Published in final edited form as:

J Am Stat Assoc. 2012 September 1; 107(499): 946–857. doi:10.1080/01621459.2012.682544.

Composite Partial Likelihood Estimation Under Length-Biased Sampling, With Application to a Prevalent Cohort Study of Dementia

Chiung-Yu Huang and

Mathematical Statistician, Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892 (huangchi@niaid.nih.gov)

Jing Qin

Mathematical Statistician, Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892 (jingqin@niaid.nih.gov)

Abstract

The Canadian Study of Health and Aging (CSHA) employed a prevalent cohort design to study survival after onset of dementia, where patients with dementia were sampled and the onset time of dementia was determined retrospectively. The prevalent cohort sampling scheme favors individuals who survive longer. Thus, the observed survival times are subject to length bias. In recent years, there has been a rising interest in developing estimation procedures for prevalent cohort survival data that not only account for length bias but also actually exploit the incidence distribution of the disease to improve efficiency. This article considers semiparametric estimation of the Cox model for the time from dementia onset to death under a stationarity assumption with respect to the disease incidence. Under the stationarity condition, the semiparametric maximum likelihood estimation is expected to be fully efficient yet difficult to perform for statistical practitioners, as the likelihood depends on the baseline hazard function in a complicated way. Moreover, the asymptotic properties of the semiparametric maximum likelihood estimator are not well-studied. Motivated by the composite likelihood method (Besag 1974), we develop a composite partial likelihood method that retains the simplicity of the popular partial likelihood estimator and can be easily performed using standard statistical software. When applied to the CSHA data, the proposed method estimates a significant difference in survival between the vascular dementia group and the possible Alzheimer's disease group, while the partial likelihood method for left-truncated and right-censored data yields a greater standard error and a 95% confidence interval covering 0, thus highlighting the practical value of employing a more efficient methodology. To check the assumption of stable disease for the CSHA data, we also present new graphical and numerical tests in the article. The R code used to obtain the maximum composite partial likelihood estimator for the CSHA data is available in the online Supplementary Material, posted on the journal web site.

Keywords

Backward and forward recurrence time; Cross-sectional sampling; Random truncation; Renewal processes

1. INTRODUCTION

Dementia is a major worldwide health problem among older people. It is an irreversible, progressive brain disease characterized by loss of memory, language, perception, complex motor skills, and other intellectual functions. Death by dementia is usually caused by secondary infections or other physical problems that result from generally poorer health of dementia patients. In North America, the overall prevalence of dementia is estimated to be 6.4% in people older than 60 years. The disease becomes more common in older individuals, with the age-specific prevalence rate reaching a level of 12.8% in people aged 80–84 and of 30.1% in people older than 85 years (Ferri et al. 2005). As the population grows old, the prevalence of dementia in the developed world is expected to increase considerably over the next 20 years. In 2001, the number of dementia cases in North America was estimated to be 3.4 millions. This number is expected to increase to 5.1 million in 2020 and to 9.2 million in 2040.

With annual costs exceeding 100 billion, dementia is the third most costly illness in the United States, second only to heart disease and cancer. Research is actively underway to develop effective interventions to reduce the prevalence and incidence of the disease, improve the quality of life for patients and their caregivers, and reduce the resources needed to provide adequate medical care. To better allocate health services, it is key to evaluate the impact of dementia on life expectancy. In the United States, Medicare beneficiaries must have an estimated life expectancy of less than 6 months to be eligible for hospice. However, information on the survival of patients with specific types of dementia is limited and is usually estimated with substantial uncertainty, particularly for vascular dementia.

Ideally, to estimate survival after onset of dementia, prospective long-term follow-up studies should be conducted by continuously monitoring the incidence of onset of dementia in initially undiseased elderly people. In many situations, studies on incidence cohorts are usually not available, mainly due to cost and time constraint; thus, data on diseased individuals are often collected and analyzed. Between February 1991 and May 1992, the Canadian Study of Health and Aging (CSHA), one of the largest epidemiology studies of dementia (McDowell, Hill, and Lindsay 2005), randomly selected 14,026 persons aged 65 and older from 36 urban and surrounding rural areas in Canada. Among the 10,263 individuals who agreed to participate, a total of 1132 persons with dementia were identified by the Modified Mini-Mental State Examination (MMSE). For each case of dementia, a diagnosis of possible Alzheimer's disease (referred to in this article as possible Alzheimer), probable Alzheimer's disease (referred to as probably Alzheimer), or vascular dementia was assigned by the study staff, and the date of dementia onset was determined by interviewing caregivers. Based on the mortality data collected before May 1997, naively applying the Kaplan-Meier estimator yields a median survival of 6.6 years from the onset of dementia (Wolfson et al. 2001), which falls in the relatively wide range of 5 years to 9.3 years reported in previous studies. Moreover, the partial likelihood method for right-censored data estimates hazard ratios of 1.22 ($P = 0.03$) and 1.34 ($P = 0.01$) when comparing probable Alzheimer and vascular dementia to possible Alzheimer.

As pointed out by Wolfson et al. (2001) the CSHA had a prevalent cohort study design, because survival data were collected from a prevalent cohort of dementia patients who had not experienced the failure event (death) at the time of recruitment. Compared with the incident cohort approach, which follows initially undiseased individuals from disease onset to failure, the prevalent cohort approach can be more efficient especially when used to study a rare disease, as it usually requires a shorter follow-up period to accumulate enough number of failure events. Individuals in the prevalent cohort, however, are not a representative sample of the target disease population, because diseased individuals who died before the

recruitment period would not be qualified to be sampled. In other words, the survival time in a prevalent cohort study is subject to left truncation, where the truncation time is the duration from disease onset to enrollment. As a result, the sampling scheme favors individuals who survive longer; that is, the enrollees tend to have a slow disease progression and thus, a longer duration between disease onset and failure. The Kaplan-Meier estimator, which fails to account for left truncation, can lead to substantial overestimation of the survival time.

To analyze right-censored survival data collected under prevalent sampling, a simple approach is to apply statistical methods for truncated data (see Andersen et al. 1993; Kalbfleisch and Prentice 2002; and references therein), among them the popular product-limit estimator (Wang 1991) for nonparametric estimation and the maximum partial likelihood estimator for semiparametric estimation of the Cox model (Kalbfleisch and Lawless 1991; Wang, Brookmeyer, and Jewell 1993). By applying the truncation product-limit estimator, the median survival after the onset of dementia is estimated to be 3.6 years, which is substantially shorter than the estimate given by the Kaplan-Meier curve. The partial likelihood method for right-censored data estimates hazard ratios of 1.03 ($P = 0.75$) and 1.12 ($P = 0.32$) when comparing probable Alzheimer and vascular dementia to possible Alzheimer. Thus, the significance in the difference between possible Alzheimer and the other two subtypes disappears after adjusting for left truncation. The aforementioned nonparametric and semiparametric estimators are fully efficient when the distribution of the random truncation time is completely unspecified (Wang, Brookmeyer, and Jewell 1993), but they can be very inefficient when the distribution of the truncation time can be parameterized.

In many applications, including the CSHA, it is reasonable to assume that the incidence of disease onset follows a Poisson process (see Zelen and Feinleib 1969; Simon 1980; Winter and Foldes 1988; de Uña-Álvarez 2004); that is, the disease incidence is stable over time. Under the stable disease condition the left-truncation variable is uniformly distributed, and the probability of a survival time being sampled is proportional to its length. In other words, the survival time in the prevalent cohort has a length-biased distribution. To emphasize this important property, the sampling of prevalent cases under the stable disease condition is termed length-biased sampling in this article. When applied to length-biased survival data, the truncation product-limit estimator and the maximum partial likelihood estimator for truncated data are consistent yet inefficient, as the estimation procedures do not exploit the special structure of length-biased sampling. On the other hand, the nonparametric and semiparametric maximum likelihood estimators under length-biased sampling (Vardi 1989; Asgharian, M' Lan, and Wolfson 2002) are efficient yet computationally intensive, as they do not have closed-form expressions. Further, their limiting distributions are intractable (Vardi and Zhang 1992; Asgharian and Wolfson 2005). Many authors, including Luo and Tsai (2009), Tsai (2009), and Qin and Shen (2010), have proposed alternative nonparametric and semiparametric methods that are formulated based on weighted risk sets. Specifically, the risk set at an event time considered by Tsai (2009) consists of all subjects who survive beyond the event time, and the weight function is proportional to the probability that a subject is in the risk set. On the other hand, the risk set considered by Qin and Shen (2010) consists of uncensored subjects who survive beyond the event time, and the weight function is proportional to the probability that a subject is uncensored and is in the risk set. These methods are easier to deal with and have relatively high efficiency; while they can be implemented in statistical software with additional programming, they cannot deal with covariate-dependent censoring.

In this article, we introduce a simple and efficient method for estimating the Cox model with survival data collected under length-biased sampling. Our methodology is motivated by the

composite likelihood method developed by Lindsay (1988) and Arnold and Strauss (1988), whose idea dates back to Besag (1974). As recently highlighted in a special issue of *Statistica Sinica* in January 2011, the composite likelihood method has drawn a good deal of attention and has been widely applied to many important areas of research, including longitudinal data analysis, spatial analysis, and statistical genetics, among others. The composite likelihood method is based on a pseudo-likelihood constructed by compounding low-dimensional marginal or conditional densities. This approach is especially useful for large dimensional data where a fully specified model may not be available or difficult to estimate, as the full likelihood usually involves high-dimensional integrals. (Readers are referred to Varin (2008) and Varin, Reid, and Firth (2011) for a comprehensive review of the composite likelihood methods.)

We explore the application of the composite likelihood method to length-biased survival data. The proposed estimator not only enjoys a simple structure as the maximum partial likelihood estimator, but also allows for covariate-dependent censoring. Importantly, the estimation procedure can be easily performed using standard statistical software without additional programming efforts. When applied to the data from the CSHA, the proposed method yields smaller standard errors in the estimation of regression coefficients in the Cox model. It is also demonstrated through a series of simulations that the proposed estimator is always at least as efficient as other alternative estimators. Owing to its computational simplicity and the gain in efficiency, the proposed composite partial likelihood estimator is expected to be an attractive method for statistical practitioners.

2. METHODOLOGY

2.1 Notation

For individuals in the susceptible population, let T^0 denote the time from the disease onset to the failure event of interest, and let $\mathbf{X}^0$ denote a $p \times 1$ vector of covariates. For the CSHA, T^0 is the survival time after the onset of dementia, and $\mathbf{X}_0 = (\mathbf{X}_1^0, \mathbf{X}_2^0)$ is a vector of indicators for probable Alzheimer and vascular dementia. The conditional density function, survival function, and hazard function of the survival time T^0 given $\mathbf{X}^0 = \mathbf{x}$ are denoted by $f(t|\mathbf{x})$, $S(t|\mathbf{x})$ and $\lambda(t|\mathbf{x})$, respectively. We assume that the survival time in the target population follows the Cox proportional hazards model (Cox 1972)

$$\lambda(t|\mathbf{x}) = \lambda(t) \exp(\boldsymbol{\beta}' \mathbf{x}),$$

where $\lambda(t)$ is an unspecified baseline hazard function and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters.

Define $\Lambda(t) = \int_0^t \lambda(u) du$; thus, Λ is the baseline cumulative hazard function.

We denote by A^0 the time between disease onset and study enrollment, and assume that A^0 is independent of T^0 . In a prevalent cohort study, a diseased subject would be qualified to be sampled only if the failure event does not occur before the sampling time (see Figure 1A), that is, $T^0 \geq A^0$. In other words, T^0 is left truncated by A^0 . Denote by T , A , and $\mathbf{X}$ the survival time, truncation time, and the covariates for individuals in the prevalent cohort. Then the triplet $(T, A, \mathbf{X})$ has the same joint distribution as $(T^0, A^0, \mathbf{X}^0) | T^0 \geq A^0$.

The observation of the survival time in the prevalent cohort is usually subject to right censoring due to study end or premature dropout. Figure 1B depicts the time variables of a censored subject in the prevalent cohort. Instead of observing the actual value of T , we observe the censored survival time $Y = \min(T, A + C)$. In many applications, it is reasonable to assume that the censoring time after enrollment, C , is independent of (T, A) given $\mathbf{X}$. Note,

however, that the survival time T and the total censoring time $A + C$ are dependent, as they share the same A . In other words, the survival time T is subject to informative censoring. Let $G(t|\mathbf{x})$ denote the survival function of C given $\mathbf{X} = \mathbf{x}$. We assume that the observed data $\{(Y^i, A^i, \mathbf{X}_i, \Delta_i), i = 1, \dots, n\}$ are independent and identically distributed copies of $(Y, A, \mathbf{X}, \Delta)$, where $\Delta = I(T \leq A + C)$.

2.2 The Partial Likelihood Method

Under length-biased sampling, the truncation variable A^0 follows a uniform distribution and the joint density function of (T, A) given $\mathbf{X} = \mathbf{x}$ evaluated at (t, a) is

$$\frac{f(t|\mathbf{x})}{\mu(\mathbf{x})} I(t \geq a) \quad (1)$$

(Lancaster 1990, chap. 3), where $\mu(\mathbf{x}) = \int u f(u|\mathbf{x}) du$ is the conditional mean of T^0 given $\mathbf{X}^0 = \mathbf{x}$. It follows from Equation (1) and the Cox model that the full likelihood is proportional to

$$\mathcal{L}_F = \prod_{i=1}^n \frac{f(Y_i|\mathbf{X}_i)^{\Delta_i} S(Y_i|\mathbf{X}_i)^{1-\Delta_i}}{\mu(\mathbf{X}_i)} = \prod_{i=1}^n \frac{\{\lambda(Y_i) \exp(\beta' \mathbf{X}_i)\}^{\Delta_i} \exp\{-\Lambda(Y_i) \exp(\beta' \mathbf{X}_i)\}}{\int_0^\infty \exp\{-\Lambda(u) \exp(\beta' \mathbf{X}_i)\} du}.$$

Direct maximization of $\mathcal{L}_F$ with respect to (β, Λ) is expected to yield fully efficient estimators; however, this approach is computationally cumbersome not only because $\mathcal{L}_F$ involves the nonparametric component Λ in a complicated way but also because it encounters high-dimensional maximization when the sample size is large.

A simple alternative to estimate β is to apply the partial likelihood method for truncated data (Kalbfleisch and Lawless 1991; Wang, Brookmeyer, and Jewell 1993), as survival data arising from length-biased sampling is a special case of truncated data. Note that the full likelihood $\mathcal{L}_F$ can be expressed as $\mathcal{L}_F = \mathcal{L}_T \times \mathcal{L}_M$, where $\mathcal{L}_T$ is the truncation likelihood of Y conditioning on the truncation time A ,

$$\mathcal{L}_T = \prod_{i=1}^n \frac{f(Y_i|\mathbf{X}_i)^{\Delta_i} S(Y_i|\mathbf{X}_i)^{1-\Delta_i}}{S(A_i|\mathbf{X}_i)} = \prod_{i=1}^n \frac{\{\lambda(Y_i) \exp(\beta' \mathbf{X}_i)\}^{\Delta_i} \exp\{-\Lambda(Y_i) \exp(\beta' \mathbf{X}_i)\}}{\exp\{-\Lambda(A_i) \exp(\beta' \mathbf{X}_i)\}},$$

and $\mathcal{L}_M$ is the marginal likelihood of A ,

$$\mathcal{L}_M = \prod_{i=1}^n \frac{S(A_i|\mathbf{X}_i)}{\mu(\mathbf{X}_i)} = \prod_{i=1}^n \frac{\exp\{-\Lambda(A_i) \exp(\beta' \mathbf{X}_i)\}}{\int_0^\infty \exp\{-\Lambda(u) \exp(\beta' \mathbf{X}_i)\} du}.$$

The truncation likelihood can be further reexpressed as $\mathcal{L}_T = \mathcal{L}_P \times \mathcal{L}_R$, where $\mathcal{L}_P$ is the partial likelihood,

$$\mathcal{L}_P = \prod_{i=1}^n \left\{ \frac{\exp(\beta' \mathbf{X}_i)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) I(A_j \leq Y_i \leq Y_j)} \right\}^{\Delta_i},$$

that only depends on β and $\mathcal{L}_R$ is the residual likelihood that depends on both β and Λ . Wang, Brookmeyer, and Jewell (1993) proposed to estimate β by maximizing $\mathcal{L}_P$. Under arbitrary random truncation, the authors showed that the maximum partial likelihood estimator is as efficient as the maximizer of $\mathcal{L}_T$, as the residual likelihood $\mathcal{L}_R$ is ancillary with respect to $\mathcal{L}_T$.

It is well-known that, for Cox model with right censored data, profiling out the baseline hazard function from the full likelihood yields Cox's partial likelihood (Johansen 1983; Murphy and van der Vaart 2000). Interestingly, we have a similar result for the truncation likelihood with truncated survival data: profiling out Λ from the truncation likelihood $\mathcal{L}_T$ also yields the partial likelihood $\mathcal{L}_P$.

2.3 The Composite Partial Likelihood Method: Complete Data

The maximum partial likelihood estimator $\hat{\beta}_P$ that maximizes $\mathcal{L}_P$ can be very inefficient because the information about β in the marginal likelihood $\mathcal{L}_M$ is not used in the estimation procedure. To better exploit the special structure of length-based sampling, various methods have been proposed in the literature. In the case where all study subjects are followed until the occurrence of the failure event, Wang (1996) considered weighting subjects by the inverse of the length of survival time. Because the assigned weight is proportional to the probability of a subject being sampled, the weighted risk set has the same probability structure as that that would be formed by an incidence cohort. In other words, the regression parameter can be estimated by maximizing a pseudo-likelihood function that is asymptotically equivalent to

$$\prod_{i=1}^n \left\{ \frac{\exp(\beta' \mathbf{X}_i)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) T_j^{-1} I(A_j \leq T_i \leq t_j)} \right\}.$$

This method, however, is not applicable in the presence of right censoring, because the weight for a censored survival time cannot be determined.

Write $V = T - A$; thus, V denotes the residual lifetime after enrollment. In the absence of censoring, it follows from Equation (1) that the pair of random variables (A, V) has an exchangeable joint density function $f(a + v | \mathbf{x}) / \mu(\mathbf{x})$ for $a \geq 0$ and $v \geq 0$, and the common marginal density function is

$$f_A(t | \mathbf{x}) = f_V(t | \mathbf{x}) = \frac{S(t | \mathbf{x})}{\mu(\mathbf{x})}.$$

A simple way to exploit this exchangeability is by applying the composite conditional likelihood method (Arnold and Strauss 1988), as the conditional density function of V given A is identical to that of A given V . As the result, the truncation density of $T = A + V$ conditioning on A is the same as the conditional density of T conditioning on V . Consider the composite conditional likelihood given by the product of the conditional likelihood of V given A and the conditional likelihood of A given V :

$$\prod_{i=1}^n \left\{ \frac{f(T_i | \mathbf{X}_i)}{S(A_i | \mathbf{X}_i)} \times \frac{f(T_i | \mathbf{X}_i)}{S(V_i | \mathbf{X}_i)} \right\} = \prod_{i=1}^n \left[\frac{\lambda(T_i) \exp(\beta' \mathbf{X}_i) \exp \{-\Lambda(T_i) \exp(\beta' \mathbf{X}_i)\}}{\exp \{-\Lambda(A_i) \exp(\beta' \mathbf{X}_i)\}} \times \frac{\lambda(T_i) \exp(\beta' \mathbf{X}_i) \exp \{-\Lambda(T_i) \exp(\beta' \mathbf{X}_i)\}}{\exp \{-\Lambda(V_i) \exp(\beta' \mathbf{X}_i)\}} \right], \quad (2)$$

where $V_i = T_i - A_i$. Hence, Equation (2) is equivalent to the product of the truncation likelihood of T conditioning on A and the conditional likelihood of T conditioning on V .

Following the profile likelihood argument for the partial likelihood $\mathcal{L}_P$ discussed in the previous section, we profile out Λ from Equation (2) to obtain the composite partial likelihood, up to a constant,

$$\prod_{i=1}^n \left[\frac{2 \exp(\beta' \mathbf{X}_i)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) \{I(A_j \leq T_i \leq T_j) + I(V_j \leq T_i \leq T_j)\}} \right]^2.$$

Thus, β can be estimated by solving

$$1 - \frac{n \sum_{i=1}^n \left[\mathbf{X}_i - \frac{\sum_{j=1}^n \mathbf{X}_j \exp(\beta' \mathbf{X}_j) \{I(A_j \leq T_i \leq T_j) + I(V_j \leq T_i \leq T_j)\}}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) \{I(A_j \leq T_i \leq T_j) + I(V_j \leq T_i \leq T_j)\}} \right]}{n \sum_{i=1}^n \exp(\beta' \mathbf{X}_i) \{I(A_i \leq T_i \leq T_i) + I(V_i \leq T_i \leq T_i)\}} = 0.$$

It is easy to see that, in the absence of censoring, the proposed maximum composite partial likelihood estimator is equivalent to the maximum partial likelihood estimator applied to the pooled left-truncated survival data $\{(T_i, A_i, \mathbf{X}_i), i = 1, \dots, n\}$ and $\{(T_i, V_i, \mathbf{X}_i), i = 1, \dots, n\}$, where in the augmented dataset V_i is treated as the truncation time for T_i .

2.4 The Composite Partial Likelihood Method: Censored Data

In the presence of right censoring, the residual survival time V in the prevalent cohort is subject to right censoring. Let $\tilde{V} = \min \{T - A, C\}$ be the observed residual lifetime, that is, $\tilde{V} = V$ for uncensored subjects and $\tilde{V} = C = Y - A$ for censored subjects. Intuitively, the joint distribution of A and $\tilde{V}$ are not exchangeable, because the truncation time is always observed but the residual lifetime could be censored. Hence, it is necessary to modify the composite conditional likelihood (Equation (2)) to accommodate censoring.

To construct a composite conditional likelihood under right censoring, we consider the bivariate failure time $(A, \tilde{V})$ for subjects with uncensored failure time. The joint density function of $(A, \tilde{V})$ conditional on $\Delta = 1$ is:

$$(A=a, \tilde{V}=v) | (\Delta=1, \mathbf{X}=\mathbf{x}) \sim \frac{f(a+v|\mathbf{x})}{\mu(\mathbf{x})} \times \frac{G(v|\mathbf{x})}{\text{pr}(\Delta=1|\mathbf{X}=\mathbf{x})}, a \geq 0, v \geq 0.$$

Thus, it can be observed that A and $\tilde{V}$ do not have an exchangeable joint distribution despite conditioning on $\Delta = 1$. Moreover, straightforward algebra yields

$$\tilde{V}=v | (\Delta=1, \mathbf{X}=\mathbf{x}) \sim \frac{S(v|\mathbf{x})}{\mu(\mathbf{x})} \times \frac{G(v|\mathbf{x})}{\text{pr}(\Delta=1|\mathbf{X}=\mathbf{x})}, v \geq 0.$$

Hence, we have

$$A=a | (\Delta=1, \tilde{V}=v, \mathbf{X}=\mathbf{x}) \sim \frac{f(a+v|\mathbf{x})}{S(v|\mathbf{x})}, a \geq 0, v \geq 0.$$

Interestingly, given that the survival time is uncensored, A conditional on $\tilde{V} = v$ has the conditional density function

$$\frac{f(a+v|x)}{S(v|x)}, a \geq 0, v \geq 0,$$

which is identical to the conditional density function of $V = T - A$ given A in the prevalent population. Hence, for uncensored survival times, we can construct a composite likelihood based on the conditional density of A conditional on $\tilde{V}$ as well as that of $\tilde{V}$ conditional on A . On the other hand, for censored survival times, the conditional density of A conditional on the censored residual lifetime $\tilde{V}$ involves the distribution of the censoring time; hence, it is difficult to be used directly in the construction of the composite likelihood.

We now construct a composite conditional likelihood for length-biased and right-censored survival data. Write $m = \sum_{i=1}^n \Delta_i$; thus, m is the total number of subjects with a uncensored failure time. For convenience, we assume the first m subjects are uncensored; that is, $\Delta_i = 1$ for $i = 1, \dots, m$ and $\Delta_i = 0$ for $i = m+1, \dots, n$. A composite conditional likelihood can be formulated as

$$\prod_{i=1}^m \left\{ \frac{f(Y_i|\mathbf{X}_i)}{S(A_i|\mathbf{X}_i)} \times \frac{f(Y_i|\mathbf{X}_i)}{S(\tilde{V}_i|\mathbf{X}_i)} \right\} \times \prod_{i=m+1}^n \frac{S(Y_i|\mathbf{X}_i)}{S(A_i|\mathbf{X}_i)} = \prod_{i=1}^n \frac{\{\lambda(Y_i)\exp(\beta' \mathbf{X}_i)^{2\Delta_i} \exp\{-(1+\Delta_i)\Lambda(Y_i)\exp(\beta' \mathbf{X}_i)\}\}}{\exp[-\{\Lambda(A_i) + \Delta_i\Lambda(\tilde{V}_i)\}\exp(\beta' \mathbf{X}_i)]} \quad (3)$$

For any fixed β , the composite conditional likelihood is maximized by

$$\hat{\Lambda}(t, \beta) = \sum_{i=1}^n \frac{2\Delta_i I(Y_i \leq t)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) \{I(A_j \leq Y_i \leq Y_j) + \Delta_j I(\tilde{V}_j \leq Y_i \leq Y_j)\}},$$

which is a Breslow-type estimator for Λ . Replacing Λ with $\hat{\Lambda}$ in Equation (3) yields the composite partial likelihood, up to a constant,

$$\mathcal{L} = \prod_{i=1}^n \times \left\{ \frac{2\exp(\beta' \mathbf{X}_i)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) \{I(A_j \leq Y_i \leq Y_j) + \Delta_j I(\tilde{V}_j \leq Y_i \leq Y_j)\}} \right\}^{2\Delta_i}.$$

Note that the composite partial likelihood only depends on β . Hence, the maximum composite partial likelihood estimator $\hat{\beta}$ can be obtained directly by solving the (normalized) estimating equation

$$U(\beta) = \frac{1}{n} \sum_{i=1}^n \Delta_i \left[\mathbf{X}_i - \left[\sum_{j=1}^n \mathbf{X}_j \exp(\beta' \mathbf{X}_j) \{I(A_j \leq Y_i \leq Y_j) + \Delta_j I(\tilde{V}_j \leq Y_i \leq Y_j)\} \right] / \left[\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) \{I(A_j \leq Y_i \leq Y_j) + \Delta_j I(\tilde{V}_j \leq Y_i \leq Y_j)\} \right] \right] = \mathbf{0}.$$

It is easy to see that the proposed estimator is equivalent to the maximum partial likelihood estimator applied to the truncated survival data pooled from $\{(Y^i, A^i, X^i, \Delta_i), i = 1, \dots, n\}$ and $\{(Y^i, \tilde{V}^i, X^i, \Delta_i = 1), i = 1, \dots, m\}$, where the augmented dataset is formed by the m uncensored subjects with $\tilde{V}^i$ being treated as the truncation time for Y^i .

In practice, we can treat the pooled-truncated survival data as clustered survival data, and use the robust variance estimator, which is available in standard statistical software, to estimate the standard errors of the estimated regression coefficients. Thus, the proposed estimator can be easily obtained using standard software. The maximum composite partial likelihood estimator is consistent and asymptotically normal. Its large-sample properties are summarized in the Appendix.

Finally, in the case where time-dependent covariates are involved in modeling the survival time distribution, theoretical justification for the composite partial likelihood method remains valid. The construction of the augmented dataset requires information about the covariate values in the entire time interval $(\tilde{V}_i^*, Y_i)$; however, in practice, it can be difficult to determine the covariate values in the subinterval $(\tilde{V}_i^*, A_i)$ when $\tilde{V}_i^* < A_i$. A careful investigation on the extension of the proposed methodology to time-dependent covariates is warranted.

3. NUMERICAL STUDIES

We conducted simulation studies to examine the finite-sample performance of the proposed methodology. We set the sampling time ξ to be 100 and simulated the onset time of a stable disease, W^0 , from a uniform distribution over $[0, 100]$. The covariates X_1^0 and X_2^0 were independently generated from the Bernoulli distribution with $\Pr(X_1^0=1)=\Pr(X_1^0=0)=0.5$ and the standard normal distribution. The failure time T^0 was generated from three Cox models with different hazard functions: (I) $2 \exp(X_1^0+X_2^0)$, (II) $2t \exp(X_1^0+X_2^0)$, and (III) $0.5(t-2)^2 \exp(X_1^0+X_2^0)$. These models correspond to constant, increasing, and U-shape hazards. To form a prevalent cohort of sample size n , realizations of (W^0, T^0, X_1^0, X_2^0) were generated repeatedly until there are n subjects satisfying the sampling constraint $W^0 + T^0 \leq \xi$. The time from enrollment ξ to loss to follow-up was generated from a uniform distribution so that the censoring rate was approximately 0%, 30%, and 50%. In each simulation, we generated 2000 datasets, each with a sample size of $n = 400$.

To compare the performance of the proposed maximum composite partial likelihood estimator to that of existing methods, we applied the proposed estimator, the popular maximum partial likelihood method for truncated survival data, the estimating equation method (EE-2) studied by Qin and Shen (2010), and the maximum pseudo-partial likelihood estimator studied by Tsai (2009). Note that the maximum partial likelihood estimator solves the partial score equation

$$U_p(\beta) = \sum_{i=1}^n \Delta_i \left[\mathbf{X}_i - \frac{\sum_{j=1}^n \mathbf{X}_j \exp(\beta' \mathbf{X}_j) I(Y_j \geq Y_i \geq A_j)}{\sum_{j=1}^n \exp(\beta' \mathbf{X}_j) I(Y_j \geq Y_i \geq A_j)} \right] = \mathbf{0},$$

while Qin and Shen (2010) proposed to solve the estimating equation

$$U_1(\beta) = \sum_{i=1}^n \Delta_i \left[\mathbf{X}_i - \frac{\sum_{j=1}^n \Delta_j \mathbf{X}_j \exp(\beta' \mathbf{X}_j) \left\{ \widehat{w}_c(Y_j) \right\}^{-1} I(Y_j \geq Y_i)}{\sum_{j=1}^n \Delta_j \exp(\beta' \mathbf{X}_j) \left\{ \widehat{w}_c(Y_j) \right\}^{-1} I(Y_j \geq Y_i)} \right] = \mathbf{0},$$

where $\widehat{w}_c(t) = \int_0^t \widehat{G}(u) du$ and $\widehat{G}(u)$ was the Kaplan-Meier estimator of the censoring time distribution. Moreover, Tsai's estimator is equivalent to the solution of the estimating equation

$$U_2(\beta) = \sum_{i=1}^n \Delta_i \left[\mathbf{X}_i - \frac{\sum_{j=1}^n \widehat{W}_t(Y_j, \Delta_j, Y_i) \mathbf{X}_j \exp(\beta' \mathbf{X}_j) I(Y_j \geq Y_i)}{\sum_{j=1}^n \widehat{W}_t(Y_j, \Delta_j, Y_i) \exp(\beta' \mathbf{X}_j) I(Y_j \geq Y_i)} \right] = \mathbf{0},$$

where

$$\widehat{W}_t(t, \delta, u) = \delta \left\{ \frac{\widehat{w}_c(t) - \widehat{w}_c(t-u)}{\widehat{w}_c(t)} + (1-\delta) \left\{ \frac{\widehat{G}(t-u) - \widehat{G}(t)}{1 - \widehat{G}(t)} \right\} \right\}.$$

Table 1 summarizes the empirical bias and the empirical standard error of these four estimators under different Cox models. The simulation result suggested that all four estimators are close to their estimands, as indicated by small empirical bias. As expected, the proposed estimator outperformed the maximum partial likelihood estimator in all scenarios, with the relative efficiency ranging from 1.08 to 1.73. The proposed estimator had smaller bootstrap standard errors than the estimating equation method studied by Qin and Shen (2010) under right-censoring, while having similar efficiency gains in the absence of censoring. Intuitively, this is because the estimation function $U_1(\beta)$ uses only uncensored subjects to form the weighted risk set, while the proposed estimator uses both censored and uncensored subjects. Overall, the proposed estimator was as or slightly more efficient than the maximum pseudo-partial likelihood estimator in all scenarios. The weight function employed in the maximum pseudo-partial likelihood estimator (Tsai 2009) involves the estimated survival function of the censoring time, and hence, it can be unstable in later time intervals. Our estimator, on the other hand, does not involve estimation of the censoring distribution, hence, is in general more stable. The average of the 2000 robust standard error estimates was very close to the standard deviation of the 2000 estimated coefficients, suggesting satisfactory performance of the robust standard error estimator. Note that we applied Breslow's method to handle ties in the failure times. In fact, in all simulation studies (results not shown), Breslow's method always yielded smaller mean square errors than the other methods that handle ties. In summary, the proposed methodology was easy to perform in standard software, and enjoyed better or similar gains in efficiency over its competitors.

4. ANALYSIS OF THE CANADIAN STUDY OF HEALTH AND AGING

In this section, we analyze the prevalent cohort survival data from the CSHA by applying the proposed method as well as its competitors. After excluding those with missing date of onset or missing dementia subtype classification and those who survived more than 20 years after dementia onset (hence, were considered unlikely to have dementia; see Asgharian, M'Land, and Wolfson 2002), a total of 807 dementia patients were included in our analysis; among them, 249 (31%) had possible Alzheimer, 388 (48%) had probable Alzheimer, and 170 (21%) had vascular dementia. In May 1997, 627 deaths (78%) were recorded; among them, 189 had a diagnosis of possible Alzheimer, 302 had probable Alzheimer, and 136 had vascular dementia at enrollment. The goal of our analysis was to evaluate the impact of different subtypes of dementia on life expectancy using the prevalent cases whose vital status were determined in May 1997.

4.1 Assessment of the Stable Disease Condition

To assess whether the incidence of each dementia subtype is constant over time, Asgharian, Wolfson, and Zhang (2006) proposed to compare the distribution of the truncation time A and that of the residual survival time V . The authors proved that, under mild conditions, A and V have identical distributions if and only if the incidence of disease is constant over time. Denote by S_A and S_V the survival functions of A and V . Let $\hat{S}_A(t)$ be the empirical survival function obtained by using $\{A_1, \dots, A_n\}$, and let $\hat{S}_V(t)$ be the Kaplan-Meier curve obtained by using $\{(\tilde{V}^1, \Delta_1), \dots, (\tilde{V}^n, \Delta_n)\}$. Hence, if the stationarity assumption with respect to disease incidence holds, the two survival curves are expected to be close to each other. The visual inspection, however, can be very arbitrary. In what follows, we describe a graphical assessment method and its corresponding numerical test based on the equality of S_A and S_V .

- Step 1.** Calculate the Kolmogorov-Smirnov test statistic $D = \sup_{t \in [0, \tau]} |\hat{S}_A(t) - \hat{S}_V(t)|$ that compares the Kaplan-Meier estimates of S_A and S_V .
- Step 2.** Apply the E-M algorithm proposed by Vardi (1989) to obtain the nonparametric maximum likelihood estimate of the survival time distribution under the stationarity condition. Let $\hat{f}$ and $\hat{F}$ be the estimated density and survival time function.
- Step 3.** Approximate the distribution of D under the null via simulation:
 - a.** For the i th individual, simulate a total survival time T_i^* from the estimated conditional density of T given the observed truncation time A_i ; that is, simulate T_i^* from the discrete density function $\hat{f}(t)/\hat{S}(A_i)I(t - A_i)$. Next, generate a censoring time C_i^* independently from the estimated censoring distribution using the Kaplan-Meier estimator. Thus, the observed residual lifetime and the the censoring indicator simulated for the i th person are $\tilde{V}_i^* = \min(T_i^* - A_i, C_i^*)$ and $\Delta_i^* = I(T_i^* - A_i \leq C_i^*)$.
 - b.** Calculate the Kolmogorov-Smirnov test statistic $D^* = \sup_{t \in [0, \tau]} |\hat{S}_A(t) - \hat{S}_V^*(t)|$, where $\hat{S}_V^*(t)$ is the Kaplan-Meier estimator based on the simulated data $(\tilde{V}_i^*, \Delta_i^*), i=1, \dots, n$.
- Step 4.** Repeat Step 3 B times, and let $D_1^*, \dots, D_B^*$ be the Kolmogorov-Smirnov test statistic in these B repetitions. Thus, an approximate p -value for the Kolmogorov-Smirnov test $D = \sup_{t \in [0, \tau]} |\hat{S}_A(t) - \hat{S}_V(t)|$ can be estimated by the empirical probability that the value of D exceeds the observed value of $D_1^*, \dots, D_B^*$ through simulation. To assess how unusual the observed residual lifetime distribution is under the stationarity condition, one may plot $\hat{S}_V$ along with the Kaplan-Meier estimate $\hat{S}_V^*$ from a few, say 20, simulated datasets.

We have applied the proposed graphical and numerical tests to data simulated from the models described in Section 3. Overall, the Type I error rate is very close to the predetermined nominal level (0.05) in all scenarios with 2000 repetitions, which supports the validity of the proposed test. We thus applied the proposed tests to data from each dementia subtype group. Figure 2 shows that, within each diagnostic group, the estimated curve of S_V falls well inside the range determined by the simulated processes. The p -values for the Kolmogorov-Smirnov test are 0.20, 0.97, and 0.62, respectively, for possible Alzheimer, probable Alzheimer, and vascular dementia with 2000 repetitions. Thus, there is

little evidence to dispute the stationarity condition required by the proposed composite partial likelihood method.

4.2 Nonparametric Estimation Within Subgroups

Ignoring the prevalent sampling design, the median survival times estimated by naive use of the Kaplan-Meier curve were 7.1, 6.6, and 5.8 years, respectively, for patients with possible Alzheimer, probable Alzheimer, and vascular dementia. On the other hand, the unbiased but inefficient truncation product-limit estimator estimated median survival times of 3.5, 3.6, and 3.3 years for the three dementia subtypes. The differences between the Kaplan-Meier estimates and the truncation product-limit estimates illustrate that failing to adjust for sampling bias in the prevalent cohort leads to substantial overestimation of survival in patients with dementia.

We also applied the truncation product-limit estimator to the augmented dataset described in Section 2.4 to estimate the survival time distribution under the stationarity assumption with respect to disease incidence. Figure 3 shows the estimated survival curves for different dementia subtypes. It can be observed that individuals with vascular dementia had the worst survival, while the estimated survival curves for the other two subtypes were very close to each other. The estimated median survival times for the three dementia subtypes were 4.0 (possible Alzheimer; 95% bootstrap confidence interval, 3.3–4.7), 3.7 (probable Alzheimer; 95% bootstrap confidence interval, 2.8–4.5), and 3.1 years (vascular dementia; 95% bootstrap confidence interval, 2.3–4.2). These estimates were not significantly different than those given by the truncation product-limit estimator, as their corresponding 95% bootstrap confidence intervals contained the estimated median survival time obtained by applying the truncation product-limit estimator.

For policy planners, it is of interest to know the life expectancy of a patient after being admitted into an assisted living facility. Based on the estimated survival curves, the median residual survival times for patients who survive longer than 5 years after onset of dementia are 2.7 years (possible Alzheimer), 2.8 years (probable Alzheimer), and 2.4 years (vascular dementia). Similarly, the estimated median residual survival times are 2.6 years (possible Alzheimer), 2.1 years (probable Alzheimer), and 1.1 years (vascular dementia) given that these patients survive longer than 8 years after onset of dementia.

4.3 Comparisons of Different Dementia Subtypes

To compare the effects of dementia subtypes on mortality, we fit a Cox proportional hazards model with indicators of possible Alzheimer and vascular dementia as covariates. The estimated regression coefficients obtained by applying different methods were summarized in Table 2. By applying the proposed methodology, it was estimated that, as compared to patients with possible Alzheimer, the risk of death increased by 12% among those with probable Alzheimer and by 24% among those with vascular dementia. Thus, vascular dementia had the worst prognosis of all causes. Note that our results are different than what were reported by Qin and Shen (2010), because their analysis included those who survived longer than 20 years after dementia onset. It could be observed that both the partial likelihood method and the pseudo-partial likelihood method (Tsai 2009) estimated nonsignificant effects of possible Alzheimer and vascular dementia, as the corresponding 95% bootstrap confidence intervals contain 0. The proposed method yielded similar estimates as the estimation equation method proposed by Qin and Shen (2010), and both estimated a significantly higher risk of death in patients with vascular dementia. The proposed composite partial method had the smallest bootstrap standard errors when compared with other estimators. For both β_1 and β_2 , the variance ratio for the competitors to the proposed method is always at least 1.45. This suggests that if a competitor method were

used in lieu of the proposed method, the CHSA would need to recruit at least 510 more subjects to achieve the same precision, thus highlighting the practical value of employing a more efficient estimator.

5. DISCUSSION

Statistical methods for left-truncated survival data have been widely applied to study important problems such as estimating the incubation period from human immunodeficiency virus (HIV) infection to the development of acquired immunodeficiency syndrome (AIDS; Brookmeyer and Gail 1987) as well as evaluating the impact of deferring initiating of the highly active antiretroviral therapies (HAART) treatment on AIDS and death (When To Start Consortium 2009). In these applications, the stationarity assumption with respect to disease incidence is likely to be violated, because HIV infection rate increased rapidly in the 1980s and then plateaued in the early 1990s due to extensive prevention and education efforts. Although the composite partial likelihood estimators yield invalid inferential results when the incidence of the disease is not constant over time, the proposed methods can be applied to a transformed dataset to obtain unbiased estimates when the information about the underlying truncation time distribution, denoted by H , is available from other sources. In fact, it is easy to verify that, by applying the monotone transformation H , the transformed survival time $H(T^0)$ is left truncated by a uniformly distributed random variable. Because the Cox model is invariant under monotone transformation, the regression coefficients can be consistently estimated by applying the proposed composite partial likelihood estimator to the transformed data $\{H(A_i), H(Y_i), \Delta^i, X^i, i = 1, \dots, n\}$.

In the Appendix, we established the strong consistency of $\hat{\Lambda}(t)$ for $t \in [0, \tau]$, where τ is the maximal support of $Y = \min(T, A + C)$. In fact, despite the fact that the censoring time C may have a smaller maximal support than T^0 , under length-biased sampling τ equals the maximal support of T^0 , denoted by τ_1 . Intuitively, this is because the truncation random variable A has the same maximal support as T^0 . Heuristically, suppose $\text{pr}(C > 0) > 0$, then for any $\varepsilon > 0$ we have

$$\Pr \left\{ \max_{1 \leq i \leq n} \Delta_i Y_i < \tau_1 - \varepsilon \right\} = \left[1 - E \left\{ \int_{\tau_1 - \varepsilon}^{\tau_1} \frac{f(u|x)}{\mu(x)} w_c(u|x) du \right\} \right]^n > 0,$$

where $w_c(t|x) = \int_0^t G(u|x) du$, and it follows that $\max_{1 \leq i \leq n} \Delta_i Y_i \rightarrow \tau_1$ as $n \rightarrow \infty$. Thus, we can show that $\tau = \tau_1$. An important implication of this result is that under length-biased sampling, the mean survival time $\mu(x)$ can be estimated consistently. This is in contrary to the conventional survival analysis where the mean survival time is usually not estimable due to right censoring.

The motivation for the use of composite likelihood is usually to avoid the computational challenges in evaluating and maximizing the likelihood for a possibly high-dimensional response vector. In this article, we applied the composite likelihood approach to improve the efficiency of the popular partial likelihood method by incorporating additional information in the distribution of the observed truncation time. Our simulation study showed that proposed estimator performed better or as well as the other alternative methods for survival data under length-biased sampling. We also observed an efficiency gain of 57% and higher for the composite partial likelihood estimator over the partial likelihood estimator in the data analysis of the CSHA.

Although we focus on semiparametric estimation of the Cox proportional hazards model in this article, a simple and efficient nonparametric estimator for the baseline hazard function can be easily derived by applying the truncation product-limit estimator (Wang, Jewell, and Tsai 1986) to the augmented dataset described above. Specifically, the new nonparametric estimator can be obtained by replacing the total mass of the risk set at time t ,

$\sum_{i=1}^n I(A_i \leq t \leq Y_i)$, in the truncation product-limit estimator with $\bar{\phi}(\hat{\beta}) = n^{-1} \sum_{i=1}^n \hat{\phi}_i(\hat{\beta})$.

Similar to the semiparametric maximum composite partial likelihood estimator, the nonparametric estimator enjoys the efficiency improvement over the product-limit estimator. In summary, the proposed methodology not only enjoys efficiency gains over its competitors but also can be easily performed using standard statistical software; hence, it is expected to be an attractive method for statistical practitioners.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Professors Ian McDowell, Masoud Asgharian, and Christina Wolfson for kindly sharing the Canadian Study of Health and Aging data. The core study was funded by 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 authors also thank the Associate Editor, the referee, Dr Dean Follmann, and Dr Michael Proschan for their comments that improved the presentation of this article.

APPENDIX: LARGE-SAMPLE PROPERTIES

To simplify the discussion, we impose the following regularity conditions. We assume that X is bounded, and the true regression parameter β_0 lies in a compact set $\mathcal{B}$. Let $S(t)$ and $G(t)$ be the marginal survival functions of T^0 and C , respectively. We assume that $\text{pr}(C > 0) > 0$ and that both $S(t)$ and $G(t)$ are absolutely continuous for $t \in [0, \tau]$, where $\tau = \sup\{t: \text{pr}(Y \leq t) > 0\}$ is the maximal support of Y . Let $F^u(t) = \text{pr}(Y \leq t, \Delta = 1)$ be the distribution function of uncensored failure times. Write $N_j(u) = \Delta_j I(Y_j \leq u)$ and $R_j(u) = 2^{-1} \{I(A_j \leq u \leq Y_j) + \Delta_j I(\tilde{V}_j \leq u \leq Y_j)\}$. Define $\sum_{i=1}^n R_i(t) = 2^{-1} \sum_{i=1}^n \{I(A_i \leq t \leq Y_i) + \Delta_i I(\tilde{V}_i \leq t \leq Y_i)\}$ for $k = 0, 1, 2$, and let $s^{(k)}(t, \beta) = E\{S^{(k)}(t, \beta)\}$. Thus, U can be reexpressed as

$$U(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ X_i - \frac{S^{(1)}(u, \beta)}{S^{(0)}(u, \beta)} \right\} dN_i(u).$$

It is easy to see that, for each fixed β , U defines a functional of the empirical processes

$S^{(k)}(t, \beta) = n^{-1} \sum_{i=1}^n X_i^{\otimes k} \exp(\beta' X_i) R_i(t)$. By the strong law of large numbers, the four empirical processes converge almost surely to their limits uniformly as $n \rightarrow \infty$. Moreover, the functional defined by U is continuous with respect to the supremum norm under regularity conditions. Hence, it follows from the extended strong law of large numbers as given in Appendix III by Andersen and Gill (1982) that $\sup_{\beta \in \mathcal{B}} |U(\beta) - \tilde{U}(\beta)| \rightarrow 0$ almost surely as $n \rightarrow \infty$, with

$$\tilde{U}(\beta) = E\{X_1 N_1(\tau)\} - \int_0^\tau \frac{s^{(1)}(u, \beta)}{s^{(0)}(u, \beta)} dF^u(u).$$

Under length-biased sampling, we have $E\{dN_1(t) | X_1 = x\} = f(t|x) \mu(x)^{-1} w_c(t|x) dt$ and

$$\begin{aligned} E[R_1(t) | X_1 = x] &= \frac{1}{2} \left\{ \text{pr}(A_1 \leq t \leq Y_1 | X_1 = x) + \text{pr}(\Delta_1 = 1, \tilde{V}_1 \leq t \leq Y_1 | X_1 = x) \right\} \\ &= \frac{1}{2} \left\{ \frac{S(t|x)}{\mu(x)} w_c(t|x) + \frac{S(t|x)}{\mu(x)} w_c(t|x) \right\} = \frac{S(t|x)}{\mu(x)} w_c(t|x), \end{aligned}$$

where $S^{(0)}(u, \beta)$, $S^{(1)}(u, \beta)$, $n^{-1} \sum_{i=1}^n N_i(u)$, and $n^{-1} \sum_{i=1}^n X_i N_i(\tau)$. It can be verified that $\tilde{U}(\beta_0) = 0$ if β_0 is the true regression parameter.

Define the stochastic process $w_c(t|x) = \int_0^t G(u|x) du R_i(u) d\Lambda(u)$ and

$$\phi_i(\beta) = \int_0^\tau \left\{ X_i - \frac{s^{(1)}(u, \beta)}{s^{(0)}(u, \beta)} \right\} dM_i(u, \beta).$$

Let $\Gamma(\beta) = -U(\beta)/\beta$ and $\tilde{\Gamma}(\beta) = -\tilde{U}(\beta)/\beta$. In light of the functional form representation of U , the large sample properties are studied by applying the functional delta method (Andersen et al. 1993, chap. II.8). The results are summarized in the following theorem.

Theorem 1. Suppose that X is bounded and the true regression parameter β_0 lies in a compact set $\mathcal{B}$. Then the maximum composite likelihood estimator $\hat{\beta}$ converges to β_0 almost surely when $n \rightarrow \infty$. Provided $\tilde{\Gamma}(\beta_0)$ is nonsingular, $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean 0 and a variance-covariance matrix given by $\tilde{\Omega}(\beta_0) = \tilde{\Gamma}(\beta_0)^{-1} \tilde{\Sigma}(\beta_0) \tilde{\Gamma}(\beta_0)^{-1}$, where $\tilde{\Sigma}(\beta_0) = E\{\phi_1(\beta_0)' \phi_1(\beta_0)\}$.

Proof. We establish the consistency of $\hat{\beta}$ as follows. Consider a log pseudo-likelihood function

$$K(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[(\beta - \beta_0)' X_i - \log \left\{ \frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)} \right\} \right] dN_i(u),$$

and define the function

$$\tilde{K}(\beta) = (\beta - \beta_0)' \cdot E\{X_1 N_1(\tau)\} - \int_0^\tau \log \left\{ \frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)} \right\} dF^u(u).$$

It is easy to verify that $U(\beta)$ and $\tilde{U}(\beta)$ are derivatives of $K(\beta)$ and $\tilde{K}(\beta)$, with $K(\beta_0) = \tilde{K}(\beta_0) = 0$. Applying Taylor expansion, we have $K(\beta) - \tilde{K}(\beta) = (\beta - \beta_0)' \{U(\beta^*) - \tilde{U}(\beta^*)\}$, where β^* lies on the line segment between β and β_0 . Hence, it follows that $\sup_{\beta \in \mathcal{B}} |U(\beta) - \tilde{U}(\beta)| \rightarrow 0$ almost surely and that $\sup_{\beta \in \mathcal{B}} |K(\beta) - \tilde{K}(\beta)| \rightarrow 0$ almost surely as $n \rightarrow \infty$.

Define $\Gamma(\beta) = dU(\beta)/d\beta$ and $\tilde{\Gamma}(\beta) = d\tilde{U}(\beta)/d\beta$, that is,

$$\Gamma(\beta) = \int_0^\tau \left[-\frac{S^{(2)}(u, \beta)}{S^{(0)}(u, \beta)} + \left\{ \frac{S^{(1)}(u, \beta)}{S^{(0)}(u, \beta)} \right\}^{\otimes 2} \right] \left\{ \frac{1}{n} \sum_{i=1}^n dN_i(u) \right\}$$

and

$$\tilde{\Gamma}(\beta) = \int_0^\tau \left[-\frac{s^{(2)}(u, \beta)}{s^{(0)}(u, \beta)} + \left\{ \frac{s^{(1)}(u, \beta)}{s^{(0)}(u, \beta)} \right\}^{\otimes 2} \right] dF^u(u).$$

Applying the Cauchy-Schwartz inequality, we can show that Γ and $\tilde{\Gamma}$ are both negative definite, and it follows that K and $\tilde{K}$ are concave. Arguing as in the proof of Lenglar's theorem (Andersen and Gill 1982, Appendix II), we can show that the unique maximum of $K, \hat{\beta}$, converges in probability to the unique maximum of $\tilde{K}$, that is, β_0 . Thus, we establish the consistency of $\hat{\beta}$.

Next, we show the asymptotic normality of $U(\beta)$ for a fixed β . The regularity conditions imply that $s^{(0)}(u, \beta)$ is bounded away from zero for $u \in [0, \tau]$ and $\beta \in \mathcal{B}$. Hence, the functional defined by U is compactly differentiable with respect to the supremum norm. The asymptotic normality of the four empirical processes in U can be easily established, as they are sums of independent and identically distributed random processes. Therefore, by the functional delta method (Andersen et al. 1993, chap. II.8), we can show that

$M_i(t, \beta) = N_i(t) - \int_0^t \exp(\beta' X_i)$. It follows the classic central limit theorem that $n^{1/2}U(\beta)$ is asymptotically normal with mean zero and variance-covariance matrix $\Sigma(\beta) = E\{\phi_1(\beta) \phi_1(\beta)'\}$.

Applying the Taylor series expansion, we have $U(\hat{\beta}) - U(\beta_0) = \Gamma(\beta^*)(\hat{\beta} - \beta_0)$, where β^* lies on the line segment between $\hat{\beta}$ and β_0 . Arguing as before, one can show that $\sup_{\beta \in \mathcal{B}} \|\Gamma(\beta) - \tilde{\Gamma}(\beta)\| \rightarrow 0$. Because $\hat{\beta}$ is consistent for β_0 and $\tilde{\Gamma}$ is continuous at β_0 under the regularity conditions, we have $\Gamma(\beta^*)$ converges to $\tilde{\Gamma}(\beta_0)$ almost surely. By Slutsky's

theorem, $U(\beta) = n^{-1} \sum_{i=1}^n \phi_i(\beta) + o_p(n^{-1/2})$ converges to a mean zero normal distribution with a variance-covariance matrix $\tilde{\Gamma}(\beta_0)^{-1} \Sigma(\beta_0) \tilde{\Gamma}(\beta_0)^{-1}$. Thus, we prove Theorem 1.

The Breslow-type estimator $\hat{\Lambda}(t, \beta)$ for the baseline cumulative hazard function can be reexpressed as $\sqrt{n}(\hat{\beta} - \beta_0)$. It is easy to see that $\hat{\Lambda}(t, \beta)$ is a continuous functional of two

empirical processes, $S^{(0)}(u, \beta)$ and $\hat{\Lambda}(t, \beta) = \int_0^t \{S^{(0)}(u, \beta)\}^{-1} d\left\{n^{-1} \sum_{i=1}^n N_i(u)\right\}$, with respect to the supremum norm. The almost surely convergence of the two processes implies

$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|\hat{\Lambda}(t, \beta) - \Lambda(t, \beta)\| \rightarrow 0$, where $n^{-1} \sum_{i=1}^n N_i(u)$. Then the strong consistency of $\hat{\Lambda}(t) \equiv \hat{\Lambda}(t, \hat{\beta})$ for $\Lambda(t) = \Lambda(t, \beta_0)$ follows the strong consistency of $\hat{\beta}$ for β_0 .

Define $\Lambda(t, \beta) = \int_0^t \{s^{(0)}(u, \beta)\}^{-1} dF^u(u)$ and let

$$\hat{\phi}_i(\beta) = \int_0^\tau \left\{ X_i - \frac{S^{(1)}(u, \beta)}{S^{(0)}(u, \beta)} \right\} d\hat{M}_i(u, \beta).$$

It can be shown that $\hat{\Sigma}(\beta_0)$ is consistently estimated by

$$\widehat{M}_i(t, \beta) = N_i(t) - \int_0^t \exp(\beta' X_i) R_i(u) d\widehat{\Lambda}(u, \beta), \text{ where}$$

$\widehat{\Sigma}(\widehat{\beta}) = n^{-1} \sum_{i=1}^n \{\widehat{\phi}_i(\widehat{\beta}) - \bar{\phi}(\widehat{\beta})\} \{\widehat{\phi}_i(\widehat{\beta}) - \bar{\phi}(\widehat{\beta})\}'$. Thus, the asymptotic variance $\tilde{\Omega}(\beta_0)$ in Theorem 1 can be consistently estimated by $\hat{\Omega}(\hat{\beta}) = \Gamma(\hat{\beta})^{-1} \hat{\Sigma}(\hat{\beta}) \Gamma(\hat{\beta})^{-1}$.

REFERENCES

- Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical Models Based on Counting Processes. New York: Springer-Verlag; 1993. [947,955,956]
2. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. The Annals of Statistics. 1982; 10:1100–1120. [955,956].
- Arnold BC, Strauss D. Bivariate Distributions With Exponential Conditionals. Journal of the American Statistical Association. 1988; 83:522–527. [947,949].
- Asgharian M, M' Lan CE, Wolfson DB. Length-Biased Sampling With Right Censoring: An Unconditional Approach. Journal of the American Statistical Association. 2002; 97:201–209. [947,952].
- Asgharian M, Wolfson DB. Asymptotic Behavior of the Unconditional NPMLE of the Length-Biased Survivor Function From Right Censored Prevalent Cohort Data. The Annals of Statistics. 2005; 33:2109–2131. [947].
- Asgharian M, Wolfson DB, Zhang X. Checking Stationarity of the Incidence Rate Using Prevalent Cohort Survival Data. Statistics in Medicine. 2006; 25:1751–1767. [952]. [PubMed: 16220462]
- Besag JE. Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society, Series B. 1974; 34:192–236. [946,947].
- Brookmeyer R, Gail MH. Biases in Prevalent Cohorts. Biometrics. 1987; 43:739–749. [954]. [PubMed: 3427161]
- Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series B. 1972; 34:187–220. (with discussion), [948].
- de Uña-Álvarez J. Nonparametric Estimation Under Length-Biased Sampling and Type I Censoring: A Moment Based Approach. The Annals of the Institute of Statistical Mathematics. 2004; 56:667–681. [947].
- Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Kathleen Hall KH, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E, Scazufca M. for Alzheimers Disease International. Global Prevalence of Dementia: A Delphi Consensus Study. Lancet. 2005; 366:2112–2117. [946]. [PubMed: 16360788]
- Johansen S. An Extension of Cox's Regression Model. International Statistical Review. 1983; 51:165–174. [949].
- Kalbfleisch JD, Lawless JF. Regression Models for Right Truncated Data With Applications to AIDS Incubation Times and Reporting Bias. Statistica Sinica. 1991; 1:19–32. [947,948].
- Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2nd ed.. New York: Wiley; 2002. [947]
- Lancaster, T. The Econometric Analysis of Transition Data. Cambridge: Cambridge University Press; 1990. [948]
- Lindsay BG. Composite Likelihood Methods. Contemporary Mathematics. 1988; 80:221–239. [947].
- Luo X, Tsai WY. Nonparametric Estimation for Right-Censored Length-Biased Data: A Pseudo-Partial Likelihood Approach. Biometrika. 2009; 96:873–886. [947].
- McDowell I, Hill G, Lindsay J. An Overview of the Canadian Study of Health and Aging. International Psychogeriatrics. 2005; 13:7–18. [946]. [PubMed: 11892976]
- Murphy SA, van der Vaart AW. On Profile Likelihood. Journal of the American Statistical Association. 2000; 95:449–465. [949].
- Qin J, Shen Y. Statistical Methods for Analyzing Right- Censored Length-Biased Data Under Cox Model. Biometrics. 2010; 66:382–392. [947,951,953]. [PubMed: 19522872]

- Simon R. Length Biased Sampling in Etiologic Studies. *American Journal of Epidemiology*. 1980; 111:444–452. [947]. [PubMed: 7377187]
- Tsai W-Y. Pseudo-Partial Likelihood for Proportional Hazards Models With Biased-Sampling Data. *Biometrika*. 2009; 96:601–615. [947,951,953]. [PubMed: 22422175]
- Vardi Y. Multiplicative Censoring, Renewal Processes, Deconvolution and Decreasing Density: Nonparametric Estimation. *Biometrika*. 1989; 76:751–761. [947,952].
- Vardi Y, Zhang C-H. Large Sample Study of Empirical Distributions in a Random-Multiplicative Censoring Model. *The Annals of Statistics*. 1992; 20:1022–1039. [947].
- Varin C. On Composite Marginal Likelihoods. *Advances in Statistical Analysis*. 2008; 92:1–28. [947].
- Varin C, Reid N, Firth Reid D. An Overview of Composite Likelihood Methods. *Statistica Sinica*. 2011; 21:5–42. [947].
- Wang M-C. Nonparametric Estimation From Cross-Sectional Survival Data. *Journal of the American Statistical Association*. 1991; 86:130–143. [947].
- Wang M-C. Hazards Regression Analysis for Length-Biased Data. *Biometrika*. 1996; 83:343–354. [949].
- Wang M-C, Brookmeyer R, Jewell N. Statistical Models for Prevalent Cohort Data. *Biometrics*. 1993; 49:1–11. [947,948,949]. [PubMed: 8513095]
- Wang M-C, Jewell N, Tsai W-Y. Asymptotic Properties of the Product Limit Estimate Under Random Truncation. *The Annals of Statistics*. 1986; 14:1597–1605. [955].
- When To Start Consortium. Timing of Initiation of Antiretroviral Therapy in AIDS-Free HIV-1-Infected Patients: A Collaborative Analysis of 18 HIV Cohort Studies. *Lancet*. 2009; 373:1352–1363. [954]. [PubMed: 19361855]
- Winter BB, Foldes A. A Product-Limit Estimator for Use With Length-Biased Data. *Canadian Journal of Statistics*. 1988; 16:337–355. [947].
- Wolfson C, Wolfson DB, Asgharian M, M'Lan CE. A Reevaluation of the Duration of Survival After the Onset of Dementia. *New England Journal of Medicine*. 2001; 344:1111–1116. [947]. [PubMed: 11297701]
- Zelen M, Feinleib M. On the Theory of Screening for Chronic Diseases. *Biometrika*. 1969; 56:601–614. [947].

Figure 1.

(A) The horizontal lines denote time from onset of dementia to death in the susceptible population, $\circ$ denotes censoring, and $\odot$ indicates the sampling time of an individual. The prevalent population (bolded lines) consists of subjects with dementia at the time of recruitment; dementia patients who died before the sampling period (dashed lines) are not qualified to be sampled. (B) A patient has disease onset at calendar time W and, if still alive, is available for enrollment at calendar time ξ into the prevalent cohort. The time to death is T and the calendar time of death is $W+T$. A patient in the prevalent cohort is followed for a period of time C ; thus, we observe $Y = \min\{T, A+C\}$. For patients in the prevalent cohort,

T must exceed $A = \xi - W$, the time from onset to enrollment, and thus, T is a left-truncated random variable.

Figure 2.

Estimated survival functions for V . Plots of the estimated survival functions for the residual lifetime in the prevalent cohort. The bold curve corresponds to the Kaplan-Meier estimator using the observed data and the gray curves are estimated survival curves from 20 (out of 2000) datasets that were simulated under the null.

Figure 3. Estimated survival curves for possible Alzheimer (dashed line), probable Alzheimer (solid line), and vascular dementia (dotted line) given by Vardi's estimator.

Table 1

Summary of simulation studies

	Proportion censored	Coef	PL		QS		PPL		CPL			
			Bias	SE	Bias	SE	Bias	SE	Bias	SE		
Scenario I: $\lambda_0(t) = 2$												
0%		$\hat{\beta}_1$	8	132	3	98	4	98	5	100	98	1.75
		$\hat{\beta}_2$	4	83	1	66	1	66	2	69	66	1.48
30%		$\hat{\beta}_1$	7	157	13	141	0	126	5	120	120	1.70
		$\hat{\beta}_2$	7	102	5	92	0	84	1	82	83	1.54
50%		$\hat{\beta}_1$	10	171	12	153	-1	141	3	130	127	1.73
		$\hat{\beta}_2$	5	108	4	100	-2	92	0	89	88	1.48
Scenario II: $\lambda_0(t) = 2$												
0%		$\hat{\beta}_1$	7	119	5	99	5	99	6	101	99	1.38
		$\hat{\beta}_2$	6	77	4	65	4	65	5	67	65	1.29
30%		$\hat{\beta}_1$	3	137	2	125	1	115	2	116	113	1.41
		$\hat{\beta}_2$	6	85	4	81	3	75	4	76	74	1.25
50%		$\hat{\beta}_1$	7	159	8	151	0	138	5	129	130	1.53
		$\hat{\beta}_2$	6	102	8	95	-1	89	4	87	85	1.40
Scenario III: $\lambda_0(t) = 5(t-2)^2$												
0%		$\hat{\beta}_1$	5	111	3	104	3	104	5	106	106	1.09
		$\hat{\beta}_2$	8	70	7	65	7	65	8	67	65	1.09
30%		$\hat{\beta}_1$	10	135	7	131	7	124	9	127	124	1.12
		$\hat{\beta}_2$	7	83	5	81	7	78	7	80	76	1.08
50%		$\hat{\beta}_1$	9	156	10	157	7	144	9	146	143	1.16
		$\hat{\beta}_2$	7	94	7	93	6	88	7	89	88	1.13

NOTE: $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated regression coefficients, where the true parameter values are (1, 1). PL, the maximum partial likelihood estimator; QS, the estimator studied by Qin and Shen (2010); PPL, the maximum pseudo-partial likelihood estimator studied by Tsai (2009); CPL, the proposed maximum composite partial estimator; Bias and ES are the empirical bias ($\times 1000$) and empirical standard

deviation ($\times 1000$) of 2000 regression parameter estimates; ASE is the averaged robust standard error estimate; RE is the empirical variance of the maximum partial likelihood estimator divided by that of the maximum composite partial likelihood estimator.

Table 2

Estimated regression coefficients of the Cox model for the CSHA study

Method	β_1 , Probable Alzheimer			β_2 , Vascular dementia		
	Coef	SE	95% CI	Estimate	SE	95% CI
PL	0.030	0.087	(-0.149, 0.195)	0.113	0.109	(-0.108, 0.314)
QS	0.159	0.093	(-0.023, 0.348)	0.260	0.119	(0.044, 0.522)
PPL	0.064	0.082	(-0.093, 0.223)	0.161	0.106	(-0.035, 0.365)
CPL	0.117	0.068	(-0.011, 0.250)	0.214	0.087	(0.052, 0.381)

NOTE: PL, the maximum partial likelihood estimator; QS, the estimator studied by Qin and Shen (2010); PPL, the maximum pseudo-partial likelihood estimator studied by Tsai (2009); CPL, the proposed maximum composite partial estimator; Coef and SE are the estimated coefficient and the empirical standard deviation of 2000 regression parameter estimates; 95% CI is the 95% bootstrap confidence interval given by the 2.5th and 97.5th percentiles of the 2000 estimates.