub-distribution Hazard Model with Multivariate Longitudinal Biomarkers

Cai Wu* Liang Li† Ruosha Li‡

Abstract

The cause-specific cumulative incidence function (CIF) quantifies the subject-specific disease risk with competing risks. With longitudinally collected biomarker data, it is of interest to dynamically update the predicted CIF by incorporating the most recent biomarker as well as the cumulating longitudinal history. Motivated by a longitudinal cohort study of chronic kidney disease, we propose a framework for dynamic prediction of end stage renal disease using multivariate longitudinal biomarkers, accounting for the competing risk of death. The proposed framework extends the landmark survival modeling to the competing risks data, and implies a distinct sub-distribution hazard regression model defined at each landmark time. The model parameters, prediction horizon, longitudinal history and at-risk population are allowed to vary over the landmark time. When the measurement times of biomarkers are irregularly spaced, the predictor may not be observed at the time of prediction. Local polynomial is used to accommodate this situation and estimate the model parameters without explicitly imputing the predictor or modeling its longitudinal trajectory. The proposed model leads to simple interpretation of the regression coefficients and closed-form calculation of the predicted CIF. The estimation and prediction can be implemented through standard statistical software, with tractable computation. We conducted simulations to evaluate the performance of the estimation procedure and predictive accuracy. The methodology is illustrated with data from the African American Study of Kidney Disease and Hypertension.

Key Words: Competing risks; Dynamic prediction; Fine-Gray model; Landmark analysis; Longitudinal biomarkers; Prediction model.

1. Introduction

Patients with chronic kidney disease (CKD) are at increased risk of kidney failure. Accurate prediction of the timing of such an adverse clinical event is of great importance in clinical research and practice to facilitate preparation for renal replacement therapy and individualize clinical decisions [26], and its use has been recommended by clinical practice guidelines [14]. The typical kidney failure risk equations are "static" prediction models in the sense that they are developed from survival regression models that relate the predictors at an earlier time point, such as baseline, to the time of kidney failure [6, 27, 11]. For example, in the development of a widely used prediction model [28], a Cox proportional hazard model with time-independent covariates was used; the predictor variables include biomarkers, clinical and demographical variables, measured at baseline, and the outcome is the time from baseline to kidney failure. Since that model was developed from electronic health records, the baseline was the initial nephrology referral or lab test recorded in the database. Longitudinal data of those biomarkers, measured over many years between baseline and kidney failure, are available and potentially informative to the disease progression, but they are not used in prediction model development.

*Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, USA

†Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

‡Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

In statistical literature, the prediction of the risk of clinical events using longitudinal data is often referred to as dynamic prediction in the sense that the prediction can be updated with accumulating longitudinal data. Important fundamental work has been published in the last decade [29, 30, 31, 34, 24, 23]. There are a number of challenges when this methodology is applied to the prediction of kidney failure among CKD population. First, CKD patients have increased chance of mortality before reaching kidney failure. In our motivating data application (Section 6), death accounts for about one third of the overall terminal events (both kidney failure and death). Treating death as independent censoring or combining death and kidney failure into a composite endpoint does not lead to optimal intepretation of the model, and proper adjustment for competing risk is often needed in CKD studies [22]. Therefore, the dynamic prediction model must be able to accommodate competing risks. Second, previous literature has identified a large number of risk factors for kidney failure, including multiple biomarkers that are known to be causally associated with the disease progression. For example, the static prediction model of Tangri *et al* includes six biomarkers: estimated glomerular filtration rate (GFR), albuminuria, serum calcium, serum phosphate, serum bicarbonate, and serum albumin [28]. Put in the longitudinal context, it requires that the dynamic prediction model for kidney failure should be able to accommodate multiple biomarkers, and use their longitudinal trajectories to generate and update the risk prediction. In addition, previous literature demonstrated that some biomarkers, such as the estimated GFR, have nonlinear longitudinal trajectories with diverse progression patterns [15, 16]. This feature could add to the complexity of the statistical analysis if the dynamic prediction model involves modeling subject-specific longitudinal trajectories. Third, as a chronic disease, the CKD often lasts many years before the patient reaches the terminal clinical events such as kidney failure or death, during which time the patient may go through multiple stages of the disease. The strength of association between the biomarkers and the disease outcome could vary over time, leading to time-varying parameters in the model. Fourth, in our motivating data application as well as many other practical situations, such as in the electronic health records of CKD, patients do not follow a common pre-specified clinical visits schedule. Instead, the observed times of clinical visits, at which the biomarker data were obtained, appear to be irregularly spaced over time. Even if these observational times are random and non-informative in the sense that they are not related to the underlying health condition of the patients, this phenomenon still presents a challenge to the development of dynamic prediction model, as will be elucidated below.

In the statistical literature on dynamic prediction with competing risks data, Rué et al. [25] and Andrinopoulou et al. [2] modeled the joint distribution of longitudinal data and the competing risks outcomes with shared random effect models, and estimated the parameters with Markov chain Monte Carlo. When a large number of random effects are needed to accommodate multiple nonlinear longitudinal trajectories per subject, as in our motivating data application, the joint models are computationally difficult to estimate [12]. Cortese and Andersen [5], Nicolaie et al. [20] and Nicolaie et al. [21] studied the dynamic prediction with competing risks data using an alternative approach called landmark modeling [29], which is computationally simpler. Motivated by the specific needs of CKD research, our proposed methodology in this paper is innovative and different from the statistical literature above in some important aspects. First, the typical landmark approach, as adopted by the literature above, involves pre-specifying a number of landmark time points distributed over the follow-up period, creating a landmark dataset at each landmark time point that consists of at-risk subjects and their predictor variables and residual time to event at the landmark times, and fitting a time-to-event model to a "super" dataset with all the landmark datasets stacked on top of each other [29]. In our data application, the observational times of the clinical visits are irregularly spaced over time without a common

schedule for all the subjects. The gap times range widely, from a few months to a few years. If the predictor variable is "current value of the biomarker", as commonly used in this kind of prediction problems, then the predictor values are unknown for majority of the landmark times. Imputing the unknown value by the last or closest measurement is also problematic because that measurement could be years apart and because the progression of CKD includes both chronic process, where biomarkers change slowly, and acute episodes, where biomarkers change more quickly. The proposed method assumes that there is a sub-distribution hazard model specified any at landmark time during the follow-up, and all the model parameters, including the baseline sub-distribution hazard function, vary as a smooth function of the landmark time. A local polynomial method is used to estimate these parameters. This method does not require pre-specification of the landmark times, which is important given that there is currently no guideline in the literature on how to set the number and locations of the landmark times. It can also accommodate the randomly spaced observational times without *ad hoc* imputation. In addition, the proposed method estimates the regression coefficients and baseline hazard functions nonparametrically, and hence is more flexible compared with many implementations of the landmark approach, which involves parametric assumptions on the regression coefficient functions or baseline hazard [20, 29]. This feature is particularly attractive for our motivating data application, the African American Study of Kidney Disease and Hypertension (AASK), because it is among the longest prospective cohort studies of CKD, with up to 12 years of follow-up. It is very likely that the association between risk factors and outcomes change considerably over long follow-up time. Finally, our approach is embedded within the framework of Fine-Gray sub-distributional hazard model [9], while the previous works were based on cause-specific hazard model [20], pseudo-observations [21] and multi-state model [5]. The Fine-Gray model imposes a parsimonious relationship between predictors and the cumulative incidence function (CIF) of the event of interest, kidney failure in our application. It is convenient to use when the predictive target is the CIF because it does not involve a model for the competing event. In our application, the cause of death is not recorded. Hence, the model linking the biomarkers and all-cause death may not fit well in light of the underlying heterogeneous among patients who died during follow-up.

This paper is organized as follows. In section 2, we introduce the dynamic prediction model for competing risk data. In section 3, we provide model estimation procedures for the quantities of interest. In section 4 we propose estimators of predictive accuracy measures within the dynamic competing risk context. In section 5, we conduct numerical studies to evaluate the performance of the proposed methodology. The application to the AASK data is presented in section 6. Discussion and future work are presented in section 7.

2. The landmark dynamic prediction model for competing risks data

2.1 The notation and data structure

Let T_i and C_i be the time to the event-of-interest and censoring for subject i, and $\epsilon_i \in (1, \ldots, K)$ be the K causes of event. We observe $\tilde{T}_i = \min(T_i, C_i)$, the time to the event or censoring, whichever comes first, and $\Delta_i = 1(T_i \leq C_i)$, the event indicator. We observe $\Delta_i \epsilon_i \in (0, \ldots, K)$, where zero indicates censoring and non-zero indicates the type of the observed event. Without loss of generality, we assume $K = 2$ throughout this paper. In the context of the data application, event 1 denotes end stage renal disease (ESRD), the clinical event of interest, and event 2 denotes death, the competing event. Let $\boldsymbol{Y}_i = \{\boldsymbol{Y}_{i1}, \boldsymbol{Y}_{i2}, \ldots, \boldsymbol{Y}_{iq}\}$ denote the $n_i \times q$ matrix for subject i, with n_i repeated measurements for the multiple longitudinal covariates $1, 2, \ldots, q$. Each $\boldsymbol{Y}_i$ is measured at subject-specific

time points $\boldsymbol{t}_i = \{t_{i1}, t_{i2}, \ldots t_{in_i}\}$ ($t_{ij} < \tilde{T}_i$), and $\boldsymbol{Y}_i$ can include both time-dependent variables such as the biomarkers, and time-independent variables such as the baseline characteristics. In our modeling framework, t_{ij} are different across subjects, but for the same subject, we assume the multiple biomarkers are measured at the same visit time. We assume that the distribution of $\boldsymbol{t}_i$ is non-informative of $\boldsymbol{Y}_i$ and T_i. As in conventional survival analysis, we assume independent censoring. At any follow-up time u ($u < \tilde{T}_i$), we denote by $\boldsymbol{\mathcal{H}}_i(u)$ the observed covariate process within a history window $[u-\tau_2, u]$, where $\boldsymbol{\mathcal{H}}_i(u) = \{\boldsymbol{Y}_i(t_{ij}), t_{ij} \mid u-\tau_2 \leq t_{ij} \leq u;\ j = 1, \ldots, n_i\}$. We observe an independent and identically distributed training data set $\mathcal{D}_n = \{\tilde{T}_i, \Delta_i\epsilon_i, \boldsymbol{Y}_i, \boldsymbol{t}_i, i = 1, \ldots, n\}$, from which the dynamic prediction model is to be developed. Our interest is to estimate, for a future individual in the same population as the training data, denoted by subscript $_o$, the probability of ESRD in the next τ_1 years given survival up to time s and the covariate information in the history window: $\pi(\tau_1|s, \boldsymbol{\mathcal{H}}_o(s)) = P(T_o \in (s, s+\tau_1], \epsilon_o = 1|\tilde{T}_o > s, \boldsymbol{\mathcal{H}}_o(s))$.

We call τ_1 the $prediction\ horizon$, which quantifies how far into the future we want to predict the individual's risk of the event. The range of τ_1 is $(0, \infty)$. A small τ_1 indicates the prediction for the immediate future and a large τ_1 indicates the prediction for the distant future. The choice of τ_1 depends on the clinical context. In our data application, we use $\tau_1 = 1$ or 3 years. Patients with high risk of ESRD within the next year may be placed on the waiting list for renal replacement therapy, and patients with high risk of ESRD within the next 3 years should be monitored for diet, vital signs, and other complications.

In the following, we describe the construction of landmark dataset. To be specific, we are interested in predicting the residual lifetime $T(s) = T - s$ at a common landmark time s using the covariate history $\boldsymbol{\mathcal{H}}(s)$ among subjects with $s < \tilde{T}$. For subject i, we define $T_{ij} = T_i - t_{ij}$ and $C_{ij} = C_i - t_{ij}$ as the subject-specific residual time to the event and time to censoring since t_{ij}. For predictions up to the horizon τ_1, we artificially censor the residual times at τ_1, i.e., we observe $\tilde{T}_{ij} = \min(T_{ij}, C_{ij}, \tau_1)$ and the event indicator $\tilde{\delta}_{ij} = 1\big(T_{ij} \leq \min(C_{ij}, \tau_1)\big) \times \epsilon_i$, where $1(\cdot)$ is the indicator function. The artificial censoring may reduce the chance of misspecifying certain model assumptions [29], such as the proportional hazards assumption when a Cox model is used, or the proportional sub-distributional hazards assumption when a Fine-Gray model is used. The predictors are extracted from the data in the history window. For prediction at a clinical visit time s, the predictors may include the biomarker value at that clinical visit, or the rate of biomarker change, variation, or maximum/minimum value during the history window period (i.e., within the past τ_2 years of the prediction time). When $\tau_2 = s$, it is equivalent to use the entire history since baseline. The landmark dataset $\mathcal{L}_m = \{\tilde{T}_{ij}, \tilde{\delta}_{ij}, \boldsymbol{\mathcal{H}}_i(t_{ij});\ i = 1, \ldots, n, j = 1, \ldots, n_i\}$ will consist of $m = \sum_{i=1}^{n} n_i$ pieces of information from each subject. As will be shown in the following sections, the landmark competing risks model is specified on the derived time scale starting from t_{ij}: $t^* = \min\Big(u - t_{ij}, \tau_1\Big)$, where u is the time since baseline and $t^* \in (0, \tau_1]$.

2.2 Sub-distribution hazard model with baseline covariates

We first briefly review Fine -Gray's sub-distribution hazard (SDH) model for competing risks [9] with a $p \times 1$ baseline covariate vector $\boldsymbol{X}$. The quantity of interest is the cumulative incidence function (CIF) $\pi_1(u; \boldsymbol{X}) = P(T \leq u, \epsilon = 1|\boldsymbol{X})$. The SDH function for event 1 is defined as the instantaneous hazard of experiencing event 1 given survival and no previous event 1 occurrence at this moment: $\lambda_1(u; \boldsymbol{X}) = \lim_{\Delta u \to 0} \frac{1}{\Delta u} P(u \leq T \leq u + \Delta u, \epsilon = 1|\{T \geq u\} \cup \{T \leq u \cap \epsilon \neq 1\}, \boldsymbol{X}) = -\frac{d\log(1 - \pi_1(u; \boldsymbol{X}))}{du}$. Such a definition can be viewed as the hazard function for an improper random variable $1(\epsilon = 1) \times T + 1(\epsilon \neq$

$1) \times \infty$. The SDH model takes the following form:

$$\lambda_1(u; \boldsymbol{X}) = \lambda_{10}(u)\exp(\boldsymbol{\alpha}^T \boldsymbol{X}),$$

where the baseline SDH function $\lambda_{10}(u)$ is an unspecified non-negative function, and the log-SDH ratio $\boldsymbol{\alpha}$ ($p \times 1$) has a direct connection to the CIF. The CIF can be estimated as $\hat{\pi}_1(u; X) = 1 - \exp\Big\{ - \exp\big(\hat{\boldsymbol{\alpha}}_{PL}^T \boldsymbol{X}\big) \cdot \hat{\Lambda}_{10}(u) \Big\}$, where $\hat{\boldsymbol{\alpha}}_{PL}$ is the estimator for $\boldsymbol{\alpha}$ from the partial likelihood, and $\hat{\Lambda}_{10}(u)$ is a modified Breslow estimator obtained by plugging in $\hat{\boldsymbol{\alpha}}_{PL}$.

2.3 Landmark proportional sub-distribution hazard model

The proposed landmark model resets the follow-up time scale at each landmark time t_{ij} and treats the time-dependent covariate as if it is measured at the new baseline t_{ij}. Extending the SDH function for the baseline model in last section, the landmark SDH model at t_{ij} using the derived follow-up time scale t^* is denoted as:

$$\begin{aligned}\lambda_1(t^*|\boldsymbol{\mathcal{H}}_i(t_{ij}), t_{ij}) =& \lim_{\Delta t^* \to 0} \frac{1}{\Delta t^*} P\big(t^* \le T_{ij} \le t^* + \Delta t^*, \epsilon_i = 1| \\ & \{T_{ij} > t^*\} \cup \{0 < T_{ij} \le t^* \cap \epsilon_i \ne 1\}, \boldsymbol{\mathcal{H}}_i(t_{ij})\big). \end{aligned} \tag{1}$$

The landmark CIF can be expressed as:

$$\begin{aligned}\pi_1(t^*|\boldsymbol{\mathcal{H}}_i(t_{ij}), t_{ij}) &= P\big(T_i - t_{ij} \le t^*, \epsilon_i = 1 | T_i > t_{ij}, \boldsymbol{\mathcal{H}}_i(t_{ij})\big) \\ &= 1 - \frac{P\Big((T_{ij} > t^*) \cup (0 < T_{ij} < t^* \cap \epsilon_i \ne 1) | \boldsymbol{\mathcal{H}}_i(t_{ij})\Big)}{P\Big(T_i > t_{ij} | \boldsymbol{\mathcal{H}}_i(t_{ij})\Big)} \\ &= 1 - \exp\Big(- \{\Lambda_1(t^*|t_{ij}, \boldsymbol{\mathcal{H}}_i(t_{ij})) - \Lambda_1(0 - |t_{ij}, \boldsymbol{\mathcal{H}}_i(t_{ij}))\}\Big) \\ &= 1 - \exp\Big(- \int_0^{t^*} \lambda_1(t|\boldsymbol{\mathcal{H}}_i(t_{ij}), t_{ij}) dt\Big). \end{aligned} \tag{2}$$

To estimate this CIF, we fit the following landmark SDH working model to the landmark data set $\mathcal{L}_m = \{\tilde{T}_{ij}, \tilde{\delta}_{ij}, \boldsymbol{\mathcal{H}}_i(t_{ij}), t_{ij};\ i = 1, \ldots, n, j = 1, \ldots, n_i\}$,

$$\lambda_1(t^*|\boldsymbol{\mathcal{H}}_i(t_{ij}), t_{ij}) = \lambda_{10}(t^*, t_{ij})\exp\Big(\boldsymbol{\beta}^T(t_{ij})\tilde{\boldsymbol{Y}}_i(t_{ij})\Big),\ t^* \in (0, \tau_1], \tag{3}$$

where $\lambda_{10}(t^*, t_{ij})$ is a bivariate baseline SDH function defined on the derived follow-up time scale $t^* \in (0, \tau_1]$ and at subject-specific landmark times t_{ij}. We use $\tilde{\boldsymbol{Y}}_i(t_{ij})$ to denote the vector of predictors at visit time t_{ij}, which are functions of $\boldsymbol{\mathcal{H}}_i(t_{ij})$. The vector of time-varying coefficients $\boldsymbol{\beta}(t_{ij})$ is assumed to be smooth functions to allow the effect of the predictors to vary with the landmark time t_{ij}. Different from Zheng and Heagerty [34], the time-varying effects are functions of the landmark time t_{ij} instead of the derived follow-up time t^*. Therefore it differs from the usual time-varying coefficient model in survival analysis that is commonly used to deal with non-proportional hazards [3]. With the artificial censoring at τ_1, the covariate effect is more likely to be constant over $t^* \in (0, \tau_1)$ (but still vary with t_{ij}) and the proportional sub-distribution assumption is more likely to hold. [18].

3. Model estimation and dynamic prediction of the CIF

For estimation, we extend the kernel approach in Li et al. [17] to the competing risk context and formalize the idea of borrowing information from lagging covariates [1, 4]. Assume

that $\boldsymbol{\beta}(.)$ has a continuous second derivative in a neighborhood of s, by local linear approximation, $\boldsymbol{\beta}(t_{ij}) \approx \boldsymbol{\beta}(s) + \boldsymbol{\beta}'(s)(t_{ij} - s)$ for subject-specific time points t_{ij} around s. Our smoothing method occurs at the individual level, which is different from that of Cai and Sun [3], where smoothing occurs at the population level and the same weights are applied to all individuals.

The landmark dataset $\mathcal{L}_m$ are clustered multivariate time-to-event data with competing events, where the n_i records from the same subject are correlated. For clustered competing risk data [35], we define the counting process for event 1 as $N_{ij}(t^*) = 1(t_{ij} \leq T_i \leq t_{ij}+t^*, \tilde{\delta}_{ij} = 1)$ and the at-risk process $R_{ij}(t^*) = 1-N_{ij}(t^*-) = 1(T_i > t_{ij}+t^*)+1(t_{ij} \leq T_i \leq t_{ij} + t^*, \tilde{\delta}_{ij} \neq 1)$. Based on a local "working independence" partial likelihood function [35], for any given landmark point s, we can estimate the parameters $\boldsymbol{\beta}(s)$ using a kernel-weighted estimation equation, by borrowing biomarker measurements from the neighboring time points, $\{t_{ij} \in (s-h, s+h)\}$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(t_{ij}-s)\int_0^{t^*} w_{ij}(t)\cdot\Big\{\tilde{\boldsymbol{Z}}_{ij}(1, t_{ij}-s) - \bar{\boldsymbol{Z}}(\boldsymbol{\beta}(s), t)\Big\}\cdot \mathrm{d}N_{ij}(t). \tag{4}$$

$K(\cdot)$ is a kernel function with bounded support on $[-1, 1]$, $K_h(x) = h^{-1}K(x/h)$ and h is the bandwidth; $\tilde{\boldsymbol{Z}}_{ij}(1, t_{ij}-s) = \tilde{\boldsymbol{Y}}(t_{ij}) \otimes (1, t_{ij}-s)$, with $\otimes$ denoting the Kronecker product. We have the notations $\bar{\boldsymbol{Z}}(\boldsymbol{\beta}(s), t) = \dfrac{\hat{\boldsymbol{S}}^{(1)}(\boldsymbol{\beta}(s), t)}{\hat{\boldsymbol{S}}^{(0)}(\boldsymbol{\beta}(s), t)}$, and

$$\begin{aligned}\hat{\boldsymbol{S}}^{(r)}(\boldsymbol{\beta}(s), t) =& n^{-1}\sum_{l=1}^{n}\sum_{m=1}^{n_i} K_h(t_{lm}-s)w_{lm}(t)R_{lm}(t)\times\tilde{\boldsymbol{Z}}_{lm}(1, t_{lm}-s)^{\otimes r}\\ &\times\exp\Big(\boldsymbol{b}^T(s)\tilde{\boldsymbol{Z}}_{lm}(1, t_{lm}-s)\Big),\end{aligned} \tag{5}$$

where $\boldsymbol{b}(s) = \{\boldsymbol{b}_0(s), \boldsymbol{b}_1(s)\} = \{\boldsymbol{\beta}(s), \boldsymbol{\beta'}(s)\}$, $\tilde{\boldsymbol{Z}}^{\otimes 0} = 1$ and $\tilde{\boldsymbol{Z}}^{\otimes 0} = \tilde{\boldsymbol{Z}}$. The coefficient $\boldsymbol{\beta}(s)$ is estimated at each landmark s using $\hat{\boldsymbol{\beta}}(s) = \hat{\boldsymbol{b}}_0(s)$. The variance of $\hat{\boldsymbol{\beta}}(s)$ can be estimated by bootstrapping, and $w_{ij}(\cdot)$ in (4) denotes the inverse probability censoring weight for competing events, modified from Fine and Gray [9]:

$$w_{ij}(t^*) = 1\big(C_{ij} \geq T_{ij}\wedge t^*\big)\frac{G(t^*|s)}{G\big(T_{ij}\wedge t^*|s\big)},$$

where $G(t^*|s) = P(C_{ij} \geq t^*|s)$ is the censoring distribution of the residual censoring time at landmark s, and $\wedge$ denotes the minimum of the two values. We use a kernel-weighted Kaplan-Meier estimator for the residual censoring distribution, estimated from the residual time to censoring around s:

$$\widehat{G}(t^*|s) = \prod_{\zeta\in\Omega,\zeta\leq t}\Big\{1 - \frac{\sum_l K_h(t_{lm}-s)\cdot 1(\tilde{C}_{lm}=\zeta, \tilde{\delta}_{lm}=0)}{\sum_l K_h(t_{lm}-s)\cdot 1(\tilde{C}_{lm}\geq\zeta)}\Big\}.$$

Once we obtain the estimates of $\boldsymbol{\beta}(s)$, the baseline cumulative SDH function at time s can be estimated by plugging in $\hat{\boldsymbol{\beta}}(s)$:

$$\hat{\Lambda}_{10}(t^*, s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(t_{ij}-s)\int_0^{t^*}\frac{1}{\hat{\boldsymbol{S}}^{(0)}(\hat{\boldsymbol{\beta}}(s), t)}\hat{w}_{ij}(t)\mathrm{d}N_{ij}(t).$$

The conditional CIF for any future subject $_o$ can be estimated as

$$\begin{aligned}\hat{\pi}_1(t^*|s, \boldsymbol{\mathcal{H}}_o(s)) &= \hat{P}(s < T_o \leq s+t^*, \epsilon_o = 1|\tilde{T}_o > s, \boldsymbol{\mathcal{H}}_o(s))\\ &= 1 - \exp\Big(-\hat{\Lambda}_{10}(t^*, s)\times\exp\Big(\hat{\boldsymbol{\beta}}^T(s)\tilde{\boldsymbol{Y}}_o(s)\Big)\Big).\end{aligned} \tag{6}$$

4. Quantifying the dynamic predictive accuracy

In this section, we propose an adaptation of two well-established predictive accuracy measures, the time-dependent receiver operating characteristic (ROC) curve, in particular the area under the ROC curve (AUC); and the Brier score (BS). In the dynamic prediction framework, the time-dependent predictive accuracy measures are functions of two time scales, the landmark time s and the prediction horizon τ_1. The procedure for estimating sensitivity, specificity, and BS follows the non-parametric kernel-weighted approach of Wu and Li [33] in the competing risk context.

4.1 The dynamic time-dependent ROC curve and AUC

At any given landmark time s, we want to evaluate how well the risk score, i.e., the estimated CIF, discriminates between subjects with event of interest within the window $(s, s+\tau_1]$ versus those without. For any at-risk subject at time s who experiences the main event within the time interval $(s, s+\tau_1]$, that occurrence epsilon is defined as a case: $D^+(s,\tau_1) = \{i : s < T_i \leq s+\tau_1, \epsilon_i = 1\}$. When a subject is event-free at $s+\tau_1$, that occurrence is defined as a control: $D^-(s,\tau_1) = \{i : T_i > s+\tau_1\}$. An alternative definition for a control is to use the complementary set $\bar{D}^+(s,\tau_1) = \{i : (s < T_i \leq s+\tau_1, \epsilon_i \neq 1) \cup (T_i > s+\tau_1)\}$, including subjects who experience a competing event within the time interval $(s, s+\tau_1]$ or remain event-free at $s+\tau_1$. We present the estimators for the former one. A similar extension can be achieved for the latter. For simplicity, we use the notation $U(\tau_1|s)$ as the risk score to denote the predicted CIF. Given a threshold value $c \in (0,1)$, the time-dependent sensitivity and specificity functions are defined as $Se(c,s,\tau_1) = P\Big(U(\tau_1|s) > c | D^+(s,\tau_1)\Big)$ and $Sp(c,s,\tau_1) = P\Big(U(\tau_1|s) \leq c | D^-(s,\tau_1)\Big)$. The estimators of sensitivity and specificity are

$$\widehat{Se}(c,s,\tau_1) = \frac{\sum_{i\in\Re_s} \hat{W}_{1i}^{dyn} \cdot 1(U_i(\tau_1|s) > c)}{\sum_{i\in\Re_s} \hat{W}_{1i}^{dyn}}$$
$$\widehat{Sp}(c,s,\tau_1) = \frac{\sum_{i\in\Re_s} (1 - \sum_{k=1}^{K} \hat{W}_{ki}^{dyn}) \cdot 1(U_i(\tau_1|s) \leq c)}{\sum_{i\in\Re_s} (1 - \sum_{k=1}^{K} \hat{W}_{ki}^{dyn})},$$

where $W_{1i}^{dyn} = P\Big(T_i(s) \in (0,\tau_1], \epsilon_i = 1 | \tilde{T}_i(s), \epsilon_i, U_i\Big) = 1(\tilde{\delta}_{ij} = 0) \cdot \frac{F_1(\tau_1|U_i,s) - F_1(\tilde{T}_i(s)|U_i,s)}{S(\tilde{T}_i(s)|U_i,s)} + 1(\tilde{\delta}_{ij} = 1)$, $T_i(s) = T_i - s$, $\tilde{T}_i(s) = \tilde{T}_i - s$ and U_i is short for $U_i(\tau_1|s)$. $\Re_s$ is risk set within the neighborhood of s which includes the most recent record at t_{ij} for each subject i $\{i : \tilde{T}_i \geq s, |t_{ij} - s| \leq |t_{ij'} - s|, \forall j' = 1,2,\ldots,n_i, t_{ij} \in (s-h, s+h)\}$. $F_1(\cdot|U_i,s) = P(T_i(s) \leq \cdot, \epsilon_i = 1 | U_i, s)$ and $S(\cdot|U_i,s) = P(T_i(s) \geq \cdot | U_i, s)$

For estimating the conditional probability weight W_{1i}^{dyn}, we treat the at-risk data set at landmark s as the new baseline data set. The time-dependent ROC curve is a plot of sensitivity $Se(c,s,\tau_1)$ over 1-specificity $1 - Sp(c,s,\tau_1)$, i.e., for $x \in [0,1]$, $\widehat{ROC}(x,s,\tau_1) = \widehat{Se}(\widehat{Sp}^{-1}(1-x,s,\tau_1),s,\tau_1)$. The AUC is estimated as $\widehat{AUC}(s,\tau_1) = \int_0^1 \widehat{ROC}(x,s,\tau_1)\mathrm{d}x$.

4.2 The dynamic time-dependent Brier score

The time-dependent BS under the dynamic competing risk framework is defined as $BS(\tau_1,s) = E\Big(1(s < T \leq s+\tau_1, \epsilon = 1) - U(\tau_1|s)\big|T > s\Big)^2$, where $1(\cdot)$ is the indicator function. Applying the weight W_{1i}^{dyn}, the BS can be estimated as

$$\widehat{BS}(\tau_1, s) = \frac{1}{n_s} \sum_{i=1}^{n_s} \Big(\hat{W}_{1i}^{dyn} \times (1 - U(\tau_1|s))^2 + (1 - \hat{W}_{1i}^{dyn}) \times (0 - U(\tau_1|s))^2 \Big),$$

where n_s is the number of subjects at risk at landmark time s.

The AUC and BS assess different aspects of the predictive model. AUC evaluates the discrimination between a case and a control, and BS quantifies the deviance of the predicted probability from the observed data. A model with perfect discrimination will have $AUC = 1$, while AUC close to 0.5 indicates poor discrimination that resembles a random guess. BS is a prediction error metric, with smaller values indicating better prediction.

5. Simulation

Our simulation addresses two goals: 1) at a single landmark time s, to evaluate the finite sample performance of the kernel-weighted estimation for the landmark competing risks model parameters at time s; and 2) to evaluate the accuracy of dynamic prediction in a longitudinal context.

Simulation at a given landmark time s It is challenging to generate data from the joint distribution of the longitudinal biomarker data and competing risks outcomes so that the SDH model holds at all the landmark times $\{t_{ij}\}$. This is a well known difficulty in the study of landmark dynamic prediction models [29, 17], though limited progress has been made in problems without competing risks [34]. It remains unclear whether a joint distribution of the longitudinal and competing risks data exists and satisfies the modeling assumption of the landmark dynamic prediction model in Section 2.3. For this reason, researchers often view the landmark dynamic model as a working model, as long as it provides adequate approximation to the data at all the landmark times of interest. This difficulty also led researchers to evaluate the numerical performance of the landmark models using data simulated from the shared random effects model [19, 13]. Since the landmark model always works under misspecification in such situations, the bias and efficiency of the estimators are difficult to interpret. In light of the similar difficulty in our competing risks problem, we resort to a simple albeit approximate approach to evaluating the quality of the proposed local linear estimation, at a single landmark time s, as described below.

We simulated a cross-sectional time-to-event data set at a given landmark s, e.g., $s = 3$, which was treated as the baseline for the purpose of this simulation. Scattered individual measurement times $\{t_{ij}\}$ and the associated biomarker values $\boldsymbol{Y}_i(t_{ij})$ were simulated within a small neighborhood of s. The landmark competing risks model in Section 2.3 was used to generate independent times to competing risks data starting from each t_{ij}, following the simulation algorithm in Fine -Gray [9]. The log-SDH $\boldsymbol{\beta}(s)$ is assumed to be a quadratic function of s (Web Appendix B Figure 1). Note that this is not a really a landmark dataset because each subject only has one t_{ij}. Nonetheless, this dataset exactly satisfies the working model (3) so that we can use it to study the numerical performance of the proposed local linear estimation in a small neighborhood of s. Specifically, we evaluate the bias of estimating $\boldsymbol{\beta}(s)$ and the baseline CIF (Web Appendix Figure 2), $\pi_0(t^*; s) = 1 - \exp\Big(- \int_0^{t^*} \lambda_{10}(t, s)dt \Big)$, as well as the selection of the kernel bandwidth.

The results are presented in Figure 1. The three columns from left to right are the plots of the estimated log-SDH ratio, absolute bias percentage and mean squared error (MSE)

against different bandwidths. The rows from top to bottom correspond to the three increasing sample sizes. For the plot of the log-SDH ratio (column 1), the mean estimated $\boldsymbol{\beta}(s)$ at $s = 3$ over $1,000$ Monte Carlo repetitions is close to the true value (red horizontal line) at small bandwidth (e.g. 0.3 and 0.5). With increased bandwidth, the mean estimator shows increasing downward bias. This is because the true $\boldsymbol{\beta}(s)$ function is concave (Web Appendix B Figure 1), and the local linear fit underestimates it at the peak as the bandwidth increases. The empirical standard errors (ESDs), shown in Figure 1 as the vertical whiskers, shrink with the increased bandwidth since more data points are included in the kernel estimation. From top to bottom, the ESDs decrease when the sample size increases. Column 2 shows that the absolute bias percentage generally increases with the bandwidth, except when the bandwidth is very small, in which case larger finite-sample bias may result due to very few data points available in the neighborhood defined by the bandwidth. In column 3, the U-shaped MSE curve is a demonstration of the typical bias-variance trade-off in kernel estimation. Overall, the percentage of absolute bias for the log-SDH ratio is very small, within 2% for middle ranged bandwidths (the horizontal dashed line in column 2). When $\boldsymbol{\beta}(s)$ is a linear function of s, the absolute bias percentage is always within 1% and is robust to bandwidth increase (Web Appendix B Figure 3). The results from this simulation suggests that the proposed local linear estimation works as expected from typical local polynomial estimators [8].

Simulation on the accuracy of dynamic prediction in the longitudinal context In this simulation, we simulated both longitudinal and competing risks data, fit a landmark SDH model, and evaluate its prediction accuracy at various landmark times. As discuss above, the landmark SDH model is a working model in this simulation. To minimize the impact of model misspecification, we considered two scenarios: the biomarker is non-informative (S1) and the biomarker is informative (S2). For S1, the longitudinal biomarker is not associated with the SDH function, and we expect to demonstrate that its contribution to the prediction accuracy in the landmark SDH model is zero. For S2, we introduce various levels of association between the longitudinal biomarker and competing risks event, and expect to demonstrate that the prediction accuracy of landmark SDH model improves with increasing association between biomarker and competing risk event. Details of the data generation process are described in the Web Appendix A.

Under S1, we consider non-informative biomarkers that have zero effect on the survival outcome, i.e., $\boldsymbol{\beta}(s) \equiv \boldsymbol{\beta} = \{\beta_{kq}\} = 0$ for event type $k = 1, 2$. The event times are generated from the model of Fine & Gray [9]. The model is a special case for our dynamic model at $\boldsymbol{\beta}(s) \equiv \boldsymbol{\beta} = 0$ and $\lambda_{10}(t, s) \equiv \lambda_{10}(t)$. We expect to see that the effect of baseline covariate X is constant and unbiased from the true value, the effects of biomarkers $\{\boldsymbol{Y}_{\cdot q}\}$ are constantly zero, and the baseline CIF at different landmark times s is close to the true conditional CIF. Under S1, we do not impose artificial censoring when estimating the baseline SDH since there is no violation of the proportional sub-distribution assumption here. In Web Appendix B Figure 4, we plot the log-SDH ratios for the three longitudinal biomarkers and the baseline covariate over the landmark time grids s. The estimated log-SDHs for the longitudinal biomarkers are consistent along the zero horizontal line where the true effect lies. The estimated log-SDH for the baseline covariate is consistent with its true value at -1.5 except for some deviation at the end of follow up, due to limited sample size. The estimated baseline CIF from 500 Monte Carlo repetitions is close to the conditional CIF at selected landmark times. The estimated baseline CIF $\pi_0(t^*; s)$ is a bivariate surface of s and t^*, as shown in the contour plot. The contour plot of the mean $\hat{\pi}_0(t^*; s)$ is almost identical to that of the true CIF except at the borders where the sample size is small.

Under S2, we simulate time-to-event data from the joint frailty model of the longitudinal biomarkers and the competing risk event times [7]. Although the covariate effect is misspecified under S2, we expect to see the biomarkers having an informative effect that is significantly different from zero. In Web Appendix Figure 5, the effects of the three biomarkers are all significantly different from zero and are in the right direction.

Under both S1 and S2, we compare the predictive performance of models with and without longitudinal biomarkers using measures of discrimination and calibration. For discrimination, we consider the true positive (TP) fraction, false positive (FP) fraction at a given threshold value, and AUC as a global summary. For calibration, we use the BS. We evaluate all the predictive accuracy measurements at three landmark points $s = 3, 5, 7$ with the prediction horizon $\tau_1 = 1, 3$. For each simulation, the proposed model is fit to a training data set and the predictive accuracy measures are calculated based on i.i.d. samples of a validation data set. The results are calculated as the mean of 1000 Monte Carlo repetitions. Table 1 presents the dynamic predictive accuracy measures for S1 and S2. The full model includes longitudinal biomarkers with time-varying effects, and the null model only includes the baseline covariate. For S1, when all biomarkers are non-informative, the predictive accuracy measures for the full model (M1) and the null model (M0) are all very similar. In contrast, when the three longitudinal biomarkers are informative, including them in the prediction model greatly increases the AUC and TP and decreases the FP and prediction error.

6. Application to the AASK data

In this section, the proposed method will be illustrated using the AASK dataset. The AASK study includes 1,094 African Americans of age 18 to 70 years who were diagnosed with hypertensive renal disease and had baseline eGFRs between $20-65mL/min/1.73m^2$[32]. Subjects were followed up every 6 months, with up to 12 years of longitudinal data collected at each visit. Each subject was closely monitored for their clinical events, such as non-terminal events of acute kidney injury, hospitalization and cardiovascular comorbidities; and more importantly the terminal event of ESRD and death. By the end of the study, 318 (29%) individuals developed ESRD and 176 (16%) died before developing ESRD. The outcome of interest is the time to ESRD, defined as the time to dialysis or transplantation; and the time to death is treated as a competing event. The median time to ESRD is 4.3 years and the median time to death is 5.2 years. We chose clinically relevant prediction horizons of $\tau_1 = 1$ or 3 years and illustrated the dynamic prediction at years 3, 5, and 7. The key longitudinal biomarker is eGFR. The number of repeated measurements for eGFR ranged from 3 to 30, with over 50% of individuals providing 17 or more measurements. In addition to the current value of eGFR at a clinical visit, we derive the rate of change in eGFR during the history window of $\tau_2 = 3$ years. The estimation of eGFR rate of change followed the approach in Li et al. [17]. Additional biomarkers included longitudinal measurements of serum albumin (Alb), urine protein to creatinine ratio (UP/Cr), serum phosphorus (Phos) and urine potassium (Upot). The predictors in our landmark SDH model include: the patient's current age, eGFR, Alb, UP/Cr, Phos and Upot at the clinical visit, the eGFR rate of change in the past $\tau_2 = 3$ years (eGFR.slope), and a binary indicator of any hospitalization within the past year.

For the competing events of ESRD and death, we fit landmark SDH models separately using the same set of candidate predictors (Web Appendix Figure 6). The eGFR, its rate of change, and log UP/Cr are significantly associated with time to ESRD but not with time to death. In contrast, age, Alb and hospitalization are risk factors related to death. This indicates that the progression to ESRD and death may be related to different pathological

processes, which justifies the proposal of modeling the competing events separately rather than as a composite outcome. After removal of the non-significant covariates, the final model for ESRD includes eGFR, eGFR.slope, log UP/Cr and Phos, and the final model for death included age, Alb, log-Upot and hospitalization (Web Appendix Figure 7). We conduct bandwidth selection using 5-fold cross-validation. Predictive accuracy metrics are evaluated in the cross-validation dataset, and the results are robust to different bandwidths (up to 3 digits after the decimal point). Therefore, we used the bandwidth of $h = 1.5$ in the final model, which provides a relatively smooth curve for the log-SDH ratio curve. The surface plots of the CIF for ESRD and death are presented in Figure 2.

Figure 3 presents the longitudinal profiles and individual dynamic predictions from three AASK subjects: subject 1 was event-free by the end of the study, subject 2 experienced ESRD after 7.5 years, and subject 3 died after 9.7 years. We demonstrated the biomarker values with real-time predicted risk probabilities for ESRD and death within the next 3 years. The risk prediction is dynamically updated using the most current biomarker values and short-term medical history (episodes of hospitalization in grey vertical bar). Subject 1 demonstrates the case of a patient with stable disease, indicated by the stable biomarkers and a result of minimum predicted risk probability. A typical progression to ESRD (subject 2) is associated with a decline in eGFR over time, and increased proteinuria, quantified by log-UP/Cr. As expected, the predicted risk for ESRD for subject 2 increases after year 5 in response to the drastically decreased eGFR and a surge of log-UP/Cr. In contrast, the risk of death for subject 2 vaguely increases, mostly explained by the Alb level and hospitalization around year 7. For subject 3, the relatively stable eGFR and log-Up/Cr decreases the subject's susceptibility to ESRD, but the frequent hospitalization and decreasing Alb level are associated with increased risk of death, possibly due to other comorbidities. We did not estimate the model parameters or calculate the predictive probabilities after year 8 because the number of observed clinical events are relatively small near the end of the follow-up period, which would result in relatively low efficiency of the estimated prediction.

Figure 4 presents the same profiles and their dynamic CIF up to 3 years given biomarkers available at landmark years $s = 3, 5, 7$. For the censored subject 1, the predicted CIF for both ESRD and death are very flat over the three landmark times. In contrast, predicted CIF for ESRD for subject 2 emerged at landmark year 5. The predicted CIF of ESRD for this subject is further elevated by year 7, followed by an event of ESRD shortly afterwards. The predicted CIF of ESRD for subject 3 remains flat overtime. In contrast, the subject is predicted to have an elevated risk of death at year 7, after frequently being hospitalized. The subject eventually died at year 9.6 without experiencing renal failure.

In Table 2, we summarize the predictive accuracy of landmark SDH models for two sets of prediction horizons $\tau_1 = 1, 3$ at three landmark years $s = 3, 5, 7$. The model for the time to ESRD achieved good discrimination in between subjects who experience ESRD in the next τ_1 years and those who are event-free, with AUCs ranged between 0.93-0.96. Using a threshold value of 0.05, the sensitivity (TP) and specificity (1-FP) can be well controlled to be 0.80-0.90 for all scenarios. Both discrimination and calibration measures are very similar in predicting outcomes 1 year and 3 years ahead. In contrast, the landmark SDH model for time to death discriminates no better than a random guess, resulting in AUCs around or lower than 0.5. The prediction errors are also at least twice as large as those from predicting ESRD. This indicates that even if we found potential prognostic factors associated with time to death, they are not necessarily good predictors. More importantly, the AUCs from the proposed model are improved from the previous studies where AUCs are around 0.8 and always less than 0.9 [17, 19]. One possible explanation is that previous studies treat ESRD and death as a composite outcome. This introduces noise and diminishes the pre-

dictive capability of predicting ESRD, since the outcome of death is difficult to predict in such a context. The ROC curves for predicting ESRD are plotted in Web Appendix Figure 8.

7. Discussion

For CKD patients, estimating the time to ESRD is crucial for the timely treatment management. Dynamic prediction is an attractive tool for this purpose because it is adaptive to the changing health condition and prognostic history of the patient. It enables real-time monitoring of the risk of the patient. In this paper, we develop novel methodology for dynamic prediction of ESRD among the CKD patients and overcome a number of analytical hurdles, including competing events of death, irregularly spaced clinical visit times, multiple biomarkers with complicated longitudinal trajectories, and time-varying at-risk population, and time-changing covariate-outcome association. Our proposed methodology is flexible because the model parameters are estimated non-parametrically. Hence, it can effectively mitigate the risk of model misspecification. This feature is very important for dynamic prediction models because, as explained in Section 5, the landmark dynamic prediction model is a working model and needs to provide adequate approximation to the data at all landmark times. Another advantage of the proposed methodology is that it is computationally simple and can be implemented through standard statistical software for competing risks analysis, regardless of how many longitudinal biomarkers are included as predictors. In this paper, the estimation process was accomplished with the available R function $coxph()$ after translating the competing events into a counting process [10]. We believe that the simplicity in computation makes the proposed methodology attractive for various practical situations, including applications with large dataset, large number of biomarkers with complicated longitudinal trajectories, and other longitudinal prognostic information that cannot be easily modeled at individual-level (e.g., hospitalization and medication history).

Our kernel-based estimation approach relies on the assumption that the clinical visit times are non-informative. Future work is needed to study dynamic prediction when the frequency of clinical visits is related to the health condition of the patients. The predictors in our proposed model framework include pre-specified features extracted from the data history. Automatic extraction of predictive features from the longitudinal history is another topic that will pursued in our future research.

Supplementary Materials

Web Appendices and Figures referenced in this paper are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article. This research was supported by grants from the U.S. National Institutes of Health (5P30CA016672 and 5U01DK103225).

References

[1] Andersen, P. K. and Liestøl, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics* **4,** 633–649.

[2] Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J., and Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research* **26,** 1787–1801.

[3] Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics* **30,** 93–111.

[4] Cao, H., Churpek, M. M., Zeng, D., and Fine, J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *Journal of the American Statistical Association* **110,** 1187–1196.

[5] Cortese, G. and Andersen, P. K. (2010). Competing risks and time-dependent covariates. *Biometrical Journal* **52,** 138–158.

[6] Echouffo-Tcheugui, J. B. and Kengne, A. P. (2012). Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Medicine* **9,** e1001344.

[7] Elashoff, R. M., Li, G., and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64,** 762–771.

[8] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall.

[9] Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* **94,** 496–509.

[10] Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics* **67,** 39–49.

[11] Greene, T. and Li, L. (2017). From static to dynamic risk prediction: Time is everything. *American Journal of Kidney Diseases* **69,** 492–494.

[12] Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology* **16,** 117.

[13] Huang, X., Yan, F., Ning, J., Feng, Z., Choi, S., and Cortes, J. (2016). A two-stage approach for dynamic prediction of time-to-event distributions. *Statistics in Medicine* **35,** 2167–2182.

[14] KDIGO CKD Work Group (2013). Kidney Disease: Improving Global Outcomes (KDIGO) 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl.* **3,** 1–150.

[15] Li, L., Astor, B. C., Lewis, J., Hu, B., Appel, L. J., Lipkowitz, M. S., Toto, R. D., Wang, X., Wright, J. T., and Greene, T. H. (2012). Longitudinal progression trajectory of gfr among patients with ckd. *American Journal of Kidney Diseases* **59,** 504–512.

[16] Li, L., Chang, A., Rostand, S. G., Hebert, L., Appel, L. J., Astor, B. C., Lipkowitz, M. S., Wright, J. T., Kendrick, C., Wang, X., et al. (2013). A within-patient analysis for time-varying risk factors of ckd progression. *Journal of the American Society of Nephrology* pages ASN–2013050464.

[17] Li, L., Luo, S., Hu, B., and Greene, T. (2017). Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease. *Statistics in Biosciences* **9,** 357–378.

[18] Liu, Q., Tang, G., Costantino, J. P., and Chang, C.-C. H. (2016). Robust prediction of the cumulative incidence function under non-proportional subdistribution hazards. *Canadian Journal of Statistics* **44,** 127–141.

[19] Maziarz, M., Heagerty, P., Cai, T., and Zheng, Y. (2017). On longitudinal prediction with time-to-event outcome: Comparison of modeling options. *Biometrics* **73,** 83–93.

[20] Nicolaie, M., Houwelingen, J., Witte, T., and Putter, H. (2012). Dynamic prediction by landmarking in competing risks. *Statistics in Medicine* **32,** 2031–2047.

[21] Nicolaie, M., van Houwelingen, J., de Witte, T., and Putter, H. (2013). Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. *Biometrics* **69,** 1043–1052.

[22] Noordzij, M., Leffondré, K., van Stralen, K. J., Zoccali, C., Dekker, F. W., and Jager, K. J. (2013). When do we need competing risks methods for survival analysis in nephrology? *Nephrology Dialysis Transplantation* **28,** 2670–2677.

[23] Proust-Lima, C. and Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics* **10,** 535–549.

[24] Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67,** 819–829.

[25] Rué, M., Andrinopoulou, E.-R., Alvares, D., Armero, C., Forte, A., and Blanch, L. (2017). Bayesian joint modeling of bivariate longitudinal and competing risks data: An application to study patient-ventilator asynchronies in critical care patients. *Biometrical Journal* **59,** 1184–1203.

[26] Tangri, N., Ferguson, T., and Komenda, P. (2017). Pro: Risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrology Dialysis Transplantation* **32,** 748–751.

[27] Tangri, N., Kitsios, G., Inker, L., Griffith, J., Naimark, D., Walker, S., Rigatto, C., Uhlig, K., Kent, D., and Levey, A. (2013). Risk prediction models for patients with chronic kidney disease: a systematic review. *Annals of Internal Medicine* **158,** 596–603.

[28] Tangri, N., Stevens, L. A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., Levin, A., and Levey, A. S. (2011). A predictive model for progression of chronic kidney disease to kidney failure. *Journal of the American Medical Association* **305,** 1553–1559.

[29] van Houwelingen, H. and Putter, H. (2011). *Dynamic prediction in clinical survival analysis*. CRC Press.

[30] Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* **34,** 70–85.

[31] van Houwelingen, H. C. and Putter, H. (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime data analysis* **14,** 447–463.

[32] Wright Jr, J. T., Bakris, G., Greene, T., Agodoa, L. Y., Appel, L. J., Charleston, J., Cheek, D., Douglas-Baltimore, J. G., Gassman, J., Glassock, R., et al. (2002). Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the aask trial. *Journal of the American Medical Association* **288,** 2421–2431.

[33] Wu, C. and Li, L. (2018). Quantifying and estimating the predictive accuracy for censored time-to-event data with competing risks. *Statistics in Medicine* **37(21),** 3106-3124.

[34] Zheng, Y. and Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics* **61,** 379–391.

[35] Zhou, B., Fine, J., Latouche, A., and Labopin, M. (2012). Competing risks regression for clustered data. *Biostatistics* **13,** 371–383.

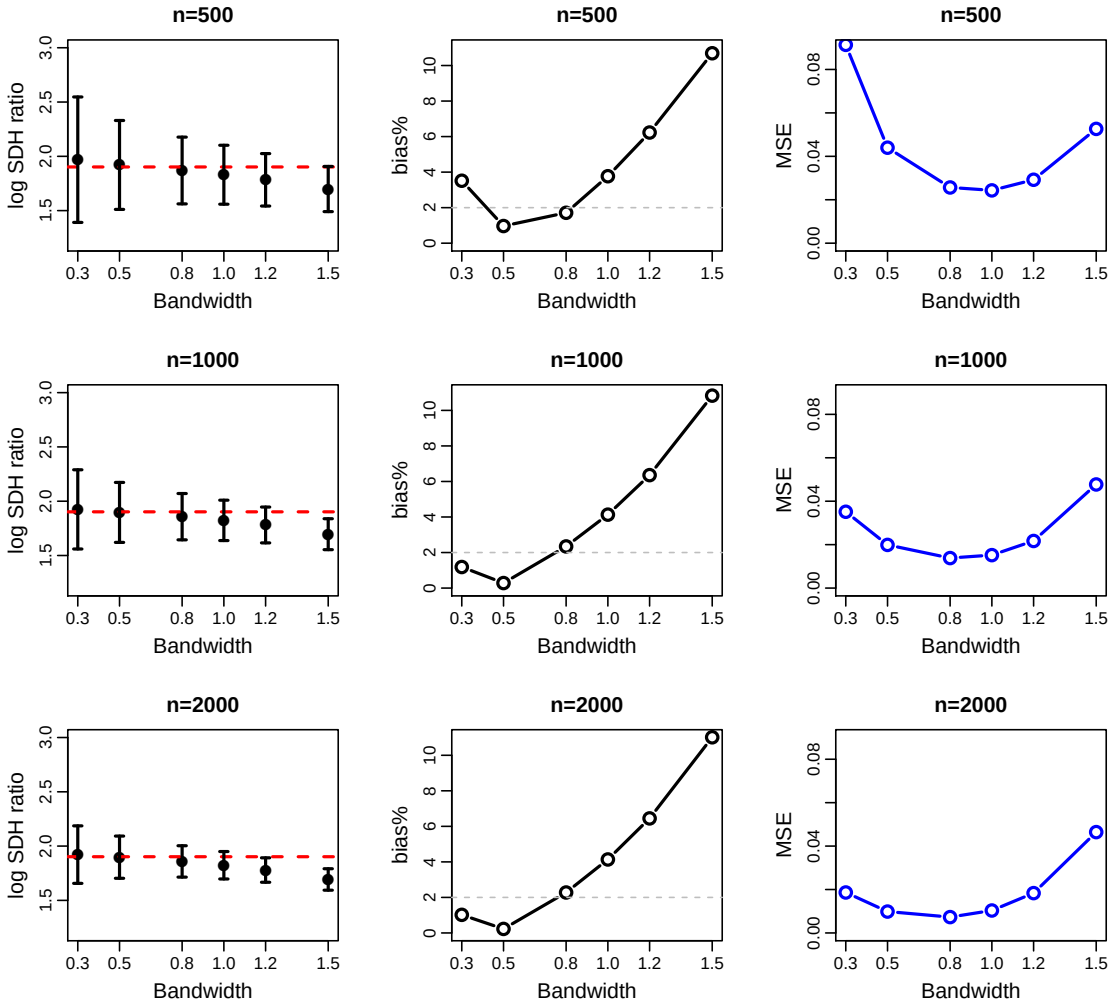

Figure 1: Simulation results for finite sample performance of kernel-weighted estimation at a single landmark time $s = 3$. The log-SDH ratio, absolute bias percentage and mean squared error (MSE) are plotted against the bandwidth on the horizontal axis under different sample sizes. The true time-varying log-SDH ratio is non-linear over time (see Web Appendix B).

Table 1: Estimates (EST) and empirical standard errors (ESD) of predictive accuracy metrics comparing the model with biomarkers M_1 and the null model without biomarkers M_0 with prediction horizon $\tau_1 = 3$. S1: scenario 1 for non-informative biomarker effects. S2: scenario 2 for informative biomarker effects. AUC: area under the ROC curve comparing the group experiencing the event of interest with the event-free group. $TP(c)$: true positive fraction at threshold c. $FP(c)$: false positive fraction at threshold c. BS: Brier score, the mean squared error for the predicted risk probability. Sample size $n = 500$, $replications = 500$.

			AUC		$TP(0.25)$		$FP(0.25)$		BS	
			M_1	M_0	M_1	M_0	M_1	M_0	M_1	M_0
S1	$s=1$	EST	0.714	0.716	0.483	0.476	0.200	0.203	0.140	0.142
		ESD	0.031	0.031	0.076	0.142	0.044	0.100	0.010	0.010
	$s=3$	EST	0.718	0.720	0.740	0.724	0.420	0.417	0.185	0.188
		ESD	0.029	0.029	0.064	0.133	0.070	0.146	0.011	0.013
	$s=5$	EST	0.722	0.727	0.736	0.707	0.369	0.378	0.184	0.193
		ESD	0.037	0.035	0.084	0.201	0.095	0.209	0.014	0.021
S2	$s=1$	EST	0.873	0.592	0.586	0.181	0.074	0.116	0.082	0.119
		ESD	0.022	0.042	0.075	0.212	0.019	0.165	0.009	0.017
	$s=3$	EST	0.840	0.540	0.742	0.479	0.158	0.416	0.135	0.195
		ESD	0.024	0.036	0.058	0.326	0.035	0.318	0.011	0.017
	$s=5$	EST	0.819	0.502	0.824	0.523	0.221	0.498	0.163	0.237
		ESD	0.029	0.041	0.061	0.390	0.061	0.388	0.014	0.021

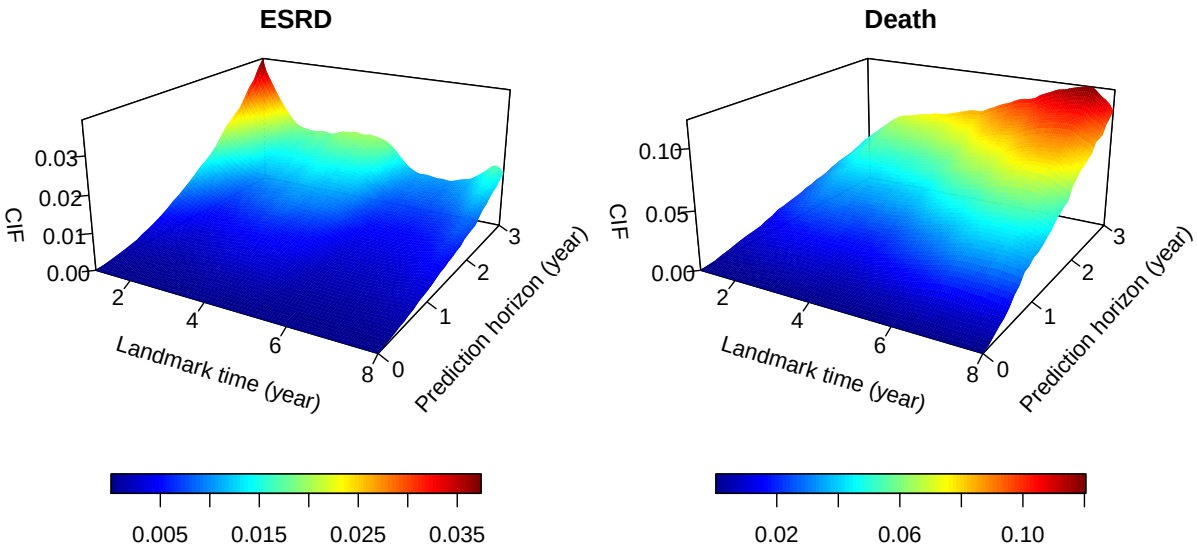

Figure 2: Estimated surface of the cumulative incidence function over the landmark years and prediction horizons. This shows an example population with age =55, eGFR =45 $(ml/min/1.73m^2)$, eGFR.slope = 0, UP/Cr = 0.3, albumin = 4 (g/dL), and hospitalized within the past year.

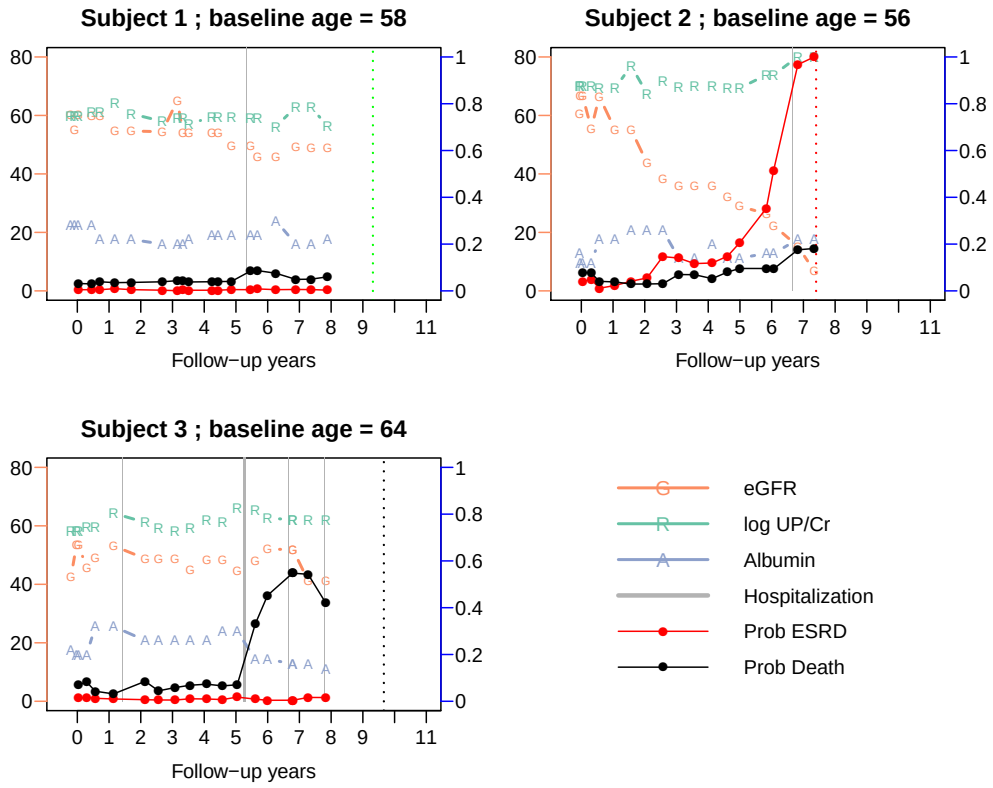

Figure 3: Individual risk predictions for three selected subjects: subject 1 censored (dashed vertical gray line), subject 2 with ESRD (dashed vertical red line) and subject 3 dead (dashed vertical black line). The three biomarkers are plotted up 8 years from baseline: "G" is eGFR $(ml/min/1.73m^2)$, "R" is log-urine protein-to-creatinine ratio (g/g), and "A" is albumin(g/dL). The connected red dots are predicted probabilities of ESRD within a horizon of $\tau_1 = 3$ years. The gray vertical bars represent episodes of hospitalization, with the two vertical borders being admission and discharge dates. The connected blue dots are the predicted probability of death within $\tau_1 = 3$ years. The left y-axis is the scale for eGFR, and the right y-axis is the scale for predicted probabilities (0 to 1). The other two biomarkers, log-UP/CR and albumin, are re-scaled to be displayed properly in the same plot with eGFR but the actual scales are not shown. The dynamic predicted probabilities of ESRD are calculated using the dynamic SDH model with four predictors: eGFR, eGFR slope in the past three years, most recent log-UP/CR and phosphorus. The dynamic predicted probabilities of death are calculated using the dynamic SDH model with four predictors: current age, serum albumin, any hospitalization within the past year, and log urine potassium.

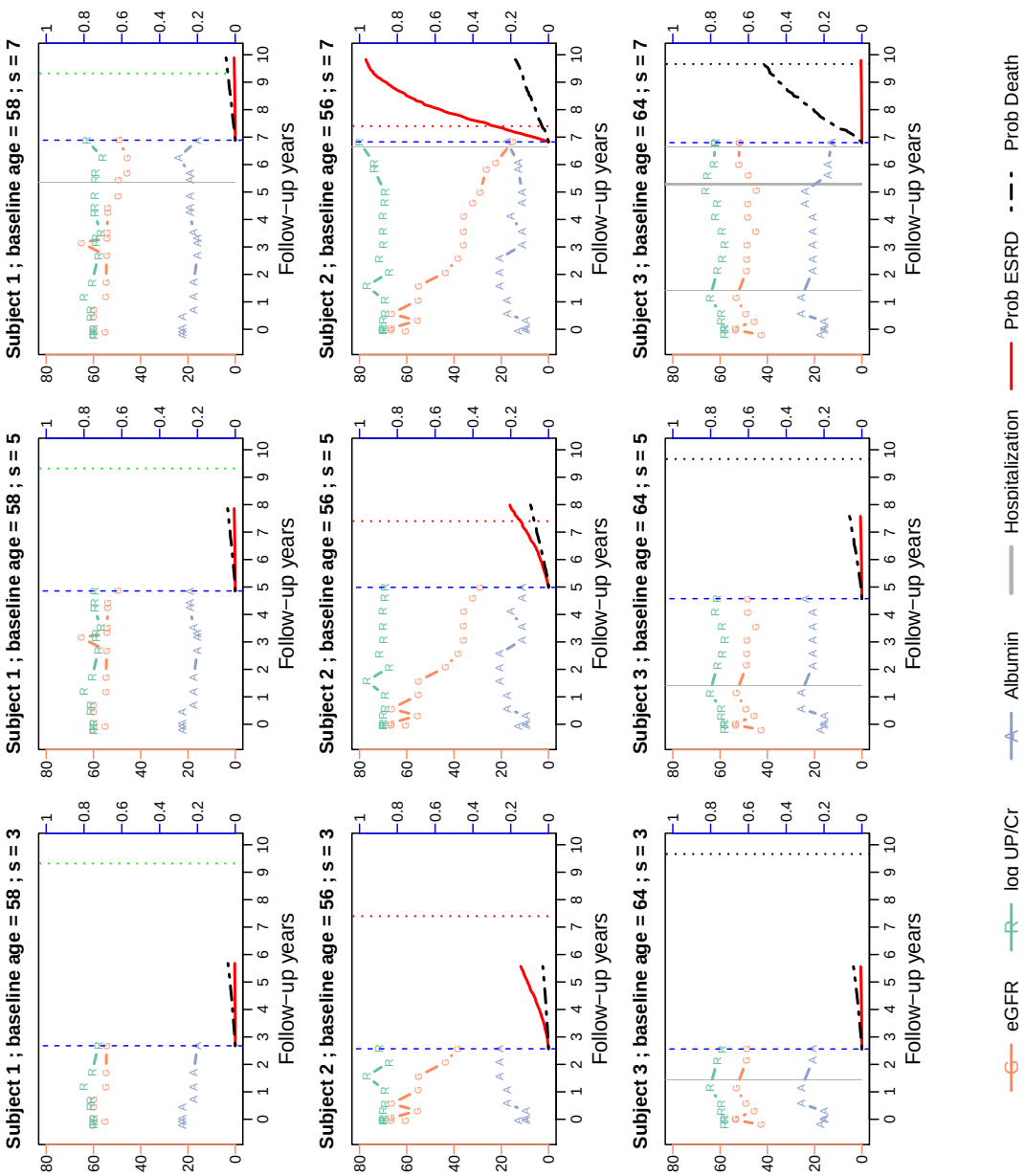

Figure 4: Individual dynamic predicted CIF for the three selected subjects in Figure 3. Each row in the panel represents one subject, the three columns are the predictions made at landmark years $s = 3, 5, 7$ with their most current biomarker data up to the blue dashed vertical lines. The prediction CIFs in $\tau_1 = 3$ years are plotted for the event of ESRD (red curve) and death (black curve). Symbols in the figure are set similarly to Figure 3.

Table 2: Measures of predictive accuracy summarizing predictions from the landmark SDH model for ESRD and death. The estimates were obtained at three landmark years, $s = 3, 5, 7$, with prediction horizon $\tau_1 = 1, 3$ years. AUC: area under the ROC curve that discriminates the subjects with ESRD from those who are event-free. $TP(c)$: true positive rate at threshold c; $FP(c)$: false positive rate at threshold c ; thresholds c are selected to be 0.1 and 0.05 for ESRD; and 0.05 and 0.01 for death. BS: Brier score, the mean squared error for the predicted risk probability.

		AUC		$TP(c)$		$FP(c)$		BS	
		ESRD	Death	ESRD	Death	ESRD	Death	ESRD	Death
$\tau_1 = 1$	$s = 3$	0.957	0.545	0.925	0.568	0.075	0.539	0.024	0.191
	$s = 5$	0.925	0.547	0.885	0.585	0.100	0.596	0.026	0.043
	$s = 7$	0.965	0.584	0.832	0.692	0.099	0.675	0.021	0.035
$\tau_1 = 3$	$s = 3$	0.944	0.558	0.876	0.468	0.119	0.378	0.054	0.119
	$s = 5$	0.943	0.520	0.852	0.498	0.140	0.549	0.052	0.093
	$s = 7$	0.957	0.492	0.863	0.343	0.131	0.405	0.048	0.110