Varying coefficient subdistribution regression for left-truncated semi-competing risks data

Ruosha Li, University of Pittsburgh
Limin Peng, Emory University

Journal Title: Journal of Multivariate Analysis
Volume: Volume 131
Publisher: Elsevier | 2014-10-01, Pages 65-78
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1016/j.jmva.2014.06.005
Permanent URL: https://pid.emory.edu/ark:/25593/v2097

Final published version: http://dx.doi.org/10.1016/j.jmva.2014.06.005

Copyright information:
© 2014 Elsevier Inc.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed December 5, 2022 6:40 AM EST
Varying coefficient subdistribution regression for left-truncated semi-competing risks data

Ruoshia Li and Limin Peng

1Department of Biostatistics, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, U.S.A
2Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, U.S.A.

Abstract

Semi-competing risks data frequently arise in biomedical studies when time to a disease landmark event is subject to dependent censoring by death, the observation of which however is not precluded by the occurrence of the landmark event. In observational studies, the analysis of such data can be further complicated by left truncation. In this work, we study a varying co-efficient subdistribution regression model for left-truncated semi-competing risks data. Our method appropriately accounts for the specifical truncation and censoring features of the data, and moreover has the flexibility to accommodate potentially varying covariate effects. The proposed method can be easily implemented and the resulting estimators are shown to have nice asymptotic properties. We also present inference, such as Kolmogorov-Smirnov type and Cramér Von-Mises type hypothesis testing procedures for the covariate effects. Simulation studies and an application to the Denmark diabetes registry demonstrate good finite-sample performance and practical utility of the proposed method.

Keywords
Cumulative incidence; Left truncation; Hypothesis testing; Observational studies; Registry data analysis; Time-varying coefficient

1. Introduction

In biomedical studies, semi-competing risks data (Fine et al., 2001) arise, for example, when time to a nonterminating landmark event of disease may be dependently censored by time to death but not vice versa. The data structure has attracted much research interest, due to its frequent occurrence in practical situations (Jiang et al. 2003; Peng and Fine 2006; Ding et al. 2009; among others). One example of the semi-competing risks data is the Denmark diabetes registry study (Andersen et al., 1993; Kofoed-Enevoldsen et al., 1987), a prospective cohort study that included 2727 type I diabetic patients referred to the Steno Memorial Hospital in Greater Copenhagen between 1931 and 1988. An intermediate
endpoint of interest is the development of diabetic nephropathy (DN), a syndrome indicating kidney failure. Since death may preclude the observation of DN but remains observable after DN onset, time to DN, say $T_1$, and time to death, say $T_2$, with diabetes diagnosis set as time origin, form a semi-competing risks structure. An important complication in this registry study, like in many other prevalence studies, is left truncation, which occurred because patients can be observed only if they were alive at the study enrollment.

With semi-competing risks data, the cumulative incidence function (i.e., subdistribution function) for $T_1$, $F_1(t) = \Pr(T_1 \leq t, T_1 < T_2)$, has been advocated to characterize the distribution of the nonterminating event time, $T_1$. This crude quantity depicts the progression of disease while accounting for the presence of death, and moreover has the virtue of being nonparametrically identifiable (Prentice et al., 1978). In the absence of left truncation, analysis of $F_1(t)$ with semi-competing risks data can follow the approaches developed for competing risks data. For example, in the one-sample case, one may estimate $F_1(t)$ by adopting the methods in Andersen et al. (1993). Under regression settings, several approaches have been developed to study the covariate effects on the cumulative incidence function. For example, a popular method is the proportional subdistribution hazards model proposed by Fine and Gray (1999), which essentially amounts to a transformation model for $F_1(t|Z) = \Pr(T_1 \leq t, T_1 < T_2|Z)$ in the form of

$$F_1(t|Z)=1-\exp\{-\Lambda_{10}(t)\exp(Z^T\gamma)\}. \quad (1)$$

Here $\Lambda_{10}(t)$ is the baseline cumulative subdistribution hazard, $Z = (Z_1, Z_2, \ldots, Z_p)^T$ is the $p \times 1$ covariate vector, and $\gamma$ is the $p \times 1$ regression coefficient. Under model (1), covariates are assumed to have linear effects on $F_1(t|Z)$ after a complementary log-log transformation.

Several other types of regression modeling of $F_1(t|Z)$ were studied by Fine (2001) and Klein and Andersen (2005), among others.

The data scenario focused in this paper is the semi-competing risks data subject to left truncation to the terminating event, as occurred in the Denmark diabetes registry example. While some efforts have been made to address the left truncation issue for subdistribution regression with competing risks data (Geskus, 2011; Zhang et al., 2011; Shen, 2012), naively adapting these methods to the semi-competing risks setting in the presence of left truncation would incur information loss from coercing semi-competing risks data into competing risks data, even in the crude analysis. For example, in the Denmark diabetes registry study, such “coercion” would require exclusion of subjects who had developed DN before the registry entry. As shown by Peng and Fine (2007), such artificial truncation can incur considerable efficiency loss. In addition, Jiang et al. (2005) have studied semi-competing risks data with the complication of left-truncation. They adopted the copula assumption to study the marginal quantity $P(T_1 \leq t|Z)$, which differs from the $F_1(t|Z)$ of interest here.

It is also worth noting that most of the existing subdistribution regression methods constrain all covariate effects to be constant. This assumption may be too restrictive in many practical situations. For example, the efficacy of a treatment may fade over time due to drug
resistance, or the gender difference may be unnoticeable at the beginning but gradually manifest itself over time. For cohort studies with long follow-up periods, such as the Denmark diabetes registry study, the validity of the constant effect assumption may be especially questionable. By these considerations, it is often of practical appeal to consider varying-coefficient regression models that can allow for changing effect patterns, thereby enabling a more comprehensive view of covariate effects, and perhaps, more accurate predictions.

Two modeling approaches have been studied for accommodating varying covariate effects on the subdistribution function of interest. One is the direct binomial regression modeling of cumulative incidence probability (Scheike et al., 2008), which directly links covariates with the cumulative incidence function while allowing for regression coefficients to vary over time. This modeling strategy shares similar spirit of the additive complementary log-log survival model studied for randomly right censored data (Peng and Huang, 2007) but more flexibly accommodate the presence of competing risks. Of note, the direct binomial regression model of subdistribution offers a strict extension of Fine and Gray (1999)’s proportional subdistribution hazard model. More recently, Peng and Fine (2009) proposed competing risks quantile regression which formulates covariate effects on a range of conditional quantiles of the cumulative incidence function. The regression coefficients are permitted to change with different quantile levels, and thus renders a less restricted model compared to a competing risks accelerated failure time model. The method of Peng and Fine (2009) was extended by Li and Peng (2011) to handle left truncation to semi-competing risks data.

A general goal of this paper is to develop a new regression approach for left-truncated semi-competing risks data, which appropriately accounts for the specific censure and truncation features of this data type, and also has the flexibility to accommodate non-constant effects of covariates. In this work, we opt for the direct binomial regression modeling of subdistribution, which may lead to the following varying-coefficient subdistribution regression model,

$$F_1(t|Z) = g(Z^T \beta_0(t)),$$  \hspace{1cm} (2)

where $g(\cdot)$ is a monotone link function, $Z = (1, Z_1^T)^T$, and $\beta_0(t)$ is a $(p+1) \times 1$ vector of time-varying coefficients. Compared to competing risks quantile regression modeling in Peng and Fine (2009) and Li and Peng (2011), model (2) provides a more direct approach to assessing the covariate effects on and predicting cumulative incidence probabilities, and thus may be preferred in scenarios where the scientific interest centers on the cumulative incidence function itself. The method developed by Scheike et al. (2008) for model (2) purely focuses on standard competing risks data. How to extend their method to handle left truncation, particularly in the semi-competing risks setting, does not seem straightforward. In our simulation studies, we observe that naively applying Scheike et al. (2008)’s method to left-truncated semi-competing risks data can incur large estimation bias.
In the rest of the paper, we propose estimation and inference procedures for the varying-coefficient subdistribution regression model (2), tailored to left-truncated semi-competing risks data. We construct monotone estimating equations that can be readily solved by existing statistical software. We establish the asymptotic properties of the proposed estimator including uniform consistency and weak convergence. The influence functions of the proposed estimators are also derived, and greatly facilitate the inference, such as Kolmogorov-Smirnov type and Cramér Von-Mises type hypothesis testing procedures for the covariate effects. Via extensive simulations, we show that the proposed method performs well with various truncation rates and link functions. An application to the Denmark diabetes registry data demonstrate the practical utility of the proposed methods.

2. Estimation and Inference Procedures

2.1. Data and Model

We begin with a formal introduction of data notation and discussions of the adopted model (2). Let $T_1, T_2,$ and $C$ denote, respectively, time to the nonterminating event, time to the terminating event, and time to administrative censoring. In the absence of left truncation, semi-competing risks data comprises $(X, Y, \delta, \eta, Z)$, where $X = \min(T_1, T_2, C)$, $Y = \min(T_2, C)$, $\delta = I(T_1 \leq Y)$, $\eta = I(T_2 \leq C)$, and $Z = (1, Z_1, \ldots, Z_p)^T$ is the $(p + 1) \times 1$ covariate vector.

When the terminating event is subject to left truncation, data is observable only when $L < Y$, with $L$ representing the truncation time. Therefore, the observed data becomes $(L^*_1, X^*_1, Y^*_1, \delta^*_1, \eta^*_1, Z^*_1)_{i=1}^n$, which are identically and independently distributed and follow the conditional distribution of $(L, X, Y, \delta, \eta, Z)$ given $L < Y$.

For the truncation and censoring schemes, we assume that the censoring time $C$ is always greater than the truncation time $L$. This assumption has been commonly adopted in literature (Wang 1991; Asgharian et al. 2002; Oakes 2008; among others), and is natural in prospective cohort studies where subjects are followed from study entry until the terminating event or censoring. We also assume that $(L, C)$ is independent of $(T_1, T_2, Z)$. Some relaxation of this assumption is discussed in Section 5.

We consider the varying-coefficient subdistribution regression model (2), which offers straightforward coefficient interpretations and great model flexibility. The first component of $\beta_0(t)$ corresponds to the $g^{-1}$ transformed baseline cumulative incidence when $Z = (1, 0)^T$, and the remaining components correspond to covariate effects on $F_1(t|Z)$, which are allowed to change with $t$. When $g(x) = 1 - \exp\{-\exp(x)\}$, model (2) is an extension of the proportional subdistribution hazards model (Fine and Gray, 1999), which corresponds to the situation where all non-intercept coefficients are constant across $t$. With this link function, model (2) assumes that covariate effects are multiplicative on $\Lambda_1(t|Z) \equiv -\log\{1 - F_1(t|Z)\}$, the cumulative subdistribution hazard at time $t$. When $g(x) = 1 - \exp(-x)$, covariates influence $\Lambda_1(t|Z)$ in an additive manner. With all non-intercept components of $\beta_0(t)$ being linear functions of $t$, model (2) becomes an additive subdistribution hazard model, a counterpart of the additive hazards model (Lin and Ying, 1994) in the competing risks setting. Another useful link function in practice may be $g(x) = \log\{x/(1-x)\}$, which would...
allow us to investigate the odds of observing the nonterminating event in the presence of death before time $t$.

2.2. Estimating Equation

The key idea for estimating $\beta_0(t)$ is to utilize the technique of inverse probability of censoring weighting (Robins et al., 1995) to account for the sampling bias from left truncation. This leads to the equality,

$$E \left\{ \frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{G(Y_i^*)/\alpha(Z_i^*)} \right| \frac{Z_i^*}{Z_i^*} = F_1(t|Z_i^*), \quad (3)$$

where $G(y) = P(L < y \leq C)$, and $\alpha(z) = P(L < Y|Z = z)$. Note, $\alpha(z)$ equals $P(L < T_2|Z = z)$ under the assumption that $L$ is always less than $C$, and represents the conditional probability that a subject is not truncated.

Let $W(y, z) = G(y)/\alpha(z)$. Suppose that we can obtain a consistent estimator of $W(Y_i^*, Z_i^*)$, denoted by $\hat{W}(Y_i^*, Z_i^*)$. Equation (3) motivates us to estimate $\beta_0(t)$ by solving the following estimating equation:

$$\frac{1}{n} \sum_{i=1}^n Z_i^* \left[ \frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{\hat{W}(Y_i^*, Z_i^*)} g(Z_i^T \beta(t)) \right] = 0. \quad (4)$$

Note, equation (4) is monotone in $\beta(t)$ (Fygenson and Ritov, 1994). This feature entails not only the global consistency of the resulting estimator in theory but also easy and stable computation. For a fixed $t$, the solution to equation (4) can be easily obtained via Newton-type root-finding methods, such as the `optim` function in R, or the `multiroot` function in R package `rootSolve`. To estimate $\beta_0(t)$ on a continuum of $t$, it is sufficient to solve (4) only at the observed event times of $T_1$. The resulting estimator, $\hat{\beta}(t)$, is a step function of $t$ that jumps only at finite time points.

Remark—One may consider substituting the first $Z_i^*$ in (4) with $Z_i^* g(Z_i^T \beta(t))$. However, this would lead to an estimating equation, which is no longer monotone. In some unreported simulation studies, adopting equation (4) was observed to outperform such an alternative.

Next, we discuss the estimation of $W(y, z)$. First, we estimate $\alpha(z)$ using the equality that $\alpha(z) = \int_{0}^{\infty} S_2(l|z) dF_L(l)$, where $S_2(l|z) = \Pr(T_2 > l|Z = z)$, $F_L(l) = \Pr(L \leq l)$. Following the lines of Li and Peng (2011), we note that $T_2$ is subject only to independent left truncation and right censoring, and thus standard regression models for left-truncated univariate survival data can be adopted to estimate $S_2(\cdot|z)$. In this work, we use the classical Cox proportional hazards model and denote the resulting estimator by $\hat{S}_2(\cdot|z)$. Alternative estimators of $S_2(\cdot|z)$ may also be adopted, provided that they render satisfactory approximations to the survival function in the presence of left-truncation. Therefore, a consistent estimator for $\alpha(z)$ can be constructed as

\[ J Multivar Anal. Author manuscript; available in PMC 2015 January 01. \]
\[ \hat{\alpha}(z) = \int_0^\nu \hat{S}_2(l|z) d\hat{F}_L(l), \]

where \( \hat{F}_L(l) = \prod_{s > l} \left\{ 1 - dF^*_L(s)/R_n(s) \right\} \) with \( F^*_L(s) = n^{-1} \sum_{i=1}^n I(L_i^* \leq s), R_n(s) = \frac{1}{n} \sum_{i=1}^n I(L_i^* \leq Y_i^*) \) The upper limit of integration, \( \nu \), is a constant satisfying regularity condition \( C_3 \) in Section 2.3. Here, \( \hat{S}_2(\cdot|Z) \) is step functions that jumps only at the observed event times and truncation times respectively. As a result, \( \hat{S}_2(\cdot|Z) \) stands for a Lebesgue-Stieljes integration, which can be calculated as a finite summation without involving complex numerical integrations. In addition, \( G(y) \) can be consistently estimated by

\[ \hat{G}(y) = \frac{1}{n} \sum_{i=1}^n \frac{I(L_i^* \leq y \leq Y_i^*) \hat{\alpha}(Z_i^*)}{\hat{S}_2(y-|Z_i^*)}. \]

Consequently, we obtain an estimator of \( W(y, z) \), \( \hat{W}(y, z) = \hat{G}(y)/\hat{\alpha}(z) \), which can be shown to be uniformly consistent in \( y \) and \( z \).

### 2.3. Asymptotic Results

For a nonnegative random variable \( T \), let \( a_T = \inf \{ t \geq 0 : P(T \leq t) > 0 \} \), \( b_T = \sup \{ t \geq 0 : P(T \leq t) < 1 \} \). For a square matrix \( A \), we use \( \text{eigmin}(A) \) to denote the minimum eigenvalue of \( A \).

Let \( b \otimes 2 \) denote \( bb^T \) for a vector \( b \). The regularity conditions are specified as follows.

**C1** The covariate space \( Z \) is bounded, i.e., \( \sup \|Z\| < \infty \).

**C2** The regularity conditions as in Anderson & Gill (1982), with the definition of at risk process modified to \( R_i(t) = I(L_i^* < t \leq Y_i^*) \).

**C3** (i) \( a_L \leq a_Y, b_L \leq b_Y \); (ii) there exists a positive constant \( \nu > 0 \), such that \( P(C = \nu) > 0 \) and \( P(C > \nu) = 0 \). (iii) \( \inf_{y \in \mathbb{R}} P(T_2 \geq 1|z) > 0 \) and \( \inf_{y \in \mathbb{R}^+} P(C > \nu) > 0 \).

**C4** (i) Let \( \mathcal{B}(\rho) = \{ \mathbf{b} \in \mathbb{R}^{p+1} : \inf_{t \in [l,u]} \| \mathbf{b} - \mathbf{b}_0(t) \| < \rho \} \), where \( [l, u] \subset (0, \infty) \) represents an interval of \( t \). There exist \( \rho_0 > 0 \) and \( k > 0 \), such that \( \inf_{\mathbf{b} \in \mathcal{B}(\rho_0)} d(\mathbf{b}) \) is Lipschitz continuous for \( t \in [l, u] \).

The assumed regularity conditions are rather mild and are expected to be met in many practical situations. More specifically, **C1** assumes the boundedness of covariates. We assume bounded covariates in **C1** to simplify theoretical arguments. This constraint on covariates can be relaxed to \( E(Z^2) < \infty \) to accommodate unbounded covariates. By **C2**, we would retain the uniform consistency and weak convergence of \( \hat{S}_2(d\mathbf{f}) \). Condition **C3** (i) poses mild assumptions on the observation range of \( L \) and \( Y \) to ensure the identifiability of \( \alpha(z) \) (He and Yang, 1998), while **C3** (ii)–(iii) simplify the asymptotic arguments, and are often met in clinical settings after some minor adjustment. For example, one may adopt **C** =...
min(C, C_U) to ensure that C^3 (ii) is always satisfied, where C_U is a constant chosen to be only slightly less than b_C. This technique is often adopted for censored data to avoid the instability of the estimators in tail regions (Peng and Fine, 2009). It is easy to see that C^4 (i) is satisfied when E(Z^* \otimes 2) is positive definite and g(\cdot) is strictly monotone. Condition C^4 (ii) requires the smoothness of \hat{\beta}_0(t), which is also reasonable in practice.

We establish the asymptotic properties of \hat{\beta}, which are stated in the following theorems.

**Theorem 1.** Under Conditions C1 – C4, \[ \sup_{t \in [l, u]} \| \hat{\beta}(t) - \beta_0(t) \| \xrightarrow{p} 0. \]

**Theorem 2.** Under Conditions C1–C4, \[ n^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} \] converges weakly to a mean zero Gaussian process with covariance matrix

\[ \sum(t', t) = d \{ \beta_0(t') \}^{-1} E[\chi(t') \chi(t)'] d \{ \beta_0(t) \}^{-T} \]

for \( t \in [l, u] \), where the definition of \( \chi(t) \) is provided in Appendix.

The detailed proofs of Theorems 1–2 are relegated to Appendix.

### 2.4. Inferences

In the proof of Theorem 2, we show that \[ n^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} \] is asymptotically equivalent to \[ n^{-1/2} \sum_{i=1}^n \xi_i(t), \] where \( \xi_i(t) \) is defined in Appendix and are i.i.d. with expectation 0. Let \( D_n \{ \beta(t) \} \equiv -n^{-1} \sum_{i=1}^n Z_i^* \otimes g \{ Z_i^T \beta(t) \} \). A consistent estimator of the asymptotic covariance matrix \( \Sigma(t, t) \) can be constructed as \[ \hat{\Sigma}(t, t) = n^{-1} \sum_{i=1}^n \hat{\xi}_i(t) \otimes 2, \] where

\[ \hat{\xi}_i(t) = -D_n \{ \hat{\beta}(t) \}^{-1} \left( Z_i \left[ I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1) g \{ Z_i^T \hat{\beta}(t) \} \right] + \frac{1}{n} \sum_{j=1}^n Z_j I(X_j^* \leq t, \delta_j^* = 1, \eta_j^* = 1) \hat{\omega}_i(Y_j^*, Z_j^*) \right). \]

Here \( \hat{\omega}_i(y, z) \) is a consistent estimator of \( w_i(y, z) \), which represents the influence function of \( \hat{W}^{-1}(y, z) \). The detailed form of \( \hat{\omega}_i(y, z) \) is provided in Supporting Information B. Based on the proposed covariance estimation, for a given time point \( t \), one can readily construct a Wald-type pointwise confidence interval for \( \beta_0(t) \).

In practice, it is often of interest to assess whether a certain covariate, say \( Z_j \), has any effect on the cumulative incidence function. To this end, the null hypothesis may be formulated as \( H_{10}^{j} : \beta_0^{(j)}(t) = 0 \) \( t \in [l, u] \), where the superscript \( (j) \) denotes the \( j + 1 \)th component of a vector.

It is important to note that \( H_{10}^{j} \) draws inference on \( \beta_0^{(j)}(t) \) simultaneously for all \( t \in [l, u] \). To test \( H_{10}^{j} \), one straightforward approach is to construct a Wald-type test based on \( \hat{\beta}_0^{(j)}(t) \) at a single or multiple prespecified time points. However, this would involve a subjective choice of time points and can miss differences from null. As a natural alternative, one may construct the test using the integral of \( \hat{\beta}_0^{(j)}(t) \) over \([l, u]\). The detailed procedure can follow...
the lines of Peng and Fine (2009). One potential issue with the resulting integral test may be lack of power when \( \hat{\beta}_0^{(j)}(t) \) changes its sign in the time interval \([l, u]\). In this work, we additionally consider two “omnibus” test statistics:

\[
\mathcal{R}_{1,\text{sup}}^j = \sup_{t \in [l, u]} \frac{n^{1/2} \hat{\beta}_0^{(j)}(t)}{\sqrt{\sum (t, t)^{(j,j)}}}, \quad \mathcal{R}_{1,\text{inte}}^j = \int_{t_l}^{t_u} \left( \frac{n^{1/2} \hat{\beta}_0^{(j)}(t)}{\sqrt{\sum (t, t)^{(j,j)}}} \right)^2 dt, \tag{5}
\]

where \( \hat{\Sigma}(t, t)^{(j,j)} \) denotes the \( j + 1 \)th diagonal element of \( \hat{\Sigma}(t, t) \). Note that \( \mathcal{R}_{1,\text{sup}}^j \) is a Kolmogorov-Smirnov (K-S) type test statistics based on the supreme absolute difference, and \( \mathcal{R}_{1,\text{inte}}^j \) is a Cramér Von-Mises (C-V) type test statistics that integrates the squared difference over time. Both \( \mathcal{R}_{1,\text{sup}}^j \) and \( \mathcal{R}_{1,\text{inte}}^j \) are sensitive to all types of departures from the null hypothesis. To obtain the \( p \)-values of the proposed test statistics, we adopt a resampling procedure following the idea of Lin et al. (1993). Specifically, one may generate \( B \) independent sets of \( \{ \hat{h}_b^{(j)} \}_{b=1}^{n} \) for a large number \( B \), where \( \{ \hat{h}_b^{(j)} \}_{b=1}^{n} \) are independent random variables from a standard normal distribution. Define \( \hat{\beta}^{(j)} = \hat{\beta}(t) + n^{-1} \sum_{i=1}^{n} \xi_i(t) \hat{h}_b^{(j)} \), it can be shown that the conditional distribution of \( n^{1/2} \{ \hat{\beta}^{(j)}(t) - \beta_0^{(j)}(t) \} \) given the observed data has the same limit as the unconditional distribution of \( n^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} \). Hence, one may approximate the asymptotic distribution of \( \mathcal{R}_{1,\text{sup}}^j \) and \( \mathcal{R}_{1,\text{inte}}^j \) under the null hypothesis \( H_{0j} \) by the empirical distribution of \( B \) resamples obtained as \( \mathcal{R}_{1,\text{sup}}^j \) and \( \mathcal{R}_{1,\text{inte}}^j \) with \( n^{1/2} \hat{\beta}(t) \) replaced by \( n^{1/2} \hat{\beta}_0(t) \). The \( p \)-values can then be computed accordingly.

One may also be interested in evaluating whether the effect of \( Z_j \) on \( F_1(t|Z) \) takes some specified parametric form. This often offers a natural means for checking the goodness-of-fit of certain classical models. For example, when \( g(x) = 1 - \exp\{-\exp(-x)\} \), rejecting the null hypothesis \( H_{0j} : \hat{\beta}_0^{(j)}(t) = c_0 \) for some \( j \) in \( \{1, \ldots, p\} \), where \( c_0 \) is an unknown constant, may suggest the inadequacy of a proportional subdistribution hazard model. With \( g(x) = 1 - \exp(-x) \), an additive subdistribution hazard model holds only when \( \beta_0^{(j)}(t) = 0 \) is a linear function of \( t \) for each \( j \) in \( \{1, \ldots, p\} \). Thus, it may be of interest to consider the null hypothesis that takes the form, \( H_{0j} : \beta_0^{(j)}(t) = c_0 t \), where \( c_0 \) is an unspecified constant. For presentation simplicity, here we only discuss the hypothesis testing on the constancy and the linearity of a covariate effect under model (2), driven by the two special examples discussed above. Nevertheless, the basic idea can be adapted to evaluate other specified parametric forms for \( \beta_0^{(j)}(t) \).

To test the parametric form of a covariate effect, we still consider the K-S and the C-V types of test statistics, which may be given by
with \( \hat{\beta}(j)(t) \) being a reasonable estimate of the regression coefficient at time \( t \) under the restriction of \( H_0 \), and \( \Xi(t,t) \) being a consistent estimator of the variance of \( n^{1/2} (\hat{\beta} - \beta) \).

To test the constancy of \( \beta_0(j)(t) \) or the linearity of \( \beta_0(j)(t) \) in \( t \), we may adopt
\[
\hat{\beta}(j)(t) = \int_{u}^{t} \{ \hat{\beta}(j)(u) \} du \text{ or } \hat{\beta}(j)(t) = \beta(j)(l) + \{ \beta(j)(u) - \beta(j)(l) \} (t - l) / (u - l) \text{ respectively.}
\]
Resampling techniques can similarly be used to approximate the asymptotic distribution of the test statistics.

### 3. Simulations

Extensive simulation studies were conducted to evaluate the finite sample performances of the proposed estimators. We set \( Z = (1, Z_1, Z_2)^T \), where \( Z_1 \) follows a normal distribution \( N(1, 0.7^2) \) truncated between 0 and 2, and \( Z_2 \) follows Bernoulli(0.5) distribution. We generated bivariate survival data \((T_1, T_2)\) such that the underlying subdistribution equals
\[
F_1(t | Z) = g\{ \beta^{(1)}_0(t) + Z_1 \beta^{(1)}_1(t) + Z_2 \beta^{(2)}_0(t) \}
\]
for \( t \leq 2.5 \), where \( \beta^{(1)}_0(t) = g[ F_1 \{ t | Z = (1, 0, 0)^T \} ] \) with \( F_1 \{ t \} = 0.45 - 0.45 \exp(-1.25t) \) and the forms of \( \beta^{(1)}_0(t) \) and \( \beta^{(2)}_0(t) \) are provided in Table 1. Furthermore, \( T_2 \) follows a Cox proportional hazards model with regression coefficient vector, \((0.3, -0.4)^T\), and baseline hazard function, \( 4.5x^{-0.1} \). Under our data generation scheme, for a fixed \( Z \), \( T_1 \) is dependent censoring by \( T_2 \) with a probability of \( P(T_1 > T_2 | Z) = 1 - F_1(2.5 | Z) \). We considered both the additive hazard model and the multiplicative hazard model by setting \( g(x) \) as \( g_1(x) = 1 - \exp(-x) \) and \( g_2(x) = 1 - \exp(-\exp(x)) \) respectively. The truncation time \( L = r_L \times L_0 \), where \( r_L \) is Bernoulli with probability \( p_L \), and \( L_0 \) is a positive random variable that is independent of \( r_L \). Such a truncation scenario mimics the Denmark diabetic registry data, where the distribution of \( L \) has a point mass at 0. We generate the censoring time as \( C = \min\{L + Unif(7, 12), 9.5\} \) such that \( C \) is always greater than \( L \). We chose different combinations of \( \{g, \beta^{(1)}_0(t), \beta^{(2)}_0(t), p_L, L_0\} \), which led to the four simulation setups in Table 1.

Table 1 shows that we examined scenarios with various combinations of truncation rates and event rates. For example, the truncation rates in setups S1A and S1M equal 20%, reflecting a rather low level of truncation, while setups S2A and S2M involve 40% truncation, corresponding to situations with moderate to heavy truncation. We also considered both time-varying and time-invariant coefficients for either continuous covariate or binary covariate. The non-constant coefficient functions in our simulation setups are increasing
with \( t \). This mimics the situation observed in the Denmark diabetes example, where covariates seem to manifest bigger effects at later times.

Under each setup, we implemented the proposed methods on 2000 simulated datasets with sample size 200 and 400. All numerical calculations were implemented in \( R \). We adopted the `coxph()` function in \( R \) package `survival` to obtain \( \hat{S}(\cdot|Z) \), and used the `mirtroot()` function in \( R \) package `rootSolve` to solve the proposed estimating equation (4). According to our numerical experiences, the implementation of our methods based on these \( R \) functions is stable and fast. In Tables 2–3, we reported the empirical bias (EBias) of \( \{ F_{10}(t), \beta^{(0)}(t), \beta^{(1)}(t), \beta^{(2)}(t) \} \) at \( t = 0.3, 0.5, 0.6, 1.1, 1.5 \), where \( F_{10}(t) \equiv g(\beta^{(0)}(t)) \) denotes the estimated baseline cumulative incidence rates. We also presented the empirical standard errors (ESD) and the average of estimated standard errors (ASD) of \( \{ \beta^{(0)}(t), \beta^{(1)}(t), \beta^{(2)}(t) \} \), as well as the empirical coverage probabilities of the 95% Wald-type confidence intervals (C95). Tables 2–3 suggests that the proposed \( \hat{\beta}(t) \) is virtually un-biased under all setups. The relatively larger bias in \( \beta^{(0)}(t) \) under S1M and S2M may be attributed to the fact that \( \log[-\log(1 - F_1(t|Z))] \) is of larger magnitude and goes to \(-\infty\) when \( t \) goes to 0. Nevertheless, the estimation of \( F_1(t|Z = (1, 0, 0)^T) \) seems rather accurate in all cases, somewhat confirming our explanation for the bias in \( \beta^{(0)}(t) \). As expected, the standard errors of the proposed estimator get larger when the truncation and censoring rates increase. Under all setups, the proposed standard error estimates closely match their empirical counterparts, and the Wald-type confidence intervals based on normal approximation provides satisfactory empirical coverages. Moreover, it is clear that the performances of the proposed estimator improve with the sample size.

For comparison purpose, we also implemented Scheike et al. (2008)'s method via the `comp.risk()` function in \( R \) package `timereg` without handling the left truncation to \( T_2 \). The empirical biases of the corresponding estimators, \( \{ F_{R,10}(t), \beta_{R}^{(0)}(t), \beta_{R}^{(1)}(t), \beta_{R}^{(2)}(t) \} \), are shown in the first four columns of Tables 2–3. These results show that ignoring the left truncation leads to biased estimation under all simulation setups, and the magnitudes of bias grow with the truncation rates. We also note that \( F_{R,10}(t) \) tends to underestimate the baseline cumulative incidence.

As suggested by one referee, we further compared the root mean squared errors (RMSE) between the proposed estimator and the naive estimator from applying Scheike et al. (2008)'s method (blue) under all simulation setups; see Figure 1. It is observed that in some cases, the naive estimator has smaller RMSE than the proposed estimator, in spite of its substantially larger bias. Our investigation on this phenomenon indicates that the smaller RMSE of the naive estimator is caused by its surprisingly small variances. These small variances of the naive estimator, coupled with its large bias, would lead to severely biased inference. To confirm our conjecture, we plot in Figure 2 the empirical coverage rates of 95% confidence intervals derived from the proposed method and those based on the naive method under all four setups. It is clear that the proposed method always yields coverage rates close to the nominal level, 95%. However the naive method generally produces markedly lower coverage probabilities. The results in Figure 2 provide a good demonstration of the serious issue with naively ignoring left truncation in registry data analysis.
We also evaluated the performance of the proposed Kolmogorov-Smirnov type test statistics, $\mathcal{F}_{i,\text{sup}}^{(j)}$ and the Cramér Von-Mises type test statistics, $\mathcal{F}_{i,\text{inte}}^{(j)} (i=1, 2, j=1, 2)$. We considered $H_{102; \alpha_0}^{(j)}(t) = c_0 t$ for setups S1A and S2A, and $H_{102; \alpha_0}^{(j)}(t) = c_0$ for setups S1M and S2M, where $c_0$ is an unspecified constant, $j = 1, 2$. For each simulated dataset, we performed the tests based on 2500 resamples. The significance levels for all tests were set as 0.05. In Table 4, we presented the empirical rejection rates (ERR) of these tests. We observe that the proposed tests have empirical sizes that are close to the nominal level, and achieve reasonably good power when the null hypotheses are violated. In addition, the Kolmogorov-Smirnov type tests statistics and the Cramér Von-Mises type tests appear to have quite similar performance.

In summary, our simulation studies demonstrate that the proposed methods work well with realistic sample sizes under different configurations of link functions, censoring and truncation rates. The proposed inference renders accurate variance estimation, confidence intervals with satisfactory coverage probabilities, and hypothesis tests with right size and good power. Furthermore, additional simulation results in Appendix A.1 illustrate the satisfactory performances of the proposed methods under smaller sample size of $n = 100$.

4. Analysis of Denmark Diabetes Registry Data

We applied the proposed methods to the Denmark diabetes registry study. We focused on the cohort of subjects that were born before 1940 and had diabetes onset age between 15 and 31. In this cohort, the median follow-up time is 27 years. Among the 858 subjects in this cohort, 249 developed DN and 343 died during the follow-up period. Here, we are interested in the cumulative incidence of DN following diabetes diagnosis, after properly accounting for the left truncation to death due to the time-lag between diagnosis and study enrollment. We fit model (2) to this dataset with covariates, diabetes onset age ($Z_1$) and gender ($Z_2$), setting $g(x) = 1 - \exp(-x)$. In this setup, a negative regression coefficient would indicate that the corresponding covariate has a beneficial protecting effect against the development of DN.

Figure 3 presents the estimated regression coefficient $\hat{\beta}^{(j)}(t)$ for $j = 1, 2$ and $t \in [5, 40]$ (years), as well as the estimate of the baseline cumulative incidence $\hat{F}_{10}(t) = 1 - \exp\{-\hat{\beta}^{(0)}(t)\}$. Also presented are the naive estimates obtained by applying Scheike et al. (2008)’s method without handling left truncation. We observe from Figure 3 that the coefficient for age is always negative, and the confidence intervals mostly exclude 0. This result suggests that patients with older diabetes onset age tend to have a lower cumulative incidence of DN in the presence of death. The effect of diabetes onset age also seems to increase over time. Regarding the gender effect on DN, we observe that the difference between males and females is not significant at small $t$, while males tend to have higher cumulative incidence of DN after 20 (years) from the diagnosis of diabetes.

Of note, the naive method leads to quite different estimates for the baseline cumulative incidence estimate and the regression coefficients from those obtained by the proposed method. Such a difference is more apparent in Figure 4, where we plot the ratio of the naive
estimator to the proposed estimator. It is shown that ignoring left truncation can lead to over 10% underestimation of cumulative incidence of DN.

We also conducted formal hypothesis tests to evaluate the significance and the parametric forms of the covariate effects for $t \in [5, 40]$ (years) based on 5000 resamples. When testing the overall significance of the age effect, both $T_{1, \text{sup}}^1$ and $T_{1, \text{inte}}^1$ yield highly significant p-values that are less than 0.001. This confirms our observation from Figure 3 that the coefficient for age is quite apart from zero for most time points. For gender, the overall significance test based on $T_{1, \text{sup}}^1$ gives a p-value of 0.077, while that based on $T_{1, \text{inte}}^1$ yields a p-value of 0.026. This result provides some marginal evidence for the influence of gender on the DN progression. Regarding the tests on parametric forms, we first assessed the constancy of covariate effects. The p-values for the age effect based on $T_{2, \text{sup}}^2$ and $T_{2, \text{inte}}^2$ are both less than 0.001, and those for the gender effect are 0.006 and 0.01 respectively. These results are consistent with our observations from Figure 3 that the impact of diabetes onset age and gender on the DN cumulative incidence seems to change over time. Next, we checked on the linearity of covariate effects. For the age effect, the p-values from $T_{2, \text{sup}}^2$ and $T_{2, \text{inte}}^2$ are smaller than 0.001. For the gender effect, the corresponding p-values equal 0.024 and 0.058 respectively. The results from the linearity tests suggest that a simple additive subdistribution hazard model may not provide an adequate fit for the data, thereby illustrating the need of adopting the proposed varying coefficient subdistribution regression model.

In conclusion, our analysis suggests a significant beneficial effect of older diabetes onset age on DN progression, which is more pronounced for long-term DN incidence. Gender may have some influence the cumulative incidence rate of DN but its effect appears to emerges rather late, about 20 years since diabetes diagnosis.

5. Discussions

In this work, we propose a flexible semiparametric regression method tailored to left-truncated semi-competing risks data. The adopted varying coefficient subdistribution regression model permits the investigation of temporal covariate effects on the cumulative incidence function. The proposed method also appropriately handles the left truncation in the semi-competing risks setting without involving unnecessary artificial truncation.

When compared to Li and Peng (2011)’s work, the proposed method directly links covariates to the cumulative incidence probabilities, and therefore it may be preferred when scientific interest lies in estimating and predicting the cumulative incidence probabilities of the nonterminating event. The method developed by Li and Peng (2011) would be useful for characterizing covariate effects on the nonterminating event in the time scale, instead of the probability scale. The proposed regression procedure for left-truncated semi-competing risks data thus serves as a useful alternative to the method by Li and Peng (2011), in the same way that a Cox proportional hazards model or other linear transformation models complement the accelerated failure time model.
Left truncated semi-competing risks data also arise in many fields other than biomedicine, such as econometrics, actuarial science, and social science. For example, in a study of life insurance, time from policy initiation to policy withdrawal or surrender may be of interest but can be dependently censored by death. At the same time, time to death is subject to left truncation by time to study enrollment. This is because subjects who die before study enrollment are naturally excluded from the study sample. The proposed methodology will be useful in such non-biomedical applications as well.

The estimation presented in Section 3 requires that \((L, C)\) is independent of \(Z\). This assumption can be easily relaxed when the distribution of \((L, C)\) differs only by some discrete covariates. Specifically, one may estimate \(W(y, z)\) following the procedure in Section 3 separately within each covariate stratum. In more general cases, additional regression modeling of \((L, C)\) given \(Z\) may be adopted to capture the dependence between \((L, C)\) and \(Z\), thereby enabling appropriate estimation of \(W(y, z)\) and thus \(\beta_0(\cdot)\).

In this work, we study a fully functional model where all regression co-efficients are functions of time. The proposed method can be extended to a partially functional model, where some of the regression coefficients are restricted to be constant. The coefficients can be estimated via an iterative procedure similar to that in Scheike and Zhang (2007) and Qian and Peng (2010).

**Acknowledgments**

The authors are grateful to the editor and referees for many helpful comments. This research has been supported by National Science Foundation under Grant Number DMS-1007660 and National Institute of Health under Award Number R01HL 113548.

**References**

Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical Models Based on Counting Processes. Springer-Verlag Inc; 1993.


**APPENDIX**

**Appendix A. Additional Simulation Results**

Here, we provide some additional simulation results with sample size $n = 100$ in Table Appendix A.1 below. The performances, though slightly worse than those with $n = 200$ and $n = 400$, still appear satisfactory in terms of bias and coverage rates.
Appendix B. Proof of Theorem 1

We first need the following lemma.

**Lemma 1**

Under regularity condition C1–C5, the estimators \( \hat{\alpha}(z) \) and \( \hat{G}(y) \) are uniformly consistent for \( z \in \mathcal{Z} \) and \( y \in \mathcal{T} \), where \( \mathcal{T} \equiv \{ay^*, \nu\} \). Write \( W(y, z) = G(y)/\alpha(z) \) and \( \hat{W}(y, z) = \hat{G}(y)/\hat{\alpha}(z) \). 1/\( \hat{W}(y, z) \) is uniformly consistent in that

\[
\sup_{z \in \mathcal{Z}, y \in \mathcal{T}} \|1/\hat{W}(y, z) - 1/W(y, z)\| = o_p(1). \tag{B.1}
\]

Furthermore, we have that

\[
n^{1/2}\left\{1/\hat{W}(y, z) - 1/W(y, z)\right\} = n^{-1/2}\sum_{i=1}^{n} w_i(y, z) + o_p(n^{1/2}), \tag{B.2}
\]

where \( w_i(y, z) \) are i.i.d. influence functions and \( o_p(n^{1/2}) \) is a term that converges to 0 in probability uniformly for \( z \in \mathcal{Z} \) and \( y \in \mathcal{T} \).

The proof of Lemma 1 follows from the Web-based supplementary materials of Li and Peng (2011).

Now define

\[
S_n(b, t) = \frac{1}{n} \sum_{i=1}^{n} Z_i \left[ \frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{W(Y_i^*, Z_i^*)} - g\{ Z_i^{T} b \} \right],
\]

and

\[
S_n^G(b, t) = \frac{1}{n} \sum_{i=1}^{n} Z_i \left[ \frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{W(Y_i^*, Z_i^*)} - g\{ Z_i^{T} b \} \right].
\]

By the boundedness of \( Z_i^* \) and \( I(X_j^* \leq t, \delta_j^* = 1, \eta_j^* = 1) \), we can combine (B.1) to get

\[
\sup_{b, t} \left\| S_n(b, t) - S_n^G(b, t) \right\| = o_p(1). \tag{B.3}
\]

Next, we can show that the function class

\[
\mathcal{F}^G \equiv \{ Z_i \left[ I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)W^{-1}(Y_i^*, Z_i^*) - g( Z_i^{T} b ) \right]; \quad Z_i^* \in \mathcal{X}, t \in [l, u], b \in \mathbb{R}^{p+1} \}
\]
is Donsker thus Glivenko-Cantelli (Kosorok, 2008). This is because \( Z_1^* \) and \( W^{-1}(Y_1^*, Z_1^*) \) are both uniformly bounded, and that the class of indicator functions is Donsker. Let
\[
\mu(b, t) = E S^G_n(b, t) = E[ Z_1^* \{ F_1(t) | Z_1^* \} - g(Z_1^T b) ],
\]
an application of the Glivenko-Cantelli theorem thus gives
\[
\sup_{b, t} \| S^G_n(b, t) - \mu(b, t) \| = o_p(1). \tag{B.4}
\]
Now using the fact that \( S_n(\beta^0, t) = \mu(\beta_0(t), t) = 0 \), and
\[
\| \mu(\hat{\beta}(t), t) - \mu(\beta_0(t), t) \| \leq \| \mu(\hat{\beta}(t), t) - S^G_n(\hat{\beta}(t), t) \| + \| S^G_n(\hat{\beta}(t), t) - S_n(\hat{\beta}(t), t) \|,
\]
we can combine (B.3) and (B.4) to see that
\[
\sup_t \| \mu(\hat{\beta}(t), t) - \mu(\beta_0(t), t) \| = o_p(1). \tag{B.5}
\]
For any \( \zeta \in \mathbb{R} \) and \( u \in \mathbb{R}^{p+1} \), with \( u \otimes 2 = 1 \), it is easy to see that \( u^T \mu(\beta_0(t) + u \zeta, t) \) is decreasing in \( \zeta \). Using similar arguments as those in Peng and Fine (2009), we can combine condition \( C4 \) and the Cauchy-Schwarz inequality to see
\[
\inf_{b \in \mathbb{R}(\rho_0), t \in [l, u]} \| \mu(b, t) - \mu(\beta_0(t), t) \| \geq k \rho_0.
\]
When combined with (B.5), this inequality further indicates that the probability that \( \{ \beta \hat{\beta}, t \in [l, u] \} \subseteq B(\rho_0) \) goes to 1 when \( n \) goes to infinity. We can now use Taylor expansion and condition \( C4 \) to get
\[
\sup_{t \in [l, u]} \| \hat{\beta}(t) - \beta_0(t) \| \leq k^{-1} \sup_t \| \mu(\hat{\beta}(t), t) - \mu(\beta_0(t), t) \| = o_p(1). \tag{B.6}
\]
This completes the proof of Theorem 1.

**Appendix C. Proof of Theorem 2**

Combine (B.2) and some algebraic manipulations, we can write
\[
\sqrt{n} \{ S_n(b, t) - S^G_n(b, t) \} = n^{-1/2} \sum_{j=1}^n Z_j^* I(X_j^* \leq t, \delta_j^* = 1, \eta_j^* = 1) \{ 1/W(Y_j^*, Z_j^*) - 1/W(Y_j^*, Z_j^*) \}
\approx n^{-1/2} \sum_{j=1}^n Z_j^* I(X_j^* \leq t, \delta_j^* = 1, \eta_j^* = 1) \{ \frac{1}{n} \sum_{i=1}^n w_i(Y_j^*, Z_j^*) \}
= n^{-1/2} \sum_{j=1}^n \frac{1}{n} \sum_{i=1}^n Z_j^* I(X_j^* \leq t, \delta_j^* = 1, \eta_j^* = 1) w_i(Y_j^*, Z_j^*).
\]

*J Multivar Anal. Author manuscript; available in PMC 2015 January 01.*
Here and in the sequel, we use \( \approx \) to denote uniform asymptotic equivalence for \( t \in [l, u] \).

We can apply Glivenko-Cantelli theorem to \( \frac{1}{n} \sum_{j=1}^{n} Z^*_j I(X^*_j \leq t, \delta^*_j = 1, \eta^*_j = 1) w_i(Y^*_j, Z^*_j) \) to see

\[
\sqrt{n} \{ S_n(b, t) - S_n^G(b, t) \} = n^{-1/2} \sum_{i=1}^{n} \chi_{1,i}(t). \tag{C.1}
\]

Here \( \chi_{1,i}(t) = E[Z^*_j I(X^*_j \leq t, \delta^*_j = 1, \eta^*_j = 1) w_i(Y^*_j, Z^*_j) | D^*_j] \) and

\[ D^*_j = \{ L^*_i, X^*_i, Y^*_i, \delta^*_i, \eta^*_i, Z^*_i \} \]

is the data from the \( i_{th} \) subject. Since \( D^*_i \) and \( D^*_j \) are independent and \( Ew(Y, z) = 0 \), we have that \( \chi_{1,i}(t) \) is also of expectation 0.

Define \( \chi_{2,i}(t) = Z^*_i I(X^*_i \leq t, \delta^*_i = 1, \eta^*_i = 1) W^{-1}(Y^*_i, Z^*_i) - g(\theta^*_i) \), we have that

\[ S_n^G(b, t) = n^{-1} \sum_{i=1}^{n} \chi_{2,i}(t) \]

and \( E \chi_{2,i}(t) = 0 \). Moreover, we can combine (C.1) and get

\[
\sqrt{n} \{ S_n(b, t) - S_n[\hat{\beta}(t), t] - S_n[\beta_0(t), t]) \} = -\sqrt{n} S_n[\beta_0(t), t] + n^{-1/2} \sum_{i=1}^{n} \chi_i(t), \tag{C.2}
\]

where \( \chi_i(t) = \chi_{1,i}(t) + \chi_{2,i}(t) \). Following standard arguments of empirical process, the function class \( \mathcal{F} = \{ \chi(t) : t \in \mathcal{T} \} \) can be shown to be Donsker. It then follows that

\[
\sqrt{n} \{ S_n[\hat{\beta}(t), t] - S_n[\beta_0(t), t] \} \]

converges weakly to a Gaussian process with asymptotic variance covariance \( \sigma^2(t, s) = E[\chi(t) \chi(s)] \).

We can now take a Taylor expansion of \( S_n[\hat{\beta}(t), t] \) around \( \beta_0(t) \), and combine the uniform consistency of \( \hat{\beta}(t) \) to \( \beta_0(t) \) to see

\[
\sqrt{n} \{ S_n[\hat{\beta}(t), t] - S_n[\beta_0(t), t] \} = D_n[\beta_0(t)] \times \sqrt{n} \{ \hat{\beta}(t) - \beta_0(t) \} + o_p^{[l, u]}(1).
\]

Since \( D_n[\beta_0(t)] = -n^{-1} \sum_{i=1}^{n} Z_i \otimes g'(Z_i^T \beta_0(t)) \) is uniformly consistent for \( d[\beta_0(t)] = -E[Z_i \otimes g'(Z_i^T \beta_0(t))] \), we can combine (C.2) and get

\[
n^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} = -n^{-1/2} \sum_{i=1}^{n} d[\beta_0(t)]^{-1} \chi_i(t) + o_p^{[l, u]}(1).
\]

Write \( \xi(t) = -d[\beta_0(t)]^{-1} \chi(t) \). As \( ||d[\beta_0(t)]^{-1}|| \) is bounded from above and \( \mathcal{F} = \{ \chi(t) : t \in \mathcal{T} \} \) is Donsker, the function \( \{ \xi(t) : t \in [l, u] \} \) can also be shown to be Donsker, as Donsker’s property is preserved under Lipschitz transformations. This immediately gives the weak convergence of \( n^{1/2} \{ \hat{\beta}(t) - \beta_0(t) \} \) to a zero-mean Gaussian process, whose covariance matrix is given by \( d[\beta_0(t)]^{-1} E[\chi(t) \chi(t)^T] d[\beta_0(t)]^{-T} \). This completes the proof of Theorem 2.
Figure 1.

Root mean squared error (RMSE) of the proposed $\hat{\beta}^{(k)}$ (black) and the $\hat{\beta}^*_B^{(k)}$ (blue) in Scheike et al. (2008). The dashed lines correspond to the situation with $n = 200$, and the solid lines correspond to $n = 400$. The first three columns represent the RMSE of the estimated regression coefficients, and the last column illustrates the RMSE of the corresponding $F_1 \{t | Z = (1, 1, 0)^T \}$ estimates. From top to bottom, the four rows correspond to Setups S1A, S2A, S1M and S2M respectively.
Figure 2.
Empirical coverage rates of the 95% confidence intervals based on the proposed methods (black) and the naive methods (blue) respectively. The dashed lines correspond to the situation with $n = 200$, and the solid lines correspond to $n = 400$. From top to bottom, the four rows correspond to Setups S1A, S2A, S1M, S2M, respectively.
Figure 3.
Analysis of Denmark diabetes registry data: estimated baseline cumulative incidence of DN, which corresponds to female subjects at age 15, and the estimated regression coefficients for diabetes onset age and gender (bold solid line: proposed estimates; dotted line: proposed 95% Wald-type confidence intervals; dashed line: naive estimates by ignoring left truncation).
Figure 4.
Analysis of Denmark diabetes registry data: ratio of the naive estimator and the proposed estimator.
Table 1

Summary of simulation setups: the choice of \( \{ g, \beta_0^{(1)}(t), \beta_0^{(2)}(t), p_L, L_0 \} \) and the resulting truncation and censoring probabilities, where \( P_1 = \Pr(T_2 < L) \), \( P_2 = \Pr(\delta^* = 1) \), and \( P_3 = \Pr(\eta^* = 1) \).

<table>
<thead>
<tr>
<th></th>
<th>( g )</th>
<th>( \beta_0^{(1)}(t) )</th>
<th>( \beta_0^{(2)}(t) )</th>
<th>( p_L )</th>
<th>( L_0 )</th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A</td>
<td>( g_1 )</td>
<td>((1 + \exp(-3t))^{-1} )</td>
<td>0.5</td>
<td>0.80</td>
<td>2.30</td>
<td>Uniform(0, 1)</td>
<td>0.2</td>
<td>0.57</td>
</tr>
<tr>
<td>S2A</td>
<td>( g_1 )</td>
<td>((1 + \exp(-3t))^{-1} )</td>
<td>0.5</td>
<td>0.85</td>
<td>3.75</td>
<td>Beta(5, 2)</td>
<td>0.4</td>
<td>0.41</td>
</tr>
<tr>
<td>S1M</td>
<td>( g_2 )</td>
<td>0.6</td>
<td>( 0.5t/(0.5t+1) )</td>
<td>0.5</td>
<td>0.80</td>
<td>2.30</td>
<td>Uniform(0, 1)</td>
<td>0.2</td>
</tr>
<tr>
<td>S2M</td>
<td>( g_2 )</td>
<td>0.6</td>
<td>( 0.5t/(0.5t+1) )</td>
<td>0.5</td>
<td>0.85</td>
<td>3.75</td>
<td>Beta(5, 2)</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Summary of simulation results under setups S1A and S2A, where the five rows under each setup correspond to \( t = 0.3, t = 0.5, t = 0.6, t = 1.1 \) and \( t = 1.5 \), respectively.

<table>
<thead>
<tr>
<th></th>
<th>EBias×10^3</th>
<th>EBias×10^3</th>
<th>ESD×10^3</th>
<th>ASD×10^3</th>
<th>CP95×10^3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta}^{(0)} )</td>
<td>( \hat{\beta}^{(1)} )</td>
<td>( \hat{\beta}^{(2)} )</td>
<td>( \hat{\beta}^{(0)} )</td>
<td>( \hat{\beta}^{(1)} )</td>
</tr>
<tr>
<td><strong>S1A, ( n = 200 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{B,10} )</td>
<td>-108</td>
<td>-116</td>
<td>-81</td>
<td>26</td>
<td>-3</td>
</tr>
<tr>
<td>( \beta_B^{(0)} )</td>
<td>-148</td>
<td>-167</td>
<td>-77</td>
<td>59</td>
<td>-5</td>
</tr>
<tr>
<td>( \beta_B^{(1)} )</td>
<td>-155</td>
<td>-180</td>
<td>-75</td>
<td>71</td>
<td>-6</td>
</tr>
<tr>
<td>( \beta_B^{(2)} )</td>
<td>-149</td>
<td>-194</td>
<td>-65</td>
<td>79</td>
<td>-9</td>
</tr>
<tr>
<td>( \beta_B^{(3)} )</td>
<td>-142</td>
<td>-196</td>
<td>-61</td>
<td>79</td>
<td>-10</td>
</tr>
<tr>
<td><strong>S1A, ( n = 400 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{B,10} )</td>
<td>-106</td>
<td>-115</td>
<td>-84</td>
<td>29</td>
<td>-3</td>
</tr>
<tr>
<td>( \beta_B^{(0)} )</td>
<td>-144</td>
<td>-166</td>
<td>-78</td>
<td>61</td>
<td>-5</td>
</tr>
<tr>
<td>( \beta_B^{(1)} )</td>
<td>-149</td>
<td>-176</td>
<td>-76</td>
<td>67</td>
<td>-5</td>
</tr>
<tr>
<td>( \beta_B^{(2)} )</td>
<td>-145</td>
<td>-194</td>
<td>-61</td>
<td>71</td>
<td>-6</td>
</tr>
<tr>
<td>( \beta_B^{(3)} )</td>
<td>-135</td>
<td>-192</td>
<td>-63</td>
<td>74</td>
<td>-6</td>
</tr>
<tr>
<td><strong>S2A, ( n = 200 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{B,10} )</td>
<td>-109</td>
<td>-119</td>
<td>-151</td>
<td>-12</td>
<td>-14</td>
</tr>
<tr>
<td>( \beta_B^{(0)} )</td>
<td>-191</td>
<td>-215</td>
<td>-194</td>
<td>21</td>
<td>-21</td>
</tr>
<tr>
<td>( \beta_B^{(1)} )</td>
<td>-231</td>
<td>-263</td>
<td>-194</td>
<td>49</td>
<td>-23</td>
</tr>
<tr>
<td>( \beta_B^{(2)} )</td>
<td>-320</td>
<td>-391</td>
<td>-170</td>
<td>139</td>
<td>-30</td>
</tr>
<tr>
<td>( \beta_B^{(3)} )</td>
<td>-333</td>
<td>-427</td>
<td>-154</td>
<td>160</td>
<td>-32</td>
</tr>
<tr>
<td><strong>S2A, ( n = 400 )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{B,10} )</td>
<td>-110</td>
<td>-120</td>
<td>-155</td>
<td>-9</td>
<td>-8</td>
</tr>
<tr>
<td>( \beta_B^{(0)} )</td>
<td>-187</td>
<td>-212</td>
<td>-202</td>
<td>23</td>
<td>-12</td>
</tr>
<tr>
<td>( \beta_B^{(1)} )</td>
<td>-228</td>
<td>-261</td>
<td>-202</td>
<td>51</td>
<td>-13</td>
</tr>
<tr>
<td>( \beta_B^{(2)} )</td>
<td>-316</td>
<td>-388</td>
<td>-184</td>
<td>140</td>
<td>-16</td>
</tr>
<tr>
<td>( \beta_B^{(3)} )</td>
<td>-328</td>
<td>-424</td>
<td>-167</td>
<td>161</td>
<td>-18</td>
</tr>
</tbody>
</table>
Table 3
Summary of simulation results under setups S1M and S2M, where the five rows under each setup correspond to \( t = 0.3, t = 0.5, t = 0.6, t = 1.1 \) and \( t = 1.5 \), respectively.

<table>
<thead>
<tr>
<th></th>
<th>( \hat{F}_{B,10} )</th>
<th>( \hat{\beta}_B^{(0)} )</th>
<th>( \hat{\beta}_B^{(1)} )</th>
<th>( \hat{\beta}_B^{(2)} )</th>
<th>( \hat{F}_{B,10} )</th>
<th>( \hat{\beta}_B^{(0)} )</th>
<th>( \hat{\beta}_B^{(1)} )</th>
<th>( \hat{\beta}_B^{(2)} )</th>
<th>( \hat{F}_{B,10} )</th>
<th>( \hat{\beta}_B^{(0)} )</th>
<th>( \hat{\beta}_B^{(1)} )</th>
<th>( \hat{\beta}_B^{(2)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A, ( n = 100 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–87</td>
<td>–1254</td>
<td>231</td>
<td>206</td>
<td>9</td>
<td>–103</td>
<td>17</td>
<td>13</td>
<td>6</td>
<td>49</td>
<td>47</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>–111</td>
<td>–936</td>
<td>183</td>
<td>303</td>
<td>6</td>
<td>–55</td>
<td>8</td>
<td>8</td>
<td>50</td>
<td>37</td>
<td>37</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>–119</td>
<td>–836</td>
<td>164</td>
<td>304</td>
<td>4</td>
<td>–51</td>
<td>11</td>
<td>9</td>
<td>45</td>
<td>34</td>
<td>34</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>–117</td>
<td>–540</td>
<td>112</td>
<td>243</td>
<td>1</td>
<td>–34</td>
<td>7</td>
<td>9</td>
<td>37</td>
<td>29</td>
<td>28</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>–115</td>
<td>–468</td>
<td>107</td>
<td>229</td>
<td>4</td>
<td>–17</td>
<td>–4</td>
<td>1</td>
<td>35</td>
<td>28</td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>S1M, ( n = 400 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–90</td>
<td>–1159</td>
<td>192</td>
<td>215</td>
<td>4</td>
<td>–49</td>
<td>10</td>
<td>6</td>
<td>45</td>
<td>34</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>–116</td>
<td>–924</td>
<td>175</td>
<td>310</td>
<td>3</td>
<td>–32</td>
<td>8</td>
<td>8</td>
<td>35</td>
<td>26</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>–122</td>
<td>–824</td>
<td>154</td>
<td>314</td>
<td>2</td>
<td>–25</td>
<td>6</td>
<td>6</td>
<td>32</td>
<td>24</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>–121</td>
<td>–543</td>
<td>113</td>
<td>240</td>
<td>1</td>
<td>–18</td>
<td>4</td>
<td>5</td>
<td>26</td>
<td>20</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>–116</td>
<td>–459</td>
<td>102</td>
<td>211</td>
<td>3</td>
<td>–6</td>
<td>–7</td>
<td>0</td>
<td>25</td>
<td>20</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>S2M, ( n = 200 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–100</td>
<td>–1702</td>
<td>192</td>
<td>–150</td>
<td>15</td>
<td>–168</td>
<td>42</td>
<td>–4</td>
<td>88</td>
<td>69</td>
<td>68</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>–164</td>
<td>–1941</td>
<td>377</td>
<td>252</td>
<td>8</td>
<td>–133</td>
<td>48</td>
<td>15</td>
<td>72</td>
<td>56</td>
<td>55</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>–195</td>
<td>–1981</td>
<td>415</td>
<td>477</td>
<td>7</td>
<td>–116</td>
<td>44</td>
<td>19</td>
<td>69</td>
<td>54</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>–242</td>
<td>–1488</td>
<td>251</td>
<td>727</td>
<td>1</td>
<td>–89</td>
<td>43</td>
<td>23</td>
<td>59</td>
<td>47</td>
<td>44</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>–254</td>
<td>–1299</td>
<td>242</td>
<td>682</td>
<td>2</td>
<td>–68</td>
<td>33</td>
<td>18</td>
<td>57</td>
<td>47</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td>S2M, ( n = 400 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>–105</td>
<td>–1634</td>
<td>166</td>
<td>–169</td>
<td>7</td>
<td>–86</td>
<td>12</td>
<td>8</td>
<td>61</td>
<td>46</td>
<td>46</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>–170</td>
<td>–1904</td>
<td>352</td>
<td>247</td>
<td>5</td>
<td>–59</td>
<td>12</td>
<td>15</td>
<td>49</td>
<td>38</td>
<td>38</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>–195</td>
<td>–1935</td>
<td>394</td>
<td>467</td>
<td>3</td>
<td>–57</td>
<td>15</td>
<td>16</td>
<td>46</td>
<td>36</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>–245</td>
<td>–1490</td>
<td>243</td>
<td>738</td>
<td>1</td>
<td>–40</td>
<td>14</td>
<td>15</td>
<td>40</td>
<td>32</td>
<td>31</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>–256</td>
<td>–1303</td>
<td>241</td>
<td>689</td>
<td>2</td>
<td>–31</td>
<td>8</td>
<td>12</td>
<td>39</td>
<td>32</td>
<td>30</td>
<td>39</td>
</tr>
</tbody>
</table>

\( \hat{F}_{B,10} \) and \( \hat{\beta}_B \) are the estimated values of the bias and treatment effect, respectively.
Table 4

Summary of simulation results: empirical rejection rates of the proposed hypothesis tests based on different test statistics.

<table>
<thead>
<tr>
<th>Setup</th>
<th>n</th>
<th>$\mathcal{R}_{1,\text{sup}}^{(1)}$</th>
<th>$\mathcal{R}_{1,\text{inte}}^{(1)}$</th>
<th>$\mathcal{R}_{1,\text{sup}}^{(2)}$</th>
<th>$\mathcal{R}_{1,\text{inte}}^{(2)}$</th>
<th>$\mathcal{R}_{2,\text{sup}}^{(1)}$</th>
<th>$\mathcal{R}_{2,\text{inte}}^{(1)}$</th>
<th>$\mathcal{R}_{2,\text{sup}}^{(2)}$</th>
<th>$\mathcal{R}_{2,\text{inte}}^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1A</td>
<td>200</td>
<td>0.543</td>
<td>0.562</td>
<td>0.043</td>
<td>0.041</td>
<td>0.340</td>
<td>0.378</td>
<td>0.041</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.809</td>
<td>0.824</td>
<td>0.050</td>
<td>0.051</td>
<td>0.593</td>
<td>0.606</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>S2A</td>
<td>200</td>
<td>0.291</td>
<td>0.286</td>
<td>0.030</td>
<td>0.030</td>
<td>0.186</td>
<td>0.222</td>
<td>0.043</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.488</td>
<td>0.494</td>
<td>0.042</td>
<td>0.042</td>
<td>0.293</td>
<td>0.342</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>S1M</td>
<td>200</td>
<td>0.535</td>
<td>0.578</td>
<td>0.243</td>
<td>0.223</td>
<td>0.056</td>
<td>0.052</td>
<td>0.197</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.834</td>
<td>0.849</td>
<td>0.496</td>
<td>0.408</td>
<td>0.055</td>
<td>0.053</td>
<td>0.339</td>
<td>0.396</td>
</tr>
<tr>
<td>S2M</td>
<td>200</td>
<td>0.203</td>
<td>0.252</td>
<td>0.089</td>
<td>0.106</td>
<td>0.052</td>
<td>0.052</td>
<td>0.209</td>
<td>0.176</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>0.445</td>
<td>0.496</td>
<td>0.181</td>
<td>0.184</td>
<td>0.066</td>
<td>0.054</td>
<td>0.298</td>
<td>0.288</td>
</tr>
</tbody>
</table>
Table Appendix A.1

Summary of simulation results, where sample size $n = 100$. The five rows under each setup correspond to $t = 0.3$, $t = 0.5$, $t = 0.6$, $t = 1.1$ and $t = 1.5$, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$EBias \times 10^3$</th>
<th>$EBias \times 10^3$</th>
<th>$ESD \times 10^2$</th>
<th>$ASD \times 10^2$</th>
<th>$CP95 \times 10^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{B,10}$</td>
<td>$\hat{\beta}^{(0)}_B$</td>
<td>$\hat{\beta}^{(1)}_B$</td>
<td>$\hat{\beta}^{(2)}_B$</td>
<td>$\hat{\beta}^{(0)}_B$</td>
</tr>
<tr>
<td>S1A, $n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2A, $n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1M, $n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2M, $n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>