Ashkan Ertefaie*, Masoud Asgharian and David A. Stephens

Double Bias: Estimation of Causal Effects from Length-Biased Samples in the Presence of Confounding

Abstract: Length bias in survival data occurs in observational studies when, for example, subjects with shorter lifetimes are less likely to be present in the recorded data. In this paper, we consider estimating the causal exposure (treatment) effect on survival time from observational data when, in addition to the lack of randomization and consequent potential for confounding, the data constitute a length-biased sample; we hence term this a double-bias problem. We develop estimating equations that can be used to estimate the causal effect indexing the structural Cox proportional hazard and accelerated failure time models for point exposures in double-bias settings. The approaches rely on propensity score-based adjustments, and we demonstrate that estimation of the propensity score must be adjusted to acknowledge the length-biased sampling. Large sample properties of the estimators are established and their small sample behavior is studied using simulations. We apply the proposed methods to a set of, partly synthesized, length-biased survival data collected as part of the Canadian Study of Health and Aging (CSHA) to compare survival of subjects with dementia among institutionalized patients versus those recruited from the community and depict their adjusted survival curves.

Keywords: Causal Inference, Length-biased Sampling, Propensity Score, Cox Proportional Hazard model, Accelerated Failure Time model.

DOI 10.1515/ijb-2014-0037

1 Introduction

In many observational studies, logistic or other constraints may render recruitment of disease-free patients for follow-up studies infeasible. In such cases, subjects who already experienced the initiation of the disease *prior* to recruitment (i.e. *prevalent* cases) are sampled. It is well known that subjects so recruited do not form a representative sample from the target population because subjects with longer survival time have greater chance to be recruited into the study. When the disease has stationary incidence, the induced bias in sampling is called *length bias* [1, 2]. This bias in sampling can lead to bias in the estimation of an exposure effect of interest.

Length-biased sampling can affect the sampling distribution of the covariates, such that covariates associated with the longer survivors have a higher chance of being selected. Recently, Bergeron et al. [3], Shen et al. [4], Qin and Shen [5] and Ning et al. [6] studied analysis of covariates under biased sampling. Studies on length-biased sampling can be traced as far back as Wicksell [7], Fisher [8], Neyman [9], Cox and Lewis [1], Zelen and Feinlein [2] and Patil and Rao [10]. An updated review of the subject can be found in Asgharian et al. [11].

A second source of potential bias in estimation of treatment or exposure effects encountered in observational studies is *confounding*. In the simple case of binary exposure, when exposure is influenced by other predictors, individuals in each exposure group may have different characteristics, which yielding imbalanced covariate distributions across the different groups. If the predictors also influence outcome (say, survival time) this may also lead to bias in the estimated exposure effect. Under an assumption of no

*Corresponding author: Ashkan Ertefaie, Department of Statistics, University of Pennsylvania, Philadelphia, PA, USA, E-mail: ertefaie@wharton.upenn.edu

Masoud Asgharian, David A. Stephens, Department of Mathematics and Statistics, McGill University, Montréal, QC, Canada, E-mail: masoud@math.mcgill.ca, d.stephens@math.mcgill.ca

unmeasured confounding, a consistent exposure effect estimator can be obtained by two well-known methods: *Inverse probability of treatment weighting* (IPTW) and *propensity score regression* (PSR). Weighted proportional hazard (PH) models for right censored data were introduced by Binder [12] and Lin [13] in the survey sampling literature. Pugh, Robins, Lipsitz and Harrington [14] also presented a weighted PH estimation equation to adjust for missing covariates [15–18].

In a recent article, Ertefaie et al. [19] developed a method for estimating the propensity score in the presence of length-biased sampling. In this paper, we address estimation of total causal effects in the presence of both length-biased sampling and confounding, which we term the *double-bias* problem, in the analysis of survival data. Specifically, we develop augmented estimating equations based on PH and accelerated failure time (AFT) models that can be used to estimate the exposure effect. In both cases, the augmentation spaces are formed using the censoring mechanism to improve the efficiency.

The rest of this paper is organized as follows. In Section 2, we introduce concepts and notation used in the manuscript. Section 3 presents our proposed estimating equation for estimating the propensity score when data are subject to length-biased sampling. In Sections 4 and 5, we present our estimating equations to deal with length-biased sampling and confounding under PH and AFT modeling assumptions, respectively. Also, the large sample properties of the estimators derived from the proposed estimation procedure are presented. We examine the performance of the proposed approach via simulation, and, in Section 7, apply our method to analyze a set of length-biased right-censored survival data collected as part of the Canadian Study of Health and Aging (CSHA) investigating the effect on survival of institutionalization; see Wolfson et al. [20].

2 Length-biased sampling

2.1 Notation

Our notation is similar to that of Ertefaie et al. [19]. Our data comprise n i.i.d samples of $(\mathbf{X}, Y, D, A, C, R)$ where D and $\mathbf{X}$ are the binary treatment variable and the vector of covariates, respectively. A is the time from the onset of the disease to the recruitment time and R covers the time from the recruitment time to the event (residual life time). Accordingly, the observed lifetime is defined as $T = A + R$. In the presence of right censoring, C is the censoring time measured from the recruitment to the loss to follow up. The observed survival time is $Y = A + \min(R, C)$. The variables with superscript pop represent the population variables; variables without pop denote the observed truncated variables. Figure 1 illustrates the different random quantities introduced in this section. The symbols $\circ$ and $\times$ denote a censored lifetime and an observed failure, respectively.

Figure 1: The data structure.

Let F and f be the distribution and density of T^{pop} , respectively. If the onset times are generated by a stationary Poisson process (the so-called *stationarity assumption*), then

$$F_{LB}(t) = \frac{\int_0^t s dF(s)}{\int_0^\infty s dF(s)} = \frac{1}{\mu} \int_0^t s dF(s) \quad \text{and} \quad f_{LB}(t) = \frac{tf(t)}{\mu}, \quad (1)$$

if F_{LB} has a corresponding absolutely continuous density f_{LB} , where μ is the mean survival time under F . Equation (1) is derived under a uniform truncation assumption.

For $t > 0$, we define the process $\{N(t)\}$ by $N(t) = \mathbb{1}(Y < t, \delta = 1)$ where δ is the censoring indicator ($\delta = 1$ indicating failure). We use small letters to refer to the possible values of the corresponding capital letter random variable. Throughout the manuscript, we make the following standard assumptions:

- A1. The variable $(T^{\text{pop}}, D^{\text{pop}}, \mathbf{X}^{\text{pop}})$ is independent of the calendar time of the onset of the disease.
- A2. The disease has stationary incidence, i.e. the disease incidence occurs at a constant rate.
- A3. The censoring time C is independent of $(A, R, D, \mathbf{X})$.

2.2 Counterfactual outcomes

We define the causal effect of interest using the counterfactual framework introduced by Rubin [21]. The counterfactual values $(A(d), R(d), Y(d))$ are representing the backward, forward recurrence times, and observed survival time, respectively, if $D = d$. Similarly, $T^{\text{pop}}(d)$ represents the counterfactual response. The observed response, T^{pop} , is defined as $DT^{\text{pop}}(1) + (1 - D)T^{\text{pop}}(0)$.

We make the following standard causal assumptions to link the counterfactual outcome and the observed data [22, 23]:

1. *Consistency*: $Y(D) = Y$.
2. *Ignorability*: $Y(d) \perp D | \mathbf{X}$.
3. *Positivity*: $p_{D|\mathbf{X}}(d|\mathbf{x}) > 0$ where $p_{D|\mathbf{X}}(d|\mathbf{x})$ is the conditional probability of receiving treatment d given $\mathbf{X} = \mathbf{x}$.

Assumption 1 means that the counterfactual outcome of a treatment corresponds to the actual outcome if assigned to that treatment. Assumption 2 means that within levels of $\mathbf{X}$, treatment D is randomized. Assumption 3 insures that there is enough overlap between treated and untreated groups for each possible value $\mathbf{x}$. In what follows, we assume that these identifiability assumptions hold.

3 Propensity score estimation under length-biased sampling

Rosenbaum and Rubin [24] adjust for differences between exposed and unexposed groups using a scalar function of the measured covariates, the propensity score, which removes the bias induced by differences between these two groups of units. The propensity score, $\pi(\mathbf{x})$, for binary exposure D is defined by $\pi(\mathbf{x}) = P(D = 1 | \mathbf{x})$, where $\mathbf{x}$ is a p -dimensional vector of covariates.

In general, the propensity score $\pi(\mathbf{x})$ is unknown and needs to be estimated; it has also been shown that even if the propensity score is known, one may gain efficiency in estimating the average treatment effect (ATE) by estimating $\pi(\cdot)$ using the data available [25]. However, estimating the propensity score using a length-biased sample does not lead to a balancing score or create the desired pseudo-population in which the exposure is independent of covariates; indeed, it may induce even more bias than leaving the confounders unadjusted [19].

Assuming a logit model for the propensity score in the target population, we have

$$\pi(\mathbf{x}, \alpha) = p(D^{\text{pop}} = 1 | \mathbf{X}^{\text{pop}} = \mathbf{x}) = \frac{\exp(\mathbf{x}\alpha)}{1 + \exp(\mathbf{x}\alpha)}, \quad (2)$$

where α is a $p \times 1$ vector of parameters. Cheng and Wang [26] develop a method that consistently estimates the parameters of the propensity score from prevalent survival data. Their method requires correct specification of the conditional hazard model given the treatment and covariates. Ertefaie et al. [19] show that under assumptions A1–A3 this requirement can be removed, and propose the following estimating equation

$$U(\alpha) = \sum_{i=1}^n \delta_i \mathbf{x}_i^T \frac{(d_i - \pi(\mathbf{x}_i, \alpha))}{\widehat{w}(y_i)} = 0, \quad (3)$$

where $\widehat{w}(y) = \int_0^y \widehat{S}_C(s) ds$ and $\widehat{S}_C$ is the Kaplan–Meier estimator of the survivor function of the residual censoring variable C . Note that the censored individuals contribute to this estimating equation through $\widehat{w}(y)$. Ertefaie et al. [19] show that

$$\frac{1}{n} U(\alpha) = \frac{1}{n} \sum_{i=1}^n \left[\delta_i \mathbf{x}_i^T \frac{(d_i - \pi(\mathbf{x}_i, \alpha))}{w(y_i)} + \mathcal{L}(y_i, c_i, a_i, d_i, \mathbf{x}_i) \right] + o_p(n^{-1/2}),$$

where $\mathcal{L}(y_i, c_i, a_i, d_i, \mathbf{x}_i)$ is the augmentation element [18]. In this manuscript, we use eq. (3) to estimate the parameters of the propensity score. The term $\mathcal{L}(y_i, c_i, a_i, d_i, \mathbf{x}_i)$ augments the failure time of the censored subjects using the observed failure times. We present the form of this augmentation term in Appendix B.

4 Cox PH models

The *hazard ratio* (HR) is defined as the ratio of hazards in the exposed and unexposed groups. Qin and Shen [5] introduce a set of estimating equations to assess the effect of covariates on the survival time in the presence of length-biased sampling. Our proposed estimating equation is an adaptation of the estimating equation introduced by Qin and Shen [5] (under the PH model) which adjusts for the confounding as well as length-biased sampling. We derive an estimating equation which estimates the marginal treatment effect without the need of estimating the effect for other covariates on the survival time.

Under A1–A3 and identifiability assumptions, the density of a counterfactual failure time observed in the study under exposure d can be expressed as

$$f(y(d), \delta = 1) = \int f(y, \delta = 1 | D = d, \mathbf{x}) f_{LB}(\mathbf{x}) d\mathbf{x} = \int \frac{f_d(y|\mathbf{x}) w(y) f(\mathbf{x})}{\mu_d} d\mathbf{x} = \frac{f_d(y) w(y)}{\mu_d},$$

where $\mu_d = \int t f_d(t) dt$ and $f_d(y)$ are the counterfactual densities of the survival time if all the individuals would have received the exposure d . The second equality follows as

$$p(Y \in (t, t + dt), \delta = 1 | d, \mathbf{x}) = \frac{f(t|d, \mathbf{x}) w(t) dt}{\mu_d(\mathbf{x})}, \quad f_{LB}(\mathbf{x}) = \frac{\mu_d(\mathbf{x}) f(\mathbf{x})}{\mu_d},$$

where $w(t) = \int_0^t S_C(s) ds$ and $\mu_d(\mathbf{x}) = \mathbb{E}[Y(d) | \mathbf{X} = \mathbf{x}] = \mathbb{E}[Y | D = d, \mathbf{X} = \mathbf{x}]$ [3–5].

Assuming the PH model for the counterfactual survival time, we have $\lambda_d(t) = \lambda_0(t) e^{\beta_0 d}$, and parameter e^{β_0} can be interpreted as a causal HR for the total effect of the treatment D . We propose the following estimating equation for β_0 ,

$$U(\beta) = \sum_{i=1}^n \int_0^s \frac{\mathbb{1}(D_i = d_i)}{p(D_i = d_i | \mathbf{X}_i)} \left[D_i - \frac{\sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)}{\sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{p(D_j = d_j | \mathbf{X}_j)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)} \right] dN_i(u), \quad (4)$$

In Appendix D, we show that eq. (4) corresponds to a score function of a pseudo-partial likelihood which can be presented as

$$\mathcal{S}_\beta = \int \frac{\mathbb{1}(D = d)}{p(D = d | \mathbf{X})} \left\{ D - \frac{\sum_j [d_j e^{\beta d_j} S_{d_j}(u) / \mu_{d_j}]}{\sum_j [e^{\beta d_j} S_{d_j}(u) / \mu_{d_j}]} \right\} dN(u)$$

where $S_d(y) = \int S(y|d, \mathbf{x})dF(\mathbf{x})$ is the counterfactual survival function of the survival time if all the individuals would have received the exposure d . The dependence of the estimating eq. (4) on the parametrization for the propensity score is shown by defining

$$U(\beta, \alpha) = \sum_{i=1}^n \int_0^s \frac{\mathbb{1}(D_i = d_i)}{p(D_i = d_i | \mathbf{X}_i, \alpha)} \left[D_i - \frac{\sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)}{\sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{p(D_j = d_j | \mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)} \right] dN_i(u), \quad (5)$$

In the proof of Theorem 1 in the Appendix, we show that $U(\beta, \alpha)$ can also be written as

$$U^M(\beta, \alpha) = \sum_{i=1}^n \int_0^s \left[D_i - \frac{\sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)}{\sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{p(D_j = d_j | \mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)} \right] dM_i(u), \quad (6)$$

where

$$dM_i(u) = \frac{\mathbb{1}(D_i = d_i)}{p(D_i = d_i | \mathbf{X}_i, \alpha)} [dN_i(u) - e^{\beta d_i} w(u) / w(Y_i) \delta_i I(Y_i > u) d\Lambda_0(u)].$$

The stochastic process $M(u)$ can be estimated by replacing the $w(\cdot)$ and $\Lambda_0(\cdot)$ by their estimates, $\hat{w}(\cdot)$ and $\hat{\Lambda}_0(\cdot)$, respectively. In the proof of Theorem 1 given in the Appendix B, we show that this stochastic process has mean zero.

The following theorem addresses the asymptotic properties of the estimator β obtained by the estimating eq. (6) when $w(y)$ and $p(D = d | \mathbf{X})$ are replaced by their estimated values $\hat{w}(y)$ and $\hat{p}(D = d | \mathbf{X}, \alpha)$, respectively. The parameters of the propensity score can be estimated using the estimating equation given in eq. (3). Define

$$M_C(s) = \mathbb{1}(Y - A < s, \delta = 0) - \int_0^s \mathbb{1}(\min(Y - A, C) > u) d\Lambda_C(u),$$

where $\Lambda_C(\cdot)$ is the cumulative hazard function of the censoring variable. The stochastic process $M_C(s)$ can be estimated by replacing the $\Lambda_C(\cdot)$ by its estimate, $\hat{\Lambda}_C(\cdot)$. The stochastic process $M_C(s)$ has mean zero,

$$\begin{aligned} \mathbb{E}[M_C(s)] &= \mathbb{E}[\mathbb{1}(C < Y - A < s)] - \int_0^s \mathbb{E}[\mathbb{1}(Y - A > u) \cdot \mathbb{1}(C > u)] d\Lambda_C(u) \\ &= \int_0^s S_C(u) \lambda_C(u) S_R(u) du - \int_0^s S_C(u) S_R(u) d\Lambda_C(u) = 0, \end{aligned}$$

where $S_R(u)$ is the survival function of the residual life time.

Theorem 1 Let $\hat{\beta}^{\text{Cox}}$ be the exposure effect estimator obtained as the root of

$$\hat{U}(\beta, \alpha) = \sum_{i=1}^n \int_0^s \frac{\mathbb{1}(D_i = d_i)}{\hat{p}(D_i = d_i | \mathbf{X}_i, \hat{\alpha})} \left[D_i - \frac{\sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \hat{\alpha})} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / \hat{w}(Y_j)}{\sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{\hat{p}(D_j = d_j | \mathbf{X}_j, \hat{\alpha})} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / \hat{w}(Y_j)} \right] dN_i(u). \quad (7)$$

Then under regularity conditions C.1 – C.4, C.5.a and C.5.c listed in Appendix A,

$$\sqrt{n}(\hat{\beta}^{\text{Cox}} - \beta) \xrightarrow{d} \mathcal{N}(0, \zeta(\beta, \alpha)),$$

where $\zeta(\beta, \alpha)$ is defined in Appendix B. Also, the estimating function $\hat{U}(\beta, \alpha)$ converges in probability to

$$U^* = \int_0^s \left[D - \frac{\mathbb{E}\{S^1(\beta, \alpha, u)\}}{\mathbb{E}\{S^0(\beta, \alpha, u)\}} \right] dM(u) + \int_0^s \frac{H_1(\beta, u)}{S_C(u)S_R(u)} dM_C(u), \quad (8)$$

where $H_1(\beta, u)$ and $S^d(\beta, \alpha, u)$ for $d = 0, 1$ are defined in Appendix B.

Proof See Appendix B.

In the absence of length-biased sampling, augmented partial likelihood estimators have been proposed in Robins et al. [27] and van der Laan and Robins [28]. The function U^* in eq. (8) generalizes this idea to length-biased sampling settings. Note the second part of the summation in U^* is the augmentation element.

Remark: Parameter β measures the marginal association between the exposure and the hazard, which is not necessarily equal to the conditional association due to non-collapsibility.

5 Accelerated failure time models

Inspired by the AFT models introduced by Cox and Oakes [29], we consider a general form of AFT models, where we do not assume a known error distribution. Assuming the AFT model for the counterfactual survival time, we have

$$\log(T^{\text{pop}}(d)) = \beta_0 d + \varepsilon,$$

and the parameter β_0 can be interpreted as a total treatment effect. Under causal identifiability assumptions and by the balancing property of the propensity score, the above model can be written in terms of the observed data as follows

$$\log(T^{\text{pop}}) = \beta_0 D + \gamma_0 \pi(\mathbf{X}) + \varepsilon. \quad (9)$$

We refer to this model as the AFT propensity score regression (AFTPSR) model [30]. Higher order and interaction terms can also be included in the model if needed. While AFT models may suffer from lack of robustness with respect to the log transformation, they are often more interpretable [31].

5.1 AFT-weighted estimating equations

Another approach for correcting the bias induced by non-random assignment was suggested by Horvitz and Thompson [32] and Hájek and Dupač [33] who introduced estimators which weight the observed outcomes. The IPTW estimator adjusts for confounding by assigning a weight to each individual proportional to their chance of receiving the exposure they actually received [34, 35].

We generalize the IPTW estimator to account for length-biased sampling. In our setting, the weights are the reciprocal of the probability of being in the exposure group to which each individual is observed to belong. The estimating equation corresponding to IPTW is given by

$$U_{\text{IPTW}}^{\text{AFT}} = \mathbb{P}_n \frac{\delta}{w(Y)} \left[\frac{D \log(Y)}{\pi} - \frac{(1-D) \log(Y)}{(1-\pi)} \right] = 0. \quad (10)$$

where $\mathbb{P}_n$ is the empirical average. This is a version of the complete case influence function introduced by Tsiatis [36] modified to take into account the censoring weight $w(Y)$.

Augmented IPTW (AIPTW), which is a more efficient version of IPTW, was introduced by Scharfstein et al. [37] and Lipsitz et al. [38]. Let $\mu_d(\mathbf{x}, \theta) = \mathbb{E}[\log(T)|D = d, \mathbf{X} = \mathbf{x}]$ for $d = 0, 1$. The corresponding estimating equation is given by

$$U_{\text{AIPTW}}^{\text{AFT}} = \mathbb{P}_n \frac{\delta}{w(Y)} [m(\mathbf{X}, D, Y) - \beta_0] + \int_0^s \frac{\kappa(t) dM_C(t)}{S_C(t) S_R(t)} = 0, \quad (11)$$

where

$$\hat{\kappa}(t) = \mathbb{P}_n \left[\frac{\delta \mathbb{I}(Y > t) [m(\mathbf{X}, D, Y)] \int_t^Y S_C(v) dv}{w^2(Y)} \right].$$

and

$$m(\mathbf{X}, D, Y) = \frac{D[\log(Y) - \mu_1(\mathbf{X}, \theta)]}{\pi(\mathbf{X}, \alpha)} - \frac{(1-D)[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1 - \pi(\mathbf{X}, \alpha))} + \mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta).$$

The causal effect estimator corresponding to an influence function in $U_{\text{AIPTW}}^{\text{AFT}}$ is called a *double robust* (DR) estimator in the sense that the estimator is consistent if either the propensity score model or the conditional response mean model is correctly specified [28, 36, 39, 40]. The influence function (11) is a member of the class of AIPTW influence functions and it has been shown that it is more efficient than eq. (10) [18]. In the proof of Theorem 2, we show how $U_{\text{AIPTW}}^{\text{AFT}}$ has been derived.

5.2 Asymptotic properties of the WEE estimator

Theorem 2 presents the asymptotic properties of the DR treatment effect estimator obtained by eq. (11) in the presence of length-biased sampling using the AFT models when both the treatment assignment and $w(\cdot)$ are replaced by their estimated values.

Theorem 2 Let $\hat{\beta}_{\text{DR}}^{\text{AFT}}$ be a DR estimator corresponding to $U_{\text{AIPTW}}^{\text{AFT}}$. Then under regularity conditions C.1 – C.4 and C.5.b, $\sqrt{n}(\hat{\beta}_{\text{DR}}^{\text{AFT}} - \beta) \xrightarrow{d} \mathcal{N}(0, \eta(\theta, \alpha))$, where $\eta(\theta, \alpha)$ is defined in the Appendix B.

Proof See Appendix B.

6 Simulation studies

We examine the performance of the proposed estimating equations for the Cox and the AFT models. In both cases, we simulate 1,000 datasets consisting of 200, 400 and 800 observations to study the performance of the proposed estimating equations for estimating the unmediated causal effect. Here, the censoring variable C is generated from a uniform distribution in the interval $(0, \tau)$ where the parameter τ is set such that it results in a desired censoring proportion. To create length-biased samples, we generate a variable A from a uniform distribution $(0, \rho)$ and ignore those whose generated unbiased failure time is less than A .

6.1 Cox model

We generated the population failure times from the hazard model $h(t|d, \mathbf{x}) = 0.2 \exp\{d - x_1 - x_2 + dx_1\}$, where $D \sim \text{Bernoulli}(\text{expit}\{-1 + 3x_1 - 1x_2\})$ with X_1 uniformly distributed on $(0, 1)$, $X_2 \sim \text{Bernoulli}(0.5)$, where $\text{expit}(t) = e^t / (1 + e^t)$. The true marginal treatment effect computed by Monte Carlo is $\theta_0 = 1.385$. We consider three different unadjusted scenarios: Unadjusted^{lc} is an estimator for which neither the length-biased nor the confounding is adjusted, Unadjusted^c is obtained by adjusting for the length-biased sampling but leaving the confounding unadjusted, and Unadjusted^l is carried out by adjusting for the confounding while the length-biased sampling is left unadjusted. The estimating equations for these unadjusted cases are listed in Appendix E.

Table 1: Cox proportional hazard model simulation study.

0% Cens. Method	<i>n</i> = 200			<i>n</i> = 400			<i>n</i> = 800		
	Bias	S.D.	MSE	Bias	S.D.	MSE	Bias	S.D.	MSE
CWE	0.032	0.237	0.057	0.030	0.175	0.031	0.022	0.122	0.015
U^{lc}	0.471	0.203	0.263	0.477	0.155	0.251	0.467	0.110	0.230
U^c	0.242	0.140	0.078	0.235	0.108	0.067	0.240	0.076	0.063
U^l	0.935	0.237	0.057	0.917	0.264	0.910	0.905	0.186	0.853
20% Cens.									
	<i>n</i> = 200			<i>n</i> = 400			<i>n</i> = 800		
CWE	0.035	0.257	0.067	0.006	0.189	0.036	0.006	0.133	0.018
U^{lc}	0.277	0.217	0.124	0.260	0.158	0.093	0.253	0.107	0.076
U^c	0.259	0.149	0.089	0.269	0.111	0.085	0.275	0.075	0.081
U^l	0.702	0.360	0.622	0.658	0.269	0.505	0.642	0.193	0.450
30% Cens.									
	<i>n</i> = 200			<i>n</i> = 400			<i>n</i> = 800		
CWE	0.023	0.254	0.065	0.039	0.193	0.039	0.048	0.139	0.022
U^{lc}	0.154	0.213	0.069	0.155	0.165	0.051	0.148	0.116	0.035
U^c	0.311	0.149	0.119	0.307	0.115	0.108	0.311	0.082	0.103
U^l	0.558	0.349	0.433	0.532	0.269	0.356	0.517	0.196	0.305

Note: CWE is the Cox-weighted estimator from eq. (7). U^{lc} is Unadjusted^{lc}: neither the length-biased nor the confounding are adjusted for. U^c is Unadjusted^c: the length biased is adjusted whereas confounding is left unadjusted. U^l is Unadjusted^l: The confounding is adjusted whereas the length biased is left unadjusted. The true marginal treatment effect is $\theta_0 = 1.385$.

Table 1 summarizes the marginal estimated treatment effects and their standard errors. Our simulation results confirm that the proposed estimating eq. (7) under Cox model assumption adjusts for both confounding and the length-biased sampling and results in smaller MSE across different sample sizes.

6.2 AFT model estimation

We consider a nonlinear failure time model and include the exposure effect modifier by adding the interaction term between the treatment and a confounder (x_2) as follows,

$$\log(T^{\text{pop}}) = 2.5d + \frac{x_2}{1 + 2x_2 + x_1} + \exp\{-x_1/2\} - 3dx_2 + \varepsilon,$$

where ε is uniformly distributed on $(-1,1)$, X_1 is uniformly distributed on $(0,1)$, $X_2 \sim \text{Bernoulli}(0.5)$, and $D \sim \text{Bernoulli}(\text{expit}\{2 - x_1 - 3x_2\})$. The estimated treatment effects and their standard errors are listed in Table 2. Similar to the previous section, we consider three different unadjusted scenarios. We have used a correct conditional mean model in the DR estimating equation. The DR estimator dominates the two other estimators in terms of the standard deviation and the MSE. Increasing the censoring proportion increases the bias in the PSR, IPTW and DR estimators while maintaining the unbiasedness. All the unadjusted estimators are biased and in our parameter setting it seems that the failure to account for the length-biased sampling leads to a more biased estimator compared to the Unadjusted^c. The estimating equations for the unadjusted cases are listed in Appendix E.

7 Real data analysis: the Canadian study of health and aging

The CSHA, initiated in 1989, is a nationwide study on aging in Canada. One of the objectives of CSHA was to study dementia. The CSHA included three phases in 1991, 1996 and 2001. In the first phase, 10,263 individuals aged 65 or over were sampled at random across Canada, from both rural and urban areas,

Table 2: Accelerated failure time simulation study.

0% Cens. Method	$n = 200$			$n = 400$			$n = 800$		
	Bias	S.D.	MSE	Bias	S.D.	MSE	Bias	S.D.	MSE
PSR	0.023	0.246	0.061	0.004	0.182	0.033	0.001	0.134	0.018
AWE	0.007	0.220	0.048	0.009	0.170	0.029	0.008	0.123	0.015
DR	0.014	0.203	0.042	0.003	0.150	0.022	0.001	0.109	0.012
U^{lc}	0.730	0.107	0.544	0.738	0.078	0.551	0.740	0.057	0.550
U^c	0.354	0.194	0.163	0.366	0.142	0.154	0.368	0.105	0.147
U^l	0.511	0.122	0.277	0.525	0.089	0.284	0.529	0.062	0.284
20% Cens.	$n = 200$			$n = 400$			$n = 800$		
PSR	0.018	0.226	0.051	0.011	0.164	0.027	0.011	0.113	0.013
AWE	0.002	0.245	0.060	0.003	0.176	0.031	0.005	0.125	0.016
DR	0.022	0.219	0.049	0.012	0.158	0.025	0.013	0.110	0.012
U^{lc}	0.713	0.127	0.525	0.722	0.089	0.529	0.724	0.062	0.529
U^c	0.344	0.214	0.164	0.355	0.150	0.148	0.348	0.107	0.132
U^l	0.490	0.141	0.260	0.503	0.098	0.262	0.512	0.067	0.269
30% Cens.	$n = 200$			$n = 400$			$n = 800$		
PSR	0.060	0.223	0.053	0.057	0.166	0.031	0.052	0.116	0.016
AWE	0.046	0.230	0.054	0.044	0.180	0.034	0.043	0.132	0.019
DR	0.069	0.216	0.051	0.063	0.161	0.030	0.058	0.113	0.016
U^{lc}	0.556	0.128	0.326	0.556	0.095	0.318	0.565	0.064	0.323
U^c	0.287	0.207	0.126	0.279	0.161	0.104	0.288	0.110	0.095
U^l	0.333	0.145	0.132	0.349	0.103	0.132	0.360	0.069	0.135

Note: PSR is the estimator based on eq. (9). AWE is the weighted estimator based on eq. (10). DR is the estimator based on eq. (11) described in Theorem 3. U^{lc} is Unadjusted^{lc}: neither the length-bias nor the confounding are adjusted for. U^c is Unadjusted^c: the length bias is adjusted whereas confounding is left unadjusted. U^l is Unadjusted^l: the confounding is adjusted whereas the length bias is left unadjusted.

from communities and institutions for the elderly. Among the participants, 1,132 people were diagnosed with dementia. The ages of dementia onset were assessed from each individual's medical history. We analyze the data collected during the first phase of the study which began in 1991 by sampling prevalent cases and examining the types of dementia: probable Alzheimer's disease, possible Alzheimer's disease and vascular dementia. The age of death or censoring were recorded for each subject from the time of screening, while the age at onset was ascertained retrospectively using CAMDEX from caregivers (Wolfson et al. [20]). Gender, level of education and the types of dementia are available as baseline covariates. The timescale for survival is set in years.

7.1 Exposure of interest: institutionalization

One of the collected covariates is the dichotomous institutionalization (exposure) indicator, which takes the value one if the subject is institutionalized at the time of sampling, and zero otherwise. We are interested in comparing survival of institutionalized subjects with dementia and subjects recruited from the community.

Since there are some covariates which confound the effect of the exposure on the survival time, the crude difference estimator will be biased. We estimate the effect of this covariate while having confounding and length-biased sampling as two sources of estimation bias using Cox PH, and semiparametric AFT models. Our data include 818 subjects (after excluding patients with missing information), of which 180 subjects were right censored [20]. The validity of the stationarity assumption has been shown to be reasonable by Addona and Wolfson [41] and Asgharian et al. [42].

In order to estimate the causal effect of institutionalization, we need to ignore those individuals that their date of institutionalization is *after* their onset of the disease. However this information was not recorded in the dataset. We address this limitation using a multiple imputation approach to generate synthetic data on which the estimating equations can be used. Using an informed model, we generate a binary variable, Z , conditional on the age at onset, X_1 , and the gender, X_2 , that attempts to reveal whether institutionalization occurred prior to onset. Specifically, we used the model $\text{logit}(Z = \mathbb{I}|x_1, x_2) = -2 + 2\mathbb{I}(x_1 > 85) + 0.65\mathbb{I}(74 < x_1 < 84) + 0.15x_2$ to generate Z , and then ignored patients with $Z = 0$, i.e. those patients that whose date of institutionalization is after their onset time. We parametrized the above model such that older patients and females, $x_2 = 1$, have more chance to be institutionalized before the onset of dementia. The value of the parameters are extracted from Carrière and Pelletier [43]. These authors estimate the relationship between sociodemographic characteristics and institutionalization of citizens of Canada. One of the limitations of our logistic model for Z is that we do not have all the covariates that are used in Carrière and Pelletier [43] such as income and marital status. We use the above model to fill the missing variable repeatedly and create a collection of 20 imputed data sets [44].

7.2 Semiparametric AFT models

We have estimated the institutionalization effect on survival time using the semiparametric estimating equation proposed in Section 5. Table 3 presents the estimated institution effect using the semiparametric estimating equations proposed in Section 5 under the AFT model. PSR is the estimator based on eq. (9), AWE is the weighted estimator based on eq. (10) and DR is the estimator based on eq. (11) described in Theorem 3. We consider three different unadjusted scenarios: Unadjusted^{lc}, Unadjusted^c and Unadjusted^l. The results reveal that the institutionalization have a significant positive effect on the survival time when estimated using AWE and PSR while it has a positive effect at the 10% level using DR estimator. The *unadjusted* estimator shows a small *negative* effect. In other words, without adjusting for either the length-biased sampling or the treatment adjustment, we might incorrectly conclude that institutionalized subjects tend to have a shorter survival time.

Table 3: CSHA data analysis (semiparametric AFT): estimation of the institutionalization effect on the survival time.

Method	Effect	S.E.	CI 95%	AFT
				CI 90%
PSR	0.28	0.12	(0.04,0.52)	(0.08,0.47)
DR	0.24	0.13	(−0.02,0.50)	(0.02,0.45)
AWE	0.27	0.13	(0.01,0.53)	(0.05,0.48)
U^{lc}	−0.14	0.10	(−0.32,0.05)	(−0.30,0.02)
U^c	0.15	0.13	(−0.10,0.40)	(−0.05,0.36)
U^l	−0.02	0.11	(−0.23,0.19)	(−0.19,0.15)

7.3 Cox PH model

Although the residual analysis shows that AFT is a suitable model for this data set (Bergeron et al. [3], we have also estimated the marginal institutionalization effect using the weighted estimating equation proposed for Cox models (Table 4). The proposed estimating equation for Cox models can be fitted using standard software, and equivalent to the following command in R for the observed subset of data, $\delta_i = 1$ for $i = 1, \dots, m$,

Table 4: CSHA data analysis (Cox model): estimation of the institutionalization effect on the survival time.

Method	Effect	S.E.	CI 95%	Cox
				CI 90%
CWE	0.07	0.12	(−0.17,0.31)	(−0.12,0.26)
U^{lc}	−0.07	0.11	(−0.29,0.15)	(−0.25,0.11)
U^c	0.19	0.10	(−0.01,0.39)	(0.03,0.35)
U^l	0.03	0.17	(−0.31,0.37)	(−0.25,0.31)

```
coxph(Surv(y,del)~d+offset(-log(hatwy)),weights=wpi,subset=(del==1))
```

where y is the observed survival time, wpi is $\hat{w}_c = d/\hat{\pi} + (1-d)/(1-\hat{\pi})$ and $hatwy$ is $\hat{w}(y)$ is the Kaplan–Meier estimate for the distribution of the censoring variable. In this parameterization, the coefficients estimated indicate the increase/decrease in the hazard while in the AFT model coefficients indicate an decrease/increase in the survival time, and hence the opposite sign of the coefficients in the AFT and PH model have the same interpretation. To determine whether a fitted Cox model adequately describe the data, we looked at the scaled Schoenfeld residuals plot, Figure 2, for the Cox model. There appears to be a trend in the scaled Schoenfeld residuals for the institution indicator variable which indicates violation of the assumption of PH.

Figure 2: CSHA data: the scaled Schoenfeld residuals obtained by Cox model against transformed time for the estimated institutionalization effect.

7.4 Survival curves

We compute adjusted and unadjusted survival curves to compare survival with dementia in the course of time between the exposure groups (Figure 3). Several methods have been proposed to adjust for the length-biased sampling such as the nonparametric maximum likelihood estimator [45–47], the truncation product-limit estimator [48] and the maximum pseudo-partial likelihood estimator [49]. Here, we use the method introduced by Huang and Qin [50] which incorporates the information from the marginal distribution of the

Figure 3: CSHA data. Survival curves for adjusted and three different unadjusted scenarios. Unadjusted^{lc}: neither the length biased nor the confounding are adjusted for. Unadjusted^c: The length biased is adjusted whereas confounding left unadjusted. Unadjusted^l: The confounding is adjusted whereas the length biased left unadjusted.

truncation time from disease onset to recruitment time. The bias induced by confounding can be adjusted by creating a pseudo-population using the inverse probability of being in the group that the individuals actually belong to [51–53]. The adjusted survival curves show that the institutionalized patients tend to live longer while the survival curves cross when unadjusted (Figure 3). Moreover, leaving the length-biased sampling unadjusted may lead to overestimate the survival times which is shown in Figure 4. This figure clearly depicts that the survival curve of the institutionalized individuals is always higher than those recruited from the community.

Figure 4: CSHA data. Adjusted (thick lines) and unadjusted (thin lines) survival curves for treated and untreated individuals.

8 Concluding remarks

We have presented two different approaches to estimate the exposure effect from right-censored length-biased samples. The estimating equations adjust for two different types of bias at the same time. Our simulation and real data analysis results highlight the importance of adjusting for the two sources of bias; failure to adjust for either the length-biased sampling or the confounding may lead to misleading results.

We have focused on the stationary case. It would, however, be of interest to extend the method to the general left truncation where the left truncation distribution is unknown. This latter approach is robust against departure from stationarity, though it is less efficient when the stationarity assumption holds [45, 46, 54].

Appendix

Here, we present the assumptions and proofs of the main and other auxiliary results.

Appendix A

The regularity conditions required for the Cox and the weighted AFT models:

- C.1 $\mu_d(\cdot)$ for $d = 0, 1$ is a twice continuously differentiable function.
- C.2 $\pi(\cdot)$ is bounded away from zero and one ($\gamma < \pi(\cdot) < 1 - \gamma$ where $\gamma > 0$).
- C.3 $\sup[t : p(R > t) > 0] \leq \sup[t : p(C > t) > 0] = s$ and $p(\delta = 1) > 0$.
- C.4 $\int_0^s [(\int_t^s S_C(v) dv)^2 / (S_C^2(t) S_R(t))] dS_C(t) < \infty$.
- C.5
 - (a) $\int_0^s \tau_1^2(t) / (S_C^2(t) S_R(t)) dS_C(t) < \infty$ and $\int_0^s \tau^2(t) / (S_C^2(t) S_R(t)) dS_C(t) < \infty$
 - (b) $\int_0^s \kappa_d^2(t) / (S_C^2(t) S_R(t)) dS_C(t) < \infty$ and $\int_0^s \kappa_2^2(t) / (S_C^2(t) S_R(t)) dS_C(t) < \infty$
 - (c) $\int_0^s v^2(t) / (S_C^2(t) S_R(t)) dS_C(t) < \infty$

where

$$\begin{aligned}\tau_1(t) &= \mathbb{E} \left[\frac{D \delta e^{\beta D} \mathbb{1}(Y > t) \int_t^Y S_C(v) dv}{p(D = 1 | \mathbf{X}) w^2(Y)} \right], \\ \tau(t) &= \mathbb{E} \left[\frac{\delta e^{\beta D} \mathbb{1}(Y > t) \int_t^Y S_C(v) dv}{p(D = d | \mathbf{X}) w^2(Y)} \right], \\ \kappa_d(t) &= \mathbb{E} \left[\frac{\mathbb{1}(D = d) \delta \mathbb{1}(Y > t) [\log(Y) - \mu_d(\pi, \theta)] \int_t^Y S_C(v) dv}{p(D = d | \mathbf{X}) w^2(Y)} \right], \\ \kappa_2(t) &= \mathbb{E} \left[\frac{\delta \mathbb{1}(Y > t) [\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] \int_t^Y S_C(v) dv}{p(D = d | \mathbf{X}) w^2(Y)} \right], \\ v(u) &= \mathbb{E} \left[\frac{\delta \mathbf{X}(D - \pi(\mathbf{X}, \alpha)) \mathbb{1}(Y > u) \int_u^Y S_C(v) dv}{w^2(Y)} \right].\end{aligned}$$

Condition C.1 is a smoothness assumption of the mean function. C.1–C.2 are termed *positivity* assumptions, meaning that there is a positive chance that a subject falls in either the treatment or the control groups and

being not censored, respectively. C.3 is an identifiability condition [54] and C.4–C.5 are required to obtain an estimator with a finite variance.

Appendix B: Proofs of Theorems 1 and 2

Proof of Theorem 1

First, we show that the stochastic process $M(s)$ has a mean of zero:

$$M_i(s) = \frac{\mathbb{1}(D_i = d)}{p(D_i = d|\mathbf{X}_i, \alpha)} \left[N_i(s) - \int_0^s e^{\beta D_i} w(y)/w(Y_i) \delta_i I(Y_i > y) d\Lambda_0(y) \right].$$

Using eqs (8), (10) and the no unmeasured confounder assumption

$$\mathbb{E}[N_i(s)] = \mathbb{E} \left[\frac{\mathbb{1}(D_i = d)}{p(D_i = d|\mathbf{X}_i, \alpha)} \mathbb{1}(Y_i < s) \delta \right] = \mathbb{E}[\mathbb{1}(Y_i(d) < s) \delta] = \int_0^s \frac{f_d(y)w(y)}{\mu_d} dy$$

and

$$\begin{aligned} & \mathbb{E} \left[\int_0^s \frac{\mathbb{1}(D_i = d)}{p(D_i = d|\mathbf{X}_i, \alpha)} e^{\beta D_i} w(y)/w(Y_i) \delta_i I(Y_i > y) d\Lambda_0(y) \right] \\ &= \mathbb{E} \left[\int_0^s e^{\beta d} w(y)/w(Y_i) \delta_i I(Y_i(d) > y) d\Lambda_0(y) \right] = \int_0^s e^{\beta d} w(y) \frac{S_d(y)}{\mu_d} d\Lambda_0(y) \end{aligned}$$

Thus, $\mathbb{E}[M_i(s)] = 0$.

In Section 5, we have shown that the estimating equation $U^M(\beta, \alpha)$ given by eq. (13) is unbiased. We need to show that the two representations (12) and (13) are equal. By the definition of the stochastic process $M_i(s)$, we have

$$\begin{aligned} U^M(\beta, \alpha) &= \sum_{i=1}^n \int_0^s \left[D_i - \frac{S^1(\beta, \alpha, t)}{S^0(\beta, \alpha, t)} \right] dM_i(u) \\ &= U(\beta, \alpha) - \int_0^s \sum_{i=1}^n \left[\frac{D_i}{\pi(\mathbf{X}_i, \alpha)} w(u) e^{\beta D_i} \delta_i \mathbb{1}(Y_i \geq u)/w(Y_i) \right] - n \frac{S^1(\beta, \alpha, t)}{S^0(\beta, \alpha, t)} S^0(\beta, \alpha, t) d\Lambda_0(u) \\ &= U(\beta, \alpha) \end{aligned}$$

where

$$\begin{aligned} S^1(\beta, \alpha, u) &= n^{-1} \sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \alpha)} w(u) e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u)/w(Y_j), \\ S^0(\beta, \alpha, u) &= n^{-1} \sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{p(D_j = d_j|\mathbf{X}_j, \alpha)} w(u) e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u)/w(Y_j). \end{aligned}$$

We have the following estimating equation when $w(y)$ is replaced by its estimate, $\hat{w}(y)$,

$$\hat{U}^M(\beta, \alpha) = \sum_{i=1}^n \int_0^s \left[D_i - \frac{\hat{S}^1(\beta, \alpha, u)}{\hat{S}^0(\beta, \alpha, u)} \right] dN_i(u).$$

where

$$\begin{aligned}\widehat{S}^1(\beta, \alpha, u) &= n^{-1} \sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \alpha)} \widehat{w}(u) e^{\beta D_j} \delta_j \mathbb{I}(Y_j \geq u) / \widehat{w}(Y_j), \\ \widehat{S}^0(\beta, \alpha, u) &= n^{-1} \sum_{j=1}^n \frac{\mathbb{I}(D_j = d_j)}{p(D_j = d_j | \mathbf{X}_j, \alpha)} \widehat{w}(u) e^{\beta D_j} \delta_j \mathbb{I}(Y_j \geq u) / \widehat{w}(Y_j).\end{aligned}$$

The estimating equation $\widehat{U}^M(\beta, \alpha)$ can be written as

$$\sum_{i=1}^n \int_0^s \left[D_i - \frac{\mathbb{E}\{S^1(\beta, \alpha, u)\}}{\mathbb{E}\{S^0(\beta, \alpha, u)\}} \right] dM_i(u) + \sum_{i=1}^n \int_0^s \left[\frac{S^1(\beta, \alpha, u)}{S^0(\beta, \alpha, u)} - \frac{\widehat{S}^1(\beta, \alpha, u)}{\widehat{S}^0(\beta, \alpha, u)} \right] dN_i(u) + o_p(1).$$

Using the strong consistency of $\widehat{w}(y)$ to $w(y)$ [55], we have

$$\frac{1}{\widehat{w}(Y_j)} = \frac{1}{w(Y_j)} \left[1 + \frac{w(Y_j) - \widehat{w}(Y_j)}{w(Y_j)} \right] + o_p(1), \quad (12)$$

and following the martingale integral representation $\sqrt{n}(\widehat{w}(Y) - w(Y))$ introduced by Shen et al. [4] and Qin and Shen [5], we can show that

$$\int_0^s \left[\frac{S^1(\beta, \alpha, u)}{S^0(\beta, \alpha, u)} - \frac{\widehat{S}^1(\beta, \alpha, u)}{\widehat{S}^0(\beta, \alpha, u)} \right] dN_i(u) = \int_0^s \frac{H_1(\beta, u)}{S_C(u)S_R(u)} dM_{Ci}(u) + o_p(1),$$

where $H_1(\beta, u) = \mathbb{E}_{Y^*} \left[\frac{\mathbb{E}[\tau(Y, u) \mathbb{I}(Y > Y^*)]}{\mathbb{E}[S^0(\beta, \alpha, Y^*)]} | Y^* = y^* \right]$ with

$$\tau(Y, u) = \frac{D \delta w(Y^*) e^{\beta D} \mathbb{I}(Y > u) \int_u^Y S_C(v) dv}{\pi(\mathbf{X}, \alpha) w^2(Y)},$$

where Y^* is independent of and identically distributed to Y . Now we can derive the asymptotic variance of our proposed estimator when α is replaced by $\hat{\alpha}$ in the propensity score model. Note

$$\begin{aligned}\tilde{\psi}(\alpha) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\psi}_i(\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \mathbf{X}_i \frac{D_i - \pi(\mathbf{X}_i, \alpha)}{\widehat{w}(Y_i)} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \mathbf{X}_i \frac{D_i - \pi(\mathbf{X}_i, \alpha)}{w(Y_i)} \left[1 + \frac{w(Y_i) - \widehat{w}(Y_i)}{w(Y_i)} \right] + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\delta_i \mathbf{X}_i \frac{D_i - \pi(\mathbf{X}_i, \alpha)}{w(Y_i)} + \int_0^s \frac{\widehat{v}(t) dM_C(t)}{S_C(t)S_R(t)} \right],\end{aligned} \quad (13)$$

where

$$\widehat{v}(t) = \frac{1}{n} \sum_{i=1}^n \left[\frac{\delta_i I(Y_i > t) \mathbf{X}_i [D_i - \pi(\mathbf{X}_i, \alpha)] \int_t^{Y_i} S_C(v) dv}{w^2(Y_i)} \right].$$

Hence, using the Taylor expansion and Theorem 1 in Pugh et al. [14],

$$\zeta(\beta, \alpha) = \left[\mathbb{E} \left\{ \frac{\partial U^*}{\partial \beta} \right\} \right]^{-2} [\mathbb{E}\{U^{*2}\} - \mathbb{E}\{U^* \tilde{\psi}_i(\alpha)'\} \mathbb{E}\{\tilde{\psi}_i(\alpha) \tilde{\psi}_i(\alpha)'\}^{-1} \mathbb{E}\{\tilde{\psi}_i(\alpha) U^{*'}\}],$$

where

$$U^* = \int_0^s \left[D - \frac{\mathbb{E}\{S^1(\beta, \alpha, u)\}}{\mathbb{E}\{S^0(\beta, \alpha, u)\}} \right] dM(u) + \int_0^s \frac{H_1(\beta, u)}{S_C(u)S_R(u)} dM_C(u).$$

Note: The asymptotic variance $\zeta(\beta, \alpha)$ may be estimated consistently by replacing the expectations in the expressions for U^* , τ_1 and τ with expectations with respect to the empirical measure.

Proof of Theorem 2

Following eq. (12) and using the martingale integral representation $\sqrt{n}(\widehat{w}(Y) - w(Y))$, we have

$$\frac{\delta}{\widehat{w}(Y)}[m(\mathbf{X}, D, Y) - \beta_0] = \frac{\delta}{w(Y)}[m(\mathbf{X}, D, Y) - \beta_0] + \int_0^s \frac{\kappa(t) dM_C(t)}{S_C(t)S_R(t)} + o_p(1)$$

where

$$\kappa(t) = \mathbb{E} \left[\frac{\delta \mathbb{I}(Y > t)[m(\mathbf{X}, D, Y)] \int_t^Y S_C(v) dv}{w^2(Y)} \right]$$

and

$$m(\mathbf{X}, D, Y) = \frac{D[\log(Y) - \mu_1(\mathbf{X}, \theta)]}{\pi(\mathbf{X}, \alpha)} - \frac{(1-D)[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))} + \mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta).$$

Hence, the generic elements of the class of influence functions $\mathcal{G}^{(\text{AFT})}$

$$\mathcal{G}^{(\text{AFT})} = \left\{ \varphi(Y, D, \mathbf{X}) : \frac{\delta}{w(Y)}[m(\mathbf{X}, D, Y) - \beta_0] + \int_0^s \frac{\kappa(t) dM_C(t)}{S_C(t)S_R(t)} \right\},$$

can be written as $\{\widehat{V}_0(\theta) - \widehat{V}_1(\theta) + \widehat{V}_2(\theta)\}$ where

$$V_0(\theta, \alpha) = \frac{(1-D)\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))w(Y)} + \int_0^s \frac{\kappa_0(t) dM_C(t)}{S_C(t)S_R(t)}$$

$$V_1(\theta, \alpha) = \frac{D\delta[\log(Y) - \mu_1(\mathbf{X}, \theta)]}{\pi(\mathbf{X}, \alpha)w(Y)} + \int_0^s \frac{\kappa_1(t) dM_C(t)}{S_C(t)S_R(t)}$$

$$V_2(\theta, \alpha) = \frac{\delta}{w(Y)}[\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] + \int_0^s \frac{\kappa_2(t) dM_C(t)}{S_C(t)S_R(t)},$$

with

$$\kappa_d(t) = \mathbb{E} \left[\frac{I(D=d)\delta I(Y>t)[\log(Y) - \mu_d(\mathbf{X}, \theta)] \int_t^Y S_C(v) dv}{p(D=d|\mathbf{X}, \alpha)w^2(Y)} \right], \quad \text{for } d = 0, 1$$

$$\kappa_2(t) = \mathbb{E} \left[\frac{\delta I(Y>t)[\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] \int_t^Y S_C(v) dv}{w^2(Y)} \right].$$

In order to show that $\mathcal{G}^{(\text{AFT})}$ results in an unbiased estimator, we need to show that $\mathbb{E}[V_0(\theta, \alpha)] = \mathbb{E}[V_1(\theta, \alpha)] = \mathbb{E}[V_2(\theta, \alpha)] = 0$. For the first expectation, we have

$$\begin{aligned} \mathbb{E}[V_0(\theta, \alpha)] &= \mathbb{E} \left[\frac{(1-D)\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))w(Y)} + \int_0^s \frac{\kappa_0(t) dM_C(t)}{S_C(t)S_R(t)} \right] \\ &= \mathbb{E} \left[\frac{(1-D)\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))w(Y)} \right] + \mathbb{E} \left[\int_0^s \frac{\kappa_0(t) dM_C(t)}{S_C(t)S_R(t)} \right] \end{aligned} \quad (14)$$

The first expectation on the RHS of eq. (14) is

$$\begin{aligned} &= \mathbb{E} \left[\int_0^\infty \int_0^\infty f(Y=y, A=a, \delta=1, D=1 | \mathbf{X}=\mathbf{x}) \times \frac{D[\log(y) - \mu_1(\mathbf{x}, \theta)]}{\pi(\mathbf{x}, \alpha)w(Y)} da dy \right] \\ &= \mathbb{E} \left[\int_0^\infty \frac{1}{\pi(\mathbf{x}, \alpha)} \frac{f(Y=y | D=1, \mathbf{X}=\mathbf{x})}{\mu_1(\mathbf{x}, \theta)} w(y) \frac{\mu_1(\mathbf{x}, \theta)}{\mu(\mathbf{x}, \theta)} \pi(\mathbf{x}, \alpha) \times \frac{[\log(y) - \mu_1(\mathbf{x}, \theta)]}{w(y)} dy \right] \\ &= \mathbb{E} \left[\frac{1}{\mu(\mathbf{x}, \theta)} \int_0^\infty f(y | \mathbf{x}, D=1) [\log(Y) - \mu_1(\mathbf{x}, \theta)] dy \right] = 0, \end{aligned}$$

where $\mu_d(\mathbf{x}, \theta) = \int y f(y | D=d, \mathbf{x}) dy$ and $\mu(\mathbf{x}, \theta) = \int_0^\infty p(T^{\text{pop}} \geq a | \mathbf{x}, \theta) da$. The second equality follows from

$$p(Y \in (t, t+dt), \delta=1 | d, \mathbf{x}) = \frac{f(t | d, \mathbf{x}) w(t) dt}{\mu_d(\mathbf{x}, \theta)},$$

and

$$p_{LB}(D=1 | \mathbf{X}=\mathbf{x}) = \frac{\mu_1(\mathbf{x}, \theta) p(D^{\text{pop}}=1 | \mathbf{X}^{\text{pop}}=\mathbf{x})}{\mu(\mathbf{x}, \theta)},$$

where $p_{LB}(D=1 | \mathbf{X}=\mathbf{x})$ is the propensity score estimated from the length-biased sample and $\pi(\mathbf{x}, \alpha) = p(D^{\text{pop}}=1 | \mathbf{X}^{\text{pop}}=\mathbf{x})$ is the true propensity score. The second expectation of the RHS of eq. (14) is also equal to zero since $\mathbb{E}[M_C(s)] = 0$. Similarly, we can show that $E[V_1(\theta, \alpha)] = E[V_2(\theta, \alpha)] = 0$.

It can be shown that $V_0(\theta, \alpha)$ and $V_1(\theta, \alpha)$ are uncorrelated. Hence the asymptotic variance of the estimator is given by

$$\eta(\theta, \alpha) = \mathbb{E}[V_1^{*2}(\theta, \alpha) + V_0^{*2}(\theta, \alpha) + V_2^{*2}(\theta, \alpha) - V_0^{*2}(\theta, \alpha)V_2^{*2}(\theta, \alpha) + V_1^{*2}(\theta, \alpha)V_2^{*2}(\theta, \alpha)].$$

where

$$\begin{aligned} \mathbb{E}[V_1^2(\theta, \alpha)] &= \mathbb{E} \left[\left\{ \frac{D\delta[\log(Y) - \mu_1(\mathbf{X}, \theta)]}{\pi(\mathbf{X}, \alpha)w(Y)} + \int_0^s \frac{\kappa_1(t) dM_C(t)}{S_C(t)S_R(t)} \right\}^2 \right] \\ \mathbb{E}[V_0^2(\theta, \alpha)] &= \mathbb{E} \left[\left\{ \frac{(1-D)\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))w(Y)} + \int_0^s \frac{\kappa_0(t) dM_C(t)}{S_C(t)S_R(t)} \right\}^2 \right] \\ \mathbb{E}[V_2^2(\theta, \alpha)] &= \mathbb{E} \left[\left\{ \frac{\delta}{w(Y)} [\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] + \int_0^s \frac{\kappa_2(t) dM_C(t)}{S_C(t)S_R(t)} \right\}^2 \right], \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[V_1(\theta, \alpha)V_2(\theta, \alpha)] &= \mathbb{E} \left[\left\{ \frac{D\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(\pi(\mathbf{X}, \alpha))w(Y)} + \int_0^s \frac{\kappa_1(t) dM_C(t)}{S_C(t)S_R(t)} \right\} \right. \\ &\quad \times \left. \left\{ \frac{\delta}{w(Y)} [\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] + \int_0^s \frac{\kappa_2(t) dM_C(t)}{S_C(t)S_R(t)} \right\} \right] \\ \mathbb{E}[V_0(\theta, \alpha)V_2(\theta, \alpha)] &= \mathbb{E} \left[\left\{ \frac{(1-D)\delta[\log(Y) - \mu_0(\mathbf{X}, \theta)]}{(1-\pi(\mathbf{X}, \alpha))w(Y)} + \int_0^s \frac{\kappa_0(t) dM_C(t)}{S_C(t)S_R(t)} \right\} \right. \\ &\quad \times \left. \left\{ \frac{\delta}{w(Y)} [\mu_1(\mathbf{X}, \theta) - \mu_0(\mathbf{X}, \theta) - \beta] + \int_0^s \frac{\kappa_2(t) dM_C(t)}{S_C(t)S_R(t)} \right\} \right]. \end{aligned}$$

Appendix C: Misspecified propensity score or mean model

In this appendix, we study the performance of the DR AFT estimator (11) when either the propensity score or the mean model is misspecified. We use the same simulation model as in Section 6.2 with only changing the treatment assignment model to $D \sim \text{Bernoulli}(\text{expit}\{2 - 3x_1 - 3x_2\})$. Our misspecified propensity score ignores the confounder X_1 . Table 5 shows results based on 500 data sets of sizes 200 and 800 with 0, 20 and 30 percent censoring. The superscript $m1$ and $m2$ represent the propensity score and mean model misspecifications, respectively. The misspecified propensity score ignores the variable x_1 and the misspecified mean model ignores the interaction term dx_2 (see Section 6.2). The results confirm that our estimator is doubly robust.

Table 5: Accelerated failure time simulation study when the propensity score is misspecified. PSR is the estimator based on eq. (9).

0% Cens.	$n = 200$			$n = 800$		
	Bias	S.D.	MSE	Bias	S.D.	MSE
PSR ^{m1}	0.421	0.238	0.234	0.423	0.127	0.195
AWE ^{m1}	0.296	0.223	0.137	0.283	0.120	0.095
DR ^{m1}	0.010	0.200	0.040	0.001	0.104	0.011
DR ^{m2}	0.076	0.277	0.083	0.055	0.208	0.046
20% Cens.			$n = 200$	$n = 800$		
PSR ^{m1}	0.286	0.210	0.126	0.274	0.103	0.086
AWE ^{m1}	0.291	0.225	0.135	0.266	0.117	0.084
DR ^{m1}	0.002	0.192	0.037	0.013	0.099	0.010
DR ^{m2}	0.057	0.252	0.067	0.020	0.134	0.019
30% Cens.			$n = 200$	$n = 800$		
PSR ^{m1}	0.242	0.203	0.100	0.222	0.111	0.061
AWE ^{m1}	0.242	0.216	0.105	0.215	0.120	0.061
DR ^{m1}	0.045	0.199	0.044	0.063	0.102	0.014
DR	0.017	0.265	0.070	0.036	0.154	0.025

Note: AWE is the weighted estimator based on eq. (10). DR is the estimator based on eq. (11) described in Theorem 3.

Appendix D: Derivation of the score function $\mathcal{S}_\beta$

The score function $\mathcal{S}_\beta$ derived from the following pseudo-partial likelihood after adjusting the risk sets for the confounding and the length-biased sampling

$$\mathcal{L}_P(\beta) = \prod_i \left\{ \frac{\exp(\beta d_i)}{\sum_{j=1}^n \exp(\beta d_j) \mathbb{I}(y_j \geq y_i \geq a_j) \mathbb{I}(D = d) / p(D_i = d_i | \mathbf{x}_i)} \right\}^{\delta_i / p(D_i = d_i | \mathbf{x}_i)};$$

where $\mathbb{I}(y_j \geq y_i \geq a_j) / p(D_j = d_j | \mathbf{x}_j)$ represents the adjusted risk set for both length-biased sampling and confounding. Followed by Shen et al. [4] and Qin and Shen [5], we estimate the denominator by

$$\sum_{j=1}^n \exp(\beta d_j) \frac{\delta_j \mathbb{I}(D = d) \mathbb{I}(y_j \geq y_i)}{p(D_i = d_i | \mathbf{x}_i) w(y_j)}$$

where the focus is on the uncensored subjects and the risk set is inversely weighted by $w(y_j)$. Note, under assumptions A1–A3, we have

$$\mathbb{E} \left[\frac{\mathbb{1}(D = d)}{p(D = d|\mathbf{X})} \cdot \frac{\delta \mathbb{1}(Y \geq y)}{w(Y)} \right] = \frac{S_d(y)}{\mu_d},$$

which justifies the form of the score function $\mathcal{S}_\beta$.

Appendix E: Cox and AFT estimating equations when either of the confounding or the length-biased sampling is ignored

1. Estimating equation for Cox model when length biased is left unadjusted:

$$\sum_{i=1}^n \int_0^s \frac{\mathbb{1}(D_i = d_i)}{p(D_i = d_i|\mathbf{X}_i, \alpha)} \left[D_i - \frac{\sum_{j=1}^n \frac{D_j}{\pi(\mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u)}{\sum_{j=1}^n \frac{\mathbb{1}(D_j = d_j)}{p(D_j = d_j|\mathbf{X}_j, \alpha)} e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u)} \right] dN_i(u) = 0.$$

2. Estimating equation for Cox model when the confounding is left unadjusted:

$$\sum_{i=1}^n \int_0^s \left[D_i - \frac{\sum_{j=1}^n D_j e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)}{\sum_{j=1}^n e^{\beta D_j} \delta_j \mathbb{1}(Y_j \geq u) / w(Y_j)} \right] dN_i(u) = 0.$$

3. Estimating equation for AFT model when length biased is left unadjusted:

$$\sum_{i=1}^n \delta_i \left[\frac{D_i \log(Y_i)}{\pi_i} - \frac{(1 - D_i) \log(Y_i)}{(1 - \pi_i)} - \beta \right] = 0.$$

4. Estimating equation for AFT model when the confounding is left unadjusted:

$$\sum_{i=1}^n \frac{\delta_i D_i}{w(Y_i)} [\log(Y_i) - \beta D_i] = 0.$$

Funding: The data reported in this article were collected as part of the Canadian Study of Health and Aging. The core study was funded by the Seniors' Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada Project 6606-3954-MC(S). Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP Project 6603-1417-302(R), Bayer Incorporated, and the British Columbia Health Research Foundation Projects 38 (93-2) and 34 (96-1). The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada. The authors would like to thank Professor Christina Wolfson for providing the data. This work was supported in part by NIDA and NSF SES-1260782 grants P50 DA010075. The second and third authors acknowledge the support of Discovery Grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Part of the work on this article was completed while the second author was on sabbatical leave at Universite de Bordeaux. He would like to thank hospitality of the Equipe de Biostatistique and in particular, Daniel Commenges and Helene Jacqmin-Gadda.

References

1. Cox DR, Lewis P. The statistical analysis of series of events. Monographs on applied probability and statistics. London: Chapman and Hall, 1966.
2. Zelen M, Feinlein M. On the theory of screening for chronic diseases. *Biometrika* 1969;56:601–14.

3. Bergeron PJ, Asgharian M, Wolfson DB. Covariate bias induced by length-biased sampling of failure times. *J Am Stat Assoc* 2008;103:737–42.
4. Shen Y, Ning J, Qin J. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *J Am Stat Assoc* 2009;104:1192–202.
5. Qin J, Shen Y. Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics* 2010;66:382–92.
6. Ning J, Qin J, Shen Y. Non-parametric tests for right-censored data with biased sampling. *J R Stat Soc Ser B (Stat Methodol)* 2010;72:609–30.
7. Wicksell SD. The corpuscle problem: a mathematical study of a biometric problem. *Biometrika* 1925;17:84–99.
8. Fisher RA. The effect of methods of ascertainment upon the estimation of frequencies. *Ann Hum Genet* 1934;6:13–25.
9. Neyman J. Statistics—servant of all science. *Science* 1955;122:401–6.
10. Patil GP, Rao CR. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* 1978;34:179–89.
11. Asgharian M, Wolfson C, Wolfson DB. Analysis of biased survival data: the Canadian study of health and aging and beyond. *Stat Action Can Outlook* 2014;193–208.
12. Binder D. Fitting Cox's proportional hazards models from survey data. *Biometrika* 1992;79:139–47.
13. Lin D. On fitting Cox's proportional hazards models to survey data. *Biometrika* 2000;87:37–47.
14. Pugh M, Robins J, Lipsitz S, Harrington D (1993): Inference in the Cox proportional hazards model with missing covariate data. Technical report, Harvard School of Public Health, Dept. of Biostatistics.
15. Chen H, Little R. Proportional hazards regression with missing covariates. *J Am Stat Assoc* 1999;94:896–908.
16. Luo X, Tsai W, Xu Q. Pseudo-partial likelihood estimators for the cox regression model with missing covariates. *Biometrika* 2009;96:617–33.
17. Qi L, Wang C, Prentice R. Weighted estimators for proportional hazards regression with missing covariates. *J Am Stat Assoc* 2005;100:1250–63.
18. Rotnitzky A, Robins JM. Inverse probability weighting in survival analysis. In: Armitage P, Coulton, T, editors. *Encyclopedia of biostatistics*, 2nd ed. New York: Wiley, 2005.
19. Ertefaie A, Asgharian M, Stephens D. Propensity score estimation in the presence of length-biased sampling: a non-parametric adjustment approach. *Stat* 2014;3:83–94.
20. Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Østbye T, Rockwood K, et al. A reevaluation of the duration of survival after the onset of dementia. *N Engl J Med* 2001;344:1111–16.
21. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat* 1978;6:34–58.
22. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat Theory Methods* 1994;23:2379–412.
23. Robins JM. Causal inference from complex longitudinal data. In: Berkane, M, etdior. *Latent variable modeling and applications to causality*. New York: Springer, 1997:69–117.
24. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
25. Hirano K, Imbens G, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 2003;71:1161–89.
26. Cheng Y, Wang M. Estimating propensity scores and causal survival functions using prevalent survival data. *Biometrics* 2012;68:707–16.
27. Robins JM, Rotnitzky A, van der Laan M. On profile likelihood: comment. *J Am Stat Assoc* 2000;95:477–82.
28. van der Laan MJ, Robins JM. *Unified methods for censored longitudinal data and causality*. New York: Springer Science & Business Media, 2003.
29. Cox DR, Oakes D. *Analysis of survival data*. Chapman & Hall/CRC, 1984.
30. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* 1992;48:479–95. URL <http://www.jstor.org/stable/2532304>.
31. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. 2nd ed. New York: Wiley, 2002.
32. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 1952;47:663–85.
33. Hájek J, Dupač V. *Sampling from a finite population*. New York: Marcel Dekker, 1981.
34. Hernán M, Brumback B, Robins J. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000;11:561–70.
35. Robins JM. Association, causation, and marginal structural models. *Synthese* 1999;121:151–79.
36. Tsiatis AA. *Semiparametric theory and missing data*. New York: Springer Verlag, 2006.
37. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc* 1999;94:1096–120.
38. Lipsitz SR, Ibrahim JG, Zhao LP. A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *J Am Stat Assoc* 1999;94:1147–60.

39. Neugebauer R, van der Laan M. Why prefer double robust estimators in causal inference? *J Stat Plan Inference* 2005;129:405–26.
40. Robins JM. Robust estimation in sequentially ignorable missing data and causal inference models. *Proc Am Stat Assoc Sec Bayesian Stat Sci* 1999;6–10.
41. Addona V, Wolfson DB. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Anal* 2006;12:267–84.
42. Asgharian M, Wolfson DB, Zhang X. Checking stationarity of the incidence rate using prevalent cohort survival data. *Stat Med* 2006;25:1751–1767.
43. Carrière Y, Pelletier L. Factors underlying the institutionalization of elderly persons in Canada. *J Gerontol Ser B Psychol Sci Soc Sci* 1995;50:S164.
44. Little R, Rubin DB. *Statistical analysis with missing data*. Vol. 539. New York: Wiley, 1987.
45. Asgharian M, M'lan CE, Wolfson DB. Length-biased sampling with right censoring. *J Am Stat Assoc* 2002;97:201–9.
46. Asgharian M, Wolfson DB. Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann Stat* 2005;33:2109–31. URL <http://www.jstor.org/stable/3448636>.
47. Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* 1989;76:751–61.
48. Wang MC, Jewell NP, Tsai WY. Asymptotic properties of the product limit estimate under random truncation. *Ann Stat* 1986;14:1597–605.
49. Luo X, Tsai W. Nonparametric estimation for right-censored length-biased data: a pseudo-partial likelihood approach. *Biometrika* 2009;96:873–86.
50. Huang CY, Qin J. Nonparametric estimation for length-biased and right-censored data. *Biometrika* 2011;98:177.
51. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed* 2004;75:45–9.
52. Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *Am J Epidemiol* 1996;143:1059.
53. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005;24:3089–110.
54. Wang MC. Nonparametric estimation from cross-sectional survival data. *J Am Stat Assoc* 1991;86:130–43.
55. Pepe, MS, Fleming, TR. Weighted Kaplan-Meier statistics: Large sample and optimality considerations. *J R Stat Soc Ser B (Methodol)* 1991;53(2):341–52.