Statistica Sinica **30** (2020), 487-509
doi:https://doi.org/10.5705/ss.202016.0268

SEMIPARAMETRIC ANALYSIS OF SHORT-TERM AND LONG-TERM HAZARD RATIO MODEL WITH LENGTH-BIASED AND RIGHT-CENSORED DATA

Na Hu, Xuerong Chen and Jianguo Sun

Boehringer Ingelheim (China) Investment Co., Southwestern University of Finance and Economics and University of Missouri

Abstract: This study examines regression analyses of length-biased and right-censored failure time data arising from the short-term and long-term hazard ratio model. Compared with some commonly used models, such as the proportional hazards models, the short-term and long-term hazard ratio model has the advantage of allowing crossing hazard functions and, thus, is more flexible. We propose two methods for estimating the regression parameters, the conditional likelihood approach, and the composite conditional likelihood approach. We establish the asymptotic and finite-sample properties of the proposed estimators, and the numerical results suggest that the methods work well in practical situations. In addition, the approaches are applied to a set of real data arising from a dementia study.

Key words and phrases: Composite likelihood, length-biased, right-censored, short-term and long-term hazard ratio

1. Introduction

In observational studies, two sampling designs are commonly used, namely, the traditional incident cohort sampling design and prevalent sampling design. In the former case, subjects are drawn without considering certain conditions or the disease of interest. In the latter case, subjects are drawn from the group of people exhibiting certain conditions or the disease of interest at the time of enrollment. It is easy to see that the latter is more efficient and practical, in general, and thus tend to be preferred. This is especially true when the disease is rare, because a traditional incident cohort sampling design would take much longer to accumulate sufficient events. On the other hand, analyzing the data arising from a prevalent sampling design is more difficult because we need to deal with the selection bias. This bias is the result of the design including only those subjects who have already experienced an initiating event and survived to the examination time. In other words, with this design, the observed survival

time within the prevalent cohort tends to be longer than that of the target population because the probability of being selected from the target population is proportional to the survival time from disease onset to failure (Vardi (1989)). It is easy to see that the latter design is a special case of left truncation, and the resulting data are usually referred to as length-biased data if the disease onset follows a stationary Poisson distribution. In the following, we focus on regression analyses of such data rather than of regular left-truncated data.

Fields that often produce length-biased data include cancer screening trials (Zelen and Feinleib (1969); Zelen (2004)), economics studies (De Una Alvarez, Otero-Giraldez and Alvarez-Llorente (2003)), epidemiological studies (Keiding (1991); Sansgiry and Akman (2000)), and renewal processes (Cox and Miller (1977); Vardi (1982, 1989)). As a result, many authors have discussed analyses of length-biased failure time data. For example, among others, Huang and Qin (2011), Vardi (1989), and Wang (1991) considered nonparametric estimations of an underlying survival function. In particular, Wang (1991) proposed a product-limit estimator, which we will discuss in greater detail below. For regression analyses, Wang (1996), Ghosh (2008), Qin and Shen (2010), and Huang and Qin (2012) investigated fitting a proportional hazards model to length-biased data, and Shen, Ning and Qin (2009) considered the problem under semiparametric transformation and accelerated failure time models.

A well-known drawback of the aforementioned regression models is that they cannot accommodate the situation in which hazard functions cross, which can occur. To address this, we consider the short-term and long-term hazard ratio model. Prior studies have developed parameter estimation methods for situations without a length bias (Tong, Zhu and Sun (2007); Yang and Prentice (2005); Yang (2011); Yang and Zhao (2012); Yang and Prentice (2015)). In addition, to allow for crossing hazard functions, this class of models is includes many commonly used models, such as the proportional hazards model, as special cases. For the model, we propose two estimation procedures, namely the conditional likelihood approach and the composite conditional likelihood approach. The former is relatively simple and applies to both general left-truncated and length-biased data. The latter takes into account the unique characteristics of length-biased data, and is expected to be more efficient.

The remainder of this paper is organized as follows. Section 2 introduces some notation and describes the length-biased data and the class of two-group short-term and long-term models. In Section 3, we present the two proposed estimation methods and establish the asymptotic properties of the resulting es-

timators. Section 4 provides the results obtained from a simulation study conducted to evaluate the finite-sample performance of the two proposed estimation procedures. These results suggest that the methods work well for practical situations. In Section 5, the methods are applied to a set of real data arising from a dementia study. Section 6 concludes the paper.

2. Length-biased Data and the Short-term and Long-term Models

2.1. Notation and length-biased data

Consider a failure time study that involves two events, an initiating event such as the onset of a disease, and a failure event such as death from the disease, and uses a prevalent sampling design. Let $\tilde{A}$ denote the time from the initiating event to the examination for the disease or the enrollment time in the study, and let $\tilde{T}$ be the time from the initiating event to the failure event, the failure time variable of interest. Then, by the assumption, the study consists only of those subjects with $\tilde{T} \geq \tilde{A} > 0$. In the following, we assume that the examination time is noninformative and that the incidence of the initiating event occurs over calendar time at a constant rate; thus, $\tilde{A}$ follows a uniform distribution.

Let T and V be defined in the same way as $\tilde{T}$ and $\tilde{A}$, respectively, but for the subjects in the study or prevalent sample. In addition, let V denote the time from the examination to the failure event, and suppose there exists a vector of covariates denoted by $\tilde{Z}$ in general, or Z for the study subjects. Note that we drop the $\sim$ to emphasize that in the prevalent sample, the failure time $T = A + V$ must exceed A and, thus, is left-truncated. The joint distribution of the triplet (T, A, Z) has the same joint distribution as $(\tilde{T}, \tilde{A}, \tilde{Z})|\tilde{T} \geq \tilde{A}$. Furthermore, note that, in practice, observations on T are usually subject to right censoring owing to the loss of follow-up or the end of the study. In other words, we only observe the censored failure time $Y = \min(T, A + C)$ and the censoring indicator $\delta = I(V \leq C)$, where C denotes the censoring time measured from the examination. Thus, the observed data from n independent subjects have the form $\{(Y_i, A_i, \delta_i, Z_i), \text{for } i = 1, \ldots, n\}$. In the following, as usual, we assume that (A, V) is independent of C, given Z. However, note that the failure time T is subject to informative censoring because it shares the same truncation time A with the total censoring time $A + C$.

Let $f(t|z)$, $F(t|z)$, and $S(t|z)$ denote the density, cumulative distribution, and survival functions of $\tilde{T}$, respectively, given the covariate $\tilde{Z}$ in the target population, and let $H(t)$ be the distribution function of the truncation time

$\tilde{A}$. Correspondingly, let $g(t|z)$ and $G(t|z)$ denote the density and cumulative distribution functions of T, respectively, given Z or the conditional density and cumulative distribution functions of $\tilde{T}$, respectively, given $\tilde{T} > \tilde{A}$ and $\tilde{Z}$. Then, it can be easily shown that the joint distribution function, say $K(a, t|z)$, of the observed (A, T) can be expressed as

$$dK(a, t|z) = \frac{dF(t|z)dH(a)I(t \geq a)}{\xi(z)}, \tag{2.1}$$

where $\xi = Pr(\tilde{T} \geq \tilde{A}|Z) = \int S(t|z)dH(t)$. It thus follows from the assumption $H(t) = t$ that the joint density function of the observed (A, T) has the form

$$f_{A,T}(a, t|z) = \frac{f(t|z)I(t \geq a)}{\mu(z)}, \tag{2.2}$$

with $\mu(z) = \int_0^\infty sf(s|z)ds$. In addition, we have the following relationship between $G(t|z)$ and $F(t|z)$:

$$dG(t|z) = \frac{tdF(t|z)}{\mu(z)}. \tag{2.3}$$

2.2. Short-term and long-term hazard ratio model

To describe the two-group short-term and long-term hazard ratio model, suppose that the study subjects come from two treatment groups, namely a control and a treatment group, and that $Z_i = 0$ if subject i is from the control group, and one otherwise. For convenience, suppose that the first n_1 subjects are from the control group, and the remaining n_2 subjects are from the treatment group, where $n_1 + n_2 = n$. Assume that the failure time $\tilde{T}$ is absolutely continuous, and let $\lambda_{\bar{P}}(t)$ and $\lambda_{\bar{T}}(t)$ denote the hazard functions of $\tilde{T}$ for the subjects with $Z = 0$ and 1, respectively. Then, the two-group short-term and long-term hazard ratio model postulates that

$$\lambda_{\bar{T}}(t) = \frac{\theta_1\theta_2}{\theta_1 + (\theta_2 - \theta_1)S_{\bar{P}}(t)}\,\lambda_{\bar{P}}(t), \quad t < \tau_0 \tag{2.4}$$

(Yang and Prentice (2005)), where $S_{\bar{P}}(t) = \exp(-\int_0^t \lambda_{\bar{P}}(s)ds)$, θ_1 and θ_2 are two unknown positive parameters, and $\tau_0 = \sup\{t : S_{\bar{P}}(t) > 0\}$. It is apparent that the above model can be rewritten as

$$\frac{\lambda_{\bar{T}}(t)}{\lambda_{\bar{P}}(t)} = \frac{\theta_1\theta_2}{\theta_1 + (\theta_2 - \theta_1)S_{\bar{P}}(t)}.$$

That is, under the model, the ratio of the two hazard functions depends on θ_1, θ_2, and $S_{\bar{P}}(t)$, and is not constant, as it is in the proportional hazards model. More specifically, if $\theta_1 < \theta_2$, the ratio is monotonically decreasing and monotonically

increasing with $\theta_1 > \theta_2$. In addition, it is easy to show that

$$\theta_1 = \lim_{t\to 0} \frac{\lambda_{\bar{T}}(t)}{\lambda_{\bar{P}}(t)}, \quad \theta_2 = \lim_{t\to \tau_0} \frac{\lambda_{\bar{T}}(t)}{\lambda_{\bar{P}}(t)}.$$

Thus, we can interpret θ_1 and θ_2 as the short-term and long-term hazard ratios, respectively.

As mentioned before, model (2.4) includes some commonly used models as special cases. For example, by taking $\theta_1 = \theta_2$, the model reduces to the proportional hazards model. We can also obtain the proportional odds model by setting $\theta_2 = 1$. Define

$$R(t) \;=\; \frac{1 - S_{\bar{P}}(t)}{S_{\bar{P}}(t)}, \quad t < \tau_0,$$

as the odds function of the control group, and $\beta^T = (\beta_1, \beta_2) = (\log(\theta_1), \log(\theta_2))$. Then, the hazard and survival functions of $\tilde{T}_i$, given Z_i, can be expressed as

$$\lambda_i(t;\beta) = \frac{1}{\gamma_{1i}(\beta) + \gamma_{2i}(\beta)R(t)} \frac{dR(t)}{dt},$$

and

$$S_i(t;\beta) = \Big(1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(t)\Big)^{-1/\gamma_{2i}(\beta)},$$

respectively, where $\gamma_{ji}(\beta) = \exp(-\beta_j Z_i)$, for $j = 1, 2$, $i = 1, \ldots, n$. In the following, we refer to $\beta = (\beta_1, \beta_2)^T$ as the regression parameters.

For convenience, in the following, we use τ to denote the longest follow-up time on $\tilde{T}$, with $\tau < \tau_0$.

3. Estimation of the Regression Parameters

In this section, we estimate the regression parameters β and present two methods. The first is the conditional likelihood approach, which can be applied in more general situations, but may be less efficient. The second is the composite conditional likelihood approach, which may can be more efficient.

3.1. Conditional likelihood estimation

In this subsection, we first present a conditional maximum likelihood estimation procedure. Note that, given the observed data, the likelihood function conditional on $\tilde{T} \geq \tilde{A}$ is proportional to

$$\mathcal{L}_C = \prod_{i=1}^{n} \frac{f(Y_i|Z_i)^{\delta_i}\, S(Y_i|Z_i)^{1-\delta_i}}{S_i(A_i|Z_i)}.$$

It thus follows that the log conditional likelihood function under model (2.4) has the form

$$\log \mathcal{L}_C(\beta, R(t)) \propto -\frac{1}{n}\sum_{i=1}^{n} \{-\delta_i \log\{\lambda_i(Y_i;\beta)\} - \log\{S_i(Y_i;\beta)\} + \log\{S_i(A_i;\beta)\}\}$$

under model (2.4). For the estimations of β and $R(t)$, it would be natural to directly maximize $\log \mathcal{L}_C(\beta, R(t))$. On the other hand, it is clear that this would be difficult and computationally intensive. To deal with this, by following Yang and Zhao (2012), we propose estimating $R(t)$ first.

To estimate $R(t)$, note that $R(t) = 1/S_{\bar{P}}(t) - 1$ and, as mentioned above, several nonparametric estimation methods have been proposed based on length-biased and right-censored data. In the following, we employ the product-limit estimator

$$\hat{S}_{\bar{P}}^{(PL)}(t) = \prod_{u\in[0,t]} \{1 - d\hat{\Lambda}(u)\} \tag{3.1}$$

given in Wang (1991), and define the estimator $\hat{\beta}$ as the value of β that maximizes the estimated log likelihood function $\log \mathcal{L}_C(\beta, \hat{R}(t))$. Here,

$$\hat{\Lambda}(t) = \int_0^t \frac{1}{K_1(u)} dN_1(u) \tag{3.2}$$

and $\hat{R}(t) = 1/\hat{S}_{\bar{P}}^{(PL)}(t) - 1$, where $N_1(t) =: \sum_{i\le n_1} N_{1i}(t) = \sum_{i\le n_1} \delta_i I(Y_i \le t)$ and $K_1(t) =: \sum_{i\le n_1} K_{1i}(t) = \sum_{i\le n_1} I(A_i \le t \le Y_i)$. Note that by treating $R(t) = \hat{R}(t)$ as known, the log conditional likelihood function can be simplified to

$$\log \mathcal{L}_C(\beta, R(t)) \propto -\frac{1}{n}\sum_{i>n_1} \left\{ \delta_i \log\Big(\gamma_{1i}(\beta) + \gamma_{2i}(\beta) R(Y_i)\Big) + \frac{1}{\gamma_{2i}(\beta)} \left[\log\left(1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(Y_i)\right) - \log\left(1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(A_i)\right)\right]\right\},$$

because the subjects from the control group do not contribute to the estimation of β.

To establish the asymptotic properties of $\hat{\beta}$, let β_0 denote the true value of β, and define $a^{\otimes 2} = aa^T$ for a vector a,

$$N_2(t) = \sum_{i>n_1} \delta_i I(Y_i \le t)\,,\ K_2(t) = \sum_{i>n_1} I(Y_i \ge t)\,,\ K_3(t) = \sum_{i>n_1} I(A_i \ge t)\,.$$

Then, the negtive estimated log likelihood function $-\log \mathcal{L}_C(\beta, \hat{R}(t))$ can be written equivalently as

$$
\begin{aligned}
L_n(\beta, \hat{R}(t)) =: & \frac{1}{n} \sum_{i>n_1} l_i(\beta, \hat{R}(t)) \\
&= \frac{1}{n}\Bigg\{ \int_0^\tau \log\Big(\gamma_1(\beta) + \gamma_2(\beta)\hat{R}(t)\Big) dN_2(t) \\
&\quad + \int_0^\tau \frac{K_2(t) d\hat{R}(t)}{\gamma_1(\beta) + \gamma_2(\beta)\hat{R}(t)} - \int_0^\tau \frac{K_3(t) d\hat{R}(t)}{\gamma_1(\beta) + \gamma_2(\beta)\hat{R}(t)} \Bigg\},
\end{aligned}
$$

where $\gamma_j(\beta) = \exp(-\beta_j)$, for $j = 1, 2$, and τ denotes the length of the follow-up time, with $\tau < \tau_0$. Define

$$
\begin{aligned}
h_i(\hat{\beta}, \hat{R}) = & \frac{1}{\sqrt{n}} \sum_{i>n_1} \frac{\partial l_i(\beta, \hat{R}(t))}{\partial \beta}\Big|_{\beta=\hat{\beta}} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_1(\hat{\beta}, \hat{R}(Y_i))}{\hat{S}_{\bar{P}}^{(PL)}(t)} \int_0^{Y_i} \frac{\hat{S}_{\bar{P}}^{(PL)}(s-)}{\hat{S}_{\bar{P}}^{(PL)}(s)} \frac{I(K_1(u) > 0)}{K_1(u)} d\hat{M}_i(s) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_2(\hat{\beta}, \hat{R}(A_i))}{\hat{S}_{\bar{P}}^{(PL)}(t)} \int_0^{A_i} \frac{\hat{S}_{\bar{P}}^{(PL)}(s-)}{\hat{S}_{\bar{P}}^{(PL)}(s)} \frac{I(K_1(u) > 0)}{K_1(u)} d\hat{M}_i(s),
\end{aligned}
$$

and $\hat{M}_i(t) = N_{1i}(t) - \int_0^t K_{1i}(u) d\hat{\Lambda}(u)$, where $W_1(\beta, R(t)), W_2(\beta, R(t))$ is provided in the proof of the following theorem.

The theorem below gives the asymptotic properties of $\hat{\beta}$. The proof is outlined in the Appendix.

Theorem 1. *Assume that conditions* (C1)-(C6) *given in the Appendix hold. Then, as* $n \to \infty$, *we have*

1. $\hat{\beta} \xrightarrow{P} \beta_0$;
2. $\sqrt{n}(\hat{\beta} - \beta_0)$ *converges weakly to a zero mean normal distribution, the covariance matrix of which can be estimated consistently as* $H(\hat{\beta})^{-1}\hat{\Sigma}H(\hat{\beta})^{-1}$, *where* $H(\hat{\beta})$ *denotes the Hessian matrix of* $L_n(\beta)$ *at* $\hat{\beta}$ *and* $\hat{\Sigma} = n^{-1} \sum_{i=1}^n [h_i(\hat{\beta}, \hat{R})]^{\otimes 2}$.

3.2. Composite conditional likelihood estimation

Note that the estimator $\hat{\beta}$ defined above applies to any left-truncated data, not just length-biased data. On the other hand, the estimator may not be efficient because it does not use the length-biased assumption that the truncation time $\tilde{A}$ follows a uniform distribution. In this subsection, we develop a composite conditional likelihood estimation procedure that does makes this assumption, and thus may be more efficient.

To develop the composite conditional likelihood estimation procedure, we follow Huang and Qin (2012), who discussed regression analyses of length-biased and right-censored data under the proportional hazards model. More specifically, define $V^0 = \min(T - A, C)$, and not that if the survival time is uncensored, the conditional density function of A given V^0 is identical to the condition density of V^0 given A in the prevalent cohort, in the sense that

$$f_{A=a|V^0=v,\delta=1,Z} = \frac{f(a+v)}{S(a|Z)} = f_{V^0=v|A=a,\delta=1,Z}\,, \quad a \geq 0, v \geq 0\,.$$

The first equality follows directly from the definition. To see the second equality, by following the arguments of Huang and Qin (2012), we can easily show that

$$f_{A,V^0|\delta=1,Z} = \frac{f(a+v|Z)}{\mu(Z)} \frac{P(C > V|Z)}{P(\delta = 1|Z)}, \quad a \geq 0, v \geq 0\,.$$

Therefore, we have

$$f_{A|\delta=1,Z} = \frac{S(a|Z)}{\mu(Z)} \frac{P(C > V|Z)}{P(\delta = 1|Z)}, a \geq 0,$$

and the second equality. This suggests that we should consider the composite conditional likelihood function

$$\begin{aligned}\mathcal{L}_{COM} &= \prod_{i=1}^{n} \left[\frac{f(Y_i|Z_i)}{S(A_i|Z_i)} \times \frac{f(Y_i|Z_i)}{S(V_i^0|Z_i)}\right]^{\delta_i} \left[\frac{S(Y_i|Z_i)}{S(A_i|Z_i)}\right]^{(1-\delta_i)} \\ &= \prod_{i=1}^{n} \frac{f(Y_i|Z_i)^{\delta_i} S(Y_i|Z_i)^{1-\delta_i}}{S(A_i|Z_i)} \times \left[\frac{f(Y_i|Z_i)}{S(V_i^0|Z_i)}\right]^{\delta_i}, \end{aligned} \tag{3.3}$$

which can be rewritten as

$$\mathcal{L}_{COM} = \mathcal{L}_C \times \prod_{i=1}^{n} \left[\frac{f(Y_i|Z_i)}{S(V_i^0|Z_i)}\right]^{\delta_i}.$$

The latter is the product of the conditional likelihood given A, based on all subjects, and that given V, based on the subjects with uncensored survival times. In other words, the composite conditional likelihood function is equivalent to the conditional likelihood function based on the pooled data $\{(A_i, Y_i, \delta_i, Z_i); i = 1, \ldots, n\}$ and $\{(V_i, Y_i, Z_i); \delta_i = 1\}$.

To estimate β, we employ the log composite conditional likelihood function given above under model (2.4), which has the form

$$\begin{aligned}\log \mathcal{L}_{COM}(\beta, R(t)) \propto -\frac{1}{n} \sum_{i>n_1} \Bigg\{ & 2\delta_i \log\Big(\gamma_{1i}(\beta) + \gamma_{2i}(\beta) R(Y_i)\Big) \\ & + \frac{1}{\gamma_{2i}(\beta)} \Bigg[(1+\delta_i) \log\left\{1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(Y_i)\right\}\end{aligned}$$

$$- \log \left\{ 1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(A_i) \right\} - \delta_i \log \left\{ 1 + \frac{\gamma_{2i}(\beta)}{\gamma_{1i}(\beta)} R(V_i^0) \right\} \Bigg] \Bigg\}$$

when treating $R(t)$ as known. Of course, in practice, we do not know $R(t)$. Thus, as in the previous subsection, we propose estimating it using the product-limit estimator based on the pooled data $\{(A_i, Y_i, \delta_i, Z_i); i = 1, \ldots, n\}$ and $\{(V_i, Y_i, Z_i); \delta_i = 1\}$. More specifically, define $\tilde{K}_1(t) =: 2^{-1} \sum_{i \le n_1} \tilde{K}_{1i}(t) = 2^{-1} \sum_{i \le n_1} \{ I(A_i \le t \le Y_i) + \delta_i I(V_i \le t \le Y_i) \}$. Then, the product-limit estimator of $S_{\bar{P}}$ has the form

$$\tilde{S}_{\bar{P}}(t) = \prod_{u \in [0,t]} \{ 1 - d\tilde{\Lambda}(u) \}, \tag{3.4}$$

where

$$\tilde{\Lambda}(t) = \int_0^t \frac{1}{\tilde{K}_1(u)} \, dN_1(u).$$

This yields $\tilde{R}(t) = 1/\tilde{S}_{\bar{P}}(t) - 1$.

Define the estimator $\tilde{\beta}$ as the value of β that maximizes the estimated log composite conditional likelihood function $\log \mathcal{L}_{COM}(\beta, \tilde{R}(t))$. To establish the asymptotic properties of $\tilde{\beta}$, define $K_4(t) = \sum_{i > n_1} \delta_i I(Y_i \ge t)$ and $K_5(t) = \sum_{i > n_1} \delta_i I(V_i^0 \ge t)$. Then, the negtive estimated log composite conditional likelihood function $-\log \mathcal{L}_{COM}(\beta, \tilde{R}(t))$ can be written as

$$\begin{aligned} L_n^*(\beta, \tilde{R}(t)) =: & \frac{1}{n} \sum_{i > n_1} l_i^*(\beta, \tilde{R}(t)) \\ = & \frac{1}{n} \Bigg\{ 2 \int_0^\tau \log \Big(\gamma_1(\beta) + \gamma_2(\beta) \tilde{R}(t) \Big) dN_2(t) \\ & + \int_0^\tau \frac{[K_2(t) + K_5(t)] d\tilde{R}(t)}{\gamma_1(\beta) + \gamma_2(\beta) \tilde{R}(t)} - \int_0^\tau \frac{K_3(t) d\tilde{R}(t)}{\gamma_1(\beta) + \gamma_2(\beta) \tilde{R}(t)} \\ & - \int_0^\tau \frac{K_5(t) d\tilde{R}(t)}{\gamma_1(\beta) + \gamma_2(\beta) \tilde{R}(t)} \Bigg\}. \end{aligned}$$

Define

$$\begin{aligned} h_i^*(\tilde{\beta}, \tilde{R}(t)) = & \frac{1}{\sqrt{n}} \sum_{i > n_1} \frac{\partial l_i^*(\beta, R(t))}{\partial \beta} \Big|_{\beta = \tilde{\beta}} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_1^*(\tilde{\beta}, \tilde{R}(Y_i))}{\tilde{S}_{\bar{P}}(t)} \int_0^{Y_i} \frac{\tilde{S}_{\bar{P}}(s-)}{\tilde{S}_{\bar{P}}(s)} \frac{I(\tilde{K}_1(u) > 0)}{\tilde{K}_1(u)} d\tilde{M}^*{}_i(s) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_2^*(\tilde{\beta}, \tilde{R}(A_i))}{\tilde{S}_{\bar{P}}(t)} \int_0^{A_i} \frac{\tilde{S}_{\bar{P}}(s-)}{\tilde{S}_{\bar{P}}(s)} \frac{I(\tilde{K}_1(u) > 0)}{\tilde{K}_1(u)} d\tilde{M}^*{}_i(s) \end{aligned}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{W_3^*(\tilde{\beta},\tilde{R}(V_i^0))}{\tilde{S}_{\bar{P}}(t)}\int_0^{V_i^0}\frac{\tilde{S}_{\bar{P}}(s-)}{\tilde{S}_{\bar{P}}(s)}\frac{I(\tilde{K}_1(u)>0)}{\tilde{K}_1(u)}d\tilde{M}^*{}_i(s),$$

and $\tilde{M}_i^*(t) = N_{1i}(t) - \int_0^t \tilde{K}_{1i}(u)d\tilde{\Lambda}(u)$, where $W_1^*(\beta,R(t)), W_2^*(\beta,R(t))$, and $W_3^*(\beta,R(t))$ are defined in the proof of the following theorem. The theorem below gives the asymptotic properties of $\tilde{\beta}$. The proof is outlined in the Appendix.

Theorem 2. *Assume that the conditions* (C1)-(C5) *and* (C6*) *given in the Appendix hold. Then, as $n \to \infty$, we have*

1. $\tilde{\beta} \xrightarrow{P} \beta_0$;
2. $\sqrt{n}(\tilde{\beta}-\beta_0)$ *converges weakly to a zero mean normal distribution, the covariance matrix of which can be estimated consistently by* $H^*(\tilde{\beta})^{-1}\,\hat{\Sigma^*}\,H^*(\tilde{\beta})^{-1}$, *where* $H^*(\tilde{\beta})$ *denotes the Hessian matrix of* $L_n^*(\beta)$ *at* $\tilde{\beta}$ *and* $\hat{\Sigma}^* = n^{-1}\sum_{i>n_1}[h_i^*(\tilde{\beta},\tilde{R}(t))]^{\otimes 2}$.

4. A Simulation Study

In this section, we present the results obtained from a simulation study conducted to evaluate the finite-sample performance of the two estimators proposed above. In the study, we generated the failure times from model (2.4) using $S_{\bar{P}}(t) = \lambda\exp(-\lambda t)$, which gives $R(t) = \lambda^{-1}\exp(\lambda t) - 1$. To generate the prevalent cohort or length-biased sample, the left truncation time $\tilde{A}_i$ was generated independently from a uniform distribution over $(0,\omega)$, with ω being larger than the upper bound of $\tilde{T}_i$ to ensure the stationary assumption. Only the pairs $(A_i,T_i) = (\tilde{A}_i,\tilde{T}_i)$, with $\tilde{A}_i \le \tilde{T}_i$, were kept until the required sample sizes were achieved for both the control and the treatment group. The censoring time C_i, was generated from a uniform distribution over $(0,a)$, where a was chosen to give the required percentage of right-censored samples. The results given below are based on 1,000 replications, with $n_1 = n_2 = 100$ or 200.

Tables 1-3 present the results of estimating β, with β_0 taking several different values, and the percentage of right-censored samples being 0%, 15%, or 30%. In particular, we considered $\beta_0^T = (0.5, 0.5)$, which gives the proportional hazards model, $\beta_0^T = (0.5, 0)$ or $(-0.5, 0)$, which correspond to the proportional odds model with an initially negative or positive effect, respectively, and $\beta_0^T = (-0.5, 0.5)$, under which the group effect is initially positive, but gradually becomes negative. Furthermore, in the latter case, the two survival functions cross each other. The results in the tables include the estimated bias (Bias),

Table 1. Summary of simulation studies: Part I.

Sample	Censor	Conditional								Composite									
		Bias		SSE		ESE		CP		Bias		SSE		ESE		CP		RE	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$\beta_1 = \beta_2 = 0$																			
	0%	0.13	-0.07	0.57	0.26	0.59	0.21	0.95	0.84	0.08	-0.07	0.50	0.24	0.54	0.21	0.97	0.92	1.14	1.11
100	15%	0.14	-0.08	0.62	0.32	0.67	0.24	0.95	0.83	0.09	-0.07	0.54	0.27	0.57	0.24	0.97	0.92	1.16	1.17
	30%	0.15	-0.08	0.66	0.40	0.67	0.31	0.94	0.83	0.13	-0.10	0.59	0.34	0.61	0.29	0.96	0.91	1.12	1.18
	0%	0.04	-0.04	0.38	0.17	0.40	0.15	0.96	0.90	0.04	-0.05	0.35	0.16	0.38	0.15	0.98	0.95	1.08	1.05
200	15%	0.06	-0.05	0.39	0.18	0.42	0.17	0.97	0.91	0.05	-0.05	0.37	0.18	0.39	0.17	0.98	0.94	1.07	1.03
	30%	0.07	-0.06	0.42	0.23	0.44	0.19	0.96	0.88	0.06	-0.06	0.39	0.21	0.41	0.20	0.97	0.95	1.07	1.06
$\beta_1 = \beta_2 = 0.5$																			
	0%	0.05	0.01	0.54	0.34	0.56	0.31	0.95	0.93	0.02	0.00	0.51	0.31	0.48	0.29	0.95	0.96	1.06	1.10
100	15%	0.06	0.01	0.58	0.40	0.61	0.36	0.95	0.94	0.03	-0.02	0.54	0.33	0.53	0.34	0.93	0.96	1.07	1.23
	30%	0.08	0.01	0.60	0.47	0.65	0.45	0.95	0.94	0.04	-0.01	0.59	0.39	0.55	0.41	0.93	0.96	1.02	1.18
	0%	0.00	0.00	0.36	0.21	0.39	0.20	0.96	0.94	-0.01	0.01	0.34	0.20	0.32	0.19	0.96	0.97	1.06	1.05
200	15%	0.01	0.00	0.38	0.23	0.41	0.23	0.95	0.94	-0.01	0.01	0.36	0.22	0.34	0.22	0.96	0.97	1.05	1.03
	30%	0.01	0.01	0.41	0.29	0.43	0.28	0.95	0.94	0.01	0.00	0.38	0.26	0.37	0.27	0.95	0.97	1.07	1.09

Table 2. Summary of simulation studies: Part II.

Sample	Censor	Conditional								Composite									
		Bias		SSE		ESE		CP		Bias		SSE		ESE		CP		RE	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$\beta_1 = 0.5, \quad \beta_2 = 0$																			
	0%	0.08	-0.06	0.60	0.25	0.69	0.21	0.96	0.88	0.10	-0.07	0.56	0.21	0.61	0.21	0.97	0.94	1.07	1.18
100	15%	0.08	-0.05	0.64	0.29	0.71	0.24	0.95	0.89	0.08	-0.06	0.60	0.25	0.65	0.24	0.97	0.95	1.05	1.16
	30%	0.09	-0.06	0.67	0.34	0.78	0.29	0.94	0.88	0.07	-0.07	0.62	0.31	0.67	0.28	0.97	0.94	1.09	1.11
	0%	0.06	-0.04	0.42	0.16	0.49	0.14	0.97	0.90	0.04	-0.03	0.37	0.15	0.43	0.14	0.98	0.97	1.14	1.08
200	15%	0.07	-0.04	0.44	0.18	0.52	0.16	0.97	0.92	0.04	-0.04	0.40	0.15	0.44	0.16	0.97	0.97	1.09	1.16
	30%	0.08	-0.05	0.46	0.21	0.54	0.19	0.97	0.92	0.05	-0.05	0.44	0.20	0.47	0.19	0.97	0.96	1.04	1.02
$\beta_1 = -0.5, \quad \beta_2 = 0$																			
	0%	0.13	-0.10	0.54	0.29	0.50	0.22	0.94	0.80	0.11	-0.10	0.53	0.26	0.49	0.22	0.96	0.89	1.03	1.12
100	15%	0.15	-0.11	0.58	0.33	0.53	0.24	0.94	0.80	0.08	-0.09	0.53	0.31	0.50	0.25	0.97	0.88	1.11	1.06
	30%	0.16	-0.13	0.63	0.42	0.56	0.30	0.92	0.77	0.15	-0.13	0.57	0.41	0.54	0.30	0.96	0.89	1.11	1.00
	0%	0.07	-0.05	0.36	0.18	0.35	0.16	0.95	0.88	0.07	-0.05	0.34	0.17	0.34	0.16	0.97	0.92	1.07	1.06
200	15%	0.08	-0.06	0.40	0.21	0.36	0.17	0.94	0.87	0.08	-0.07	0.35	0.20	0.35	0.18	0.97	0.93	1.13	1.09
	30%	0.10	-0.07	0.43	0.27	0.39	0.21	0.94	0.84	0.10	-0.08	0.38	0.23	0.37	0.21	0.97	0.92	1.14	1.17

Table 3. Summary of simulation studies: Part III.

Sample	Censor	Conditional								Composite									
		Bias		SSE		ESE		CP		Bias		SSE		ESE		CP		RE	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$\beta_1 = -0.5, \quad \beta_2 = 0.5$																			
	0%	0.01	0.00	0.47	0.38	0.40	0.32	0.91	0.91	0.03	0.00	0.45	0.34	0.38	0.31	0.94	0.94	1.04	1.09
100	15%	0.06	-0.01	0.49	0.42	0.42	0.36	0.90	0.89	0.04	0.00	0.45	0.40	0.40	0.37	0.95	0.94	1.08	1.05
	30%	0.06	-0.03	0.51	0.51	0.45	0.46	0.92	0.85	0.07	-0.04	0.49	0.48	0.43	0.43	0.95	0.93	1.04	1.05
	0%	0.01	0.00	0.31	0.23	0.27	0.21	0.93	0.94	0.01	0.01	0.29	0.21	0.26	0.21	0.94	0.95	1.05	1.06
200	15%	0.01	0.00	0.34	0.27	0.29	0.24	0.91	0.93	0.02	-0.02	0.31	0.24	0.28	0.24	0.95	0.97	1.10	1.11
	30%	0.03	-0.01	0.35	0.32	0.31	0.28	0.92	0.90	0.04	-0.03	0.33	0.30	0.30	0.28	0.96	0.96	1.05	1.09
$\beta_1 = 0.5, \quad \beta_2 = 1$																			
	0%	0.02	0.08	0.51	0.53	0.48	0.54	0.93	0.95	0.01	0.05	0.51	0.52	0.40	0.50	0.91	0.96	1.00	1.02
100	15%	0.02	0.08	0.53	0.60	0.51	0.65	0.92	0.95	0.04	0.05	0.49	0.48	0.42	0.59	0.94	0.97	1.09	1.23
	30%	0.06	0.05	0.55	0.67	0.55	0.79	0.94	0.95	0.07	0.05	0.57	0.62	0.48	0.73	0.93	0.96	0.98	1.08
	0%	0.00	0.03	0.34	0.32	0.34	0.32	0.95	0.97	0.00	0.03	0.35	0.31	0.27	0.29	0.91	0.96	0.98	1.03
200	15%	0.01	0.04	0.36	0.38	0.35	0.37	0.94	0.97	0.02	0.03	0.35	0.34	0.29	0.33	0.92	0.96	1.02	1.12
	30%	0.02	0.04	0.38	0.45	0.38	0.45	0.94	0.96	0.02	0.04	0.38	0.43	0.32	0.41	0.92	0.96	1.01	1.04

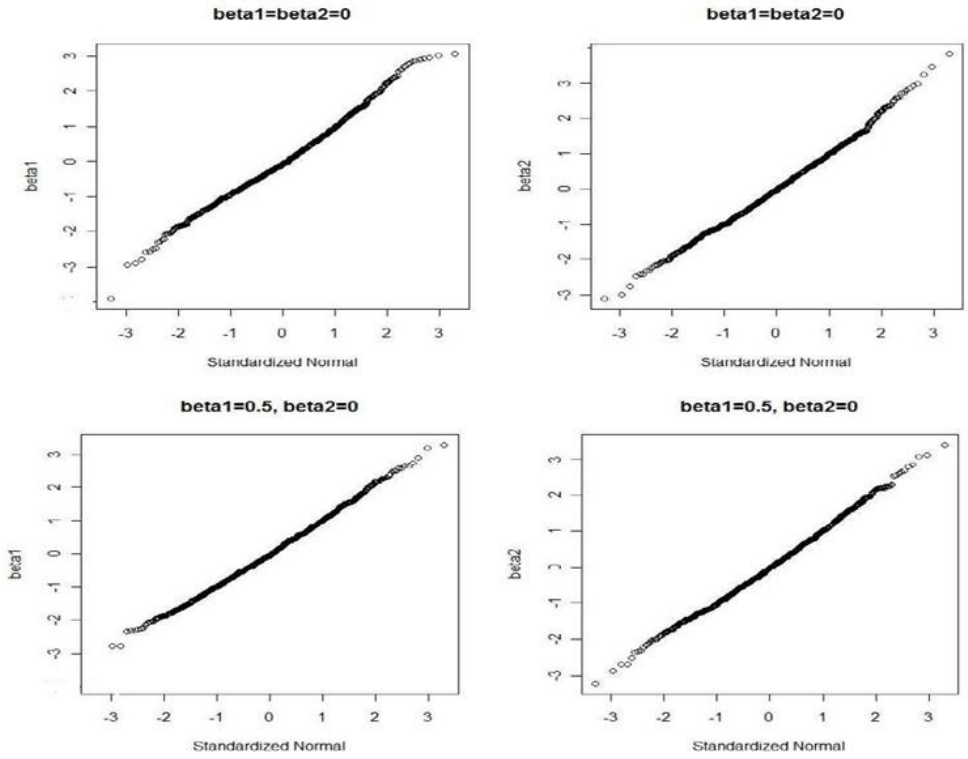

Figure 1. Quantile plots of the standardized maximum composite conditional likelihood estimator for $n_1 = n_2 = 200$ and a censoring rate of 15%.

the sample standard error (SSE) of the estimates, the average of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). The results in the tables show that both of the proposed estimators seem to be unbiased and that the variance estimation seems to be appropriate. When the sample size increases or the right-censoring percentage decreases, both the bias and the estimated standard error become smaller, as expected. In addition, the tables show that the normal approximation to the distributions of the two proposed estimators appears to be reasonable, in general, when the sample size is large. In addition, as discussed above and expected, the composite conditional likelihood estimator is more efficient than the conditional likelihood estimator. However, there exist some low CP, especially on the conditional likelihood estimates of β_2 with $n_1 = n_2 = 100$, reinforcing the notion that we should employ the composite conditional likelihood estimator, in general.

To further assess the normal approximation to the distributions of the two

proposed estimators, we compared the quantile plots of the standardized estimates with those of the standard normal distribution. Figure 1 presents four such plots, corresponding to the results given in Table 1 with $\beta_1 = \beta_2 = 0$, and those shown in Table 2 with $\beta_1 = 0.5$ and $\beta_2 = 0$. In all cases, $n_1 = n_2 = 200$ and we use a 15% censoring rate. The results again indicate that the normal approximation seems to be reasonable for the situations investigated. We also considered other setups for generating length-biased data, and obtained similar results.

Note that in all situations here,we assume that the truncation time $\tilde{A}$ follows a uniform distribution (i.e., we have length-biased data). Following the advice of a reviewer, we also considered situations where $\tilde{A}$ follows some other distribution, or where we have regular left-truncated data only. Table 4 shows some of the results given by the two proposed methods with the truncation time $\tilde{A}$. It also shows the results for the uniform distribution generated from an exponential distribution with a rate of 0.5, gamma distribution $\Gamma(1, 1)$, chi square distribution $\mathcal{X}^2(1)$, and beta distribution $B(1, 3)$. Here, the other setups were the same as those in Table 1-3, and the table includes the estimated bias, sample standard error of the estimates, and mean square error (MSE). The results suggest that, when the stationary assumption is satisfied, the composite conditional likelihood method will usually outperform the conditional likelihood method. On the other hand, if the stationary assumption is violated or if we have regular length-biased data, the former may yield significantly biased estimates, making the latter preferable for parameter estimation.

5. An Application

Now, we apply the estimation procedures proposed in the previous sections to length-biased and right-censored data arising from the Canadian Study of Health and Aging on patients diagnosed with dementia (Wolfson et al. (2001); Addona and Wolfson (2006); Shen, Ning and Qin (2009)). The study is a large prevalent cohort study that began with the random selection of more than 14,000 subjects aged 65 or older from throughout Canada to take part in a health survey. Of these, 10,263 subjects agreed to participate. All patients were screened for dementia in 1991, with 1,132 participants identified as having the disease. It is easy to see that some patients with a serious prognosis of dementia may die before the screening test, and thus will not be included in the study, implying the existence of left truncation. For most patients, the dates of disease onset

Table 4. Comparison of two methods with $n_1 = n_2 = 200$.

(β_1, β_2)	Censor	Conditional						Composite					
		Bias		SSE		MSE		Bias		SSE		MSE	
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
(0.5, -0.5)	$A \sim$ Uniform distribution												
	0%	0.01	-0.06	0.39	0.13	0.15	0.02	0.01	-0.04	0.36	0.11	0.13	0.01
	15%	0.02	-0.05	0.41	0.17	0.17	0.03	0.05	-0.05	0.38	0.13	0.15	0.02
	30%	-0.01	-0.06	0.44	0.22	0.19	0.05	-0.03	-0.03	0.39	0.16	0.15	0.03
	$A \sim$ Exponential distribution												
	0%	0.06	-0.04	0.47	0.14	0.22	0.02	0.21	0.22	0.38	0.12	0.19	0.06
	15%	0.06	-0.04	0.49	0.17	0.24	0.03	0.21	0.22	0.40	0.14	0.21	0.07
	30%	0.06	-0.03	0.51	0.20	0.26	0.04	0.18	0.24	0.42	0.22	0.15	0.08
	$A \sim$ Gamma distribution												
	0%	0.05	-0.04	0.44	0.16	0.20	0.03	0.31	0.33	0.38	0.14	0.24	0.13
	15%	0.05	-0.03	0.47	0.18	0.22	0.04	0.30	0.34	0.41	0.16	0.26	0.14
	30%	0.05	-0.02	0.50	0.20	0.25	0.04	0.28	0.36	0.44	0.17	0.27	0.16
	$A \sim$ Chi square distribution												
	0%	0.06	-0.06	0.51	0.16	0.26	0.03	0.57	0.37	0.46	0.15	0.54	0.16
	15%	0.05	-0.05	0.53	0.19	0.28	0.04	0.60	0.37	0.47	0.16	0.57	0.16
	30%	0.02	-0.04	0.53	0.23	0.27	0.05	0.57	0.37	0.49	0.20	0.57	0.18
	$A \sim$ Beta distribution												
	0%	0.05	-0.04	0.46	0.18	0.21	0.03	0.47	0.52	0.39	0.16	0.37	0.30
	15%	0.05	-0.05	0.46	0.20	0.22	0.04	0.50	0.53	0.39	0.17	0.37	0.31
	30%	0.03	-0.02	0.49	0.25	0.24	0.06	0.44	0.56	0.43	0.23	0.38	0.36
(0.5, 1)	$A \sim$ Uniform distribution												
	0%	0.02	0.01	0.34	0.30	0.12	0.09	0.03	0.01	0.30	0.27	0.09	0.07
	15%	0.02	0.02	0.36	0.35	0.13	0.12	0.03	0.02	0.32	0.32	0.10	0.10
	30%	0.03	0.02	0.38	0.42	0.15	0.18	0.03	0.02	0.34	0.37	0.12	0.14
	$A \sim$ Exponential distribution												
	0%	0.03	0.01	0.33	0.34	0.11	0.11	0.15	0.08	0.31	0.30	0.12	0.10
	15%	0.04	0.02	0.35	0.39	0.13	0.15	0.15	0.09	0.32	0.34	0.13	0.12
	30%	0.05	0.02	0.39	0.52	0.15	0.27	0.16	0.11	0.35	0.47	0.15	0.23
	$A \sim$ Gamma distribution												
	0%	0.03	0.05	0.39	0.40	0.15	0.16	0.22	0.18	0.35	0.35	0.17	0.15
	15%	0.04	0.05	0.42	0.46	0.17	0.22	0.23	0.19	0.38	0.44	0.19	0.22
	30%	0.05	0.08	0.45	0.59	0.21	0.36	0.24	0.21	0.41	0.53	0.22	0.33
	$A \sim$ Chi square distribution												
	0%	0.02	0.02	0.44	0.47	0.20	0.22	0.51	0.16	0.45	0.41	0.46	0.20
	15%	0.04	0.03	0.48	0.58	0.23	0.34	0.53	0.12	0.50	0.49	0.54	0.26
	30%	0.07	0.01	0.51	0.72	0.26	0.52	0.56	0.07	0.53	0.60	0.59	0.37
	$A \sim$ Beta distribution												
	0%	0.01	0.03	0.40	0.47	0.16	0.22	0.40	0.38	0.37	0.43	0.29	0.34
	15%	0.02	0.08	0.43	0.69	0.19	0.48	0.40	0.43	0.40	0.60	0.33	0.54
	30%	0.07	-0.00	0.47	0.73	0.23	0.53	0.44	0.43	0.45	0.75	0.40	0.75

Table 5. Estimated hazard ratios for the CSHA study.

	β_1 (short-term)			β_2 (long-term)		
Method	Estimate	ESE	95%CI	Estimate	ESE	95%CI
Probable Alzheimer vs. Possible Alzheimer (baseline)						
I[a]	-0.275	0.269	(-0.803, 0.252)	0.096	0.165	(-0.227, 0.419)
II[b]	-0.129	0.252	(-0.623, 0.366)	0.166	0.179	(-0.185, 0.518)
Probable Alzheimer vs. Vascular (baseline)						
I	0.473	0.495	(-0.497, 1.442)	-0.215	0.159	(-0.527, 0.097)
II	0.708	0.426	(-0.126, 1.543)	-0.319	0.142	(-0.598, -0.040)
Possible Alzheimer vs. Vascular (baseline)						
I	0.053	0.396	(-0.723, 0.829)	0.076	0.217	(-0.350, 0.503)
II	0.488	0.411	(-0.318, 1.294)	0.066	0.196	(-0.450, 0.318)

[a]I: conditional likelihood estimates

[b]II: composite conditional likelihood estimates

can be obtained from their medical records. Another variable of interest is the overall survival time from disease onset to death. Here, we have right-censored observations because the subjects were only followed until the end of the study in 1996.

In the analyses below, following other authors, we focus on the 818 patients for whom disease onset dates and the classification of three dementia subtypes are known. Of those, 393 subjects had probable Alzheimer's disease, 252 had possible Alzheimer's disease, and 173 had vascular dementia. By the end of this study, 638 patients had died as a result of the disease and the others were right censored. For the data, Addona and Wolfson (2006) validated the stationarity assumption that the incidence of dementia does not change over time, using the method proposed by Wang (1991). In other words, it seems reasonable to treat the observed prevalent cohort data from the study as length-biased data.

For our analysis, we conduct a pairwise comparison of the effects of the dementia subtypes on mortality. The results after applying the two proposed estimation procedures are presented in Table 5. In most cases, the results from the two methods are quite similar, suggesting that there are no significant differences between the short-term and long-term hazard ratios of patients with different dementia subtypes. On the other hand, the composite conditional likelihood estimation procedure indicates that there is a some significant difference between the short-term hazards of patients with probable Alzheimer's disease and those with vascular dementia. To further investigate this, we obtained the nonparametric estimators $\tilde{S}_{\bar{P}}(t)$ given in (3.1) corresponding to the three subtypes (see Figure 2). It seems that the three survival functions cross each other,

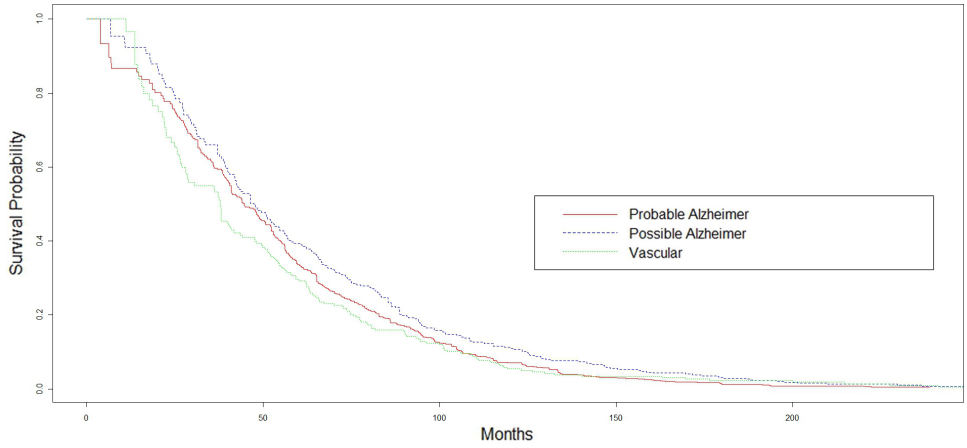

Figure 2. Estimated survival curves for different subtypes of dementia.

and that the main difference between probable Alzheimer's disease and vascular dementia occurred in the early stage.

6. Conclusion

This paper discussed regression analyses of length-biased and right-censored failure time data arising from the short-term and long-term hazard ratio model. Two estimation procedures were proposed. As discussed above and shown in the simulation study, the conditional likelihood approach applies to general left-truncated data, but may be less efficient, whereas the composite conditional likelihood approach is more efficient than the first method, in general, but only applies to length-biased data. In other words, we need to be careful when using the second method, and should check the stationarity assumption before applying it, as in Section 5. For both methods, we established the asymptotic properties of the estimators of the regression parameters. The results of the simulation study suggest that the eatimators work well for practical situations.

There exist several directions for future research. First, model (2.4) applies to the two treatment-group situation only. In some situations, there may exist more than two groups or other covariates. In other words, it would be useful to generalize model (2.4) to a recession setting. For further details, see Yang and Prentice (2005). As noted in Yang (2011), in the case of heavy censoring, $\hat{\theta}_2$ cannot be interpreted as the estimated long-term hazard ratio, because this interpretation requires that $\lim t \to \tau_0$, and there are no data near τ_0, owning to the heavy right censoring. As such, it is currently more complicated and difficult

to obtain an accurate assessment of the treatment effect. Therefore, this is left to future research. we have focused here on right-censored data. However, instead of right censoring, the observed data may suffer from interval censoring, which is a more general form of censoring that includes right censoring as a special case and occurs commonly in medical follow-up studies, among others (Sun (2006)). Thus, another direction for future research is to develop estimation procedures for model (2.2) based on length-biased and interval-censored data.

Acknowledgments

We wish to thank the Editor and two reviewers for their helpful and insightful comments and suggestions. Chen's work was supported by the National Natural Science Foundation of China (NSFC)(11871402), and the Fundamental Research Funds for the Central Universities (JBK140507, JBK1806002).

Appendix: Proofs of Asymptotic Properties

In this appendix, we will sketch the proofs of the asymptotic results given above. For this, we will first describe the conditions needed, which are similar to those used in Yang and Prentice (2005) and Yang and Zhao (2012).

(C1) The true parameter $\beta_{\mathbf{0}}$ lies in a compact set $\mathcal{B}$.

(C2) Assume that $\lim n_1/n = \rho \in (0, 1)$.

(C3) The data range of interest is $[0, \tau]$, where $\tau < \tau_0$ and τ_0 is the upper boundary of the support of the control distribution.

(C4) Assume that $W(t)$ the marginal survival function of the residual censoring time C. $W(t)$ is absolutely continuous for $t \in [0, \tau]$ and $W(\tau) > 0$.

(C5) The survival functions $S_{\bar{P}}(t)$ and $S_T(t)$ of the two comparison groups are absolutely continuous and $S_{\bar{P}}(\tau)S_T(\tau) > 0$.

(C6) The matrix $\partial^2 l_i(\beta, R(t))/\partial\beta\partial\beta$ is continuous and nonsingular in $\mathcal{B}$, $E[(\partial l_i(\beta)/\beta)^{\otimes 2}]$ is finite at β_0.

(C6*) The matrix $\partial^2 l_i^*(\beta, R(t))/\partial\beta\partial\beta$ is continuous and nonsingular in $\mathcal{B}$, $E[(\partial l_i^*(\beta)/\beta)^{\otimes 2}]$ is finite at β_0.

For the proof, define $L(\beta) = E[L_n(\beta)]$. In the following, we will focus on the proof of the results given in Theorem 1 and the proof for Theorem 2 will be similar and most of it thus omitted.

Proof of Theorem 1.

Proof of Consistency. To prove the consistency, by Theorem 5.7 in Van der Vaart (1998), it is enough to verify that

(i) $\sup_{\beta\in\mathcal{B}} \|L_n(\beta, \hat{R}(t)) - L(\beta, R(t))\| = o_p(1)$.

(ii) $\inf_{\beta:d(\beta,\beta_0)>\epsilon} L(\beta, R(t)) > L(\beta_0, R(t))$.

Let's prove condition (i) hold firstly. Note that $\hat{R}(t) - R(t) = o_p(1)$,

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}} \|L_n(\beta, \hat{R}(t)) - L(\beta, R(t))\| \\
&\leq \sup_{\beta\in\mathcal{B}} \|L_n(\beta, \hat{R}(t)) - L_n(\beta, R(t))\| + \sup_{\beta\in\mathcal{B}} \|L_n(\beta, R(t)) - L(\beta, R(t))\|, \\
&=: I_1 + I_2,
\end{aligned}$$

and

$$\begin{aligned}
I_1 = \sup_{\beta\in\mathcal{B}} \Bigg\| \frac{1}{n} \sum_{i>n_1} \Bigg\{ & \delta_i \log\left(1 - \frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)}\right) \\
& + \frac{1}{\gamma_2(\beta)} \Bigg[\log\left(1 - \frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)}\right) \\
& - \log\left(1 - \frac{\gamma_2(\beta)[\hat{R}(A_i) - R(A_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(A_i)}\right)\Bigg]\Bigg\}\Bigg\| \\
\leq \sup_{\beta\in\mathcal{B}} \Bigg\| \frac{1}{n} \sum_{i>n_1} \Bigg\{ & \delta_i \left(\frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)} + o\left(\frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)} \right) \right) \\
& + \frac{1}{\gamma_2(\beta)} \Bigg[\left(\frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)} + o\left(\frac{\gamma_2(\beta)[\hat{R}(Y_i) - R(Y_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(Y_i)} \right) \right) \\
& - \left(\frac{\gamma_2(\beta)[\hat{R}(A_i) - R(A_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(A_i)} + o\left(\frac{\gamma_2(\beta)[\hat{R}(A_i) - R(A_i)]}{\gamma_1(\beta) + \gamma_2(\beta)R(A_i)} \right) \right) \Bigg] \Bigg\} \Bigg\| = o_p(1),
\end{aligned}$$

it is sufficient to prove that $\mathcal{E} = \{l_i(\beta, R(t)) : \beta \in \mathcal{B}\}$ is G-C(Glivenko-Cantelli) class. The classes $\{R(t), t \in [0, \tau_0]\}$ is G-C classes because it is monotone increasing function. The classes $\{K_2(t), t \in [0, \tau_0]\}$ and $\{K_3(t), t \in [0, \tau_0]\}$ are G-C classes because they are bounded variation functions. The classes $\{\gamma_1(\beta) : \beta \in \mathcal{B}\}$ and $\{\gamma_2(\beta) : \beta \in \mathcal{B}\}$ are G-C classes because the exponential function $\exp(\beta_j), j = 1, 2$ is monotone function. And it is easy to see that, there exist positive constants c_1 and c_2 such that $c_1 < \gamma_j(\beta) < c_2$ since $\mathcal{B}$ is compact. By Kosorok (2008, Thm. 9.2.6), we have $\mathcal{E} = \{l_i(\beta, R(t)) : \beta \in \mathcal{B}\}$ is G-C class because $l_i(\beta, R(t))$ is continuous function of $\gamma_1(\beta)$ and $\gamma_2(\beta)$. Thus condition (i) holds.

For condition (ii), note that it is equivalent to that the $\nabla_\beta L(\beta, R(t)) = 0$ has unique solution. We just need to verify that

$$\inf_{\beta: d(\beta,\beta_0)>\epsilon} \left\| \frac{\partial L(\beta, R(t))}{\partial \beta} \right\| > 0.$$

Note by condition (C6), $(\partial L(\beta, R(t))/\partial\beta)|_{\beta=\beta_0} = 0$ and mean value theorem, we have

$$\begin{aligned}
\inf_{\beta: d(\beta,\beta_0)>\epsilon} \left\| \frac{\partial L(\beta, R(t))}{\partial \beta} \right\| &= \inf_{\beta: d(\beta,\beta_0)>\epsilon} \left\| \frac{\partial L(\beta, R(t))}{\partial \beta} - \frac{\partial L(\beta, R(t))}{\partial \beta}\Big|_{\beta=\beta_0} \right\| \\
&= \inf_{\beta: d(\beta,\beta_0)>\epsilon} \left\| \frac{\partial^2 L(\beta, R(t))}{\partial\beta\partial\beta}\Big|_{\beta=\beta^*} (\beta - \beta_0) \right\| \\
&= \inf_{\beta: d(\beta,\beta_0)>\epsilon} \left\| E\frac{\partial^2 l_i(\beta, R(t))}{\partial\beta\partial\beta}\Big|_{\beta=\beta^*} (\beta - \beta_0) \right\| > 0,
\end{aligned}$$

where β^* lies between β and β_0. Therefore, the proof of consistency is completed.

Proof of asymptotic normality. By consistency, we have $\hat\beta - \beta_0 = o_p(1)$. By Taylor expansion, we have

$$\begin{aligned}
0 &= \frac{\partial L_n(\beta, \hat R(t))}{\partial\beta}\Big|_{\beta=\hat\beta} \\
&= \frac{1}{n}\sum_{i>n_1} \frac{\partial l_i(\beta, \hat R(t))}{\partial\beta}\Big|_{\beta=\hat\beta} \\
&= \frac{1}{n}\sum_{i>n_1} \frac{\partial l_i(\beta, \hat R(t))}{\partial\beta}\Big|_{\beta=\beta_0} + \frac{1}{n}\sum_{i>n_1} \frac{\partial^2 l_i(\beta, \hat R(t))}{\partial\beta\partial\beta}\Big|_{\beta=\beta_0} (\hat\beta - \beta_0) + o_p(n^{-1/2}).
\end{aligned}$$

Hence,

$$\sqrt{n}(\hat\beta - \beta_0) = -\left(\frac{1}{n}\sum_{i>n_1} \frac{\partial^2 l_i(\beta, \hat R(t))}{\partial\beta\partial\beta}\Big|_{\beta=\beta_0}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i>n_1} \frac{\partial l_i(\beta, \hat R(t))}{\partial\beta}\Big|_{\beta=\beta_0} + o_p(1).$$

Note that, by the similar method of Flemming and Harrington (1991, pp. 97-98), we have

$$\hat R(t) - R(t) = \frac{1}{\hat S_{\bar P}^{(PL)}(t)} \int_0^t \frac{\hat S_{\bar P}^{(PL)}(s-)}{S_{\bar P}(s)} \frac{I(K_1(u) > 0)}{K_1(u)} dM(s),$$

where $M(t) = N_1(t) - \int_0^t K_1(u)d\Lambda(u) = \sum_{i=1}^n M_i(t) = \sum_{i=1}^n N_{1i}(t) - \int_0^t K_{1i}(u)$ $d\Lambda(u)$. By above equation and Taylor expansion, after some calculation, we have

$$\frac{1}{\sqrt{n}}\sum_{i>n_1} \frac{\partial l_i(\beta, \hat R(t))}{\partial\beta}\Big|_{\beta=\beta_0}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{i>n_1} \frac{\partial l_i(\beta, R(t))}{\partial \beta}\Big|_{\beta=\beta_0} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{W_1(\beta_0, R(Y_i))}{\hat{S}_{\bar{P}}^{(PL)}(t)} \int_0^{Y_i} \frac{\hat{S}_{\bar{P}}^{(PL)}(s-)}{S_{\bar{P}}(s)} \frac{I(K_1(u)>0)}{K_1(u)} dM_i(s) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{W_2(\beta_0, R(A_i))}{\hat{S}_{\bar{P}}^{(PL)}(t)} \int_0^{A_i} \frac{\hat{S}_{\bar{P}}^{(PL)}(s-)}{S_{\bar{P}}(s)} \frac{I(K_1(u)>0)}{K_1(u)} dM_i(s) + o_p(1) \\
&=: \frac{1}{n} \sum_{i=1}^{n} h_i(\beta_0, R(t)) + o_p(1) \xrightarrow{D} N(0, \Sigma),
\end{aligned}
$$

where $\Sigma = E[h_i(\beta_0, \hat{R}(t))]^{\otimes 2}$ and $W_1(\beta, R(t)) = (W_{11}(\beta, R(t)), W_{12}(\beta, R(t)))^T$, $W_2(\beta, R(t)) = (W_{21}(\beta, R(t)), W_{22}(\beta, R(t)))^T$, in which

$$
\begin{aligned}
W_{11}(\beta, R(t)) &= \frac{1}{n} \sum_{i>n_1} \frac{\delta_i \gamma_1(\beta)\gamma_2(\beta) + \gamma_1(\beta)}{(\gamma_1(\beta) + \gamma_2(\beta) R(t))^2}, \\
W_{12}(\beta, R(t)) &= \frac{1}{n} \sum_{i>n_1} \frac{-\delta_i \gamma_1(\beta)\gamma_2(\beta) + 2\gamma_1(\beta) + \gamma_2(\beta) R(t)}{(\gamma_1(\beta) + \gamma_2(\beta) R(t))^2}, \\
W_{21}(\beta, R(t)) &= \frac{1}{n} \sum_{i>n_1} \frac{-\gamma_1(\beta)}{(\gamma_1(\beta) + \gamma_2(\beta) R(t))^2}, \\
W_{22}(\beta, R(t)) &= \frac{1}{n} \sum_{i>n_1} \frac{-\gamma_2(\beta) R(t)}{(\gamma_1(\beta) + \gamma_2(\beta) R(t))^2}.
\end{aligned}
$$

Note that

$$
\frac{1}{n} \sum_{i>n_1} \frac{\partial^2 l_i(\beta, \hat{R}(t))}{\partial \beta \partial \beta}\Big|_{\beta=\beta_0} \xrightarrow{P} H(\beta_0),
$$

thus we have

$$
\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, H(\beta_0)^{-1} \Sigma H(\beta_0)^{-1}).
$$

This complete the proof of Theorem 1.

Proof of Theorem 2. The proofs of consistency of $\tilde{\beta}$ are similar to these of $\hat{\beta}$, and we omit them here.

Proof of asymptotic normality. Note that, by similar derivation of the results on P97-98 of Flemming and Harrington (1991), we have

$$
\tilde{R}(t) - R(t) = \frac{1}{\tilde{S}_{\bar{P}}(t)} \int_0^t \frac{\tilde{S}_{\bar{P}}(s-)}{S_{\bar{P}}(s)} \frac{I(\tilde{K}_1(u)>0)}{\tilde{K}_1(u)} d\tilde{M}(s),
$$

where $M^*(t) = N_1(t) - \int_0^t \tilde{K}_1(u) d\Lambda(u) = \sum_{i=1}^n M_i^*(t) = \sum_{i=1}^n N_{1i}(t) - \int_0^t \tilde{K}_{1i}(u)$

$d\Lambda(u)$. By above equation and Taylor expansion, after some calculation, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i>n_1}\frac{\partial l_i^*(\beta,\hat{R}(t))}{\partial\beta}\Big|_{\beta=\beta_0}\\
&=\frac{1}{\sqrt{n}}\sum_{i>n_1}\frac{\partial l_i^*(\beta,R(t))}{\partial\beta}\Big|_{\beta=\beta_0}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{W_1^*(\beta_0,R(Y_i))}{\tilde{S}_{\bar{P}}(t)}\int_0^{Y_i}\frac{\tilde{S}_{\bar{P}}(s-)}{S_{\bar{P}}(s)}\frac{I(\tilde{K}_1(u)>0)}{\tilde{K}_1(u)}dM_i^*(s)\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{W_2^*(\beta_0,R(A_i))}{\tilde{S}_{\bar{P}}(t)}\int_0^{A_i}\frac{\tilde{S}_{\bar{P}}(s-)}{S_{\bar{P}}(s)}\frac{I(\tilde{K}_1(u)>0)}{\tilde{K}_1(u)}dM_i^*(s)\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{W_3^*(\beta_0,R(V_i^0))}{\tilde{S}_{\bar{P}}(t)}\int_0^{V_i^0}\frac{\tilde{S}_{\bar{P}}(s-)}{S_{\bar{P}}(s)}\frac{I(\tilde{K}_1(u)>0)}{\tilde{K}_1(u)}dM_i^*(s)+o_p(1)\\
&=:\frac{1}{n}\sum_{i=1}^{n}h_i^*(\beta_0,R(t))+o_p(1)\xrightarrow{D}N(0,\Sigma^*),
\end{aligned}$$

where $\Sigma^*=E[h_i^*(\beta_0,\hat{R}(t))]^{\otimes 2}$ and $W_1^*(\beta,R(t))=(W_{11}^*(\beta,R(t)),\ W_{12}^*(\beta,R(t)))^T$, $W_2^*(\beta,R(t))=(W_{21}^*(\beta,R(t)),\ W_{22}^*(\beta,R(t)))^T$ and $W_3^*(\beta,R(t))=(W_{31}^*(\beta,R(t)),\ W_{32}^*(\beta,R(t)))^T$, in which

$$\begin{aligned}
W_{11}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{2\delta_i\gamma_1(\beta)\gamma_2(\beta)+(1+\delta_i)\gamma_1(\beta)}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2},\\
W_{12}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{-2\delta_i\gamma_1(\beta)\gamma_2(\beta)+2(1+\delta_i)\gamma_1(\beta)+(1+\delta_i)\gamma_2(\beta)R(t)}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2},\\
W_{21}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{-\gamma_1}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2},\\
W_{22}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{-\gamma_2(\beta)R(t)}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2},\\
W_{31}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{-\delta_i\gamma_1(\beta)}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2},\\
W_{32}^*(\beta,R(t))&=\frac{1}{n}\sum_{i>n_1}\frac{-\delta_i\gamma_2(\beta)R(t)}{(\gamma_1(\beta)+\gamma_2(\beta)R(t))^2}.
\end{aligned}$$

Then we can obtain conclusion by similar derivation steps of proof of last theorem.

References

Addona, V. and Wolfson, D. B. (2006). A formal test for the stationary of the incidence rate

using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis* **12**, 267–284.

Cox, D. R. and Miller, H. D. (1977). *The Theory of Stochastic Processes*. New York: CRC, London.

De Una Alvarez, J., Otero-Giraldez, M. S. and Alvarez-Llorente, G. (2003). Estimation under length-bias and right-censoring: An application to unemployment duration analysis for married women. *Journal of Applied Statistics* **30**, 283–291.

Flemming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Ghosh, D. (2008). Proportional hazards regression for cancer studies. *Biometrics* **64**, 141–148.

Huang, C. Y. and Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. *Biometrika* **98**, 177–186.

Huang, C. Y. and Qin, J. (2012). Composite partial likelihood estimation under length-biased sampling, with application to a prevalent cohort study of dementia. *Journal of the American Statistical Association* **107**, 946–957.

Keiding, N. (1991). Age-specified incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **154**, 371–412.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science, New York, USA.

Qin, J. and Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under cox model. *Biometrics* **66**, 382–392.

Sansgiry, P. and Akman, O. (2000). Transformations of the lognormal distribution as a selection model. *The American Statistician* **54**, 307–309.

Shen, Y., Ning, J. and Qin, J. (2009). Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association* **104**, 1192–1202.

Sun, J. G. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York.

Tong, X., Zhu, C. and Sun, J. (2007). Semiparametric regression analysis of two-sample current status data, with applications to tumorigenicity experiments. *The Canadian Journal of Statistics* **35**, 575–584.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York, USA.

Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics* **10**, 616–620.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751–761.

Wang, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86**, 130–143.

Wang, M. C. (1996). Hazards regression analysis for length-biased data. *Biometrika* **83**, 343–354.

Wolfson, C., Wolfson, D. B., Asgharian, M., M'Lan, C. E., Ostbye, T., Rockwood, K. and Hogan, D. B. (2001). A reevaluation of the duration of survival after the onset of dementia. *New England Journal of Medicine* **344**, 1111–1116.

Yang, S. (2011). Estimation of the 2-sample hazard ratio function using a semiparametric model. *Biostatistics* **12**, 354–368.

Yang, S. and Prentice, R. (2005). Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika* **92**, 1–17.

Yang, S. and Prentice, R. (2015). Assessing potentially time-dependent treatment effect from clinical trials and observational studies for survival data, with applications to the women's health initiative combined hormone therapy trial. *Statistics in Medicine* **34**, 1801–1817.

Yang, S. and Zhao, Y. (2012). Checking the short-term and long-term hazard ratio model for survival data. *Scandinavian Journal of Statistics* **39**, 554–567.

Zelen, M. (2004). Forward and backward recurrence times and length biased sampling: Age specific models. *Lifetime Data Analysis* **10**, 325–334.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56**, 601–614.

Boehringer Ingelheim (China) Investment Co., Ltd., Shanghai, 200040, China.

E-mail: nh2hd@mail.missouri.edu

Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China.

E-mail: chenxuerong@swufe.edu.cn

Department of Statistics, University of Missouri, Columbia, Missouri, 65201, U.S.A.

E-mail: sunj@missouri.edu

(Received June 2016; accepted May 2018)