

MARGINAL STRUCTURAL COX MODELS WITH CASE-COHORT SAMPLING

Hana Lee¹, Michael G. Hudgens², Jianwen Cai² and Stephen R. Cole²

¹*Brown University* and ²*University of North Carolina at Chapel Hill*

Abstract: A common objective of biomedical cohort studies is assessing the effect of a time-varying treatment or exposure on a survival time. In the presence of time-varying confounders, marginal structural models fit using inverse probability weighting can be employed to obtain a consistent and asymptotically normal estimator of the causal effect of a time-varying treatment. This article considers estimation of parameters in the semiparametric marginal structural Cox model (MSCM) from a case-cohort study. Case-cohort sampling entails assembling covariate histories only for cases and a random subcohort, which can be cost effective, particularly in large cohort studies with low outcome rates. Following Cole et al. (2012), we consider estimating the causal hazard ratio from a MSCM by maximizing a weighted-pseudo-partial-likelihood. The estimator is shown to be consistent and asymptotically normal under certain regularity conditions. Finite sample performance of the proposed estimator is evaluated in a simulation study. In the corresponding supplementary document, computation of the estimator using standard survival analysis software is presented.

Key words and phrases: Case-cohort study, causal inference, Cox model, marginal structural model, survival analysis.

1. Introduction

Biomedical cohort studies are often conducted with the goal of assessing the effect of a time-varying treatment (or exposure) on a survival time. In such studies there may exist time-dependent covariates which are simultaneously (i) confounders and (ii) affected by prior treatment. In the presence of time-varying confounders affected by prior treatment, standard methods such as Cox regression modeling with time-varying covariates do not in general yield consistent estimators of the causal effect of treatment (Robins (1986, 1998); Robins and Rotnitzky (1992); Hernán, Brumback, and Robins (2001)). On the other hand, marginal structural models (MSM) fit using inverse probability weighting can be employed to obtain consistent estimators of the causal effect of a time-varying treatment on an outcome of interest, even if there are time-varying confounders affected by prior treatment (Robins (1999); Hernán, Brumback, and Robins (2001)).

Recently, Cole et al. (2012) considered fitting MSCMs via inverse probability weighting in the presence of case-cohort sampling. The case-cohort study design is a cost-efficient approach to estimate treatment effects in large cohorts with low event rates when treatment or covariate information is expensive. The design entails randomly selecting a subcohort from the entire cohort. Covariate information is then collected only from the random subcohort and from individuals that are observed to experience an event (i.e., cases), saving cost and effort relative to obtaining covariate information from the full cohort. In addition to being cost efficient, the case-cohort design enjoys other benefits. For instance, the subcohort can serve as a basis for real-time covariate monitoring during the course of the study. Also, because the subcohort is chosen randomly, survival times to different diseases can be analyzed using the same subcohort (Self and Prentice (1988)).

In the presence of case-cohort sampling, Cole et al. (2012) considered estimating the causal hazard ratio of a MSCM via inverse probability weighting. Simulation studies indicated that their estimator could perform well empirically, but no formal justification for their estimator has been developed to date. Following Cole et al. (2012), we consider estimating the causal hazard ratio of a MSCM via inverse probability weighting in case-cohort studies and establish consistency and asymptotic normality for the estimator that maximizes a weighted-pseudo-partial-likelihood (WPPL) under certain regularity conditions.

The approach utilized in this paper entails standard counting process and martingale theory. This formulation readily enables practical implementation of the methods using existing survival analysis software. Framing the problem using counting processes may also be helpful in future work, e.g., in fitting MSCMs to data from nested case-control studies or in the presence of competing risks. In the special situation that the subcohort is the full cohort, the proposed inverse probability weighted estimator is asymptotically equivalent to the estimator in Robins (1999). In this case our proof gives an alternative consistency and normality proof to the one in Robins (1999) that does not utilize the usual counting process framework. We also derive a new variance estimator that arises from the counting process formulation under both full and case-cohort settings. Empirical results presented here indicate that, in certain scenarios, the proposed variance estimator may be preferred to the so-called “robust” variance estimator (Lin and Ying (1993)) employed in Cole et al. (2012).

The outline of the remainder of this paper is as follows. In Section 2, estimators of the hazard ratio of a MSCM in the presence of case-cohort sampling are introduced, including the estimator proposed by Cole et al. (2012). Consistency and asymptotic normality results are presented in Section 3 and a simulation study is in Section 4. Some additional considerations are in Section 5. Regularity conditions and discussion of the conditions are deferred to the Appendix.

The supplementary document accompanying this paper includes detailed proofs for Theorems 3.1–3.6, a description of how a MCSM can easily be fit via inverse probability weighting for either the full cohort or case-cohort setting using standard survival analysis software, such as R or SAS, additional simulation study results including performance of the baseline cumulative hazard estimator, and a summary of notation.

2. Marginal Structural Cox Model Estimators

2.1. Notation, assumptions, and model

Consider an observational cohort study where the outcome of interest is a survival time T , based on the time from study entry until some particular outcome occurs. We assume T is continuous so that there are no tied failure times. During the study, individuals may dropout or discontinue participation so that T is right censored. Individuals may or may not elect to receive treatment at various points of time during the study; let $A_i(t)$ indicate whether subject i is on treatment at time t . If more than one treatment is available, $A_i(t)$ is a vector of treatment indicator variables corresponding to the treatment levels. We assume $A_i(t)$ is a $p \times 1$ vector and treatment variation is irrelevant (VanderWeele (2009)): for fixed values of $A_i(t)$, additional variation in the treatment (e.g., dose, timing) does not affect the potential outcome. The subscript i is often suppressed when there is no ambiguity, as we assume random vectors are drawn independently from a distribution common to all subjects. Let $L(t)$ denote a vector of covariates at time t and $L(0)$ represent baseline covariates. Overbars are used to represent history up to and including time t , with $\bar{A}(t) = \{A(u) : 0 \leq u \leq t\}$ and $\bar{L}(t)$ defined analogously. Assume decisions related to treatment at t are made after obtaining the covariate information at t , so $L(t)$ is temporally prior to $A(t)$. For a case-cohort study, the time varying covariates $L(t)$ and treatment $A(t)$ are by design observed only for the cases and individuals in the random subcohort (while under study); $L(t)$ and $A(t)$ are missing for all other individuals. Corresponding to the subcohort, let \tilde{C} denote the set of indices of size $\tilde{n} \leq n$ that are randomly selected without replacement from the set $\{1, \dots, n\}$ corresponding to the entire cohort.

Let \bar{a} denote a possible (static) treatment plan, $\bar{a} = \{a(t) : 0 \leq t \leq \tau\}$, where τ is the study duration. Assume $\tau = 1$ without loss of generality. Each possible value of \bar{a} can be interpreted as a prespecified treatment plan. Given a single treatment, examples of \bar{a} are never treat ($a(t) = 0$ for all $t \in [0, 1]$), treat starting at a prespecified time $t_1 < 1$ ($a(t) = I\{t \geq t_1\}$ where $I\{\cdot\}$ is the usual indicator function), treat from baseline ($a(t) = 1$ for all $t \in [0, 1]$). Let $T_{\bar{a}}$ be a subject's potential failure time had (possibly contrary to what was observed

in the actual study) the subject been treated according to \bar{a} . Let \perp denote statistical independence. Assume

$$T = T_{\bar{a}} \quad \forall \bar{a} \text{ such that } a(t) = A(t) \quad \forall t \leq T, \quad (2.1)$$

$$T_{\bar{a}} \perp A(t) | \bar{A}(t^-), \bar{L}(t) \quad \forall \bar{a}, \quad (2.2)$$

$$\text{pr}[A(t) | \bar{A}(t^-), \bar{L}(t)] > 0 \quad \forall t \in [0, 1] \text{ such that } \text{pr}[\bar{A}(t^-), \bar{L}(t)] > 0. \quad (2.3)$$

These are referred to as the *causal consistency*, *conditional exchangeability*, and *positivity* assumptions, respectively. Assumption (2.1) states that, in the absence of censoring, the observed failure time T equals the potential failure time $T_{\bar{a}}$ for all treatment plans \bar{a} compatible with the observed treatment up to time T . Assumption (2.2) ensures no unmeasured confounding. Assumption (2.3) states that the conditional probability of receiving any particular treatment is greater than zero. Of the three, only (2.3) can be tested empirically. Sensitivity analysis may be useful in assessing the robustness of inference drawn to violations of (2.2) (Robins, Rotnitzky, and Scharfstein (1999)).

Consider the MSCM

$$\lambda_{T_{\bar{a}}}(t) = \lambda_0(t) \exp\{\beta_0' f(\bar{a}(t))\},$$

where $\lambda_{T_{\bar{a}}}(t)$ is the hazard of failure at time t if all individuals in the population had followed treatment plan \bar{a} through time t , $\lambda_0(t)$ is an unspecified baseline hazard function corresponding to the hazard if all individuals had been untreated through time t , $f(\bar{a}(t))$ is a specified function of treatment history up to time t , and β_0 is an unknown parameter vector. Hereafter, we consider the MSCM

$$\lambda_{T_{\bar{a}}}(t) = \lambda_0(t) r\{\beta_0' a(t)\} \quad (2.4)$$

where, for notational convenience, we let $r\{\cdot\} = \exp\{\cdot\}$. For example, if we are interested in the causal effect of current AZT treatment on mortality of HIV-positive homosexual men, then $r(\beta_0)$ is the ratio of the hazard of death at time t had all subjects in the population alive at time t been exposed to AZT compared to being unexposed at time t . Here (2.4) focuses on the effect of current treatment status only; however, the results presented below are valid for any specified $f(\bar{a}(t))$.

We employ the counting process framework to study the large sample behavior of estimators of β_0 . All processes discussed hereafter refer to observed processes. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space and let $\{\mathcal{F}_t : t \in [0, 1]\}$ be an increasing right-continuous family of sub σ -algebras of \mathcal{F} consisting of failure times, covariates, and treatment histories up to time t , and censoring histories up to time t^+ for all subjects in a cohort of size n . Thus the filtration with respect to the probability space is the same as usual, except that treatment histories are

now listed separately from covariate histories. Let $N_i(\cdot)$ be a counting process adapted to \mathcal{F}_t representing the number of failures of subject i by time t so that $dN_i(t)$ indicates the number of events of subject i that occurred in $[t, t + dt)$ for sufficiently small dt . Because failures are assumed to occur in continuous time, we only allow jumps of size 1 and no simultaneous jumps can occur in $[t, t + dt)$. Let $C_i(t) = 0$ indicate that subject i remained uncensored prior to time t and $C_i(t) = 1$ otherwise. The treatment process $A_i(\cdot)$ and the censoring process $C_i(\cdot)$ are assumed to be piece-wise constant point processes with cadlag (right-continuous with left-hand limits) step-function sample paths. The processes $A(\cdot)$ and $C(\cdot)$ are assumed to have jumps that can occur at no more than a finite number of time points. Informally, all participants follow (approximately) the same visit schedule. This assumption should be reasonable in studies with regularly scheduled follow-up visits (e.g., every six months) and good study compliance. We refer to censoring as ignorable (or noninformative) if the cause-specific hazard of being censored at t among subjects alive and uncensored does not depend on the failure times $T_{\bar{a}}$ given prior treatment/covariate history $\bar{A}(t^-)$ and $\bar{L}(t^-)$ (Hernán, Brumback, and Robins (2001)). Let $Y_i(t) = I\{N_i(t) = C_i(t) = 0\}$ denote whether an individual is at-risk of being observed to fail at time t , having left-continuous sample paths, and assume $\text{pr}[Y(1) > 0] > 0$.

2.2. Inverse probability weights

If we can correctly model the probability of receiving treatment at time t given the past treatment history and covariate history, then we can consistently estimate the weights

$$W^T(t) = \prod_{k \leq t} \frac{\text{pr}[A(k) | \bar{A}(k^-)]}{\text{pr}[A(k) | \bar{A}(k^-), \bar{L}(k)]}. \quad (2.5)$$

These are referred to as *stabilized* inverse-probability-of-treatment-weights (IPTWs). If the numerator probabilities in (2.5) were replaced with 1, then the weights are referred to as *unstabilized* IPTWs. We can consistently estimate the numerator probabilities in (2.5) based on sample proportions because $A(\cdot)$ is assumed to have at most a finite number of jumps over the study period. Under (2.2) to (2.3), in the absence of censoring, Robins (1999) showed that a consistent estimator of the unknown parameter β_0 in (2.4) can be obtained by fitting an ordinary time-dependent Cox model with the contribution of subject i to the risk set at time t weighted by estimates of (2.5). Informally we can think of the analysis via IPTWs as reweighting the observed data set such that it has the same properties as a random sample, with respect to the measured confounders L , from a population where $\bar{L}(t) \perp A(t) | \bar{A}(t^-)$ holds at time t . The weighted study population is sometimes called a *pseudo-population*.

Dropout may introduce selection bias if it is associated with exposure and the outcome. In the presence of such censoring, we can still obtain a consistent estimator of β_0 by fitting the ordinary Cox model, but weighting a subject alive and uncensored at time t by estimates of $W^T(t) \times W^C(t)$, where

$$W^C(t) = \prod_{k \leq t} \frac{\text{pr}[C(k) = 0 | \bar{C}(k^-) = 0, \bar{A}(k^-)]}{\text{pr}[C(k) = 0 | \bar{C}(k^-) = 0, \bar{A}(k^-), \bar{L}(k)]}. \quad (2.6)$$

This is under the assumption of no unmeasured confounders for censoring, an analogous assumption to (2.3) for censoring, and assuming that we can correctly model the denominator probabilities in (2.6) (Robins (1999)). Here the weighted study population can be thought of as a pseudo-population in which there is no confounding due to measured covariates or selection bias due to censoring. In Section 2.3, we make use of the stabilized weights defined by $W(t) \equiv W^T(t) \times W^C(t)$ after modifying (2.5) by adding $C(k) = 0$ to the conditioning events in both the numerator and the denominator (Hernán, Brumback, and Robins (2000)). Hereafter $W(t)$ is referred to as inverse-probability-weights (IPWs). Here (2.5) and (2.6) are finite products and (2.3) ensures non-zero probabilities in their denominators; hence the IPWs are bounded at all t . Our results are not limited to a specific form of the weights $W(t)$. The proposed methods are applicable to different inverse probability weighting analysis provided that the IPWs (or IPTWs in the absence of censoring) are bounded, such as when truncated (Cole and Hernán (2008)) and normalized (Xiao, Abrahamowicz, and Moodie (2010)) weights are employed. Under the assumption of finite support of the treatment and censoring processes, unstabilized weights (where the numerators probabilities of both (2.5) and (2.6) are replaced with 1) are also bounded, but are highly variable and monotone increasing functions of t . Other weights such as stabilized, truncated, and normalized weights are generally recommended in practice as they lead to more efficient estimators of the causal treatment effect.

We now briefly describe estimation of the random weights $W(t)$, denoted by $\hat{W}(t)$. One can specify a pooled logistic model (treating each person-visit as an observation) to estimate the probability in the denominators of (2.5) and (2.6) at each time (for example, at each visit), then plug in the estimated probabilities (Hernán, Brumback, and Robins (2000, 2001)). In the presence of case-cohort sampling, the same model can be used to obtain $\hat{W}(t)$ after weighting subcohort controls by the inverse probability of subcohort selection. Example SAS code to obtain $\hat{W}(t)$ using case-cohort data is provided in Section 2 of the supplement. We assume throughout that the models to estimate denominator probabilities in the IPWs are correctly specified. In practice, investigators may wish to explore

the sensitivity of treatment effect estimates to different model specifications for estimating the weights.

2.3. Weighted-pseudo-partial-likelihood

We consider two weighted-pseudo-partial-likelihoods (WPPL) that form the basis for obtaining consistent estimators of β_0 in the presence of case-cohort sampling. They are formed by weighting individual contributions to the usual partial likelihoods by $W_i(t)$ assuming that $W_i(t)$ is known.

The log-WPPL created by individual-time-specific weights at time t under the full cohort setting is given by

$$l(\beta, t; W) = \sum_{i=1}^n \int_0^t W_i(u) \left[\beta' A_i(u) - \log \sum_{l=1}^n W_l(u) Y_l(u) r\{\beta' A_l(u)\} \right] dN_i(u); \tag{2.7}$$

it is motivated by the weighted estimating equations proposed by Robins (1993). The log-WPPL in the case-cohort setting is

$$\tilde{l}(\beta, t; W) = \sum_{i=1}^n \int_0^t W_i(u) \left[\beta' A_i(u) - \log \sum_{l \in \tilde{C}} W_l(u) Y_l(u) r\{\beta' A_l(u)\} \right] dN_i(u). \tag{2.8}$$

This is slightly different from the log-WPPL proposed by Cole et al. (2012),

$$l^*(\beta, t; W) = \sum_{i=1}^n \int_0^t W_i(u) \left[\beta' A_i(u) - \log \sum_{l \in \tilde{C} \cup \{i\}} W_l(u) Y_l(u) r\{\beta' A_l(u)\} \right] dN_i(u). \tag{2.9}$$

Here (2.8) and (2.9) differ only in whether a case outside the subcohort \tilde{C} contributes to the risk set. If $W_i(u) = 1$ for all i and u , (2.8) reduces to the log-likelihood considered by Self and Prentice (1988) and (2.9) reduces to the log-likelihood considered by Prentice (1986).

Let $\hat{\beta}$, $\tilde{\beta}$, and β^* be solutions to $\partial l(\beta, 1; \hat{W})/\partial \beta = 0$, $\partial \tilde{l}(\beta, 1; \hat{W})/\partial \beta = 0$, and $\partial l^*(\beta, 1; \hat{W})/\partial \beta = 0$, respectively. Based on arguments as in Andersen and Gill (1982), in Theorems 3.1–3.2 we show $\hat{\beta}$ and $\tilde{\beta}$ are consistent estimators of β_0 . That $\beta^* \rightarrow_p \beta_0$ can be shown analogously. Asymptotic normality of $\hat{\beta}$ and $\tilde{\beta}$ will be shown via asymptotic normality of score statistics corresponding to (2.7) and (2.8).

To make use of counting process and martingale theory, under (2.1) each (observed) counting process $N_i(\cdot)$ ($i = 1, \dots, n$) can be uniquely decomposed into the sum of its intensity process λ_i and a local square integrable martingale M_i ,

$$N_i(t) = \int_0^t \lambda_i(u) du + M_i(t), \quad t \in [0, 1], \tag{2.10}$$

where the intensity process is given by

$$\lambda_i(t) = Y_i(t)r\{\beta'_0 A_i(t)\}\lambda_0(t), \quad (2.11)$$

which embodies the same parameters as in (2.4).

3. Asymptotic Properties

In this section, we present the main results: consistency and asymptotic normality of the estimators $\hat{\beta}$, $\tilde{\beta}$, and β^* . Sufficient conditions for these results are stated in the Appendix, followed by discussion of the conditions. Proofs of the theorems are given in S1 of the supplementary document.

Theorem 1. (*Consistency of $\hat{\beta}$ under full cohort*) Under conditions A–F, $\hat{\beta} \rightarrow_p \beta_0$.

Theorem 2. (*Consistency of $\tilde{\beta}$ under the case-cohort*) Under conditions A–G, $\tilde{\beta} \rightarrow_p \beta_0$.

It is straightforward to show that the estimator based on (2.9) converges in probability to the same limit as $\tilde{\beta}$. An individual case's contribution to \tilde{C} at its failure time (which is weighted by its IPWs) is asymptotically negligible in the sense that IPWs are bounded at all times and weighted subcohort averages are asymptotically stable (see conditions B and G-3 in the Appendix).

Theorem 3. Under conditions A–G, $\tilde{\beta} - \beta^* \rightarrow_p 0$.

We need some additional notation. Let $c^{\otimes 0} = 1$, $c^{\otimes 1} = c$, $c^{\otimes 2} = cc'$, and c_i denote the i -th element of c for a $p \times 1$ column vector c . Let $r^{(j)}\{\beta' A(t)\} = A(t)^{\otimes j} r\{\beta' A(t)\}$, $j = 0, 1, 2$. Full cohort averages are defined by $S^{(j)}(\beta, t) = n^{-1} \sum_{i=1}^n Y_i(t)r^{(j)}\{\beta' A_i(t)\}$ for $j = 0, 1, 2$, as given in Andersen and Gill (1982) with covariates $Z_i(t)$ being replaced by the treatment process $A_i(t)$. Similarly, subcohort averages are defined by $\tilde{S}^{(j)}(\beta, t) = n^{-1} \sum_{i \in \tilde{C}} Y_i(t)r^{(j)}\{\beta' A_i(t)\}$. Limits of $S^{(j)}(\beta, t)$ and $\tilde{S}^{(j)}(\beta, t)$ are given by $s^{(j)}(\beta, t)$, formally defined in the Appendix. Analogously, let weighted full cohort averages be $S_{W^{(k)}}^{(j)}(\beta, t) = n^{-1} \sum_{i=1}^n W_i(t)^k Y_i(t)r^{(j)}\{\beta' A_i(t)\}$ and weighted subcohort averages be $\tilde{S}_{W^{(k)}}^{(j)}(\beta, t) = \tilde{n}^{-1} \sum_{i \in \tilde{C}} W_i(t)^k Y_i(t)r^{(j)}\{\beta' A_i(t)\}$ for $j = 0, 1, 2$ and $k = 1, 2$, with limits $s_{W^{(k)}}^{(j)}(\beta, t)$. Let $E = S^{(1)}/S^{(0)}$ and $\tilde{E} = \tilde{S}^{(1)}/\tilde{S}^{(0)}$ with limit $e = s^{(1)}/s^{(0)}$, and let $E_{W^{(k)}} = S_{W^{(k)}}^{(1)}/S_{W^{(k)}}^{(0)}$ and $\tilde{E}_{W^{(k)}} = \tilde{S}_{W^{(k)}}^{(1)}/\tilde{S}_{W^{(k)}}^{(0)}$ with limits $e_{W^{(k)}}$. Lastly, let $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$ and $v_{W^{(k)}} = s_{W^{(k)}}^{(2)}/s_{W^{(k)}}^{(0)} - e_{W^{(k)}}^{\otimes 2}$, and let $\Sigma = \int_0^1 v(\beta_0, t)s^{(0)}(\beta_0, t)\lambda_0(t)dt$ and $\Sigma_{W^{(k)}} = \int_0^1 v_{W^{(k)}}(\beta_0, t)s_{W^{(k)}}^{(0)}(\beta_0, t)\lambda_0(t)dt$.

Theorem 4 (Asymptotic normality of the full cohort MSCM score statistic). *Let $U(\beta_0, t) = \partial l(\beta, t)/\partial \beta|_{\beta=\beta_0}$ be the full cohort MSCM score process at time t . Under conditions A–F,*

$$n^{-1/2}U(\beta_0, 1) \rightarrow_d N(0, \Sigma_U),$$

where $\Sigma_U = \Sigma_{W(2)} + \Delta_{W(1),W(2)}$ and

$$\Delta_{W(1),W(2)} = \int_0^1 \{e_{W(2)}(\beta_0, u) - e_{W(1)}(\beta_0, u)\}^{\otimes 2} s_{W(2)}^{(0)}(\beta_0, u) \lambda_0(u) du.$$

Theorem 4 can be used to show asymptotic normality of the MSCM case-cohort score statistic:

Theorem 5 (Asymptotic normality of the case-cohort MSCM score statistic). *Let $\tilde{U}(\beta_0, t) = \partial \tilde{l}(\beta, t)/\partial \beta|_{\beta=\beta_0}$ be the case-cohort MSCM Score process at time t . Under conditions A–G,*

$$n^{-1/2}\tilde{U}(\beta_0, 1) \rightarrow_d N(\mathbf{0}, \Sigma_{\tilde{U}}),$$

where $\Sigma_{\tilde{U}} = \Sigma_U + \Delta_\alpha$,

$$\Delta_\alpha = \int_0^1 \int_0^1 G(\beta_0, x, v) \lambda_0(x) \lambda_0(v) dx dv,$$

$$G(\beta_0, x, v) = (1 - \alpha)\alpha^{-1} \left[h^{(1)}(\beta_0, x, v) - e_{W(1)}(\beta_0, x) h^{(2)}(\beta_0, x, v)' - h^{(2)}(\beta_0, v, x) e_{W(1)}(\beta_0, v)' + e_{W(1)}(\beta_0, x) e_{W(1)}(\beta_0, v)' h^{(0)}(\beta_0, x, v) \right],$$

$$h^{(0)}(\beta, x, v) = q^{(0)}(\beta, x, v) - s_{W(1)}^{(0)}(\beta, x) s_{W(1)}^{(0)}(\beta, v),$$

$$h^{(1)}(\beta, x, v) = q^{(1)}(\beta, x, v) - s_{W(1)}^{(1)}(\beta, x) s_{W(1)}^{(1)}(\beta, v)', \text{ and}$$

$$h^{(2)}(\beta, x, v) = q^{(2)}(\beta, x, v) - s_{W(1)}^{(0)}(\beta, x) s_{W(1)}^{(1)}(\beta, v),$$

with $q^{(j)}(\beta, \cdot, \cdot)$ defined in condition G-2 in the Appendix.

Theorem 6 (Asymptotic normality of $\tilde{\beta}$). *Under conditions A–G,*

$$n^{1/2}(\tilde{\beta} - \beta_0) \rightarrow_d N(\mathbf{0}, \Sigma_{W(1)}^{-1} \Sigma_{\tilde{U}} \Sigma_{W(1)}^{-1})$$

where $\Sigma_{\tilde{U}}$ is given in Theorem 5.

Based on Theorem 6, we propose a new variance estimator

$$\widehat{\text{var}}(\tilde{\beta}) = n^{-1} \hat{\Sigma}_{W(1)}^{-1} (\hat{\Sigma}_{W(2)} + \hat{\Delta}_{W(1),W(2)} + \hat{\Delta}_\alpha) \hat{\Sigma}_{W(1)}^{-1}, \tag{3.1}$$

where

$$\tilde{\mathcal{I}}(\tilde{\beta}, 1; W) = -\frac{\partial^2 \tilde{l}(\beta, 1; W)}{\partial \beta^2 |_{\beta=\tilde{\beta}}},$$

$$\hat{\Sigma}_{W(1)} = n^{-1} \tilde{\mathcal{I}}(\tilde{\beta}, 1; W = \hat{W}), \quad (3.2)$$

$$\hat{\Sigma}_{W(2)} = n^{-1} \tilde{\mathcal{I}}(\tilde{\beta}, 1; W^2 = \hat{W}^2), \quad (3.3)$$

$$\hat{\Delta}_{W(1), W(2)} = n^{-1} \sum_{i=1}^{\tilde{n}} \int_0^1 \hat{W}_i(u)^2 \left[\tilde{E}_{\{W(2)=\hat{W}^2\}}(\tilde{\beta}, u), -\tilde{E}_{\{W(1)=\hat{W}\}}(\tilde{\beta}, u) \right]^{\otimes 2} dN_i(u), \quad (3.4)$$

$$\hat{\Delta}_\alpha = n^{-2} \int_0^1 \int_0^1 \hat{G}(\tilde{\beta}, x, v) \tilde{S}_{W(1)}^{(0)}(\tilde{\beta}, x)^{-1} \tilde{S}_{W(1)}^{(0)}(\tilde{\beta}, v)^{-1} d\bar{N}_{\hat{W}}(x) d\bar{N}_{\hat{W}}(v). \quad (3.5)$$

Here $\tilde{E}_{\{W(2)=\hat{W}^2\}}$ and $\tilde{E}_{\{W(1)=\hat{W}\}}$ denote that the IPWs in $\tilde{E}_{W(2)}$ and $\tilde{E}_{W(1)}$ are replaced by \hat{W}^2 or \hat{W} , $\bar{N}_{\hat{W}}(t)$ is defined by $\sum_i \hat{W}_i(t) N_i(t)$, and $\hat{G}(\cdot, \cdot, \cdot)$ is $G(\cdot, \cdot, \cdot)$ in Theorem 5 with $q^{(j)}(\beta_0, \cdot, \cdot)$ and $s_{W(1)}^{(j)}(\beta_0, \cdot)$ in $h^{(j)}(\beta_0, \cdot, \cdot)$ replaced by $\tilde{Q}^{(j)}(\tilde{\beta}, \cdot, \cdot)$ (defined in condition G-2 in the Appendix) and $\tilde{S}_{W(1)}^{(j)}(\tilde{\beta}, \cdot)$ along with $e_{W(1)}(\beta_0, \cdot)$ replaced by $\tilde{E}_{W(1)}(\tilde{\beta}, \cdot)$. Estimators (3.2), (3.3), and (3.4) are consistent estimators of $\Sigma_{W(1)}$, $\Sigma_{W(2)}$, and $\Delta_{W(1), W(2)}$ in view of condition A and the fact that $\sup_{\beta, t} |n^{-1} \{\mathcal{I}(\beta, t) - \tilde{\mathcal{I}}(\beta, t)\}| \rightarrow_p 0$ where $\mathcal{I}(\beta, t) = -\partial^2 l(\beta, t) / \partial \beta^2$ (see S1-2 of the supplementary document). Estimator (3.5) is a consistent estimator of Δ_α in view of conditions A, G-1(ii), that $n^{-1} \bar{N}_{\hat{W}}(t)$ uniformly converges to $\int_0^t s_{W(1)}^{(0)}(\beta_0, u) \lambda_0(u) du$, and that $n^{-1} \bar{N}_{\hat{W}}(1)$ is bounded in probability.

The proposed variance estimator (3.1) is different from the robust estimator proposed by Lin and Ying (LY, Lin and Ying (1993)) that is used in most MSM analyses. Both (3.1) and the LY estimator are sandwich-type estimators where the “bread” of sandwich ($\hat{\Sigma}_{W(1)}^{-1}$) is the same. However, the “meat” is different. In particular, the “meat” of the LY estimator is based on (weighted) score residuals whereas the “meat” of (3.1) is given by an estimator of $\Sigma_{\tilde{U}} + \Delta_\alpha$. Simulation results reported in §4 (and S2-2 in the supplement document) indicate that (3.1) can be more accurate when the size of subcohort is small.

4. Simulation

A simulation study was conducted to examine the finite sample bias of $\tilde{\beta}$ and β^* , and the performance of the proposed variance estimator (3.1) as well as the LY variance estimator. Simulations were conducted as in Cole et al. (2012). Briefly, potential survival times were generated according to the MSCM given in

(2.4), and observed survival times were generated by stochastically generating time varying exposures and confounders for cohorts of size $n = 1,000$ (see Cole et al. (2012) for details). While they considered only one scenario in which 25% of individuals were cases and a 20% subcohort fraction, $\tilde{n}n^{-1} \times 100 = 20$, we considered 36 scenarios by varying both the subcohort fraction and the event rate from 5 to 30%. Censoring times were generated from uniform distributions with support chosen to achieve the desired event rate. We did not incorporate (2.6) when calculating IPWs because the censoring times were generated independently of the exposure and potential survival times. Following Cole et al. (2012), stabilized weights were used to calculate IPWs. For each scenario 5,000 data sets were generated under the null $\beta_0 = 0$ and the alternative $\beta_0 = \log(1/2)$.

Results from the simulation study are summarized in Table 1. Only results obtained from six scenarios under the null are presented; results from other scenarios and under the alternative were similar (see S2 of the supplementary document for results under the alternative). For all scenarios, under both the null and alternative, $\tilde{\beta}$ and β^* were nearly unbiased; that the two estimators performed similarly is not surprising in light of Theorem 3. Under the null, the proposed variance estimator was always less biased than the LY variance estimator when the subcohort fraction was only 5%, regardless of the event rate. Similarly, (3.1) was less biased regardless of the subcohort fraction when the event rate was 5%. Both the proposed and the LY variance estimators were approximately unbiased when the subcohort fraction and event rate were both at least 20%. Wald confidence intervals (CIs) using the LY variance estimator tended to undercover when the subcohort fraction was 5%, whereas Wald CIs using (3.1) exhibited coverage close to the nominal level for all scenarios considered. In summary, both $\tilde{\beta}$ and β^* , along with the proposed variance estimator and CIs, exhibited good finite sample properties for the scenarios considered, while performance of the LY variance estimator depended on subcohort size and event rate.

5. Additional Considerations

5.1. Baseline cumulative hazard estimation

In addition to the treatment effect β , it may be of interest to estimate the cumulative baseline hazard function. Similar to the cumulative baseline hazard estimator proposed in Self and Prentice (1988), a consistent estimator of $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ with case-cohort sampling is

$$\tilde{\Lambda}_{\tilde{W}}(\tilde{\beta}, t) = \tilde{n}n^{-1} \int_0^t \left[\sum_{i \in \tilde{C}} \hat{W}_i(u) Y_i(u) r \{ \tilde{\beta}' A_i(u) \} \right]^{-1} \sum_{i=1}^n \hat{W}_i(u) dN_i(u). \quad (5.1)$$

This estimator is equivalent to Self and Prentice's estimator when $\hat{W}_i(t) = 1$ for all i and t . A consistent estimator of the cumulative baseline hazard can also

Table 1. Summary of simulation study. Bias denotes the empirical bias of the different estimators of β_0 . ESE denotes the empirical standard errors. ASE denotes the average estimated standard errors and Coverage denotes the empirical coverage of 95% Wald-type confidence intervals using either (3.1) or the LY variance estimator. † Of the 5,000 estimates of β^* and $\tilde{\beta}$, one pair was excluded because some of the stabilized IPWs were greater than 10^6 .

Sub-cohort(%)	Event rate(%)	Estimator	Bias	ESE	ASE		Coverage	
					proposed	LY	proposed	LY
5	5 †	β^*	-0.12	0.49	0.55	0.42	0.97	0.91
		$\tilde{\beta}$	-0.20	0.68	0.66	0.47	0.96	0.90
	25	β^*	-0.03	0.37	0.36	0.31	0.94	0.91
		$\tilde{\beta}$	-0.04	0.44	0.37	0.35	0.92	0.91
10	5	β^*	-0.05	0.40	0.42	0.37	0.97	0.94
		$\tilde{\beta}$	-0.06	0.44	0.43	0.39	0.96	0.94
	25	β^*	-0.02	0.27	0.26	0.26	0.93	0.93
		$\tilde{\beta}$	-0.02	0.29	0.26	0.26	0.93	0.93
20	5	β^*	-0.02	0.35	0.36	0.34	0.96	0.95
		$\tilde{\beta}$	-0.02	0.36	0.36	0.35	0.96	0.95
	25	β^*	-0.01	0.21	0.20	0.20	0.94	0.94
		$\tilde{\beta}$	-0.01	0.21	0.20	0.20	0.94	0.94

obtained by using β^* instead of $\tilde{\beta}$ in (5.1). In the supplement (5.1) is shown to be consistent under conditions A-G and its performance is examined in a simulation study; see S1 and S2 of the supplement.

5.2. Limitations and future directions

While Cole et al. (2012) extensively discussed limitations of the MSCM with case-cohort sampling, we reiterate some of the issues that are important from a theoretical point of view. First, the asymptotic results presented here correct model specification for both the treatment assignment model (2.5) and the censoring model (2.6). If (2.2) holds, flexible parametric regression models fit adjusting for all measured confounders and their histories should provide a good approximation to denominator probabilities in (2.5) and (2.6). However, (2.2) is an untestable assumption and thus in practice analysts may want to explore sensitivity of treatment effect estimates to departures from this assumption.

Second, the proposed MSCM estimators for the case-cohort study, which are based on Prentice (1986) and Self and Prentice (1988) type estimators, are not expected to be fully efficient. After Prentice (1986) and Self and Prentice (1988), various methods have been proposed as means of improving the efficiency of the hazard ratio estimator in the standard case-cohort Cox regression analysis. Those methods seek to improve efficiency mostly by using weighted partial-likelihood estimation. For example, Barlow (1994), Barlow et al. (1999),

Kong, Cai, and Sen (2006), and Kong and Cai (2009) utilize fixed and time-varying inverse-probability-sampling weights to account for subcohort sampling to improve efficiency. Borgan et al. (2000), Kulich and Lin (2004), and Breslow et al. (2009a,b) consider leveraging phase 1 covariates (available from the entire cohort) to improve efficiency. These methods could potentially be extended to develop more efficient MSCM estimators in the presence of case-cohort sampling.

Supplementary Materials

The online supplementary material contains proofs corresponding to consistency and asymptotic normality of MSCM parameter estimators along with consistency of the cumulative baseline hazard estimator proposed in Section 5.1, implementation of the method using standard survival analysis software such as R or SAS, and additional simulation study results including performance of the proposed baseline cumulative hazard estimator. A summary of notation introduced in the main text and the supplement is presented.

Acknowledgement

We thank Professor Danyu Lin for expert advice and the Editor and three anonymous reviewers for helpful comments that improved the paper. The authors were partially supported by the Lifespan/Tufts/Brown Center for AIDS Research (CFAR), University of North Carolina CFAR, and National Institutes of Health grants P30 AI042853, P30 AI50410, R01 AI085073, UL1RR025747, R01 ES021900, P01 CA142538, and R01 AI100654. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix: Regularity Conditions

In what follows, norms are defined by $\|c^{\otimes 2}\| = \sup_{i,j} |c_i c_j|$, $\|c\| = \sup_i |c_i|$, and $|c| = (\sum c_i^2)^{1/2} = (c'c)^{1/2}$.

A (Uniform consistency of estimated weights)

$$\sup_{\substack{i \in \{1, \dots, n\} \\ t \in [0, 1]}} |\hat{W}_i(t) - W_i(t)| \equiv M_{\hat{W}} \rightarrow_p 0.$$

B (Stability of weights) Individual time-specific weights $W_i(t)$ and the corresponding estimators $\hat{W}_i(t)$ are strictly positive and bounded, with positive real numbers M_1 and M_2 such that

$$\sup_{\substack{i \in \{1, \dots, n\} \\ t \in [0, 1]}} W_i(t) \leq M_1, \quad \text{and} \quad \sup_{\substack{i \in \{1, \dots, n\} \\ t \in [0, 1]}} \hat{W}_i(t) \leq M_2.$$

C (Finite interval) $\int_0^1 \lambda_0(t) dt < \infty$.

D (Asymptotic stability)

(i) There exists a neighborhood \mathcal{B}_0 of β_0 and functions $s^{(0)}$, $s^{(1)}$, and $s^{(2)}$ defined on $\mathcal{B}_0 \times [0, 1]$ such that

$$\sup_{\substack{\beta \in \mathcal{B}_0 \\ t \in [0, 1]}} \|S^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \rightarrow_p 0, \quad j = 0, 1, 2.$$

(ii) There exists a neighborhood \mathcal{B} of β_0 , $\mathcal{B} \subseteq \mathcal{B}_0$, and functions $s_{W^{(k)}}^{(j)}$ defined on $\mathcal{B} \times [0, 1]$ such that

$$\sup_{\substack{\beta \in \mathcal{B} \\ t \in [0, 1]}} \|S_{W^{(k)}}^{(j)}(\beta, t) - s_{W^{(k)}}^{(j)}(\beta, t)\| \rightarrow_p 0, \quad j = 0, 1, 2; k = 1, 2.$$

(iii) $n^{1/2}\{S_{W^{(1)}}^{(0)}(\beta_0, t) - s_{W^{(1)}}^{(0)}(\beta_0, t)\} \rightarrow_d N(0, \sigma^2(t))$ uniformly in $t \in [0, 1]$, for some $\sigma^2(t)$.

E (Lindeberg condition) For any $\epsilon > 0$, $j = 1, \dots, p$

$$\begin{aligned} n^{-1} \int_0^1 \sum_{i=1}^n W_i(u)^2 [A_{ij}(u) - E_{W^{(1)}}(\beta_0, u)_j]^2 Y_i(u) r\{\beta'_0 A_i(u)\} \\ \times I\{n^{-1/2} W_i(u) |A_{ij}(u) - E_{W^{(1)}}(\beta_0, u)_j| > \epsilon\} \lambda_0(u) du \rightarrow_p 0. \end{aligned}$$

F (Asymptotic regularity conditions) $s^{(j)}(\beta, t)$ and $s_{W^{(k)}}^{(j)}(\beta, t)$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, 1]$ that are bounded on $\mathcal{B} \times [0, 1]$ for $j = 0, 1, 2$ and $k = 1, 2$. For all $(\beta, t) \in \mathcal{B} \times [0, 1]$ and $m = 0, 1$,

$$s^{(m+1)}(\beta, t) = \frac{\partial s^{(m)}(\beta, t)}{\partial \beta}, \quad s_{W^{(k)}}^{(m+1)}(\beta, t) = \frac{\partial s_{W^{(k)}}^{(m)}(\beta, t)}{\partial \beta}.$$

Here $s^{(0)}$ and $s_{W^{(k)}}^{(0)}$ are bounded away from zero and the matrices Σ and $\Sigma_{W^{(k)}}$ are positive definite.

G-1 (Stability of subcohort average)

(i) (Nontrivial subcohort) $\tilde{n}n^{-1} \rightarrow_p \alpha$ for some $\alpha \in (0, 1]$.

(ii) (Asymptotic normality of subcohort averages at β_0) For any $\epsilon > 0$

$$\sup_{t \in [0, 1]} n^{-1} \sum_{i=1}^n W_i(t)^2 Y_i(t) r\{\beta'_0 A_i(t)\}^2 I\{n^{-1/2} W_i(t) Y_i(t) r\{\beta'_0 A_i(t)\} > \epsilon\} \rightarrow_p 0,$$

$$\begin{aligned} \sup_{t \in [0, 1]} n^{-1} \sum_{i=1}^n W_i(t)^2 Y_i(t) \|r^{(1)}\{\beta'_0 A_i(t)\}\|^2 I\{n^{-1/2} W_i(t) Y_i(t) \|r^{(1)}\{\beta'_0 A_i(t)\}\| \\ > \epsilon\} \rightarrow_p 0, \end{aligned}$$

and the sequences of distributions of $n^{1/2}\{\tilde{E}(\beta_0, t) - E(\beta_0, t)\}$ are tight on the product space of cadlag functions equipped with the product Skorohod topology, and so are $n^{1/2}\{\tilde{E}_{W_{(1)}}(\beta_0, t) - E_{W_{(1)}}(\beta_0, t)\}$.

G-2 (Asymptotic stability and regularity of covariance function) There exists a neighborhood \mathcal{B} of β_0 and functions $q^{(j)}(\beta, t, u)$ for $j = 0, 1, 2$, defined on $\mathcal{B} \times [0, 1]^2$ such that $q^{(j)}(\beta, t, u)$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $(t, u) \in [0, 1]^2$, the $q^{(j)}$ are bounded on $\mathcal{B} \times [0, 1]^2$ and

$$\sup_{\substack{\beta \in \mathcal{B} \\ (t,u) \in [0,1]^2}} \|Q^{(j)}(\beta, t, u) - q^{(j)}(\beta, t, u)\| \rightarrow_p 0, \quad j = 0, 1, 2, \quad \text{where}$$

$$Q^{(0)}(\beta, t, u) = n^{-1} \sum_{i=1}^n W_i(t) Y_i(t) r\{\beta'_0 A_i(t)\} W_i(u) Y_i(u) r\{\beta'_0 A_i(u)\},$$

$$Q^{(1)}(\beta, t, u) = n^{-1} \sum_{i=1}^n W_i(t) Y_i(t) r^{(1)}\{\beta'_0 A_i(t)\} W_i(u) Y_i(u) r^{(1)}\{\beta'_0 A_i(u)\}',$$

$$Q^{(2)}(\beta, t, u) = n^{-1} \sum_{i=1}^n W_i(t) Y_i(t) r\{\beta'_0 A_i(t)\} W_i(u) Y_i(u) r^{(1)}\{\beta'_0 A_i(u)\}.$$

Moreover, $\sup_{n \geq 1} \mathcal{E}[Q^{(j)}(\beta, t, u)]$ for $j = 0, 1, 2$ are bounded sequences where \mathcal{E} denote expectation.

G-3 (Asymptotic stability of subcohort averages) If $\tilde{Q}^{(j)}(\beta, t, u)$ are covariance functions based on subcohort members $i = 1, \dots, \tilde{n}$, then

$$\sup_{\substack{\beta \in \mathcal{B} \\ t \in [0,1]}} \|\tilde{S}_{W^{(k)}}^{(0)}(\beta, t) - s_{W^{(k)}}^{(0)}(\beta, t)\| \rightarrow_p 0 \quad k = 1, 2,$$

the subcohort average converges to the limit of the full cohort in probability uniformly in $\beta \in \mathcal{B}$ and $t \in [0, 1]$, and

$$\sup_{\substack{\beta \in \mathcal{B} \\ (t,u) \in [0,1]^2}} \|\tilde{Q}^{(j)}(\beta, t, u) - q^{(j)}(\beta, t, u)\| \rightarrow_p 0, \quad j = 0, 1, 2,$$

the subcohort covariance functions converge in probability uniformly in $\beta \in \mathcal{B}$ and $t \in [0, 1]$ to the full cohort covariance functions. In addition,

$$n^{1/2}\{\tilde{S}_{W_{(1)}}^{(0)}(\beta_0, t) - s_{W_{(1)}}^{(0)}(\beta_0, t)\} \rightarrow_d N(0, \tilde{\sigma}^2(t)), \quad \text{uniformly in } t \in [0, 1]$$

for some $\tilde{\sigma}^2(t)$.

All conditions except A and B are extensions of the regularity conditions from Self and Prentice (1988) by incorporating IPWs. Conditions A and B are required for IPWs to ensure asymptotic properties of the proposed estimators.

Condition A: $\hat{W}(\cdot)$ and $W(\cdot)$ are assumed to be predictable with respect to the filtration \mathcal{F}_t because weights are determined by predictable processes: $A(\cdot)$, $L(\cdot)$, and their histories. Then along with the assumption of no misspecification of the model used to estimate denominator probabilities in $W(\cdot)$, the finite number of jumps assumption on the treatment and censoring processes are sufficient for this condition to hold. From a practical point of view, having a finite number of time points when treatment status can change or when censoring might occur may be reasonable to assume. For instance, studies often have planned visits at finite discrete intervals when a patient can have treatment altered. Similarly, the censoring time for a subject is often assumed to be the last observed visit time before the subject became lost-to-follow-up.

Condition B: All weights discussed in Section 2.2 satisfy conditions A and B in general, except the unstabilized weights. Unstabilized weights satisfy conditions A and B only when the assumption of finite support of $A(\cdot)$ and $C(\cdot)$ is met. The IPWs are strictly positive by (2.3).

Conditions C and D: Condition C is the same as in Self and Prentice (1988). When the IPWs are equal to 1, $S_{W^{(k)}}^{(j)}$ in D is $S^{(j)}$ in Self and Prentice (1988) for all $j = 0, 1, 2$ and $k = 1, 2$. Then $\Sigma_{W^{(k)}}$ defined in Section 3 and Δ_α defined in Theorem 3.5 are Σ and Δ defined in Self and Prentice (1988), respectively. Hence, $\Sigma_{\tilde{U}}$ in Theorem 3.5 is $\Sigma + \Delta$, the asymptotic covariance matrix of the case-cohort score statistic in Self and Prentice (1988), in the absence of IPWs.

Condition E: If the treatment process $A(\cdot)$ is bounded (as assumed here) and B and F are satisfied, then E holds trivially.

Condition F: $e_{W^{(k)}}$ can be interpreted as the weighted average of a treatment function with the weights taking an inverse-probability weighted exponential form. The positive definite condition on Σ in Andersen and Gill (1982) can easily be extended to the $\Sigma_{W^{(k)}}$ assuming $W(t)$ are bounded away from zero on $t \in [0, 1]$.

Condition G: This is the same as condition G in Self and Prentice (1988), incorporating individual-specific time-varying weights $W_i(t) (i = 1, \dots, n)$. Conditions A-F are sufficient to prove consistency and asymptotic normality of $\hat{\beta}$. To show asymptotic properties of $\hat{\beta}$ and β^* , this additional condition is required to ensure asymptotic stability of certain quantities estimated using subcohort data.

References

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- Barlow, W. E. (1994). Robust variance estimation for the case-cohort design. *Biometrics* **50**, 1064-1072.

- Barlow, W. E., Ichikawa, L., Rosner, D. and Izumi, S. (1999). Analysis of case-cohort designs - A prospective cohort study. *J. Clin. Epidemiol.* **52**, 1165-1172.
- Borgan, O., Langholz, B., Samuelsen, S. O., Goldstein, L. and Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Anal.* **6**, 39-58.
- Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich, M. (2009a). Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Statist. Biosc.* **1**, 32-49.
- Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich, M. (2009b). Using the whole cohort in the analysis of case-cohort data. *Amer. J. Epidemiology* **169**, 1398-1405.
- Cole, S. R. and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *Amer. J. Epidemiology* **168**, 656-664.
- Cole, S. R., Hernán, M. A., Robins, J. M., Anastos, K., Chmiel, J., Detels, R., Ervin, C., Feldman, J., Greenblatt, R., Kingsley, L., Lai, S., Young, M., Cohen, M. and Munoz, A. (2003). Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Amer. J. Epidemiology* **158**, 687-694.
- Cole, S. R., Hudgens, M. G., Tien, P. C., Anastos, K., Kingsley, L., Chmiel, J. S. and Jacobson, L. P. (2012). Marginal structural models for case-cohort study designs to estimate the association of antiretroviral therapy initiation with incident AIDS or death. *Amer. J. Epidemiology* **175**, 381-390. [erratum: 175 (7), 732]
- Hernán, M. A., Brumback, B. and Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561-570.
- Hernán, M. A., Brumback, B. and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Amer. Statist. Assoc.* **96**, 440-448.
- Kulich, M. and Lin, D. (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *J. Amer. Statist. Assoc.* **99**, 832-844.
- Kong, L., Cai, J. and Sen, P. K. (2006). Asymptotic results for fitting semiparametric transformation models to failure time data from case-cohort studies. *Statist. Sinica* **16**, 135-151.
- Kong, L. and Cai, J. (2009). Case-cohort analysis with accelerated failure time model. *Biometrics* **65**, 135-142
- Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *J. Amer. Statist. Assoc.* **88**, 1341-1349.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period - Application to control of the healthy worker survivor effect. *Math. Modelling* **7**, 1393-1512.
- Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24-33.
- Robins, J. M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine* **17**, 269-302.
- Robins, J. M. (1999). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models Epidemiology: The Environment and Clinical Trials* (Edited by M. E. Halloran and D. A. Berry), 95-134. Springer-Verlag, NY.

- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology - Methodological Issues* (Edited by N. P. Jewell, K. Dietz, and V. T. Farewell), 297-331. Birkhäuser, Boston.
- Robins, J. M., Rotnitzky, A. and Scharfstein, D. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology: The Environment and Clinical Trials* (Edited by M. Halloran and D. Berry), 1-92. Springer-Verlag, NY.
- Self, S. G. and Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64-81.
- Therneau, T. M. and Li, H. (1999). Computing the Cox model for case cohort designs. *Lifetime Data Anal.* **5**, 99-112.
- Therneau, T. (2013). A Package for Survival Analysis in S. R package version 2.37-4, URL: <http://CRAN.R-project.org/package=survival>.
- VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. *Epidemiology* **20**, 880-883.
- Xiao, Y., Abrahamowicz, M. and Moodie, E. E. M. (2010). Accuracy of conventional and marginal structural Cox model estimators: A simulation study. *Internat. J. Biostatist.* **6**, Article 13.

Department of Biostatistics, Brown University, Providence, Rhode Island 02912, USA.

E-mail: hlee@stat.brown.edu

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

E-mail: mhudgens@email.unc.edu

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

E-mail: cai@bios.unc.edu

Department of Epidemiology, University of North Carolina at Chapel Hill.

E-mail: cole@unc.edu

(Received January 2014; accepted June 2015)