

# Likelihood approaches for the invariant density ratio model with biased-sampling data

BY YU SHEN, JING NING

*Department of Biostatistics, MD Anderson Cancer Center, The University of Texas, Houston, Texas 77030, U.S.A.*

yshen@mdanderson.org    jning@mdanderson.org

AND JING QIN

*Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892, U.S.A.*

jingqin@niaid.nih.gov

## SUMMARY

The full likelihood approach in statistical analysis is regarded as the most efficient means for estimation and inference. For complex length-biased failure time data, computational algorithms and theoretical properties are not readily available, especially when a likelihood function involves infinite-dimensional parameters. Relying on the invariance property of length-biased failure time data under the semiparametric density ratio model, we present two likelihood approaches for the estimation and assessment of the difference between two survival distributions. The most efficient maximum likelihood estimators are obtained by the EM algorithm and profile likelihood. We also provide a simple numerical method for estimation and inference based on conditional likelihood, which can be generalized to  $k$ -arm settings. Unlike conventional survival data, the mean of the population failure times can be consistently estimated given right-censored length-biased data under mild regularity conditions. To check the semiparametric density ratio model assumption, we use a test statistic based on the area between two survival distributions. Simulation studies confirm that the full likelihood estimators are more efficient than the conditional likelihood estimators. We analyse an epidemiological study to illustrate the proposed methods.

*Some key words:* Conditional likelihood; Density ratio model; EM algorithm; Length-biased sampling; Maximum likelihood approach.

## 1. INTRODUCTION

Length-biased sampling has been recognized in the statistics and scientific literature for decades (Zelen, 2004; Scheike & Keiding, 2006). An epidemiological study serves as an illustration: in a prevalent cohort study, individuals age 65 and older were recruited and screened for dementia (Wolfson et al., 2001). Individuals who were confirmed to have dementia were prospectively followed to assess their time to death or censoring. Individuals who survived until the time of recruitment were sampled into the study, whereas those who died before recruitment were not included. This resulted in selection bias, because the observed time intervals from the onset of dementia to death were longer for individuals in the study cohort compared with those dementia patients in the general population. With three different dementia diagnoses, it is of interest to compare the survival distributions from different groups of patients.

For conventional survival data, the partial likelihood is rank invariant with a binary covariate under the proportional hazards model (Peto & Peto, 1972). However, when right-censored survival data are subject to length-biased sampling, the invariance is lost because the size of the failure time also plays a key role in the inference. The proportional hazards assumption does not hold in the hazard functions of the observed biased failure time data, even though the unbiased failure time satisfies the proportional hazards model. For uncensored length-biased data, Chen (2010) and Mandel & Ritov (2010) proposed estimation approaches that rely on the invariance of the covariate effects under the accelerated failure time model. For censored length-biased data, the accelerated failure time model does not enjoy invariance due to the presence of dependent right censoring.

The density ratio model, which relates the distributions of two arms under a flexible semi-parametric structure, has been explored for uncensored data and traditional right-censored data (Anderson, 1979; Qin et al., 2002; Shen et al., 2007), but to our knowledge has never been investigated for length-biased survival data. In fact, the density ratio model has the invariance property for length-biased data; this naturally facilitates the construction of likelihood functions.

Assume  $\tilde{T}$  to be the time interval measured from an initiating event to failure within the population of interest. Let  $f_u(t)$  and  $f(t)$  define the probability density function for unbiased time,  $\tilde{T}$  and the length-biased time,  $T$ , respectively, in the first group. Then

$$f(t) = \frac{t f_u(t)}{\mu_f}, \quad \mu_f = \int t f_u(t) dt. \quad (1)$$

Let  $F_u$  be the cumulative distribution function of  $\tilde{T}$ . For the second group, let  $g_u$  and  $G_u$  denote the unbiased density and the cumulative distribution functions. The corresponding  $f$  and  $g$  represent the density functions for length-biased samples, and  $F$  and  $G$  denote the distribution functions of  $T$  in each group. Under the density ratio model, the ratio of two density functions of  $\tilde{T}$  in the target population is assumed to have the form

$$g_u(t) = \exp\{\alpha_u + \beta\gamma(t)\} f_u(t), \quad (2)$$

where  $\gamma(t)$  is a known monotone function for transforming the observed time  $t$ . Different forms of  $\gamma(t)$  correspond to some conventionally used densities; the common forms of  $\gamma$  in applications include  $\gamma(t) = t$  or  $\gamma(t) = \log(t)$  (Anderson, 1979; Kay & Little, 1987; Qin et al., 2002). When  $\gamma(t) = t$ ,  $\beta$  is the log odds ratio for density of  $\tilde{T} = t + 1$  against density of  $\tilde{T} = t$  between the two groups. In this case, model (2) is also equivalent to a semiparametric model in which the log ratio of two density functions is linear in  $t$ , where  $\beta$  is the slope in the linear function. When  $\gamma(t) = \log(t)$  and  $0 < \beta \leq 1$ ,  $g_u(t)$  is proportional to its length of  $t$  with the power of  $\beta$ . When both density functions are exponential, they belong to the intersection of the density ratio models and proportional hazards models. For any form of  $\gamma(t)$ ,  $\beta = 0$  leads to the equality of the two density functions, because  $\alpha_u = \alpha_u(\beta)$ . A unique property of (2) is its invariance for length-biased  $T$ . The density ratio of biased samples can be associated with the same model structure, except for a different intercept,  $g(t) = \exp\{\alpha + \beta\gamma(t)\} f(t)$ , where  $\alpha = \alpha_u + \log(\mu_f) - \log(\mu_g)$  and  $\mu_g = \int t g_u(t) dt$ .

In the present paper, we propose a conditional estimating equation approach to estimating the nonparametric baseline density function  $f_u(t)$  and parameters  $(\alpha_u, \beta)$  for length-biased data under this model. We also develop a full likelihood approach under the density ratio model. We obtain the maximum likelihood estimators by extending Vardi's EM algorithm for nonparametric estimation under the renewal process together with the profile likelihood method. We assess the adequacy of the density ratio model using a test based on the area between two survival curves.

## 2. CONDITIONAL LIKELIHOOD METHOD

Suppose that  $T = A + V$  is the observed failure time and  $U = A + C$  is the censoring time, where  $V$  is the time from examination to failure, and  $C$  is the time from examination to censoring. Here,  $A$  is called the truncation variable and  $V$  is the residual survival time. Assume that residual censoring time  $C$  and  $(A, V)$  are independent. The joint density distributions of  $(A, V)$  and  $(A, T)$  have the same formula when there is no censoring,

$$f_{A,V}(a, v) = \frac{f_u(t)}{\mu_f} \quad (t = a + v, t > a > 0). \quad (3)$$

Given (3), the probability of observing a biased failure time is

$$\text{pr}(A = a, T = t, U > t) = \text{pr}(A = a, V = t - a)\text{pr}(C > t - a).$$

The conditional density function of observing uncensored  $T$  can be expressed as

$$f_1(t) = \frac{\lim_{\Delta \rightarrow 0^+} \text{pr}(t \leq T < t + \Delta, U > t \mid Z = 0)/\Delta}{\text{pr}(T < U \mid Z = 0)} = \frac{f_u(t)w_f(t)}{\mu_f\lambda_f}, \quad (4)$$

where  $Z = 0$  or  $1$  is the group indicator,  $w_f(t) = \int_0^t S_c(u \mid Z = 0) du$ ,  $S_c(\cdot \mid Z)$  is the survival distribution of  $C$  given  $Z$ ,  $\lambda_f = \text{pr}(T < U \mid Z = 0)$  and

$$g_1(t) = \frac{\lim_{\Delta \rightarrow 0^+} \text{pr}(t \leq T < t + \Delta, U > t \mid Z = 1)/\Delta}{\text{pr}(T < U \mid Z = 1)} = \frac{g_u(t)w_g(t)}{\mu_g\lambda_g}, \quad (5)$$

where  $w_g(t) = \int_0^t S_c(u \mid Z = 1) du$  and  $\lambda_g = \text{pr}(T < U \mid Z = 1)$ . Under the density ratio model assumption in (2), the ratio of the two conditional density functions retains a model framework similar to that of the marginal density function for the unbiased observation  $\tilde{T}$ ,

$$\frac{g_1(t)}{f_1(t)} = \frac{g_u(t)\mu_f w_g(t)\lambda_f}{f_u(t)\mu_g w_f(t)\lambda_g} = \exp\{\alpha_1 + \beta\gamma(t) + w(t)\},$$

where  $\alpha_1 = \alpha_u + \log(\lambda_f\mu_f) - \log(\lambda_g\mu_g)$  and  $w(t) = \log\{w_g(t)\} - \log\{w_f(t)\}$ . When the censoring distribution is the same for two arms, there is a further simplification with  $w(t) = 0$ .

Let  $Y_i = \min(T_i, U_i)$ ,  $A_i$ ,  $\delta_i = I(T_i \leq U_i)$ ,  $Z_i$  ( $i = 1, \dots, n$ ) be the observed data, equivalently denoted by  $(Y_i, A_i, \delta_i, Z_i)$  for the  $i$ th subject. Conditional on the observed failure times:  $\{s_1, \dots, s_{m_1}, \dots, s_m\}$ ,  $m = m_1 + m_2$  where  $m_1$  is the number of failures among  $n_1$  subjects in arm 1, and  $m_2$  is the number of failures among  $n_2$  subjects in arm 2, and the likelihood for  $F_1$  and  $(\alpha_1, \beta)$  is proportional to

$$L_c(\alpha_1, \beta, F_1) = \prod_{i=1}^m p_{1i} \prod_{j=m_1+1}^m \exp\{\alpha_1 + \beta\gamma(s_j) + w(s_j)\},$$

where  $p_{1i} = dF_1(s_i)$  and  $F_1(t) = \sum_{s_i \leq t} p_{1i}$ . The corresponding loglikelihood function is

$$\ell_c(\alpha_1, \beta, F_1) = \sum_{i=1}^m \log(p_{1i}) + \sum_{j=m_1+1}^m \{\alpha_1 + \beta\gamma(s_j) + w(s_j)\}, \quad (6)$$

subject to the constraints  $p_{1i} \geq 0 (i = 1, \dots, m)$ ,  $\sum_{i=1}^m p_{1i} = 1$  and

$$\sum_{i=1}^m p_{1i} [\exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\} - 1] = 0,$$

where the second constraint comes from the fact that  $\int g_1(t) dt = 1$ . After profiling  $p_{1i}$ , we have the logistic type of loglikelihood (Qin & Zhang, 1997),

$$\ell_c(\alpha_1, \beta) = \sum_{j=m_1+1}^m \{\alpha_1 + \beta\gamma(s_j) + w(s_j)\} - \sum_{i=1}^m \log[1 + \rho \exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\}],$$

where  $\rho = m_2/m_1$ . If  $w(t)$  is known, the score equations for  $(\alpha_1, \beta)$  are

$$\frac{\partial \ell_c}{\partial \alpha_1} = m_2 - \sum_{i=1}^m \frac{\rho \exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\}}{1 + \rho \exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\}}, \tag{7}$$

$$\frac{\partial \ell_c}{\partial \beta} = \sum_{i=m_1+1}^m \gamma(s_i) - \sum_{i=1}^m \frac{\rho \gamma(s_i) \exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\}}{1 + \rho \exp\{\alpha_1 + \beta\gamma(s_i) + w(s_i)\}}. \tag{8}$$

The expectations for the right side of equations (7) and (8) are zero under model (2), so the estimating equation from the score function is unbiased. If  $f_1$  and  $g_1$  have finite mean, finite and nonzero variance,  $0 < E_{f_1}\{\gamma(t)\}^2 < \infty$  and  $0 < E_{g_1}\{\gamma(t)\}^2 < \infty$ , we can derive the existence of  $E_{f_1}\{m(t)\}^2$  and  $E_{g_1}\{m(t)\}^2$ , where  $m(t) = \gamma(t) \exp\{\alpha_1 + \beta\gamma(t) + w(t)\} / [1 + \rho \exp\{\alpha_1 + \beta\gamma(t) + w(t)\}]$ . Under the above regularity conditions, the derivatives of (7) and (8) with respect to  $(\alpha_1, \beta)$  yield a positive definite matrix, so the solution is unique and leads to consistent estimators of  $(\alpha_1, \beta)$ .

When  $w(t)$  is unknown, we replace it by its consistent estimator  $\tilde{w}(t) = \log\{\tilde{w}_g(t)\} - \log\{\tilde{w}_f(t)\}$ , where  $\tilde{w}_f$  and  $\tilde{w}_g$  are the area under the Kaplan–Meier estimators of the residual censoring time of each arm. We obtain the estimators,  $(\tilde{\alpha}_1, \tilde{\beta})$  by solving equations (7) and (8). Subsequently, we can estimate  $F_1(t)$  by  $\tilde{F}_1(t) = \sum_{i:s_i \leq t} \tilde{p}_{1i}$ , from

$$\tilde{p}_{1i} = m_1^{-1} [1 + \rho \exp\{\tilde{\alpha}_1 + \tilde{\beta}\gamma(s_i) + \tilde{w}(s_i)\}]^{-1},$$

and estimate  $G_1(t) = \int_0^t g_1(v) dv$  by  $\tilde{G}_1(t) = \sum_{i:s_i \leq t} \tilde{q}_{1i}$ , where

$$\tilde{q}_{1i} = \exp\{\tilde{\alpha}_1 + \tilde{\beta}\gamma(s_i) + \tilde{w}(s_i)\} \tilde{p}_{1i}.$$

The quantities,  $\lambda_f$  and  $\lambda_g$  can be consistently estimated by  $\tilde{\lambda}_f = \sum_{i=1}^n I(\delta_i = 1, Z_i = 0) / n_1$  and  $\tilde{\lambda}_g = \sum_{i=1}^n I(\delta_i = 1, Z_i = 1) / n_2$ . Because  $f_u$  and  $g_u$  are proper density functions with  $\int f_u(t) dt = 1$  and  $\int g_u(t) dt = 1$ , consistent estimators of  $\mu_f$  and  $\mu_g$  can be derived from (4) and (5) as

$$\tilde{\mu}_f = \frac{1}{\tilde{\lambda}_f} \left\{ \int \frac{d\tilde{F}_1(t)}{\tilde{w}_f(t)} \right\}^{-1}, \quad \tilde{\mu}_g = \frac{1}{\tilde{\lambda}_g} \left\{ \int \frac{d\tilde{G}_1(t)}{\tilde{w}_g(t)} \right\}^{-1}.$$

Provided there is a positive follow-up time  $t$ ,  $w_f(t)$  and  $w_g(t)$  can be consistently estimated by  $\tilde{w}_f(t)$  and  $\tilde{w}_g(t)$ , which are bounded away from zero. Given these consistent estimators, the

consistent estimators of  $F_u(t)$  and  $G_u(t)$  can subsequently be obtained through (4) and (5),

$$\tilde{F}_u(t) = \sum_{i:s_i \leq t} \frac{\tilde{\mu}_f \tilde{\lambda}_f}{\tilde{w}_f(s_i)} \tilde{p}_{1i}, \quad \tilde{G}_u(t) = \sum_{i:s_i \leq t} \frac{\tilde{\mu}_g \tilde{\lambda}_g}{\tilde{w}_g(s_i)} \tilde{q}_{1i}.$$

Our motivating example involves three dementia diagnoses in the study cohort; a comparison of the survival distributions among the three types is of scientific interest. We generalize the proposed two-arm estimation and inference procedures to those for  $k$ -arm studies where  $k > 2$ . Details for the global  $k$ -arm estimation and inference methods can be found in the Supplementary Material.

### 3. FULL LIKELIHOOD APPROACH

#### 3.1. Maximum likelihood estimate

Conditioning on the observed failure times, the proposed estimation procedure is straightforward and can be implemented easily, as shown in § 2. However, conditional inference is often less efficient than inference based on the full likelihood (Wang, 1991; Zeng & Lin, 2007). We formulate the full likelihood function under the density ratio model by generalizing the nonparametric maximum likelihood estimation method of Vardi (1989).

From (3), the full likelihood for the observed data can be formally expressed as

$$\prod_{i=1}^n \left\{ \frac{dF_u(Y_i)}{\mu_f} \right\}^{\delta_i(1-Z_i)} \left\{ \frac{\bar{F}_u(Y_i)}{\mu_f} \right\}^{(1-\delta_i)(1-Z_i)} \left\{ \frac{dG_u(Y_i)}{\mu_g} \right\}^{\delta_i Z_i} \left\{ \frac{\bar{G}_u(Y_i)}{\mu_g} \right\}^{(1-\delta_i)Z_i},$$

where  $\bar{F}_u$  and  $\bar{G}_u$  are the survivor functions of  $\tilde{T}$  for the two arms. Because of the relationship between the probability density function of the unbiased time and that of the biased time given in (1),

$$\begin{aligned} \frac{Y_i dF_u(Y_i)}{\mu_f} &= dF(Y_i), & \frac{\bar{F}_u(Y_i)}{\mu_f} &= \int_{Y_i}^{\infty} \frac{1}{t} dF(t), \\ \frac{Y_i dG_u(Y_i)}{\mu_g} &= dG(Y_i), & \frac{\bar{G}_u(Y_i)}{\mu_g} &= \int_{Y_i}^{\infty} \frac{1}{t} dG(t). \end{aligned}$$

Under the density ratio model,  $dG(t) = \exp\{\alpha + \beta\gamma(t)\} dF(t)$  and removing the constant  $Y_i$ , the full likelihood of the observed data is proportional to

$$\begin{aligned} L_f(\alpha, \beta, F) &= \prod_{i=1}^n \{dF(Y_i)\}^{\delta_i(1-Z_i)} \left\{ \int_{Y_i}^{\infty} u^{-1} dF(u) \right\}^{(1-\delta_i)(1-Z_i)} \\ &\times [dF(Y_i) \exp\{\alpha + \beta\gamma(Y_i)\}]^{\delta_i Z_i} \left[ \int_{Y_i}^{\infty} u^{-1} \exp\{\alpha + \beta\gamma(u)\} dF(u) \right]^{(1-\delta_i)Z_i}. \quad (9) \end{aligned}$$

To find the nonparametric maximum likelihood estimator for  $F$  and the unknown parameters  $(\alpha, \beta)$  from (9), it is sufficient to maximize the discrete version of  $F$  on the point mass support of  $0 < t_1 < \dots < t_h$ , where  $\{t_1, \dots, t_h\}$  denote the uniquely observed failure and censoring times from both arms in increasing order (Vardi, 1989). Let  $p_i = dF(t_i)$  ( $i = 1, \dots, h$ ) and  $q_i = dG(t_i) = p_i \exp\{\alpha + \beta\gamma(t_i)\}$ .

The discrete version of the full likelihood can be expressed as

$$L_f(\alpha, \beta, F) = \prod_{j=1}^h p_j^{\xi_j^f} \left( \sum_{k=j}^h t_k^{-1} p_k \right)^{\zeta_j^f} \prod_{j=1}^h q_j^{\xi_j^g} \left( \sum_{k=j}^h t_k^{-1} q_k \right)^{\zeta_j^g}, \tag{10}$$

where

$$\begin{aligned} \xi_j^f &= \sum_{i=1}^n \delta_i I(Y_i = t_j, Z_i = 0), & \zeta_j^f &= \sum_{i=1}^n (1 - \delta_i) I(Y_i = t_j, Z_i = 0), \\ \xi_j^g &= \sum_{i=1}^n \delta_i I(Y_i = t_j, Z_i = 1), & \zeta_j^g &= \sum_{i=1}^n (1 - \delta_i) I(Y_i = t_j, Z_i = 1) \end{aligned}$$

are the multiplicities of the failure times and censoring times at  $t_j$  in each arm. Thus, maximizing (9) reduces to maximizing (10) with respect to  $p_j (j = 1, \dots, h)$ ,  $\beta$  and  $\alpha$  subject to the constraints

$$p_j \geq 0, \quad \sum_{j=1}^h p_j = 1, \quad \sum_{j=1}^h p_j [\exp\{\alpha + \beta\gamma(t_j)\} - 1] = 0 \quad (j = 1, \dots, h). \tag{11}$$

When  $\alpha = \beta = 0$ , the likelihood (10) is equivalent to the nonparametric likelihood for length-biased data proposed by Vardi (1989).

### 3.2. Implementation of the EM algorithm

Vardi (1989) showed a natural connection between the nonparametric likelihood of length-biased data and the likelihood under multiplicative censorship. For a right-censored observation,  $U$  can be considered as incomplete failure time data in an EM algorithm and equivalently expressed as  $U = \phi T$ ,  $\phi \sim \text{Un}(0, 1)$ ,  $T \sim dF(t)$ , where  $T$  and  $\psi$  are independent and  $T$  is the biased failure time for a subject with  $Z = 0$ . The joint density of  $(U, T)$  and the marginal density of  $U$  are, respectively,

$$(U, T) \sim \frac{1}{t} dF(t) \quad (t \geq u), \quad U \sim \int_u^\infty \frac{1}{t} dF(t).$$

The conditional density function of  $T$  given right-censored time  $U$  can thus be expressed as

$$f(t | U = u) = \frac{1}{t} dF(t) \left\{ \int_u^\infty \frac{1}{t} dF(t) \right\}^{-1} \quad (t \geq u). \tag{12}$$

For subjects in the treatment group with  $Z = 1$ , similar equations can be derived given  $T \sim dG(t)$ . Applying the EM algorithm in Vardi (1989), we can estimate the nonparametric density function  $F$  and parameters  $(\alpha, \beta)$  from the likelihood function (10).

Denote the observed data by  $\mathcal{O} = \{Y_i, \delta_i, Z_i : i = 1, \dots, n\}$ . A key step in the EM algorithm is to replace the current estimated probability mass at  $t_j$  by its expected and normalized conditional multiplicity of the complete data observation at  $t_j$  using (12). At the E-step, we start with arbitrary  $p_i^0 (i = 1, \dots, h)$ ,  $\alpha^0$  and  $\beta^0$  satisfying (11). Since not all  $T_i$  are observable, we thus replace

those unobserved  $I(T_i = t_j)$  by their expectations. The expectation of  $\sum_{i=1}^n I(T_i = t_j)$  in each arm given the observed data and current parameter estimators is

$$\begin{aligned} w_j^f &= E_f \left\{ \sum_{i=1}^n \delta_i I(T_i = t_j, Z_i = 0) + \sum_{i=1}^n (1 - \delta_i) I(T_i = t_j, Z_i = 0) \middle| \mathcal{O} \right\} \\ &= \xi_j^f + t_j^{-1} p_j^0 \sum_{k=1}^j \zeta_k^f \left( \sum_{i=k}^h t_i^{-1} p_i^0 \right)^{-1}, \\ w_j^g &= E_g \left\{ \sum_{i=1}^n \delta_i I(T_i = t_j, Z_i = 1) + \sum_{i=1}^n (1 - \delta_i) I(T_i = t_j, Z_i = 1) \middle| \mathcal{O} \right\} \\ &= \xi_j^g + t_j^{-1} q_j^0 \sum_{k=1}^j \zeta_k^g \left( \sum_{i=k}^h t_i^{-1} q_i^0 \right)^{-1}, \end{aligned}$$

where  $q_i^0 = p_i^0 \exp\{\alpha^0 + \beta^0 \gamma(t_i)\}$ . The conditional expectation of the complete data loglikelihood is

$$\ell = \sum_{j=1}^h w_j^f \log p_j + \sum_{j=1}^h w_j^g \log q_j = \sum_{j=1}^h w_j \log p_j + \sum_{j=1}^h w_j^g \{\alpha + \beta \gamma(t_j)\}, \quad (13)$$

where  $w_j = w_j^f + w_j^g$ . Define the sum of the  $w_j$  as  $w^+ = \sum_{j=1}^h w_j$ .

At the M-step, we maximize the loglikelihood (13) under the constraints specified in (11) by including the Lagrange multipliers in the weighted likelihood (Owen, 1988) as follows:

$$\begin{aligned} \ell_L &= \sum_{j=1}^h w_j \log p_j + \sum_{j=1}^h w_j^g \{\alpha + \beta \gamma(t_j)\} + \lambda_1 \left( 1 - \sum_{j=1}^h p_j \right) \\ &\quad - w^+ \lambda_2 \sum_{j=1}^h p_j [\exp\{\alpha + \beta \gamma(t_j)\} - 1]. \end{aligned}$$

Taking the derivative of  $\ell_L$  with respect to  $p_j$ , we have

$$w_j p_j^{-1} - \lambda_1 - w^+ \lambda_2 [\exp\{\alpha + \beta \gamma(t_j)\} - 1] = 0,$$

together with the constraints (11), leading to  $\lambda_1 = w^+$ . Consequently,

$$p_j = \frac{w_j}{w^+ (1 + \lambda_2 [\exp\{\alpha + \beta \gamma(t_j)\} - 1])}. \quad (14)$$

After profiling over the  $p_j$ s in (13) and ignoring the constant terms, we have

$$\ell_L(\alpha, \beta) = - \sum_{j=1}^h w_j \log(1 + \lambda_2 [\exp\{\alpha + \beta \gamma(t_j)\} - 1]) + \sum_{j=1}^h w_j^g \{\alpha + \beta \gamma(t_j)\}. \quad (15)$$

Taking the derivative of  $\ell_L$  with respect to  $\alpha$ ,

$$\sum_{j=1}^h w_j^g - \lambda_2 \sum_{j=1}^h \frac{w_j \exp\{\alpha + \beta\gamma(t_j)\}}{1 + \lambda_2[\exp\{\alpha + \beta\gamma(t_j)\} - 1]} = 0,$$

and together with (14), we obtain  $\lambda_2 = \sum_{j=1}^h w_j^g/w^+$ . After inserting the Lagrange multipliers into (15), the loglikelihood function with respect to  $\alpha$  and  $\beta$  under the constraints is

$$\ell_L(\alpha, \beta) = \sum_{j=1}^h w_j^g \{\alpha + \beta\gamma(t_j)\} - \sum_{j=1}^h w_j \log[1 - \exp(c) + \exp\{\alpha + c + \beta\gamma(t_j)\}], \quad (16)$$

where  $c = \log(\sum_{j=1}^h w_j^g/w^+)$ . We can estimate  $\alpha, \beta$  from (16) and then derive  $p_j$  from (14). Given the estimated  $\alpha, \beta$  and  $p_j, w_j^f$  and  $w_j^g$  are updated by repeating the E-step. The iteration is repeated until a prespecified criterion is satisfied. The maximum likelihood estimate is denoted by  $\hat{\theta}_n = (\hat{\alpha}, \hat{\beta}, \hat{F})$ , where  $\hat{F}(t) = \sum_{t_j \leq t} \hat{p}_j$ . The corresponding unbiased distribution function is estimated by  $\hat{F}_u(t) = \int_0^t x^{-1} d\hat{F}(x) / \int_0^\infty x^{-1} d\hat{F}(x)$ .

### 3.3. Large sample properties

Let  $\theta_0 \equiv (\alpha_0, \beta_0, F_0)$  denote the true values of  $\theta \equiv (\alpha, \beta, F)$ . Under the regularity conditions specified in the Appendix, we can establish the asymptotic properties of the estimated maximum likelihood estimator  $\hat{\theta}_n$  for  $t \in [0, \tau]$ , where  $\tau$  is a finite upper bound of the support for the population survival times.

**THEOREM 1.** *Under Conditions A1–A4, the maximum likelihood estimator  $\hat{\theta}_n$  is consistent, defined by*

$$|\hat{\alpha} - \alpha_0| + |\hat{\beta} - \beta_0| + \sup_{t \in [0, \tau]} |\hat{F}(t) - F_0(t)|$$

converging to 0 almost surely and uniformly as  $n \rightarrow \infty$ .

**THEOREM 2.** *Under Conditions A1–A4,  $n^{1/2}(\hat{\theta}_n - \theta_0)$  converges weakly to a tight zero-mean Gaussian process  $-\phi'_{\psi_0}\{\dot{U}_0^{-1}(\mathbb{W})\}$  as  $n \rightarrow \infty$ .*

Definitions of  $\phi'_{\psi_0}\{\dot{U}_0^{-1}(\mathbb{W})\}$  and  $\mathbb{W}$  are provided in the Appendix. The consistency of  $\hat{\theta}_n$  is proved by the Kullback–Leibler information approach (Murphy, 1994; Parner, 1998). The weak convergence of  $\hat{\theta}_n$  is established by applying the Z-theorem for the infinite-dimensional estimating equations (van der Vaart & Wellner, 1996). We present the proof of Theorem 1 and an outline proof for Theorem 2 in the Appendix. Although the variance matrix is derived from the weak convergence of  $\hat{\theta}_n$ , its computation is complex. Alternatively, we use the simple bootstrap resampling method to compute the variances for the estimated parameters (Bilker & Wang, 1996; Gross & Lai, 1996; Asgharian et al., 2002). Given the established weak convergence for the proposed estimators in Theorem 2 under the regularity conditions, this bootstrap is valid for estimating the variances of the estimators (van der Vaart & Wellner, 1996, Ch. 3.6).



## 4. INFERENCE

## 4.1. Hypothesis testing

To test the difference in the survival distributions, we describe likelihood ratio tests under the conditional likelihood framework, when the censoring distributions in the two arms are the same,  $w(t) = 0$ . For the conditional probability density function,  $f_1(t) = g_1(t)$  is equivalent to  $f_u(t) = g_u(t)$ , when  $w(t) = 0$ . If  $f_1(t) = g_1(t)$  and  $w_f(t) = w_g(t)$ , then  $\lambda_f = \lambda_g$  because both  $f(t)$  and  $g(t)$  are proper density functions. Thus,  $f(t) = g(t)$ , and  $f_u(t) = g_u(t)$ . We therefore use the conditional likelihood ratio statistic to test the equality  $f_1(t) = g_1(t)$ , i.e.,  $H_0 : \alpha_1 = \beta = 0$ . Let  $\ell_c$  denote the loglikelihood of (6). Under the null hypothesis  $H_0$ ,  $\ell_c$  reduces to  $\ell_c(0, 0, \hat{F}_1)$ , where  $\hat{F}_1$  is the nonparametric estimator of  $F_1$  under  $\alpha_1 = \beta = 0$ . The asymptotic distribution of  $R_c = 2\{\ell_c(\tilde{\alpha}_1, \tilde{\beta}, \tilde{F}_1) - \ell_c(0, 0, \hat{F}_1)\}$  is  $\chi_1^2$  because  $\beta = 0$  implies  $\alpha_1 = 0$ . When the censoring distributions of the two arms are not equal,  $w(t) \neq 0$ , the asymptotic distribution of  $R_c$  for testing  $\beta = 0$  will not be chi-squared in general, but we can obtain a critical value for the test statistic using the bootstrap. Intuitively, the full likelihood ratio statistic  $R_f$ , which is defined similarly to  $R_c$ , may be used to test  $H_0$ . Our simulation results showed that this test statistic can be well approximated by a  $\chi_1^2$  distribution when  $w(t) = 0$ , though we cannot prove this property theoretically.

## 4.2. Model checking

To assess the parametric component for the link function  $\gamma(t)$  imposed on the ratio of the two density functions in model (2), we use a general model checking method similar to that in Shen et al. (2007). The test statistic is based on a quadratic form of the cumulative difference between the nonparametric estimate of the distribution function and its estimate under the density ratio model assumption. The goodness-of-fit test is

$$A_n = \int_0^\tau \{\hat{F}_u(t) - \hat{F}_v(t)\}^2 dt, \quad (17)$$

where  $\hat{F}_u$  is the estimated distribution function of  $\tilde{T}$  using the full likelihood under the density ratio model assumption, and  $\hat{F}_v$  is Vardi's nonparametric estimate using data from arm 1 only. Similarly, we can replace  $\hat{F}_u(t)$  by  $\tilde{F}_u(t)$  in (17) to assess the goodness-of-fit for the conditional likelihood estimator. Because the derivation of the asymptotic distribution of  $A_n$  is tedious and not the focus of this paper, we use the bootstrap to find its critical values.

## 5. NUMERICAL STUDIES AND APPLICATIONS

## 5.1. Simulations

We compare the performance of the estimation and inference of the full likelihood and the conditional likelihood, under various scenarios. We use a Weibull distribution for  $f_u$ , so that  $f$  is also Weibull. The density  $g_u$  is  $\exp\{\alpha_u + \beta\gamma(t)\}f_u(t)$  for  $Z = 1$ , where  $\gamma(t) = t$ .

For  $Z = 0$  or 1, we generate independent pairs  $(\tilde{A}, \tilde{T})$ , then keep those with  $\tilde{A} < \tilde{T}$ , where  $\tilde{A}$  is generated independently from a uniform distribution on  $(0, \tau_0)$  and  $\tilde{T}$  is generated from the density  $f_u$  or  $g_u$  depending on  $Z$ , where  $\tau_0$  is the upper bound of  $\tilde{T}$ . The censoring times  $C$  are generated independently from uniform distributions corresponding to censoring percentages ranging from 15 to 43%. The censoring indicator is  $\delta = I(T \leq C + A)$ . For each scenario, we use sample sizes  $n_1 = n_2 = 50$  or 100, and repeat the simulation 2000 times.

Table 1. *Estimated standard deviations of the parameter estimators under the density ratio model, size and power of the likelihood ratio tests under the nominal level of 1 and 5%*

Cens.	Full likelihood				Conditional likelihood			
	Standard deviation		$R_f : \chi^2$ -test		Standard deviation		$R_c : \chi^2$ -test	
	$\hat{\alpha}_u$	$\hat{\beta}$	1%	5%	$\tilde{\alpha}_u$	$\tilde{\beta}$	1%	5%
	$\alpha_u = \beta = 0, n_0 = n_1 = 50$							
15%	0.17	0.17	1.4	6.5	0.20	0.19	1.1	6.3
25%	0.18	0.17	1.2	5.8	0.22	0.21	0.8	5.6
43%	0.19	0.18	1.4	5.6	0.31	0.30	1.2	6.8
	$\alpha_u = \beta = 0, n_0 = n_1 = 100$							
15%	0.11	0.11	1.5	6.4	0.12	0.12	0.8	5.6
25%	0.11	0.11	1.3	6.0	0.14	0.14	1.0	4.8
43%	0.12	0.12	1.5	6.2	0.19	0.19	1.0	6.0
	$\alpha_u = -0.405, \beta = 0.333, n_0 = n_1 = 50$							
15%	0.21	0.15	58.6	79.6	0.23	0.17	41.1	66.1
25%	0.21	0.15	54.1	75.4	0.25	0.18	35.4	60.1
43%	0.23	0.16	38.6	59.6	0.33	0.25	20.0	41.8
	$\alpha_u = -0.405, \beta = 0.333, n_0 = n_1 = 100$							
15%	0.13	0.10	91.0	97.5	0.15	0.11	81.8	93.5
25%	0.15	0.11	87.2	94.8	0.16	0.12	68.0	87.6
43%	0.15	0.11	62.8	73.8	0.21	0.16	41.5	67.3

In Table 1, we summarize the standard deviation of the estimated parameters for both methods. We also provide the size and power of the conditional likelihood ratio statistics for testing the null hypothesis  $H_0 : \alpha_u = \beta = 0$ . The empirical bias slightly increases with increasing censoring for the conditional likelihood estimator. For both methods, the point estimators are fairly unbiased, and the biases are not significantly different from zero, so we do not present them. The standard error of each estimate is found using 500 bootstrap replicates. Compared with the conditional likelihood approach, the estimators from the full likelihood are more efficient; the standard deviations for the former approach are up to 1.65 times greater. The efficiency gain is more significant when censoring increases. The significance levels of the conditional likelihood ratio tests are maintained around the nominal levels. We also compare the estimated mean and the baseline cumulative distributions,  $\hat{F}_u(t)$  and  $\tilde{F}_u(t)$  at the 50 and 75% quantiles. The standard deviations of the full likelihood estimator are smaller than those of the conditional likelihood estimators when data have heavy censoring, but the difference in efficiency is otherwise small. For both methods, the estimation efficiency for  $(\beta, F)$  increases with the percentage of uncensored data. For the conditional likelihood, although  $w_f(t)$  and  $w_g(t)$  may not be estimated with great accuracy when most of data are uncensored,  $\hat{w}_f(t)$  and  $\hat{w}_g(t)$  are always robust and have relatively small impact on the estimation efficiency for  $(\beta, F)$ .

### 5.2. Example

In a multicentre prospective observational study conducted throughout Canada during 1991–1996, more than 14 000 individuals aged 65 years or older participated in a health survey (Wolfson et al., 2001). Among them, 1132 participants who were identified to have dementia were followed until their death or censoring in 1996. We have complete data on a total of 818 individuals, consisting of the date of dementia onset, date of screening for dementia, date of death or censoring and the dementia diagnosis. Three categories of dementia diagnosis were identified among these patients: probable Alzheimer’s disease, with  $n = 393$ , possible Alzheimer’s disease,

Table 2. Parameter estimators, standard deviations and  $p$ -values from the conditional likelihood ratio test under the density ratio model for the dementia study

$\gamma(t)$	$\tilde{\alpha}_{u1}$ (sd)	$\tilde{\beta}_1$ (sd)	$\tilde{\alpha}_{u2}$ (sd)	$\tilde{\beta}_2$ (sd)	$p$ -value, $R_c$
$t$	0.16 (0.11)	-0.03 (0.02)	0.28 (0.16)	-0.05 (0.03)	0.11
$\log(t)$	0.22 (0.21)	-0.15 (0.15)	0.54 (0.24)	-0.42 (0.18)	0.05

Baseline group is ‘Possible Alzheimer’s’;  $\alpha_{u1}$  and  $\beta_1$ , the coefficients for ‘Probable Alzheimer’s’ vs. ‘Possible Alzheimer’s’;  $\alpha_{u2}$  and  $\beta_2$ , the coefficients for ‘Vascular’ vs. ‘Possible Alzheimer’s’;  $p$ -value, the critical value of  $R_c$  is determined by  $\chi^2$  statistic with one degree of freedom; sd, standard deviation.

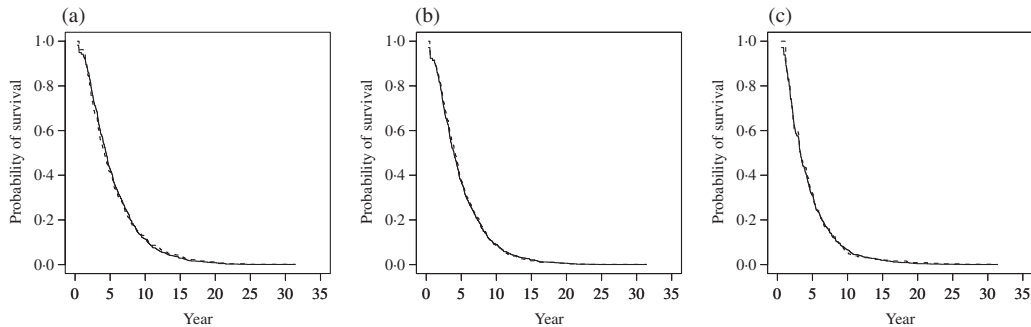


Fig. 1. Estimated survival curves using Vardi’s method (dotted line) and by the full likelihood approach under the density ratio model with  $\gamma(t) = \log(t)$  (solid line) for (a) possible Alzheimer’s disease, (b) probable Alzheimer’s disease and (c) vascular dementia.

with  $n = 252$  and vascular dementia, with  $n = 173$ . Several studies have confirmed the stationarity assumption, indicating that the dataset collected from this cohort of individuals represents a collection of length-biased samples (Asgharian et al., 2006). The distributions of the censoring time measured from the date of screening for dementia to censoring are essentially the same for the three diagnosis groups.

Applying the methods proposed for an overall analysis, we use the group diagnosed with possible Alzheimer’s disease as the baseline cohort with a link function of  $\gamma(t) = t$  or  $\log(t)$ . The corresponding parameter estimates and their standard errors are given in Table 2. We test the hypothesis that all three diagnoses of dementia are equidistributed. Under the log-transformation  $\gamma(t) = \log(t)$ , the conditional likelihood ratio test indicates a marginal significant difference in the survival distributions,  $p$ -value = 0.05. The model-based survival distribution estimates for the three groups with  $\gamma(t) = \log(t)$  are provided in the Supplementary Material.

Using the test statistic described in § 4.2, we check the fit of the density ratio model with  $\gamma(t) = t$  and  $\gamma(t) = \log(t)$ . The tests indicate that both link functions fit the data well; the  $p$ -values from the goodness-of-fit test range from 0.59 to 0.92. Comparing the density ratio model-based survival curve for each arm against Vardi’s nonparametric survival estimate, we find that the model with the link function  $\gamma(t) = \log(t)$  leads to a slightly better fit in all three pairwise comparisons. As shown in Fig. 1, the survival distribution estimate using the full likelihood approach under the density ratio model with  $\gamma(t) = \log(t)$  is almost identical to Vardi’s nonparametric estimate of the survival distribution for each disease diagnosis.

Under the density ratio model assumption, the estimated parameters for the three pairwise comparisons among the diagnosis groups are listed in Table 3. The estimates using the full likelihood and conditional likelihood methods are similar, whereas the standard deviations of the full likelihood estimators are smaller than those obtained from the conditional likelihood.

Table 3. *Pairwise models: parameter estimators, standard deviations and p-values from the likelihood ratio test statistics for full and conditional likelihood methods for the dementia study*

	Vascular vs. probable		Vascular vs. possible		Probable vs. possible	
	Full	Conditional	Full	Conditional	Full	Conditional
$\gamma(t) = t$						
$\hat{\alpha}_H(\text{sd})$	0.09 (0.11)	0.11 (0.15)	0.24 (0.10)	0.25 (0.14)	0.19 (0.09)	0.18 (0.13)
$\hat{\beta}(\text{sd})$	-0.02 (0.02)	-0.02 (0.03)	-0.05 (0.02)	-0.05 (0.03)	-0.04 (0.02)	-0.03 (0.02)
<i>p</i> -value	0.32	0.37	0.01	0.06	0.03	0.13
$\gamma(t) = \log(t)$						
$\hat{\alpha}_H(\text{sd})$	0.29 (0.18)	0.34 (0.21)	0.50 (0.20)	0.52 (0.24)	0.26 (0.17)	0.21 (0.20)
$\hat{\beta}(\text{sd})$	-0.24 (0.14)	-0.27 (0.17)	-0.39 (0.15)	-0.39 (0.17)	-0.19 (0.12)	-0.15 (0.14)
<i>p</i> -value	0.08	0.09	0.01	0.02	0.13	0.29

Possible, possible Alzheimer's disease; Probable, probable Alzheimer's disease; *p*-value, conditional likelihood ratio test  $R_c$  or full likelihood ratio test  $R_f$  for assessing the equality of two survival distributions; sd, standard deviation.

## 6. DISCUSSION

Other than applications in standard case-control studies, the two-sample density ratio model has been used for the estimation of malaria attributable fractions (Qin & Leung, 2005) and the analysis of genetic linkage studies (Qin & Zhang, 2005; Anderson, 1979). In this paper, by taking advantage of the invariance property, we have proposed two likelihood approaches for analysing length-biased failure data. We generalized Vardi's EM algorithm together with the profile likelihood method, so that the unknown parameters and baseline nonparametric distribution of  $T$  can be jointly estimated under the density ratio model. Inference based on the full likelihood method is more efficient than that based on the conditional likelihood, because the censored data are fully utilized in the EM algorithm. Moreover, the full maximum likelihood estimator of  $F$  is on the support of both failure and censoring times, whereas the conditional maximum likelihood estimator of  $F$  is on the failure times only. Finally, the full likelihood method is more robust to various censoring patterns, because it does not require the estimation of  $w(t)$ , which is a function of the residual censoring distribution. Although the full likelihood estimator has efficiency and robustness advantages over the conditional likelihood estimate, the conditional approach is appealing due to its straightforward computation under the density ratio model. Kalbfleisch (1978) recommended the conditional approach in a partially specified generalized linear model, of which the two-sample density ratio model is a special case.

There is a unique feature for length-biased data different from conventional survival data. The normalizing factor in the length-biased density function,  $\mu_f$ , can be estimated consistently if there is a positive follow-up time, regardless of how short this is. Because of symmetry between the truncation time  $A$  and the residual survival time  $V$  under length-biased sampling, information of  $V$  can be recovered from  $A$ , which is treated as free follow-up time. Such a data structure facilitates the estimability of the mean of the population failure times.

As suggested by one reviewer, it would be interesting to generalize the density ratio model to semiparametric regression models by taking covariates other than group indicators into consideration. One possibility is to model covariates in a parametric form, which may be incorporated into a general density ratio structure.

## ACKNOWLEDGEMENT

This work was supported in part by grants from the National Institutes of Health. We are grateful to three referees, an associate editor and the editor for helpful comments and suggestions.

We thank Professors Asgharian and Wolfson and the investigators from the Canadian Study of Health and Aging for providing us with the dementia data.

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes a proof of the estimability of the mean of  $\hat{T}$ , the inference method for  $k$ -arms, additional simulation results and a detailed proof of Theorem 2.

#### APPENDIX

We assume the following regularity conditions.

*Condition A1.* The true parameters  $(\xi_0, F_0)$  belong to  $\mathcal{B} \times \mathcal{F} = \{(\xi, F) : \xi \in \mathcal{B} \text{ and } F \text{ is a distribution function with continuous and differentiable density function}\}$ , where  $\xi = (\alpha, \beta)^\top$  and  $\mathcal{B}$  is a compact set in  $\mathbb{R}^2$ .

*Condition A2.* The known monotone function  $\gamma(t)$  is differentiable and bounded for  $t \in [0, \tau]$ .

*Condition A3.* The residual censoring time  $C$  has a continuous survival function and satisfies  $\text{pr}(C > V) > 0$  and  $E(C | Z) > 0$ .

*Condition A4.* The Fisher information  $-E\{\partial^2 \ell_f(\xi, \hat{F}) / \partial \xi^2\}$  evaluated at  $\xi_0$  is positive, where  $\ell_f(\xi, F)$  is defined in expression (A2).

Condition A1 ensures the smoothness of the underlying distribution functions. Condition A3 states that at least some subjects are uncensored at the end of the study. The condition on  $E(C | Z) > 0$  ensures a positive follow-up time in each arm, so that the upper bound for the finite support of failure time denoted by  $\tau$  can be consistently estimated by  $\hat{\tau} = \max_{1, \dots, n} \{Y_i\}$ . Condition A4 is a classical condition for semi-parametric models (Andersen et al., 1992, Condition VII2.1.(e), p. 497) and implies that the information matrix for  $\xi$  is positive definite when the cumulative distribution function is known.

*Proof of Theorem 1.* Denote the maximum likelihood estimator based on  $n$  observed samples by  $\hat{\theta}_n = (\hat{\alpha}, \hat{\beta}, \hat{F})$ . Estimation of the cumulative distribution function  $F$  is via a discretized version  $\hat{F}(t) = \sum_{t_i \leq t} \hat{p}_i$ . To obtain the maximum likelihood estimator of  $F$  subject to the constraint  $\sum_{j=1}^h p_j = 1$  and  $\sum_{j=1}^h p_j [\exp\{\alpha + \beta\gamma(t_j)\} - 1] = 0$ , we consider the unconstrained loglikelihood function  $\ell_f(\alpha, \beta, F, \lambda_3, \lambda_4)$  by incorporating the Lagrange multipliers  $\lambda_3$  and  $\lambda_4$ ,

$$\sum_{i=1}^n \left( \sum_{j=1}^h \delta_i I(Y_i = t_j) [\log(p_j) + Z_i \{\alpha + \beta\gamma(t_j)\}] + (1 - \delta_i) \log \left[ \sum_{t_j \geq Y_i} t_j^{-1} e^{Z_i \{\alpha + \beta\gamma(t_j)\}} p_j \right] \right) - \lambda_3 \left( \sum_{j=1}^h p_j - 1 \right) - \lambda_4 \left( \sum_{j=1}^h p_j [\exp\{\alpha + \beta\gamma(t_j)\} - 1] \right). \quad (\text{A1})$$

Taking the derivative of  $\ell_f(\alpha, \beta, F, \lambda_3, \lambda_4)$  with respect to  $p_j$ , setting it equal to zero, multiplying  $p_j$  on both sides, summing over  $j$  and taking the constraints into account, we obtain  $\lambda_3 = n$ . Taking the derivative of  $\ell_f(\alpha, \beta, F, \lambda_3, \lambda_4)$  with respect to  $\alpha$ , together with two constraints, we obtain  $\lambda_4 = n_2$ . After inserting

the Lagrange multipliers into (A1), the loglikelihood function  $\ell_f(\alpha, \beta, F)$  under the constraints is

$$\begin{aligned} \sum_{i=1}^n \left( \sum_{j=1}^h \delta_i I(Y_i = t_j) \left[ \log(p_j) + Z_i \{\alpha + \beta \gamma(t_j)\} \right] + (1 - \delta_i) \log \left[ \sum_{t_j \geq Y_i} t_j^{-1} e^{Z_i \{\alpha + \beta \gamma(t_j)\}} p_j \right] \right) \\ - n \left( \sum_{j=1}^h p_j - 1 \right) - n_2 \left( \sum_{j=1}^h p_j [\exp\{\alpha + \beta \gamma(t_j)\} - 1] \right). \end{aligned} \tag{A2}$$

Let  $P \equiv (p_1, \dots, p_h)$ . The maximizer of  $\ell_f(\alpha, \beta, F)$  for given  $(\alpha, \beta)$  is denoted by  $\hat{P}(\alpha, \beta)$  and the corresponding estimator of  $F$  is  $\hat{F}(\alpha, \beta)$ . The loglikelihood function  $\ell_f(\alpha, \beta, F)$  is strictly concave in  $P$ , implying a unique maximizer  $\hat{P}(\alpha, \beta)$  for each  $(\alpha, \beta)$  in the compact set  $\mathcal{B}$ . The compactness of  $\mathcal{B}$  and the continuity of the profile likelihood  $\ell_f\{\alpha, \beta, F(\alpha, \beta)\}$  implies the existence of the maximum likelihood estimator for  $(\alpha, \beta, F)$ . Condition A4 ensures the uniqueness of the maximum likelihood estimator (Rothenberg, 1971).

We prove the consistency of  $\hat{\theta}_n$  by the Kullback–Leibler information approach (Murphy, 1994; Parner, 1998). Since the maximum likelihood estimator  $\hat{\theta}_n$  is bounded, by Helly’s selection theorem, there exists a convergent subsequence of  $\hat{\theta}_n$ , denoted by  $\hat{\theta}_{nk} \equiv (\hat{\alpha}_{nk}, \hat{\beta}_{nk}, \hat{F}_{nk})$ , such that  $(\hat{\alpha}_{nk}, \hat{\beta}_{nk})$  converges to some  $(\alpha^*, \beta^*)$  in  $\mathcal{B}$  and  $\hat{F}_{nk}$  converges to a distribution function  $F^*$  in  $\mathcal{F}$ . The second step is to show that any such convergent subsequence of  $\hat{\theta}_n$  must converge to  $\theta_0$ , i.e.,  $\theta^* \equiv (\alpha^*, \beta^*, F^*) = \theta_0$ .

The empirical Kullback–Leibler distance  $\ell_f(\hat{\alpha}_{nk}, \hat{\beta}_{nk}, \hat{F}_{nk}) - \ell_f(\bar{\alpha}, \bar{\beta}, \bar{F})$  is always nonnegative for any  $(\bar{\alpha}, \bar{\beta}, \bar{F})$  in the parameter space, since  $(\hat{\alpha}_{nk}, \hat{\beta}_{nk}, \hat{F}_{nk})$  maximizes the loglikelihood function  $\ell_f(\alpha, \beta, F)$ . We choose  $(\bar{\alpha}, \bar{\beta}) = (\alpha_0, \beta_0)$  and let

$$\bar{F}(t) = \sum_{t_j \leq t} \frac{w_j^{f_0} + w_j^{g_0}}{\sum_{k=1}^h w_k^{f_0} + \sum_{k=1}^h w_k^{g_0} \exp\{\alpha_0 + \beta_0 \gamma(t_j)\}},$$

and  $E_0$  is the expectation under the true parameter values. If  $\theta_0$  was used as the initial value in the EM algorithm in § 3.2, then  $\bar{F}(t)$  is the one-step estimator of the cumulative distribution function. It can be shown that  $\bar{F}(t)$  converges to  $F_0$  almost surely and uniformly in  $t$ . By the properties of the maximum likelihood estimator,

$$\ell_f(\hat{\theta}_{nk}) - \ell_f(\alpha_0, \beta_0, \bar{F}) \geq 0. \tag{A3}$$

By taking limits on both sides, we can show that  $(nk)^{-1} \{\ell_f(\hat{\theta}_{nk}) - \ell_f(\alpha_0, \beta_0, \bar{F})\}$  must converge to the negative Kullback–Leibler distance between  $P_{\theta^*}$  and  $P_{\theta_0}$  almost surely by the strong law of large numbers, where  $P_\theta$  is the probability measure under the parameter  $\theta$ . Expression (A3) implies that the Kullback–Leibler distance between  $P_{\theta^*}$  and  $P_{\theta_0}$  is nonpositive. However, the limiting version of  $\ell_f$  is maximized at  $\theta_0$ , implying that  $P_{\theta^*} = P_{\theta_0}$  almost surely. Thus, the model identifiability yields  $\theta^* = \theta_0$ . Since every convergent subsequence of  $\hat{\theta}_n$  converges to  $\theta_0$ , the entire sequence  $\hat{\theta}_n$  must converge to  $\theta_0$  for any  $t \in [0, \tau]$ . The convergence is almost sure, since we only use the strong law of large numbers at most countably many times. The continuity and monotonicity of  $F_0$  ensure the uniform convergence of  $\hat{F}(t)$  in  $t$ .  $\square$

*Proof of Theorem 2.* The technical details are provided in the Supplementary Material. We only list the main steps here. For theoretical developments, it is helpful to express the likelihood in terms of the hazard function. Let  $\Lambda$  denote the cumulative hazard function for the length-biased time  $T$ . Using counting process notation,  $N_i(t) = I(Y_i \leq t)\delta_i$ , for  $i = 1, \dots, n$ , the loglikelihood for  $\psi \equiv (\alpha, \beta, \Lambda)$  is

$$\ell_f(\psi) = \sum_{i=1}^n \left[ \int_0^\tau \{\log d\Lambda(u) - \Lambda(u) + Z_i X^\top(u)\xi\} dN_i(u) + \log \int_0^\tau K_i(u) d\Lambda(u) \right],$$

where  $\xi = (\alpha, \beta)^\top$ ,  $X(u) = (1, \gamma(u))^\top$  and  $K_i(u) = (1 - \delta_i)I(u \geq Y_i) \exp\{Z_i X^\top(u)\xi - \Lambda(u)\} d\Lambda(u)$ . Denote the maximum likelihood estimator based on  $n$  observed samples by  $\hat{\psi}_n = (\hat{\alpha}, \hat{\beta}, \hat{\Lambda})$ . We prove

the weak convergence of  $\hat{\psi}_n$  by using the Z-theorem (van der Vaart & Wellner, 1996, Theorem 3.3.1) for infinite-dimensional estimating equations.

Denote the score equation vector of the loglikelihood  $\ell_f(\psi)$  by  $U(\xi, \Lambda) \equiv \{U_1(\xi, \Lambda), U_2(t; \xi, \Lambda)\}$ , where  $U_1(\xi, \Lambda)$  is the score function of  $\xi$ ,

$$U_1(\xi, \Lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \int_0^\tau Z_i X(u) dN_i(u) + \frac{\int_0^\tau K_i(u) X(u) d\Lambda(u)}{\int_0^\tau K_i(u) d\Lambda(u)} \right\},$$

and  $U_2(t; \xi, \Lambda)$  is the score function for the infinite-dimensional parameter  $\Lambda$ . To derive  $U_2(t; \xi, \Lambda)$ , we consider a one-dimensional submodel given as  $d\Lambda_{\eta, h} = (1 + \eta h) d\Lambda$ , where  $h$  is a bounded and integrable function. When setting  $h(\cdot) = I(\cdot \leq t)$  and evaluating the derivative of the loglikelihood with respect to  $\eta$  at  $\eta = 0$ ,

$$U_2(t; \xi, \Lambda) = \frac{1}{n} \sum_{i=1}^n \left( \int_0^t \left[ dN_i(u) - Y_i(u) d\Lambda(u) + \frac{\int_0^t \{K_i(u) - \int_u^\tau K_i(v) d\Lambda(v)\} d\Lambda(u)}{\int_0^\tau K_i(u) d\Lambda(u)} \right] \right),$$

where  $Y_i(u) = I(Y_i \geq u)$  is the at risk function for the  $i$ th subject. Denote the expectation of score function under the true parameter values  $(\alpha_0, \beta_0, \Lambda_0)$  by  $U_0(\xi, \Lambda) \equiv \{U_{10}(\xi, \Lambda), U_{20}(t; \xi, \Lambda)\}$ . We prove Theorem 2 by verifying the required conditions in Theorem 3.3.1 of van der Vaart & Wellner (1996). We first verify that  $U_0(\xi, \Lambda)$  is Fréchet differentiable and its Fréchet derivative can be calculated through the one-parameter submodel given by  $\xi_\eta = \xi_0 + \eta\xi$  and  $\Lambda_\eta = \Lambda_0 + \eta\Lambda$ :

$$\dot{U}_0(\xi, \Lambda) \equiv - \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} \xi \\ \Lambda \end{pmatrix} \equiv -\{d_{11}(\xi) + d_{12}(\Lambda), d_{21}(\xi)(\cdot) + d_{22}(\Lambda)(\cdot)\},$$

where

$$d_{11}(\xi) = \partial U_{10}(\xi_\eta, \Lambda_0) / \partial \eta|_{\eta=0}, \quad d_{12}(\Lambda) = \partial U_{10}(\xi_0, \Lambda_\eta) / \partial \eta|_{\eta=0}, \\ d_{21}(\xi)(t) = \partial U_{20}(t, \xi_\eta, \Lambda_0) / \partial \eta|_{\eta=0}, \quad d_{22}(\Lambda)(t) = \partial U_{20}(t, \xi_0, \Lambda_\eta) / \partial \eta|_{\eta=0}.$$

In the Supplementary Material, we prove that the inverse operator  $\dot{U}_0^{-1}$  exists and takes the following form:

$$\dot{U}_0^{-1}(\xi, \Lambda) \equiv \begin{pmatrix} d_{11}^{-1} + d_{11}^{-1} d_{12} \Phi^{-1} d_{21} d_{11}^{-1} & -d_{11}^{-1} d_{12} \Phi^{-1} \\ \Phi^{-1} d_{21} d_{11}^{-1} & \Phi^{-1} \end{pmatrix} \begin{pmatrix} \xi \\ \Lambda \end{pmatrix},$$

where  $\Phi = d_{22}(\Lambda) - d_{21} d_{11}^{-1} d_{12}$ .

We then show the weak convergence of the score equation vector evaluated at the true value,  $U(\xi_0, \Lambda_0)$ . Last we prove the following stochastic approximation for the score equation vector

$$n^{1/2}\{(U - U_0)(\hat{\psi}_n) - (U - U_0)(\psi_0)\} = o_p(1).$$

We conclude that under the Conditions A1–A4,  $n^{1/2}(\hat{\psi}_n - \psi_0)$  converges weakly to a tight mean zero Gaussian process  $-\dot{U}_0^{-1}(\mathbb{W})$ . By the functional delta method,  $n^{1/2}(\hat{\theta}_n - \theta_0)$  converges weakly to a tight mean zero Gaussian process  $-\phi'_{\psi_0}\{\dot{U}_0^{-1}(\mathbb{W})\}$ , where  $\phi$  is the transformation from  $\mathcal{B} \times \mathcal{A} \rightarrow \mathcal{B} \times \mathcal{F}$ , and  $\phi'_{\psi_0}$  is its Hadamard derivative at  $\psi_0$ .  $\square$

## REFERENCES

- ANDERSON, J. A. (1979). Multivariate logistic compounds. *Biometrika* **66**, 17–26.  
 ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. & KEIDING, N. (1992). *Statistical Models Based on Counting Processes*. New York: Springer.  
 ASGHARIAN, M., M'LAN, C. E. & WOLFSON, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Am. Statist. Assoc.* **97**, 201–9.  
 ASGHARIAN, M., WOLFSON, D. B. & ZHANG, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statist. Med.* **25**, 1751–67.

- BILKER, W. B. & WANG, M. C. (1996). A semiparametric extension of the Mann–Whitney test for randomly truncated data. *Biometrics* **50**, 10–20.
- CHEN, Y. Q. (2010). Semiparametric regression in size-biased sampling. *Biometrics* **66**, 149–58.
- GROSS, S. T. & LAI, T. L. (1996). Bootstrap methods for truncated and censored data. *Statist. Sinica* **6**, 509–30.
- KALBFLEISCH, J. D. (1978). Likelihood methods and nonparametric tests. *J. Am. Statist. Assoc.* **73**, 167–70.
- KAY, R. & LITTLE, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika* **74**, 495–501.
- MANDEL, M. & RITOV, Y. (2010). The accelerated failure time model under biased sampling. *Biometrics* **66**, 149–58.
- MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.
- PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.
- PETO, R. & PETO, J. (1972). Asymptotically efficient rank invariant test procedures. *J. R. Statist. Soc. A* **135**, 185–207.
- QIN, J. & LEUNG, D. (2005). A semiparametric two-component ‘compound’ mixture model and its application to estimating malaria attributable fractions. *Biometrics* **61**, 456–64.
- QIN, J. & ZHANG, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609–18.
- QIN, J. & ZHANG, B. (2005). Marginal likelihood, conditional likelihood and empirical likelihood: connections and applications. *Biometrika* **92**, 251–70.
- QIN, J., BERWICK, M., ASHBOLT, R. & DWYER, T. (2002). Quantifying the change of melanoma incidence by Breslow thickness. *Biometrics* **58**, 665–70.
- ROTHENBERG, T. J. (1971). Identification in parametric models. *Econometrica* **39**, 577–91.
- SCHEIKE, T. H. & KEIDING, N. (2006). Design and analysis of time-to-pregnancy. *Statist. Meth: Med. Res.* **15**, 127–40.
- SHEN, Y., QIN, J. & COSTANTINO, J. (2007). Inference of tamoxifen’s effects on prevention of breast cancer from a randomized controlled trial. *J. Am. Statist. Assoc.* **102**, 1235–44.
- VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer.
- VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751–61.
- WANG, M. C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Am. Statist. Assoc.* **86**, 130–43.
- WOLFSON, C., WOLFSON, D. B., ASGHARIAN, M., M’LAN, C., OSTBYE, T., ROCKWOOD, K., HOGAN, D. & THE CLINICAL PROGRESSION OF DEMENTIA STUDY GROUP (2001). A reevaluation of the duration of survival after the onset of dementia. *New Engl. J. Med.* **344**, 1111–6.
- ZELEN, M. (2004). Forward and backward recurrence times and length biased sampling: age specific models. *Lifetime Data Anal.* **10**, 325–34.
- ZENG, D. & LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Statist. Soc. B* **69**, 507–64.

[Received June 2010. Revised December 2011]