Published in final edited form as:

Stat Med. 2016 June 15; 35(13): 2167–2182. doi:10.1002/sim.6860.

A Two-Stage Approach for Dynamic Prediction of Time-to-Event Distributions

Xuelin Huang^{a,*}, Fangrong Yan^{a,d}, Jing Ning^a, Ziding Feng^a, Sangbum Choi^b, and Jorge Cortes^c

^a Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77230, USA.

^b Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

^c Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX 77230, USA.

^d Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing 210009, P.R. China.

Abstract

Dynamic prediction uses longitudinal biomarkers for real-time prediction of an individual patient's prognosis. This is critical for patients with an incurable disease such as cancer. Biomarker trajectories are usually not linear, nor even monotone, and vary greatly across individuals. Therefore, it is difficult to fit them with parametric models. With this consideration, we propose an approach for dynamic prediction that does not need to model the biomarker trajectories. Instead, as a trade-off, we assume that the biomarker effects on the risk of disease recurrence are smooth functions over time. This approach turns out to be computationally easier. Simulation studies show that the proposed approach achieves stable estimation of biomarker effects over time, has good predictive performance, and is robust against model misspecification. It is a good compromise between two major approaches, namely, (1) joint modeling of longitudinal and survival data, and (2) landmark analysis. The proposed method is applied to patients with chronic myeloid leukemia. At any time following their treatment with tyrosine kinase inhibitors, longitudinally measured *BCR-ABL* gene expression levels are used to predict the risk of disease progression.

Keywords

Biomarker; Dynamic prediction; Landmark analysis; Longitudinal data; Survival analysis; Time-dependent covariate

*Correspondence to: Xuelin Huang, Department of Biostatistics - Unit 1411, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX, 77230-1402. xluang@mdanderson.org.

1. Introduction

Medical advancements have allowed many types of cancer to be managed as a chronic disease. However, achieving a complete cancer cure remains much more challenging. Cancer recurrence is a common problem and patients are scheduled for many follow-up visits once their primary treatment concludes. It is important to use the longitudinal biomarker data collected during such visits to make real-time predictions for the patient's disease prognosis so that new treatments can be initiated early to prevent disease progression. Thus, the prediction is conducted at each patient visit, and the measure of interest is the time to the next failure event. Such a real-time prediction system is called dynamic prediction and is critical for physicians and patients to be able to monitor disease progression and plan for early prevention and treatment, thus improving the patient's likelihood of survival.

This article is motivated by the need to predict the distribution of time to relapse for patients with chronic myeloid leukemia (CML) [1]. The disease burden of these patients can be represented by the expression level of the gene *BCR-ABL*. Tyrosine kinase inhibitors (TKIs) achieve good responses in these patients, as defined by both symptoms and the *BCR-ABL* expression level. Such patients take the TKIs continuously and are scheduled for regular follow-up visits. However, CML may recur during the years following treatment. The common clinical practice is to wait until a patient shows symptoms of disease relapse before starting a new treatment. However, for many patients, the *BCR-ABL* expression levels increase before clinical symptoms of disease relapse appear. Thus, it would be helpful to use *BCR-ABL* levels to predict the time to relapse so that patients and physicians can initiate new treatments early to minimize disease relapse.

The literature on traditional survival analysis provides many models for estimating the time to an event of interest. Most such models can be used for the purpose of prediction. However, many of them incorporate only baseline covariates [2, 3], so they can be used to predict survival at the baseline, but are not useful for dynamic prediction at any time beyond that.

As a tool for dynamic prediction, it is natural to consider using the Cox [4] proportional hazards model, with longitudinal biomarker values as time-dependent covariates. We denote by T_i the time to an event of interest, such as disease progression, for subject i , with $i = 1, \dots, n$. Hereafter, we simply call T_i the survival time for convenience. We denote the baseline covariates as Y_i and time-dependent covariates at time t as $Z_i(t)$. The Cox model specifies $h_i(t)$, the hazard function for T_i , as follows.

$$h_i(t|Y_i, Z_i(t)) = h_0(t) \exp\{\theta' Y_i + \gamma' Z_i(t)\}, \quad (1)$$

where $h_0(t)$ is an arbitrary non-negative function, and θ and γ are unknown parameters. At time t , conditional on $T_i \geq t$, the future survival distribution for this patient can be computed as follows.

$$\Pr[T_i \geq t+v | T_i \geq t, Y_i, \{Z_i(t+u), 0 \leq u \leq v\}] = \exp\left[-\int_0^v h_0(t+u) \exp\{\theta' Y_i + \gamma' Z_i(t+u)\} du\right]. \quad (2)$$

Using the above model, the prediction conducted at time t needs to use the future biomarker values $\{Z_A(t+u), 0 \leq u \leq v\}$. However, these values of $Z_A(t+u)$ are not yet available at time t . In order to use the above equation to make such predictions, a commonly adopted approach is to model the longitudinal biomarker trajectories, and use the model-based biomarker values in the above Cox model (1). This approach of jointly modeling longitudinal and time-to-event data has been adopted by many researchers [5, 6, 7, 8, 9, 10, 11, 12, 13]. Nice reviews of this topic are available in the literature [14, 15].

In real applications, longitudinal biomarker trajectories can be any shape, and can vary greatly from patient to patient. This can be seen from Figure 1, which shows the trajectories of the *BCR-ABL* expression levels over time for three patients with CML in the study we described. It is difficult to find satisfactory parametric models to fit longitudinal biomarker data well. In this article, we propose a method that does not specify a model for the changing patterns of the longitudinal biomarkers. It is an information-cumulating model for continuous predictive analysis over time. It performs dynamic predictions through a two-stage procedure. In the first stage, we provide a prediction based on only baseline covariate information and ignore post-baseline time-dependent covariates. In the second stage, we use information from post-baseline time-dependent covariates to improve the prediction results obtained from the first stage. Through this process, our approach is computationally easier than the joint modeling approach, and we avoid any need to impute unobserved biomarker values.

Another approach to dynamic prediction is landmark analysis, which fits a marginal survival model at fixed prediction times known as landmarks [17, 18, 19, 20]. However, in practice, we need to make predictions any time a new biomarker measurement is obtained, not only at a few selected time points. As shown in Figure 2, the time points for patients' follow-up visits and longitudinal biomarker measurements are scattered irregularly throughout the time interval following the initial treatment. Consequently, we need dynamic prediction models that are applicable at any post-treatment time point. Our proposed approach can perform such dynamic predictions continuously over time. Our approach exploits the following inherent property of hazard functions, which seems to have been neglected in this research area. That is, suppose the hazard function for T is $\lambda_T(u)$. Here and below, when needed, we use a subscript for each hazard function to indicate the time variable with which the hazard function corresponds. At time t , given $T \geq t$, denote the hazard function for $T - t$ by $\lambda_{T-t}(u)$, then we have

$$\lambda_{T-t}(u) = \lambda_T(t+u). \quad (3)$$

With this equality, the hazard functions used by landmark analysis at different time points, say $\lambda_{T-3}(u)$ for prediction at $t = 3$ given $T \geq 3$, $\lambda_{T-6}(u)$ and $\lambda_{T-9}(u)$ (defined similarly), can be expressed as different shifts from a single function, namely, $\lambda_T(3+u)$, $\lambda_T(6+u)$ and $\lambda_T(9+u)$. This helps to greatly reduce the number of variables. Moreover, this equality enables us to specify the conditional hazard function of T given $T \geq t$, at any time t . That is to say, making a prediction is no longer limited to a few selected time points. This equality makes it possible to conduct predictions continuously over time.

Another challenge when conducting landmark analysis is that it requires biomarker measurements for all the patients at each landmark time. This requirement is almost impossible to satisfy in practical applications. Figure 2 illustrates this difficulty. Using interpolation or some other imputation techniques to estimate the biomarker values may result in bias. Our two-stage method described in the next Section avoids this difficulty. It enables us to simply use every biomarker measurement at the time it is measured, and does not require any two or more patients to have their biomarkers measured at the same time post-baseline. This makes our approach easy to use and free from biomarker imputation errors.

Our proposed approach is described in Section 2, and is compared with joint modeling and landmark analysis through simulations in Section 3. It is then applied to a CML study in Section 4. Related topics about our approach are discussed in Section 4. Mathematical details are provided in the Appendix.

2. Method

The proposed model for dynamic prediction is implemented through a two-stage procedure. In the first stage, we provide a prediction based on the survival information and baseline covariate information of all the subjects, but ignore the information from their post-baseline time-dependent covariates. In the second stage, we use information from all of the post-baseline time-dependent covariates to improve the prediction results obtained from the first stage. Our two-stage information-cumulating approach proceeds as follows.

2.1. Stage I: Prediction using only survival and baseline covariate information

In the first stage, we suppose that the hazard function $\lambda_i(t)$ for patient i with baseline covariate Y_i satisfies the following proportional hazards model,

$$\lambda_i(t|Y_i) = \lambda_0(t) \exp\{\alpha' Y_i\}, \quad (4)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and α is an unknown parameter. Here, the hazard function $\lambda_i(t|Y_i)$ is conditional on baseline covariates Y_i only. It does not depend on time-dependent covariates $Z_i(t)$. Thus, it is different from the hazard function $h_i(t|Y_i, Z_i(t))$ introduced in the previous section, which is conditional on both Y_i and $Z_i(t)$. Let

$S_0(t) = \exp\left\{-\int_0^t \lambda_0(s) ds\right\}$ and $S_i(t|Y_i) = \Pr(T_i > t|Y_i)$. Given that $T_i > t$, the prediction of whether $T_i > t+u$ can be obtained through the following equation,

$$\Pr(T_i > t+u|T_i > t, Y_i) = S_i(t+u|T_i \geq t, Y_i) = \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp(\alpha' Y_i)}. \quad (5)$$

This prediction can be conducted at any post-baseline time point t .

Denoting the censoring time for patient i by C_i , $X_i = \min(T_i, C_i)$ and $I_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. We use lower case letters (x_i , y_i , δ_i , etc.) to denote the realized values of the corresponding random variables. For convenience, we assume that the time

variables $x_i, i = 1, \dots, n$, have been sorted in ascending order without ties. We then use the following two-stage estimation procedure.

In the first stage, we estimate α and $S_0(t), t > 0$, in equations (4) and (5). The estimator $\hat{\alpha}$ is obtained by maximizing the following partial likelihood,

$$PL(\alpha) = \prod_{i=1}^n \left\{ \frac{\exp(\alpha' y_i)}{\sum_{j: x_j \geq x_i} \exp(\alpha' y_j)} \right\}^{\delta_i}. \quad (6)$$

Then the Breslow estimator is used to estimate $S_0(t), t > 0$:

$$\begin{aligned} \hat{\lambda}_0(x_i) &= \frac{\delta_i}{\sum_{j: x_j \geq x_i} \exp(\hat{\alpha}' y_j)}, i=1, \dots, n; \\ \hat{S}_0(t) &= \exp \left\{ - \sum_{i: x_i \leq t} \hat{\lambda}_0(x_i) \right\}. \end{aligned} \quad (7)$$

2.2. Stage II: Using longitudinal data to improve predictions from stage I

In the second stage, at time t and with time-dependent covariates $Z(t)$, we should be able to provide a more accurate prediction than that obtained by equation (5) in stage I. For this purpose, given the property of hazard functions in equation (3), we have

$$\lambda_{T_i-t}(u|T_i \geq t, Y_i) = \lambda_{T_i}(t+u|Y_i). \quad (8)$$

We use the hazard function property in (3) to avoid specifying a hazard function $\lambda(u)$ for $T_i - t|T_i \geq t$ at each time point t . Instead, the values of $\lambda(u), t \geq 0$ are expressed as $\lambda(t+u)$, a shift of the baseline hazard function. This makes our proposed prediction model parsimonious. According to equation (4),

$$\lambda_{T_i}(t+u|Y_i) = \lambda_0(t+u) \exp\{\alpha' Y_i\}. \quad (9)$$

Then (8) and (9) together imply $\lambda_{T_i-t}(u|T_i \geq t, Y_i) = \lambda_0(t+u) \exp\{\alpha' Y_i\}$. Based on this equation, with additional information in $Z(t)$, we postulate that

$$\lambda_{T_i-t}(u|T_i \geq t, Y_i, Z_i(t)) = \lambda_0(t+u) \exp\{\alpha' Y_i\} \exp\{\beta'(t) Z_i(t)\}. \quad (10)$$

In the above model, we regard the parameters $\lambda_0(t+u)$ and α as being the same as in equations (9) and (4), so they can be estimated by equations (6) and (7). The only unknown parameter in the above model (10) is $\beta(t)$, which is a smooth vector function that describes the time-varying effects of covariates $Z(t)$ on future survival, with $\beta'(t)$ denoting its transpose. With the above model (10), the prediction at time t can be provided as

$$Pr(T_i \geq t+u|T_i \geq t, Y_i, Z_i(t)) = \left\{ \frac{S_0(t+u)}{S_0(t)} \right\}^{\exp\{\alpha' Y_i + \beta'(t) Z_i(t)\}}, \quad (11)$$

for which the estimation of α and $S_0(\cdot)$ is provided by equation (6) and (7), and the estimation of $\beta(t)$ is provided hereafter.

If we have biomarker measurements taken at or prior to baseline, we include such information in Y_i for $i = 1, \dots, n$. That is to say, for the same biomarker, such as the expression level of *BCR-ABL*, its baseline measurement is included in Y_i (together with other demographic information that does not change over time). Only its post-baseline information is included in $Z_i(t)$, $t > 0$. We do not use $Z_i(0)$ in this article. This makes it convenient for $Z_i(t)$ to include some commonly used covariates such as biomarker changes from previous measurements or change rates. That is to say, $Z_i(t)$ is not limited to being biomarker values at time t . It can be summary statistics (features, patterns) of all the biomarker information up to time t . Using different variables Y and $Z(t)$, $t > 0$ for baseline and post-baseline covariates respectively (i.e., including “ $Z(0)$ ” in Y) avoids the difficulty that arises when, for example, the changing slopes of biomarker expression levels are not defined at baseline $t = 0$.

In the second stage, we estimate $\beta(t)$. Various smoothness constraints, such as splines and fractional polynomials, can be placed on $\beta(t)$. Binder *et al.* [21] used simulations to compare these two approaches, and concluded that fractional polynomials better recover simpler functions, whereas splines better recover more complex functions. As illustration, for a scalar time-dependent covariate $Z(t)$, denoting $v = t + 1$, we assume

$$\begin{aligned} \beta'(t) Z(t) &= \beta_0(t) + \beta_1(t) Z(t) \\ &= \beta_{00} + \beta_{01} \ln(v) \\ &\quad + \beta_{02} \sqrt{v} + \beta_{03} \frac{1}{\sqrt{v}} + \beta_{04} v + \beta_{05} \frac{1}{v} + \beta_{06} v^2 + \beta_{07} \frac{1}{v^2} + Z(t) \left(\beta_{10} + \beta_{11} \ln(v) + \beta_{12} \sqrt{v} + \beta_{13} \frac{1}{\sqrt{v}} + \beta_{14} v + \beta_{15} \frac{1}{v} + \beta_{16} v^2 + \beta_{17} \frac{1}{v^2} \right) \end{aligned} \quad (12)$$

where $v = t + 1$ is used instead of the original t to avoid $\log(0)$ and $1/0$. The higher order polynomials in equation (12) may be removed by backward variable elimination.

Note that $Z(t)$ can represent values other than the biomarker values measured at time t , and may be a summary statistic of all the biomarker information observed up to (and including) time t . $Z(t)$ may include slope variables to describe the changes or rate of change in the biomarker values. It may be another variable that indicates a category of changing patterns of biomarker expression. The specification of $Z(t)$ requires a careful exploration of the data.

We include an intercept variable $\beta_0(t)$ in the expression of $\beta'(t)Z(t)$ in equation (12). If needed, we may also include interactions between baseline covariates Y and $Z(t)$. Thus, we may allow $\beta'(t)Z(t)$ to be

$$\beta_0(t) + \beta_1'(t) Y + \beta_2'(t) Z(t) + \beta_3'(t) Y Z(t).$$

This kind of flexible parameterization means that the proportional hazards assumption that seems to be implied by our model (11) does not need to hold for the model to have a good

approximation of reality. For example, if $\beta_1(t) = \eta \ln(t)$, then the hazard ratio of $Y=1$ versus $Y=0$ is $t^\eta e^{\alpha}$, instead of being a constant e^{α} over time.

Suppose patient i has biomarker measurements taken at time points $t_{ij}, j=1, \dots, n_i$ with $0 < t_{i1} < \dots < t_{i,n_i} < x_i$. At each time point t_{ij} by our model (11), the likelihood function connecting the biomarker value $z_i(t_{ij})$ and the final survival outcome (x_i, δ_i) is

$$L_{ij} \{ \beta(t) \} = L \{ X_i = x_i, \Delta_i = \delta_i | T_i > t_{ij}, y_i, z_i(t_{ij}), \alpha, \lambda_0(\cdot) \} \\ = \left[\lambda_0(x_i) \exp \{ \alpha' y_i + \beta'(t_{ij}) z_i(t_{ij}) \} \right]^{\delta_i} \cdot \exp \left[- \sum_{k: t_{ij} < x_k \leq x_i} \lambda_0(x_k) \exp \{ \alpha' y_i + \beta'(t_{ij}) z_i(t_{ij}) \} \right]. \quad (13)$$

We denote $l_{ij}(\beta(t)) = \log \{ L_{ij}(\beta(t)) \}$. Then, using a working independence assumption among the different time points $t_{ij}, j=1, \dots, n_i$, we have the following “working” log-likelihood function:

$$l \{ \beta(t) \} = \sum_{i=1}^n l_i \{ \beta(t) \} = \sum_{i=1}^n \sum_{j=1}^{n_i} l_{ij} \{ \beta(t) \} \\ = \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_i \left[\log \{ \lambda_0(x_i) \} + \alpha' y_i + \beta'(t_{ij}) z_i(t_{ij}) \right] - \exp \{ \alpha' y_i + \beta'(t_{ij}) z_i(t_{ij}) \} \sum_{k: t_{ij} < x_k \leq x_i} \lambda_0(x_k). \quad (14)$$

Then we plug in the stage 1 estimators $\alpha = \hat{\alpha}$ and $\lambda_0(\cdot) = \hat{\lambda}_0(\cdot)$, maximize the above log-likelihood function by setting its derivative with respect to β to zero, and solve the resulting estimating equations to obtain an estimator $\hat{\beta}(t)$.

The variance formulae for $\hat{\alpha}$ and $\hat{S}_0(\cdot)$ can be obtained in the same way as in the Cox model [22]. The variance for the coefficients in $\hat{\beta}(t)$ is estimated by a two-stage estimator [23], which is provided in the Appendix.

Given an existing data set, to apply the above two-stage approach for dynamic prediction, we first obtain estimators $\hat{\alpha}, \hat{S}_0(t)$ using only baseline covariates and survival information. We then estimate $\beta(t)$ for $t=0$, as described above. For a new patient (not in the data set) who has baseline covariate Y_{new} and covariate value $Z_{new}(t)$, we predict the new patient's survival time T_{new} as

$$Pr(T_{new} \geq t+u | T_{new} > t, Y_{new}, Z_{new}(t)) = \left[\frac{\hat{S}_0(t+u)}{\hat{S}_0(t)} \right]^{\exp \{ \hat{\alpha}' Y_{new} + \hat{\beta}'(t) Z_{new}(t) \}}. \quad (15)$$

Assuming a parametric form for $\beta(t)$ and using a two-stage approach, we achieve the following advantages. First, there is no need to specify a biomarker trajectory model, which is usually difficult. This also avoids the biased prediction that may result from a misspecified trajectory model. Second, the proposed approach uses only the observed biomarker values; therefore, there is no need to require all patients to have biomarker measurements taken at any common post-baseline time point. This avoids the need to impute biomarker values, even when their distributions are scattered throughout irregular time intervals. Third, the proposed method can conduct predictive analyses continuously over time (i.e.,

predictions can be conducted at any time $t \geq 0$). These advantages and the convenience of our method come with a price: we need to assume a parametric shape for the time-varying biomarker effect coefficient $\beta(t)$. We believe this is a good trade-off. It is safer to assume a parametric shape for $\beta(t)$ than for $Z_i(t)$, $i = 1, \dots, n$. Biomarker values can fluctuate over time for each patient, and their changing patterns vary greatly among different patients. This can be seen in Figure 1. On the other hand, it is reasonable to believe that $\beta(t)$ has a smooth shape.

3. Simulation Studies

Two major approaches to dynamic prediction are currently available in the literature: (1) joint modeling of longitudinal and survival data, and (2) landmark analysis. We briefly describe these approaches, compare them with our proposed method, and explain their advantages and disadvantages.

3.1. Joint modeling approach

Most approaches for jointly modeling time-to-event and longitudinal data are based on the Cox proportional hazards model with time-dependent covariates, as shown in equation (1). The full likelihood of such a joint model includes the following partial likelihood equation,

$$L_{td}(\beta) = \prod_{i=1}^n \left[\frac{\exp\{\theta' y_i + \gamma' z_i(x_i)\}}{\sum_{j: x_j \geq x_i} \exp\{\theta' y_j + \gamma' z_j(x_i)\}} \right]^{\delta_i}. \quad (16)$$

From this likelihood function, we can see that at the time of each event x_i , we need to know $Z_j(x_i)$ for all the patients j with $x_j \geq x_i$ (i.e., patients who are still at risk at this time x_i). In most studies, biomarker values are measured only at discrete time points. It is almost impossible to have biomarker values available for all patients in a study at all the random event times. Thus, the above condition is usually not satisfied in practical applications.

Some approaches will try to solve this problem by filling in the biomarker values according to ad hoc methods such as the last observation carried forward, the next observation carried backward, using the closest observation, or using the mean of the previous and next observations. However, such methods can give biased results [8]. The joint modeling approach solves this problem by assuming parametric models for the longitudinal biomarker values. For example,

$$Z_i(t) = f_i(t) + \varepsilon_i(t), \quad i=1, \dots, n, \quad (17)$$

where $f_i(t)$, $i = 1, \dots, n$ are usually related to a common parametric function $f(t)$ through some individual level random effects. Then, in the partial likelihood equation (16), instead of using $Z_j(x_i)$ for the patients j with $x_j \geq x_i$, the joint modeling approach maximizes the following partial likelihood that involves $f_j(x_i)$,

$$L_{jm}(\beta) = \prod_{i=1}^n \left[\frac{\exp\{\theta' y_i + \gamma' f_i(x_i)\}}{\sum_{j: x_j \geq x_i} \exp\{\theta' y_j + \gamma' f_j(x_i)\}} \right]^{\delta_i} \quad (18)$$

Using this joint modeling approach, even though $Z_j(x_i)$ is not observed, $f_j(x_i)$ can still be estimated from model (17), and the likelihood in (18) can still be computed. An advantage of the joint modeling approach is that the model-based value $f_j(t)$ is not as prone to measurement errors as $Z_j(t)$. However, this advantage is critically dependent on the correct specification of model (17) for the longitudinal biomarker trajectories, which is not an easy task.

3.2. Landmark analysis

In landmark analysis [24, 25], a fixed time point after the initiation of therapy is selected as a landmark for conducting the analysis of survival. Only patients who are alive and have not experienced the event of interest at the landmark time are included in the analysis. Many researchers have used a series of landmark analyses over time as an approach to dynamic prediction [17, 19], such as that shown here:

$$h_i(t|Z_i(s)) = h_s(t) \exp\left\{\theta_{LM}(s) Y_i + \beta'_{LM}(s) Z_i(s)\right\}, t > s. \quad (19)$$

In equation (19), for any two distinct time point $s \neq s'$, the hazard functions $h_s(t)$ and $h_{s'}(t)$ are different hazard functions without any constraints between them. Similarly, $\theta_{LM}(s)$ and $\theta_{LM}(s')$ have no constraints between them, nor do $\beta'_{LM}(s)$ and $\beta'_{LM}(s')$. Smoothness constraints between each pair of them can be imposed by using cubic splines or other techniques [17].

An advantage of landmark analysis is that it does not specify a model such as model (17) for the biomarker trajectories. However, separate and unconstrained landmark analyses on discrete time points ignore the correlation between neighboring time intervals. Thus, such analyses are not efficient and may result in large variations in the predictions. When biomarker measurements are irregularly scattered throughout the follow-up period, one must perform ad hoc grouping of biomarker measurements into different time intervals and conduct biomarker imputation so that the analyses can be performed on some common time points. This imputation process may introduce bias [8].

3.3. Simulation set-up

We performed a series of simulations to evaluate the proposed method for dynamic prediction and compared it with the approaches of landmark analysis and jointly modeling of time-to-event and longitudinal data. We assumed that a study involved 200 patients who had been followed for a period of 15 years, and for whom longitudinal biomarker measurements had been taken at baseline and every year thereafter. We used a linear mixed effects model to generate the longitudinal biomarker measurements. To generate survival times, we applied a Cox proportional hazards model, using the longitudinal biomarkers as time-dependent covariates. We evaluated the robustness of the dynamic prediction

approaches by considering different linear mixed effects models for the longitudinal data. Specifically, we generated data by the following Models I and II, which we call Scenarios I and II, respectively. Then, we analyzed each data set by each of the four models listed below to make predictions. Model III is a landmark analysis model. Model IV is our proposed approach. Models III and IV do not specify a data generating mechanism.

$$\begin{aligned}
 \text{Model I: } h_i(t|m_i(t)) &= h_0(t) \exp\{\gamma' m_i(t)\}, \\
 m_i(t) &= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t, \\
 Z_i(t) &= m_i(t) + \varepsilon_i(t). \\
 \text{Model II: } h_i(t|m_i(t)) &= h_0(t) \exp\{\gamma' m_i(t)\}, \\
 m_i(t) &= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t^3, \\
 Z_i(t) &= m_i(t) + \varepsilon_i(t). \\
 \text{Model III: } h_i(t|Z_i(s)) &= h_s(t) \exp\{\beta'_{LM}(s) Z_i(s)\}, t > s. \\
 \text{Model IV: } h_i(t|Z_i(0)) &= h_0(t) \exp\{\alpha' Z_i(0)\}, \\
 h_i(t|Z_i(s)) &= h_0(t) \exp\{\alpha' Z_i(0) + \beta'(s) Z_i(s)\}, t \geq s > 0.
 \end{aligned}$$

When we used Models I and II to generate data, we let $h_0(t) = at^{\sigma-1}$, a Weibull hazard function. However, when we used these models for data analysis and then prediction, we assumed that their $h_0(t)$ functions were nonparametric. We let $\beta_0 = 3$, $\beta_1 = 2$, $\gamma = 0.8$ or 1.6 . Here, γ controls how the survival time is influenced by the longitudinal biomarker values; a larger value of γ results in higher correlation. The random effects (b_{i0}, b_{i1}) have a

multivariate normal distribution with mean 0 and covariance matrix $\begin{pmatrix} 4 & 0.1 \\ 0.1 & 2 \end{pmatrix}$. We measured the biomarkers at $t = 1, 2, 3, \dots, 10$. At each time point t , the distribution of the measurement error term $\varepsilon_i(t)$ is $N(0, 0.6^2)$. During data generation, the shape parameter for the Weibull function $\sigma = 0.5$. We adjusted the scale parameter a to achieve a 50% censoring rate for each scenario using a uniform $(0, 28)$ distribution for the censoring time. In Model III, each s represents a landmark time point, and the baseline hazard function at time s is $h_s(t)$, $t > s$, which is an arbitrary non-negative function. For different landmark time points s_1, s_2 , there is no constraint between $h_{s_1}(t)$ and $h_{s_2}(t)$, nor between the regression coefficients $\beta_{LM}(s_1)$ and $\beta_{LM}(s_2)$. In Model IV, $\beta(t)$ is a fractional polynomial, and $\beta'(t)Z_i(t)$ has the formulation described in equation (12).

3.4. Measures to assess predictive performance

To evaluate the predictive performance of the above methods, we focused on calibration and discrimination. We assessed calibration by calculating the root mean squared prediction errors (RMSEs) between the true survival probabilities and the predictions obtained from the four models. We used $S_i(t) = \exp\left\{-\int_0^t h_i(s) ds\right\}$ with $t = 10$ to calculate the true survival probabilities, in which $h_i(s)$ were known in the simulations.

To measure the discriminative capability of a longitudinal biomarker, we focused on a time interval of clinical relevance, $(t, t + \delta)$, within which the occurrence of events is of interest and physicians can take action to improve the patient's chance of survival. In this setting, a

useful property of the model is the ability to successfully discriminate between patients who will and those who will not experience the event within this time period. For a randomly chosen pair of patients $\{i, j\}$ both of whom have provided measurements up to time t , the discriminative capability of the assumed model can be assessed by the area under the receiver operating characteristic curve (AUC) as

$$AUC(t, \Delta t) = Pr[\pi_i(t + \Delta t | t) \leq \pi_j(t + \Delta t | t) | \{T_i \in (t, t + \Delta t]\} \cap \{T_j > t + \Delta t\}],$$

where $\pi_i(t + \Delta t | t)$ represents the survival probability for patient i at $t + \Delta t$. That is, if patient i experiences the event within the relevant timeframe, and patient j does not, then we would expect to assign a higher probability of surviving longer than $t + \Delta t$ to patient j . We follow the work of Rizopoulos *et al.* [16], Harrell, Lee, and Mark [27], Zheng and Heagerty [28], and Antolini, Boracchi, and Biganzoli [29] to calculate this discrimination measurement.

The mean values and standard deviations of the RMSEs (Table 1) and AUCs (Table 2) from 1000 simulated data sets are presented for Scenarios I and II, under two values of γ (which controls the effects of the biomarkers on survival). Our proposed Model IV performed almost the same as the true model in each scenario, and outperformed all the other models. In Scenario I, the RMSEs (and AUCs) obtained from our method are only slightly larger (smaller) than those obtained from Model I (the true model in this Scenario), and are much better than those obtained from Models II and III. In Scenario II, the RMSEs (and AUCs) obtained from our method are only slightly larger (smaller) than those obtained from Model II (the true model in this Scenario), and are much better than those obtained from Models I and III. When the joint modeling approach used the correct biomarker trajectory model for prediction (Model I in Scenario I and Model II in Scenario II), it performed the best, as expected, with the smallest RMSE and largest AUC. However, when a misspecified trajectory model was used (Model I in Scenario II, or Model II in Scenario I), the joint modeling approach for prediction was clearly outperformed by our proposed model. In all the scenarios, the landmark analysis approach performed worse than the other approaches on average, and its AUCs showed large variations.

4. Application to a CML study

We used our proposed method to predict the distribution of time to disease progression for patients with CML [1]. The data were obtained from a study of 670 patients with CML who had failed their front-line therapy with imatinib, and then enrolled in a trial to receive dasatinib. Only 567 patients had *BCR-ABL* measurements taken both before and after the dasatinib treatment, and thus were included in our data analysis by the proposed prediction model. By variable selection, we decided to use two baseline variables and one time-dependent covariate as follows: (1) Age60, a binary indicator of whether at baseline the patient is 60 years of age or older; (2) Bcrbase, the baseline *BCR-ABL* level; and (3) Bcrcurr(t), the *BCR-ABL* level at time $t > 0$. We normalized all the *BCR-ABL* levels to be between 0 and 100, and measured time in months.

4.1. Stage 1: Prediction using only baseline covariate information

First, we fit a Cox proportional hazards model as shown in equation (4), using the two baseline variables and the time to disease progression, as follows,

$$\lambda_i(t|Age60_i, Bcrbase_i) = \lambda_0(t) \exp(\alpha_1 Age60_i + \alpha_2 Bcrbase_i). \quad (20)$$

We obtained estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{S}_0(\cdot)$ as a function of $\hat{\lambda}_0(\cdot)$, shown in (7). Then, without using the post-treatment biomarker observations, we used the following formula to predict the time to disease progression:

$$\hat{Pr}(T_i \geq t+u|T_i \geq t, Age60_i, Bcrbase_i) = \left\{ \frac{\hat{S}_0(t+u)}{\hat{S}_0(t)} \right\}^{\exp\{\hat{\alpha}_1 Age60_i + \hat{\alpha}_2 Bcrbase_i\}} \quad (21)$$

Suppose a patient with CML who was younger than 60 years had $BCR-ABL = 80$ at baseline. At 10 months post-treatment, the disease has not progressed for this individual. At that time, without using information about the post-treatment $BCR-ABL$ measures, the future risk of disease progression for this patient is predicted by the solid line in Figure 3. This prediction can be viewed as an average for all the patients at that time, $t = 10$ months. The dashed and dotted lines in the Figure are not confidence intervals. They are generated by using different values of $BCR-ABL$ at 10 months, as described hereafter.

4.2. Stage II: Using longitudinal data to improve predictions from stage I

After treatment, patients were scheduled to have follow-up visits at 3, 6, 9, 12, 18 and 24 months and annually thereafter. Their $BCR-ABL$ expression levels were measured at these visits. Some patients did not adhere to this follow-up schedule and some missed scheduled visits. Most patients had irregular intervals between visits. Figure 2 shows all the available $BCR-ABL$ measurements.

With the longitudinal measurements $Bcrcurr_t(t)$ at time points t_{ik} , $k = 1, \dots, n_i$, we assume that

$$\begin{aligned} Pr(T_i \geq t_{ik}+u|T_i \geq t_{ik}, Age60_i, Bcrbase_i, Bcrcurr_t(t_{ik})) \\ = \left\{ \frac{S_0(t_{ik}+u)}{S_0(t_{ik})} \right\}^{\exp\{\alpha_1 Age60_i + \alpha_2 Bcrbase_i + \beta_0(t_{ik}) + \beta_1(t_{ik})Bcrcurr_t(t_{ik})\}} \end{aligned} \quad (22)$$

with $\beta(t) = (\beta_0(t), \beta_1(t))'$ specified in (12). At any time t , the effect of the current $BCR-ABL$ value, $Bcrcurr_t(t)$, on progression-free survival is a power term $\exp\{\beta_0(t) + \beta_1(t)Bcrcurr_t(t)\}$ on the conditional survival function. Fitting this model by maximizing the working likelihood in (14), the resulting estimators for $\beta_0(t)$ and $\beta_1(t)$ are plotted in Figure 4 (a) and (b), respectively. Those plots show that the value of $\beta_1(t)$ is small initially but increases over time and then reaches a plateau. Recall that $\beta_1(t)$ describes the effects of the $BCR-ABL$ expression level at time t on the future risk of disease progression. The small value of $\beta_1(t)$ in the first few months after treatment indicates that a high expression level of $BCR-ABL$ in the first few months does not have much impact on progression-free survival. Six months after treatment, the large value of $\beta_1(t)$ indicates that a high expression level of $BCR-ABL$ at

that time is a sign of a short progression-free survival time. This difference suggests that it takes a few months for patients to respond to the TKI treatment and for their *BCR-ABL* expression levels to be reduced. However, if a patient has high *BCR-ABL* expression levels six months after treatment, that patient is not responding well to the treatment and has a high risk of disease progression in the near future.

We also apply the landmark analysis approach to analyze the CML data set. The resulting effects of *BCR-ABL* expression levels on the risk of disease progression are shown in Figure 4 (c). It can be seen that $\beta_1(t)$ by our proposed approach, shown in Figure 4 (b), is a smoothed version of that by landmark analyses. It is hard to believe that the real effects of *BCR-ABL* on disease progression can be so bumpy over time, as shown by the landmark analysis. One may try to get estimates for a series of discrete landmark time points first, and then smooth them by splines or other techniques. However, this kind of ad hoc method is arbitrary and inefficient. By putting our estimation in a systematic framework, we borrow information between neighbor time points on biomarker effects on disease progression and thus are more efficient. This is reflected in the smaller standard deviations of the estimated AUC's by our method (Table 2).

With all the unknown parameters in the above equation (22) having been estimated, we now illustrate how to conduct the dynamic prediction by our proposed two-stage approach. We already provided stage-1 prediction in the previous Subsection for a patient who was younger than 60 years and had *BCR-ABL*=80 at baseline. That prediction used only baseline information. Then at 10 months post baseline, the patient's disease has not progressed and the current *BCR-ABL* measure is *Bcrcurr*(10). The prediction of this patient's future progression-free survival is as follows.

$$\hat{\Pr}(T \geq 10+u | T \geq 10, \text{Age60}, \text{Basebcr}, \text{Bcrcurr}(10)) = \left\{ \frac{\hat{S}_0(10+u)}{\hat{S}_0(10)} \right\}^{\exp\{\hat{\alpha}_1 \text{Age60} + \hat{\alpha}_2 \text{Basebcr} + \hat{\beta}_0(10) + \hat{\beta}_1(10) \text{Bcrcurr}(10)\}} \quad (23)$$

For two cases, *Bcrcurr*(10) = 1 and *Bcrcurr*(10) = 30, the predicted progression survival curves are plotted in Figure 3 (dashed and dotted lines, respectively). Figure 3 (solid line) also shows the prediction conducted at 10 months, computed by (21), using only baseline information, without using time-varying *BCR-ABL* measurements. This figure shows that the longitudinal biomarker information helps to further distinguish patients within each subgroup as defined by baseline covariates, and thus provides more accurate prediction.

4.3. Generating nomograms for medical decision making

Suppose, at any time t beyond the baseline measure, a physician would like to initiate new treatments for patients whose risk of disease progression in the next 12 months is 20% or greater. Then, based on the model obtained above, we can determine the cut-off values for the *BCR-ABL* expression levels over time that correspond to such a risk level. That is to say, after plugging all the estimated values $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_0(t)$, $\hat{\beta}_1(t)$, and $\hat{S}_0(t)$ into equation (22), letting $u = 12$ months, we would like to determine *Bcrcurr*(t) such that

$$\begin{aligned}
 &1 - \hat{\Pr}(T \geq t+12 | T \geq t, \text{Age60}, \text{Basebcr}, \text{Bcrrt}(t)) \\
 &= 1 - \left\{ \frac{\hat{S}_0(t+12)}{\hat{S}_0(t)} \right\}^{\exp\{\hat{\alpha}_1 \text{Age60} + \hat{\alpha}_2 \text{Basebcr} + \hat{\beta}_0(t) + \hat{\beta}_1(t) \text{Bcrrt}(t)\}} \geq 0.2. \quad (24)
 \end{aligned}$$

For a patient who is 60 years of age or younger, and $\text{Basebcr} = 80$, we solve the above equation to obtain a solution for $\text{Bcrrt}(t)$ for each $t > 0$ (we actually only need to work on those event times). We repeated this procedure, replacing the 20% with different risk levels of 25% and 30%. The solutions for $\text{Bcrrt}(t)$ depend on $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}_0(t)$, $\hat{\beta}_1(t)$, and $\hat{S}_0(t)$. Although $\hat{\beta}_0(t)$ and $\hat{\beta}_1(t)$ are smooth, $\hat{S}_0(t)$ is not. Consequently, the solutions for $\text{Bcrrt}(t)$ are not smooth. We then applied smoothing techniques and plotted the results (see Figure 5). This nomogram shows that at any time post-baseline, the expression levels of *BCR-ABL* correspond to, respectively, 20% (dotted line), 25% (dashed line), and 30% (solid line) risk of disease progression in the next 12 months. For example, at the 20th month after the beginning of treatment, if the above patient's *BCR-ABL* level is 20 (or 35, or 45), then the patient's risk of disease progression in the next 12 months is about 20% (or 25%, or 30%, respectively). Similarly, we can use this nomogram to perform risk prediction at any other time such as the 30th or 40th month. Patients can use this nomogram to evaluate whether their situation is improving or becoming worse over time. This is similar to the way pediatricians use growth curve nomograms to evaluate the growth of babies and children. Our nomogram is as convenient to use as growth curves.

4.4. Predictions by different approaches

We conducted predictions for patients with CML using the joint modeling approach, landmark analysis, and the proposed two-stage approach. To evaluate the performance of these three approaches and verify their validity, we computed Kaplan-Meier estimators (Figure 6, d) for the four subgroups of patients with *BCR-ABL* at 10 months falling within the ranges (0, 0.1], (0.1, 1.0], (1.0, 10] and (10, 100] respectively, and compared them with the corresponding results obtained from the aforementioned three methods. These three methods use continuous *BCR-ABL* values to fit a model for each subgroup. For comparison, predictive results for the median *BCR-ABL* value within each of the four subgroups are respectively displayed in Figure 6 (a-c). If we consider the distances between the four survival curves obtained from each of the approaches, the joint modeling approach shrinks the curves the most, while the landmark analysis shrinks the curves the least. Our proposed approach gives results comparable with the empirical Kaplan-Meier curves. This shows that our approach achieves a desirable balance between joint modeling and landmark analysis.

5. Discussion

It is important for individuals who have had cancer to be monitored for signs of disease recurrence, even after they have been successfully treated. Monitoring patients may involve measuring specific biomarkers, such as prostate-specific antigen for prostate cancer, CA-125 for ovarian cancer, and *BCR-ABL* expression for CML.

We have provided an information-cumulating model for continuously conducting predictive analyses over time. Our model comprises a two-stage approach that uses longitudinal biomarker data to conduct dynamic real-time prediction of the time to a specific event. The first stage provides predictions that are conditional on only the baseline covariates. These predictions describe the distributions of the time to the event of interest in the absence of post-treatment biomarker information. In the second stage, the post-treatment biomarker information is used to improve the prediction from the first stage. An inherent property of hazard functions is used to provide parsimonious prediction models over time. Our proposed method has a few advantages. The first is that it does not need to specify a model for the longitudinal biomarker, which avoids the bias caused by the mis-specification of such a model. A second advantage is that it does not need to have the complete history of the biomarker, and does not need to perform biomarker value imputation. Without using random effect models for the longitudinal biomarkers, the computation required by the proposed method is much easier than that of other dynamic prediction methods, such as jointly modeling time-to-event and longitudinal data.

It is important to optimize the decision-making tool we have proposed by using receiver operating characteristic (ROC) curves [26]. It would be desirable to find some change points that can achieve the best ROC properties. There are a large number of publications on the optimization of ROC curves. Our next step will be to apply the optimization of ROC curves to our dynamic prediction method to achieve its best performance.

Acknowledgements

This research was supported by the US National Institutes of Health, grants U54 CA096300, U01 CA152958 and 5P50 CA100632.

A. Appendix: Variance Estimation

Maximizing the proposed working likelihood in equation (14) by replacing α and $S_0(\cdot)$ with their consistent estimators gives a consistent estimator of $\beta(t)$ [23]. We next derive the asymptotic distribution of $\hat{\beta}(t)$. For simplicity, we assume a simple linear form $\beta(t) = \beta \times t$; however, the arguments can be easily generalized to a general parametric function.

To study the asymptotic variance of $\hat{\beta}$, we explore the first derivative of the working log-likelihood,

$$U(\beta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left[\beta \delta_i t_{ij} z_i(t_{ij}) + t_{ij} z_i(t_{ij}) \exp \left\{ \hat{\alpha}' y_i + \beta t_{ij} z_i(t_{ij}) \right\} \left(\log \hat{S}_0(x_i) - \log \hat{S}_0(t_{ij}) \right) \right]. \quad (25)$$

Denote $\tilde{U}(\beta)$ as the equation derived from $U(\beta)$ by replacing $\hat{\alpha}$ and $\hat{S}(\cdot)$ with their true values. Then the score equation $U(\beta)$ evaluated at the true value β_0 of β can be rewritten as

$$U(\beta_0) = \tilde{U}(\beta_0) + U(\beta_0) - \tilde{U}(\beta_0) = \tilde{U}(\beta_0) + I + II, \quad (26)$$

where

$$\begin{aligned}
 I &= \sum_{i=1}^n \sum_{j=1}^{n_i} t_{ij} z_i(t_{ij}) \log \left\{ \hat{S}_0(x_i) / \hat{S}_0(t_{ij}) \right\} \cdot \left[\exp \left\{ \hat{\alpha}' y_i + \beta t_{ij} z_i(t_{ij}) \right\} - \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} \right] \\
 &= \sum_{i=1}^n y_i \sum_{j=1}^{n_i} \left[t_{ij} z_i(t_{ij}) \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} \log \left\{ \frac{\hat{S}_0(x_i)}{\hat{S}_0(t_{ij})} \right\} \right] (\hat{\alpha} - \alpha) + o_p(\sqrt{n})
 \end{aligned} \quad (27)$$

$$II = \sum_{i=1}^n \sum_{j=1}^{n_i} t_{ij} z_i(t_{ij}) \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} \cdot \left[\log \left\{ \hat{S}_0(x_i) / \hat{S}_0(t_{ij}) \right\} - \log \left\{ S_0(x_i) / S_0(t_{ij}) \right\} \right]. \quad (28)$$

Given the asymptotic martingale expressions of $\hat{\alpha}$ and $\log \hat{S}_0(\cdot)$,

$$\sqrt{n} \{ \hat{\alpha} - \alpha \} = \frac{1}{\sqrt{n}} \Gamma_{\alpha} \sum_{i=1}^n \int_0^{\infty} I(u \leq x_i) \left(y_i - \bar{y}(u) \right) dM_i(u; \alpha)$$

and

$$\begin{aligned}
 &\sqrt{n} \left[\log \left\{ \hat{S}_0(x_i) / \hat{S}_0(t_{ij}) \right\} - \log \left\{ S_0(x_i) / S_0(t_{ij}) \right\} \right] \\
 &= \frac{-1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} I(t_{ij} < u \leq x_i) \cdot \left[1 / \sum_{k=1}^n \exp(\alpha' y_k) I(x_k \geq u) + \left(y_i - \bar{y}(u) \right) \right] dM_i(u; \alpha),
 \end{aligned}$$

we have

$$\frac{1}{\sqrt{n}} I = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \left\{ \Gamma_{\alpha} D_1(u) \left(y_i - \bar{y}(u) \right) \right\} dM_i(u; \alpha) + o_p(1), \quad (29)$$

$$\frac{1}{\sqrt{n}} II = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \left\{ D_2(u) + D_3(u) \left(y_i - \bar{y}(u) \right) \right\} dM_i(u; \alpha) + o_p(1), \quad (30)$$

where Γ_{α} is the expectation of the Jacobian matrix of the partial likelihood, and

$$\begin{aligned}
 \hat{y}(u) &= \frac{\sum_{i=1}^n I(x_i \geq u) y_i \exp(\alpha' y_i)}{\sum_{i=1}^n I(x_i \geq u) \exp(\alpha' y_i)}, \\
 M_i(u; \alpha) &= I(x_i \leq u, \delta_i = 1) - \int_0^t I(x_i > s) \exp(\alpha' y_i) \lambda_0(s) ds, \\
 D_1(u) &= E \left(y_i \sum_{j=1}^{n_i} \left[t_{ij} z_i(t_{ij}) \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} \log S_0(t_{ij}) \right] \right), \\
 D_2(u) &= E \left(- \sum_{j=1}^{n_i} \frac{t_{ij} z_i(t_{ij}) \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} I(t_{ij} < u \leq x_i)}{\sum_{k=1}^n \exp(\alpha' y_k) I(x_k \geq u)} \right), \\
 D_3(u) &= E \left(- \sum_{j=1}^{n_i} \left[t_{ij} z_i(t_{ij}) \exp \left\{ \alpha' y_i + \beta t_{ij} z_i(t_{ij}) \right\} I(t_{ij} < u \leq x_i) \right] \right).
 \end{aligned}$$

By equations (26), (27), (28), (29) and (30) and the martingale central limit theorem, $\frac{U(\beta_0)}{\sqrt{n}}$ converges to a normal distribution with zero mean and variance matrix Σ_{β} which can be

estimated by the sum of the products of the terms from $\tilde{U}(\beta)$, I and II . Then by the Taylor series expansion,

$$\frac{1}{\sqrt{n}}U(\hat{\beta}) = n^{-1/2}U(\hat{\beta}_0) + \frac{1}{n}\Gamma_n(\beta_0)\sqrt{n}(\hat{\beta} - \beta_0) + o_p(1),$$

where $\sqrt{n}(\hat{\beta} - \beta_0)$ converges weakly to a normal distribution with mean zero and variance-covariance matrix $\Gamma_\beta^{-1}\Sigma_\beta\Gamma_\beta^{-1}$. Here, Γ_β is the expectation of the Jacobian matrices

$$\frac{1}{n}\Gamma_n(\beta_0) = \frac{1}{n}\partial U(\beta)/\partial\beta|_{\beta=\beta_0}.$$

References

1. Quintas-Cardama A, Choi S, Kantarjian H, Jabbour E, Huang X, Cortes J. Predicting outcomes in patients with chronic myeloid leukemia at any time during tyrosine kinase inhibitor therapy. *Clinical Lymphoma, Myeloma & Leukemia*. 2014; 14(4):327–334. doi: 10.1016/j.clml.2014.01.003.
2. Zheng Y, Cai T, Feng Z. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*. 2006; 62(1):279–287. [PubMed: 16542256]
3. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102(478):527–537.
4. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodology)*. 1972; 34(2):187–220.
5. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53(1):330–339. [PubMed: 9147598]
6. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1(4):465–480. [PubMed: 12933568]
7. Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*. 2001; 96(455):895–905.
8. Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*. 2001; 88(2):447–458.
9. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*. 2002; 58:742–753. [PubMed: 12495128]
10. Huang X, Liu L. A joint frailty model for survival and gap times between recurrent events. *Biometrics*. 2007; 63(2):389–397. [PubMed: 17688491]
11. Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*. 2008; 64(3):950–958. [PubMed: 18162110]
12. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2009; 58(1):65–81.
13. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011; 67(3):819–829. doi:10.1111/j.1541-0420.2010.01546.x. [PubMed: 21306352]
14. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*. 2004; 14:809–834.
15. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-event Data: With Applications in R*. Chapman and Hall/CRC; Boca Raton: 2012.
16. Rizopoulos D, Hatfield L, Carlin P, Takkenberg J. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association*. 2014; 109(508):1385–1397.

17. Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. 2005; 61(2):379–391. [PubMed: 16011684]
18. van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*. 2007; 34:7–85.
19. van Houwelingen HC, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis*. 2008; 14:447–463. [PubMed: 18836831]
20. van Houwelingen, H.; Putter, H. *Dynamic Prediction in Clinical Survival Analysis*. CRC Press; Boca Raton: 2012.
21. Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. *Statistics in Medicine*. 2013; 32(13):2262–2277. [PubMed: 23034770]
22. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982; 10(4):1110–1120.
23. Shih JH, Louis TA. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*. 1995; 51:1384–1399. [PubMed: 8589230]
24. Anderson J, Cain K, Gelber R. Analysis of survival by tumor response. *Journal of Clinical Oncology*. 1983; 1:710–719. [PubMed: 6668489]
25. Dafni U. Landmark analysis at the 25-year landmark point. *Circulation: Cardiovascular Quality and Outcomes*. 2011; 4(3):363–371. doi: 10.1161/CIRCOUTCOMES.110.957951. [PubMed: 21586725]
26. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005; 61(1):92–105. [PubMed: 15737082]
27. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*. 1996; 15:361–387. [PubMed: 8668867]
28. Zheng Y, Heagerty PJ. Prospective accuracy for longitudinal markers. *Biometrics*. 2007; 63(2): 332–341. [PubMed: 17688486]
29. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Statistics in Medicine*. 2005; 24(24):3927–3944. [PubMed: 16320281]

Figure 1.
BCR-ABL expression values over time for three selected patients in the CML data set. The *BCR-ABL* values are standardized to lie within the range (0,100).

Figure 2.
BCR-ABL expression values over time for all patients with CML in the data set.

Figure 3.

Predictions of future progression-free survival at 10 months for patients with baseline age 60 years, baseline *BCR-ABL*=80: (1) Patient # 1: *BCR-ABL* expression level at 10 months is 1 (dashed line), (2) Without using *BCR-ABL* information at 10 months (solid line), and (3) Patient # 2: *BCR-ABL* expression level at 10 months is 30 (dotted line).

Figure 4.

Estimation of the time-varying effects of *BCR-ABL* expression levels on the risk of disease progression: (a) $\beta_0(t)$ by proposed method, (b) $\beta_1(t)$ by proposed method, and (c) $\beta_{LM}(t)$ by landmark analysis, which is the counterpart of $\beta_1(t)$ in the proposed method. Landmark analysis does not use $\beta_0(t)$.

Nomogram for Risk Prediction

Figure 5.

For patients with a baseline age = 60 years, baseline $BCR-ABL=80$, at any time t post-baseline, the expression level of $BCR-ABL$ at t corresponding to the risk of disease progression within $(t, t + 12)$ months being 20% (dotted line), 25% (dashed line), or 30% (solid line).

Figure 6.

Prediction of future progression-free probability for patients who were at-risk at 10 months, by four different methods: (a) joint modelling with a linear mixed model for biomarker measurements over time, (b) landmark analysis, (c) proposed two-stage approach, and (d) Kaplan-Meier estimators; with each method applied to each of the four subgroups of patients classified by their expression levels of *BCR-ABL* at 10 months: solid line for range (0, 0.1], dashed line for range (0.1, 1], dotted line for range (1, 10] and dot-dashed line for range (10, 100].

Table 1

Mean value (and standard deviation) of the root mean squared prediction error (RMSE) in the simulations. Models I and II are joint models, Model III is landmark analysis. Model IV is the proposed method.

	Data generated by Model I		Data generated by Model II	
	$\gamma = 0.8$	$\gamma = 1.6$	$\gamma = 0.8$	$\gamma = 1.6$
Prediction by Model I	0.346 (0.021)	0.601 (0.021)	0.429 (0.030)	0.645 (0.009)
Prediction by Model II	0.408 (0.019)	0.622 (0.017)	0.387 (0.023)	0.616 (0.012)
Prediction by Model III	0.564 (0.020)	0.659 (0.009)	0.545 (0.024)	0.649 (0.008)
Prediction by Model IV	0.344 (0.035)	0.608 (0.021)	0.390 (0.031)	0.623 (0.017)

Table 2

Mean value (and standard deviation) of the area under the receiver operating characteristic curve (AUC) in the simulations. Models I and II are joint models, Model III is landmark analysis. Model IV is the proposed method.

	Data generated by Model I		Data generated by Model II	
	$\gamma = 0.8$	$\gamma = 1.6$	$\gamma = 0.8$	$\gamma = 1.6$
Prediction by Model I	0.622 (0.045)	0.667 (0.126)	0.564 (0.046)	0.538 (0.109)
Prediction by Model II	0.573 (0.049)	0.545 (0.110)	0.598 (0.062)	0.653 (0.120)
Prediction by Model III	0.518 (0.260)	0.434 (0.353)	0.504 (0.270)	0.466 (0.355)
Prediction by Model IV	0.605 (0.047)	0.655 (0.134)	0.596 (0.043)	0.608 (0.106)