Published in final edited form as:

Lifetime Data Anal. 2012 January ; 18(1): 80–93. doi:10.1007/s10985-011-9203-3.

On estimation of linear transformation models with nested case–control sampling

Wenbin Lu and

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

Mengling Liu

Division of Biostatistics, New York University, New York, NY 10016, USA

Wenbin Lu: lu@stat.ncsu.edu

Abstract

Nested case–control (NCC) sampling is widely used in large epidemiological cohort studies for its cost effectiveness, but its data analysis primarily relies on the Cox proportional hazards model. In this paper, we consider a family of linear transformation models for analyzing NCC data and propose an inverse selection probability weighted estimating equation method for inference. Consistency and asymptotic normality of our estimators for regression coefficients are established. We show that the asymptotic variance has a closed analytic form and can be easily estimated. Numerical studies are conducted to support the theory and an application to the Wilms' Tumor Study is also given to illustrate the methodology.

Keywords

Linear transformation models; Nested case–control sampling; Weighted estimating equation

1 Introduction

Thomas (1977) introduced the nested case–control (NCC) design, by which along the follow-up of a cohort, subjects with observed failure event are deemed as cases and for each case, controls are randomly selected from the at-risk set at the case's failure time. Then covariates of interest are ascertained for cases and selected controls only. Therefore, NCC studies contain time-to-event information as cohort studies while being cost-effective as case–control studies.

There exists an intrinsic connection between the conditional likelihood for matched case–control data and the partial likelihood function for NCC data by the Cox proportional hazards model (Prentice and Breslow 1978; Oakes 1981). Thus, the Cox proportional hazards model has been extensively used for NCC studies and its theoretical properties have been established using the counting process and martingale theory (Goldstein and Langholz 1992; Borgan et al. 1995, among others). A more efficient estimator was further developed by Chen (2004) using a kernel-weighted local averaging method. In addition, Samuelsen (1997) proposed an inverse selection probability weighted estimator based on a pseudo-likelihood approach.

For many medical applications, the proportional hazards assumption may be too restrictive. Therefore, alternative models that relax this assumption are useful. For example, the proportional odds model provides a better fit if the hazard functions of two treatment groups converge to the same limit (e.g. Pettitt 1982, 1984; Bennett 1983; Dabrowska and Doksum 1988; Murphy et al. 1997). In this paper, we consider a class of linear transformation models (Clayton and Cuzick 1985; Bickel et al. 1993; Cheng et al. 1995; Fine et al. 1998), which can be specified by

$$\log H(T) = -\beta'Z + \varepsilon, \quad (1)$$

where T denotes the survival time; $H(\cdot)$ is an unknown monotone increasing function with $H(0) = 0$; β is a p -dimensional regression parameter vector; the error term ε is assumed to follow a known continuous distribution independent of covariate vector Z . Let $\lambda(t)$ and $A(t)$ denote the hazard and cumulative hazard functions for e^ε , respectively. Note that model (1) with $A(t) = t$ corresponds to the proportional hazards model and $A(t) = \log(1 + t)$ corresponds to the proportional odds model. The formulation of linear transformation models provides a unified way of representing a family of semiparametric models and is amenable to development of unified estimation procedures. But one should note that statistical inference of the linear transformation model necessitates the specification of error distribution and the interpretation of regression coefficients β also depends on the distributional assumption.

In survival analysis literature, the linear transformation model has been employed as an alternative to the Cox model for case-cohort design (Prentice 1986; Self and Prentice 1988), which is another cost-effective sampling method. For example, Kong et al. (2004) proposed a weighted U -statistics estimation method for the linear transformation model with case-cohort design but their method required censoring distribution being independent of covariates. To relax this restrictive censoring distribution assumption, Lu and Tsiatis (2006) proposed an inverse selection probability weighted estimating equation approach based on the martingale integral representation for the linear transformation model (Chen et al. 2002). However, to the best of our knowledge, the estimation and inference methods for the linear transformation model under NCC design have not been studied. The contributions of our paper are developing an estimation method for the linear transformation model with NCC data, establishing asymptotic properties of the resulting estimators, and providing a valid inference procedure. In addition, comparing with the simple constant selection probability in case-cohort design, the selection of controls in NCC design is tied to each specific case and the selection probability is much more complicated depending on subjects' survival times. Thus, the inference for our proposed inverse selection probability weighted estimators is more challenging than that of Lu and Tsiatis (2006).

The rest of the paper is organized as follows. In Sect. 2, we present the inverse selection probability weighted estimating equations. The theoretical properties for the resulting estimators are established in Sect. 3. In Sect. 4, simulation studies are conducted to evaluate the finite-sample performance of the proposed approach, and an application to the Wilms' Tumor Study is also given to illustrate the methodology. Some conclusions and discussions are given in Sect. 5. All technical derivations are contained in Appendix.

2 Weighted estimating equations

Consider a cohort of size N . Let (T_i, C_i, Z_i) , $i = 1, \dots, N$, denote N independent and identically distributed random triplets of failure time, censoring time and covariate vector of dimension p . The failure times T_i 's are assumed to follow the linear transformation model

(1). Define $\tilde{T}_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$, $N_i(t) = \delta_i I(\tilde{T}_i \leq t)$ and $Y_i(t) = I(\tilde{T}_i \geq t)$. Let β_0 and H_0 denote the true values of β and H . In addition, define

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) d\Lambda \left\{ H_0(s) e^{\beta_0' Z_i} \right\}. \quad (2)$$

It is easy to see that $M_i(t)$ is a martingale process by the counting process and martingale theory (Fleming and Harrington 1991; Andersen et al. 1993).

Consider that m controls are sampled without replacement from the at-risk set at each observed failure time. Following the notation of Samuelsen (1997), let V_{0i} be the indicator that individual i is ever selected as a control. Given the observed event times and censoring indicators $\mathcal{F} = \{(\tilde{T}_i, \delta_i); i = 1, \dots, N\}$, the probability that individual i is ever selected as a control in the NCC study is

$$p_{0i} = 1 - \prod_{\tilde{T}_j < \tilde{T}_i} \left(1 - \frac{m \delta_j}{\sum_{k=1}^N Y_k(\tilde{T}_j) - 1} \right). \quad (3)$$

Furthermore, we define the weight $\pi_i = \delta_i + (1 - \delta_i) V_{0i} / p_{0i}$ for each individual in the full cohort by the inverse selection probability.

Motivated by the martingale-based estimation method for the linear transformation model (Chen et al. 2002) and the inverse probability weighted technique for sample selections (Horvitz and Thompson 1952), we propose to estimate H and β using the following weighted estimating equations,

$$\sum_{i=1}^N \pi_i \left[dN_i(t) - Y_i(t) d\Lambda \left\{ H(t) e^{\beta' Z_i} \right\} \right] = 0, \quad t \geq 0, \quad (4)$$

$$\sum_{i=1}^N \int_0^\tau \pi_i Z_i \left[dN_i(t) - Y_i(t) d\Lambda \left\{ H(t) e^{\beta' Z_i} \right\} \right] = 0, \quad (5)$$

where τ is the maximum follow-up length. Equations 4 and 5 can be solved iteratively using a similar algorithm proposed by Lu and Tsiatis (2006). Let $(\hat{\beta}, \hat{H})$ be the solution of (4) and (5), where $\hat{H}$ is a nondecreasing step function with jumps only at the observed failure times. Note that for the proportional hazards model, our estimator $\hat{\beta}$ is equivalent to Samuelsen's maximum pseudo-likelihood estimator (Samuelsen 1997).

3 Asymptotical properties

In this section, we will establish the consistency and asymptotic normality for our proposed estimators of regression parameters. We first introduce additional notation. Let $\psi(t) = \dot{\lambda}(t) / \lambda(t)$. Hereafter, $\dot{f}$ of a function f always denotes its derivative. For any $t, s \in (0, \tau]$, define

$$\begin{aligned}
B(t, s) &= \exp \left(\int_s^t \frac{E \left[\lambda \left\{ H_0(u) e^{\beta_0' Z} \right\} H_0(u) e^{\beta_0' Z} + \lambda \left\{ H_0(u) e^{\beta_0' Z} \right\} e^{\beta_0' Z} Y(u) \right]}{E \left[\lambda \left\{ H_0(u) e^{\beta_0' Z} \right\} e^{\beta_0' Z} Y(u) \right] H_0(u)} \times dH_0(u) \right), \\
\mu_Z(t) &= \frac{E \left[Z \lambda \left\{ H_0(\tilde{T}) e^{\beta_0' Z} \right\} H_0(\tilde{T}) e^{\beta_0' Z} Y(t) B(t, \tilde{T}) \right]}{E \left[\lambda \left\{ H_0(t) e^{\beta_0' Z} \right\} H_0(t) e^{\beta_0' Z} Y(t) \right]}, \\
A &= \int_0^\tau E \left(\{Z - \mu_Z(t)\} Z' \left[\lambda \left\{ H_0(t) e^{\beta_0' Z} \right\} H_0(t) e^{\beta_0' Z} + \lambda \left\{ H_0(t) e^{\beta_0' Z} \right\} \right] \times e^{\beta_0' Z} Y(t) \right) dH_0(t), \\
\Sigma &= E \left(\left\{ \delta + \frac{1-\delta}{s_0} \right\} \left[\int_0^\tau \{Z - \mu_Z(t)\} dM(t) \right]^{\otimes 2} \right) - m \int_0^\tau \frac{\bar{\lambda}(t)}{\bar{y}(t)} \times \left(E \left[Y(t) (1 - \delta) \frac{1-s_0}{s_0} \int_0^\tau \{Z - \mu_Z(t)\} dM(t) \right] \right)^{\otimes 2} dH_0(t),
\end{aligned}$$

where, for any real vector a , $a^{\otimes 2} = aa'$, $\bar{\lambda}(t) = E[Y(t) \lambda \{H_0(t) e^{\beta_0' Z}\} e^{\beta_0' Z}]$, $\bar{y}(t) = E\{Y(t)\}$ and for $i = 1, \dots, N$,

$$s_{0i} = \lim_{N \rightarrow \infty} p_{0i} = 1 - \exp \left\{ -m \int_0^{\tilde{T}_i} \frac{\bar{\lambda}(t)}{\bar{y}(t)} dH_0(t) \right\}, \quad (6)$$

with s_0 being a random copy of s_{0i} 's. The equality in (6) was established in Samuelsen (1997) following the consistency proof of the Kaplan–Meier estimator (Andersen et al. 1993).

To achieve consistency and asymptotic normality, the following regularity conditions are assumed:

- (C1) β_0 is an interior point of a known compact set $\mathcal{B}_0$ and H_0 is strictly increasing and continuously differentiable with $H_0(0) = 0$.
- (C2) The covariate Z is bounded, and if there exists a vector α and a deterministic function $\alpha_0(t)$ such that $\alpha_0(t) + \alpha' Z = 0$ with probability one, then $\alpha = 0$ and $\alpha_0(t) \equiv 0$.
- (C3) $\lambda(\cdot)$ is positive and has continuous second derivative, with $\lim_{t \rightarrow 0} \lambda(t) = 0$ and $\lim_{t \rightarrow 0} \psi(t) = 0$.
- (C4) τ is finite and $\text{pr}(\tilde{T} \geq \tau | Z) > 0$ for all the possible values of Z .
- (C5) The matrices A and Σ are finite and nondegenerate.

Theorem 1

Assume the above conditions (C1)–(C5) hold. Then

$$N^{1/2}(\hat{\beta} - \beta_0) \rightarrow N\{0, A^{-1} \sum (A^{-1})'\}, \quad (7)$$

in distribution. Moreover, consistent estimators of A and Σ can be obtained by the usual plug-in method. Specifically,

$$\begin{aligned}\widehat{A} &= \frac{1}{N} \sum_{i=1}^N \int_0^\tau \pi_i \{Z_i - \bar{Z}(t)\} Z_i' \left[\lambda \left\{ \widehat{H}(t) e^{\beta' Z_i} \right\} \widehat{H}(t) e^{\beta' Z_i} + \lambda \left\{ \widehat{H}(t) e^{\beta' Z_i} \right\} \right] e^{\beta' Z_i} \times Y_i(t) d\widehat{H}(t), \\ \widehat{\Sigma} &= \frac{1}{N} \sum_{i=1}^N \pi_i^2 \left[\int_0^\tau \{Z_i - \bar{Z}(t)\} d\widehat{M}_i(t) \right]^{\otimes 2} - m \int_0^\tau \frac{\sum_{j=1}^N Y_j(t) \lambda \left\{ \widehat{H}(t) e^{\beta' Z_j} \right\} e^{\beta' Z_j}}{\sum_{j=1}^N Y_j(t)} \times \left[\frac{1}{N} \sum_{i=1}^N (\pi_i^2 - \pi_i) Y_i(t) \int_0^\tau \{Z_i - \bar{Z}(t)\} d\widehat{M}_i(t) \right]^{\otimes 2} d\widehat{H}(t),\end{aligned}$$

where

$$\begin{aligned}\bar{Z}(t) &= \frac{\sum_{i=1}^N \pi_i Z_i \lambda \left\{ \widehat{H}(\bar{T}_i) e^{\beta' Z_i} \right\} \widehat{H}(\bar{T}_i) e^{\beta' Z_i} Y_i(t) \widehat{B}(t, \bar{T}_i)}{\sum_{i=1}^N \pi_i \lambda \left\{ \widehat{H}(t) e^{\beta' Z_i} \right\} \widehat{H}(t) e^{\beta' Z_i} Y_i(t)}, \\ \widehat{B}(t, s) &= \exp \left(\int_s^t \frac{\sum_{i=1}^N \left[\lambda \left\{ \widehat{H}(u) e^{\beta' Z_i} \right\} \widehat{H}(u) e^{\beta' Z_i} + \lambda \left\{ \widehat{H}(u) e^{\beta' Z_i} \right\} \right] e^{\beta' Z_i} Y_i(u)}{\sum_{i=1}^N \pi_i \lambda \left\{ \widehat{H}(u) e^{\beta' Z_i} \right\} \widehat{H}(u) e^{\beta' Z_i} Y_i(u)} \times d\widehat{H}(u) \right), \\ \widehat{M}_i(t) &= N_i(t) - \int_0^t Y_i(s) d\Lambda \left\{ \widehat{H}(s) e^{\beta' Z_i} \right\},\end{aligned}$$

for $s, t \in (0, \tau]$.

Note that for the proportional hazards model, the derived sandwich variance estimator is the same as that for Samuelsen's maximum pseudo-likelihood estimator (Samuelsen 1997).

4 Numerical studies

4.1 Simulations

We have conducted simulation studies to assess the finite-sample performance of our proposed inverse selection probability weighted estimators, denoted by ISPW. For the hazard function of $e^{\beta' Z}$, we assumed $\lambda(t) = 1/(1 + \gamma t)$, with $\gamma = 0, 1$ or 2 , where the proportional hazards model and the proportional odds model correspond to $\gamma = 0$ and $\gamma = 1$, respectively. The transformation function $H(t) = t$ for $\gamma = 0$, $H(t) = e^t - 1$ for $\gamma = 1$, and $H(t) = 0.5e^{2t} - 0.5$ for $\gamma = 2$. For each model, two independent covariates Z_1 and Z_2 were generated with Z_1 following a $\text{Un}(0, 1)$ and Z_2 following a Bernoulli distribution with the success probability of 0.5 .

The first set of simulation studies were designed for covariate independent censoring. The regression parameters $\beta = (\beta_1, \beta_2)' = (1, -1)'$ and the censoring times were generated from a $\text{Un}(0, c)$, where c was chosen such that the average proportion of censoring is 80% in the full cohort of 1,000 subjects. We considered two NCC sampling schemes with different numbers of controls, i.e. $m = 1$ or 2 . For comparison, we also included Thomas's estimator based on the proportional hazards model for NCC data (denoted by TM), Samuelsen's maximum pseudo-likelihood estimator (denoted by SS) and the martingale equations-based estimator of Chen et al. (2002) as a benchmark for the full cohort analysis (denoted by CJY). Note that for the proportional hazards model, the proposed ISPW estimator is the same as SS estimator. So we only report the results of SS estimators for the cases $\gamma = 1$ or 2 . Under each scenario, 500 runs of simulations were conducted and the results were summarized in Table 1.

Results from Table 1 indicate that ISPW estimators are essentially unbiased under all the scenarios (Bias); the means of estimated standard errors (SE) are close to the empirical standard deviations (SD) of the parameter estimators; and the 95% Wald-type confidence intervals have proper coverage rate (CP). When the number of selected controls increases

from $m = 1$ to $m = 2$, the efficiency of ISPW estimators improves and quickly approaches that of the full cohort CJY estimators. On the other hand, Thomas's estimators are unbiased only for the proportional hazards model ($\gamma = 0$), but for other hazard models, TM estimators show marked biases. Moreover, even for the proportional hazards model, TM estimators are less efficient than ISPW estimators for all cases. Samuelsen's estimators also show biases under the cases $\gamma = 1$ and $\gamma = 2$.

The second set of simulation studies were conducted for covariate dependent censoring. Here we allowed the censoring time to depend on continuous covariate Z_1 and generated the censoring time from a Cox proportional hazards model $\lambda_c(t|Z_1) = \lambda_{0c} \exp(0.2Z_1)$, where the constant baseline hazard function λ_{0c} was chosen to obtain a 80% censoring rate. Simulation results were summarized in Table 2. Similar findings were observed as from Table 1.

4.2 Application to Wilms' Tumor Study

We applied our proposed ISPW estimation method to the Wilms's Tumor Study (D'Angio et al. 1989; Green et al. 1998). Wilms' tumor is a rare renal cancer and typically occurs in children. This dataset contains complete information of 3,915 subjects participating in the third (NWTSG-3) and fourth (NWTSG-4) Wilms' tumor studies. Among them, 669 (17%) patients had cancer recurrence, which were treated as cases in our analysis. Considering six covariates: NWTSG-4 (1 for NWTSG-4 and 0 for NWTSG-3), Stage II–IV (three binary variables), UH (unfavorable central histology: 1 or 0), and tumor diameter, we fitted the linear transformation model (1) in which the error distribution assumed the same hazard function as studied in the simulations with $\gamma = 0, 1$ and 2 . We conducted NCC sampling from the full cohort with the number of controls $m = 1$ or 2 . We also compared the proposed ISPW estimator with Thomas' estimator (TM) and the full-cohort estimator (CJY) of Chen et al. (2002).

The results are summarized in Table 3. We observe that for all the fitted models, the ISPW estimators achieve consistent conclusions as the full-cohort CJY estimators: that effects of covariates NWTSG-4, Stage II–IV and UH are significant while tumor diameter is not. When the number of selected controls increases from 1 to 2, the SE of the ISPW estimators decrease and become close to those of the CJY estimators. However, the TM estimators do not show NWTSG-4 being significant for either 1 or 2 controls while Stage III is not significant when $m = 1$, mainly due to low efficiency of the TM estimator.

5 Discussion

In this paper, we studied the linear transformation model for NCC data to provide alternative inference tools for NCC studies. We proposed a class of inverse selection probability weighted martingale-based estimating equations for parameter estimation. The proposed inference approach generalized Thomas' partial likelihood estimator and Samuelsen's pseudo-likelihood estimator by allowing non-proportional hazards models. The new estimators are easy to compute, with a closed-form variance–covariance matrix that can be computed by the usual plug-in method.

Due to the linear transformation model presentation given in (1), the proposed estimation method can only accommodate time-independent covariates. It is interesting to investigate the extension to the counting process based semiparametric transformation model proposed by Zeng and Lin (2006) that allows time-dependent covariates. In addition, it is practically more interesting to consider a class of linear transformation models, in which the error distribution is indexed by a parameter γ , and use data to estimate the optimal value of γ . In the literature, Cai and Cheng (2004) proposed a method by minimizing a measure of discrepancy between the observed and fitted values to estimate the parameter γ in the error

distribution. Zeng and Lin (2006) proposed a nonparametric maximum likelihood estimation method and suggested to estimate the optimal value of γ by a grid search. However, for both methods, the asymptotical properties of the estimators for β and H are derived under the assumption that γ is known. Under NCC sampling, it becomes even more challenging to estimate γ together with the parameters in the semiparametric transformation models. These are warranted for future research.

In practice, left-truncation sometimes occur in cohort studies; for example when only subjects who are disease-free enter the study. To incorporate the left truncation into NCC studies, controls need to be drawn from an adjusted risk set defined as $R(t) = \{i : L_i < t \leq \tilde{T}_i\}$, where L_i denotes the left-truncation time for individual i . Then the selection probability formula (3) needs to be modified as follows

$$p_{0i} = 1 - \prod_{L_i < \tilde{T}_j < \tilde{T}_i} \left(1 - \frac{m\delta_j}{\sum_{k=1}^N Y_k(\tilde{T}_j) - 1} \right), \quad (8)$$

where the at-risk indicator function $Y_i(t) = I(L_i < t \leq \tilde{T}_i)$ for subject i . With the modified selection probability formula and modified at-risk indicator, our proposed inverse selection probability weighted martingale-based estimating equations can accommodate left-truncated and right-censored NCC data.

Acknowledgments

The authors would like to thank the two referees for their insightful and constructive comments. The research was supported in part by NIH grants RO3 CA153083 and RO1 CA140632.

References

- Andersen, PK.; Borgan, Ø.; Gill, R.; Keiding, N. Statistical models based on counting processes. Springer; New York: 1993.
- Bennett S. Analysis of survival data by the proportional odds model. *Stat Med.* 1983; 2:273–277. [PubMed: 6648142]
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press; Baltimore: 1993.
- Borgan Ø, Goldstein L, Langholz B. Methods for the analysis of sampled data in the Cox proportional hazards model. *Ann Stat.* 1995; 33:1749–1778.
- Cai T, Cheng S. Semi-parametric regression analysis for doubly censored data. *Biometrika.* 2004; 91:277–290.
- Chen K. Statistical estimation in the proportional hazards model with risk set sampling. *Ann Stat.* 2004; 32:1513–1532.
- Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. *Biometrika.* 2002; 89:659–668.
- Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika.* 1995; 82:835–845.
- Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). *J R Stat Soc A.* 1985; 148:82–117.
- Dabrowska DM, Doksum KA. Estimation and testing in the two-sample generalized odds rate model. *J Am Stat Assoc.* 1988; 83:744–749.
- D’Angio GJ, Breslow N, Beckwith JB, Evans A, Baum H, de Lorimier A, Ferbach D, Hrabovsky E, Jones G, Kelalis P. Treatment of Wilms’ tumor. Results of the third national Wilms tumor study. *Cancer.* 1989; 64:349–360. [PubMed: 2544249]

- Fine J, Ying Z, Wei LJ. On the linear transformation model for censored data. *Biometrika*. 1998; 85:980–986.
- Fleming, TR.; Harrington, DP. Counting processes and survival analysis. Wiley; New York: 1991.
- Goldstein L, Langholz B. Asymptotic theory of nested case–control sampling in the Cox regression model. *Ann Stat*. 1992; 20:1903–1928.
- Green DM, Breslow NE, Beckwith JB, Finklestein JZ, Grundy PG, Thomas PRM, Kim T, Shochat S, Haase GM, Ritchey ML, Kelalis PP, D’Angio GJ. Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms’ tumor: a report from the national Wilms’ tumor study group. *J Clin Oncol*. 1998; 16:237–245. [PubMed: 9440748]
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc*. 1952; 47:663–685.
- Kong L, Cai J, Sen PK. Weighted estimating equations for semiparametric transformation model with censored data from a case–cohort design. *Biometrika*. 2004; 91:305–319.
- Lu W, Tsiatis AA. Semiparametric transformation models for case–cohort study. *Biometrika*. 2006; 93:207–214.
- Murphy SA, Rossini AJ, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *J Am Stat Assoc*. 1997; 92:968–976.
- Oakes D. Survival times: aspects of partial likelihood (with discussion). *Int Stat Rev*. 1981; 49:235–264.
- Pettitt AN. Inference for the linear model using a likelihood based on ranks. *J R Stat Soc B*. 1982; 44:234–243.
- Pettitt AN. Proportional odds model for survival data and estimates using ranks. *Appl Stat*. 1984; 33:169–175.
- Pollard, D. NSF-CBMS regional conference series in probability and statistics. Vol. 2. Hayward; 1990. Empirical processes: theory and applications.
- Prentice RL. A case–cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
- Prentice RL, Breslow NE. Retrospective studies and failure time models. *Biometrika*. 1978; 65:153–158.
- Samuelsen SO. A pseudo-likelihood approach to analysis of nested case–control studies. *Biometrika*. 1997; 84:379–394.
- Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case–cohort studies. *Ann Stat*. 1988; 16:64–81.
- Thomas DC. Addendum to ‘methods of cohort analysis: appraisal by application to asbestos mining’ by F. D. K. Liddell, J. C. McDonald and D. C. Thomas. *J R Stat Soc A*. 1977; 140:469–491.
- Zeng D, Lin DY. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006; 93:627–640.
- Zhang L. A functional central limit theorem for asymptotically negatively dependent random fields. *Acta Math Hung*. 2000; 86:237–259.

Appendix

Proof of Theorem 1

We first establish the consistency of the estimators $\hat{\beta}$ and $\hat{H}$ as solutions to Eqs. 4 and 5. Throughout the Appendix, the notion $\|\cdot\|$ for a vector means the L^2 norm. Define a fixed ball in R^p centered at β_0 , i.e. $D_\varepsilon = \{\beta : \|\beta - \beta_0\| \leq \varepsilon\}$. Note that $E(V_{0i} | \mathcal{F}) = p_{0i}$ and $E(\pi_i | \mathcal{F}) = 1$. By the strong law of large numbers and the fact that $E\{\pi_i M_i(t)\} = 0$, we have, for large N , $\beta \in D_\varepsilon$ and $t \in [0, \tau]$,

$$\frac{1}{N} \sum_{i=1}^N \pi_i \left(N_i(t) - \Lambda \left[H_0 \left\{ \min(t, \tilde{T}_i) \right\} e^{a+\beta' Z_i} \right] \right) < 0, \quad (9)$$

$$\frac{1}{N} \sum_{i=1}^N \pi_i \left(N_i(t) - \Lambda \left[H_0 \left\{ \min(t, \tilde{T}_i) \right\} e^{-a+\beta' Z_i} \right] \right) > 0, \quad (10)$$

for sufficiently large a . Thus, there exists a unique $\hat{H}(t; \beta)$ such that for $t \in [0, \tau]$ and $\beta \in D_\varepsilon$

$$\frac{1}{N} \sum_{i=1}^N \pi_i \left[dN_i(t) - Y_i(t) d\Lambda \left\{ \hat{H}(t; \beta) e^{\beta' Z_i} \right\} \right] = 0. \quad (11)$$

Since (9) and (10) hold for any $a > 0$ when and only when $\beta = \beta_0$, we have that $\sup_{t \in [0, \tau]} |\hat{H}(t; \beta_0) - H_0(t)|$ converges to zero almost surely.

Take the derivative with respect to β on both sides of (11) and apply the strong law of large numbers, we have

$$B_2(t) d \left(\frac{\partial \log \hat{H}(t; \beta)}{\partial \beta} \Big|_{\beta=\beta_0} \right) + \left(\frac{\partial \log \hat{H}(t; \beta)}{\partial \beta} \Big|_{\beta=\beta_0} \right) B_1(t) dH_0(t) = -B_z(t) dH_0(t) + o_p(1),$$

where

$$\begin{aligned} B_2(t) &= E[\lambda \{H_0(t) e^{\beta_0' Z}\} e^{\beta_0' Z} Y(t) | H_0(t)], \\ B_1(t) &= E([\lambda \{H_0(t) e^{\beta_0' Z}\} H_0(t) e^{\beta_0' Z} + \lambda \{H_0(t) e^{\beta_0' Z}\}] e^{\beta_0' Z} Y(t)), \\ B_z(t) &= E(Z[\lambda \{H_0(t) e^{\beta_0' Z}\} H_0(t) e^{\beta_0' Z} + \lambda \{H_0(t) e^{\beta_0' Z}\}] e^{\beta_0' Z} Y(t)). \end{aligned}$$

For $t \in [0, \tau]$, define $\lambda^* \{\log H_0(t)\} = B(t, 0)$ and $\Lambda^*(t) = \int_{-\infty}^t \lambda^*(s) ds$. Note that $\lambda^* \{\log H_0(t)\} B_1(t) dH_0(t) - B_2(t) d\lambda^* \{\log H_0(t)\} = 0$. Then we have

$$\frac{\partial \log \hat{H}(t; \beta)}{\partial \beta} \Big|_{\beta=\beta_0} = - \int_0^t \frac{\lambda^* \{\log H_0(s)\} B_z(s)}{\lambda^* \{\log H_0(s)\} B_2(s)} dH_0(s) + o_p(1). \quad (12)$$

Under the conditions of Theorem 1, we can show that $\frac{\partial \log \hat{H}(t; \beta)}{\partial \beta} \Big|_{\beta=\beta_0}$ is bounded on $[0, \tau]$. Moreover, the equality established in (12) can be extended to a small neighborhood $D_{\varepsilon 0}$ of β_0 and $\sup_{t \in [0, \tau]} \frac{\partial \log \hat{H}(t; \beta)}{\partial \beta}$ is bounded for $\beta \in D_{\varepsilon 0}$. Note that $\hat{H}(t; \beta)$ is monotone increasing in t . Therefore, $\hat{H}(t; \beta_n)$ converges uniformly to $H_0(t)$ on $[0, \tau]$ provided β_n converges to β_0 . In particular, $\sup_{t \in [0, \tau]} |\hat{H}(t; \beta) - H_0(t)|$ converges to zero provided β is a consistent

estimator. We now plug the estimator $\hat{H}(t; \beta)$ into the Eq. 5, and define the estimating equation for β as

$$U(\beta) = \frac{1}{N} \sum_{i=1}^N \int_0^\tau \pi_i Z_i \left[dN_i(t) - Y_i(t) d\Lambda \left\{ \hat{H}(t; \beta) e^{\beta' Z_i} \right\} \right] = 0.$$

Based on (12) and the strong law of large numbers, we have

$$\dot{U}(\beta_0) = \frac{\partial}{\partial \beta} U(\beta)|_{\beta=\beta_0} = -A + o_p(1). \quad (13)$$

Moreover, by repeatedly applying the uniform law of large numbers (Pollard 1990), we can show that, for $\beta \in D_{\varepsilon_0}$, $U(\beta)$ and $\dot{U}(\beta)$ converge uniformly to $u(\beta)$ and $\dot{u}(\beta)$, where $\dot{u}(\beta)$ is the derivative of $u(\beta)$. Note that $u(\beta_0) = 0$ and $\dot{u}(\beta_0) = -A$. Following the similar techniques of Chen et al. (2002), Eq. 13 can be strengthened as, when $N \rightarrow \infty$ and then $\varepsilon_0 \rightarrow 0$,

$$\sup_{\beta \in D_{\varepsilon_0}} \left\| \dot{U}(\beta) - A \right\| \rightarrow 0 \quad (14)$$

in probability. Consider $U(\beta)$ as a random mapping from an arbitrarily small but fixed ball D_{ε_0} to another open connected set in R^p . Since A is assumed to be finite and nondegenerate, Eq. 14 implies that, with probability tending to 1, the mapping $U(\beta)$ is a homeomorphism from D_{ε_0} to its image, denoted as B_n . The convergence of $U(\beta)$ to zero indicates that B_n contains $0 \in R^p$ with probability tending to 1. This proves that $\hat{\beta}$ is consistent since $U(\hat{\beta}) = 0$ and D_{ε_0} is centered at β_0 and can be arbitrarily small.

Next, we derive the asymptotic distribution of $\hat{\beta}$. Compared to the estimator studied in Lu and Tsiatis (2006) for case-cohort sampling, the asymptotical distribution of the proposed estimator for NCC sampling is more difficult to derive. One main reason is that NCC sampling scheme is a dynamic process and the probability of being selected as a control is neither a constant nor independent, even in the asymptotic sense. Therefore, the conventional central limit theory for independent data cannot be applied. Here we use the central limit theory for asymptotically negatively dependent random variables (Zhang 2000). Following Step A.2 in the appendix of Chen et al. (2002), we can obtain the following asymptotical representation

$$\begin{aligned} U(\beta_0) &= \frac{1}{N} \sum_{i=1}^N \int_0^\tau \pi_i \{Z_i - \mu_Z(t)\} dM_i(t) + o_p(1) \\ &\equiv U_1(\beta_0) + U_2(\beta_0) + o_p\left(\frac{1}{\sqrt{N}}\right), \end{aligned}$$

where $N^{1/2} U_1(\beta_0) = N^{-1/2} \sum_{i=1}^N \int_0^\tau \{Z_i - \mu_Z(t)\} dM_i(t) \rightarrow N(0, \sum_1)$ as $N \rightarrow \infty$ by the martingale central limit theorem. Since $E(\pi_i - 1 | \mathcal{F}) = 0$, it is easy to show that $U_1(\beta_0)$ and $U_2(\beta_0)$ are uncorrelated. However, π_i and π_j ($i \neq j$) are correlated even after conditioning on $\mathcal{F}$ due to the NCC sampling scheme. Note that $(\pi_i - 1)^2 = (1 - \delta_i)(V_{0i} - p_{0i}^2)/p_{0i}^2$. Then the

$E\{(\pi_i - 1)^2 | \mathcal{F}\} = (1 - \delta_i)(1 - p_{0i})/p_{0i}$. Then the conditional variance of $N^{1/2}U_2(\beta_0)$ can be written as

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \frac{1-p_{0i}}{p_{0i}} (1 - \delta_i) \left[\int_0^\tau \{Z_i - \mu_Z(t)\} dM_i(t) \right]^{\otimes 2} \\ & + \frac{1}{N} \sum_{i \neq j} E \left\{ \left(\frac{V_{0i}}{p_{0i}} - 1 \right) \left(\frac{V_{0j}}{p_{0j}} - 1 \right) | \mathcal{F} \right\} \\ & \times (1 - \delta_i)(1 - \delta_j) \int_0^\tau \{Z_i - \mu_Z(t)\} dM_i(t) \left[\int_0^\tau \{Z_j - \mu_Z(t)\} dM_j(t) \right]'. \end{aligned}$$

From Samuelsen (1997, Section 3.1), we have, for $i \neq j$, $Cov(V_{0i}, V_{0j} | \mathcal{F}) = \rho_{ij}(1 - p_{0i})(1 - p_{0j})$, where

$$\rho_{ij} = -\frac{m}{N} \int_0^{\min(\tilde{T}_i, \tilde{T}_j)} \frac{\bar{\lambda}(t)}{\bar{y}(t)} dH_0(t) + O_p(N^{-2}).$$

After some algebra, the variance $\text{Var}\{N^{1/2}U_2(\beta_0) | \mathcal{F}\}$ can be written as

$$\frac{1}{N} \sum_{i=1}^N \frac{1-p_{0i}}{p_{0i}} (1 - \delta_i) \left[\int_0^\tau \{Z_i - \mu_Z(t)\} dM_i(t) \right]^{\otimes 2} - m \int_0^\tau \left[\frac{1}{N} \sum_{i=1}^N Y_i(t) \frac{1-p_{0i}}{p_{0i}} (1 - \delta_i) \int_0^\tau \{Z_i - \mu_Z(t)\} dM_i(t) \right]^{\otimes 2} \frac{\bar{\lambda}(t)}{\bar{y}(t)} dH_0(t) + o_p(1).$$

By the strong law of large numbers and (6), we have

$$\begin{aligned} \sum_2 & \equiv \lim_{N \rightarrow \infty} \text{Var}\{N^{1/2}U_2(\beta_0) | \mathcal{F}\} \\ & = E \left(\frac{1-s_0}{s_0} (1 - \delta) \left[\int_0^\tau \{Z - \mu_Z(t)\} dM(t) \right]^{\otimes 2} \right) \\ & - m \int_0^\tau \frac{\bar{\lambda}(t)}{\bar{y}(t)} \left(E \left[Y(t) (1 - \delta) \frac{1-s_0}{s_0} \int_0^\tau \{Z - \mu_Z(t)\} dM(t) \right] \right)^{\otimes 2} dH_0(t). \end{aligned}$$

By the central limit theory for asymptotically negatively dependent random variables (Zhang 2000), we have $N^{1/2}U_2(\beta_0) \rightarrow N(0, \Sigma_2)$ as $N \rightarrow \infty$. Thus,

$$N^{1/2}U(\beta_0) \rightarrow N(0, \sum_1 + \sum_2), \quad (15)$$

in distribution as $N \rightarrow \infty$. It is easy to show that $\Sigma_1 + \Sigma_2 = \Sigma$.

It follows from the Taylor expansion and the consistency of $\hat{\beta}$ that

$$0 = N^{1/2} U(\widehat{\beta}) = N^{1/2} U(\beta_0) + \dot{U}(\beta_0) N^{1/2} (\widehat{\beta} - \beta_0) + o_p(1).$$

Then the desired asymptotic normality of $\widehat{\beta}$ follows from (14) and (15). Recall that $E(\pi_i|\mathcal{F})=1$, $E(\pi_i^2|\mathcal{F})=\delta_i+(1-\delta_i)/p_{0i}$ and $E(\pi_i^2-\pi_i|\mathcal{F})=(1-\delta_i)(1-p_{0i})/p_{0i}$. By the strong law of large numbers and the consistency of $\widehat{\beta}$ and $\widehat{H}$, A and Σ can be consistently estimated by $\widehat{A}$ and $\widehat{\Sigma}$ respectively.

Table 1

Simulation results when censoring is covariate independent

m	β	ISPW			TM			CJY			SS		
		Bias	SD	SE	CP	Bias	SD	Bias	SD	Bias	SD		
$\gamma = 0$													
1	β_1	-0.028	0.325	0.345	96.4	-0.010	0.393	-0.027	0.224				
	β_2	-0.012	0.206	0.201	94.2	-0.001	0.226	-0.015	0.158				
2	β_1	0.012	0.281	0.296	95.8	0.050	0.347	-0.027	0.224				
	β_2	-0.022	0.177	0.178	96.2	-0.021	0.203	-0.015	0.158				
$\gamma = 1$													
1	β_1	0.005	0.372	0.405	97.2	-0.134	0.366	-0.022	0.270	-0.141	0.313		
	β_2	-0.023	0.232	0.230	94.8	0.103	0.219	-0.000	0.177	0.108	0.201		
2	β_1	-0.009	0.319	0.346	96.4	-0.123	0.313	-0.022	0.270	-0.152	0.270		
	β_2	-0.008	0.189	0.201	97.0	0.118	0.190	-0.000	0.177	0.120	0.165		
$\gamma = 2$													
1	β_1	-0.024	0.432	0.461	96.0	-0.239	0.382	-0.018	0.306	-0.267	0.317		
	β_2	-0.020	0.257	0.259	95.6	0.208	0.230	0.000	0.201	0.214	0.197		
2	β_1	0.005	0.371	0.396	96.0	-0.223	0.326	-0.018	0.306	-0.246	0.271		
	β_2	-0.018	0.213	0.225	96.8	0.205	0.189	0.000	0.201	0.214	0.164		

SD sample standard deviation, SE mean of estimated standard error, CP empirical coverage probability of 95% confidence interval

Table 2

Simulation results when censoring is covariate dependent

m	β	ISPW			TM			CJY			SS		
		Bias	SD	SE	CP	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\gamma = 0$													
1	β_1	0.010	0.321	0.334	96.0	0.029	0.404	-0.009	0.235				
	β_2	-0.014	0.197	0.195	96.0	-0.009	0.238	-0.011	0.156				
2	β_1	-0.029	0.280	0.287	95.8	0.002	0.341	-0.009	0.235				
	β_2	-0.022	0.161	0.172	96.6	-0.021	0.176	-0.011	0.156				
$\gamma = 1$													
1	β_1	0.022	0.384	0.407	96.6	-0.120	0.382	-0.020	0.289	-0.158	0.312		
	β_2	-0.007	0.228	0.232	96.0	0.159	0.225	-0.011	0.175	0.157	0.189		
2	β_1	-0.002	0.333	0.351	95.2	-0.157	0.316	-0.020	0.289	-0.177	0.270		
	β_2	-0.013	0.203	0.203	95.2	0.152	0.198	-0.011	0.175	0.152	0.169		
$\gamma = 2$													
1	β_1	0.017	0.440	0.474	96.4	-0.258	0.376	-0.020	0.340	-0.287	0.303		
	β_2	-0.021	0.270	0.268	94.6	0.260	0.227	-0.013	0.200	0.270	0.190		
2	β_1	0.040	0.391	0.410	96.0	-0.249	0.324	-0.020	0.340	-0.269	0.267		
	β_2	-0.018	0.229	0.234	95.8	0.268	0.176	-0.013	0.200	0.268	0.163		

SD sample standard deviation, SE mean of estimated standard error, CP empirical coverage probability of 95% confidence interval

Table 3

Analysis of Wilms' tumor data

	CJY		ISPW		TM			
			$m = 1$		$m = 2$		$m = 1$	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\gamma = 0$								
NWTSG-4	-0.217	0.076	-0.246	0.114	-0.274	0.095	0.156	0.126
Stage II	0.395	0.101	0.429	0.148	0.425	0.123	0.508	0.163
Stage III	0.500	0.104	0.335	0.150	0.364	0.128	0.249	0.162
stage IV	1.006	0.114	0.897	0.175	0.865	0.147	0.776	0.188
UH	1.323	0.086	1.381	0.146	1.443	0.118	1.307	0.169
Diameter	-0.014	0.011	-0.008	0.016	0.014	0.013	0.024	0.015
$\gamma = 1$								
NWTSG-4	-0.234	0.086	-0.282	0.129	-0.295	0.108		
Stage II	0.448	0.113	0.500	0.168	0.493	0.139		
Stage III	0.564	0.117	0.385	0.170	0.420	0.147		
stage IV	1.152	0.132	1.033	0.202	1.003	0.171		
UH	1.556	0.107	1.627	0.183	1.725	0.150		
Diameter	-0.013	0.013	-0.006	0.018	0.023	0.015		
$\gamma = 2$								
NWTSG-4	-0.250	0.096	-0.314	0.145	-0.320	0.122		
Stage II	0.503	0.126	0.571	0.188	0.552	0.155		
Stage III	0.628	0.131	0.436	0.190	0.469	0.164		
stage IV	1.295	0.151	1.164	0.228	1.130	0.194		
UH	1.800	0.129	1.884	0.221	2.013	0.185		
Diameter	-0.011	0.014	-0.003	0.020	0.030	0.017		

Ext. estimates, SE estimated standard error