Published in final edited form as:
Biometrics. 2009 June ; 65(2): 405–414. doi:10.1111/j.1541-0420.2008.01077.x.

Marginal Hazards Regression for Retrospective Studies within Cohort with Possibly Correlated Failure Time Data

Sangwook Kang[1,*] and **Jianwen Cai**[2,**]

[1]Department of Epidemiology and Biostatistics, University of Georgia, Athens, Georgia 30602, U.S.A.

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A.

Summary

A retrospective dental study was conducted to evaluate the degree to which pulpal involvement affects tooth survival. Due to the clustering of teeth, the survival times within each subject could be correlated and thus the conventional method for the case–control studies cannot be directly applied. In this article, we propose a marginal model approach for this type of correlated case–control within cohort data. Weighted estimating equations are proposed for the estimation of the regression parameters. Different types of weights are also considered for improving the efficiency. Asymptotic properties of the proposed estimators are investigated and their finite sample properties are assessed via simulations studies. The proposed method is applied to the aforementioned dental study.

Keywords

Correlated case–control studies within cohort; Correlated failure times; Marginal hazard model; Survival analysis; Weighted estimating equations

1. Introduction

Case–control study design is an efficient and economic method to ascertain a large number of cases in a relatively short period of time. Often, the case–control study is conducted within a well-defined cohort. For example, in occupational epidemiology, a commonly used approach is to conduct a case–control study nested within a cohort that has already been enumerated. The reason for conducting a case–control study even when a cohort can be enumerated is usually that more information is needed than is readily available from records and it would be too expensive to seek this information for everyone in the cohort (Rothman, 2002). Thus, such a case–control study could greatly reduce the cost while achieving the same goals as a cohort study. Failure time models from such retrospective case–control studies have been studied in the literature (Thomas, 1977; Prentice and Breslow, 1978; Binder, 1992; Borgan, Goldstein, and Langholz, 1995; Samuelsen, 1997; Chen and Lo, 1999; Lin, 2000; Chen, 2001). An important assumption for these conventional case–control studies is the statistical independence among subjects. However, in many biomedical studies, this assumption might not hold. For example, in a retrospective cohort dental study (Caplan and Weintraub, 1997; Caplan et al., 2005), it was of interest to evaluate the degree to which pulpal involvement affects tooth survival. Root canal filled (RCF) teeth were used as an indicator of pulpal involvement.

* skang@uga.edu . ** cai@bios.unc.edu .

In this study, cases were defined as those who lost the RCF tooth, while controls were defined as those who did not lose the RCF tooth during the study period. After cases and controls were sampled, a non-RCF tooth was matched to the RCF tooth within each subject. The survival times of the two teeth within the same subject could be correlated and thus the independence assumption might not be valid. The primary goal of the study is to evaluate the effect of pulpal involvement on tooth survival. The fact that the survival times of the teeth from the same individual are correlated is considered as a nuisance. In such case, a marginal model approach is appealing. Examples like this one are very common in biomedical studies. For example, case–control family studies have been frequently used to assess familial aggregation of a disease and the relationship between the disease and genetic or environmental risk factors. In such studies, independent cases and controls are identified and information is collected for both cases and controls and their relatives. Because related individuals share common genetic or environmental factors, their failure times could be correlated.

There is an extensive literature of statistical methods for correlated failure time data from prospective cohort studies (Wei, Lin, and Weissfeld, 1989; Lee, Wei, and Amato, 1992; Lin, 1994; Cai and Prentice, 1995; Cai and Prentice, 1997; Spiekerman and Lin, 1998; Clegg, Cai, and Sen, 1999). However, these methods cannot be directly applied to correlated failure time data from case–control within cohort studies. Work for correlated failure time data from case–control within cohort studies has been limited. Some efforts have been made to analyze failure time data from case–control family studies (e.g., Li, Yang, and Schwartz, 1998; Hsu et al., 1999; Shih and Chatterjee, 2002; Hsu et al., 2004). For the case–control family studies, investigators are usually interested in estimating the strength of dependence of failure times within family. Consequently, most of the methods concentrated on frailty models or parametric approach, with the exception of Shih and Chatterjee (2002). In Shih and Chatterjee (2002), the authors considered a quasipartial-likelihood approach for simultaneously estimating the parameters in the marginal proportional hazard model and the association among family members. However, the asymptotic properties of the proposed estimator were not clear and estimation of the variance of their estimator relied on a bootstrap method. When the correlation of the failure times is not of interest, as in the aforementioned dental study, a statistical inference procedure that is easy to conduct and has good asymptotic properties remains to be developed. Recently, Lu and Shih (2006) proposed case–cohort designs for clustered failure time data. Although our study and Lu and Shih's (2006) design both involve correlated failure time data, they are different designs. Lu and Shih's (2006) design involves a random sampling of clusters (subcohort) regardless of the failure status of the members within clusters and the subsequent addition of all the remaining failures in the entire cohort (design B). For their design A, they randomly sample r individuals (where r is a prespecified number) per cluster from every cluster regardless of the failure status of the members within clusters (subcohort) and then all failures from the entire cohort. On the other hand, in our study design, the sampling of clusters depends on the failure status of the index member of the cluster. In addition, unlike Lu and Shih's (2006) design, not all the failures are necessarily sampled. These differences distinguish our study design from that of Lu and Shih (2006). While methods for the case–cohort design for clustered failure time data have been proposed by Lu and Shih (2006), it is desirable to develop hazard regression models for the correlated failure time data from case–control within cohort studies that account for the possible correlation within subject while avoiding the specification of the correlation structure.

For univariate failure time data from complex sampling designs, Binder (1992) proposed an estimating equation approach for fitting Cox's proportional hazards models for complex survey data. Lin (2000) studied the theoretical aspects of estimating procedures by Binder (1992) and extended it to the superpopulation approach. The sampling weights used in Binder (1992) are proportional to the inverse of the sampling probability. Samuelsen (1997) considered the same type of weights when nested case–control design is involved and Chen (2001) proposed a more

efficient estimator by using different forms of weights. Recently, Breslow and Wellner (2007) proposed inverse probability weighted estimators under semi-parametric models with two-phased stratified samples. All these methods assume independent failure times and cannot be directly applied to multivariate failure time data.

In this article, we propose a weighted estimating equation approach for estimating the parameters in the marginal hazards regression models for the correlated failure time data from case–control studies within cohort.

The rest of this article is organized as follows. The proposed model and method of estimation are presented in Section 2, followed by the study of the asymptotics in Section 3. In Section 4, we report some simulation results. The methodology is illustrated in Section 5 using the aforementioned dental study. In Section 6, we give a few concluding remarks.

2. Modeling and Estimation

2.1 Marginal Hazards Model

Let i indicate cluster and k indicate member within cluster. Let T_{ik} denote the failure time for member k of cluster i. In the aforementioned retrospective dental study example, i would indicate the patient, k would indicate the tooth within the patient, and T_{i1}, T_{i2} would represent the failure time of the index tooth and the matching tooth, respectively, for patient i. Let C_{ik} denote the censoring time. We assume that C_{ik} is independent of the failure process conditional on covariates. Without loss of generality, we assume that there are K members in each cluster. Varying cluster sizes can be accommodated by setting the corresponding C_{ik} to be equal to zero. In many practical cases, $C_{ik} = C_i$ for $k = 1,...,K$. The observed time is $X_{ik} = \min(T_{ik}, C_{ik})$ and $\Delta_{ik} = I(T_{ik} \leq C_{ik})$ is an indicator for failure. Note that the "at risk" indicator process is given by $Y_{ik}(t) = I(X_{ik} \geq t)$ for member k of cluster i and let $N_{ik}(t) = I(X_{ik} \leq t, \Delta_{ik} = 1)$ denote the counting process corresponding to T_{ik}. Let $\lambda_{ik}(t)$ and

$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) \exp\left\{\beta_0^T \boldsymbol{Z}_{ik}(u)\right\} \lambda_0(u)\, du$ denote the corresponding marginal hazards function and a martingale with respect to the marginal filtration $\mathcal{F}_{ik}(t) = \sigma\{N_{ik}(s), Y_{ik}(s), \boldsymbol{Z}_{ik}(s) : 0 \leq s \leq t\}$. Note that $M_{ik}(t)$ are not martingales with respect to the joint filtration generated by all of the failure, censoring, and covariate history up to time t, $\mathcal{F}(t) = \vee_{i=1}^{n} \vee_{k=1}^{K} \mathcal{F}_{ik}(t)$, due to the intraclass dependence. Let τ denote the study end time.

Suppose that T_{ik} arises from a marginal intensity process model of the form (Lee et al., 1992)

$$\lambda_{ik}(t) = Y_{ik}(t)\, \lambda_0(t) \exp\left\{\beta_0^T \boldsymbol{Z}_{ik}(t)\right\}, \tag{1}$$

where $\boldsymbol{Z}_{ik}(t) = (Z_{1ik}(t),...,Z_{pik}(t))^T$ is a p-dimensional vector of covariates for member k of cluster i, and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of fixed and unknown parameters. We assume that all the time-dependent covariates in $\boldsymbol{Z}_{ik}(t)$ are "external," i.e., they are not affected by the disease processes, as described by Kalbfleisch and Prentice (2002).

2.2 Estimation of Regression Parameters and Cumulative Baseline Hazard Function

Under the correlated case–control within cohort study design, the index member of a cluster is specified. The case–control sample is drawn based on the failure status of the index member. Suppose we select $\tilde{n}_1$ case clusters and $\tilde{n}_0$ control clusters from the n_1 case clusters and n_0 control clusters, respectively, in the population. Let $n = n_1 + n_0$ and $\tilde{n} = \tilde{n}_1 + \tilde{n}_0$. Case clusters are defined as those clusters with the index member having experienced failure while control clusters are defined as those clusters with the index member having not experienced failure. Note that the control cluster could have nonindex members who experience the failure. Each

case (control) cluster has the same probability $\tilde{n}_1/n_1(\tilde{n}_0/n_0)$ to be selected. Let π_i denote this inclusion probability for the ith cluster and ξ_i denote the indicator for being selected. Note that by the study design, the inclusion probability π_i depends on the failure status of the index member, i.e., $\pi_i = \Pr(\xi_i = 1|\Delta_{i1})$ and the K members in the ith stratum have the same inclusion statuses, i.e., $\pi_{ik} = \pi_i$ and $\xi_{ik} = \xi_i$ for $k = 1,...,K$. We will use $k = 1$ to indicate the index members.

We propose the following weighted estimating equations for estimating $\boldsymbol{\beta}_0$:

$$\widehat{\boldsymbol{U}}(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}^{(1)}(\beta,t)}{\widehat{S}^{(0)}(\beta,t)} \right\} dN_{ik}(t) = 0, \tag{2}$$

where

$$\begin{aligned} w_i &= \frac{\xi_i}{\pi_i}, \\ \widehat{\boldsymbol{S}}^{(d)}(\beta,t) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(t)\,\boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\beta^T \boldsymbol{Z}_{ik}(t)} \quad (d=0,1), \end{aligned}$$

and

$$\boldsymbol{a}^{\otimes 0} = 1, \boldsymbol{a}^{\otimes 1} = \boldsymbol{a}, \quad \text{and} \quad \boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^T \quad \text{for a vector} \quad \boldsymbol{a}.$$

It is assumed that $\pi_i > 0$ for all i.

Let $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$. To predict the t-year survival probability for future patients with specific covariates, a Breslow–Aalen type estimator of the baseline cumulative hazard function is proposed and is given by:

$$\widehat{\Lambda}_0\left(\widehat{\beta},t\right) = \int_0^t \frac{\sum_{i=1}^{n}\sum_{k=1}^{K} w_i \quad dN_{ik}(s)}{\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(s)\, e^{\widehat{\beta}^T \boldsymbol{Z}_{ik}(s)}}, \tag{3}$$

where $w_i = \xi_i/\pi_i$ and $\widehat{\beta}$ is a solution for equation (2).

Note that when $K = 1$, i.e., when failure time data are from traditional case–control studies without correlated components from the same cluster, the proposed estimators reduce to the ones studied by Binder (1992) and Lin (2000) for complex survey data for univariate failure time. When all the subjects are sampled, i.e., $\xi_i = 1$, $\pi_i = 1$, $i = 1,...,n$, the proposed estimators reduce to the one studied by Lee et al. (1992) for random samples.

Suppose the information on the observed failure times of all the cohort members are available. Under such situation, using only the inclusion probability π_i might not be efficient because it does not fully use the available information. Note that calculation of π_i only requires the cohort size n_0, n_1 and the sample size $\tilde{n}_0$, $\tilde{n}_1$. Thus, in an attempt to increase the efficiency of the estimator, a different type of weight that uses all the available information is desired. To this end, we consider a local average estimator. The idea of the local average estimator is to replace each missing covariate term by an appropriate local average. This estimator was considered

by Chen (2001) for independent data. We propose the following weighted estimating equations for estimating $\boldsymbol{\beta}_0$. The form of the weighted estimating equations is the same as the one with inclusion probabilities except that we replace π_i with a local average. Specifically,

$$\widehat{U}_c(\beta) = \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} w_i \left\{ \boldsymbol{Z}_{ik}(t) - \frac{\widehat{\boldsymbol{S}}_c^{(1)}(\beta,t)}{\widehat{S}_c^{(0)}(\beta,t)} \right\} dN_{ik}(t) = 0, \tag{4}$$

where

$$\begin{aligned} w_i &= \frac{\xi_i}{r_n(X_{i1},\Delta_{i1})}, \\ \widehat{\boldsymbol{S}}_c^{(d)}(\beta,t) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} w_i Y_{ik}(t)\, \boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\beta^T \boldsymbol{Z}_{ik}(t)}, (d=0,1) \end{aligned}$$

and

$$\begin{aligned} r_n(t,d) &= \frac{\sum_{j=1}^{n}\Delta_{j1}\xi_j I\{X_{j1}\in[t_{l-1},t_l)\}}{\sum_{j=1}^{n}\Delta_{j1} I\{X_{j1}\in[t_{l-1},t_l)\}} \\ &\qquad \text{if} \quad d=1 \quad \text{and} \quad t\in[t_{l-1},t_l), \\ &= \frac{\sum_{j=1}^{n}(1-\Delta_{j1})\xi_j I\{X_{j1}\in[s_{m-1},s_m)\}}{\sum_{j=1}^{n}(1-\Delta_{j1}) I\{X_{j1}\in[s_{m-1},s_m)\}} \\ &\qquad \text{if} \quad d=0 \quad \text{and} \quad t\in[s_{m-1},s_m) \end{aligned}$$

for some $1 \le l \le a_n$ and $1 \le m \le b_n$, where $0 = t_0 \le t_1 \le \cdots \le t_{a_n} = \tau$ and $0 = s_0 \le s_1 \le \cdots s_{b_n} = \tau$ are two partitions of $[0, \tau)$. With additional assumptions, the estimator using this weight function is expected to be more efficient than the previous one using the inclusion probability in the sense that the former results in a parameter estimator with smaller asymptotic variance. Note that as pointed out by Samuelsen, Anestad, and Skrondal (2007), this local average method can be described by a procedure called "poststratification" in survey sampling literature. Specifically, after cases and controls are sampled, we divide the cohort as well as the sampled data into strata constructed from using the additional information (in this case, the failure times and censoring times for all the cohort members). Then, we construct the weighted estimating functions as if the data were collected originally by stratified case–control sampling.

The Breslow–Aalen type estimator of the baseline cumulative hazard function $\widehat{\Lambda}_0^c\left(\widehat{\beta}_c, t\right)$ will be in the form of equation (3) with $w_i = \xi_i / r_n(X_{i1}, \Delta_{i1})$ and $\widehat{\beta}_c$ is a solution for equation (4).

3. Asymptotic Properties

In this section, we describe the asymptotic properties of the proposed estimates. We introduce the following notation for convenience:

$$
\begin{aligned}
\boldsymbol{S}^{(d)}(\beta,t) &= n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K} Y_{ik}(t)\,\boldsymbol{Z}_{ik}(t)^{\otimes d} e^{\beta^T \boldsymbol{Z}_{ik}(t)},\\
\boldsymbol{s}^{(d)}(\beta,t) &= \mathrm{E}\left\{\boldsymbol{S}^{(d)}(\beta,t)\right\} (d=0,1,2),\\
\mathbf{e}(\beta,t) &= \frac{\boldsymbol{s}^{(1)}(\beta,t)}{s^{(0)}(\beta,t)},\\
\boldsymbol{v}(\beta,t) &= \frac{\boldsymbol{s}^{(2)}(\beta,t)s^{(0)}(\beta,t)-\boldsymbol{s}^{(1)}(\beta,t)^{\otimes 2}}{s^{(0)}(\beta,t)^2},\\
\tilde{\boldsymbol{Z}}_{ik}(\beta,t) &= \boldsymbol{Z}_{ik}(t)-\boldsymbol{e}(\beta,t),\\
\boldsymbol{M}_{\tilde{\boldsymbol{z}},ik}(\beta) &= \int_0^{\tau}\tilde{\boldsymbol{Z}}_{ik}(\beta,t)\,dM_{ik}(t) \quad \text{and}\\
\boldsymbol{A}(\beta) &= \int_0^{\tau}\boldsymbol{v}(\beta,t)\,s^{(0)}(\beta,t)\,\lambda_0(t)\,dt.
\end{aligned}
$$

The regularity conditions needed to derive the asymptotic properties are given in the supplementary material.

3.1 Asymptotic Properties of $\widehat{\beta}$ and $\widehat{\Lambda}(\widehat{\beta},t)$

We summarize the asymptotic behavior of the regression parameter estimator under the inclusion probability approach in the following theorem:

THEOREM 1. *Under the regularity conditions,* $\widehat{\beta}$ *solving equation* (2) *is a consistent estimator of* $\boldsymbol{\beta}_0$. *Also* $n^{1/2}\left(\widehat{\beta}-\beta_0\right)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \boldsymbol{V}(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ where

$$
\begin{aligned}
\boldsymbol{Q}(\beta) &= \mathrm{E}\left[\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\beta)\right)^{\otimes 2}\right],\\
\boldsymbol{V}(\beta) &= \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Var}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\beta)\,\middle|\,\Delta_{11}\right)\right]\\
&\quad \textit{and} \quad \alpha(\Delta_{11})|_{\Delta_{11}=s}=\alpha_s=\lim_{n\to\infty}\frac{\tilde{n}_s}{n_s}\,(s=0 \quad \textit{and} \quad 1).
\end{aligned}
$$

Note that $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ has two sources of variations: $\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{Q}\times(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the cohort and $\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{V}(\boldsymbol{\beta}_0)\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ is the variation due to the sampling of the case–control sample from the cohort. We summarize the asymptotic properties of $\widehat{\Lambda}_0(\widehat{\beta},t)$ in the following theorem:

THEOREM 2. *Under the regularity conditions,* $\widehat{\Lambda}_0(\widehat{\beta},t)$ *is a onsistent estimator of* $\Lambda_0(t)$. *Also,* $n^{1/2}\left(\widehat{\Lambda}_0(\widehat{\beta},t)-\Lambda_0(t)\right)$ converges weakly to a zero-mean Gaussian process with covariance function $\varphi(t_1, t_2)(\boldsymbol{\beta}_0) + \sigma(t_1, t_2)(\boldsymbol{\beta}_0)$ *at* (t_1, t_2) *where*

$$
\begin{aligned}
\phi(t_1,t_2)(\beta) &= \mathrm{E}\left[\left(\sum_{k=1}^{K}\phi_{1k}(\beta,t_1)\right)\left(\sum_{m=1}^{K}\phi_{1m}(\beta,t_2)\right)\right],\\
\sigma(t_1,t_2)(\beta) &= \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\right.\\
&\qquad \left.\times\,\mathrm{Cov}\left(\sum_{k=1}^{K}\phi_{1k}(\beta,t_1),\sum_{m=1}^{K}\phi_{1m}(\beta,t_2)\,\middle|\,\Delta_{11}\right)\right],\\
\phi_{ik}(\beta,t) &= \int_0^t \frac{dM_{ik}(u)}{s^{(0)}(\beta,u)}\boldsymbol{r}(\beta,t)^T\boldsymbol{A}^{-1}(\beta)\,\boldsymbol{M}_{\tilde{\boldsymbol{z}},ik}(\beta), \quad \text{and}\\
\boldsymbol{r}(\beta,t) &= -\int_0^t \boldsymbol{e}(\beta,u)\,d\Lambda_0(u).
\end{aligned}
$$

3.2 Asymptotic Properties of $\widehat{\beta}_c$ and $\widehat{\Lambda}^c\left(\widehat{\beta}_c, t\right)$

Now, we describe the asymptotic properties of the model parameter estimators under the local average approach. The asymptotic variance–covariance matrix is shown to have the form of a proportionally allocated stratified sample. This is due to the poststratification argument when the original sampling is either simple random sampling or stratified simple random sampling (Cochran, 1977). Our original sampling scheme is a stratified simple random sampling where the strata are defined by case status. Thus, we do not need the additional assumptions imposed in Chen (2001) for local average method while those assumptions are needed for other sampling schemes such as nest case–control sampling (Samuelsen et al., 2007).

We summarize the asymptotic behavior of the regression parameter estimator and $\widehat{\Lambda}_0^c\left(\widehat{\beta}_c, t\right)$ under the local average approach in the following two theorems:

THEOREM 3. *Under the regularity conditions,* $\widehat{\beta}_c$ *solving equation* (4) *is a consistent estimator of* $\boldsymbol{\beta}_0$. *Also,* $n^{1/2}\left(\widehat{\beta}_c - \beta_0\right)$ is asymptotically normally distributed with mean zero and with variance matrix of the form $\boldsymbol{\Sigma}_c(\boldsymbol{\beta}_0) = \boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)\{\boldsymbol{Q}(\boldsymbol{\beta}_0) + \boldsymbol{V}_c(\boldsymbol{\beta}_0)\}\boldsymbol{A}^{-1}(\boldsymbol{\beta}_0)$ where

$$\boldsymbol{V}_c(\beta) = \mathrm{E}\left[\frac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Var}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\beta)\,\middle|\,X_{11},\Delta_{11}\right)\right].$$

Note that $\boldsymbol{V}_c(\boldsymbol{\beta}_0)$ is not larger than $\boldsymbol{V}(\boldsymbol{\beta}_0)$ because

$$\begin{aligned}\boldsymbol{V}_c(\beta_0) &= \mathrm{E}\left[\tfrac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{E}\right.\\ &\qquad \times \left.\left\{\mathrm{Var}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\beta)\,\middle|\,X_{11},\Delta_{11}\right)\middle|\,\Delta_{11}\right\}\right]\\ &\le \mathrm{E}\left[\tfrac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Var}\left(\sum_{k=1}^{K}\boldsymbol{M}_{\tilde{\boldsymbol{z}},1k}(\beta)\,\middle|\,\Delta_{11}\right)\right].\end{aligned}$$

Hence, $\boldsymbol{\Sigma}_c(\boldsymbol{\beta}_0)$ is not larger than $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$.

THEOREM 4. *Under the regularity conditions,* $\widehat{\Lambda}_0^c\left(\widehat{\beta}_c, t\right)$ *is a consistent estimator of* $\Lambda_0(t)$. *Also,* $n^{1/2}\left(\widehat{\Lambda}_0^c\left(\widehat{\beta}_c, t\right) - \Lambda_0(t)\right)$ converges weakly to a zero-mean Gaussian process with covariance function $\varphi(t_1, t_2)(\boldsymbol{\beta}_0) + \sigma_c(t_1, t_2)(\boldsymbol{\beta}_0)$ *at* (t_1, t_2) *where*

$$\begin{aligned}&\sigma_c(t_1,t_2)(\beta)\\ &\quad=\mathrm{E}\left[\tfrac{1-\alpha(\Delta_{11})}{\alpha(\Delta_{11})}\mathrm{Cov}\left(\sum_{k=1}^{K}\phi_{1k}(\beta,t_1),\right.\right.\\ &\qquad\qquad\left.\left.\sum_{m=1}^{K}\phi_{1m}(\beta,t_2)\,\middle|\,X_{11},\Delta_{11}\right)\right].\end{aligned}$$

The outline of the proofs of the Theorems 1 and 2 as well as the explicit forms of consistent variance estimators are provided in the separate supplementary material.

4. Simulations

Extensive Monte Carlo simulations have been conducted to examine the finite sample properties of the proposed procedures. For each cluster, mimicking the setup for our motivating

dental study example, failure times for two members ($K = 2$) were generated via a multivariate extension of the Clayton and Cuzick (1985) study, in which the joint survival function for (T_1, T_2) given $(\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ is $S(t_1, t_2; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = \{S_1(t_1; \boldsymbol{Z}_1)^{-1/\theta} + S_2(t_2; \boldsymbol{Z}_2)^{-1/\theta} - 1\}^{-\theta}$, where $S_k(t; \boldsymbol{Z}) = \Pr(T_k > t|\boldsymbol{Z}_k)$, $k = 1, 2$, is the marginal survival function for T_k given covariate $\boldsymbol{Z}_k$. We considered a binary covariate with all the first members having 1 and second members having 0, which is the case for the dental study example. We also considered a continuous covariate where the continuous covariate was generated from the standard normal distribution. An exponential distribution was considered for the marginal distribution of the failure times. The parameter θ represents the degree of dependence of T_1 and T_2. The relationship between Kendall's tau and θ is $\tau_\theta = 1/(2\theta + 1)$. Values of 4, 1.25, 0.67, and 0.1 were considered for θ where the smaller the value of θ, the stronger the dependence between T_1 and T_2. The corresponding Kendall's tau values are 0.09, 0.29, 0.43, and 0.83. We used values of 0 and log (2) for the regression parameter β_0. λ_0 was set to 1. Cohort sizes of $n = 1000$ and 2000 were considered. We conducted simple random sampling without replacement for case clusters and for control clusters independently. Approximately 80% and 90% of censoring proportions were considered for each setup and 90% of the case clusters and the same number of control clusters were sampled. The censoring times were generated from uniform(0, c) independently from the failure times, where c was determined to achieve the desired censoring proportions. For each of the configurations studied, 2000 simulations were carried out.

Table 1 presents simulation summary statistics with marginal distribution with $\lambda_0 = 1$ and for the binary covariate where the value of the first member is equal to 1 and the value of the second member is equal to 0 ($Z_{i1} = 1$ and $Z_{i2} = 0$). The "mean $\widehat{\beta}$" denotes the average of the estimates of β_0, "indep. s.e." denotes the average of the estimates of standard errors based on independence assumption, "proposed s.e." denotes the average of the estimates of standard errors based on the proposed method, "true S.D." denotes the sample standard deviation of the 2000 estimates, and "95% coverage" denotes the coverage rate of the nominal 95% confidence interval. Note that the sample size for the case–control sample increases with increasing event proportion in our setup because we sample 90% of the case clusters and the same number of control clusters. The simulation results suggest that the coefficient estimates are approximately unbiased for the samples considered when $\beta_0 = 0$, while the coefficient estimates are relatively biased (4–12%) when $\beta_0 = \log(2)$ with small cohort and sample sizes ($n = 1000$, event proportion = 10%). However, as the cohort size or sample size increases, the coefficient estimates improve and are approximately unbiased. The proposed estimated standard errors provide a very good estimate of the true variability of $\widehat{\beta}$ while standard errors based on independence assumption do not. As expected, the variance of $\widehat{\beta}$ decreases as cohort size or sample size increases. The coverage rate of the nominal 95% confidence intervals using the proposed method are in the 93–96% range in most of the cases considered.

Next, we considered a different simulation setup to evaluate the performance of our proposed methods in situations where the sampling depends on the status of the index member in a cluster and the covariate of interest is a continuous random variable. The covariate values for the first member and the second member were generated independently from the standard normal distribution. Table 2 provides simulation summary statistics for this setup. The findings are similar to those of Table 1.

We have also conducted simulations to compare the estimates with inclusion probability and the local average method. Exponential failure times were generated with $\lambda_0 = 0.4$ and 0.25 for $\beta_0 = 0$ and log(2), respectively. The covariate Z was uniformly distributed on five points $m/5$, $1 \leq m \leq 5$. We considered the situation when the censoring times were dependent on covariates. For each cluster i, the censoring time was generated from uniform distributions on the interval with length 0.4 and centered at $m^*/5$, where m^* was chosen such that it satisfies

$$(m^* - 1)/5 < \sum_{k=1}^{K} Z_{ik}/K \le m^*/5 \quad (m^* = 1, \ldots, 5).$$

Cohort sizes of $n = 1000$ and 2000 were considered. Under these setups, the proportion of failures is about 0.230 when $\beta_0 = 0$ and is about 0.228 when $\beta_0 = \log(2)$. For the local average approach, the partitions of the time interval $[0, \tau)$ were defined as [0, 0.1), [0.1,0.2), [0.2,0.3), [0.3, 0.4), [0.4, 0.5), [0.5, 0.6), [0.6, 0.7), and $[0.7, \tau)$ for case clusters and [0, 0.4), [0.4,0.5), [0.5,0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 0.9), [0.9, 1.0), and $[1.0, \tau)$ for control clusters where τ was set to a value bigger than the maximum value of the failure and censoring times of the first members, say $\max_i(\boldsymbol{T}_{i1}, \boldsymbol{C}_{i1}, i = 1,...,n) + 0.1$. Eighty percent of the case clusters were sampled and the same number of control clusters were sampled. Table 3 displays a comparison between the estimators using inclusion probabilities and local average method. "Mean" $\widehat{\beta}^*$ denotes the average of $\widehat{\beta}_S$ for inclusion probabilities and the average of $\widehat{\beta}_{cS}$ for local averages. Both methods perform reasonably well under the settings considered. The results indicate that the local average method is more efficient than the inclusion probabilities method when the censoring time depends on the covariate, especially when the correlation of the failure times within a cluster is very high ($\theta = 0.1$).

5. Application to the Retrospective Dental Study

We applied our proposed method to data from the retrospective dental study of pulpal involvement and tooth survival described in Section 1. The sample was drawn from the population of enrollees in the Kaiser Permanente Dental Care Program (KPDCP), a dental health maintenance organization (HMO) located in Portland, Oregon, United States (Caplan and Weintraub, 1997). Enrollees are current or retired employees (or their dependents) of companies with dental insurance through KPDCP. As an indicator of pulpal involvement, RCF teeth were used. *Case patients* were defined as those who lost the RCF tooth during the 1987–1994 period, while *control patients* were defined as those who did not lose the RCF tooth during that period. In this case, a patient served as a cluster and the RCF tooth served as the index tooth. After case patients and control patients were sampled, a non-RCF tooth was matched to the RCF tooth within each subject. For a matched non-RCF tooth, the contralateral tooth was selected if it was present. If that tooth was missing or already had RCF on the RCF tooth's access date (index date), the tooth of the same type (anterior, premolar, or molar) adjacent to the contralateral tooth was selected. Note that a *control patient* could have the matched non-RCF tooth lost during the follow-up period. A total of 406 charts were requested, including 232 randomly selected from among 272 case patients, and 174 randomly selected from among 1523 control patients. A total of 202 subjects were identified following the study eligibility criteria. Each of them had one root canal treated (RCT) tooth and a matched non-RCT tooth. Subject- and tooth-level covariates were then ascertained for the RCF tooth and the matching non-RCF tooth from the electronic databases and from radiographs (bitewing, periapical, panoramic) and clinical periodontal recordings taken most recently before the RCF tooth's access date. Databases and charts were examined to determine all treatment received by the study teeth between the index date and December 31, 1994, and the most recent radiograph was examined to validate extraction status. For both RCF and non-RCF teeth, follow-up started on the index date and continued through the date of extraction or December 31, 1994, whichever came first. If an initially non-RCF tooth was accessed endodontically during that interval, the tooth was censored on its endodontic access date.

We applied the proposed method to this data set to investigate the effect of RCF on tooth survival. We also analyzed the data using the unweighted method, where the sampling scheme was not taken into consideration. For the analyses, we included RCF status, tooth type, interaction between RCF status and tooth type, proximal contacts (PCs), and number of pockets

≥ 5 mm as covariates and studied the effect of RCF on tooth survival. Tooth type is molar andnonmolar. There were 176 molars and 228 nonmolars among 202 subjects. PCs are the areas of a tooth that are closely connected with adjacent teeth in the same arch. PCs were divided into four mutually exclusive categories: bridge abutment (PCABUT), nonbridge abutment with zero PCs (PC0), nonbridge abutment with one PC (PC1), and nonbridge abutment with two PCs (PC2). Ninety percent of the sampled teeth fall either in PC1 or PC2. Periodontal pockets are the spaces between the teeth and gums. Pocket depths had been recorded at six sites per tooth. Only periodontal pockets ≥ 5 mm were counted. Two hundred and seventy-nine teeth (70%) did not have any periodontal pockets 5 mm.

Table 4 provides ≥hazard ratio (HR) estimates, the estimated standard errors, and the associated p-values for the proposed method and naive (unweighted) method. The results show strong evidence of significant RCF effect among molars. It indicates that for molars, the hazard rate with RCF is approximately seven times as high as that without RCF. However, no statistically significant effect was seen among nonmolars. The HR estimates for the molars and nonmolars using naive method are biased and are 1.5 to 3 times higher than those using the proposed method. For other variables, the teeth with two PCs and the number of pockets show statistically significant effect. The hazard rate with the teeth with two PCs is approximately one tenth of those with zero PCs. Having one more pocket ≥ 5 mm increases the hazard rate by approximately 30%; however, this effect is marginal (p-value = 0.09).

Figure 1 displays the estimated cumulative hazards functions for RCF tooth and non-RCF tooth with nonbridge abutment with zero PC (PC0) and no periodontal pocket by tooth type. The associated 95% pointwise confidence limits at some selected time points (500, 1000, 1500, 2000, and 2500 days) are also displayed. From the figure, we can see that for molars, the cumulative hazards for RCF tooth are much higher than those for non-RCF tooth; while for nonmolars, the differences are smaller.

6. Concluding Remarks

Motivated by the aforementioned dental study (Caplan and Weintraub, 1997; Caplan et al., 2005), we proposed methods of fitting marginal hazard regression models for the multivariate failure time data from correlated case–control within cohort studies. The primary interest of the study was to evaluate the effect of pulpal involvement on tooth survival. The correlation between two teeth within the same subject is considered as a nuisance. This naturally led us to consider marginal hazard regression models. Weighted estimating equations are proposed for the estimation of the regression parameter. A Breslow–Aalen type estimator is proposed for the cumulative baseline hazard functions. The proposed estimators are shown to be consistent and asymptotically normally distributed. Two types of weights were considered in estimation: the inverse of the inclusion probabilities and the local average. The latter requires the additional information on the observed failure times of all the cohort members. It is more efficient than the inclusion probability estimator when the censoring time is dependent on some covariates which the failure time is also dependent on.

The proposed design can be implemented using existing statistical software such as S-plus or R. For the estimation of the regression parameter, one may use coxph function with weights option. The variance estimator can be obtained using resid function with dfbeta option. An example of S-plus/R script is given in the supplementary material.

We have assumed right-censored survival data throughout this article. A more general study design that allows for left truncation by staggered entry can be easily accommodated. Let L_{ik} be the entry time for the kth member of ith cluster. Then, it can be shown that the main results of this article still hold by using modified $Y_{ik}(t)$, where $Y_{ik}(t) = I(X_{ik} \geq t, L_{ik} < t)$ with slight modification in the derivations.

7. Supplementary Materials

The outline of proofs for Theorems 1 and 2, the consistent estimators for the asymptotic variances, a sample S-plus/R script, and an additional table of simulation results are available from the *Biometrics* website http://www.biometrics.tibs.org.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Daniel Caplan in the College of Dentistry at the University of Iowa and the Kaiser Permanente Dental Care Program for providing KPDCP data. We also thank Guosheng Yin from MD Anderson Cancer Center for helpful discussions. This work was partially supported by the National Institutes of Health grant R01-HL57444.

References

Binder DA. Fitting Cox's proportional hazards models from survey data. Biometrika 1992;79:139–147.

Borgan O, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. The Annals of Statistics 1995;23:1749–1778.

Breslow NE, Wellner JA. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. Scandinavian Journal of Statistics 2007;34:86–102.

Cai J, Prentice R. Estimating equations for hazard ratio parameters based on correlated failure time data. Biometrika 1995;82:151–164.

Cai J, Prentice R. Regression analysis for correlated failure time data. Lifetime Data Analysis 1997;3:197–213. [PubMed: 9384652]

Caplan D, Weintraub J. Factors related to loss of root canal filled teeth. Journal of Public Health Dentistry 1997;57:31–39. [PubMed: 9150061]

Caplan D, Cai J, Yin G, White BA. Root canal filled versus non-root canal filled teeth: A retrospective comparison of survival times. Journal of Public Health Dentistry 2005;65:90–96. [PubMed: 15929546]

Chen K. Generalized case-cohort sampling. Journal of the Royal Statistical Society, Series B 2001;63:791–809.

Chen K, Lo S. Case-cohort and case-control analysis with Cox's model. Biometrika 1999;86:755–764.

Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). Journal of the Royal Statistical Society, Series A 1985;148:82–117.

Clegg LX, Cai J, Sen PK. A marginal mixed baseline hazards model for multivariate failure time data. Biometrics 1999;55:805–812. [PubMed: 11315010]

Cochran, WG. Sampling Techniques. 3rd edition. Wiley; New York: 1977.

Hsu L, Prentice R, Zhao L, Fan J. On dependence estimation using correlated failure time data from case-control family studies. Biometrika 1999;86:743–753.

Hsu L, Chen L, Gorfine M, Malone K. Semiparametric estimation of marginal hazard function from case-control family studies. Biometrics 2004;60:936–944. [PubMed: 15606414]

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2nd edition. John Wiley & Sons; New York: 2002.

Lee, EW.; Wei, LJ.; Amato, DA. Cox-type regression analysis for large numbers of small groups of correlated failure time observations.. In: Klein, JP.; Goel, PK., editors. Survival Analysis: State of the Art. Kluwer Academic Publishers; Dordrecht: 1992. p. 237-247.

Li H, Yang P, Schwartz AG. Analysis of age of onset data from case-control family studies. Biometrics 1998;54:1030–1039. [PubMed: 9750249]

Lin DY. Cox regression analysis of multivariate failure time data: The marginal approach. Statistics in Medicine 1994;13:2233–2247. [PubMed: 7846422]

Lin DY. On fitting Cox's proportional hazards models to survey data. Biometrika 2000;87:37–47.

Lu S, Shih JH. Case-cohort designs and analysis for clustered failure time data. Biometrics 2006;62:1138–1148. [PubMed: 17156289]

Prentice R, Breslow N. Retrospective studies and failure time models. Biometrika 1978;65:153–158.

Rothman, KJ. Epidemiology: An Introduction. Oxford University Press; New York: 2002.

Samuelsen SO. A pseudolikelihood approach to analysis of nested case-cohort studies. Biometrika 1997;84:379–394.

Samuelsen SO, Ånestad H, Skrondal A. Stratified case-cohort analysis of general cohort sampling designs. Scandinavian Journal of Statistics 2007;34:103–119.

Shih JH, Chatterjee N. Analysis of survival data from case-control family studies. Biometrics 2002;58:502–509. [PubMed: 12229984]

Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. Journal of the American Statistical Association 1998;93:1164–1175.

Thomas DC, Liddell FDK, Mcdonald JC, Thomas DC. Addendum to 'Methods of cohort analysis: Appraisal by application to asbestos mining'. Journal of the Royal Statistical Society, Series A 1977;140:483–485.

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 1989;84:1065–1073.

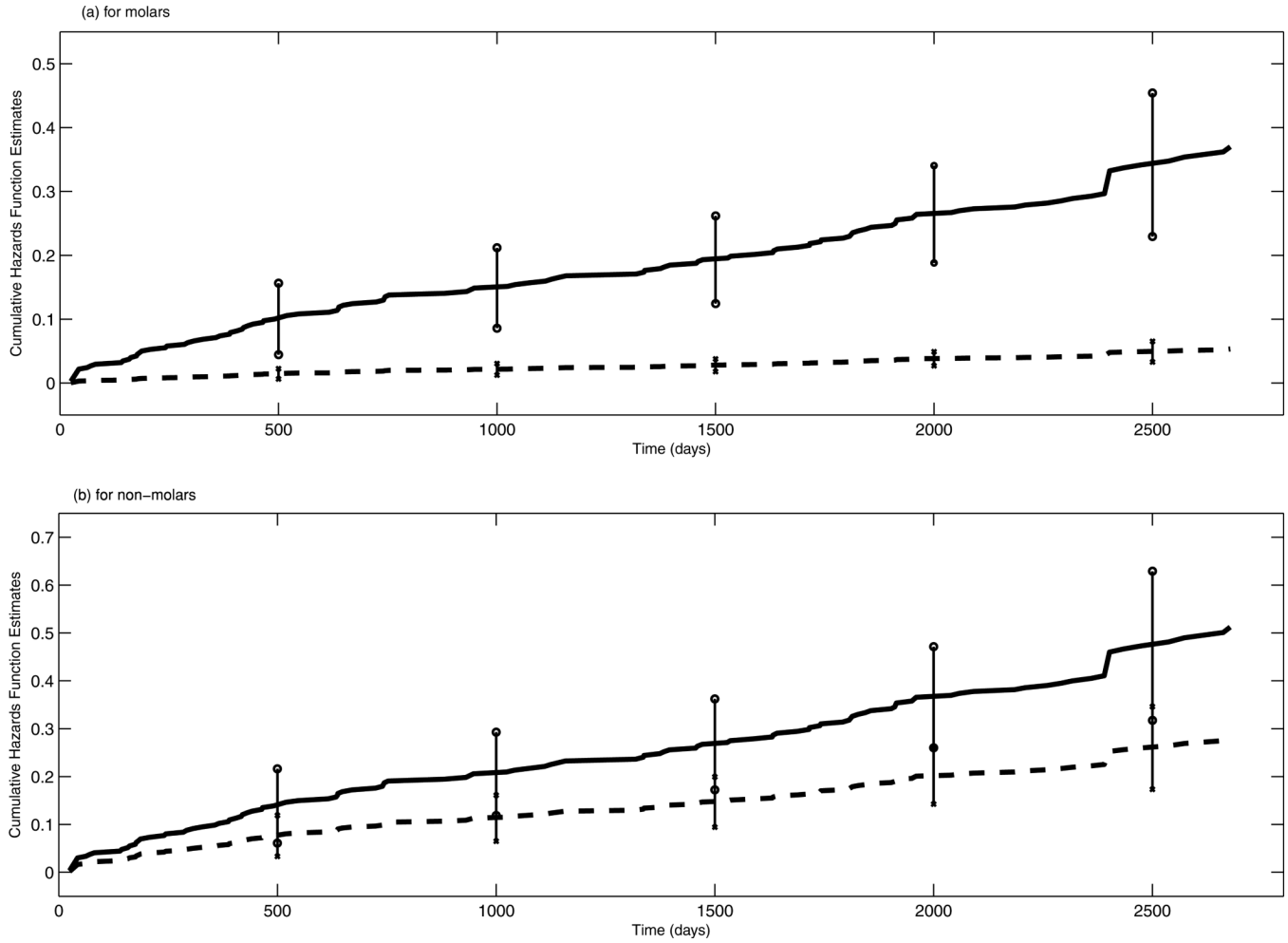

Figure 1.

Cumulative hazards function estimates for RCF tooth and non-RCF tooth by tooth type. Solid lines = cumulative hazards function estimates for RCF tooth; dashed lines = cumulative hazards function estimates for non-RCF tooth; vertical bars with "o"s = 95% pointwise confidence limits for RCF tooth at some selected time points; and vertical bars with "x"s = 95% pointwise confidence limits for non-RCF tooth at some selected time points.

Table 1

Summary of simulation results. $Z_{i1} = 1$ *and* $Z_{i2} = 0$.

β_0	n	Event proportion	ñ	τ_0	Mean $\hat{\beta}$	Indep. s.e.	Proposed s.e.	True S.D.	95% coverage
0	1000	10%	179	0.83	0.019	0.1434	0.2094	0.2177	0.913
			179	0.43	0.035	0.1446	0.2976	0.3096	0.944
			180	0.29	0.048	0.1450	0.3110	0.3268	0.948
			179	0.09	0.038	0.1451	0.3234	0.3337	0.951
		20%	360	0.83	0.001	0.1003	0.0822	0.0839	0.931
			359	0.43	–0.003	0.1003	0.1348	0.1339	0.954
			360	0.29	0.007	0.1004	0.1435	0.1437	0.947
			359	0.09	0.001	0.1005	0.1522	0.1507	0.948
	2000	10%	360	0.83	0.010	0.1005	0.1491	0.1589	0.919
			360	0.43	0.019	0.1009	0.2090	0.2101	0.942
			360	0.29	0.030	0.1013	0.2181	0.2139	0.955
			360	0.09	0.024	0.1011	0.2248	0.2240	0.949
		20%	720	0.83	0.003	0.0708	0.0579	0.0577	0.943
			719	0.43	0.001	0.0708	0.0952	0.0960	0.945
			721	0.29	0.005	0.0708	0.1013	0.1024	0.948
			720	0.09	0.003	0.0709	0.1071	0.1061	0.948
0.693	1000	10%	183	0.83	0.734	0.1741	0.2483	0.2659	0.826
			183	0.43	0.768	0.1795	0.3886	0.4203	0.927
			184	0.29	0.789	0.1810	0.4136	0.4423	0.946
			183	0.09	0.781	0.1818	0.4347	0.4564	0.950
		20%	363	0.83	0.699	0.1195	0.1028	0.1027	0.949
			363	0.43	0.706	0.1201	0.1708	0.1730	0.947
			364	0.29	0.709	0.1201	0.1821	0.1852	0.942
			362	0.09	0.699	0.1200	0.1933	0.1940	0.946
	2000	10%	369	0.83	0.708	0.1209	0.1800	0.1891	0.913
			368	0.43	0.734	0.1229	0.2747	0.2859	0.939
			368	0.29	0.737	0.1232	0.2884	0.2964	0.945
			369	0.09	0.732	0.1232	0.2990	0.3164	0.939
		20%	728	0.83	0.693	0.0842	0.0725	0.7235	0.948

β_0	n	Event proportion	$\tilde{n}$	τ_0	Mean $\hat{\beta}$	Indep. s.e.	Proposed s.e.	True S.D.	95% coverage
			727	0.43	0.694	0.0843	0.1203	0.1172	0.960
			728	0.29	0.702	0.0845	0.1284	0.1286	0.950
			727	0.09	0.698	0.0845	0.1359	0.1363	0.945

Table 2

Summary of simulation results. $Z_{ik} \sim N(0, 1)$.

β_0	n	Event proportion	$\tilde{n}$	τ_0	Mean $\hat{\beta}$	Indep. s.e.	Proposed s.e.	True S.D.	95% coverage
0	1000	10%	179	0.83	0.003	0.0722	0.1359	0.1382	0.946
			179	0.43	0.001	0.0723	0.1558	0.1727	0.916
			180	0.29	–0.005	0.0724	0.1589	0.1679	0.934
			179	0.09	0.002	0.0721	0.1616	0.1653	0.933
		20%	360	0.83	0.002	0.0504	0.0742	0.0744	0.941
			359	0.43	0.000	0.0503	0.0805	0.0816	0.949
			360	0.29	–0.001	0.0503	0.0814	0.0839	0.933
			359	0.09	–0.002	0.0505	0.0833	0.0846	0.948
	2000	10%	360	0.83	–0.001	0.0504	0.0956	0.0952	0.949
			360	0.43	–0.003	0.0505	0.1114	0.1116	0.945
			360	0.29	0.001	0.0507	0.1137	0.1193	0.930
			360	0.09	–0.003	0.0506	0.1157	0.1133	0.951
		20%	720	0.83	0.001	0.0355	0.0524	0.0525	0.950
			720	0.43	–0.001	0.0355	0.0567	0.0566	0.950
			721	0.29	0.001	0.0355	0.0576	0.0583	0.946
			720	0.09	–0.003	0.0355	0.0585	0.0587	0.952
0.693	1000	10%	174	0.83	0.708	0.0763	0.1573	0.1678	0.922
			174	0.43	0.709	0.0767	0.1717	0.1863	0.914
			174	0.29	0.704	0.0763	0.1735	0.1873	0.919
			174	0.09	0.706	0.0761	0.1763	0.1895	0.917
		20%	363	0.83	0.698	0.0526	0.0836	0.0843	0.939
			362	0.43	0.700	0.0527	0.0865	0.0891	0.937
			363	0.29	0.695	0.0525	0.0866	0.0864	0.950
			362	0.09	0.700	0.0527	0.0873	0.0890	0.938
	2000	10%	349	0.83	0.705	0.0529	0.1133	0.1166	0.937
			348	0.43	0.702	0.0531	0.1239	0.1293	0.932
			349	0.29	0.704	0.0529	0.1244	0.1279	0.931
			349	0.09	0.695	0.0530	0.1265	0.1298	0.934
		20%	727	0.83	0.692	0.0370	0.0595	0.0609	0.940

β_0	n	Event proportion	$\tilde{n}$	τ_0	Mean $\hat{\beta}$	Indep. s.e.	Proposed s.e.	True S.D.	95% coverage
			727	0.43	0.694	0.0370	0.0610	0.0617	0.944
			727	0.29	0.696	0.0370	0.0616	0.0594	0.952
			726	0.09	0.695	0.0370	0.0617	0.0635	0.940

Table 3

Summary of simulation results. Comparison between the inclusion probabilities and local averages. The covariate is uniformly distributed on five points, m/5, 1 < m < 5.

n	β_0	ñ	τ_0	Approach	Mean $\hat{\beta}^*$	Proposed s.e.	True S.D.	95% coverage
1000	0	366	0.83	Inclusion probabilities	0.004	0.2407	0.2399	0.949
				Local average	–0.004	0.2214	0.2251	0.945
		367	0.43	Inclusion probabilities	0.001	0.2728	0.2703	0.949
				Local average	0.001	0.2624	0.2674	0.944
		367	0.29	Inclusion probabilities	0.002	0.2779	0.2803	0.953
				Local average	0.004	0.2682	0.2779	0.948
		366	0.09	Inclusion probabilities	–0.006	0.2846	0.2845	0.954
				Local average	–0.009	0.2759	0.2816	0.943
	log(2)	364	0.83	Inclusion probabilities	0.702	0.2478	0.2503	0.950
				Local average	0.699	0.2288	0.2343	0.944
		364	0.43	Inclusion probabilities	0.702	0.2821	0.2864	0.946
				Local average	0.693	0.2708	0.2800	0.940
		364	0.29	Inclusion probabilities	0.706	0.2866	0.2899	0.948
				Local average	0.692	0.2760	0.2839	0.943
		364	0.09	Inclusion probabilities	0.710	0.2928	0.2981	0.952
				Local average	0.697	0.2835	0.2875	0.952
2000	0	734	0.83	Inclusion probabilities	–0.001	0.1697	0.1712	0.949
				Local average	–0.000	0.1568	0.1571	0.949
		734	0.43	Inclusion probabilities	0.004	0.1930	0.1883	0.956
				Local average	0.003	0.1866	0.1839	0.954
		733	0.29	Inclusion probabilities	–0.000	0.1961	0.1940	0.956
				Local average	–0.002	0.1903	0.1912	0.951
		734	0.09	Inclusion probabilities	–0.001	0.2002	0.1962	0.957
				Local average	–0.005	0.1955	0.1946	0.952
	log(2)	729	0.83	Inclusion probabilities	0.693	0.1747	0.1761	0.948
				Local average	0.696	0.1621	0.1641	0.946
		728	0.43	Inclusion probabilities	0.691	0.1992	0.2027	0.943
				Local average	0.695	0.1928	0.1943	0.945

n	β_0	$\tilde{n}$	τ_0	Approach	Mean $\hat{\beta}^*$	Proposed s.e.	True S.D.	95% coverage
		728	0.29	Inclusion probabilities	0.690	0.2023	0.2052	0.953
				Local average	0.694	0.1966	0.1990	0.945
		728	0.09	Inclusion probabilities	0.692	0.2070	0.2104	0.951
				Local average	0.692	0.2020	0.2062	0.946

Table 4

Data analysis for KPDCP data

		Proposed method			Unweighted method		
Variable	**Level**	**HR**	**s.e.**	**p-value**	**HR**	**s.e.**	**p-value**
RCF (molar)		6.9	0.44	<0.01	9.1	0.40	<0.01
RCF (nonmolar)		1.8	0.57	0.30	4.7	0.30	<0.01
PCs							
	PC1	0.3	0.81	0.17	0.5	0.47	0.10
	PC2	0.1	0.81	0.02	0.2	0.46	<0.01
	PCABUT	0.5	0.97	0.44	0.6	0.54	0.33
Number of pockets ≥ 5 mm		1.3	0.15	0.09	1.2	0.09	0.08