Published in final edited form as:

Stat Med. 2012 September 20; 31(21): . doi:10.1002/sim.5369.

Nonparametric Multi-state Representations of Survival and Longitudinal Data with Measurement Error

Bo Hu^{1,*}, Liang Li¹, Xiaofeng Wang¹, and Tom Greene²

¹ Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH.

² Division of Clinical Epidemiology, University of Utah, Salt Lake City, UT.

Abstract

This paper proposes a nonparametric procedure to describe the progression of longitudinal cohorts over time from a population averaged perspective, leading to multi-state probability curves with the states defined jointly by survival and longitudinal outcomes measured with error. To account for the challenges of informative dropout and nonlinear shapes of the longitudinal trajectories, a bias corrected penalized spline regression is applied to estimate the unobserved longitudinal trajectory for each subject. The multi-state probability curves are then estimated based on the survival data and the estimated longitudinal trajectories. Simulation Extrapolation (SIMEX) is further used to reduce the estimation bias caused by the randomness of the estimated trajectories. A bootstrap test is developed to compare multi-state probability curves between groups. We present theoretical justification of the estimation procedure along with a simulation study to demonstrate finite sample performance. The procedure is illustrated by data from the African American Study of Kidney Disease and Hypertension, and it can be widely applied in longitudinal studies.

Keywords

Multi-state representations; penalized spline; SIMEX

1. Introduction

In cohort studies and clinical trials, longitudinal analysis is often complicated by attrition due to death or other terminating clinical events that either prevent subsequent longitudinal measurements or else so substantially alter the definition of the longitudinal outcome that subsequent values must be excluded from the target of inference. In this setting, analysis of the longitudinal data alone may lead to conceptual inconsistencies since the longitudinal outcome is undefined at follow-up visits among subjects who experience the terminating events prior to those visits [1, 2], as well as biased inference due to informative dropout [3, 4], which occurs when the terminating events are associated with the parameters defining the probability distribution of the longitudinal data. On the other hand, survival analysis of the survival data alone does not fully utilize the information and can result in loss of efficiency, especially when the event rate is low. This situation calls for joint analysis of longitudinal and survival data. See, for example, [5, 6, 7, 8] and many others.

The most common joint analysis approach is the shared parameter model (e.g., [9]), which links the longitudinal and survival data by subject-specific frailties that represent features such as intercepts and slopes of the longitudinal trajectories. However, when the longitudinal trajectories demonstrate diverse nonlinear shapes, it is not easy to properly define simple features of the trajectories and computation can be intensive [10].

In this paper, we consider joint analysis of longitudinal and survival data from a population averaged perspective. Our method is built upon multi-state representations, which generalize the conventional competing-risk models to a set of mutually exclusive clinical states defined jointly by the longitudinal and survival outcomes. For example, in an analysis of quality of life in a cohort experiencing a high rate of mortality, states may be defined as death, alive with a decline in a quality of life score from baseline, and alive with a stable or increasing quality of life score from baseline. By estimating the state probabilities over designated time periods, the multi-state representations are able to characterize the progression pattern over the full cohort. The designation of multiple states corresponding to different levels of the longitudinal outcome allows the analysis to address questions pertaining to preservation of potentially favorable states defined by the longitudinal outcome in addition to adverse outcomes defined by the terminating events. The multi-state representations we propose differ from much of the literature on multi-state models, in which transition probabilities between states are assumed to depend on the state prior to transition, but not on other continuous variables. In the approach pursued in this paper, the longitudinal outcome is modeled over time as a continuous variable, and the classification into discrete states is made based on the observed survival outcome and the estimated longitudinal trajectory. Previous treatments of multi-state models have often considered directly observed states (e.g., [11, 12]). In this paper we assume that the longitudinal measurements are subject to random error, which may reflect measurement error or subject-specific fluctuation. Hence states defined by the longitudinal outcome are not directly observed. We also allow for the possibility that the longitudinal measurements are subject to informative dropout when a terminating event occurs.

We develop a nonparametric approach to estimate the multi-state probabilities. To account for nonlinear shapes of the longitudinal trajectories, a bias corrected penalized spline regression is applied to estimate a smooth trajectory for each subject. Based on the estimated trajectories and the survival data, each subject is classified into one of several states of clinical interests at each time point. A modified simulation-extrapolation (SIMEX) method [13] is proposed to further correct the bias in classification caused by the randomness of the estimated trajectories. We prove that the estimation procedure is approximately consistent under the condition that the SIMEX extrapolation is exact. We also provide a bootstrap procedure to compare multi-state probability curves between different treatment or exposure groups.

The remainder of the paper is organized as follows. Section 2 defines the multi-state representations. Section 3 describes the estimation procedure for the multi-state probability curves and the bootstrap test for comparing different groups. Asymptotic results are provided with the proof deferred to the appendix. Simulation results are presented in Section 4. We also illustrate the method with an application to a longitudinal chronic kidney disease study, where high proportions of patients died during the follow-up period, precluding further measurement of longitudinal outcomes altogether, or reached end stage kidney disease (ESKD), which alters the interpretation of the longitudinal outcomes so greatly that post-ESKD measurements must be excluded. Section 5 concludes the paper with some brief discussions.

2. Multi-State Representations

For subject i , $i = 1, \dots, n$, let $\mathbf{X}_i(t)$ be the true longitudinal process and Y_{ij} be the measured longitudinal outcome at time points t_{ij} , $j = 1, \dots, n_i$. Meanwhile let T_i denote the minimum event time of s_1 competing risk events that each terminate the longitudinal measurements, and let C_i denote a common censoring time that applies to each risk event. An indicator variable δ_i takes a value of 0 if subject i is censored, or a value of k if he reaches the k th competing risk event for $k = 1, \dots, s_1$. The observed outcome data for subject i can be expressed as $\{\min(T_i, C_i), \delta_i, Y_{i1}, \dots, Y_{in_i}\}$.

The survival data would naturally imply a set of clinical states related with the cohort progression. For example, in chronic kidney disease (CKD) studies, investigators are often interested at the survival outcomes of death and ESKD. In addition to these clinical events, they may also study disease states corresponding to different levels of the longitudinal outcome. Glomerular filtration rate (GFR) is a key surrogate of kidney function, which is typically measured longitudinally in CKD studies. According to the guidelines of National Kidney Foundation [14] for staging renal disease, individuals with GFR lower than 60 mL/min/1.73m² are classified as having Stage III chronic kidney disease. Those with GFR lower than 30 are classified as having Stage IV kidney disease. Thus one may consider five clinical states for CKD studies, including two survival states (death and ESKD) and three states based on GFR levels (< 30 , $30 \sim 60$ and > 60).

To generalize the definition, we introduce $b_1 < \dots < b_{s_2+1} = \infty$ as a sequence of thresholds for defining the states based on the longitudinal outcome. Thus there are a total of $m = s_1 + s_2$ states. For subject i at time t , the state can be defined as a multinomial variable

$$Z_i(t) = k, \quad \text{where} \quad \begin{aligned} k &= 1, \dots, s_1, & \text{if } T_i \leq t, \delta_i = k; \\ &= s_1 + 1, \dots, s_1 + s_2, & \text{if } T_i > t, \quad b_{k-s_1} < \mathbf{X}_i(t) \leq b_{k-s_1+1}. \end{aligned} \quad (1)$$

The states $1, \dots, s_1$ are defined by the survival data only while the states $s_1 + 1, \dots, m$ are defined by both the survival and longitudinal data.

Under the assumption that the outcome variables $\{T_i, C_i, \mathbf{X}_i(t)\}_{i=1}^n$ are independent and identically distributed (i.i.d.), the state variables $Z_i(t)$, $i = 1, \dots, n$, are also i.i.d. Furthermore we assume that the censoring time C_i is independent of both T_i and $\mathbf{X}_i(t)$. Since $Z_i(t)$ is defined on the underlying process $\mathbf{X}_i(t)$, it is not always observed.

Our goal is to estimate the probability that an individual is in each state at a set of time points $\mathcal{T}$, which could be a time interval $[0, \infty]$ or a set of designated time points. The multi-state probabilities can be written as, for $k = 1, \dots, m$,

$$p_k(t) = P(Z_i(t) = k), t \in \mathcal{T}. \quad (2)$$

In the following section, we propose a two-step estimation procedure that estimates the probabilities for the states ($k = 1, \dots, s_1$) and states ($k = s_1 + 1, \dots, m$) sequentially.

3. Estimation Procedure

3.1. Estimating Probabilities for States $1, \dots, s_1$

The probabilities associated with the first s_1 states can be written as $p_k(t) = P(T_i \leq t, \delta_i = k)$ for $k = 1, \dots, s_1$, which are the cause-specific cumulative incidence probabilities. Since $p_k(t)$

depends only on $\{\min(T_i, C_i) \delta_i\}_{i=1}^n$ that are in the form of competing risk data, they can be consistently estimated with standard competing risk models [15].

Let $\#A$ denote the number of elements in set A . The number of subjects reaching state k at time t is expressed as

$$d_{kt} = \# \{i | \min(T_i, C_i) = t, \delta_i = k\}, k=1, \dots, s_1.$$

Denote the risk set at time t as

$$R(t) = \{i=1, \dots, n | \min(T_i, C_i) \geq t\}. \quad (3)$$

Then the cause-specific hazard rate at t can be estimated as $\hat{h}_k(t) = \frac{d_{kt}}{\#R(t)}$, and $p_k(t)$ as:

$$\hat{p}_k(t) = \int_0^t \hat{h}_k(s) \hat{S}(s) ds, \quad (4)$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate of the overall event-free probability $S(t) = P(T_i > t)$. As the number of subjects goes to infinity, $p_k(t)$ in (4) is asymptotically consistent and normally distributed [16].

If there is a single event ($s_1 = 1$), the probability $p_k(t)$ for this terminating event can be estimated with the standard Kaplan-Meier method. In some applications, it may be desirable to model the survival outcome using a semi-competing risk rather than a competing risk framework, in which one has the option of applying the estimation approach of [17]. For example, in the chronic kidney disease example, if data on mortality after ESKD are available, we may want to include deaths following ESKD in the death state, rather than including only deaths prior to ESKD, which is implied in the competing risk framework. Our multi-state representations can be considered as an extension of the cumulative incidence curves for the termination clinical events by incorporating additional states defined by the continuous longitudinal outcome. We describe the estimation of the probabilities for these states in the next section.

3.2. Estimating Probabilities for States $s_1 + 1, \dots, s_1 + s_2$

The estimation of the probability curves for the s_2 states associated with the longitudinal outcome poses a somewhat greater challenge, since the underlying longitudinal trajectories $\mathbf{X}_i(t)$ are unobserved. For simplicity, we describe how to estimate $p_m(t)$ and the estimation of the probabilities of other states follows straightforwardly. From (1) and (2), $p_m(t) = P(X_i(t) > b_{s_2}, T_i > t) = q_m(t)P(T_i > t)$, where

$$q_m(t) = P(\mathbf{X}_i(t) > b_{s_2} | T_i > t). \quad (5)$$

Since $P(T_i > t)$ is the overall survival probability, it can be consistently estimated by

$$1 - \sum_{k=1}^{s_1} \hat{p}_k(t). \text{ Then the remaining task is to estimate the conditional probability } q_m(t).$$

Since $q_m(t)$ depends on the underlying $\mathbf{X}_i(t)$, it is necessary to estimate $\mathbf{X}_i(t)$ based on the observed longitudinal data. Figure 1 plots the trajectories of the longitudinal outcome (GFR) of six randomly selected subjects in the AASK study (more details can be found in Section 4.1). The wide range of the nonlinear shapes of the longitudinal trajectories suggests that a flexible nonparametric regression method is desired to provide an adequate fit for the longitudinal data. While several smoothing techniques might be used for this purpose, we consider a penalized spline regression approach with a first-order bias correction in this

paper. For simplicity, we drop the subscript i in the remainder of this section unless specifically noted. But it is worth reiterating that the penalized spline regressions are subject specific.

Suppose that the observed longitudinal trajectory $Y(t)$ follows

$$Y(t) = \mathbf{X}(t) + e(t), \quad (6)$$

where $e(t)$ is a Gaussian white noise process with mean zero and variance σ_e^2 . The penalized spline regression method approximates the smooth function $\mathbf{X}(t)$ as

$$\mathbf{X}(t) = \mathbf{B}(t)^T \theta, \quad (7)$$

where θ is a vector of parameters and $\mathbf{B}(t)$ is a sufficiently flexible spline basis that is determined by the data $\{Y_{ij}, t_{ij}\}_{j=1}^{n_i}$. For example, $\mathbf{B}(t) = (1, t, |t - t_1|^3, \dots, |t - t_K|^3)^T$ for cubic thin-plate spline basis with knots $t_1, \dots, t_K$. The parameter θ shall be estimated by minimizing a penalized quadratic loss function

$$L(\theta) = (\mathbf{Y} - \mathbf{B}\theta)^T (\mathbf{Y} - \mathbf{B}\theta) + \lambda \theta^T D \theta,$$

where $\mathbf{Y} = (Y_{i1}, \dots, Y_{in_i})^T$ and $\mathbf{B} = (\mathbf{B}(t_{i1}), \dots, \mathbf{B}(t_{in_i}))^T$; λ is a smoothing parameter and $D = \text{diag}(0, 0, 1_K)$ is a penalty matrix. See [18] for details on the selection of knots and penalty parameters. The resulting estimate of θ is

$$\hat{\theta} = H_\lambda^{-1} \mathbf{B}^T \mathbf{Y}$$

with $H = \mathbf{B}^T \mathbf{B} + D$, and the associated estimate of $\mathbf{X}(t)$ is then $\hat{\mathbf{X}}(t) = \mathbf{B}(t)^T \hat{\theta}$.

If the true trajectory $\mathbf{X}_i(t)$ was known, $q_{mk}(t)$ could be consistently estimated by

$\frac{\#\{i \in R(t); \mathbf{X}_i(t) > b_{s_2}\}}{\#\{R(t)\}}$, the proportion of subjects with $\mathbf{X}_i(t)$ greater than b_{s_2} in the risk set. Since $\mathbf{X}_i(t)$ is unknown, a natural approach is to replace it with the smoothed estimate $\hat{\mathbf{X}}_i(t)$. However, $\hat{\mathbf{X}}_i(t)$ is biased due to penalized spline smoothing. More specifically, under models (6) and (7), we have

$$\begin{aligned} \hat{\mathbf{X}}(t) &= \mathbf{B}(t)^T \theta + \mathbf{B}(t)^T (\hat{\theta} - \theta) \\ &= \mathbf{X}(t) + \mathbf{B}(t)^T (H_\lambda^{-1} \mathbf{B}^T \mathbf{Y} - H_\lambda^{-1} H_\lambda \theta) \\ &= \mathbf{X}(t) - \lambda \mathbf{B}(t)^T H_\lambda^{-1} D \theta + \mathbf{B}(t)^T H_\lambda^{-1} \mathbf{B}^T \mathbf{e} \\ &= \mathbf{X}(t) + b(t) + U(t) \end{aligned} \quad (8)$$

where $\mathbf{e} = (e(t_{i1}), \dots, e(t_{in_i}))^T$, $b(t) = -\lambda \mathbf{B}(t)^T H_\lambda^{-1} D \theta$ and $U(t) = \mathbf{B}(t)^T H_\lambda^{-1} \mathbf{B}^T \mathbf{e}$.

It can be shown that the systematic bias $b(t)$ has the stochastic order of $O_p(1/(n_i + 1))$ [19]. Correcting bias in penalized spline regression has been studied extensively in the literature (see e.g., [19]). We here consider a simple correction procedure because this is not the focus of the paper. In principle, one may use any other correction procedure such as Wild Bootstrap [20]. Replacing the unknown parameter θ in $b(t)$ with $\hat{\theta}$ leads to an estimate of the systematic bias $\hat{b}(t) = -\lambda \mathbf{B}(t)^T H_\lambda^{-1} D \hat{\theta}$. A corrected smoothing estimate can be obtained by subtracting $\hat{b}(t)$ from $\hat{\mathbf{X}}(t)$, that is,

$$\hat{\mathbf{X}}_c(t) = \hat{\mathbf{X}}(t) - \hat{b}(t) = \hat{\mathbf{X}}(t) + \lambda \mathbf{B}(t)^T H_\lambda^{-1} D \hat{\theta}. \quad (9)$$

Then we propose the following estimator for the conditional probability $q_m(t)$,

$$\hat{q}_m^c(t) = \frac{\# \{i \in R(t); \hat{\mathbf{X}}_{ci}(t) > b_{s_2}\}}{\# \{R(t)\}}. \quad (10)$$

Similarly, for $k = s_1 + 1, \dots, m - 1$, $\hat{q}_k^c(t)$ can be obtained by replacing $\mathbf{X}_{ci}(t) = b_{s_2}$ in (10) with $b_{k-s_1} < \mathbf{X}_{ci}(t) < b_{k-s_1+1}$. For the multi-state probabilities associated with the states $k = s_1 + 1, \dots, m$, their estimates are then

$$\hat{p}_k^c(t) = \hat{q}_k^c(t) \hat{S}(t), \quad (11)$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate of the overall event-free probability.

Remark 1. The bias of the original smoothing estimate $\mathbf{X}(t)$ is caused by the introduction of the penalty term in the likelihood. The corrected estimate $\mathbf{X}_c(t)$ removes the first order bias of $\mathbf{X}(t)$, that is $E(\mathbf{X}_c(t)) = \mathbf{X}(t) + o_p(1/n_j)$. On the other hand, if the penalty term is not imposed ($\lambda = 0$), the spline model reduces to regression spline and the resulting estimate of the trajectory $\mathbf{X}(t)$ is unbiased. However, in the case that there are sparse regions between spline knots, non-convergence problem could arise for regression spline since the design matrix is not of full rank. Also, the regression spline may produce a wiggly fit of $\mathbf{X}(t)$.

There exists a convolution issue for the estimator in (11), even with perfect bias correction in (9). In the case that the smoothing parameter λ is 0, $b(t) = 0$ in (8) and $\mathbf{X}_c(t) = \mathbf{X}(t) + U(t)$. The presence of $U(t)$ leads to a misclassification in (10) because the indicator $\mathbf{I}\{\mathbf{X}_c(t) > b_{s_2}\}$ is biased for $\mathbf{I}\{\mathbf{X}(t) > b_{s_2}\}$. This is similar to the classical measurement error problem, where $U(t)$ can be considered as an additive measurement error. The misclassification rate depends on the variance of $U(t)$, or equivalently the variance of $\mathbf{X}_c(t)$. To further correct the bias due to misclassification, we describe the application of the SIMEX method by [13] in the following section. SIMEX is an attractive approach to this problem because it does not require distributional assumption on $\mathbf{X}(t)$ and works very well when the variance of the measurement error is small.

3.3. SIMEX Correction

Suppose that the remaining higher-order bias in (9) can be ignored, we can write the corrected estimate as

$$\hat{\mathbf{X}}_c(t) \approx \mathbf{X}(t) + U_c(t), \quad (12)$$

where the Gaussian random variable $U_c(t)$ has mean zero and variance

$$\sigma_u^2 = \sigma_e^2 \mathbf{B}(t)^T \left[\mathbf{C} H_\lambda^{-1} \mathbf{B}^T \mathbf{B} H_\lambda^{-1} \mathbf{C} \right] \mathbf{B}(t)$$

with $\mathbf{C} = [\mathbf{I} + \lambda H_\lambda^{-1} D]$. The variance is usually larger than the variance of $U(t)$ in (8) since more variation is introduced through bias correction (9). Under (12),

$$P(\hat{\mathbf{X}}_c(t) > b_{s_2}) \approx P(\mathbf{X}(t) + U_c(t) > b_{s_2}) \neq P(\mathbf{X}(t) > b_{s_2})$$

causes misclassification bias. Since $U_c(t)$ only depends on the random error in model (6), it is independent of $\mathbf{X}(t)$. The misclassification problem can be considered as a convolution problem with heteroscedastic measurement errors [21] since the variance of $U_c(t)$ varies with subjects.

For a set of pre-determined positive numbers $\mathcal{G} = \{0 < \gamma_1 < \dots < \gamma_G\}$, the SIMEX algorithm can be summarized as follows:

1. *Simulation Step.* For each γ in $\mathcal{G}$, generate a large number of error-inflated pseudo data. For $b = 1, \dots, B$,

$$\tilde{\mathbf{X}}_{ib}(t; \gamma) = \hat{\mathbf{X}}_{ci}(t) + \gamma^{1/2} u_{ib}(t),$$

where u_{ib} is randomly drawn from $N(0, \sigma_{ui}^2)$, the distribution of $U_c(t)$.

- 2.

$$\text{Estimation Step. Compute } \hat{q}_m(t; \gamma) = \frac{1}{B} \sum_{b=1}^B \frac{\# \{i \in R(t); \tilde{\mathbf{X}}_{ib}(t; \gamma) > b_{s_2}\}}{\# \{R(t)\}}.$$

3. *Extrapolation Step.* Fit a quadratic model based on $q_m(t; \gamma)$, $\gamma \in \mathcal{G}$, and extrapolate at $\gamma = -1$. The extrapolation, denoted by $\hat{q}_m^s(t)$, is the final estimate of $q_m(t)$.

Remark 2. The simulation step can be bypassed as in Stefanski and Bay (1996). By the strong Law of Large Numbers, as B goes to ∞ , the limit of $q_m(t; \gamma)$ is

$1 - \frac{1}{n_t} \sum_{i \in R(t)} \Phi\left(\frac{b_{s_2} - \hat{\mathbf{X}}_{ci}(t)}{\sigma_{ui} \gamma^{1/2}}\right)$, where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. σ_{ui}^2 is unknown and can be estimated by

$\hat{\sigma}_{ei}^2 \mathbf{B}_i(t)^T [\mathbf{C} \mathbf{H}_\lambda^{-1} \mathbf{B}^T \mathbf{B} \mathbf{H}_\lambda^{-1} \mathbf{C}] \mathbf{B}_i(t)$, where $\hat{\sigma}_{ei}^2$ is the estimated residual variance in fitting model (6). Since the target $q_m(t)$ is a probability, we shall truncate the extrapolation $\hat{q}_m^s(t)$ on the interval $[0, 1]$.

Asymptotic results of the SIMEX procedure have been established in [22]. Proposition 1 generalizes their results to $\hat{q}_m^s(t)$, as an estimator for the conditional probability $q_m(t)$. The proof is deferred to the Appendix.

Proposition 1. Suppose that (i) $\sigma_{ui} > 0$ for all i ; (ii) the quadratic extrapolation is exact. Then the estimator $\hat{q}_m^s(t)$ is asymptotically consistent under (12).

In practice, the success of the SIMEX method depends on the fact that $E(q_m(t; \gamma))$ can be well approximated by a quadratic function of γ . Conventional choices of γ are positive numbers that evenly span an interval, say $[0, 3]$. The quadratic extrapolation is very common in SIMEX methods. See more discussion about it in [23].

Consequently, for the original multi-state probabilities, their final estimates are

$$\hat{p}_k^s(t) = \hat{q}_k^s(t) \hat{S}(t) \quad (13)$$

for $k = s_1 + 1, \dots, m$, where $\hat{S}(t)$ is the Kaplan-Meier estimate of the overall event-free probability. The consistency results of $\hat{p}_k^s(t)$ follow from Proposition 1 and the consistency results of $\hat{S}(t)$. By estimating the multinomial probabilities at each time point in $\mathcal{T}$, we can obtain the multi-state probability curves that graphically summarize the distribution of subjects in these clinical states over time.

3.4. Two-group Comparison of Multi-state Probability Curves and Confidence Intervals

Let $\mathbf{P}^l(t) = (p_1^l(t), \dots, p_m^l(t))^T$ be the multi-state probability curves for group l , $l = 1, 2$. The general null hypothesis can be formulated as

$$H_0: L\mathbf{P}^1(t) = L\mathbf{P}^2(t), t \in \mathcal{T}, \quad (14)$$

where L is a design matrix that is specified by the investigators. One may use $L = (w_1, \dots, w_m)^T$ with w_k being a rank score of disease severity associated with the clinical state k . For example, w_k may be the utility measurement that is often used in computing quality-adjusted life years [24].

Denote $\mathbf{p}^l(t)$ as the estimate of $\mathbf{P}^l(t)$ obtained by using the method described in the earlier sections. For the states associated with both longitudinal and survival data, the probabilities can be estimated from either (11) or (13). The test statistic is defined as

$$T = \int_{t \in \mathcal{T}} \|L(\hat{\mathbf{p}}^1(t) - \hat{\mathbf{p}}^2(t))\|^2 dt, \quad (15)$$

where $\|\cdot\|$ is L^2 -norm. Computing the p -value requires the sampling distribution of the T -statistic, which is determined by the joint sampling distribution of probability vector $\mathbf{p}(t)$. However, due to the complication in estimating the probabilities in (13), deriving the closed form of the sampling distribution of $\mathbf{p}(t)$ becomes intractable. Thus we suggest using bootstrap approach to approximate the null distribution of the T -statistic.

Bootstrap should be carried out by resampling the subjects and multi-state probabilities will be computed based on the bootstrap subjects. Since the longitudinal trajectory $\mathbf{X}_i(t)$ has been estimated for each subject in computing the observed T -statistic, it does not need to be re-estimated if subject i is sampled in a bootstrap simulation, which saves significant computation time. The p -value can be empirically computed as the percentage that the bootstrap T -statistics exceed the observed one. Similarly, confidence intervals of the multi-state probabilities can be obtained by using the bootstrap approach.

In our bootstrap procedure, convergence of the bootstrap statistic to an appropriate limit requires certain smoothness conditions on the functional corresponding to the estimator. Employing asymptotic normality for the statistic requires the existence of an Edgeworth expansion for its distribution function [25]. The theoretical development for the bootstrap procedure is beyond the focus of the current paper and will be investigated in our future research.

4. Numerical Results

4.1. Application to the AASK Study

The African American study of Kidney Disease (AASK) was a randomized clinical trial with a follow-on cohort study in 1094 African Americans with chronic hypertensive kidney disease [26]. Patients were randomized into two blood pressure control treatment arms: low vs usual. The follow-up period was up to 12 years. The AASK study database includes an

extensive array of longitudinal outcomes, and provides a rare opportunity to study the long-term progression of kidney disease. Almost half of the cohort (494 patients) reached end stage kidney disease (ESKD) or death prior to the end of the study.

In this paper we focus on the longitudinal outcome of estimated glomerular filtration rate (eGFR), as determined by serum creatinine, age, and gender [26]. The GFR levels for young individuals with healthy kidneys typically exceed 100 mL/min/1.73m², but with CKD progressively decline until a value of approximately 10 mL/min/1.73m² is reached, at which point the patient's kidneys would no longer support life without renal replacement therapy in the form of dialysis or kidney transplantation. ESKD is defined to occur on the date of initiation of renal replacement therapy.

In the AASK study, eGFR was assessed at baseline, 3 months, 6 months, and approximately every 6 months thereafter until the patient died, reached ESKD, or was otherwise lost to follow-up. The maximum number of eGFR measurements was 24. We note that eGFR does not determine the “true” GFR exactly, but is subject to a variety of sources of error. These include measurement error in the serum creatinine assay, as well as deviations resulting from variations among patients in creatinine generation and creatinine secretion [27].

We consider five clinical states, including death, ESKD, and three states based on GFR (30, 30 ~ 60 and > 60), that is, $b_1 = -$, $b_2 = 30$, $b_3 = 60$ and $b_4 =$. By ESKD we mean occurrence of ESKD prior to death, and by death we mean occurrence of death prior to ESKD, as the AASK database does not contain complete information on mortality following ESKD. Hence, both death and ESKD are treated as absorbing states, and we ignore the partial transitions from ESKD to death observed in the study. Patients may repeatedly transition among the three higher states defined by different levels of GFR.

Figure 2 compares the estimated multi-state probability curves by two treatment groups (low BP vs. usual BP). The probabilities for the non-absorbing states were obtained from the SIMEX corrected estimator in (13). It appears that the two groups show similar disease progression pattern. The bootstrap test yields a p-value of 0.71 in comparing the probability curves between the two groups.

Proteinuria has been recognized as an important risk factor for chronic renal disease [28], which is measured by the ratio of urinary protein and urinary creatinine (UP/Cr). Patients are usually classified as having low proteinuria if UP/Cr is less than 0.22, and having high proteinuria otherwise [26]. Figure 3 shows the multi-state probability curves for patients with low and high baseline proteinuria, respectively. The bootstrap p -value is less than 0.001, which indicates that the disease progresses significantly faster in patients with high proteinuria.

4.2. Simulation Study

The first part of the simulation study was to examine the accuracy of the estimated multi-state probability curves. The following quadratic-spline model was assumed for the longitudinal process,

$$\begin{aligned} Y_i(t) &= \mathbf{X}_i(t) + e_i(t), \\ \mathbf{X}_i(t) &= b_{0i} + b_{1i}t + b_{2i}t^2 + b_{3i}(t - 30)_+^2 + b_{4i}(t - 60)_+^2 = \mathbf{B}(t)^T \mathbf{b}_i, \end{aligned}$$

where $e_i(t)$ is a Gaussian white noise process with standard deviation $\sigma_e = 4$. Longitudinal data were simulated at time points $t_{jj} = 6(j - 1)$, $j = 1, \dots, 21$. For simplicity, univariate

survival outcome was considered in the simulation study and the survival time was simulated from the following accelerated failure time (AFT) model,

$$\log(T_i) = \alpha_0 + \alpha_1^T \mathbf{b}_i + \delta_i,$$

where $\delta_i \sim N(0, \sigma_T^2)$ is independent of $\mathbf{b}_i$ and $e_i(t)$. The censoring time C_i was independently simulated from a truncated Gaussian distribution $N(120, 30)\mathbf{I}_{(-\infty, 120)}$. A simulated Y_{ij} was discarded if $t_{ij} > \min(T_i, C_i)$ so that the problem of informative dropout exists in the simulation data.

For subject i , the value of $\mathbf{b}_i = (b_{0i}, \dots, b_{4i})^T$ was randomly drawn from a 5-dimensional Gaussian distribution with mean vector μ_b and variance-covariance matrix Σ_b . μ_b and Σ_b were selected so that the simulation study approximates the AASK study. On the other hand, the parameters α_0 and $\alpha_1 = 0.2$ in the AFT model were chosen to yield an approximate cumulative event rate of 45% at $t = 120$. Numerical values of μ_b , Σ_b and σ_T are given in the appendix. The censoring rate was about 20% as in the AASK study. The number of repeated measurements, n_i , ranges from 3 to 21. For subjects with only three measurement, their trajectories were fit with linear regression models. Similarly, we also categorize the longitudinal outcome into 3 states as the AASK study with $b_1 = -$, $b_2 = 30$, $b_3 = 60$ and $b_4 =$.

We considered two sample sizes $n = 50$ and $n = 200$. Table 1 shows the simulation results of the estimated multi-state probabilities at time points 30, 60, 90 and 120. The estimated probabilities are the averages across 500 simulations. The true probabilities can be easily computed since they depend on linear transformations of the multivariate normal vector

$(\mathbf{b}_i^T, \delta_i)^T$. We compared the performance of the two proposed estimators, $\hat{p}_k^c(t)$ in (11) and $\hat{p}_k^s(t)$ in (13), and a raw estimator. For simplicity, we refer to the estimation methods for $\hat{p}_k^c(t)$ and $\hat{p}_k^s(t)$ as “method I” and “method II”, respectively. For the raw estimator, we treat the measured $Y_i(t)$ as if there is no random error. For example, the raw estimator of $q_m(t)$ is $\hat{q}_m^r(t) = \frac{\#\{i \in R(t); Y_i(t) > 60\}}{\#\{R(t)\}}$. Note that the raw estimator can be only computed at the designed time points (every 6 months) when $Y_i(t)$ was simulated.

For state 1, the estimates are basically the Kaplan-Meier estimates for all three methods. For the other states, it appears that the estimates by method II are closest to the true probabilities at all four time points. The estimates by method I have slightly larger biases than method II. The difference is not large since the variance of the corrected smoothing estimate $\mathbf{X}_c(t)$ is much smaller than the mean and variance of $\mathbf{X}(t)$ when $\sigma_e = 4$. The misclassification rate (e.g., $P(\mathbf{X}_c(t) > 60) - P(\mathbf{X}(t) > 60)$) is thus small and might be ignored. However, such phenomenon may not always hold in general. In situations where the residual variance is large and/or number of repeated measurements per subject is small, $\mathbf{X}_c(t)$ may have larger variance and method I would have non-ignorable bias in comparison with method II. For example, we increased the variance ratio of noise and signal by letting $\sigma_e = 8$. Meanwhile we also reduced n_i by simulating longitudinal measurements every 12 months, i.e., $n_i = 11$ if subject i was not censored. Figure 4 compares the estimated probabilities for state 4 ($n = 200$) by methods I and II over the time interval $[0, 120]$ for $\sigma_e = 4$ and 8. The probability curves obtained from method II are both close to the true curve. On the other hand, apparent bias is observed for the curve estimated by using method I when $\sigma_e = 8$. Similar results were observed for states 2 and 3 and are not shown here. Although the SIMEX correction offers limited improvement when the residual variance is relatively small as in the AASK study, it has advantages in longitudinal studies with larger residual variance. Method II is also more

rigorous in theory and only adds little computation complexity. The raw estimates generally have substantial bias since they use the raw measurement $Y_k(t)$.

For each probability estimate, the number in the parenthesis represents the standard deviation of the 500 simulated estimates. The standard deviations for $n = 200$ are about a half of those for $n = 50$. A similar trend is also observed in the integrated mean squared error (IMSE), which is defined as $\int_0^{120} (p_k(t) - \hat{p}_k(t))^2 dt$ for state k . For the raw estimator, IMSE is not available since it only estimates the state probabilities at the prespecified time points.

The second part of the simulation study was to evaluate the validity and power of the bootstrap test in comparing the multi-state probability curves between two groups ($n = 100$ per group). Two hypotheses were tested: (i) $p_1^1(t) = p_1^2(t)$ and $p_k^1(t) = p_k^2(t)$, $k = 1$, where hypothesis (i) was about the equivalence of survival curves only while hypothesis (ii) was about the equivalence of other curves. The power was computed as the percentage of rejecting the null hypothesis among 500 simulations. For each simulation run, the bootstrap test was performed with 1000 bootstrap replicates.

The data for group 1 were generated from the model used in the first part of the simulation. To generate data for group 2, we vary the linear slope of the longitudinal process, the second component of the vector μ_b . All other parameters were same for the two groups. Figure 5 plots the power curves of the bootstrap test in testing the two hypotheses. For the bootstrap test, we used the estimates by method II. Similar results were obtained for method I, but are not shown here. The horizontal axis represents the ratio of the linear slopes of the two groups. When this ratio is 1, the simulated power of the bootstrap test is 0.048 for both hypotheses, which supports the validity of the bootstrap test. As a comparison, the power of the log-rank test is 0.046 for the first hypothesis. When the ratio is greater than 1, the power of the bootstrap test gradually increases from 0.15 to 0.81 for the second hypothesis. The bootstrap test has similar power as the log-rank test for the first hypothesis.

5. Conclusion and Discussion

We propose a nonparametric procedure to estimate multi-state probabilities defined jointly by unobserved longitudinal trajectories and observed survival data. In most applications, the observed longitudinal data deviate from the true trajectory by the amount of random error due to biological fluctuation and measurement error. Penalized spline regression with a bias correction procedure is used to estimate the longitudinal trajectories with diverse nonlinear shapes. A modified SIMEX procedure is developed to further correct the classification bias caused by the randomness of the estimated longitudinal trajectories. Source code is available online at <http://stat.cwru.edu/~bxh123>. In the simpler setting in which the true longitudinal outcome is observed, the SIMEX procedure is not needed, and the multi-state probabilities can be readily estimated as the empirical percentages of subjects in the pre-specified states according to different levels of the longitudinal outcome.

Informative dropout is an analytical challenge for many longitudinal studies, where termination of follow-up for the longitudinal outcome by terminating clinical events may lead to biased estimates of parameters characterizing the mean longitudinal trajectories prior to those events. In this paper, we overcome this challenge by fitting the longitudinal trajectory for each subject separately and then averaging across all the subjects to obtain the estimated probabilities. This method is similar to the unweighted least squares method used in [5] and [8], which is more robust than the common joint-modeling approach [30] in the sense that it requires less assumptions, which are typically needed in specifying the likelihood for the joint models. It is particularly useful when the trajectories are not linear. Since we advocate the multi-state probability plot as a descriptive tool, we are in favor of

using the proposed approach for estimating the multi-state probabilities, though the likelihood-based approaches should also be applicable. In fitting the longitudinal trajectories, it may be necessary to assume linear or constant trajectories for subjects with small n_j (e.g., less than 3). Since these subjects typically drop out of the risk set early, they have little effect on the estimated multi-state probability curves for most of the later follow-up period, where the state probabilities are of most clinical interests.

When the longitudinal trajectories demonstrate diverse shapes and are correlated with the terminating clinical events, it is difficult to define a clinically meaningful comparison for different treatment groups solely based on the longitudinal or the survival data. For example, comparing the change (slope) of the longitudinal outcome might not be meaningful if the incidence rates of the clinical events differ systematically. In contrast, comparing the incidence rates of the clinical events, without making use of the longitudinal data, might result in inefficient analysis. The loss of efficiency may be considerable when the rate of the clinical events is low. Furthermore, for nonlinear longitudinal trajectories, it is difficult to develop simple statistics such as intercept and slope that are unaffected by informative dropout. The proposed methodology avoids the aforementioned difficulties and can be useful in many longitudinal cohort studies where both the longitudinal biomarkers and clinical events are informative. The multi-state probability curve can be viewed as a composite outcome that bears easy interpretation, and an intuitive graphical visualization of the progression of the cohort over time.

We want to emphasize that the theoretical and empirical results in this paper rely on the assumption that the total number of subjects is large, but not on the assumption that the number of repeated measurements per subject, n_j , is large. Although a larger n_j would improve the results by reducing the approximation biases of both the penalized spline regression and SIMEX, the simulation shows that under realistic conditions similar to the application, the empirical performance of the proposed estimators, with proper bias correction, is satisfactory. Biomedical data usually involve many sources of variation, and it is often reasonable to assume a fairly smooth underlying trajectory. Thus a small number of knots might be sufficient to ensure small approximation error. When n_j is very small for all subjects, it may be inadequate to identify any nonlinear pattern efficiently. In this case, if the repeated biomarker measurements are made at design-fixed time points, the SIMEX method can be used to estimate the multi-state probabilities at each design-fixed time point. If the time points are irregularly spaced but a linear trajectory can be assumed, one can replace the penalized spline regression with linear regression. We discuss briefly in the appendix about adapting the estimation procedure for the multi-state probabilities to the transition probabilities between different states.

The bootstrap method suggested for comparing multi-state probability curves only applies to a single group variable. A desired extension is to allow for covariate adjustment. A working project is to adapt the inverse probability weighting approach for estimating the Kaplan-Meier curves of survival data [29] to the multi-state probability curves. Details of the extension are beyond the scope of this paper and will be reported elsewhere.

Acknowledgments

We thank the editor and two reviewers for their helpful comments, which significantly improve the quality of this paper. This study was partially supported by NIH grants: U01 DK048652, UL1 RR024989 and R01 DK090046-01.

Appendix

Proof of Proposition 1: Denote $n_t = \# \{R(t)\}$. By the definition of $R(t)$, $\lim_{n \rightarrow \infty} \frac{n_t}{nS(t)} = 1$. By the strong Law of Large Numbers, as $B \rightarrow \infty$,

$$\begin{aligned} \hat{q}_m(t; \gamma) &\rightarrow_{a.s.} \frac{1}{n_t} \sum_{i \in R(t)} E \left(\tilde{\mathbf{X}}_{ib}(t; \gamma) > b_{s_2} | \hat{\mathbf{X}}_{ci}(t) \right) \\ &= 1 - \frac{1}{n_t} \sum_{i \in R(t)} \Phi \left(\frac{b_{s_2} - \hat{\mathbf{X}}_{ci}(t)}{\sigma_{ui} \gamma^{1/2}} \right), \end{aligned}$$

where the last equality follows from the distribution of u_{ib} ; and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Therefore the simulation step can be bypassed in the SIMEX algorithm as the conventional CDF estimation in Stefanski and Bay (1996).

As $n_t \rightarrow \infty$, by the Law of Large Numbers, $\frac{1}{n_t} \sum_{i \in R(t)} \Phi \left(\frac{b_{s_2} - \hat{\mathbf{X}}_{ci}(t)}{\sigma_{ui} \gamma^{1/2}} \right)$ converges to its expectation

$$\frac{1}{n_t} \sum_{i \in R(t)} E \left(\Phi \left(\frac{b_{s_2} - \hat{\mathbf{X}}_{ci}(t)}{\sigma_{ui} \gamma^{1/2}} \right) \right),$$

where the expectation is taken over the distribution of $\mathbf{X}_{ci}(t)$. Under (12), we have

$$\begin{aligned} E \left(\Phi \left(\frac{b_{s_2} - \hat{\mathbf{X}}_{ci}(t)}{\sigma_{ui} \gamma^{1/2}} \right) \right) &\approx E \left(\Phi \left(\frac{b_{s_2} - \mathbf{X}_i(t) - \sigma_{ui} U}{\sigma_{ui} \gamma^{1/2}} \right) \right) \\ &= \int \int \Phi \left(\frac{b_{s_2} - \mathbf{X}_i(t) - \sigma_{ui} u}{\sigma_{ui} \gamma^{1/2}} \right) f_{\mathbf{X}}(x) \phi(u) dx du \\ &= \int \int F_{\mathbf{X}} \left(b_{s_2} - \sigma_{ui} (u + \gamma^{1/2} v) \right) \phi(v) \phi(u) dv du \\ &= E \left(F_{\mathbf{X}} \left(b_{s_2} - \sigma_{ui} (1 + \gamma)^{1/2} Z \right) \right), \end{aligned}$$

where $Z \sim N(0, 1)$; $f_{\mathbf{X}}(\cdot)$ and $F_{\mathbf{X}}(\cdot)$ are the PDF and CDF of $\mathbf{X}_i(t)$ conditioning on $T_i > t$, respectively; and $\phi(\cdot)$ is the density function of the standard normal distribution. By letting

$$-1, \lim_{\gamma \rightarrow -1} E \left(F_{\mathbf{X}} \left(b_{s_2} - \sigma_{ui} (1 + \gamma)^{1/2} Z \right) \right) = F_{\mathbf{X}}(b_{s_2}) \text{ under condition (i).}$$

Since $\mathbf{X}_{ci}(t)$, $i \in R(t)$, are independent of each other, $\lim_{\gamma \rightarrow -1} \hat{q}_m(t; \gamma)$ is asymptotically consistent with $1 - F_{\mathbf{X}}(b_{s_2}) = P(\mathbf{X}_i(t) > b_{s_2} | T_i > t)$ as $n_t \rightarrow \infty$. Under condition (ii),

$E(\hat{q}_m(t; \gamma)) = \beta_0 + \beta_1 \gamma + \beta_2^2 \gamma^2$. Then,

$$\beta_0 - \beta_1 + \beta_2 = \lim_{\gamma \rightarrow -1} E(\hat{q}_m(t; \gamma)) = P(\mathbf{X}_i(t) > b_{s_2} | T_i > t). \text{ The extrapolation}$$

$\hat{q}_m^c(t) = \hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2$ is asymptotically consistent with $P(\mathbf{X}_i(t) > b_{s_2} | T_i > t)$. This completes the proof.

Numerical values of μ_b , σ_b and γ in the simulation study: $\mu_b = (55, -0.10, -0.0012, -0.0020, 0.001)^T$, $\sigma_b = ((3.5, 0.026, 1, 1, 1, 1))^T$ and $\gamma = \text{diag}(144, 0.04, 3.75e^{-7}, 1.03e^{-6}, 2.5e^{-7})$.

Estimation of Transition Probabilities: We illustrate the estimation procedure by describing how to estimate the transition probability from state $s_1 + 1$ at time t to state s_1 at time $u > t$,

that is, $P_{s_1+1, s_1}(t, u) = P(Z_i(u) = s_1 | Z_i(t) = s_1 + 1) = \frac{P(Z_i(t) = s_1 + 1, Z_i(u) = s_1)}{P(Z_i(t) = s_1 + 1)}$, where the numerator $P(Z_i(t) = s_1 + 1, Z_i(u) = s_1)$ can be written as

$$\begin{aligned} & P(\mathbf{X}_i(t) \leq b_2, t < T_i \leq u, \delta_i = s_1) \\ &= P(\mathbf{X}_i(t) \leq b_2, T_i \leq u, \delta_i = s_1) \\ &\quad - P(\mathbf{X}_i(t) \leq b_2, T_i \leq t, \delta_i = s_1) \\ &= P(\mathbf{X}_i(t) \leq b_2 | T_i \leq u, \delta_i = s_1) P(T_i \leq u, \delta_i = s_1) \\ &\quad - P(\mathbf{X}_i(t) \leq b_2, T_i \leq t, \delta_i = s_1) P(T_i \leq t, \delta_i = s_1). \end{aligned}$$

Since $P(T_i \leq t, \delta_i = s_1)$ can be estimated as $p_{s_1}(t)$ in (4) and $P(Z_i(t) = s_1 + 1)$ can be estimated as $\hat{p}_{s_1+1}^s(t)$ in (13), the remaining task is to estimate the conditional probabilities $P(\mathbf{X}_i(t) \leq b_2 | T_i \leq t, \delta_i = s_1)$ and $P(\mathbf{X}_i(t) \leq b_2 | T_i \leq u, \delta_i = s_1)$. Let $R_2(t) = \{i = 1, \dots, n | \min(T_i, C_i) \leq t, \delta_i = s_1\}$. Then the estimation of $P(\mathbf{X}_i(t) \leq b_2 | T_i \leq t, \delta_i = s_1)$ follows the derivation of $\hat{q}_{s_1+1}^s(t)$ by replacing the risk set $R(t)$ in (10) by $R_2(t)$.

References

1. McConnell S, Stuart E, Devaney B. What to do in an experimental evaluation when the outcome is not always measured. *Evaluation Review*. 2008; 32:157–186. The truncation by death problem. [PubMed: 18319423]
2. Rubin D. Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death. *Statistical Science*. 2006; 21:299–309.
3. Wu MC, Bailey KR. Analyzing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*. 1988; 7:337–346. [PubMed: 3281208]
4. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*. 1992; 11:1861–1870. [PubMed: 1480878]
5. Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
6. Tsiatis AA, DeGruttola V, Wulfson 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.
7. Schluchter MD, Greene T, Beck GJ. Analysis of change in the presence of informative censoring - application to a longitudinal clinical trial of progressive renal disease. *Statistics in Medicine*. 2001; 20:989–1007. [PubMed: 11276031]
8. Li L, Hu B, Greene T. A semiparametric joint model for longitudinal and survival data with application to Hemodialysis study. *Biometrics*. 2009; 65:737–745. [PubMed: 19173700]
9. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*. 2006; 25:143–63. [PubMed: 16025541]
10. Song X, Wang CY. Semiparametric Approaches for Joint Modeling of Longitudinal and Survival Data with Time-Varying Coefficients. *Biometrics*. 2008; 64:557–566. [PubMed: 17725812]
11. Klein JP, Shu YY. Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research*. 2002; 11:1171–139.
12. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*. 2007; 26:2389–2430. [PubMed: 17031868]
13. Cook JR, Stefanski LA. Simulation extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*. 1994; 89:1314–1328.
14. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease. 2002
15. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. 2nd ed.. Wiley; New York: 2002.

16. Lin DY. Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*. 1997; 6:901–910. [PubMed: 9160487]
17. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88:907–919.
18. Ruppert, D.; Wand, M.; Carroll, RJ. *Semiparametric Regression*. Cambridge University Press; 2003.
19. Kauermann G, Claeskens G, Opsomer JD. Bootstrapping for penalized spline regression. *Journal of Computational and Graphical Statistics*. 2009; 18:126–146.
20. Härdle W, Marron JS. Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*. 1991; 19:778–796.
21. Wang XF, Fan Z, Wang B. Estimating smooth distribution function in the presence of heteroscedastic measurement errors. *Computational Statistics and Data Analysis*. 2010; 54:25–36. [PubMed: 20160998]
22. Stefanski LA, Bay JM. Simulation extrapolation deconvolution of finite population cumulative distribution function estimators. *Biometrika*. 1996; 83:407–417.
23. Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, C. *Measurement error in nonlinear models: A modern perspective*. Chapman Hall; New York: 2006.
24. Drummond, MF.; Sculpher, MJ.; Torrance, GW.; O'Brien, BJ.; Stoddart, GL. *Methods for the economic evaluation of health care programmes*. 3rd edition. Oxford University Press; 2005. Cost-utility analysis..
25. Hall, P. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag; New York: 1992.
26. Wright J, Agodoa L, Greene T, Appel L, Bakris G, Charleston J, Cheek D, Douglas-Baltimore J, Gassman J, Glasscock R, Hebert L, Jamerson K, Lewis J, Middleton J, Phillips R, Rostand S, Toto R. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: Results from the AASK Trial. *Journal of the American Medical Association*. 2002; 288:2421–2431. [PubMed: 12435255]
27. Stevens LA, Schmid C, Greene T, Li L, Beck G, Joffe M, Froissart M, Kusek J, Zhang Y, Coresh J, Levey A. Factors other than glomerular filtration rate affect serum cystatin C levels. *Kidney International*. 2009; 75:652–660.
28. Eknoyan G, Hostetter T, Bakris GL, Hebert L, Levey AS, Parving HH, Stees MW, Toto R. Proteinuria and other markers of chronic kidney disease: a position statement of the national kidney foundation (NKF) and the national institute of diabetes and digestive and kidney diseases (NIDDK). *Am J Kidney Dis*. 2003; 42:617–22. [PubMed: 14520612]
29. Xie J, Liu CF. Adjusted KaplanMeier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Statistics in Medicine*. 2005; 24:3089–3110.
30. Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G. *Longitudinal Data Analysis*. (Chapman Hall/CRC); 2009.

Figure 1. Observed longitudinal (GFR) trajectories of 6 randomly selected subjects in the AASK study (section 4.1). The solid lines represent nonparametric smoothing curves of the data.

Figure 2.

Estimated multi-state probability curves by treatment groups in AASK study

Figure 3.
Estimated multi-state probability curves by proteinuria groups in AASK study

Figure 4.
Comparison of the estimated curves by methods I and II for $\sigma_e = 4$ and 8 ($n = 200$)

Figure 5.
Power curves of the bootstrap and log-rank tests for two-group comparison

Table 1

Simulation results: estimation accuracy of the multi-state probabilities

State	Method	p_k (30)	p_k (60)	p_k (90)	p_k (120)	IMSE
$n = 50$						
1	True	.000	.041	.218	.462	
	Method I	.000 (.002)	.041 (.029)	.213 (.058)	.458 (.081)	.217
2	True	.060	.189	.303	.295	
	Method I	.064 (.034)	.196 (.051)	.305 (.068)	.291 (.078)	.370
	Method II	.062 (.034)	.192 (.055)	.305 (.069)	.295 (.078)	.361
	Raw	.068 (.038)	.198 (.037)	.307 (.069)	.298 (.078)	
3	True	.691	.613	.391	.191	
	Method I	.692 (.063)	.611 (.068)	.391 (.074)	.194 (.070)	.574
	Method II	.693 (.062)	.613 (.068)	.392 (.072)	.192 (.070)	.554
	Raw	.673 (.065)	.600 (.071)	.387 (.072)	.189 (.072)	
4	True	.249	.157	.089	.052	
	Method I	.243 (.060)	.150 (.048)	.086 (.040)	.054 (.038)	.325
	Method II	.245 (.058)	.153 (.049)	.085 (.040)	.051 (.037)	.312
	Raw	.259 (.061)	.164 (.053)	.089 (.041)	.050 (.039)	
$n = 200$						
1	True	.000	.041	.218	.462	
	Method I	.000 (.001)	.040 (.015)	.219 (.031)	.461 (.038)	.053
2	True	.060	.189	.303	.295	
	Method I	.064 (.019)	.200 (.029)	.302 (.036)	.289 (.040)	.102
	Method II	.061 (.018)	.196 (.029)	.303 (.035)	.294 (.040)	.098
	Raw	.069 (.020)	.202 (.029)	.305 (.036)	.296 (.040)	
3	True	.691	.613	.391	.191	
	Method I	.692 (.032)	.607 (.035)	.389 (.036)	.191 (.034)	.148
	Method II	.692 (.031)	.609 (.035)	.389 (.036)	.190 (.034)	.141
	Raw	.674 (.034)	.599 (.037)	.383 (.036)	.187 (.034)	
4	True	.249	.157	.089	.052	
	Method I	.244 (.030)	.152 (.026)	.090 (.022)	.056 (.021)	.092
	Method II	.245 (.030)	.154 (.026)	.089 (.022)	.054 (.020)	.090
	Raw	.257 (.032)	.163 (.028)	.092 (.022)	.054 (.021)	