Published in final edited form as:

Biometrics. 2017 March ; 73(1): 83–93. doi:10.1111/biom.12562.

On Longitudinal Prediction with Time-to-Event Outcome: Comparison of Modeling Options

Marlena Maziarz^a, Patrick Heagerty^a, Tianxi Cai^b, and Yingye Zheng^{c,†}

^aDepartment of Biostatistics, University of Washington, Seattle, Washington 98195

^bDepartment of Biostatistics, Harvard School of Public Health, Boston, MA 0211

^cFred Hutchinson Cancer Research Center, 1100 Fairview Ave N., M2-B500, Seattle, Washington 98109

Summary

Long term follow-up is common in many medical investigations where the interest lies in predicting patients' risks for a future adverse outcome using repeatedly measured predictors over time. A key quantity is the likelihood of developing an adverse outcome among individuals who survived up to time s given their covariate information up to time s . Simple, yet reliable, methodology for updating the predicted risk of disease progression using longitudinal markers remains elusive. Two main approaches have been considered in the literature. One approach, based on joint modeling (JM) of failure time and longitudinal covariate process (Tsiatis and Davidian, 2004), derives such longitudinal predictive probability from the joint probability of a longitudinal marker and an event at a given time. A second approach, the partly conditional (PC) modeling (Zheng and Heagerty, 2005), directly models the predictive probability conditional on survival up to a landmark time and information accrued by that time. In this paper we propose new PC models for longitudinal prediction that are more flexible than joint modeling and improve the prediction accuracy over existing PC models. We provide procedures for making inference regarding future risk for an individual with longitudinal measures up to a given time. In addition, we conduct simulations to evaluate both JM and PC approaches in order to provide practical guidance on modeling choices. We use standard measures of predictive accuracy adapted to our setting to explore the predictiveness of the two approaches. We illustrate the performance of the two approaches on a dataset from the End Stage Renal Disease Study (ESRDS).

Keywords

joint model; longitudinal data analysis; partly conditional model; risk prediction; survival analysis

[†]Corresponding author: yzheng@fhcrc.org.

8. Supplementary Materials

Web Appendices, Tables, and Figures referenced in Sections 4, 5 and 6 and R codes are available with this paper at the *Biometrics* website on Wiley Online Library.

1. Introduction

Long-term followup is common in many medical investigations where the interest lies in predicting patients' risks for a future adverse outcome. Serially measured biomarkers provide information both on the marker level and its changes over time, and thus are often more informative regarding the occurrence of an outcome of interest at a future time, compared with prediction based only on information collected at baseline. A key quantity of risk prediction in longitudinal settings is the probability of developing an adverse outcome in the next τ_0 time interval given survival up to time s and covariate information available up to time s : $R(\tau_0|s, \mathbf{H}(s)) = P(T \leq s + \tau_0 | T > s, \mathbf{H}(s))$, where T is the time to the outcome of interest and $\mathbf{H}(s)$ contains both time-constant and longitudinal covariate history up to time s . Such information can be used to identify subjects at high risk who can be targeted for preventive strategy or treatment plan. Similarly, less frequent follow-up may be recommended to low-risk individuals. The primary inferential objective is to identify predictive algorithms for risk of an adverse outcome at a future time based on relevant longitudinal processes up to the time of prediction, and to assess the utility of such an algorithm for use in aiding medical decisions regarding disease monitoring plans and treatment choices.

The use of repeated biomarkers presents some unique challenges in outcome prediction analysis. The effects of biomarkers often vary with time. Variation due to measurement error in biomarkers often leads to attenuated estimates of the relative risk parameters in a Cox model that are biased towards zero (Prentice, 1982), and this bias affects predictions of future risk. In survivorship analysis, serial markers are typically incorporated in a time-varying covariate Cox regression model (Crowley and Hu, 1977). Since the marker is usually measured only at discrete time points, time-specific measurements for the marker may not be available for members in the risk set when a failure occurs. To address such difficulties, joint modeling (JM) for both the marker process and survival data have been developed in recent years (Tsiatis et al., 1995; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997). Estimation of the parameters can be done through a two-stage approach (Pawitan, 1993; Bycott and Taylor, 1998) or a full likelihood-based approach (Henderson et al., 2000; Wang and Taylor, 2001). Both approaches involve specifying the longitudinal marker process via a random effect submodel or a latent class submodel, and then linking the failure time process via a time-varying covariate hazard submodel of the latent biomarker values without measurement error.

The main focus in the JM literature has been on inferring how biomarkers change over time and the estimation of the association between the marker and the disease. Adopting the JM approach for prediction of disease progression has recently been considered in the literature (Proust-Lima and Taylor, 2009; Rizopoulos, 2011; Taylor et al., 2013). However the calculation of updated risks, $R(\tau_0|s, \mathbf{H}(s))$, in this setting is not straightforward and involves fairly complicated model specification and a computational procedure that requires integration over the marker process. When multiple biomarkers are under consideration, the computational burden increases drastically and the requirement of correct specification of both submodels within the JM framework can be critical for estimation. Methodology that is

computationally simple yet sufficiently flexible to capture updated risks over time is appealing from a practical standpoint.

To develop a survival model that directly facilitates the estimation of $R(\tau_0|s, \mathbf{H}(s))$ for any specific pair of (s, τ_0) , Zheng and Heagerty (2005) proposed a class of semiparametric models, called “partly conditional survival models”(PC), where the conditional hazard of failure is specified by conditioning on only a part of the marker trajectory at select ‘landmark’ times. Specifically, a marker measured at time s , $Y(s)$, generates a data record with time origin for T reset to s , i.e., the time scale is now in terms of time since measurement. The marker measurement at time s can now be regarded as a baseline measure. As a result, there can be multiple derived event times for each individual corresponding to repeatedly measured marker values. Thus, the PC approach casts the problem of using longitudinal marker values for predicting survival within the general framework of multivariate survival models, and it avoids using a time-varying covariate Cox model for risk prediction. The PC approach does not require a modeling specification regarding the event time as a function of the entire marker process. Implementation of the partly conditional model is relatively simple using slightly modified standard statistical software (Zheng and Heagerty, 2005). Similar approaches based on the idea of a landmark time without setting starting times to s were considered (van Houwelingen, 2007; van Houwelingen and Putter, 2012). A thorough investigation of different approaches to prediction using longitudinal markers is needed in order to provide practical guidance to researchers in terms of the choice of estimation methods. To make a comparison among longitudinal prediction tools, it is important to gauge model performance based on their intended clinical utilities. The evaluation of the performance of a medical test has been based on Receiver Operating Characteristic (ROC) curves and calculation of time-dependent ROC curves to evaluate a longitudinal marker has been considered in Zheng and Heagerty (2004). In this manuscript we will compare the performance of updated prediction rules developed based on either a JM or various PC approaches, adopting metrics for assessing the performance of prediction algorithms to the setting with longitudinal markers and survival outcomes.

The goals of this manuscript are to provide flexible tools for calculating updated risk for patients under medical monitoring over time, and to investigate the performance of predictive algorithms derived from either the JM or the PC modeling approaches. In particular, we propose two simple, yet flexible, modeling methods within the PC modeling framework. The methods have the advantage of ease of implementation and are robust to violations in modeling assumptions. To provide practical guidance in the choice among the proposed methods, we conduct extensive numerical studies to compare prediction rules derived from the different approaches using an array of summary measures of predictive performance.

We illustrate the performance of our methods on the End Stage Renal Disease Study (ESRDS) conducted at the Community Health Network in San Francisco. The ESRDS is a retrospective cohort of 698 individuals aged 18–60 years with severe non-dialysis requiring chronic kidney disease (CKD). Each individual provided 1–24 measurements with more than half providing four or more marker measurements. The outcome of interest was time to

ESRD (initiation of dialysis or kidney transplantation) or death. Predicting the conditional probability of progression to ESRD or death within a clinically meaningful timeframe, τ_0 , given event-free survival and covariate information up to one or two years has important consequences for the individuals. A subject found to be at high risk of ESRD in a one-year timeframe would likely be recommended to meet with their healthcare team to start preparing for dialysis. A three-year timeframe would provide a subject at high risk with preventive treatment options, starting with controlling comorbid conditions such as hypertension and diabetes, lifestyle changes, among others.

We introduce notation and general models in Section 2. We describe estimation and inference procedures of proposed methods for the PC approach in Sections 3 and 4. Numerical studies of finite sample performance and predictive performance comparison are presented in Section 5. The illustrative example and discussion are presented in Sections 6 and 7.

2. Modeling Longitudinal Data with Failure Time Outcome

2.1 Notation

Let T_i and C_i be random variables with values measuring the time to failure and time to censoring, respectively, for subject i . We may not observe T_i for all subjects, but rather we observe $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. We use $\mathbf{Z}_i$ to denote time-constant covariates such as gender. The longitudinal biomarker for subject i is denoted by $\mathbf{Y}_i = \{Y_i(s_{i1}), \dots, Y_i(s_{im_i})\}$ measured at times $\mathbf{s}_i = \{s_{i1}, \dots, s_{im_i} : s_{im_i} < X_i\}$, $i = 1, \dots, n$, $j = 1, \dots, m_i$ where $Y_i(s_{ij})$ is the j^{th} measurement for subject i observed at time s_{ij} . The observation times are assumed to be specified by a study protocol, although deviations from protocol are not unusual. The observed covariate information for subject i consists of $\mathbf{H}_i = \{\mathbf{Z}_i, \mathbf{Y}_i, \mathbf{s}_i\}$. At time $u = 0$, the history of the covariate process for subject i is $\mathbf{H}_i(u) = \{\mathbf{Z}_i, \mathbf{Y}_i(u), \mathbf{s}_i(u)\}$, where $\mathbf{Y}_i(u) = \{Y_i(s_{ij}) : 0 \leq s_{ij} \leq u, j = 1, \dots, m_i, u < X_i\}$ and $\mathbf{s}_i(u) = \{s_{ij} : 0 \leq s_{ij} \leq u, j = 1, \dots, m_i, u < X_i\}$. The full data is denoted by $\mathcal{D}_n = \{\mathbf{X}_i, \mathbf{H}_i, i = 1, \dots, n\}$. A subscript $\cdot$ is used to denote a new, or future, individual. We review two approaches in the literature for modeling such data next.

2.2 Joint Model

Joint modeling framework for longitudinal and time-to-event data (Tsiatis et al., 1995; Wulfsohn and Tsiatis, 1997; Henderson et al., 2000) allows the full representation of the joint likelihood given the observed data, $\mathcal{D}_n$ (Tsiatis and Davidian, 2004). It comprises two linked sub-models, one for the underlying ‘true’ biomarker process as a function of time, and one linking the failure time T_i to the underlying ‘true’ biomarker process.

To model the biomarker process, a linear mixed effects model (LMEM) is often considered: $\mathbf{Y}_i = \mathbf{Y}_i^* + \mathbf{e}_i$, where $\mathbf{Y}_i^* = \mathbf{U}_i \boldsymbol{\beta} + \mathbf{W}_i \mathbf{b}_i$ is a vector of the underlying ‘true’, thus unobserved, biomarker values for subject i , $\boldsymbol{\beta}$ denotes a $(p \times 1)$ vector of fixed effects, $\mathbf{b}_i$ is a $(q \times 1)$ vector of the random effects assumed to be normally distributed with mean 0 and variance $\boldsymbol{\Sigma}$, $\mathbf{U}_i$ is a $(m_i \times p)$ matrix of covariates and $\mathbf{W}_i$ is a $(m_i \times q)$ matrix of covariate data for individual i . The measurement error $\mathbf{e}_i = \{e_i(s_{i1}), \dots, e_i(s_{im_i})\}$ and $e_i(s_{ij})$ is assumed to be

distributed $N(0, \sigma_e^2)$, $\text{cov}(e_{ij}, e_{i'j'}) = 0$ for $j \neq j'$, and is independent of $\mathbf{b}_i$. The failure time process is usually modeled using a time-varying covariate proportional hazards model (Cox, 1972) with hazard function

$$\lambda_i(t|Y_i^*(t), \mathbf{Z}_i) = \frac{1}{\Delta t} \lim_{\Delta t \rightarrow 0} P(t \leq T_i < t + \Delta t | T_i \geq t, Y_i^*(t), \mathbf{Z}_i), \text{ specified as}$$

$$\lambda_i(t) = \lambda_0(t) \exp(\eta Y_i^*(t) + \gamma' \mathbf{Z}_i) \text{ (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2004).}$$

Since both model specifications involve unknown quantities, fitting involves iteration between the linear mixed effects submodel and the survival submodel. For a future individual with $\mathbf{H}_o(s) = \{\mathbf{Z}_o, \mathbf{Y}_o(s), \mathbf{s}_o(s)\}$ and event free at time s , one may predict $R(\tau_0|s, \mathbf{H}_o(s))$ as $R^{\text{JM}}(\tau_0|s, \mathbf{H}_o(s)) = P(s < T_o \leq s + \tau_0 | T_o > s, \mathbf{H}_o(s) = \mathbf{H}_o(s), \mathcal{D}_n, \theta)$, with θ denoting the parameter vector of the joint model. Rizopoulos (2011) showed that

$$R^{\text{JM}}(\tau_0|s, \mathbf{H}_o(s)) = 1 - P(T_o > s + \tau_0 | T_o > s, \mathbf{H}_o(s), \mathcal{D}_n, \theta) = 1 - \int \frac{S(s + \tau_0 | \mathbf{Y}_o^*(s + \tau_0), \mathbf{Z}_o, \theta)}{S(s | \mathbf{Y}_o^*(s), \mathbf{Z}_o, \theta)} p(\mathbf{b} | T_o > s, \mathbf{H}_o(s), \theta) d\mathbf{b}$$

where $S(t) = \exp \left[- \int_0^t \lambda(u | Y_o^*(u), \mathbf{Z}_o, \theta) du \right]$ and $\mathbf{b}$ is a vector of random effects. Point and interval estimates can be obtained using Monte Carlo simulations (Proust-Lima and Taylor, 2009; Rizopoulos, 2011), thus prediction for an individual requires complex computation.

2.3 PC Models

In a partly conditional survival model framework, at time s , we are interested in predicting the residual lifetime $\mathcal{T}_s = T - s | \mathcal{T}_s > 0$ using observed covariate history up to s based on $\mathbf{H}(s)$. To directly model such updated risks, the analysis time of a PC model is such that the measurement time becomes the baseline time and estimation is based on the derived failure times calculated from each measurement time. Specifically, at s_{ij} , observed data vector for subject i is $\mathcal{O}_{ij} = \{X_{ij}, \mathbf{H}(s_{ij})\}$, where $X_{ij} = T_i - s_{ij}$. A PC model then specifies the relationship between $\mathcal{T}_{ij} = T_i - s_{ij}$ and $\mathbf{H}(s_{ij})$ without having to specify the full marker process. Such an approach is simple and robust, since we only need to specify a working model for the residual lifetime as a function of s_{ij} and $\mathbf{H}(s_{ij})$. Computationally it is also much easier to implement compared to a JM model.

In the following section we introduce two new estimating procedures for PC models that vary in efficiency and flexibility.

3. Estimating Approaches for PC Models

3.1 Partly Conditional Cox-type model (PC_{Cox})

A key consideration in PC models is the specification of a flexible working model that captures the time-varying relationship between the future outcome and a measurement at a given point in time. To make a prediction about an event within $\mathcal{T}_s - \tau_0$ for a given τ_0 , one approach considered in Zheng and Heagerty (2005) is to assume

$$\lambda(\tau | \mathbf{H}_i(s_{ij})) = \lambda_0(\tau) \exp(\alpha' \mathbf{B}(s_{ij}) + \gamma' \mathbf{Z}_i + \eta(\tau)' h(\mathbf{Y}_i(s_{ij}))) \quad (1)$$

where $\lambda_0(\cdot)$ is an unknown baseline hazard, $\mathbf{B}(s)$ is a spline basis function of s which captures the potentially non-linear effect of measurement time s_{ij} on the risk; $\eta(\tau)$ allows the

marker effect to vary over τ , the time since measurement; $h(\mathbf{Y}_i(s_{ij}))$ is a functional of $\mathbf{Y}_i(s_{ij})$, for example it can be simply the last observed marker value, $Y_i(s_{ij})$, or a predicted response profile approximating the true marker value at s_{ij} , $Y_i^*(s_{ij})$, based on LMEM estimates of fixed and random effects, their variance components and observed data up to time s_{ij} for individual i (see Section 3.3). A kernel-based estimating procedure (Cai and Sun, 2003) can be used to obtain estimators in model (1). To further reduce the computational burden of fitting the model, we consider instead a similar model, where $\eta(\tau)$ is parametrized as $\eta' \mathbf{B}(\tau)$, where $\mathbf{B}(\tau)$ is a spline basis of time since measurement τ .

$$\lambda(\tau | \mathbf{H}_i(s_{ij})) = \lambda_0(\tau) \exp(\boldsymbol{\alpha}' \mathbf{B}(s_{ij}) + \boldsymbol{\gamma}' \mathbf{Z}_i + \eta' \mathbf{B}(\tau) h(\mathbf{Y}_i(s_{ij}))) = \lambda_0(\tau) \exp(\boldsymbol{\theta}'_{\text{Cox}} \mathbf{H}_i^B(s_{ij}, \tau))$$

where $\boldsymbol{\theta}_{\text{Cox}} = [\boldsymbol{\alpha}', \boldsymbol{\gamma}', \eta']'_{p \times 1}$ and $\mathbf{H}_i^B(s_{ij}, \tau) = [\mathbf{B}(s_{ij})', \mathbf{Z}_i', \mathbf{B}(\tau) h(\mathbf{Y}_i(s_{ij}))']'$.

Estimators of the model parameters, $\boldsymbol{\theta}_{\text{Cox}}$, can be obtained by maximizing the following marginal log partial likelihood $\mathcal{L}(\boldsymbol{\theta}_{\text{Cox}})$:

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \int \left(\boldsymbol{\theta}'_{\text{Cox}} \mathbf{H}_i^B(s_{ij}, u) - \log \left(\sum_{k=1}^n \sum_{l=1}^{m_k} I(X_{kl} \geq u) \exp(\boldsymbol{\theta}'_{\text{Cox}} \mathbf{H}_k^B(s_{kl}, u)) \right) \right) dN_{ij}(u) \quad (2)$$

where $N_{ij}(u) = I(i = 1, X_{ij} \geq u)$. For a new subject with covariate history $\mathbf{H}_o(s)$ up to time s , our key predictive probability within time τ_0 from s can be estimated as

$$\hat{R}_{\text{Cox}}^{\text{PC}}(\tau_0 | s, \mathbf{H}_o(s), \hat{\boldsymbol{\theta}}_{\text{Cox}}) = 1 - \exp(-\hat{\Lambda}(\tau_0 | s, \mathbf{H}_o(s), \hat{\boldsymbol{\theta}}_{\text{Cox}})),$$

where

$$\hat{\Lambda}(\tau_0 | \mathbf{H}_o(s), \hat{\boldsymbol{\theta}}_{\text{Cox}}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^{\tau_0} \frac{\exp(\hat{\boldsymbol{\theta}}'_{\text{Cox}} \mathbf{H}_o^B(s_{ij}, u))}{\sum_{k=1}^n \sum_{l=1}^{m_k} I(X_{kl} \geq u) \exp(\hat{\boldsymbol{\theta}}'_{\text{Cox}} \mathbf{H}_k^B(s_{kl}, u))} dN_{ij}(u). \text{ In}$$

practice, one may apply administrative censoring at τ_0 for robustness of the model (van Houwelingen and Putter, 2012).

3.2 Partly Conditional Generalized Linear Model (PCGLM)

The above model based on a Cox-type partly conditional model can be further relaxed to allow effect of measurement time to vary with prediction time τ . To this end, we propose to approximate $R(\tau_0 | s, \mathbf{H}_o(s))$ through a working *marginal* partly conditional generalized linear model with a binary outcome (PCGLM),

$$P(\mathcal{T}_{ij} \leq \tau_0 | s_{ij}, \mathbf{H}_i(s_{ij})) = g(\boldsymbol{\alpha}' \mathbf{B}(s_{ij}) + \boldsymbol{\gamma}' \mathbf{Z}_i + \eta' h(\mathbf{Y}_i(s_{ij}))) = g(\boldsymbol{\theta}'_{\text{GLM}} \mathbf{H}_i^B(s_{ij})), \quad (3)$$

where $\mathbf{H}_i^B(s_{ij}) = [\mathbf{B}(s_{ij})', \mathbf{Z}_i', h(\mathbf{Y}_i(s_{ij}))']'$, $g(\cdot)$ is a prespecified link function such as anti-logit, and $\boldsymbol{\theta}_{\text{GLM}} = [\boldsymbol{\alpha}', \boldsymbol{\gamma}', \boldsymbol{\eta}']_{p \times 1}'$. We may also include possible interactions between $\mathbf{B}(s_{ij})$ and the other predictors to allow for time-varying effects of the predictors. In the presence of censoring, we propose to estimate $\boldsymbol{\theta}_{\text{GLM}}$ by solving a weighted estimating equation $\mathcal{U}(\boldsymbol{\theta}_{\text{GLM}})$:

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \hat{w}_{ij}^{(C)} \frac{\dot{g}(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij}))}{g(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij}))(1 - g(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij})))} (I(X_{ij} \leq \tau_0) - g(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij}))) \quad (4)$$

where $\dot{g}(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij})) = \frac{\partial g(\boldsymbol{\theta}_{\text{GLM}}' \mathbf{H}_i^B(s_{ij}))}{\partial \boldsymbol{\theta}_{\text{GLM}}}$, $\hat{w}_{ij}^{(C)} = \Delta_i \frac{I(T_i \leq s_{ij} + \tau_0)}{\hat{G}(T_i)} + \frac{I(X_i > s_{ij} + \tau_0)}{\hat{G}(s_{ij} + \tau_0)}$ accounts for censoring and $\hat{G}(\cdot)$ is the Kaplan-Meier estimator for $G(\cdot)$, the survival function of C . Such time-specific logistic model has been previously studied, for example, by Zheng et al. (2006) and Uno et al. (2007) but only with baseline covariates. An in-depth discussion of binomial regression models for time-to-event outcome can be found in Grøn and Gerds (2014).

Let $\hat{\boldsymbol{\theta}}_{\text{GLM}}$ denote the final estimator of $\boldsymbol{\theta}_{\text{GLM}}$. Then for a future subject with covariate history up time s , $\mathbf{H}_\circ(s)$, the probability of an event within τ_0 time from s , $P(\mathcal{T}_s \leq \tau_0 | \mathcal{T}_s > 0, \mathbf{H}_\circ(s))$ can be estimated as $\hat{R}_{\text{GLM}}^{\text{PC}}(\tau_0 | s, \mathbf{H}_\circ(s), \hat{\boldsymbol{\theta}}_{\text{GLM}}) = g(\hat{\boldsymbol{\theta}}_{\text{GLM}}' \mathbf{H}_\circ^B(s))$.

3.3 Novel Two-stage PC Models

Biomarkers are often, if not always, measured with error. When the error term is nontrivial, using observed values at the time of prediction, $Y_\lambda(s_{ij})$, often leads to attenuated effects and naturally it might impact the risk prediction as well. In the longitudinal setting, considering a functional of $Y_\lambda(s_{ij})$ that is less prone to error may result in better approximation of the true underlying risk in the working model. We therefore propose a two-stage procedure, where in the first stage we model the longitudinal process and calculate a fitted $\hat{Y}_\lambda(s_{ij})$ based on the *best linear unbiased predictor* (BLUP) estimator (Figure 1). Specifically, under the following model for $\mathbf{Y}_i$: $\mathbf{Y}_i = \mathbf{U}_i \boldsymbol{\beta} + \mathbf{W}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$, $i = 1, \dots, n$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}(\boldsymbol{\phi}))$, $\mathbf{Y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ where $\mathbf{U}_{i(m_i \times p)}$ and $\mathbf{W}_{i(m_i \times q)}$ are matrices of covariates, which are subsets of $(\mathbf{Z}_i, \mathbf{s}_i)$ for individual i , $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)'$ denotes the parameter vector of the variance components of the random effects and $\sum_i = \sigma^2 \mathbf{I} + \mathbf{W}_i \mathbf{D}(\boldsymbol{\phi}) \mathbf{W}_i'$. Therefore the BLUP estimator is

$$\hat{\mathbf{Y}}_i = \mathbf{U}_i \hat{\boldsymbol{\beta}} + \mathbf{W}_i \mathbf{D}(\hat{\boldsymbol{\phi}}) \mathbf{W}_i' (\sum_i)^{-1} (\mathbf{Y}_i - \mathbf{U}_i \hat{\boldsymbol{\beta}}). \quad (5)$$

In the second stage, we obtain BLUP-based estimators of risk model parameters $\hat{\boldsymbol{\theta}}_{\text{Cox}}^{\text{BLUP}}$ and $\hat{\boldsymbol{\theta}}_{\text{GLM}}^{\text{BLUP}}$ by replacing $h(\mathbf{Y}_\lambda(s_{ij}))$ with $h(\hat{\mathbf{Y}}_\lambda(s_{ij}))$ in Equation (2) or Equation (4), respectively.

For a new subject with $\mathbf{H}_\circ(s)$ observed up to time $s_{\hat{j}}$, the predicted random effect is

$$\hat{\mathbf{b}}_o|s_{oj} = \mathbf{D}(\hat{\phi})\mathbf{W}_o'(s_{oj})(\hat{\sigma}^2\mathbf{I} + \mathbf{W}_o(s_{oj})\mathbf{D}(\hat{\phi})\mathbf{W}_o'(s_{oj}))^{-1}(\mathbf{Y}_o(s_{oj}) - \mathbf{U}_o(s_{oj})\hat{\beta}), \quad (6)$$

where $\mathbf{Y}_o(s_{oj})$, $\mathbf{U}_o(s_{oj})$ and $\mathbf{W}_o(s_{oj})$ denote the covariate data up to time s_{oj} for the new subject 'o'. Then the fitted marker value based on covariate data only up to time s_{oj} is

$$\hat{Y}_o(s_{oj}) = \mathbf{U}_o(s_{oj})\hat{\beta} + \mathbf{W}_o(s_{oj})(\hat{\mathbf{b}}_o|s_{oj}). \quad (7)$$

Iterating through each marker measurement of each subject, one can obtain vectors of BLUP fitted marker values for each subject denoted by $\hat{\mathbf{Y}}_o$ throughout the manuscript. We emphasize that $\hat{\mathbf{Y}}_o$ is not the same as $\mathbf{Y}_o^*$ from Section 2.2 (see Figure 1).

Replacing $\mathbf{H}_o(s)$ with $\hat{\mathbf{H}}_o(s) = \{\mathbf{Z}_o, \hat{\mathbf{Y}}_o(s), \mathbf{s}_o(s)\}$ in $\hat{R}_{\text{Cox}}^{\text{PC}}(\tau_0|\mathbf{H}_o(s))$ or $\hat{R}_{\text{GLM}}^{\text{PC}}(\tau_0|\mathbf{H}_o(s))$ we can obtain BLUP-based predicted risk estimators $\hat{R}_{\text{Cox}}^{\text{BLUP}}(\tau_0|\hat{\mathbf{H}}_o(s))$ or $\hat{R}_{\text{GLM}}^{\text{BLUP}}(\tau_0|\hat{\mathbf{H}}_o(s))$ for the new subject. Note that such a two-stage approach is different from the two-stage estimation procedure in the classic JM literature, in that a PC model rather than a time-varying covariate survival model is considered here for prediction.

4. Inference

We note here that PC models are working models that can be subject to misspecification. Under mild regularity conditions, it can be shown that the estimates from such models converge to a constant vector as $n \rightarrow \infty$ (Hjort, 1992). To construct confidence intervals (CI) for proposed risk estimators, we suggest perturbation resampling-based methods. By decomposition of an estimator into a sum of asymptotically i.i.d. terms, computational burden can be alleviated by perturbing directly the influence function of the estimator. For BLUP estimators, our inference procedure accounts for the variations generated across all stages. Details are given in the Web Supplement Appendix A.

5. Simulations

Our simulations addressed two goals: to evaluate the finite sample performance of PC model-based estimation and inference procedures proposed in Section 3 and Web Supplement Appendix A; and to compare the prediction performance of the proposed PC models with that of the JM in terms of accuracy and discrimination.

5.1 Data Generation

Throughout we consider a cohort with size $n = 500$. The longitudinal marker data was generated following the general setup used in the JM literature (e.g., Tsiatis et al. (1995)) and similar to that described in Zheng and Heagerty (2005). See the Web Supplement Appendix B for details on data generation for all simulation studies.

5.2 Inference for Longitudinal Prediction

In the first set of simulations we investigated the point and interval estimates of predicted partly conditional risk (PCR) as derived by PC_{GLM} , PC_{Cox} and the two-stage PC model illustrated with PC_{GLM} BLUP. For PC_{GLM} and PC_{Cox} we compare the $R(\tau_0|s, \mathbf{H}(s))$ for an individual with $Y_s = 0$ or 1 at time s , for four sets of s and τ_0 , the time of prediction and length of the prediction window, respectively. We also varied the magnitude of the measurement error ($\sigma_e = \{0.1, 1.0\}$). For PC_{GLM} and PC_{Cox} the true values of $R(\tau_0|s, Y_s = 0)$ and $R(\tau_0|s, Y_s = 1)$ were obtained empirically, by simulating a large dataset ($n = 5,000,000$) without censoring and empirically estimating the proportion of survivors at $s + \tau_0$ given survival up to time s , for subjects with marker values close to 0, (0 ± 0.05) and 1, (1 ± 0.05) . As seen in Table 1 and Web Table 1, both PC_{GLM} and PC_{Cox} , even though they are ‘working’ models and only approximating the true model, produce PCRs that are quite close to the truth, and the perturbation based asymptotic standard error (ASE) estimators approximate the empirical standard errors (ESD) closely. PC_{GLM} tends to have smaller bias, however PC_{Cox} tends to have a smaller mean squared error (MSE). The performance of the two models is similar for $\sigma_e = 0.1$ (Web Table 1) and $\sigma_e = 1.0$ (Table 1).

We then evaluated the point and interval estimates of PCR based on a two-stage PC_{GLM} (BLUP) model. For PC_{GLM} BLUP we estimated $R(\tau_0|s)$ for two randomly selected subjects with marker trajectories up to $s = 12$ and $s = 24$ months, and $6 \leq \tau_0 \leq 36$ months. Based on the individual’s observed values up to $s = 12$ and 24, we obtained the BLUP marker values and made risk predictions for each individual based on those. To approximate true values, we estimated PCRs for the two subjects based on their trajectories up to s using JM methods that are specified according to the true underlying model and used a sample size of $n=10,000$. The PCRs are quite close to the truth and the perturbation-based point-wise CIs trace the empirical CIs estimated over 500 simulated datasets quite well. This adds confidence that the proposed variance estimation procedure can be used in practice for making inference regarding patient’s risk at the personal level (Figure 2).

5.3 Performance Comparison

In the second set of simulations we compare the predictive performance of five longitudinal models for estimating the conditional survival probability (PC_{Cox} , PC_{GLM} , PC_{Cox} BLUP, PC_{GLM} BLUP and a joint model), using measures of calibration and discrimination. We first briefly describe the metrics in the following. To measure how well a dynamic prediction rule is calibrated, we considered a version of the *prediction error* (PE), or Brier score (Brier, 1950) extended to the time-dependent situation: $PE = E(I(s < T_i - s + \tau_0) - R(\tau_0|s, \mathbf{H}(s)))^2$ (Schoop et al., 2008). Another type of key summary indices for characterizing the performance of a risk prediction model are the risk threshold specific true and false positive fractions (TPF and FPF), and are defined as $TPF(\tau_0|s, \psi) = P(R(\tau_0|s) \geq \psi | s < T_i - s + \tau_0)$ and $FPF(\tau_0|s, \psi) = P(R(\tau_0|s) \geq \psi | T_i > s + \tau_0)$. When estimated over the full range of risk thresholds, these can be displayed using an ROC curve: $ROC(\tau_0|s, \cdot) = \{TPF(\tau_0|s, \psi), FPF^{-1}(\tau_0|s, \psi), \psi \in (0, 1)\}$. We consider the area under such an ROC curve (AUC) as a global summary of discrimination in our performance comparison below. In the simulations we consider $PE(\tau_0|s)$, $TPF(\tau_0|s, \psi)$ and $FPF(\tau_0|s, \psi)$ for $\psi = 0.25$, $AUC(\tau_0|s)$. For the joint model fitting we used a function provided in the R package JM (Rizopoulos, 2010). In all

simulations considered, JMs were fit with specifications that corresponded to the true models. For each simulation, all five models were fit to a training dataset. A validation dataset was generated with the same parameters as the corresponding training dataset. PCRs were predicted for individuals in the validation set who were at risk at time s . We used the last available marker value (marker at time s) for prediction in the PC models, and used all the marker information up to time s in prediction using the two-stage PC_{Cox} BLUP model and the JM. Accuracy performance metrics were calculated nonparametrically conditioning on time s . Censoring was accounted for with the inverse censoring probability weighted approach similar to that considered in Section 3.2.

In the scenario where measurement error is small, $\sigma = 0.1$, the predictive performance of all models is comparable (Web Table 2). In the case of a large measurement error, $\sigma = 1.0$, JM using individual trajectories up to time s performed better in terms of both calibration and discrimination compared to the PC_{GLM} , PC_{Cox} , PC_{GLM} BLUP and PC_{Cox} BLUP models. It also yielded prediction rules with larger PCF and smaller PNF, compared with PC_{GLM} and PC_{Cox} models. However, we observed a substantial improvement in predictive performance of the PC_{GLM} BLUP and PC_{Cox} BLUP over PC_{GLM} and PC_{Cox} . The performance of the PC_{GLM} BLUP and PC_{Cox} BLUP models was often comparable to that of the JM, and was achieved with much simpler computational steps at model fitting and risk prediction stages (Table 2).

We conclude that PC models are adequate for use in practice. We recommend using PC_{GLM} BLUP or PC_{Cox} BLUP models when the biomarkers under consideration have large measurement errors and when the marker trajectory can be modeled well. When there is no strong variation observed over time, the simpler approach of PC_{GLM} or PC_{Cox} might be desirable. Computationally, PC models are much faster to fit and in our experience have much higher convergence rates compared to the JM when fit with the R package JM version 1.4–2 (Rizopoulos, 2010). See Web Tables 3 and 4 in Web Supplement Appendix C for a comparison of the computational burden of the five models. Also, see Web Supplement Appendix D for performance comparison of the models under misspecification.

6. Analysis of the ESRD Study (ESRDS)

The ESRDS dataset

The ESRDS is a retrospective cohort of 698 individuals aged 18–60 years with severe non-dialysis requiring chronic kidney disease (CKD) who received ambulatory care in the Community Health Network (CHN) in San Francisco from January 1, 1996 to February 29, 2008. CKD was defined based on at least two outpatient estimated glomerular filtration rate (eGFR) measurements $< 60 \text{ mL/min/1.73}^2$ that were separated by at least three months; from these subjects we identified those with severe CKD defined based on the baseline eGFR measurement of $< 30 \text{ mL/min/1.73}^2$. The outcome of interest was time to ESRD (initiation of dialysis or kidney transplantation) or death. During the study, there were 321 (46%) events observed (220 (31.5%) ESRD events and 101 (14.5%) deaths without prior ESRD diagnosis). The average follow-up time was 6.2 years (max = 14.3 years) with the average time to event of 2.8 years. We used clinically relevant prediction timeframes, τ_0 , of 1 and 3 years basing our predictions on data available up to 1 or 2 years. A subject found to be at

high risk of ESRD in a 1-year timeframe would likely be recommended to meet with their healthcare team to start preparing for dialysis. A three-year timeframe would provide a subject at high risk with preventive treatment options, starting with controlling comorbid conditions such as hypertension and diabetes, lifestyle changes, among others. The longitudinal biomarker of interest was eGFR. The number of measurements available for a given individual ranged from 1–24, with over 50% of the individuals providing four or more measurements. Other covariates available were demographic (age, sex, ethnicity) and comorbidities at baseline (diabetes, hypertension, chronic viral disease, substance abuse). We were interested in predicting the conditional probability of progression to ESRD or death within a given time frame, τ_0 , given event-free survival and covariate information up to time s , $R(\tau_0|s, \mathbf{H}_\cdot(s)) = P(T_{\tau_0} \leq s + \tau_0 | T_{\tau_0} > s, \mathbf{H}_\cdot(s))$, where $\mathbf{H}_\cdot(s) = \{\mathbf{eGFR}_\cdot(s), \mathbf{s}_\cdot(s)\}$, the eGFR measurements and measurement times up to time s .

Statistical methods

To make risk predictions for every subject in our dataset, but avoid overfitting we used cross-validation. We split the dataset into five mutually exclusive subsets using stratified sampling stratifying on baseline eGFR level ($[0, 15]$, $(15, 30]$). Each of the five subsets was treated as a validation set, with the remaining 4/5 of the data used as a training set. We fit the five models introduced in Sections 2 and 3 to each of the five training sets. The first model was a PC generalized linear model with the logit link function with eGFR as the main predictor and measurement time modeled with a natural cubic spline with $df = 3$ (PC_{GLM}). Our second model was a PC Cox-type model with measurement time modeled with a natural cubic spline with $df = 3$ and eGFR as the main predictor (PC_{Cox}). The third model was the two-stage PC_{Cox} BLUP model. For each individual in the training and validation sets, fitted marker values were obtained via the empirical BLUP using REML estimates of covariance parameters from a LMEM fit to the training set (see Equations 6 and 7, Figure 1). The LMEM modeled the marker with a linear spline with a knot at 0.5 years for the fixed effects and a random intercept and slope, since the observed marker trajectories were non-linear with a tendency to increase shortly after entry into the cohort and stabilizing after about 0.5 years (PC_{Cox} BLUP). The fourth model was a two-stage PC_{GLM} model, which also used the BLUP marker values and the PC_{GLM} model for fitting (PC_{GLM} BLUP). The fifth model was a joint model, with the LME sub-model as described for model 3, the two-stage PC_{Cox} BLUP (JM). For each training set we fit each model and used it to obtain risk predictions for each individual in the corresponding validation set (for each s and τ_0 combination). The last observation prior to s was used to make predictions using PC_{Cox} , PC_{GLM} and JM. The BLUP marker values prior to s were used for prediction using PC_{Cox} BLUP and PC_{GLM} BLUP. Using the estimators of prediction evaluation measures (Section 4), we estimated the prediction error (PE), the true positive fraction (TPF) and the false positive fraction (FPF) at a risk threshold of 0.25, and the area under the ROC curve (AUC). The standard errors were estimated by perturbation.

Results

The mean (sd) eGFR among the subjects in the study cohort was $19.6 (7.6) \text{ mL/min/1.73}^2$, mean age was $45.7 (9.4)$ years, 64.3% of the subjects were male, 38.5% black, 27.0% had a history of hypertension, 14.0% were diabetic, 32.8% had a history of substance abuse and

35.0% had chronic viral disease. The estimates of the prediction capacity measures were summarized in Table 3. The prediction errors for the three year time frame were twice those for predictions for one year, with PC_{Cox} and PC_{Cox} BLUP achieving the lowest prediction errors of the five methods (Table 3). TPF and FPF at the risk threshold of 0.25 were considerably different between the methods, suggesting that the risk distributions of those with events and those without events differed between the methods. Despite these differences in risk distributions, the ability of the models to discriminate between the two groups of subjects was relatively comparable. The best discrimination between those with and without events was achieved for $s = 2$ years and $\tau_0 = 1$ year with the $AUC = 0.71$ for the JM and 0.78–0.86 for the PC models (Table 3). We note that in this example considering BLUP estimators in either of the PC models did not offer an advantage compared with ordinary PC models. Like the JM, the BLUP-based model also relies on correct specification of the longitudinal model. In practice, when longitudinal trajectory is difficult to capture well, one may observe diminished predictive performance for the BLUP models and the JM. Time-dependent ROC curves estimated using the five models are shown in Web Figure 2.

The plots of predicted risk for one example individual (Figure 3) with predictions up to 9 years based on marker data showed that the PC models behaved similarly, with the JM predictions differing substantially for $s = 2$ and 4 years. Predictions based on two and four years of data increasingly differed with increasing τ_0 . Given six years of data, the risk predictions based on the different methods ranged from 0.3 to 0.7 at nine years. This patient experienced an event at 11.4 years after study entry, therefore the PC models, assigning higher risks at later measurement times, are more appealing than the JM model for this particular individual. The trends were consistent with results summarized in Table 3.

7. Discussion

Updated predictions of risk using accumulated information among patients under active surveillance are important for disease management. The derivation of dynamic prediction rules is often challenging, requiring the consideration of both longitudinal process of the predictors and their relationship with disease outcome that can change over time. Statistical methods focusing on longitudinal prediction have appeared in the literature in recent years, however, much work is still needed.

We provide flexible approaches that can accommodate complex practical settings, and we evaluate modeling choices in order to provide practical guidance for using the available methods in outcome prediction research. Specifically, we propose estimation and inference procedures for estimating partly conditional risk, $R(\tau_0|s, \mathbf{H}(s))$, based on the covariate history $\mathbf{H}(s)$ up to time s . When large within-individual variability is observed for longitudinal covariates, and assumed to be non-systematic, predictions based on the two-stage PC_{GLM} BLUP and PC_{Cox} BLUP models, when specified well, are nice tools for smoothing the marker data prior to estimation, which can improve the prediction performance of the marker. The partly conditional models provide a versatile, flexible and robust framework for longitudinal risk prediction and are a practical alternative to joint models. One of the strengths of the PC models is that they are computationally relatively simple, and can be easily extended to more complex situations as dictated by the applied

problem at hand. For example, multiple longitudinal biomarkers can be easily incorporated without having to resort to multi-level integration, as can markers with complex longitudinal trajectories that cannot be modeled adequately with simple linear mixed effects models. Future work is needed to adopt the proposed PC models when a large number of longitudinal predictors are available but the sample size is relatively small.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work is supported by grants U01-CA86368, P01-CA053996, R01-GM085047, R01-GM079330 and R01 HL072966 awarded by the National Institutes of Health.

References

- Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*. 1950; 78:1–3.
- Bycott P, Taylor J. A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. *Statistics in Medicine*. 1998; 17:2061–2077. [PubMed: 9789914]
- Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*. 2003; 30:93–111.
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972; 34:187–220.
- Crowley J, Hu M. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*. 1977; 72:27–36.
- Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*. 1996; 15:1663–1685. [PubMed: 8858789]
- Grøn, R., Gerds, T. Handbook of survival analysis, chapter Binomial Regression Models. Chapman and Hall/CRC; 2014. p. 221–242.
- Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1:465–480. [PubMed: 12933568]
- Hjort, NL. International Statistical Review/Revue Internationale de Statistique. 1992. On inference in parametric survival data models; p. 355–387.
- Pawitan, Yudi SS. Modeling disease marker processes in AIDS. *Journal of the American Statistical Association*. 1993; 88:719–726.
- Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982; 69:331–342.
- Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*. 2009; 10:535–549. [PubMed: 19369642]
- Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*. 2010; 35:1–33. [PubMed: 21603108]
- Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011; 67:819–829. [PubMed: 21306352]
- Schoop R, Graf E, Schumacher M. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*. 2008; 64:603–610. [PubMed: 17764480]

- Taylor JM, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*. 2013; 69:206–213. [PubMed: 23379600]
- Tsiatis A, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*. 1995; 90:27–37.
- Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*. 2004; 14:809–834.
- Uno H, Cai T, Tian L, Wei LJ. Theory and methods - evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102:527–537.
- van Houwelingen, H., Putter, H. *Dynamic Prediction in Clinical Survival Analysis*. CRC Press; 2012.
- van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*. 2007; 34:70–85.
- Wang Y, Taylor JMG. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*. 2001; 96:895–905.
- Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330–339. [PubMed: 9147598]
- Zheng Y, Cai T, Feng Z. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*. 2006; 62:279–287. [PubMed: 16542256]
- Zheng YY, Heagerty PJ. Semiparametric estimation of time-dependent roc curves for longitudinal marker data. *Biostatistics*. 2004; 5:615–632. [PubMed: 15475423]
- Zheng YY, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. 2005; 61:379–391. [PubMed: 16011684]

Figure 1.

An illustration of the differences in fitting behavior between a linear mixed effects (LME) model (naïve fit) and the best linear unbiased predictor (BLUP) fit for a new individual (shown) and mean and variance structures from an LME fit to the study population. The BLUP approach uses only the *past* information to obtain BLUP values at any given observation time and provides smoothing to the individual marker data by *shrinking* the individual's marker trajectory towards the population-averaged marker trajectory. Thus, this approach requires iteration over a given individual's observations (see Section 3.3 for details). In contrast, the naïve LME fit uses past and future information of a new individual to obtain their fitted marker trajectory. The difference between the approaches is evident by comparing the trajectory of triangles (BLUP fit) with the squares (naïve fit), with the observed data shown with circles. Throughout the paper, observed data = $\mathbf{Y}$, BLUP fit = $\hat{\mathbf{Y}}$ (see Section 3.3), naïve fit = $\hat{\mathbf{Y}}^*$ (see Section 2.2). This figure appears in color in the electronic version of this article.

Figure 2.

Simulation results for individual partly conditional survival probability predictions using two-stage BLUP estimators for two example subjects. This figure appears in color in the electronic version of this article.

Figure 3.

Individual risk predictions for subject 531 from the ESRD Study. We fit five models to the ESRDS dataset excluding subject 531, and obtained risk predictions at 0.5 year increments starting at time $s = 2, 4$ and 6 years (rows) with maximum prediction time of 9 years. The curves on the left show the observed data the predicted response profiles of eGFR the risk prediction curves shown on the right were based on. The five models fit were the partly conditional logistic model(PC GLM), partly conditional Cox-type model (PC Cox), two-stage PC models (PC Cox BLUP and PC GLM BLUP) and the joint model (JM). This figure appears in color in the electronic version of this article.

Table 1

The conditional predicted risk ($R(\tau_0|s, \mathbf{H}_*(s))$) for four sets of s and τ_0 for subjects with marker = $\{0, 1\}$ estimated using the partly conditional generalized linear model with the logit link function (PC_{GLM}) and the partly conditional Cox-type model (PC_{Cox}). The true conditional risks (True) were obtained empirically. Empirical standard errors (ESD), asymptotic standard errors (ASE) and mean-squared errors ($\times 1000$) (MSE) are also provided. Data simulated with standard deviation of the measurement error $\sigma_e = 1.0$, 500 replications, $n = 500$.

$\tau_0 = 12$									
		$s = 24$		$s = 48$		$s = 24$		$s = 48$	
		$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$	$y = 0$	$y = 1$
True		0.189	0.131	0.157	0.109	0.333	0.237	0.280	0.207
PC _{GLM}									
EST		0.190	0.123	0.166	0.107	0.334	0.230	0.298	0.201
ESD		0.017	0.013	0.021	0.016	0.026	0.021	0.031	0.025
ASE		0.017	0.012	0.021	0.015	0.026	0.021	0.032	0.025
MSE		0.289	0.210	0.538	0.23	0.671	0.497	1.323	0.658
PC _{Cox}									
EST		0.191	0.125	0.171	0.111	0.332	0.229	0.301	0.206
ESD		0.011	0.009	0.009	0.007	0.019	0.015	0.016	0.013
ASE		0.011	0.008	0.009	0.007	0.018	0.015	0.015	0.013
MSE		0.116	0.113	0.287	0.055	0.317	0.292	0.662	0.163

Table 2

Estimates (EST) and empirical standard errors (ESD) of measures of predictive capacity summarizing predictions based on the partly conditional generalized linear model (PC_{GLM}), partly conditional Cox-type model (PC_{Cox}) using observed data and smoothed marker values (PC_{GLM} BLUP, PC_{Cox} BLUP), and the joint model (JM). The estimates were obtained for four sets of s and τ_0 for large measurement error ($\sigma_e = 1.0$). TPF(p) = true positive fraction at $R(\tau_0|s) = p$, FPF(p) = false positive fraction at $R(\tau_0|s, \mathbf{H}_s(s)) = p$, PE = prediction error, AUC = area under the ROC curve ($n = 500$, replications = 500).

	$\tau_0 = 12$		$\tau_0 = 24$	
	$s = 24$	$s = 48$	$s = 24$	$s = 48$
	EST (ESD)	EST (ESD)	EST (ESD)	EST (ESD)
PC_{GLM}				
PE	0.209 (0.014)	0.196 (0.023)	0.248 (0.012)	0.254 (0.018)
TPF(0.25)	0.759 (0.085)	0.608 (0.150)	0.975 (0.023)	0.945 (0.046)
FPF(0.25)	0.572 (0.089)	0.443 (0.129)	0.918 (0.045)	0.859 (0.083)
AUC	0.639 (0.042)	0.614 (0.063)	0.651 (0.039)	0.623 (0.056)
PC_{Cox}				
PE	0.209 (0.014)	0.193 (0.023)	0.248 (0.011)	0.253 (0.017)
TPF(0.25)	0.717 (0.074)	0.543 (0.117)	0.985 (0.015)	0.960 (0.036)
FPF(0.25)	0.527 (0.073)	0.384 (0.086)	0.941 (0.037)	0.892 (0.065)
AUC	0.636 (0.041)	0.609 (0.064)	0.650 (0.037)	0.619 (0.057)
PC_{GLM} BLUP				
PE	0.194 (0.014)	0.183 (0.022)	0.224 (0.013)	0.228 (0.020)
TPF(0.25)	0.782 (0.071)	0.679 (0.124)	0.956 (0.028)	0.927 (0.054)
FPF(0.25)	0.491 (0.074)	0.394 (0.103)	0.802 (0.069)	0.734 (0.099)
AUC	0.709 (0.038)	0.693 (0.056)	0.727 (0.034)	0.709 (0.053)
PC_{Cox} BLUP				
PE	0.195 (0.013)	0.180 (0.022)	0.223 (0.012)	0.229 (0.018)
TPF(0.25)	0.762 (0.065)	0.629 (0.122)	0.972 (0.020)	0.947 (0.042)
FPF(0.25)	0.466 (0.065)	0.344 (0.078)	0.852 (0.061)	0.788 (0.082)
AUC	0.708 (0.038)	0.692 (0.060)	0.729 (0.033)	0.705 (0.053)
JM				
PE	0.192 (0.014)	0.183 (0.020)	0.223 (0.014)	0.227 (0.019)
TPF(0.25)	0.794 (0.057)	0.715 (0.094)	0.961 (0.020)	0.938 (0.039)
FPF(0.25)	0.506 (0.053)	0.430 (0.073)	0.813 (0.054)	0.759 (0.073)
AUC	0.710 (0.039)	0.694 (0.059)	0.728 (0.037)	0.710 (0.053)

Table 3

(ESRD Study analysis results) Estimates (EST) and standard errors (ESD) of measures of predictive capacity summarizing predictions based on a partly conditional logistic model with the logit link function (PC_{GLM}), partly conditional Cox-type model using observed data (PC_{Cox}), smoothed marker values obtained using predicted random effects based on estimates from a linear mixed effects model and individual's data up to time s (PC_{GLM} BLUP, PC_{Cox} BLUP) and the joint model (JM). The estimates were obtained for four sets of s and τ_0 . TPF(p) = true positive fraction at $R(\tau_0|s, \mathbf{H}_s(s)) = p$, FPF(p) = false positive fraction at $R(\tau_0|s) = p$, PE = prediction error, AUC = area under the ROC curve. The number of events between s and $s + \tau_0$, and the number of subjects at risk at time s are denoted by n_e and n , respectively.

	$\tau_0 = 1$ year		$\tau_0 = 3$ years	
	$s = 1$ years ($n_e/n = 55/574$)	$s = 2$ years ($n_e/n = 31/519$)	$s = 1$ years ($n_e/n = 114/574$)	$s = 2$ years ($n_e/n = 77/519$)
PC_{GLM}				
PE	0.082 (0.015)	0.063 (0.017)	0.159 (0.038)	0.154 (0.040)
TPF(0.25)	0.745 (0.062)	0.710 (0.080)	0.736 (0.041)	0.708 (0.048)
FPF(0.25)	0.245 (0.013)	0.201 (0.014)	0.380 (0.016)	0.357 (0.017)
AUC	0.790 (0.023)	0.838 (0.028)	0.748 (0.022)	0.707 (0.026)
PC_{Cox}				
PE	0.075 (0.009)	0.053 (0.011)	0.132 (0.024)	0.128 (0.031)
TPF(0.25)	0.582 (0.060)	0.710 (0.085)	0.719 (0.043)	0.694 (0.050)
FPF(0.25)	0.135 (0.012)	0.135 (0.013)	0.333 (0.018)	0.355 (0.020)
AUC	0.791 (0.033)	0.861 (0.041)	0.771 (0.024)	0.731 (0.029)
PC_{GLM} BLUP				
PE	0.100 (0.016)	0.076 (0.018)	0.179 (0.033)	0.165 (0.038)
TPF(0.25)	0.691 (0.061)	0.645 (0.088)	0.816 (0.046)	0.777 (0.049)
FPF(0.25)	0.312 (0.016)	0.264 (0.016)	0.600 (0.021)	0.502 (0.022)
AUC	0.714 (0.034)	0.782 (0.039)	0.650 (0.027)	0.682 (0.030)
PC_{Cox} BLUP				
PE	0.078 (0.006)	0.057 (0.011)	0.139 (0.017)	0.130 (0.028)
TPF(0.25)	0.436 (0.058)	0.645 (0.090)	0.719 (0.046)	0.707 (0.052)
FPF(0.25)	0.064 (0.013)	0.121 (0.015)	0.366 (0.021)	0.374 (0.023)
AUC	0.777 (0.039)	0.851 (0.047)	0.756 (0.028)	0.731 (0.032)
JM				
PE	0.086 (0.007)	0.069 (0.011)	0.147 (0.020)	0.139 (0.023)
TPF(0.25)	0.218 (0.060)	0.323 (0.084)	0.728 (0.044)	0.635 (0.047)
FPF(0.25)	0.096 (0.015)	0.201 (0.016)	0.386 (0.021)	0.350 (0.020)
AUC	0.706 (0.038)	0.708 (0.048)	0.727 (0.027)	0.691 (0.031)