Statistica Sinica **22** (2012), 489-508
doi:http://dx.doi.org/10.5705/ss.2010.228

ANALYSIS OF COHORT SURVIVAL DATA WITH TRANSFORMATION MODEL

Kani Chen, Liuquan Sun and Xingwei Tong

Hong Kong University of Science and Technology, Chinese Academy of Sciences and Beijing Normal University

Abstract: Existing methods for analysis of survival data arising from cohort sampling are largely based on Cox's model and pertained to a certain type of sampling design. This paper applies the general linear transformation model, which includes Cox's model and proportional odds model as special cases, to a class of sampling designs including nested case-control, case-cohort and classical case-control designs. A simple likelihood-based method is developed, and the resulting estimator of the regression coefficient is shown to be consistent and asymptotic normal. The computation and inference procedures are straightforward. In addition to the simplicity and generality of the method, it also has minimal loss of efficiency as the observations with missing covariates that are not used contain little information about the regression parameter. The proposed estimation performs well in simulation studies and is applied to analyze the Colorado Plateau uranium miners cohort data.

Key words and phrases: Complete case analysis, generalized case-cohort sampling, linear transformation model, missing at random.

1. Introduction

For reasons such as costs of covariate ascertainment or disease rareness, cohort sampling design becomes an important issue in epidemiological studies and clinical trials. When a time-to-failure response subject to censoring is involved, nested case-control (N-C-C), case-cohort (C-C), and classical case-control (C-C-C) designs are widely used sampling designs, categorized as generalized case-cohort (G-C-C) designs in Chen (2001). This paper proposes a likelihood method of regression analysis of G-C-C sampling via linear transformation models. Throughout the paper, sampling always means sampling for covariate ascertainment.

Consider a cohort of size n. Let $(Y_i, \delta_i, X_i), i = 1, \ldots, n$, be the full cohort data where, for subject i, Y_i represents event time, δ_i failure/censoring index, and X_i the covariate. Let $n_1 = \sum_{i=1}^{n} \delta_i$ be the total number of failures and $n_0 = n - n_1$ the total number of non-failures of the cohort. A C-C-C design takes m_1 failures and m_0 non-failures without replacement; a C-C design takes all n_1

failures and m subjects from the entire cohort without replacement; a N-C-C design takes all n_1 failures and m subjects without replacement from each risk set of a failure time. A G-C-C design, as defined in Chen (2001), consists of K independent sampling steps with the k-th step taking m_k subjects without replacement from a certain specified subcohort. The subcohorts and K must only depend on $(Y_i, \delta_i), i = 1, \ldots, n$. Then, G-C-C covers C-C-C, C-C and N-C-C designs as special cases and offers more flexibility.

There are many publications addressing the regression analysis of cohort data with the above designs; see Thomas (1977), Oakes (1981), Prentice (1986), Self and Prentice (1988), Kalbfleisch and Lawless (1988), Langholz and Thomas (1990, 1991), Goldstein and Langholz (1992), Bednarski (1993), Sasieni (1993), Barlow (1994), Borgan, Goldstein, and Langholz (1995), Langholz and Goldstein (1996), Breslow (1996), Samuelsen (1997), Suissa, Edwards, and Biovin (1998), Chen and Lo (1999), Lawless, Kalbfleisch, and Wild (1999), Chen (2001, 2004), Kulich and Lin (2000, 2004) and Nan, Yu, and Kalbfleisch (2006) among many others. All publications except for Chen (2001) address only one of C-C-C, C-C, or N-C-C designs. In addition, Chen (2001) only considers Cox's proportional hazards model and the estimation method therein cannot be generalized for the more general transformation model. Moreover, most of the articles use Cox's model and little is reported using the transformation model. Indeed, Kong, Cai, and Sen (2004), Zeng et al. (2006), Lu and Tsiatis (2006) and Chen and Zucker (2009) applied the transformation model but only to C-C or N-C-C designs, and these methods take advantage of the simple structure of C-C designs and cannot be readily extended to more general sampling designs.

This paper applies the linear transformation model to the G-C-C designs and proposes a simple likelihood-based estimation method. Compared with the existing ones, the proposed method is superior not only in simplicity but also in generality. It applies the transformation model, more general than Cox's model, to the G-C-C designs. The computation of the estimator, through maximization of a likelihood function, is straightforward. The inference procedure is also easily available with the variance estimator in closed form. Although the method is not based on a full likelihood but based on a likelihood of the complete cases, it does not cause much loss of efficiency. In most practical designs, failures are all sampled while the non-complete cases, the cohort members that are not sampled, are largely censored and censoring times alone, without observed covariates, contain little information about the regression parameter. Since the full likelihood method encounters the curse of dimensionality arising from the dependence of censoring time on covariates and is difficult to implement, the proposed method offers a simple and general alternative which may be near optimal, for example, in a N-C-C design with Cox model and with m controls matching each case. The

relative efficiency of the N-C-C partial likelihood estimation, which is not even the best (e.g., Chen (2004)), can be as high as $m/(m+1)$. If cases are rare, which is often the situation for rare disease or large cohort, the sample used in the N-C-C design has a much smaller size than the size of the full cohort. This implies that the vast majority of censored observations, even with covariate identified, contribute relatively little to improve the accuracy of the estimation. An intuitive point of view is that every censored observation plays only the role of comparison in estimation. Thus, every additional censored observation contributes less to the improvement of the estimation accuracy.

The next section introduces the linear transformation model, proposes the estimation method, and provides theoretical results regarding consistency and asymptotic normality. Some simulation results are reported in Section 3 to evaluate the proposed method and compare it with some existing methods. In Section 4, the method is illustrated with an analysis of the Colorado Plateau uranium miners cohort data. Technical proofs are relegated to the Appendix.

2. Estimation and Inference

Let (T, C) be the pair of failure and censoring times that are conditionally independent given the p-dimensional covariate X. The linear transformation model is

$$\log H(T) = -\beta' X + \epsilon, \tag{2.1}$$

where H is an unknown increasing function with $H(0) = 0$, β is an unknown p-dimensional regression parameter of interest, and ϵ is a continuous random variable with known distribution and is independent of (X, C). The event time $Y = \min(T, C)$ and the censoring index $\delta = I(T \le C)$. Let $(Y_i, \delta_i, X_i, \epsilon_i, C_i)$, for $i = 1, \ldots, n$, be i.i.d copies of $(Y, \delta, X, \epsilon, C)$.

Recall that $K = K_n$ is the number of steps of sampling in a G-C-C design, and note that K is 2, 2, and $n_1 + 1$ for C-C-C, C-C and N-C-C designs, respectively. For $1 \le k \le K$ and $1 \le j \le n$, let Δ_{kj} be the indicator of subject j being sampled for covariate ascertainment at step k. By definition of G-C-C, $\{\Delta_{kj}, j = 1, \ldots, n\}$ are independent of $\{\Delta_{lj}, j = 1, \ldots, n\}$ for $l \ne k$, and $\Delta_j = 1 - \prod_{k=1}^{K}(1 - \Delta_{kj})$ is the index of j being ever sampled. Then, the observed data of a G-C-C design can be written as $(Y_i, \delta_i, \Delta_{ki}, \Delta_i X_i), k = 1, \ldots, K; i = 1, \ldots, n$. Let π_j be the conditional probability of individual j being sampled given the longitudinal data of full cohort, $\pi_j = P\{\Delta_j = 1 \,|\, (Y_i, \delta_i), 1 \le i \le n\}$. Note that a subscript n is suppressed in the notations π_j and Δ_j.

Let $\lambda(\cdot)$ and $\Lambda(\cdot)$ be the known hazard and cumulative hazard functions of e^{ϵ}, respectively. With the linear transformation model (1), the log-likelihood

function of the full cohort data is

$$\sum_{i=1}^{n}\left(\delta_i\Big[\log\lambda\{H(Y_i)e^{\beta' X_i}\}+\beta' X_i+\log h(Y_i)\Big]-\Lambda\{H(Y_i)e^{\beta' X_i}\}\right), \tag{2.2}$$

where $h(\cdot)$ is the derivative function of $H(\cdot)$. Zeng and Lin (2006) considered maximizing the log-likelihood function with a discretization of $H(\cdot)$. Specifically, let q_j represent size of increment of $H(\cdot)$ at the j-th smallest failure times, say s_j, $j=1,\ldots,n_1$. Set

$$H(t)=\sum_{j=1}^{n_1} q_j I(s_j\le t), \qquad \text{and} \qquad h(t)=\sum_{j=1}^{n_1} q_j I(s_j=t).$$

Then the maximization of (2.2) over $(\beta, q_1,\ldots,q_{n_1})$ leads to consistent, asymptotic normal and semiparametric efficient estimators of β.

In a G-C-C design, let $d=\sum_{i=1}^{n}\delta_i\Delta_i$ be the total number of *sampled* failures and t_j be the j-th smallest *sampled* failure time. With G-C-C data, we propose maximizing

$$\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}l_i(\beta,q)\equiv\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\left(\delta_i\Big[\log\lambda\{H(Y_i)e^{\beta' X_i}\}+\beta' X_i+\log h(Y_i)\Big]-\Lambda\{H(Y_i)e^{\beta' X_i}\}\right), \tag{2.3}$$

where $q=(q_1,\ldots,q_d)$,

$$H(t)=\sum_{j=1}^{d} q_j I(t_j\le t), \qquad \text{and} \qquad h(t)=\sum_{j=1}^{d} q_j I(t_j=t).$$

The maximizer is denoted as $(\hat\beta,\hat q_1,\ldots,\hat q_d)$. Let $\hat H(t)=\sum_{j=1}^{d}\hat q_j I(t_j\le t)$.

Remark. The above estimation procedure is essentially a complete case analysis and usually cannot produce efficient estimation. For cohort sampling, however, this may not be a severe drawback. First, without covariate, the event times normally do not contain much information. Second, in practice all or most failures are sampled and the information contained in subjects that are not sampled may be minimal as they are all or largely censoring times with unobserved covariates. Third, the proposed method is based on a likelihood for complete cases, which contain nearly all information about the regression parameter. Moreover, unless there are further restrictive assumptions on the censoring variable, efficient estimation cannot be obtained because of curse of dimensionality involved in the conditional distribution of the censoring variable given the covariate. In addition, the cohort could be loosely defined, and the event times of the subjects that are

not sampled are not as reliable as those that are. More comments may be found in Chen and Lo (1999).

Let β_0 and H_0 be the true values of β and H. Consider the following regularity conditions.

(C1) The function $H_0(t)$ is strictly increasing and continuously differentiable with $H_0(0) = 0$, and β_0 lies in the interior of a compact set $\mathcal{B}$.

(C2) $\lambda(t) > 0$, and $P(Y \geq \tau | X) > 0$.

(C3) X is bounded, and if there exists a vector γ and a deterministic function $\gamma_0(t)$ such that $\gamma_0(t) + \gamma' X = 0$ with probability one, then $\gamma = 0$ and $\gamma_0(t) = 0$.

(C4) For any positive c_0, $\limsup_{x\to\infty}[\log\{x \sup_{y\leq x} \lambda(y)\}/\Lambda(c_0 x)] = 0$.

These conditions are similar to those used by Zeng and Lin (2006) for counting processes, and condition (C4) is only used in the consistency proof of $\hat{\beta}$ and $\hat{H}$.

Theorem 1. *If* (C1)−(C4) *hold,* $|\hat{\beta} - \beta_0| \to 0$ *and* $\|\hat{H} - H_0\|_{l^\infty[0,\tau]} \to 0$ *in probability, where* $\|\cdot\|_{l^\infty[0,\tau]}$ *is the supremum norm in the interval* $[0, \tau]$.

To describe the asymptotic distribution of $\hat{\beta}$ and $\hat{H}$, let $\mathcal{D} = \{\phi(t) : \phi(t) \in BV[0,\tau], \|\phi\|_{BV[0,\tau]} \leq 1\}$, where $BV[0,\tau]$ denotes the set of functions with bounded total variations and $\|\phi\|_{BV[0,\tau]}$ denotes the total variation of $\phi(t)$ in $[0,\tau]$. Hence $\hat{H}$ can be considered as a bounded linear functional in $l^\infty(\mathcal{D})$ by the definition $\hat{H}(\phi) = \int_0^\tau \phi(t) d\hat{H}(t)$. Thus $(\hat{\beta} - \beta_0, \hat{H} - H_0)$ is treated as a random element in the metric space $\mathcal{R}^p \times l^\infty(\mathcal{D})$.

Theorem 2. *If* (C1)−(C4) *hold,* $n^{1/2}(\hat{\beta} - \beta_0, \hat{H} - H_0)$ *converges weakly to a zero-mean Gaussian process in the metric space* $\mathcal{R}^p \times l^\infty(\mathcal{D})$.

In order to obtain a variance estimate of $\hat{\beta}$, let $\dot{l}_i \equiv \dot{l}_i(\beta, q)$ and $\ddot{l}_i \equiv \ddot{l}_i(\beta, q)$ be the first and second derivatives of $l_i(\beta, q)$ with respect to (β, q), respectively. Then, $\dot{l}_i$ is a vector of $p + d$ dimension and $\ddot{l}_i$ is a $(p + d) \times (p + d)$ matrix. Let

$$\pi_{ij} = E\{\Delta_i \Delta_j | (Y_1, \delta_1), \ldots, (Y_n, \delta_n)\}$$

and let τ denote the duration of the study. For any function w with bounded total variation in $[0,\tau]$ and a real vector b, we show in the Appendix that the asymptotic variance of

$$n^{1/2} b'(\hat{\beta} - \beta_0) + n^{1/2} \int_0^\tau w(t) d[\hat{H}(t) - H_0(t)] \tag{2.4}$$

can be estimated by $(b', w')\hat{A}^{-1}\hat{B}\hat{A}^{-1}(b', w')'$, where

$$\hat{A} = n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\ddot{l}_i(\hat{\beta},\hat{q}), \tag{2.5}$$

$$\hat{B} = n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i^2}\dot{l}_i(\hat{\beta},\hat{q})\dot{l}_i(\hat{\beta},\hat{q})' + n^{-1}\sum_{1\le i\ne j\le n}\frac{\Delta_i\Delta_j(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}\pi_i\pi_j}\dot{l}_i(\hat{\beta},\hat{q})\dot{l}_j(\hat{\beta},\hat{q})', \tag{2.6}$$

and $w = (w(t_1), \ldots, w(t_d))'$. Thus, the variance estimator of $n^{1/2}(\hat{\beta} - \beta_0)$ is the upper left $p \times p$ submatrix of $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$. We note that it is not possible, although quite appealing, to derive a sandwich formula without involving a high-dimensional matrix, as in Zeng and Lin (2006), which paper deals with the transformation model with full cohort data.

With the structure of G-C-C sampling, the computation of π_i and π_{ij} are not difficult, although a universal form is not available because it relies on the specification of the subcohorts. For example, with C-C-C data, $\pi_i = m_1/n_1$ (m_0/n_0) if subject i is a failure (non-failure), and

$$\pi_{ij} = \begin{cases} \frac{m_1(m_1-1)}{n_1(n_1-1)} & \text{if subjects } i \text{ and } j \text{ are both failures;} \\ \frac{m_0(m_0-1)}{n_0(n_0-1)} & \text{if subjects } i \text{ and } j \text{ are both non-failures;} \\ \frac{m_1 m_0}{n_1 n_0} & \text{else.} \end{cases}$$

For C-C design, $\pi_i = 1$ (m/n) if subject i is a failure (nonfailure), and

$$\pi_{ij} = \begin{cases} 1 & \text{if subjects } i \text{ and } j \text{ are both failures;} \\ \frac{m(m-1)}{n(n-1)} & \text{if subjects } i \text{ and } j \text{ are both non-failures;} \\ \frac{m}{n} & \text{if one is a failure and the other is not.} \end{cases}$$

For N-C-C design, $\pi_i = 1$ if subject i is a failure, and

$$1 - \prod_{t<Y_i}\Big(1 - \frac{m}{n_t}\Big)^{dN(t)} \qquad \text{if not,}$$

where $n_t = \sum_{j=1}^{n} I(Y_j > t)$ denotes the size of the risk set at time t and, using the counting process notation, $dN(t) = \sum_{i=1}^{n}\delta_i I(Y_i = t)$ and

$$\pi_{ij} = \begin{cases} 1 & \text{if subjects } i \text{ and } j \text{ are both failures;} \\ \pi_j & \text{if } i \text{ is a failure and } j \text{ is not;} \\ \pi_i & \text{if } j \text{ is a failure and } i \text{ is not.} \end{cases}$$

In the case both subjects i and j are non-failures,

$$\pi_{ij} = \pi_i + \pi_j - 1 + \prod_{t<\min(Y_i,Y_j)} \left(\frac{n_t - m - 1}{n_t - 1}\right)^{dN(t)} \prod_{t<\max(Y_i,Y_j)} \left(\frac{n_t - m}{n_t}\right)^{dN(t)}.$$

3. Simulation Study

Extensive simulation studies have been carried out to assess the performance of the likelihood-based estimators for C-C-C, C-C and N-C-C designs. The linear transformation model was specified as

$$\log H(T) = -\beta_1 X_1 - \beta_2 X_2 + \epsilon,$$

where $H(t) = t/2$, X_1 Bernoulli with success probability 0.5 and X_2 uniform on $[0, 1]$. We set $\beta_1 = 1$ and $\beta_2 = -1$. The hazard function of ϵ was

$$\lambda_0(t) = \frac{\exp(t)}{1 + r\,\exp(t)},$$

with $r = 0, 0.5, 1$ and 2 (Dabrowska and Doksum (1988); Chen, Jin, and Ying (2002)). Note that the proportional hazards and proportional odds models correspond to $r = 0$ and $r = 1$, respectively. The censoring time C was independent of the covariates, exponential with parameter adjusted for a censoring rate of about 80%. In other words, about 20% of all the failure times were observed. The sample size n was set to be 500 and all simulations were based on 1,000 replications. For C-C design, we took all n_1 failures and $m = 0.2n$ subjects from the entire cohort without replacement. For C-C-C design, we took $m_1 = n_1$ for failures, and $m_0 = n_1$ for non-failures. For N-C-C design, we took $m = 2$.

Table 1 below summarizes the simulation results of the estimation of β_1 and β_2. It includes the averages (Mean), sample standard deviations (SSD), and averages of the estimated standard errors (ESE) of the estimates. It also contains the coverage probabilities (CP) for β_1 and β_2 at level 95%. It is seen that the proposed estimation procedures performed well in all cases. The bias of the estimation was negligible. The estimated and empirical standard errors agreed with each other. The coverage probabilities were generally close to the nominal level 95%. Other simulation studies showed similar results.

We compared the proposed method with widely used methods in the literature, such as the partial likelihood of Thomas (1977) for N-C-C sampling, the pseudo-likelihood estimation of Prentice (1986) for C-C sampling, and the inclusion probability method of Samuelsen (1997) for N-C-C sampling. The methods of Thomas and Prentice cannot be generalized to treat G-C-C sampling while

Table 1. Summary of simulation results.

C-C design

r	Mean		SSD		ESE		CP	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
0	1.031	-1.046	0.306	0.531	0.313	0.525	0.949	0.937
0.5	1.015	-1.038	0.325	0.536	0.334	0.568	0.952	0.950
1	1.019	-1.035	0.336	0.560	0.349	0.596	0.959	0.957
2	1.009	-1.026	0.367	0.619	0.377	0.646	0.961	0.956

C-C-C design

r	Mean		SSD		ESE		CP	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
0	1.022	-1.032	0.287	0.487	0.292	0.491	0.952	0.950
0.5	1.019	-1.033	0.299	0.496	0.313	0.532	0.961	0.960
1	1.019	-1.014	0.308	0.513	0.325	0.555	0.961	0.961
2	1.011	-1.023	0.336	0.567	0.355	0.607	0.967	0.958

N-C-C design

r	Mean		SSD		ESE		CP	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
0	1.014	-1.016	0.224	0.386	0.216	0.382	0.931	0.933
0.5	1.010	-1.019	0.246	0.395	0.257	0.440	0.960	0.964
1	1.015	-1.017	0.256	0.425	0.276	0.472	0.965	0.959
2	1.009	-1.021	0.294	0.488	0.302	0.524	0.954	0.961

that of Samuelsen can. The local average method of Chen (2001) for G-C-C sampling was also considered for comparison. We note that all the four methods were designed for Cox's model rather than the transformation model. The setup of the simulation was similar to that reported in Table 1. The bias (BIAS) and sample standard errors (SSD) of the estimation are reported in Table 2.

It is seen in Table 2 that, when Cox's model is true ($r = 0$), the proposed estimators are comparable to the existing ones in the sense that the bias and standard errors are close to one another. When r is increased from 0 to 2, implying the true model deviates from Cox's model, the bias of the proposed estimators remains very small, but those of other estimators all become rather large. As a result, the existing methods become invalid under a transformation model other than Cox's model. On the other hand, the proposed method performs well in all cases.

4. Application: Colorado Plateau Uranium Miners Cohort

In this section, we consider the application of the proposed method to the Colorado Plateau uranium miners cohort data. This data set was gathered for

Table 2. Comparing BIAS (SSD) of the proposed method with existing ones ($\times 10^3$).

C-C design

	$r=0$		$r=0.5$		$r=1$		$r=2$	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Proposed	31(306)	-46(531)	15(325)	-38(536)	19(336)	-35(560)	9(367)	-26(619)
Chen	26(320)	-33(563)	-66(328)	57(562)	-127(318)	131(559)	-234(317)	237(566)
Prentice	15(338)	-28(621)	-77(337)	61(602)	-137(330)	137(595)	-244(324)	240(583)
Samuelsen	36(323)	-42(566)	-59(329)	50(566)	-122(319)	126(561)	-232(318)	233(567)

C-C-C design

	$r=0$		$r=0.5$		$r=1$		$r=2$	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Proposed	22(287)	-32(487)	19(299)	-33(496)	19(308)	-14(513)	11(336)	-23(567)
Chen	21(304)	-20(522)	-60(305)	63(524)	-130(291)	148(509)	-235(288)	238(515)
Samuelsen	28(306)	-28(525)	-56(305)	59(527)	-126(292)	145(510)	-233(289)	236(516)

N-C-C design

	$r=0$		$r=0.5$		$r=1$		$r=2$	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Proposed	14(224)	-16(386)	10(246)	-19(395)	15(256)	-17(425)	9(294)	-21(488)
Chen	27(248)	-23(431)	-65(255)	66(430)	-131(242)	145(423)	-238(247)	249(431)
Thomas	83(217)	-95(474)	-17(267)	74(463)	- 87(257)	89(455)	-203(259)	209(463)
Samuelsen	17(244)	-14(420)	-72(249)	74(418)	-136(237)	151(413)	-242(243)	251(421)

Full cohort data

	$r=0$		$r=0.5$		$r=1$		$r=2$	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Full^{lt}	8(196)	-7(328)	3(211)	-10(334)	7(222)	-10(363)	3(255)	-11(416)
Full^{cox}	10(219)	-8(360)	-79(219)	84(358)	-144(210)	158(358)	-248(213)	259(361)

'Proposed','Chen','Thomas','Prentice','Samuelsen' refer to the estimators proposed by our method, Chen (2001), Thomas (1977), Prentice (1986), Samuelsen (1997), respectively. Full^{lt} and Full^{cox} refer to the estimators proposed by linear transformation model and proportional hazards model using the full cohort data, respectively.

the study of the effects of radon exposure and smoking on the rates of lung cancer; it has been described in detail in Lubin et al. (1994), Langholz and Goldstein (1996), and Langholz et al. (1999). Here we compare the results using the full cohort analysis to those based on C-C-C, C-C and N-C-C designs.

The cohort consisted of 3347 (n) Caucasian male miners who worked underground at least one month in the uranium mines of the four-state Colorado Plateau area. For each subject, the information included the age at entry to the study, the age at exit from the study, the death time if death occurred during the study, the cumulative radon exposure, and cumulative smoking in number of packs and the death information. In this study, a total of 258 (n_1) miners died of

lung cancer. Subjects who died of lung cancer were taken to be the failures and all other were censored at their exit times. Let X_1 denote the cumulative radon exposure in 100 working level months (WLMs), X_2 be the cumulative smoking in 1,000 packs, and $X = (X_1, X_2)'$.

For the analysis, we used the model

$$\log H(T) = -g(X;\beta) + \epsilon, \tag{4.1}$$

where the hazards function of ϵ is $\exp(t)/(1 + r\exp(t))$ with r unknown, and the function g was used to describe different models. Following Thomas et al. (1994) and Langholz and Goldstein (1996), we considered four models of g as a function of radon and smoking:
Radon:

$$g(X;\beta) = \beta_1 X_1; \tag{4.2}$$

Smoking:

$$g(X;\beta) = \beta_2 X_2; \tag{4.3}$$

Radon and smoking:

$$g(X;\beta) = \beta_1 X_1 + \beta_2 X_2; \tag{4.4}$$

Interaction:

$$g(X;\beta) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2. \tag{4.5}$$

With $\pi_i = 1$ and $\pi_{ij} = 1$ for $i, j = 1, \cdots, n$, we obtain the full cohort analysis. For the full cohort data, the choices of r were determined as follows. First, for given $r \geq 0$, we used the proposed estimation procedure in Section 2 (e.g., Zeng and Lin (2006)) to obtain the estimators $\hat{\beta}$ and $\hat{H}$. Second, we computed the estimated observed log-likelihood $\hat{l}(r; \hat{\beta}, \hat{H})$ defined in (2.2). Finally, we chose $\hat{r}$ as the estimation of r, where $\hat{r}$ maximizes the $\hat{l}$. The estimates of r for the four models are included in Table 3.

The results of fitting the four models (4.2)−(4.5) using the full cohort data and the three G-C-C sampling data are given in Table 3. The regression results using the full cohort data suggest strong association between radon and smoking and lung cancer mortality rates. For the interaction model (4.5), the interaction parameter β_3 is negative with p-value of 10^{-11}, there is significant evidence of the joint negative effect of the two exposures on the mortality rates. For all the G-C-C sampling, we used the same setups as in the simulation studies, and took all cases of 258 failures. That is, we took $m = 0.2n$ for the C-C design, $m_1 = n_1$ and $m_0 = n_1$ for the C-C-C design, and $m = 2$ for the N-C-C design. For each design, we sampled the data 1,000 times and obtained the averages of the parameter estimates, their standard errors and p-values. We drew similar

Table 3. Comparison of parameter estimation (Standard error, approximated p-value) for radon and smoking models using different sampling methods [a].

Univariate models

	Radon (β_1)[b]	Smoking (β_2)[c]
Full cohort	0.04286 (0.00177, 1.623e-129)	6.665 (0.3582, 3.012e-77)
C-C	0.04128 (0.00449, 2.527e-013)	7.089 (0.7120, 4.587e-11)
C-C-C	0.04247 (0.00513, 3.069e-007)	7.461 (0.9703, 1.849e-06)
N-C-C	0.03495 (0.00357, 1.975e-020)	7.036 (0.7521, 1.877e-09)

Adjusted model[d]

	Radon (β_1)	Smoking (β_2)
Full cohort	0.03899 (0.00206, 3.734e-80)	6.959 (0.3845, 3.303e-73)
C-C	0.04002 (0.00484, 4.028e-09)	7.190 (0.7299, 5.467e-14)
C-C-C	0.04104 (0.00540, 3.282e-05)	7.519 (0.9882, 4.263e-08)
N-C-C	0.02720 (0.00395, 1.044e-10)	6.736 (0.6992, 1.250e-12)

Interaction model[e]

	Radon (β_1)	Smoking (β_2)	Interaction (β_3)
Full cohort	0.06043 (0.0030, 8.96e-90)	8.552 (0.363, 1.78e-122)	-0.1292 (0.0199, 9.62e-11)
C-C	0.06657 (0.0069, 1.98e-10)	9.508 (0.831, 5.33e-13)	-0.1590 (0.0426, 0.0238)
C-C-C	0.07066 (0.0077, 5.59e-07)	10.31 (1.092, 4.03e-08)	-0.1753 (0.0472, 0.0322)
N-C-C	0.05406 (0.0067, 1.22e-13)	9.244 (0.865, 3.19e-11)	-0.1576 (0.0379, 9.21e-4)

[a] Randon slopes given as per 100 WLMs. Smoking slopes given as per 1,000 cumulative packs of cigarettes.
[b] Univariate radon: $g(X;\beta) = \beta_1 X_1, \hat{r} = 0$.
[c] Univariate smoking: $g(X;\beta) = \beta_2 X_2, \hat{r} = 0.05$.
[d] Adjusted model: $g(X;\beta) = \beta_1 X_1 + \beta_2 X_2, \hat{r} = 0$.
[e] Interaction model: $g(X;\beta) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2, \hat{r} = 0$.

conclusions about the statistical relationships between Radon, smoking, their joint exposure, and cancer mortality rates.

For comparisons, we also show the results of fitting model (4.4) in Table 4 for $r = 0$, $r = 0.5$, and $r = 1$. Note that the proportional hazards and proportional odds models correspond to $r = 0$ and $r = 1$, respectively. From Table 4, we see results similar to those in Table 3.

Our study shows that analysis of G-C-C designs using the proposed estimation method can effectively draw the same conclusion as that of full cohort data, while saving the costs of covariate ascertainment. In many practical situations, accurately identifying the covariates, such as genotype, for every individual of a large cohort can be quite expensive, especially in the case of rare diseases. In these situations, G-C-C designs may be the ideal alternative, and this paper provides a statistical methodology for data analysis through linear transformation

Table 4. Comparison of parameter estimation (Standard error, approximated p-value) for adjusted model using different r.

$r = 0$, Proportional hazards model

	Radon (β_1)	Smoking (β_2)
Full cohort	0.03899 (0.00206, 3.734e-80)	6.959 (0.3845, 3.303e-73)
C-C	0.04002 (0.00484, 4.028e-09)	7.190 (0.7299, 5.467e-14)
C-C-C	0.04104 (0.00540, 3.282e-05)	7.519 (0.9882, 4.263e-08)
N-C-C	0.02720 (0.00395, 1.044e-10)	6.736 (0.6992, 1.250e-12)

$r = 0.5$

	Radon (β_1)	Smoking (β_2)
Full cohort	0.05594 (0.00345, 5.514e-59)	8.439 (0.4457, 6.103e-80)
C-C	0.05700 (0.00943, 9.522e-06)	8.496 (0.8664, 5.258e-13)
C-C-C	0.05956 (0.01088, 0.002077)	8.931 (1.1910, 1.579e-06)
N-C-C	0.05728 (0.00889, 7.011e-09)	11.36 (0.9637, 6.210e-20)

$r = 1$, Proportional odds model

	Radon (β_1)	Smoking (β_2)
Full cohort	0.07150 (0.00423, 5.497e-64)	9.833 (0.4990, 2.007e-86)
C-C	0.07280 (0.01075, 2.608e-06)	9.998 (0.9579, 4.609e-15)
C-C-C	0.07562 (0.01418, 0.001579)	10.33 (1.3560, 2.406e-06)
N-C-C	0.08391 (0.01092, 8.965e-13)	15.55 (1.1700, 2.397e-30)

models.

5. Concluding Remarks

The existing statistical methodologies are either focused on a special type of cohort sampling or are valid only under Cox's model. This paper presents an effective and unified approach to a broad of class of sampling designs (G-C-C) using linear transformation models. The computation procedure and variance estimation are straightforward. The estimation and inference proposed in this paper can also be generalized in a straightforward fashion to slightly more general models with $H(t) = g(\beta' X, \epsilon)$, where g is a known smooth function.

Acknowledgement

The authors thank the Editor, Professor Kung-Yee Liang, an associate editor, and a referee for their insightful comments and suggestions that greatly improved the article. Kani Chen's research was supported by Hong Kong RGC grants 600307 and 600509. Liuquan Sun's research was supported by the National Natural Science Foundation of China Grants (No. 11171330, 10731010, 10971015 and 11021161), the National Basic Research Program of China (No. 2007CB814902 of

Program 973) and Key Laboratory of RCSDS, CAS (No.2008DP173182). Xingwei Tong 's research was supported by the National Natural Science Foundation of China Grant (No. 10971015).

Appendix: Proofs of Theorems

Proof of Theorem 1. Mimicking the consistency proof of Zeng and Lin (2006), we first show that the jump sizes of $\hat{H}$ are finite. Note that $\sup_{\beta\in\mathcal{B},t\in[0,\tau]}|\beta' X|\le d_0$ almost surely, where d_0 is a constant. Then it follows from that the ith term in (2.3) is bounded above by

$$\frac{\Lambda(H(Y_i)e^{-d_0})}{\pi_i}\Big[\frac{\log\{H(Y_i)e^{d_0}\sup_{y\le H(Y_i)e^{d_0}}\lambda(y)\}}{\Lambda(H(Y_i)e^{-d_0})}-1\Big],$$

which diverges to $-\infty$ by (C4) if $H(Y_i)$ is infinite for some Y_i. Thus, the jump sizes of $\hat{H}$ must be finite. Next we show that $\hat{H}$ is bounded almost surely. For this, let

$$l_n(\beta,H)=\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\bigg(\delta_i\Big[\log\lambda\{H(Y_i)e^{\beta' X_i}\}+\beta' X_i+\log h(Y_i)\Big]-\Lambda\{H(Y_i)e^{\beta' X_i}\}\bigg).$$

Since $l_n(\beta,H)$ is maximized at $(\hat{\beta},\hat{H})$, we have

$$n^{-1}[l_n(\hat{\beta},\alpha_n\bar{H})-l_n(\hat{\beta},\bar{H})]\ge 0, \tag{A.1}$$

where $\alpha_n=\hat{H}(\tau)$ and $\bar{H}=\hat{H}/\alpha_n$. From (A.1), we obtain that

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\Big[\delta_i\log\big\{\alpha_n\lambda\{\alpha_n\bar{H}(Y_i)e^{\hat{\beta}' X_i}\big\}-\Lambda\{\alpha_n\bar{H}(Y_i)e^{\hat{\beta}' X_i}\}\Big]\\&\quad\ge n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\Big[\delta_i\log\lambda\{\bar{H}(Y_i)e^{\hat{\beta}' X_i}\}-\Lambda\{\bar{H}(Y_i)e^{\hat{\beta}' X_i}\}\Big].\end{aligned} \tag{A.2}$$

Note that the right-hand side of (A.2) is bounded from below by

$$n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\Big[\delta_i\log\{\min_{y\le e^{d_0}}\lambda(y)\}-\Lambda(e^{d_0})\Big]>-\infty.$$

However, the left-hand side is bounded from above by

$$n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\Big[\log\{\alpha_n\sup_{y\le\alpha_n e^{d_0}}\lambda(y)\}-\Lambda(\alpha_n e^{-d_0})I(Y_i\ge\tau)\Big].$$

Suppose that $\alpha_n \to \infty$ for some subsequence. Condition (C4) implies that for any $\nu > 0$ when n is sufficiently large,

$$\log\{\alpha_n \sup_{y \le \alpha_n e^{d_0}} \lambda(y)\} \le \nu \Lambda(\alpha_n^{-d_0}).$$

Thus

$$n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\pi_i} \big[\nu - I(Y_i \ge \tau)\big] \Lambda(\alpha_n e^{-d_0}) > -\infty.$$

If we choose ν such that $\nu \le P(Y \ge \tau)/2$, the left-hand side diverges to $-\infty$ when $\alpha_n \to \infty$. This is a contradiction. Hence $\hat{H}$ is bounded with probability one. Then Helly's Selection Theorem yields that there exists a convergent subsequence such that $\hat{\beta} \to \beta^*$ and $\hat{H} \to H^*$ weakly. Finally we show that $\beta^* = \beta_0$ and $H^* = H_0$. By taking derivatives of $l_n(\beta, H)$ with respect to $h(Y_i)$ to zero, we get

$$\hat{H}(t) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\pi_i} \int_0^t \frac{dN_i(u)}{\Phi_n(u, \hat{\beta}, \hat{H})},$$

where

$$\begin{aligned}\Phi_n(u, \beta, H) = {} & n^{-1} \sum_{j=1}^{n} \frac{\Delta_j}{\pi_j} \lambda\{H(Y_j) e^{\beta' X_j}\} e^{\beta' X_j} I(Y_j \ge u) \\ & - n^{-1} \sum_{j=1}^{n} \frac{\Delta_j}{\pi_j} \frac{\dot{\lambda}\{H(Y_j) e^{\beta' X_j}\}}{\lambda\{H(Y_j) e^{\beta' X_j}\}} \delta_j e^{\beta' X_j} I(Y_j \ge u),\end{aligned}$$

and the superscript dot denotes derivative. Let

$$\tilde{H}(t) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\pi_i} \int_0^t \frac{dN_i(u)}{\Phi_n(u, \beta_0, H_0)}.$$

It follows from Proposition in Appendix 1 of Kulich and Lin (2000) that $\tilde{H}(t)$ converges to $H_0(t)$ uniformly in t in probability. Also, following Step 3 in the Appendix of Zeng and Lin (2006), we have that $\hat{H}(t)$ is absolutely continuous with respect to $\tilde{H}(t)$, and that $d\hat{H}(t)/d\tilde{H}(t)$ converges to a bounded measurable function. Thus, $H^*(t)$ is absolutely continuous with respect to Lebesgue measure, its derivative is denoted as $h^*(t)$. Note that $l_n(\hat{\beta}, \hat{H}) - l_n(\beta_0, \tilde{H}) \ge 0$. By taking the limits on both sides, we obtain that the Kullback-Leibler information between the density indexed by (β^*, H^*) and the true density is negative. Therefore, with probability one,

$$\begin{aligned}& \delta\Big[\log \lambda\{H(Y) e^{\beta^{*\prime} X}\} + \beta^{*\prime} X + \log h^*(Y)\Big] - \Lambda\{H(Y) e^{\beta^{*\prime} X}\} \\ & \quad = \delta\Big[\log \lambda\{H_0(Y) e^{\beta_0' X}\} + \beta_0' X + \log h_0(Y)\Big] - \Lambda\{H_0(Y) e^{\beta_0' X}\},\end{aligned}$$

where $h_0(\cdot)$ is the derivative function of $H_0(\cdot)$. Equality holds when $\delta = 0$, and also when $\delta = 1$. The difference between the equalities from these two cases entails that $\Lambda\{H(Y)e^{\beta^{*'}X}\} = \Lambda\{H_0(Y)e^{\beta_0'X}\}$. Thus,

$$H(Y)e^{\beta^{*'}X} = H_0(Y)e^{\beta_0'X}.$$

It then follows from Condition (C3) that $\beta^* = \beta_0$ and $H^* = H_0$. Hence we have shown that $\hat{\beta} \to \beta_0$ and $\hat{H}(t) \to H_0(t)$ in probability. The continuity and monotonicity of H_0 imply that the convergence of $\hat{H}(t)$ can be strengthened to uniform convergence in $t \in [0, \tau]$.

Proof of Theorem 2. We choose ρ small enough and let $\mathcal{A} = \{(\beta, H) : |\beta - \beta_0| < \rho, \|H - H_0\|_{l^\infty[0,\tau]} < \rho\}$. Define a map $W_n = (W_{n1}, W_{n2})$ from $\mathcal{A}$ to $\mathcal{R}^p \times l^\infty(\mathcal{D})$ as follows: for any $\phi(t) \in \mathcal{D}$,

$$\begin{aligned}
W_{n1}(\beta, H) &= n^{-1}\frac{\partial l_n(\beta, H)}{\partial \beta} \\
&= n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\left(\delta_i\frac{\dot{\lambda}\{H(Y_i)e^{\beta' X_i}\}}{\lambda\{H(Y_i)e^{\beta' X_i}\}}e^{\beta' X_i} + 1 - \lambda\{H(Y_i)e^{\beta' X_i}\}e^{\beta' X_i}\right)X_i, \\
W_{n2}(\beta, H)[\phi] &= n^{-1}\frac{\partial l_n(\beta, H(t) + \varepsilon\int_0^t \phi(u)dH(u))}{\partial \varepsilon}|_{\varepsilon=0} \\
&= n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i}\left(\delta_i e^{\beta' X_i}\frac{\dot{\lambda}\{H(Y_i)e^{\beta' X_i}\}}{\lambda\{H(Y_i)e^{\beta' X_i}\}}\int_0^{Y_i}\phi(u)dH(u)\right. \\
&\qquad \left. + \delta_i\phi(Y_i) - e^{\beta' X_i}\lambda\{H(Y_i)e^{\beta' X_i}\}\int_0^{Y_i}\phi(u)dH(u)\right).
\end{aligned}$$

Let W_1 and W_2 be the limits of W_{n1} and W_{n2}, respectively, and $W = (W_1, W_2)$. Clearly, $W_n(\hat{\beta}, \hat{H}) = 0$ and $W(\beta_0, H_0) = 0$. By Proposition in Appendix 1 of Kulich and Lin (2000),

$$n^{1/2}(W_n - W)(\hat{\beta}, \hat{H}) - n^{1/2}(W_n - W)(\beta_0, H_0) = o_p(1)$$

in the metric space $\mathcal{R}^p \times l^\infty(\mathcal{D})$. Following the proof of weak convergence in the Appendix of Zeng and Lin (2006), it can be verified that W is Fréchet-differentiable at (β_0, H_0) and that the derivative is continuously invertible in the set $\mathcal{A}$. Thus, it follows from Theorem 3.3.1 of van der Vaart and Wellner (1996) that $n^{1/2}(\hat{\beta} - \beta_0, \hat{H} - H_0)$ converges weakly to a zero-mean Gaussian process in the metric space $\mathcal{R}^p \times l^\infty(\mathcal{D})$. Furthermore,

$$n^{1/2}\dot{W}\begin{pmatrix}\hat{\beta} - \beta_0 \\ \hat{H} - H_0\end{pmatrix}\left[\begin{pmatrix} b \\ \phi\end{pmatrix}\right] = n^{-1/2}\left(l_\beta^{(n)}(\beta_0, H_0)'b + l_H^{(n)}(\beta_0, H_0)[\int \phi dH_0]\right) + o_p(1), \tag{A.3}$$

where $\dot{W}$ is the Fréchet derivative of W at (β_0, H_0), $l_\beta^{(n)}(\beta, H)$ is the score vector for β, and

$$l_H^{(n)}(\beta, H)[\phi] = \lim_{\varepsilon \to 0} \frac{l_n(\beta, H + \varepsilon\phi) - l_n(\beta, H)}{\varepsilon}.$$

Asymptotic variance for (2.4)**.** Let

$$l_{HH}^{(n)}(\beta, H)[\phi_1, \phi_2] = \lim_{\varepsilon \to 0} \frac{l_H^{(n)}(\beta, H + \varepsilon\phi_2)[\phi_1] - l_H^{(n)}(\beta, H)[\phi_1])}{\varepsilon},$$

and $l_{\beta\beta}^{(n)}(\beta, H)$ denote the Hessian matrix of $l_n(\beta, H)$ with respect to β, with $l_{\beta H}^{(n)}(\beta, H)[\phi]$ and $l_{H\beta}^{(n)}(\beta, H)[\phi]$ defined similarly. Let $l_{\beta\beta}(\beta, H)$, $l_{\beta H}(\beta, H)[\phi]$, $l_{H\beta}(\beta, H)[\phi]$, and $l_{HH}(\beta, H)[\phi_1, \phi_2]$ be the limits of $n^{-1} l_{\beta\beta}^{(n)}(\beta, H)$, $n^{-1} l_{\beta H}^{(n)}(\beta, H)[\phi]$, $n^{-1} l_{H\beta}^{(n)}(\beta, H)[\phi]$ and $n^{-1} l_{HH}^{(n)}(\beta, H)[\phi_1, \phi_2]$, respectively. A straightforward calculation yields, for any (β, H) and (b, ϕ),

$$\dot{W}\begin{pmatrix} \beta - \beta_0 \\ H - H_0 \end{pmatrix}\left[\begin{pmatrix} b \\ \phi \end{pmatrix}\right] = -\begin{pmatrix} l_{\beta\beta}(\beta_0, H_0) & l_{\beta H}(\beta_0, H_0) \\ l_{H\beta}(\beta_0, H_0) & l_{HH}(\beta_0, H_0) \end{pmatrix}\left[\begin{pmatrix} \beta - \beta_0 \\ H - H_0 \end{pmatrix}, \begin{pmatrix} b \\ \int \phi dH_0 \end{pmatrix}\right],$$

which, combined with (A.3), implies

$$\begin{aligned} &-\begin{pmatrix} l_{\beta\beta}(\beta_0, H_0) & l_{\beta H}(\beta_0, H_0) \\ l_{H\beta}(\beta_0, H_0) & l_{HH}(\beta_0, H_0) \end{pmatrix}\left[\begin{pmatrix} n^{1/2}(\hat{\beta} - \beta_0) \\ n^{1/2}(\hat{H} - H_0) \end{pmatrix}, \begin{pmatrix} b \\ \int \phi dH_0 \end{pmatrix}\right] \\ &\quad = n^{-1/2}\left(l_\beta^{(n)}(\beta_0, H_0)'b + l_H^{(n)}(\beta_0, H_0)[\int \phi dH_0]\right) + o_p(1). \end{aligned} \tag{A.4}$$

This approximation holds uniformly for ϕ with bounded variation and b with bounded norm. Take $\tilde{H}_0(t)$ as a step function with jumps at the *sampled* failure times $\{t_1, \ldots, t_d\}$ with jump size at t_j equal to $\tilde{h}(t_j) \equiv H_0(t_j) - \max_{t_k < t_j} H_0(t_k)$. Clearly, $\tilde{H}_0(t_j) = H_0(t_j)$. For any bounded vector $\{p_1, \ldots, p_d\}$ and bounded vector $b \in \mathcal{R}^p$, let the step function $p(t)$ jump only at t_j with $p(t_j) = p_j$, and let η be the vector consisting of $p_j \hat{h}(t_j)$, where $\hat{h}(t) = \sum_{j=1}^d \hat{q}_j I(t_j = t)$. By the definition of $\hat{A}$ in (2.5),

$$(b', \eta')\hat{A}\begin{pmatrix} b \\ \eta \end{pmatrix} = -n^{-1}\begin{pmatrix} l_{\beta\beta}^{(n)}(\hat{\beta}, \hat{H}) & l_{\beta H}^{(n)}(\hat{\beta}, \hat{H}) \\ l_{H\beta}^{(n)}(\hat{\beta}, \hat{H}) & l_{HH}^{(n)}(\hat{\beta}, \hat{H}) \end{pmatrix}\left[\begin{pmatrix} b \\ \int p d\hat{H} \end{pmatrix}, \begin{pmatrix} b \\ \int p d\hat{H} \end{pmatrix}\right]$$

which converges to

$$-\begin{pmatrix} l_{\beta\beta}(\beta_0, H_0) & l_{\beta H}(\beta_0, H_0) \\ l_{H\beta}(\beta_0, H_0) & l_{HH}(\beta_0, H_0) \end{pmatrix}\left[\begin{pmatrix} b \\ \int p dH_0 \end{pmatrix}, \begin{pmatrix} b \\ \int p dH_0 \end{pmatrix}\right] > 0$$

uniformly in any bounded function $p(t)$ and b. This means that $\hat{A}$ is positive definite for large n. On the other hand, let $\Delta\hat{H} = (\hat{h}(t_1), \ldots, \hat{h}(t_d))'$ and $\Delta\tilde{H}_0 = (\tilde{h}(t_1), \ldots, \tilde{h}(t_d))'$. Then it follows from (A.4) that

$$\begin{aligned}
&-n^{1/2}(b', \eta')\hat{A}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \Delta\hat{H} - \Delta\tilde{H}_0 \end{pmatrix} \\
&\quad = -n^{-1/2}\begin{pmatrix} l^{(n)}_{\beta\beta}(\hat{\beta}, \hat{H}) & l^{(n)}_{\beta H}(\hat{\beta}, \hat{H}) \\ l^{(n)}_{H\beta}(\hat{\beta}, \hat{H}) & l^{(n)}_{HH}(\hat{\beta}, \hat{H}) \end{pmatrix}\left[\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{H} - \tilde{H}_0 \end{pmatrix}, \begin{pmatrix} b \\ \int p d\hat{H} \end{pmatrix}\right] \\
&\quad = -n^{1/2}\begin{pmatrix} l_{\beta\beta}(\beta_0, H_0) & l_{\beta H}(\beta_0, H_0) \\ l_{H\beta}(\beta_0, H_0) & l_{HH}(\beta_0, H_0) \end{pmatrix}\left[\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{H} - H_0 \end{pmatrix}, \begin{pmatrix} b \\ \int p d H_0 \end{pmatrix}\right] + o_p(1) \\
&\quad = n^{-1/2}\Big(l^{(n)}_{\beta}(\beta_0, H_0)'b + l^{(n)}_{H}(\beta_0, H_0)[\int p d H_0]\Big) + o_p(1) \\
&\quad = n^{-1/2}\Big(l^{(n)}_{\beta}(\beta_0, H_0)'b + l^{(n)}_{H}(\beta_0, H_0)[\int p d \hat{H}]\Big) + o_p(1). \qquad \text{(A.5)}
\end{aligned}$$

Since $\hat{A}$ is invertible, for any bounded vector $w = (w_1, \ldots, w_d)'$ and $\tilde{b}$, we can choose η and b such that $\hat{A}(b', \eta')' = (\tilde{b}', w')'$. With such choices, (A.5) implies that

$$\begin{aligned}
&n^{1/2}\tilde{b}'(\hat{\beta} - \beta_0) + n^{1/2}\sum_{i=1}^{d} w_i(\hat{h}(t_i) - \tilde{h}(t_i)) \\
&\qquad = n^{-1/2}\Big(l^{(n)}_{\beta}(\beta_0, H_0)'b + l^{(n)}_{H}(\beta_0, H_0)[\int p d \hat{H}]\Big) + o_p(1),
\end{aligned}$$

which converges to normal with covariance matrix

$$\begin{aligned}
V = &\lim_{n\to\infty}\Big\{n^{-1}\sum_{i=1}^{n} E\Big[\frac{1}{\pi_i}\Big(l^{(i)}_{\beta}(\beta_0, H_0)'b + l^{(i)}_{H}(\beta_0, H_0)[\int p d \hat{H}]\Big)^2\Big]\Big\} \\
&+ \lim_{n\to\infty}\Big\{n^{-1}\sum_{1\le i\ne j\le n} E\Big[\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\Big(l^{(i)}_{\beta}(\beta_0, H_0)'b + l^{(i)}_{H}(\beta_0, H_0)[\int p d \hat{H}]\Big) \\
&\times\Big(l^{(j)}_{\beta}(\beta_0, H_0)'b + l^{(i)}_{H}(\beta_0, H_0)[\int p d \hat{H}]\Big)\Big]\Big\},
\end{aligned}$$

where $l^{(i)}_{\beta}(\beta, H)$ is the derivative of $l_i(\beta, H) = \delta_i[\log\lambda\{H(Y_i)e^{\beta' X_i}\} + \beta' X_i + \log h(Y_i)] - \Lambda\{H(Y_i)e^{\beta' X_i}\}$ with respect to β, $l^{(i)}_{H}(\beta, H)[\int p d\hat{H}]$ is the derivative of $l_i(\beta, H)$ with respect to H along the path $H + \epsilon\int p d\hat{H}$. It is easy to see that

V can be consistently estimated by

$$n^{-1}\sum_{i=1}^{n}\frac{\Delta_i}{\pi_i^2}\Big(l_\beta^{(i)}(\hat\beta,\hat H)'b+l_H^{(i)}(\hat\beta,\hat H)[\int pd\hat H]\Big)^2$$
$$+n^{-1}\sum_{1\le i\ne j\le n}\frac{\Delta_i\Delta_j(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}\pi_i\pi_j}\Big(l_\beta^{(i)}(\hat\beta,\hat H)'b+l_H^{(i)}(\hat\beta,\hat H)[\int pd\hat H]\Big)$$
$$\times\Big(l_\beta^{(j)}(\hat\beta,\hat H)'b+l_H^{(j)}(\hat\beta,\hat H)[\int pd\hat H]\Big),$$

which is equal to $(b',\eta')\hat B(b',\eta')'$, where $\hat B$ is defined in (2.6). Thus, the asymptotic variance for

$$n^{1/2}\tilde b'(\hat\beta-\beta_0)+n^{1/2}\sum_{i=1}^{d}w_i(\hat h(t_i)-\tilde h(t_i))$$

can be consistently estimated by $(b',\eta')\hat B(b',\eta')'=(\tilde b',w')\hat A^{-1}\hat B\hat A^{-1}(\tilde b',w')'$. That is, for any vector $\tilde b$ and any function w with bounded total variation in $[0,\tau]$ such that $w(t_i)=w_i$, the asymptotic variance for

$$n^{1/2}\tilde b'(\hat\beta-\beta_0)+n^{1/2}\int_0^\tau w(t)d[\hat H(t)-H_0(t)]$$

can be estimated by $(\tilde b',w')\hat A^{-1}\hat B\hat A^{-1}(\tilde b',w')'$, where $w=(w(t_1),\dots,w(t_d))'$.

References

Barlow, W. (1994). Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064-1072.

Bednarski, T. (1993). Robust estimation in Cox's regression model. *Scand. J. Statist.* **20**, 213-225.

Borgan, O., Goldstein, L. and Langholz, B. (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann. Statist.* **23**, 1749-1778.

Breslow, N. E. (1996). Statistics in epidemiology: the case-control studies. *J. Amer. Statist. Assoc.* **91**, 14-28.

Chen, K. (2001). Generalized case-cohort sampling. *J. Roy. Statist. Soc. Ser. B* **63**, 791-809.

Chen, K. (2004). Statistical estimation in the proportional hazards model with risk set sampling. *Ann. Statist.* **32**, 1513-1532.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659-668.

Chen, K. and Lo, S-H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika* **86**, 755-764.

Chen, Y.-H. and Zucker, D. M. (2009). Case-cohort analysis with semiparametric transformation models. *J. Statist. Plann. Inference* **139**, 3706-3717.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in the two-sample generalized odds rate model. *J. Amer. Statist. Assoc.* **83**, 744-749.

Goldstein, L. and Langholz, B. (1992). Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann. Statist.* **20**, 1903-1928.

Kalbfleisch, J. D. and Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statist. Medicine* **7**, 147-160.

Kong, L., Cai, J. and Sen, P. K. (2004). Weighted estimating equations for semiparametric transformation models with censored data from a case-cohort design. *Biometrika* **91**, 305-319.

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika* **87**, 73-87.

Kulich, M. and Lin, D. Y. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *J. Amer. Statist. Assoc.* **99**, 832-844.

Langholz, B. and Goldstein, L. (1996). Risk set sampling in epidemiologic cohort studies. *Statist. Sci.* **11**, 35-53.

Langholz, B. and Thomas, D. C. (1990). Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *Amer. J. Epidemiology* **31**, 169-176.

Langholz, B. and Thomas, D. C. (1991). Efficiency of cohort sampling designs: Some surprising results. *Biometrics* **47**, 1563-1571.

Langholz, B., Thomas, D., Xiang, A. and Stram, D. (1999). Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. *Amer. J. Industrial Medicine* **35**, 246-256.

Lawless, J. F., Kalbfleisch, J. D. and Wild, C. J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *J. Roy. Statist. Soc. Ser. B* **61**, 413-438.

Lu, W. and Tsiatis, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93**, 207-214.

Lubin, J., Boice, J., Edling, C., Hornung, R., Howe, G., Kunz, E., Kusiak R, Morrison, H., Radford, E., Samet, J., Tirmarche, M., Woodward, A., Xiang, Y. and Pierce, D. (1994). Radon and lung cancer risk: A joint analysis of 11 underground miners studies. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. NIH Publication 94-3644.

Nan, B., Yu, M. and Kalbfleisch, J. D. (2006). Censored linear regression for case-cohort studies. *Biometrika* **93**, 747-762.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.

Oakes, D. (1981). Survival times: Aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* **49**, 235-264.

Samuelsen, S. (1997). A pseudo-likelihood approach to analysis of nested case-control data. *Biometrika* **84**, 379-394.

Sasieni, P. (1993). Some new estimators for Cox regression. *Ann. Statist.* **31,** 1721-1759.

Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64-81.

Suissa, S., Edwards, M. and Biovin, J-F. (1998). External comparisons from nested case-control designs. *Epidemiology* **9**, 72-78.

Thomas, D. C. (1977). Appendum to "Methods of cohort analysis: appraisal by application to asbestos mining," by Liddell, F. D. K., McDonald, J. C., and Thomas, D. C.. *J. Roy. Statist. Soc. Ser. A* **140**, 469-490.

Thomas, D., Pogoda, J., Langholz, B. and Mack, W. (1994). Temporal modifiers of the radon-smoking interaction. *Health Physics* **66**, 257-262.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627-640.

Zeng, D., Lin, D. Y., Avery, C. L., North, K. E. and Bray, M. S. (2006). Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. *Biostatistics* **7**, 486-502.

Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong.

E-mail: makchen@ust.hk

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R.China.

E-mail: slq@amt.ac.cn

School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China.

E-mail: xweitong@bnu.edu.cn

(Received October 2010; accepted March 2011)