



HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2017 August 18.

Published in final edited form as:

J Am Stat Assoc. 2016 ; 111(514): 787–799. doi:10.1080/01621459.2015.1044090.

Efficient Estimation of the Cox Model With Auxiliary Subgroup Survival Information

Chiung-Yu Huang [Associate Professor],

Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, Maryland 21205

Jing Qin [Mathematical Statistician], and

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Huei-Ting Tsai [Assistant Professor]

Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC 20007

Abstract

With the rapidly increasing availability of data in the public domain, combining information from different sources to infer about associations or differences of interest has become an emerging challenge to researchers. This paper presents a novel approach to improve efficiency in estimating the survival time distribution by synthesizing information from the individual-level data with t -year survival probabilities from external sources such as disease registries. While disease registries provide accurate and reliable overall survival statistics for the disease population, critical pieces of information that influence both choice of treatment and clinical outcomes usually are not available in the registry database. To combine with the published information, we propose to summarize the external survival information via a system of nonlinear population moments and estimate the survival time model using empirical likelihood methods. The proposed approach is more flexible than the conventional meta-analysis in the sense that it can automatically combine survival information for different subgroups and the information may be derived from different studies. Moreover, an extended estimator that allows for a different baseline risk in the aggregate data is also studied. Empirical likelihood ratio tests are proposed to examine whether the auxiliary survival information is consistent with the individual-level data. Simulation studies show that the proposed estimators yield a substantial gain in efficiency over the conventional partial likelihood approach. Two sets of data analysis are conducted to illustrate the methods and theory.

Keywords

Information synthesis; Meta-analysis; SEER cancer registries; Subgroup analysis

1. Introduction

Combining information from different sources to infer about associations or differences of interest is an important area of research known as meta-analysis. Results of meta-analyses have been used to guide the design of future studies, aid the development of regulatory

recommendations, and even modify clinical practice. A PubMed search of the word “meta-analysis” in article titles found 7231 articles in just 2013. With the rapidly increasing availability of data in the public domain, taking full advantage of available information while saving considerable resources has become an emerging challenge to researchers.

An ideal meta-analysis would be an analysis of pooled individual-level data, where the raw data from each study are obtained and analyzed directly. In clinical studies, the use of pooled individual-level data enables researchers to conduct subgroup analysis to investigate whether patient characteristics are related to treatment effects. In most applications, however, synthesis of the information is conducted by analyzing summary statistics, such as means, standard deviations, proportions, odds ratios, and relative risks, from each individual study. Specifically, meta-analysis calculates a weighted average of the summary statistics across studies to provide an overall measure of the association or difference of interest. The major drawback of this approach is that covariate-treatment interactions are usually not provided in the reports of primary analysis findings, thus making subgroup analysis difficult to perform (Simmonds and Higgins, 2007).

On the other hand, methods for combining information from both individual-level data and published aggregate data have drawn much attention (Kovalchick, 2013; Liu et al., 2014). This research is particularly inspired by the growing interest in exploiting the population-based cancer survival statistics made available by the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute (NCI). The SEER program consists of 18 cancer registries covering approximately 28% of the U.S. population. The registries began collecting demographics and cancer factors on all types of incident cancer patients in 1973, and the database links to state death certificates for patient survival information, including cause of death. The SEER program updates their data annually and has been used by thousands of researchers, clinicians, public health officials, policy makers, community groups and public for cancer incidence and survival statistics in the United States. The SEER Cancer Statistics Review 1973-2010 (http://seer.cancer.gov/csr/1975_2010/) reports the 5-year survival after cancer diagnosis by race, sex, age, and year of diagnosis for the major cancer sites and for all cancers combined using data from the population-based cancer registry. For example, the 5-year survival among ovarian cancer patients diagnosed before age 65 is 57%, and is 27.7% among patients diagnosed after age 65. This paper presents a novel approach for combining data from clinical studies, that is, individual-level data, and the published subgroup t -year survival probabilities, that is, aggregate data. Individual-level data provide estimates of the treatment-covariate interactions or effects of biomarkers that are not reported in existing publications, but the sample size may be too small to provide accurate estimates. Properly combining with survival information available from external sources is expected to yield more efficient estimates of the effects of interest as well as more accurate prediction of the risk of the failure event.

Our approach to synthesize information from different sources is motivated by the empirical likelihood method that was first introduced by Thomas and Grunkemeier (1975) to obtain confidence intervals for the Kaplan-Meier estimator. Later, Owen (1988, 1990) studied empirical likelihood-based confidence regions for the mean or other functions. It has been

shown that the empirical likelihood ratio has a limiting chi-squared distribution under mild regularity conditions, and that the empirical likelihood is Bartlett correctable in many applications (DiCiccio et al., 1991), leading to an advantage over the bootstrap method in the construction of confidence regions. Many researchers have applied the empirical likelihood method to more general settings. In particular, Qin and Lawless (1994) made connections to estimating equations and demonstrated that the empirical likelihood method can yield the most efficient estimator by making optimal use of the estimating equations. Although the use of empirical likelihood methods has been very popular in fields such as survey sampling, the main interest is usually to improve the estimation of the mean or other functions of the distribution function. For example, Chen and Qin (1993), Chen and Wu (2002), Chen et al. (2002), and Wu and Sitter (2001) applied the empirical likelihood method to incorporate auxiliary covariate information to improve efficiency of estimation. Imbens (2002) provided a very nice discussion on how the empirical likelihood methods can be used as an alternative for the generalized method of moments.

Application of the empirical likelihood approach to survival data has received much attention because variances of statistical estimates can be very difficult to estimate in the presence of right censoring. The empirical likelihood method can efficiently establish joint confidence regions without directly estimating the corresponding asymptotic variances, and can significantly improve coverage accuracy. In a review paper, Li et al. (2005) summarized the result of empirical likelihood analysis for censored survival time data. Ren and Zhou (2011) investigated the properties of the maximum empirical likelihood estimator and compared with the maximum partial likelihood method. Zhou (2006) applied the empirical likelihood method to improve estimation of the Cox model with partial information on the baseline hazard function.

The proposed application of the empirical likelihood method deals with a nonconventional scenario. Under the proportional hazards model, the auxiliary t -year survival probabilities amount to a system of nonlinear estimating equations that involve the regression parameters, the infinite-dimensional baseline hazard function, and the infinite-dimensional marginal distribution function of the covariate X , making it difficult to derive the constrained maximum likelihood estimator. To tackle this difficulty, two empirical likelihoods, one based on the conditional likelihood of the survival time T given X and another based on the marginal likelihood of X , are constructed to combine information from different sources. In the same spirit of Breslow's estimator (Breslow, 1972), the baseline hazard function and the marginal distribution function of X are profiled out of the conditional likelihood and the marginal likelihood, respectively. To the best of our knowledge, this is the first paper that considers a double empirical likelihood approach.

This paper is organized as follows. In Section 2 we introduce notation and summarize the landmark survival information as unbiased estimating equations. The main results are presented in Section 3, where a double empirical likelihood approach is proposed to synthesize the auxiliary survival information. Because the auxiliary survival information may not be consistent with the individual-level data due to inclusion/exclusion criteria of the clinical study, in Section 4 we extend the proposed double empirical likelihood method to allow the population from which the aggregate survival information is derived to have a

different baseline risk. We present results of simulation studies in Section 5 and illustrate the proposed methods with two sets of data analysis in Section 6. Some concluding remarks are given in Section 6.

2. Model Setup

Let T denote the time from disease onset to a failure event or an event of interest. Assume that T is absolutely continuous, that is, T has a probability density. Let \mathbf{X} denote a $p \times 1$ vector of baseline covariates. Denote by $f(t|\mathbf{x})$ and $S(t|\mathbf{x})$ the conditional density function and the conditional survival function of T given $\mathbf{X} = \mathbf{x}$. We assume that the survival time T follows the proportional hazards model (Cox, 1972)

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

where β is a vector of $p \times 1$ regression parameters and $\lambda(t)$ is an unspecified baseline hazard function. Let $\Lambda(t) = \int_0^t \lambda(u) du$ be the corresponding baseline cumulative hazard function. The observation of the survival time is usually subject to right censoring due to study end or premature dropout. Thus, instead of observing the actual value of the survival time T , we observe the possibly censored survival time $Y = \min(T, C)$, where C is the time of censoring. In many applications, it is reasonable to assume that C is independent of T given the observed covariates \mathbf{X} .

Our goal is to derive an efficient estimator of the Cox model by incorporating published t -year survival probabilities. To express the auxiliary information on survival at the time point t^* , we use Ω_k , $k = 1, \dots, K$ to denote the k th subgroup whose t^* -year survival is provided. In the aforementioned ovarian cancer example, we set \mathbf{Z} to be the age of diagnosis, where \mathbf{Z} is a subset of the covariate of interest \mathbf{X} , which may include biomarkers and other risk factors of ovarian cancer. Write $\mathbf{X} = (\mathbf{Z}, \mathbf{W})$. Thus the two subgroups of ovarian cancer patients are $\Omega_1 = \{(\mathbf{Z}, \mathbf{W}) : \mathbf{Z} < 65\}$ and $\Omega_2 = \{(\mathbf{Z}, \mathbf{W}) : \mathbf{Z} \geq 65\}$, and the auxiliary 5-year survival probabilities obtained from the SEER Cancer Statistics Review 1973-2010 are given by $\text{pr}(T > 5 | \mathbf{X} \in \Omega_1) = 0.56$ and $\text{pr}(T > 5 | \mathbf{X} \in \Omega_2) = 0.277$.

A general expression of the auxiliary survival information for subgroup k at the time point t^* is

$$\text{pr}(T > t^* | \mathbf{X} \in \Omega_k) = \phi_k, k = 1, \dots, K,$$

or, equivalently, $\text{pr}(T > t^* | \mathbf{X} \in \Omega_k) - \phi_k \times \text{pr}(\mathbf{X} \in \Omega_k) = 0$. By double expectation and under the assumed Cox model, we can derive

$$E[I(\mathbf{X} \in \Omega_k)\{S(t^*|\mathbf{X}) - \phi_k\}] = E\{I(\mathbf{X} \in \Omega_k)[\exp\{-\Lambda(t^*)\exp(\beta' \mathbf{X})\} - \phi_k]\} = 0.$$

Define the estimating function

$$\psi_k(\mathbf{X}, \boldsymbol{\beta}, \Lambda) = I(\mathbf{X} \in \Omega_k) [\exp\{-\Lambda(t^*) \exp(\boldsymbol{\beta}' \mathbf{X})\} - \phi_k].$$

Then the subgroup survival information at t^* is summarized by

$$E\{\psi_k(\mathbf{X}, \boldsymbol{\beta}, \Lambda)\} = 0, \quad k=1, \dots, K, \quad (1)$$

where the random function $\psi_k(\mathbf{X}, \boldsymbol{\beta}, \Lambda)$, $k = 1, \dots, K$, is bounded by 2. Note that the estimating equations only involve the regression parameter $\boldsymbol{\beta}$ and the baseline cumulative hazard function evaluated at t^* . Hence, by setting $\alpha = \Lambda(t^*)$, equation (1) can be reexpressed as $E\{\psi_k(\mathbf{X}, \boldsymbol{\beta}, \alpha)\} = 0$, $k = 1, \dots, K$.

3. Method

In this section, we introduce a double empirical likelihood method to synthesize information from different sources. Under the Cox model, the density of the bivariate random variable (T, \mathbf{X}) , relative to the product of the Lebesgue measure and the marginal distribution of \mathbf{X} , is given by $\exp(\boldsymbol{\beta}' \mathbf{x}) \lambda(t) \exp\{-\Lambda(t) \exp(\boldsymbol{\beta}' \mathbf{x})\} dG(\mathbf{x})$, where G is the distribution function of \mathbf{X} . Assume that the observed data (Y_i, \mathbf{X}_i) , $i = 1, \dots, n$, on n subjects are independent and identically distributed realizations of (Y, \mathbf{X}) . Dropping the factors involving the censoring time distribution, the log full likelihood based on the observed data is

$$\ell_F = \sum_{i=1}^n \Delta_i [\boldsymbol{\beta}' \mathbf{X}_i + \log\{d\Lambda(Y_i)\}] - \Lambda(Y_i) \exp(\boldsymbol{\beta}' \mathbf{X}_i) + \log\{dG(\mathbf{X}_i)\}.$$

Define the functions $S^{(k)}(t, \boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n I(Y_j \geq t) \exp(\boldsymbol{\beta}' \mathbf{X}_j) \mathbf{X}_j^{\otimes k}$, $k = 0, 1, 2$, with $\mathbf{x}^{\otimes 2} = \mathbf{x}' \mathbf{x}$. Following the empirical likelihood method of Owen (1988) and Qin and Lawless (1994), we denote by λ_i the jump of Λ at Y_i and by p_i the jump of G at \mathbf{X}_i . The full empirical likelihood can then be decomposed as the product of the conditional likelihood of (Y_i) given \mathbf{X} and the marginal likelihood of \mathbf{X} , where the log conditional likelihood is

$$\ell_C = \sum_{i=1}^n \Delta_i (\boldsymbol{\beta}' \mathbf{X}_i + \log \lambda_i) - n \sum_{i=1}^n \lambda_i S^{(0)}(Y_i, \boldsymbol{\beta}),$$

and the log marginal likelihood is

$$\ell_M = \sum_{i=1}^n \log(p_i).$$

For a fixed β , differentiating the log conditional likelihood \mathcal{L} with respect to λ_i and setting the derivative to 0 yields

$$\lambda_i = \frac{1}{n} \times \frac{\Delta_i}{S^{(0)}(Y_i, \beta)},$$

which leads to the Breslow (1972) estimator for the baseline cumulative hazard function

$$\hat{\Lambda}_B(t, \beta) = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i I(Y_i \leq t)}{S^{(0)}(Y_i, \beta)}.$$

Another well-known result is that replacing Λ with $\hat{\Lambda}_B$ in \mathcal{L} yields the (log) partial likelihood function.

Assume that the auxiliary survival information is consistent with the individual-level data, that is, individuals in the clinical study are a representative sample of the population from which the aggregate survival information is derived. A simple idea to combine auxiliary information is applying the empirical likelihood method to maximize the full likelihood with respect to the constraints

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad \text{and} \quad \sum_{i=1}^n p_i \psi_k(\mathbf{X}_i; \beta, \Lambda) = 0, \quad k=1, \dots, K.$$

Because the estimating function ψ_k only involves the value of $\Lambda(t)$ at $t = t^*$, intuitively, one may replace $\Lambda(t^*)$ in ψ_k with its Breslow-type estimator $\hat{\Lambda}_B(t^*, \beta)$. However, simulation studies suggest that this simple approach yields biased estimation, because the Breslow-type estimator involves unknown parameter β .

We propose to combine the auxiliary subgroup survival information to estimate the Cox model by formulating two empirical likelihoods – one of which is derived from the conditional likelihood and the other is derived from the marginal likelihood. Our idea is to treat $\alpha = \Lambda(t^*)$ as a nuisance parameter and construct an empirical likelihood for α . We then formulate the usual empirical likelihood for β and Λ using the auxiliary survival information which depends on β and α .

The steps to construct the double empirical likelihood are described below. By definition, $\alpha = \Lambda(t^*) = \int_0^\infty I(u \leq t^*) d\Lambda(u)$. We propose to maximize the log conditional likelihood \mathcal{L} subject to the constraint

$$\sum_{i=1}^n \lambda_i I(Y_i \leq t^*) - \alpha = 0.$$

Applying Lagrange multipliers ν , the objective function to be maximized is

$$\sum_{i=1}^n \Delta_i (\boldsymbol{\beta}' \mathbf{X}_i + \log \lambda_i) - n \sum_{i=1}^n \lambda_i S^{(0)}(Y_i, \boldsymbol{\beta}) - n\nu \left\{ \sum_{i=1}^n \lambda_i I(Y_i \leq t^*) - \alpha \right\}.$$

Taking derivative of the objective function with respect to λ_i and setting the derivative to 0 yields

$$\lambda_i = \frac{1}{n} \times \frac{\Delta_i}{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)}, \quad (2)$$

where the Lagrange multiplier is determined by

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\Delta_i I(Y_i \leq t^*)}{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)} \right\} - \alpha = 0.$$

Substituting (2) back to the objective function yields, up to a constant,

$$\sum_{i=1}^n \Delta_i [\boldsymbol{\beta}' \mathbf{X}_i - \log \{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)\}] + n\nu\alpha.$$

Hence the marginal empirical score function for $\boldsymbol{\beta}$ is

$$\sum_{i=1}^n \Delta_i \left\{ \mathbf{X}_i - \frac{S^{(1)}(Y_i, \boldsymbol{\beta})}{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)} \right\}.$$

Next, we maximize the log marginal likelihood ℓ_M with respect to the constraints $p_i \geq 0$,

$\sum_{i=1}^n p_i = 1$, and $\sum_{i=1}^n p_i \psi_k(\mathbf{X}_i, \boldsymbol{\beta}, \alpha) = 0$ for $k = 1, \dots, K$. Write $\boldsymbol{\psi}(\mathbf{x}, \boldsymbol{\beta}, \alpha) = \{\psi_1(\mathbf{x}, \boldsymbol{\beta}, \alpha), \dots, \psi_K(\mathbf{x}, \boldsymbol{\beta}, \alpha)\}'$. Given $\boldsymbol{\beta}$ and α , a unique maximum exists provided $\mathbf{0}$ lies in the convex hull of $\boldsymbol{\psi}(\mathbf{X}_1, \boldsymbol{\beta}, \alpha), \dots, \boldsymbol{\psi}(\mathbf{X}_n, \boldsymbol{\beta}, \alpha)$. Applying the classic empirical likelihood argument, we have

$$p_i = \frac{1}{n} \times \frac{1}{1 + \boldsymbol{\xi}' \boldsymbol{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha)}$$

and the constrained log marginal likelihood, up to a constant,

$$-\sum_{i=1}^n \log\{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)\}$$

where the Lagrange multipliers $\xi = (\xi_1, \dots, \xi_K)'$ are determined by

$$\frac{1}{n} \sum_{i=1}^n \frac{\psi(\mathbf{X}_i, \beta, \alpha)}{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)} = 0.$$

Combing the two constrained log likelihoods yields the constrained log full likelihood function ℓ where, up to a constant,

$$\ell(\beta, \xi, \nu, \alpha) = \sum_{i=1}^n \Delta_i [\beta' \mathbf{X}_i - \log\{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)\}] + n\nu\alpha - \sum_{i=1}^n \log\{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)\}.$$

The procedure described above enables us to change an infinite dimension problem to a finite-dimension problem at the expense of introducing an additional $(K + 2)$ -dimensional parameters.

To estimate β , we solve a system of empirical score equations:

$$\begin{aligned} U_1(\beta, \xi, \nu, \alpha) &= \sum_{i=1}^n \Delta_i \left\{ \mathbf{X}_i - \frac{S^{(1)}(Y_i, \beta)}{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)} \right\} - \sum_{i=1}^n \frac{\xi' \frac{\partial \psi}{\partial \beta}(\mathbf{X}_i, \beta, \alpha)}{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)}, \\ U_2(\beta, \xi, \nu, \alpha) &= \sum_{i=1}^n \frac{\psi(\mathbf{X}_i, \beta, \alpha)}{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)}, \\ U_3(\beta, \xi, \nu, \alpha) &= \sum_{i=1}^n \left\{ \frac{\Delta_i I(Y_i \leq t^*)}{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)} - \alpha \right\}, \\ U_4(\beta, \xi, \nu, \alpha) &= \sum_{i=1}^n \left\{ \frac{\xi' \frac{\partial \psi}{\partial \alpha}(\mathbf{X}_i, \beta, \alpha)}{1+\xi' \psi(\mathbf{X}_i, \beta, \alpha)} - \nu \right\}, \end{aligned}$$

with the usual convention $0/0 = 0$. These empirical score functions are derived by taking derivative of the empirical likelihood ℓ with respect to $\theta = (\beta, \xi, \nu, \alpha)$. Let β_0, Λ_0 , and α_0 be the true parameter values, and denote $\theta_0 = (\beta_0, \mathbf{0}, 0, \alpha_0)$. Define $U_{0k} = U_k(\beta_0, \mathbf{0}, 0, \alpha_0)$, $k = 1, 2, 3, 4$, that is, U_{0k} is the value of U_k evaluated at θ_0 . Define $U' = (U'_1, U'_2, U_3, U_4)'$ and $U'_0 = (U'_{01}, U'_{02}, U_{03}, U_{04})$, where $U_{04} = 0$. We show in the Appendix that, under some regularity conditions, $n^{-1/2}U_0$ converges in distribution to a zero-mean multivariate normal distribution with variance-covariance matrix

$$\Omega = \begin{pmatrix} \Sigma & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where Σ, \mathbf{J}, K are defined in the appendix. Note, for convenience, we use $\mathbf{0}$ and \mathbf{I} to denote a matrix of 0's and an identity matrix with proper dimensions.

Denote by $\hat{\theta} = (\hat{\beta}, \hat{\xi}, \hat{v}, \hat{a})$ the solution to $U(\theta) = \mathbf{0}$. The large-sample properties of $\hat{\theta}$ are presented in Theorem 1, the proof of which is given in the Appendix.

Theorem 1 Assume that \mathbf{X} is bounded, the true regression parameter β_0 lies in a compact set, and both T and C are absolutely continuous. Moreover, assume that $E\{\psi(\mathbf{X}, \beta_0, \alpha_0)\psi(\mathbf{X}, \beta_0, \alpha_0)'\}$ is positive definite and $\alpha_0 = \Lambda_0(t^*) < \infty$. Then $n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution to a zero mean multivariate normal distribution with variance-covariance matrix $\Gamma^{-1} = (\Sigma + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}')^{-1}$, provided Γ is non-singular, where \mathbf{B} and \mathbf{Q} are specified in the Appendix.

Interestingly, Σ^{-1} is the asymptotic covariance-covariance matrix of $n^{1/2}(\hat{\beta}_{PL} - \beta_0)$, where $\hat{\beta}_{PL}$ is the maximum partial likelihood estimator. Hence Theorem 1 implies that the proposed estimator $\hat{\beta}$ is asymptotically more efficient than the maximum partial likelihood estimator $\hat{\beta}_{PL}$.

When the subgroups involved in the auxiliary survival information are determined by a subset of covariates, the efficiency gain in the estimated coefficients for other covariates is expected to be minimum. To see this, consider a simple case where $\mathbf{X} = (X_1, X_2)$ and the subgroups are determined only based on X_1 . Then the auxiliary survival information for the k th subgroup can be reexpressed as

$$\begin{aligned} & E\{\psi_k(\mathbf{X}, \beta, \alpha)\} \\ &= \int \int \{I(x_1 \in \Omega_k) [\exp\{ \\ & \quad - \alpha \exp(\beta_1 x_1 \\ & \quad + \beta_2 x_2)\} - \phi_k]\} dG(x_1, x_2) = \int \int \{I(x_1 \in \Omega_k) [\exp\{ \\ & \quad - \alpha \exp(\beta_1 x_1 \\ & \quad + u)\} \\ & \quad - \phi_k]\} dG(x_1, \beta_2^{-1}u). \end{aligned}$$

Because the proposed estimation procedure allows for an arbitrary distribution function G for (X_1, X_2) , after reparameterization, it is equivalent to maximize the log marginal likelihood $\sum_{i=1}^n \log(p_i^*)$ with respect to the constraints

$$p_i^* \geq 0, \sum_{i=1}^n p_i^* = 1 \text{ and } \sum_{i=1}^n p_i^* \{I(X_{i1} \in \Omega_k) [\exp\{-\alpha \exp(\beta_1 X_{i1} + X_{i2})\} - \phi_k]\},$$

where $X_i = (X_{i1}, X_{i2})$ and p_i^* is the jump of G at $(X_{i1}, X_{i2}/\beta_2)$. It is easy to see that, after profiling out p_i^* , the auxiliary survival information does not involve β_2 . Thus the proposed estimation procedure is expected to significantly improve the efficiency in the estimation of β_1 but has only limited impact on the estimation of β_2 .

To estimate the baseline cumulative hazard function $\Lambda(t)$, we consider the following empirical likelihood-based estimator to incorporate the auxiliary survival information:

$$\hat{\Lambda}(t) = \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i I(Y_i \leq t)}{S^{(0)}(Y_i, \hat{\beta}) + \hat{\nu} I(Y_i \leq t^*)}.$$

Applying the functional delta method, we can show that $n^{1/2}\{\hat{\Lambda}_{EL}(t) - \Lambda(t)\}$ converges to a zero-mean Gaussian process on $[0, \tau]$. A sketch of the proof is given in the Appendix.

The validity of the proposed method holds when the t -year survival probabilities summarized by (1) are consistent with the individual-level data. To test the conformity of the auxiliary survival information, an empirical likelihood ratio test statistic can be constructed in the spirit of Corollary 4 of Qin and Lawless (1994) and Qin and Lawless (1995). Specifically, we consider the test statistic

$$R = 2 \left\{ \sup_{\beta, \xi, \nu, \alpha} l(\beta, \xi, \nu, \alpha) - \sup_{\beta, \nu, \alpha} l(\beta, \mathbf{0}, \nu, \alpha) \right\}.$$

Note that when the conformity assumption holds, that is, $\xi = \mathbf{0}$, the likelihood $l(\beta, \alpha, \mathbf{0}, \nu)$ is maximized by $(\beta, \nu, \alpha) = (\hat{\beta}_{PL}, \mathbf{0}, \hat{\alpha}_{PL})$, where $\hat{\alpha}_{PL} = \hat{\Lambda}_B(t^*, \hat{\beta}_{PL})$ is the Breslow-type estimator of the baseline cumulative hazard function at time t^* . Theorem 2 summarizes the asymptotic properties of empirical log-likelihood ratio statistic R .

Theorem 2 Under the regularity conditions specified in Theorem 1 and the null hypothesis that $\xi = \mathbf{0}$, the empirical log-likelihood ratio R converges in distribution to a χ^2 random variable with K degrees of freedom as $n \rightarrow \infty$.

4. An Extension

As discussed before, a major limitation of the estimation procedure described in Section 3 is that the auxiliary information must be consistent with the individual-level data. However, due to study inclusion/exclusion criteria, subjects enrolled in the clinical study may not be a representative sample of the population from which the aggregate survival information is derived. Hence it is desired to allow the aggregate data to have a different survival time model. To this end, we propose to accommodate the inconsistency by assuming that the

hazard function of the survival time in the aggregate data follows the Cox model $\lambda^*(t)\exp(\beta'x)$, where

$$\lambda^*(t) = \rho\lambda(t), \quad \rho > 0, \quad (3)$$

so that the potential differences in the two data sources are characterized by a scale factor ρ . Of note, $\rho = 1$ indicates that survival time model for the aggregate data is the same as that for the individual-level data.

Similar to the discussions in Section 3, the auxiliary survival information $\text{pr}(T > t^* | \mathbf{X} \in \Omega_k) = \phi_k$, $k = 1, \dots, K$, can be summarized by the estimating equations

$$E\{\tilde{\psi}_k(\mathbf{X}, \boldsymbol{\beta}, \alpha, \rho)\} = 0, \quad k = 1, \dots, K,$$

where, under model (3), $\hat{\psi}_k(\mathbf{X}, \boldsymbol{\beta}, \alpha, \rho) = I(\mathbf{X} \in \Omega_k)[\exp\{-\rho\alpha\exp(\boldsymbol{\beta}'\mathbf{X})\} - \phi_k]$. The constrained log full likelihood function is given by, up to a constant,

$$\tilde{l}(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) = \sum_{i=1}^n \Delta_i [\boldsymbol{\beta}'\mathbf{X}_i - \log\{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)\}] + n\nu\alpha - \sum_{i=1}^n \log\{1 + \boldsymbol{\xi}'\tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)\}.$$

Taking derivative of the empirical likelihood $\tilde{\ell}$ with respect to $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho)$, we reach a system of empirical score equations:

$$\begin{aligned} \tilde{U}_1(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) &= \sum_{i=1}^n \Delta_i \left\{ \mathbf{X}_i - \frac{S^{(1)}(Y_i, \boldsymbol{\beta})}{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)} \right\} - \sum_{i=1}^n \frac{\boldsymbol{\xi}' \frac{\partial \tilde{\psi}}{\partial \boldsymbol{\beta}}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}{1 + \boldsymbol{\xi}' \tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}, \\ \tilde{U}_2(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) &= \sum_{i=1}^n \frac{\tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}{1 + \boldsymbol{\xi}' \tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}, \\ \tilde{U}_3(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) &= \sum_{i=1}^n \left\{ \frac{\Delta_i I(Y_i \leq t^*)}{S^{(0)}(Y_i, \boldsymbol{\beta}) + \nu I(Y_i \leq t^*)} - \alpha \right\}, \\ \tilde{U}_4(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) &= \sum_{i=1}^n \left\{ \frac{\boldsymbol{\xi}' \frac{\partial \tilde{\psi}}{\partial \boldsymbol{\xi}}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}{1 + \boldsymbol{\xi}' \tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)} - \nu \right\}, \\ \tilde{U}_5(\boldsymbol{\beta}, \boldsymbol{\xi}, \nu, \alpha, \rho) &= \sum_{i=1}^n \frac{\boldsymbol{\xi}' \frac{\partial \tilde{\psi}}{\partial \rho}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}{1 + \boldsymbol{\xi}' \tilde{\psi}(\mathbf{X}_i, \boldsymbol{\beta}, \alpha, \rho)}, \end{aligned}$$

with the usual convention $0/0 = 0$. Let ρ_0 be the true parameter value for ρ , and denote $\tilde{\boldsymbol{\theta}}_0 = (\boldsymbol{\theta}'_0, \rho_0)' = (\boldsymbol{\beta}_0, \mathbf{0}, 0, \alpha_0, \rho_0)$. Define $\tilde{\boldsymbol{U}}_{0k} = \tilde{\boldsymbol{U}}_k(\tilde{\boldsymbol{\theta}}_0)$, $k = 1, \dots, 5$, that is, $\tilde{\boldsymbol{U}}_{0k}$ is the value of $\tilde{\boldsymbol{U}}_k$ evaluated at $\tilde{\boldsymbol{\theta}}_0$. Define $\tilde{\boldsymbol{U}}' = (\tilde{\boldsymbol{U}}'_1, \tilde{\boldsymbol{U}}'_2, \tilde{\boldsymbol{U}}'_3, \tilde{\boldsymbol{U}}'_4, \tilde{\boldsymbol{U}}'_5)$ and $\tilde{\boldsymbol{U}}'_0 = (\tilde{\boldsymbol{U}}'_{01}, \tilde{\boldsymbol{U}}'_{02}, \tilde{\boldsymbol{U}}'_{03}, \tilde{\boldsymbol{U}}'_{04}, \tilde{\boldsymbol{U}}'_{05})$, where $\tilde{\boldsymbol{U}}'_{04} = \tilde{\boldsymbol{U}}'_{05} = 0$. Let $\hat{\boldsymbol{\theta}}_\rho$ be the solution to $\tilde{\boldsymbol{U}}(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ and $\hat{\boldsymbol{\beta}}_\rho$ be the corresponding estimated regression coefficient. The large-sample properties of $\hat{\boldsymbol{\theta}}_\rho$ are presented in Theorem 3, with the proof given in the Appendix.

Theorem 3 Assume that the matrix $E\{\tilde{\Psi}(X, \beta_0, \alpha_0, \rho_0)\tilde{\Psi}(X, \beta_0, \alpha_0, \rho_0)'\}$ is positive definite. Then, under the same regularity conditions in Theorem 1, $n^{1/2}(\hat{\beta}_p - \beta_0)$ converges in distribution to a zero mean multivariate normal distribution with variance-covariance matrix $\tilde{\Gamma}^{-1} = (\Sigma + \tilde{B}\tilde{Q}^{-1}\tilde{B}')^{-1}$, provided $\tilde{\Gamma}$ is non-singular, where \tilde{B} and \tilde{Q} are specified in the Appendix.

Theorem 3 implies that the extended double empirical likelihood estimator $\hat{\beta}_p$ is asymptotically more efficient than the maximum partial likelihood estimator $\hat{\beta}_{PL}$. Moreover, it is easy to see that, $\hat{\theta} = (\hat{\beta}, \hat{\xi}, \hat{\nu}, \hat{\alpha})$ is the maximizer of $\ell(\beta, \xi, \nu, \alpha, \rho \equiv 1)$. In the proof of Theorem 3, we also show that $\hat{\beta}_p$ is less efficient than $\hat{\beta}$. To test if the same baseline hazard function is shared by the individual-level data and the aggregate data, we consider the empirical log-likelihood ratio statistic

$$\tilde{R} = 2\left\{ \sup_{\beta, \xi, \nu, \alpha, \rho} \tilde{\ell}(\beta, \xi, \nu, \alpha, \rho) - \sup_{\beta, \xi, \nu, \alpha} \tilde{\ell}(\beta, \xi, \nu, \alpha, 1) \right\}.$$

Under minor regularity conditions and the null hypothesis that $\rho_0 = 1$, the empirical log-likelihood ratio \tilde{R} converges in distribution to a χ^2 random variable with 1 degrees of freedom as $n \rightarrow \infty$. The proof closely follows that for Theorem 2, and thus is omitted.

5. Numerical Studies

5.1 Monte Carlo Simulations

We conducted two sets of Monte Carlo simulations to examine the finite-sample performance of the proposed methods. In both simulation studies, we generated X_1 from the standard normal random variable and X_2 from a Bernoulli distribution with $\text{pr}(X_2 = 1) = \text{pr}(X_2 = 0) = 0.5$. The survival time T in the individual-level data was generated from the proportional hazards models (A) $\lambda(t | X_1, X_2) = \lambda(t) \exp(\beta_1 X_1 + \beta_2 X_2)$ with $(\beta_1, \beta_2) = (-0.5, 0.5)$, and (B) $\lambda(t | X_1, X_2) = \lambda(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$ with $(\beta_1, \beta_2, \beta_3) = (-0.5, 1, -0.5)$, where we set $\lambda(t) = 2t$ for both models. The censoring time C was generated from an uniform distribution so that the censoring rate was approximately 0%, 30%, and 50%. In each simulation, we generated 1000 datasets, each with a sample size of $n = 100$ and $n = 400$.

In the first set of simulations, we considered the case where the individual-level and the aggregate data share the same survival time model, that is, $\rho = 1$ in model (3). We derived the auxiliary survival information at $t = 0.5$ for subgroups $\Omega_1 = \{(X_1, X_2) : X_1 \leq 0, X_2 = 0\}$ and $\Omega_2 = \{(X_1, X_2) : X_1 > 0, X_2 = 0\}$ under the assumed Cox model. This specification aims to mimic the situation where in a randomized clinical trial we exploit the information about the 6-month survival probabilities in the standard-of-care control group ($X_2 = 0$) and where X_1 is a baseline risk factor. The 6-month survival probabilities for the two subgroups are approximately 0.68 and 0.84 under Models (A) and (B).

Tables 1 and 2 summarize the empirical bias and empirical standard deviation of the maximum partial likelihood estimator $\hat{\beta}_{PL}$, the double empirical likelihood estimator $\hat{\beta}$ and

the extended double empirical likelihood estimator $\hat{\beta}_\rho$ that allows for a different baseline hazard function for the aggregate data. All three estimators are close to their estimands under Models (A) and (B). Compared with the maximum partial likelihood estimator $\hat{\beta}_{PL}$, the two double empirical likelihood estimators $\hat{\beta}$ and $\hat{\beta}_\rho$ enjoy substantial efficiency gains. Under Model (A), the relative efficiency ranges from 4.86 to 9.66 for the estimated coefficient of baseline risk factor and from 1.02 to 1.84 for the estimated treatment effect. Under Model (B), the relative efficiency in the treatment-covariate interaction ranges from 1.50 to 2.13, suggesting that the use of proposed methods can substantially reduce the sample size requirement by about 18%–58% in investigating treatment heterogeneity. As expected, $\hat{\beta}_\rho$ is slightly less efficient than $\hat{\beta}$ and the estimated value of ρ is close to 1 using the extended double empirical likelihood approach. We also reported the estimated baseline cumulative hazard function at $t = 0.3$ and 0.7 . As expected, the double empirical likelihood approach enjoys a substantial gain in efficiency in the estimation of $\Lambda(t)$, while the efficiency gain for the extended approach is minimal because it allows for a different baseline risk in the aggregate data. Finally, our simulations (results not shown) also show that the efficiency gain increases with the number of constraints.

In the second set of simulations, we assume that the auxiliary survival information is derived from the Cox model with a different baseline hazard function $\lambda^*(t) = 1.5\lambda(t)$, that is, we set $\rho = 1.5$ in model (3). The 6-month survival probabilities for the same subgroups are approximately 0.57 and 0.77 under both Model (A) and Model (B). Tables 3 and 4 give the summary statistics of the simulation results. As expected, the double empirical likelihood estimator $\hat{\beta}$ which assumes consistency between the individual-level data and the aggregate data yields biased estimates. On the other hand, the extended double empirical likelihood estimator $\hat{\beta}_\rho$ performs very well in terms of bias and efficiency gain, and the estimated value of ρ is very close to its true value 1.5.

5.2 Data Example

5.2.1 Example 1: Prostate cancer study—To demonstrate how the proposed methods can improve efficiency by incorporating auxiliary survival information with the individual-level data, we investigated the comparative effectiveness of two modes of androgen deprivation therapy (ADT), intermittent and continuous ADT (IADT and CADT), on survival outcomes in men with advanced prostate cancer. Continuous ADT has been the conventional palliative approach in the U.S. for the control of advanced prostate cancer, and intermittent ADT has been proposed as an alternative to CADT for the potential advantages of improved quality of life, reduced cost, and reduced risk of side effects. However, it remains unclear whether IADT has a survival benefit comparable to CADT. While age and prostate specific antigen (PSA) level have been shown to be important prognostic factors in advanced prostate cancer, it remains unclear whether the comparative effectiveness of IADT versus CADT differs by PSA level at diagnosis after adjustment for men's age at diagnosis. Although clinical trials and meta-analysis have been conducted to investigate this issue, the answer remains inconclusive due to lack of statistical power needed for subgroup analysis (Tsai et al., 2013; Hussain et al., 2013).

We used the linked Surveillance, Epidemiology and End Result (SEER)-Medicare dataset in this example. The SEER-Medicare dataset matches incident cancer patients identified from SEER registries to their data from Medicare, the major insurer in the U.S. for people 65 years and older, to obtain longitudinal inpatient and outpatient treatment information and determine patients' receipt of ADT. Our study population was defined as men 66 years and older diagnosed with advanced prostate cancer anytime from January 1, 2004 to December 31, 2009 who received ADT anytime during 2004 to 2010. After excluding men who did not have continuous Medicare coverage from 2004 to 2010, who did not receive either CADT or IADT, and who were missing diagnosis age and PSA measures, a total of 4548 patients were included in this illustrative data example. Among these patients, 71.7% of them received continuous ADT treatment and 45.0% died before December 31, 2010. The median age at diagnosis was 75 years and the median PSA level was 16.2 ng/mL. To illustrate the proposed estimation procedure, we randomly selected 300 cases from the complete dataset, that is, 6.6% of the available 4548 cases. The selected subset has a median age at diagnosis of 75 years and a median PSA level of 17.1 ng/mL. Additionally, 68.7% of the cases in the subset received continuous ADT treatment and 46% died before December 31, 2010.

We fitted a Cox model to analyze survival after diagnosis of prostate cancer using diagnosis age (66-75, 76-80, and > 80) and PSA level (0-40 and > 40) as categorical covariates. Table 5 shows the maximum partial likelihood estimators and their standard errors by using the complete dataset with 4548 cases and the subset with 300 selected cases. We also applied the two double empirical likelihood estimators to synthesize information from the subset data with three sets of auxiliary survival information at the 5-year landmark: (I) The five year survival probabilities were 70.5% and 34.5% for IADT-treated patients whose PSA levels were below and above 40 ng/mL at diagnosis; (II) The five year survival probabilities were 64.2% and 20.1% for CADT-treated patients whose PSA level were below and above 40 ng/mL at diagnosis; (III) The aforementioned survival information are available in both IADT and CADT groups. Note that these auxiliary survival probabilities were derived by applying the Kaplan-Meier estimator to the corresponding subgroups from the complete dataset. To obtain the standard errors for the estimated regression coefficients, we adopted a nonparametric bootstrap method by sampling 300 subjects with replacements from the subset data. The resampling procedure was repeated 1000 times, and the standard errors were estimated with the standard deviation of the 1000 estimates.

As expected, incorporating auxiliary survival information in the data analysis using the proposed methods yields smaller standard errors than the conventional partial likelihood approach using the subset data, and the largest efficiency gains are observed in Scenario III where a greater amount of information is incorporated. Of note, the estimator $\hat{\beta}_\rho$ that allows for a different baseline hazard function for the aggregate data gives almost identical results as the one that assumes consistency between the two populations. As expected, the estimated value of ρ is very close to 1 in all scenarios because the selected subset is a random sample of the complete data. Interestingly, the analysis with complete data shows the comparative effectiveness of IADT versus CADT to differ significantly by patients' PSA level after adjusting for age at diagnosis, yet the significance disappears when applying conventional analysis in the subset data but is revealed in our proposed approaches with auxiliary survival

information, thus highlighting the practical value of employing a more efficient methodology.

5.2.2 Example 2: Pancreatic cancer study—We now apply the extended double empirical likelihood estimator to another example where the auxiliary survival information may be inconsistent with individual-level data. We analyzed data from a pancreatic cancer study conducted at the Johns Hopkins Hospital to study risk factors affecting survival following pancreatectomy. Pancreatic ductal adenocarcinoma (PDAC), the most common histological subtype of pancreatic malignancy, is a very resilient disease with a very poor prognosis. To date, radical surgical resection remains the only treatment for PDAC that offers clinical benefit in terms of overall survival. Unfortunately, less than 20% of pancreatic cancer patients have surgically resectable disease at the time of diagnosis, and the majority of resected pancreatic cancer recurs within 5 years. Despite advances in the treatment of cancer during the past few decades, improvement in long-term survival of PDAC patients has been modest. The major risk factors influencing survival after pancreatic cancer surgery are tumor characteristics. Favorable prognostic factors include negative resection margin, negative lymph node, and absence of perineural invasion.

This data example is from a retrospective cohort study of 209 consecutive patients who had surgical resection of PDAC and follow-up at the Johns Hopkins Hospital from January 9, 1998 to June 13, 2007. Thorough chart reviews were conducted to ascertain patient's demographics and results of laboratory test, clinical and pathological exams. Treatment data were collected from the electronic medical records. Disease recurrence was determined clinically through imaging studies (computed tomography, positron emission tomography) or pathological diagnosis (CT-guided biopsy, wedge resection or lobectomy). All-cause and cancer-specific deaths and dates of death were determined by a combined review of clinical follow-up information, Social Security Death Index, and the National Cancer Database.

We fitted a Cox model to evaluate the effects of presence of lymph nodes, positive resection margins, presence of perineural invasion (PNI), age at surgery (≤ 65 and > 65), and gender on overall survival. Table 6 shows the estimated covariate effects using the partial likelihood method. We also applied the double empirical likelihood estimators $\hat{\beta}_\rho$ to synthesize three sets of auxiliary survival information reported in Cameron et al. (2006): (I) the three-year survival probabilities for node-negative and node-positive patients were 40% and 26%, respectively. (II) the three-year survival probabilities for margin-negative and margin-positive patients were 35% and 20%, respectively. (III) All the four survival probabilities given in (I) and (II). These survival probabilities were estimated from 1000 consecutive pancreatectomies performed by a single surgeon between March 1969 and May 2003.

We only reported the results of $\hat{\beta}_\rho$ in Table 6 because the baseline hazard function in the aggregate data is expected to be different due to potential differences in patients' characteristics. As before, incorporating auxiliary survival information in the data analysis using the proposed methods yields smaller standard errors than the conventional partial likelihood approach, and the largest efficiency gains are observed in Scenario III where a greater amount of information is incorporated. Note that the effect of positive lymph nodes is only marginally significant when applying the partial likelihood method, but becomes

statistically significant when combining with the auxiliary survival information. The estimated value of ρ is 0.8 and is significantly lower than 1, indicating that the baseline risk in patients in the study reported by Cameron et al. (2006) is lower than that in patients in our clinical study.

6. Remarks

In this paper, we have proposed two double empirical likelihood approaches to synthesize information from both patient-level right-censored survival data and the auxiliary survival information. We first construct an efficient estimation procedure by imposing consistency between the individual-level data and the aggregate data, and then extend the estimation procedure to allow for potential differences in the survival models for the two data sources. Many researchers, including Imbens and Lancaster (1994), Hellerstein and Imbens (1999), and Chaudhuri et al. (2008), have considered estimation of the general/generalized linear models by imposing additional moment restrictions. Most of the existing work deals with complete data, while this paper considers estimation of the semiparametric Cox proportional hazards model using right-censored survival data by incorporating a system of nonlinear constraints. The simulation studies show large gains in efficiency by incorporating marginal moments from published survival probability information. We believe that the proposed methodologies will have a significant impact on the practice of meta-analysis.

The proposed approaches are more flexible than the conventional meta-analysis in the sense that they can automatically combine survival information for different subgroups and the information may be derived from different studies. For example, one study may publish survival probabilities for different age groups, while the other study may publish survival probabilities for different disease stages. The proposed double empirical likelihood methods provide a unified framework to incorporate all the available information in the form of nonlinear constraints.

The sample size of the external data source, such as disease registries, is in general much larger than that of the individual-level data. As a result, the variability in the published survival information is usually negligible compared to the variability in the parameter estimates using the individual-level data. In the case where the variability is not negligible, a higher-order Taylor expansion needs to be employed to summarize the auxiliary survival information as estimating equations. This will be explored elsewhere.

Acknowledgments

The authors thank the editor, the associate editor, and two referees for their constructive comments. Thanks are due to Dr. Lei Zheng for kindly sharing the pancreatic cancer data. The authors also acknowledge the efforts of the Applied Research Program, NCI; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. This work is supported by National Institutes of Health.

Appendix: Large-Sample Properties

Proof of Asymptotic Normality for U_0 . Define $N_i(t) = \int_0^t I(Y_j \geq t)$ and $M_i(t) = N_i(t) - \Lambda_0(t) \exp(\beta_0' X_i)$. Thus we have

$$\begin{aligned} U_{01} &= \sum_{i=1}^n \int_0^\infty \left\{ X_i - \frac{S^{(1)}(u, \beta_0)}{S^{(0)}(u, \beta_0)} \right\} dM_i(u), \\ U_{02} &= \sum_{i=1}^n \psi(X_i, \beta_0, \alpha_0), \\ U_{03} &= \sum_{i=1}^n \int_0^\infty \frac{I(u \leq t^*)}{S^{(0)}(u, \beta_0)} dM_i(u), \\ U_{04} &= 0. \end{aligned}$$

For convenience, let $s^{(k)}(t, \beta)$ denote the limit of $S^{(k)}(t, \beta)$, that is, $s^{(k)}(t, \beta) = \lim_{n \rightarrow \infty} S^{(k)}(t, \beta)$ for $k \in \{0, 1, 2\}$. Because $M_i(t)$ is a local square-integrable martingale and $X_i - S^{(1)}(u, \beta_0)/S^{(0)}(u, \beta_0)$ and $I(u \leq t^*)/S^{(0)}(u, \beta_0)$ are both predictable quadratic variation processes, we have $E(U_{01}) = \mathbf{0}$, $E(U_{03}) = 0$,

$$\Sigma \equiv \text{var}(n^{-1/2}U_{01}) = \int_0^\infty \left[s^{(2)}(u, \beta_0) - \{s^{(0)}(u, \beta_0)\}^{-1} \{s^{(1)}(u, \beta_0)\}^{\otimes 2} \right] d\Lambda_0(u),$$

$$K \equiv \text{var}(n^{-1/2}U_{03}) = \int_0^{t_0} \{s^{(0)}(u, \beta_0)\}^{-1} d\Lambda_0(u),$$

and

$$\text{cov}(U_{01}, U_{03}) = \sum_{i=1}^n \int_0^\infty E \left[\left\{ X_i - \frac{S^{(1)}(u, \beta_0)}{S^{(0)}(u, \beta_0)} \right\} \frac{I(u \leq t^*)}{S^{(0)}(u, \beta_0)} \exp(\beta_0' X_i) I(Y_i \geq u) \right] d\Lambda_0(t) = 0.$$

Moreover, by double expectation, it can be shown that $E(U_{01}'U_{02}) = \mathbf{0}$ and $E(U_{03}U_{02}) = \mathbf{0}$. Define $J = \text{var}(n^{-1/2}U_{02})$. It is easy to see that $|\psi_k(X, \beta_0, \alpha_0)| \leq 2$ for $k = 1, \dots, K$. Thus, by the martingale central limit theorem and the classic central limit theorem, U_0 converges in distribution to a zero mean multivariate normal distribution with the variance-covariate matrix Ω as $n \rightarrow \infty$.

Proof of Theorem 1. Arguing as in the proof of Lemma 1 in Qin and Lawless (1994), under some regularity conditions, we can show that the full constraint empirical likelihood attains the maximum at some point $(\hat{\beta}, \hat{\alpha})$ in the interior of the ball $\{(\beta, \alpha) : \|(\beta, \alpha) - (\beta_0, \alpha_0)\| \leq n^{-1/3}\}$ with probability 1. Next, straightforward algebra yields

$$-n^{-1}E\left(\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) = \begin{pmatrix} \sum & \mathbf{B} \\ -\mathbf{B}' & \mathbf{Q} \end{pmatrix}, \quad (4)$$

Where

$$\begin{aligned} \mathbf{B} &= -n^{-1}\left\{E\left(\frac{\partial U_1}{\partial \xi}\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, E\left(\frac{\partial U_1}{\partial \nu}\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, E\left(\frac{\partial U_1}{\partial \alpha}\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\} \\ &= \left[E\left\{\frac{\partial \psi}{\partial \boldsymbol{\beta}}(X_1, \beta_0, \alpha_0)\right\}, -\int_0^{t_0} \frac{s^{(1)}(u, \boldsymbol{\beta}_0)}{s^{(0)}(u, \boldsymbol{\beta}_0)} d\Lambda_0(u), 0\right] \\ &:= (\mathbf{B}_1, \mathbf{B}_2, \mathbf{0}), \end{aligned}$$

and, similarly,

$$\mathbf{Q} = \begin{pmatrix} \mathbf{J} & \mathbf{0} & \mathbf{H} \\ \mathbf{0} & \mathbf{K} & 1 \\ \mathbf{H}' & 1 & 0 \end{pmatrix},$$

with $\mathbf{H} = -n^{-1}E\{\boldsymbol{\psi}' \alpha(X_1, \boldsymbol{\beta}_0, \alpha_0)\}$. It follows from $\mathbf{H}'\mathbf{K}\mathbf{H} + \mathbf{J}$ being positive definite that \mathbf{Q} is negative definite. By singular value decomposition, one can derive

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & -\mathbf{J}^{-1}\mathbf{H} \\ \mathbf{0} & 1 & -\mathbf{K}^{-1} \\ \mathbf{0} & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{J}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -(\mathbf{H}'\mathbf{J}^{-1}\mathbf{H} + \mathbf{K}^{-1})^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \\ -\mathbf{H}'\mathbf{J}^{-1} & -\mathbf{K}^{-1} & 1 \end{pmatrix}.$$

Define $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} + \mathbf{B}\mathbf{Q}^{-1}\mathbf{B}'$, then

$$\begin{pmatrix} \sum & \mathbf{B} \\ -\mathbf{B}' & \mathbf{Q} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Gamma}^{-1} & -\boldsymbol{\Gamma}^{-1}\mathbf{B}\mathbf{Q}^{-1} \\ \mathbf{Q}^{-1}\mathbf{B}'\boldsymbol{\Gamma}^{-1} & (\mathbf{Q} + \mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1} \end{pmatrix}.$$

By definition, $\hat{\boldsymbol{\theta}}$ is the solution to $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Write $\mathbf{U} = (\mathbf{U}'_{01}, \mathbf{U}'^{*})'$ with $\mathbf{U}^* = (\mathbf{U}'_{02}, \mathbf{U}_{03}, 0)'$. Then, by (4) and a Taylor series expansion, we have

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \begin{pmatrix} \sum & \mathbf{B} \\ -\mathbf{B}' & \mathbf{Q} \end{pmatrix}^{-1} \begin{pmatrix} n^{-1/2}\mathbf{U}_{01} \\ n^{-1/2}\mathbf{U}^* \end{pmatrix} + o_p(1). \quad (5)$$

Thus we establish the asymptotic representation $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{\Gamma}^{-1}(n^{-1/2}\mathbf{U}_{01}) - \boldsymbol{\Gamma}^{-1}\mathbf{B}\mathbf{Q}^{-1}(n^{-1/2}\mathbf{U}^*) + o_p(1)$. Because \mathbf{U}_{01} and \mathbf{U}^* are orthogonal, the asymptotic variance of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is given by

$$\begin{aligned}
 & \Gamma^{-1} \sum \Gamma^{-1} \\
 & + \Gamma^{-1} B Q^{-1} \begin{pmatrix} J & 0 & 0 \\ 0 & K & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} B' \Gamma^{-1} \\
 & = \Gamma^{-1} (\Gamma \\
 & \quad - B Q^{-1} B') \Gamma^{-1} + \Gamma^{-1} B Q^{-1} \begin{pmatrix} J & 0 & 0 \\ 0 & K & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} B' \Gamma^{-1} \\
 & = \Gamma^{-1} - \Gamma^{-1} B Q^{-1} \begin{pmatrix} 0 & 0 & H \\ 0 & 0 & 1 \\ H' & 1 & 0 \end{pmatrix} Q^{-1} B' \Gamma^{-1}.
 \end{aligned}$$

Let $G = H' J^{-1} H + K^{-1}$. It can be verified that

$$Q^{-1} \begin{pmatrix} 0 & 0 & H \\ 0 & 0 & 1 \\ H' & 1 & 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} 0 & 0 & J^{-1} H G^{-1} \\ 0 & 0 & K^{-1} G^{-1} \\ G^{-1} H' J^{-1} & G^{-1} K^{-1} & -2G^{-1} \end{pmatrix}.$$

Straightforward algebra yields

$$B Q^{-1} \begin{pmatrix} 0 & 0 & H \\ 0 & 0 & 1 \\ H' & 1 & 0 \end{pmatrix} Q^{-1} B' = 0. \tag{6}$$

Hence the asymptotic variance of $\{n^{1/2}(\hat{\beta} - \beta_0)\}$ is given by $\Gamma^{-1} = (\Sigma + B Q^{-1} B')^{-1} \Sigma^{-1}$.

Proof of Asymptotic Normality for $\hat{\Lambda}(t)$. To establish the asymptotic normality of $\hat{\Lambda}(t)$, we first note that the double empirical likelihood-based estimator for the baseline cumulative hazard function can be reexpressed as $\hat{\Lambda}(t, \hat{\beta}, \hat{\nu})$, where

$$\hat{\Lambda}(t, \beta, \nu) = \int_0^t \frac{n^{-1} \sum_{i=1}^n dN_i(u)}{S^{(0)}(u, \beta) + \nu I(u \leq t^*)}.$$

Note that $\hat{\Lambda}$ defines a functional of two empirical processes $n^{-1} \sum_{i=1}^n dN_i(u)$ and $S^{(0)}(u, \beta) + \nu I(u \leq t^*)$, and the mapping defined by $\hat{\Lambda}$ is compactly differentiable. Define $F^u(t) = E\{N_1(t)\}$ and let $\Lambda_0(t)$ be the true baseline cumulative hazard function. It can be shown that $\hat{\Lambda}(t, \beta_0, 0)$ converges almost surely to $\int_0^t dF^u(u) / s^{(0)}(u, \beta_0) = \Lambda_0(t)$.

A Taylor series expansion of $\hat{\Lambda}(t, \hat{\beta}, \hat{\nu})$ about $(\beta_0, 0)$ yields

$$\hat{\Lambda}(t, \hat{\beta}, \hat{\nu}) = \hat{\Lambda}(t, \beta_0, 0) + \left\{ \frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \beta} \Big|_{(\beta_t^*, 0)} \right\}' (\hat{\beta} - \beta_0) + \frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \nu} \Big|_{(\beta_0, \nu_t^*)} (\hat{\nu} - 0), \tag{7}$$

where (β_t^*, ν_t^*) lies between $(\hat{\beta}, \hat{\nu})$ and $(\beta_0, 0)$. Given the consistency of $\hat{\theta}$ and by the Glivenko-Cantelli theorem, we can show that almost surely

$$\frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \beta} \Big|_{(\beta_t^*, 0)} \rightarrow E \left\{ \frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \beta} \Big|_{(\beta_0, 0)} \right\} = - \int_0^t \frac{s^{(1)}(u, \beta_0)}{s^{(0)}(u, \beta_0)^2} dF^u(u)$$

and

$$\frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \nu} \Big|_{(\beta_0, \nu_t^*)} \rightarrow E \left\{ \frac{\partial \hat{\Lambda}(t, \beta, \nu)}{\partial \nu} \Big|_{(\beta_0, 0)} \right\} = - \int_0^t \frac{I(u \leq t^*)}{s^{(0)}(u, \beta_0)^2} dF^u(u)$$

as $n \rightarrow \infty$. Moreover, applying the functional delta method to $\hat{\Lambda}(t, \beta_0, 0)$ yields

$$\hat{\Lambda}(t, \beta_0, 0) - \Lambda_0(t) = \int_0^t \frac{n^{-1} \sum_{i=1}^n dN_i(u)}{s^{(0)}(u, \beta_0)} - \int_0^t \frac{S^{(0)}(u, \beta) + \nu I(u \leq t^*)}{s^{(0)}(u, \beta_0)^2} dF^u(u) + o_p(n^{-1/2}). \tag{8}$$

It follows from (5), (7), and (8) that $n^{1/2}\{\hat{\Lambda}(t, \hat{\beta}, \hat{\nu}) - \Lambda_0(t)\}$ is asymptotically equivalent to a sum of i.i.d. monotone processes with bounded second moments, and thus, following example 2.11.16 of van der Vaart and Wellner (1996), converges weakly to a mean zero gaussian process.

Proof of Theorem 2. Note that $(\tilde{\beta}, \tilde{\alpha})$ is the solution to $U_1(\beta, \mathbf{0}, 0, \alpha) = 0$, and $U_3(\beta, \mathbf{0}, 0, \alpha) = 0$. Hence $(\tilde{\beta}, \tilde{\alpha})$ has the following asymptotic representation

$$n^{1/2} \begin{pmatrix} \tilde{\beta} - \beta_0 \\ \tilde{\alpha} - \alpha_0 \end{pmatrix} = n^{-1/2} \begin{pmatrix} \Sigma & 0 \\ -B_2' & 1 \end{pmatrix}^{-1} \begin{pmatrix} U_{01} \\ U_{03} \end{pmatrix} + o_p(1) = n^{-1/2} \begin{pmatrix} \Sigma^{-1} & 0 \\ B_2' \Sigma^{-1} & 1 \end{pmatrix} \begin{pmatrix} U_{01} \\ U_{03} \end{pmatrix} + o_p(1).$$

It follows from (5) and $n^{1/2}(\mathbf{H}' \hat{\xi} + \hat{\nu}) = n^{-1/2} U_{04} + o_p(1) = o_p(1)$ that

$$n^{1/2} \begin{pmatrix} \hat{\beta} - \tilde{\beta} \\ \hat{\xi} \\ \hat{\nu} \\ \hat{\alpha} - \tilde{\alpha} \end{pmatrix} = n^{-1/2} \begin{pmatrix} \Sigma^{-1}(-B_1 + B_2 H') \\ \mathbf{I} \\ -H' \\ -B_2' \Sigma^{-1} B_1 + (B_2' \Sigma^{-1} B_2 + K) H' \end{pmatrix} \hat{\xi} + o_p(1).$$

Expanding the partial likelihood $l(\tilde{\beta}, \mathbf{0}, 0, \tilde{\alpha})$ at $\hat{\theta}$ yields

$$\begin{aligned} R &= 2\{l(\tilde{\beta}, \mathbf{0}, 0, \tilde{\alpha}) - l(\hat{\beta}, \hat{\xi}, \hat{\nu}, \hat{\alpha})\} \\ &= n(\hat{\beta}' - \tilde{\beta}', \hat{\xi}', \hat{\nu}, \hat{\alpha} - \tilde{\alpha}) \begin{pmatrix} \sum & B \\ -B' & Q \end{pmatrix} (\hat{\beta}' - \tilde{\beta}', \hat{\xi}', \hat{\nu}, \hat{\alpha} - \tilde{\alpha})' + o_p(1) \\ &= n^{1/2} \hat{\xi}' \{(B_1' - HB_2') \Sigma^{-1} (B_1 - B_2 H') + J + H K H'\} n^{1/2} \hat{\xi} + o_p(1). \end{aligned}$$

Define $W = (B_1' - HB_2') \Sigma^{-1} (B_1 - B_2 H') + J + H K H'$ and denote the asymptotic variance of $n^{1/2} \xi$ by

$$V_{\xi} = (\mathbf{0}, I, 0, 0) \begin{pmatrix} \sum & B \\ -B' & Q \end{pmatrix}^{-1} \Omega \begin{pmatrix} \sum & B \\ -B' & Q \end{pmatrix}^{-1} (\mathbf{0}, I, 0, 0)'$$

By tedious algebra, we can show that $VWVWV = VWV$ and $\text{rank}(WV) = K$, that is, the quadratic form of the asymptotically normally distributed random variable ξ satisfies the conditions of Ogasawara-Takahashi Theorem (Rao, 1973, page 188). Thus we prove that R converges in distribution a χ^2 distribution with K degrees of freedom as $n \rightarrow \infty$.

Proof of Theorem 3. The proof of Theorem 3 closely follows that of Theorem 1. Hence we only highlight their differences. It is easy to see that $\tilde{U}_{01} = U_{01}$,

$$\tilde{U}_{02} = \sum_{i=1}^n \tilde{\psi}(X_i, \beta_0, \alpha_0, \rho_0),$$

$\tilde{U}_{03} = U_{03}$, $\tilde{U}_{04} = U_{04} = 0$, $\tilde{U}_{05} = U_{05} = 0$. Arguing as before,

$\tilde{U}_0 = (\tilde{U}'_{01}, \tilde{U}'_{02}, \tilde{U}'_{03}, \tilde{U}'_{04}, \tilde{U}'_{05})'$ converges in distribution to a zero mean multivariate normal distribution with the variance-covariate matrix $\tilde{\Omega}$ as $n \rightarrow \infty$, where

$$\tilde{\Omega} = \begin{pmatrix} \sum & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & K & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\tilde{J} = \text{var}(n^{-1/2} \tilde{U}_{02})$.

Define the matrices

$$\Phi = \begin{pmatrix} \tilde{Q} & L \\ L' & 0 \end{pmatrix},$$

where

$$\begin{aligned} L &= -n^{-1} \left\{ E \left(\frac{\partial \tilde{U}_5}{\partial \xi} \right) \Big|_{\theta=\tilde{\theta}_0}, E \left(\frac{\partial \tilde{U}_5}{\partial \nu} \right) \Big|_{\theta=\tilde{\theta}_0}, E \left(\frac{\partial \tilde{U}_5}{\partial \alpha} \right) \Big|_{\theta=\tilde{\theta}_0} \right\}' \\ &= \left[-E \left\{ \frac{\partial \tilde{\psi}}{\partial \rho} (X_1, \beta_0, \alpha_0, \rho_0) \right\}, 0, 0 \right]' := (L', 0, 0)', \end{aligned}$$

$$\tilde{Q} = \begin{pmatrix} J & 0 & \tilde{H} \\ 0 & K & 1 \\ \tilde{H}' & 1 & 0 \end{pmatrix},$$

and $\tilde{H} = -n^{-1} E \{ \tilde{\psi}' / a(X_1, \beta_0, \alpha_0, \rho_0) \}$. Straightforward algebra yields

$$-n^{-1} E \left(\frac{\partial \tilde{U}}{\partial \theta} \Big|_{\theta=\tilde{\theta}_0} \right) = \begin{pmatrix} \Sigma & \kappa \\ -\kappa' & \Phi \end{pmatrix},$$

where $\kappa = (\tilde{B}', 0)$ with $\tilde{B} = \left[E \left\{ \frac{\partial \tilde{\psi}}{\partial \beta} (X_1, \beta_0, \alpha_0, \rho_0) \right\}, B_2, 0 \right]$.

By a Taylor series expansion, we have

$$n^{1/2}(\hat{\theta}_\rho - \tilde{\theta}_0) = \begin{pmatrix} \Sigma & \kappa \\ -\kappa' & \Phi \end{pmatrix}^{-1} \begin{pmatrix} n^{-1/2} \tilde{U}_{01} \\ n^{-1/2} \tilde{U}^* \end{pmatrix} + o_p(1),$$

where $\tilde{U}^* = (\tilde{U}'_{02}, \tilde{U}'_{03}, 0, 0)'$. Define $\tilde{\Gamma} = \Sigma + \kappa \Phi^{-1} \kappa'$. Arguing as before, we establish the asymptotic representation $n^{1/2}(\hat{\beta}_\rho - \beta_0) = \tilde{\Gamma}^{-1} (n^{-1/2} \tilde{U}_{01}) - \tilde{\Gamma}^{-1} \kappa \Phi^{-1} (n^{-1/2} \tilde{U}^*) + o_p(1)$.

Because \tilde{U}_{01} and \tilde{U}^* are orthogonal, the asymptotic variance of $n^{1/2}(\hat{\beta}_\rho - \beta_0)$ is given by

$$\begin{aligned}
 & \tilde{\Gamma}^{-1} \sum \tilde{\Gamma}^{-1} \\
 & + \tilde{\Gamma}^{-1} \kappa \Phi^{-1} \begin{pmatrix} J & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & K & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi^{-1} \kappa' \tilde{\Gamma}^{-1} \\
 & = \tilde{\Gamma}^{-1} \\
 & (\tilde{\Gamma} - \kappa \Phi^{-1} \kappa') \tilde{\Gamma}^{-1} + \tilde{\Gamma}^{-1} \kappa \Phi^{-1} \begin{pmatrix} \tilde{J} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & K & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi^{-1} \kappa' \tilde{\Gamma}^{-1} \\
 & = \tilde{\Gamma}^{-1} - \tilde{\Gamma}^{-1} \kappa \Phi^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{H} & L_1 \\ \mathbf{0}' & \mathbf{0} & 1 & \mathbf{0} \\ \tilde{H}' & 1 & \mathbf{0} & \mathbf{0} \\ L_1' & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi^{-1} \kappa' \tilde{\Gamma}^{-1}.
 \end{aligned}$$

Arguing as in the proof of Theorem 1, we can show that

$$L' \tilde{Q}^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{H} \\ \mathbf{0}' & \mathbf{0} & 1 \\ \tilde{H}' & 1 & \mathbf{0} \end{pmatrix} \tilde{Q}^{-1} L = \mathbf{0}, \quad \text{and} \quad \tilde{B} \tilde{Q}^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{H} \\ \mathbf{0}' & \mathbf{0} & 1 \\ \tilde{H}' & 1 & \mathbf{0} \end{pmatrix} \tilde{Q}^{-1} L = \mathbf{0}.$$

Together with (7) and

$$\begin{aligned}
 \Phi^{-1} & = \begin{pmatrix} \tilde{Q} & L \\ L' & \mathbf{0} \end{pmatrix}^{-1} \\
 & = \begin{pmatrix} \tilde{Q}^{-1} - \tilde{Q}^{-1} L (L' \tilde{Q}^{-1} L)^{-1} L' \tilde{Q}^{-1} & \tilde{Q}^{-1} L (L' \tilde{Q}^{-1} L)^{-1} \\ (L' \tilde{Q}^{-1} L)^{-1} L' \tilde{Q}^{-1} & -(L' \tilde{Q}^{-1} L)^{-1} \end{pmatrix}
 \end{aligned}$$

we have

$$\begin{aligned}
 & \kappa \Phi^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{H} & L_1 \\ \mathbf{0}' & \mathbf{0} & 1 & \mathbf{0} \\ \tilde{H}' & 1 & \mathbf{0} & \mathbf{0} \\ L_1' & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi^{-1} \kappa' \\
 & = (\tilde{B}, \mathbf{0}) \begin{pmatrix} \tilde{Q} & L \\ L' & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{H} & L_1 \\ \mathbf{0}' & \mathbf{0} & 1 & \mathbf{0} \\ \tilde{H}' & 1 & \mathbf{0} & \mathbf{0} \\ L_1' & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{Q} & L \\ L' & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{B}' \\ \mathbf{0} \end{pmatrix} = \mathbf{0}.
 \end{aligned}$$

Hence we prove that $\text{var}\{n^{1/2}(\hat{\beta}_\rho - \beta_0)\} = \tilde{\Gamma}^{-1} = (\Sigma + \kappa\Phi^{-1}\kappa')^{-1} \Sigma^{-1}$, where Σ^{-1} is the variance-covariance matrix for the maximum partial likelihood estimator $\hat{\beta}_{PL}$. Moreover, because $\kappa\Phi^{-1}\kappa' = \tilde{B}\tilde{Q}^{-1}\tilde{B}' - (\tilde{B}\tilde{Q}^{-1}L)(L'\tilde{Q}^{-1}L)^{-1}(\tilde{B}\tilde{Q}^{-1}L)'$, we also prove that $\hat{\beta}_\rho$ is less efficient than $\hat{\beta}$ which is the maximum empirical likelihood estimator for β when the true parameter value $\rho_0 = 1$ is known.

References

- Breslow NE. Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B.* 1972; 34:216–217.
- Cameron JL, Riall TS, Coleman J, Belcher KA. One Thousand Consecutive Pancreaticoduodenectomies. *Annals of Surgery.* 2006; 244:10. [PubMed: 16794383]
- Chaudhuri S, Handcock MS, Rendall MS. Generalized Linear Models Incorporating Population Level Information: An Empirical-Likelihood-Based Approach. *Journal of the Royal Statistical Society, Series B.* 2008; 70:311–328.
- Chen J, Qin J. Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information. *Biometrika.* 1993; 80:107–116.
- Chen J, Sitter RR, Wu C. Using Empirical Likelihood Methods to Obtain Range Restricted Weights in Regression Estimators for Surveys. *Biometrika.* 2002; 89:230–237.
- Chen J, Wu C. Estimation of Distribution Function and Quantiles Using the Model-Calibrated Pseudo Empirical Likelihood Method. *Statistica Sinica.* 2002; 12:1223–1239.
- Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B.* 1972; 34:187–220.
- DiCiccio T, Hall P, Romano J. Empirical Likelihood is Bartlett-Correctable. *The Annals of Statistics.* 1991; 19:1053–1061.
- Hellerstein JK, Imbens GW. Imposing Moment Restrictions from Auxiliary Data by Weighting. *Review of Economics and Statistics.* 1999; 81:1–14.
- Hussain M, Tangen CM, Berry DL, Higano CS, Crawford ED, Liu G, Wilding G, Prescott S, Kanaga Sundaram S, Small EJ, Dawson NA, Donnelly BJ, Venner PM, Vaishampayan UN, Schellhammer PF, Quinn DI, Raghavan D, Ely B, Moinpour CM, Vogelzang NJ, Thompson IM. Intermittent Versus Continuous Androgen Deprivation in Prostate Cancer. *New England Journal of Medicine.* 2013; 368:1314–1325. [PubMed: 23550669]
- Imbens GW. Generalized Method of Moments and Empirical Likelihood. *Journal of Business & Economic Statistics.* 2002; 20:493–506.
- Imbens GW, Lancaster T. Combining Micro and Macro Data in Microeconomic Models. *The Review of Economic Studies.* 1994; 61:655–680.
- Kovalchick SA. Aggregate-Data Estimation of an Individual Patient Data Linear Random Effects Meta-Analysis With a Patient Covariate-Treatment Interaction Term. *Bio-statistics.* 2013; 14:273–283.
- Li, G.; Li, R.; Zhou, M. Empirical Likelihood in Survival Analysis. In: Fang, K.; Fan, J.; Li, G., editors. *Contemporary Multivariate Analysis and Design of Experiments.* Singapore: World Scientific Publishing Co. Inc.; 2005. p. 337-350.
- Liu D, Zheng Y, Prentice RL, Hsu L. Estimating Risk with Time-to-Event Data: An Application to the Womens Health Initiative. *Journal of the American Statistical Association.* 2014; 109:514–524. [PubMed: 25018574]
- Owen AB. Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika.* 1988; 75:237–249.
- Owen AB. Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics.* 1990; 18:90–120.
- Qin J, Lawless J. Estimating Equations, Empirical Likelihood and Constraints on Parameters. *Canadian Journal of Statistics.* 1995; 23:145–159.

- Qin J, Lawless JF. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*. 1994; 22:300–325.
- Rao, CR. *Linear Statistical Inference and Its Applications*. 2nd. New York: Wiley; 1973.
- Ren JJ, Zhou M. Full Likelihood Inferences in the Cox Model: An Empirical Likelihood Approach. *Annals of the Institute of Statistical Mathematics*. 2011; 63:1005–1018.
- Simmonds MC, Higgins JP. Covariate Heterogeneity in Meta-Analysis: Criteria for Deciding Between Meta-Regression and Individual Patient Data. *Statistics in Medicine*. 2007; 26:2982–2999. [PubMed: 17195960]
- Thomas DR, Grunkemeier GL. Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of the American Statistical Association*. 1975; 70:865–871.
- Tsai HT, Penson DF, Makambi KH, Lynch JH, Van Den Eeden SK, Potosky AL. Efficacy of Intermittent Androgen Deprivation Therapy vs Conventional Continuous Androgen Deprivation Therapy for Advanced Prostate Cancer: a Meta-Analysis. *Urology*. 2013; 82:327–334. [PubMed: 23896094]
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. New York: Springer; 1996.
- Wu C, Sitter RR. A Model-Calibration Approach to Using Complete Auxiliary Information From Survey Data. *Journal of the American Statistical Association*. 2001; 96:185–193.
- Zhou, M. The Cox Proportional Hazards Model With Partially Known Baseline. In: Hsiung, AC.; Ying, Z.; Zhang, CH., editors. *Random Walk, Sequential Analysis and Related Topics*. Singapore: World Scientific Publishing Co.; 2006. p. 215-232.

Table 1

Summary statistics for the estimation of Model (A) $\lambda(t | X_1, X_2) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2)$ when $\rho = 1$.

| Proportion censored | n | Coef | β_1 | | | β_2 | | | ρ | | | $\Lambda(0.3)$ | | | $\Lambda(0.7)$ | | |
|---------------------|-----|---------------|-----------|----|------|-----------|----|------|--------|----|----|----------------|------|------|----------------|------|------|
| | | | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE |
| 0% | 100 | PL | -2 | 12 | - | 3 | 22 | - | - | - | - | 0 | 3 | - | 0 | 10 | - |
| | | DEL | -1 | 5 | 4.96 | 0 | 18 | 1.48 | - | - | - | 0 | 2 | 1.67 | 0 | 6 | 2.59 |
| | | DEL $_{\rho}$ | -1 | 5 | 4.86 | 2 | 22 | 1.03 | 7 | 26 | 0 | 3 | 1.03 | 0 | 10 | 1.05 | |
| 400 | 400 | PL | 0 | 6 | - | 1 | 11 | - | - | - | 0 | 1 | - | 0 | 5 | - | |
| | | DEL | 0 | 2 | 5.62 | 0 | 9 | 1.47 | - | - | - | 0 | 1 | 1.86 | 0 | 3 | 2.55 |
| | | DEL $_{\rho}$ | 0 | 2 | 5.52 | 0 | 11 | 1.02 | 1 | 11 | 0 | 1 | 1.05 | 0 | 5 | 1.05 | |
| 30% | 100 | PL | -1 | 14 | - | 3 | 26 | - | - | - | 0 | 3 | - | 0 | 11 | - | |
| | | DEL | -1 | 5 | 6.84 | 1 | 20 | 1.67 | - | - | - | 0 | 2 | 1.76 | 0 | 7 | 2.50 |
| | | DEL $_{\rho}$ | -1 | 5 | 6.68 | 2 | 25 | 1.04 | 7 | 31 | 0 | 3 | 1.05 | 0 | 11 | 1.05 | |
| 400 | 400 | PL | -1 | 7 | - | 1 | 13 | - | - | - | 0 | 2 | - | 0 | 5 | - | |
| | | DEL | 0 | 2 | 7.10 | 0 | 10 | 1.60 | - | - | - | 0 | 1 | 1.87 | 0 | 4 | 2.36 |
| | | DEL $_{\rho}$ | 0 | 2 | 6.98 | 0 | 13 | 1.02 | 1 | 13 | 0 | 1 | 1.05 | 0 | 5 | 1.04 | |
| 50% | 100 | PL | -2 | 17 | - | 0 | 31 | - | - | - | 0 | 3 | - | 1 | 13 | - | |
| | | DEL | -1 | 5 | 9.64 | -2 | 23 | 1.84 | - | - | 0 | 2 | 1.81 | 1 | 8 | 2.54 | |
| | | DEL $_{\rho}$ | -1 | 5 | 9.61 | -1 | 30 | 1.04 | 6 | 32 | 0 | 3 | 1.01 | 2 | 13 | 1.03 | |
| 400 | 400 | PL | -1 | 8 | - | 1 | 15 | - | - | - | 0 | 2 | - | 0 | 6 | - | |
| | | DEL | 1 | 2 | 9.66 | 0 | 12 | 1.72 | - | - | 0 | 1 | 1.87 | 0 | 4 | 2.50 | |
| | | DEL $_{\rho}$ | 1 | 2 | 9.56 | 0 | 15 | 1.02 | 1 | 14 | 0 | 2 | 1.05 | 0 | 6 | 1.04 | |

NOTE: β_1 and β_2 are the regression coefficients, where the true parameter values are $(-0.5, 0.5)$; $\Lambda(t) = \hat{\rho}$ is the baseline cumulative hazard function evaluated at t ; PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; DEL, the double empirical likelihood estimator $\hat{\beta}_{DEL}$, the extended double empirical likelihood estimator $\hat{\beta}_{\rho}$ that allows for a different baseline hazard function for the aggregate data; Bias and ES, empirical bias ($\times 100$) and empirical standard deviation ($\times 100$) of 1,000 regression parameter estimates; RE, the empirical variance of the maximum partial likelihood estimator divided by that of the double empirical likelihood estimators.

Table 2

Summary statistics for the estimation of Model (B) $\lambda(t | X_1, X_2) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$ when $\rho = 1$.

| Proportion censored | n | Coef | β_1 | | | β_2 | | | β_3 | | | ρ | | | $\Lambda(0.3)$ | | | $\Lambda(0.7)$ | | | |
|---------------------|-----|---------------|-----------|----|------|-----------|----|------|-----------|----|------|--------|----|----|----------------|------|------|----------------|------|------|---|
| | | | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | |
| 0% | 100 | PL | -1 | 17 | - | 2 | 24 | - | -2 | 24 | - | - | - | - | - | 0 | 3 | - | 0 | 10 | - |
| | | DEL | 0 | 6 | 9.45 | 0 | 18 | 1.80 | -1 | 17 | 1.84 | - | - | - | 0 | 2 | 1.84 | 0 | 6 | 2.54 | |
| | | DEL $_{\rho}$ | 0 | 6 | 9.38 | 2 | 23 | 1.02 | -2 | 18 | 1.72 | 6 | 26 | 0 | 3 | 1.01 | 0 | 10 | 1.04 | | |
| 30% | 400 | PL | -1 | 8 | - | 1 | 12 | - | 0 | 11 | - | - | - | - | 0 | 1 | - | 0 | 5 | - | |
| | | DEL | 0 | 3 | 9.44 | 1 | 9 | 1.77 | 0 | 8 | 1.64 | - | - | - | 0 | 1 | 2.01 | 0 | 3 | 2.35 | |
| | | DEL $_{\rho}$ | 0 | 3 | 9.41 | 1 | 12 | 1.01 | 0 | 9 | 1.50 | 1 | 12 | 0 | 1 | 1.03 | 0 | 5 | 1.03 | | |
| 50% | 100 | PL | -2 | 21 | - | 3 | 29 | - | -2 | 28 | - | - | - | - | 0 | 3 | - | 0 | 12 | - | |
| | | DEL | -1 | 6 | 13.8 | 0 | 21 | 1.92 | -1 | 20 | 1.93 | - | - | - | 0 | 2 | 1.96 | 0 | 8 | 2.44 | |
| | | DEL $_{\rho}$ | -1 | 6 | 13.7 | 2 | 28 | 1.04 | -2 | 21 | 1.84 | 8 | 31 | 0 | 3 | 1.03 | 0 | 12 | 1.05 | | |
| 50% | 400 | PL | -1 | 9 | - | 1 | 14 | - | 0 | 13 | - | - | - | - | 0 | 2 | - | 0 | 6 | - | |
| | | DEL | 0 | 3 | 13.7 | 0 | 10 | 1.88 | -1 | 10 | 1.78 | - | - | - | 0 | 1 | 2.04 | 0 | 4 | 2.42 | |
| | | DEL $_{\rho}$ | 0 | 3 | 13.6 | 1 | 13 | 1.04 | -1 | 10 | 1.70 | 2 | 14 | 0 | 1 | 1.04 | 0 | 5 | 1.06 | | |
| 50% | 100 | PL | -2 | 27 | - | 4 | 35 | - | -2 | 36 | - | - | - | - | 0 | 3 | - | 0 | 14 | - | |
| | | DEL | 0 | 6 | 23.2 | -1 | 24 | 2.12 | -4 | 25 | 2.13 | - | - | - | 0 | 2 | 2.05 | 1 | 9 | 2.45 | |
| | | DEL $_{\rho}$ | 0 | 6 | 23.0 | 2 | 33 | 1.09 | -4 | 25 | 2.11 | 9 | 35 | 0 | 3 | 1.05 | 0 | 14 | 1.08 | | |
| 50% | 400 | PL | -1 | 12 | - | 1 | 16 | - | 0 | 16 | - | - | - | - | 0 | 2 | - | 0 | 7 | - | |
| | | DEL | 1 | 3 | 21.0 | 0 | 12 | 1.94 | -2 | 11 | 1.95 | - | - | - | 0 | 1 | 2.04 | 0 | 4 | 2.35 | |
| | | DEL $_{\rho}$ | 1 | 3 | 21.0 | 0 | 16 | 1.03 | -2 | 11 | 1.91 | 2 | 15 | 0 | 2 | 1.04 | 0 | 6 | 1.06 | | |

NOTE: β_1 , β_2 , and β_3 are the regression coefficients, where the true parameter values are $(-0.5, 1, -0.5)$. $\Lambda(t) = \int_0^t \lambda_0(u) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2) du$ is the baseline cumulative hazard function evaluated at t . PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; DEL, the double empirical likelihood estimator $\hat{\beta}_{DEL}$; DEL $_{\rho}$, the extended double empirical likelihood estimator $\hat{\beta}_{DEL_{\rho}}$ that allows for a different baseline hazard function for the aggregate data; Bias and ES, empirical bias ($\times 100$) and empirical standard deviation ($\times 100$) of 1,000 regression parameter estimates; RE, the empirical variance of the maximum partial likelihood estimator divided by that of the double empirical likelihood estimators.

Table 3

Summary statistics for the estimation of Model (A) $\lambda(t | X_1, X_2) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2)$ when $\rho = 1.5$.

| Proportion censored | n | Coef | β_1 | | | β_2 | | | ρ | | | $\Lambda(0.3)$ | | | $\Lambda(0.7)$ | | |
|---------------------|-----|----------------------------------|-----------|----|------|-----------|----|------|--------|----|----|----------------|------|------|----------------|------|------|
| | | | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE |
| 0% | 100 | PL | -1 | 12 | - | 2 | 22 | - | - | - | - | 0 | 3 | - | 0 | 10 | - |
| | | DEL | 0 | 5 | 5.99 | -22 | 15 | 1.88 | - | - | - | 4 | 3 | 0.76 | 15 | 7 | 1.92 |
| | 400 | DEL _{ρ} | -1 | 5 | 4.71 | 1 | 22 | 1.02 | 8 | 38 | 0 | 3 | 1.02 | 0 | 10 | 1.03 | |
| | | PL | 0 | 6 | - | 1 | 11 | - | - | - | - | 0 | 1 | - | 0 | 5 | - |
| | | DEL | 2 | 2 | 6.26 | -21 | 8 | 1.81 | - | - | - | 4 | 2 | 0.88 | 15 | 3 | 1.90 |
| | | DEL _{ρ} | 1 | 2 | 5.68 | 0 | 11 | 1.02 | 2 | 17 | 0 | 1 | 1.04 | 0 | 5 | 1.05 | |
| 30% | 100 | PL | -2 | 14 | - | 2 | 27 | - | - | - | 0 | 3 | - | 0 | 12 | - | |
| | | DEL | 1 | 5 | 8.39 | -25 | 19 | 1.99 | - | - | - | 4 | 3 | 0.82 | 16 | 8 | 2.01 |
| | 400 | DEL _{ρ} | 0 | 2 | 5.20 | 0 | 11 | 1.01 | 2 | 17 | 0 | 1 | 1.03 | 0 | 5 | 1.04 | |
| | | PL | -1 | 7 | - | 1 | 13 | - | - | - | - | 0 | 2 | - | 0 | 5 | - |
| | | DEL | 1 | 2 | 7.59 | -23 | 9 | 1.92 | - | - | - | 4 | 2 | 0.90 | 15 | 4 | 1.88 |
| | | DEL _{ρ} | 0 | 2 | 7.01 | 0 | 13 | 1.02 | 2 | 19 | 0 | 1 | 1.04 | 0 | 5 | 1.04 | |
| 50% | 100 | PL | -2 | 16 | - | 2 | 32 | - | - | - | 0 | 3 | - | 0 | 13 | - | |
| | | DEL | 1 | 5 | 8.84 | -28 | 22 | 2.09 | - | - | - | 4 | 3 | 0.97 | 16 | 10 | 1.84 |
| | 400 | DEL _{ρ} | 0 | 6 | 8.29 | 1 | 31 | 1.04 | 12 | 50 | 0 | 3 | 1.02 | 1 | 13 | 1.03 | |
| | | PL | -1 | 8 | - | 1 | 15 | - | - | - | - | 0 | 2 | - | 0 | 6 | - |
| | | DEL | 2 | 2 | 9.99 | -27 | 11 | 2.01 | - | - | - | 4 | 2 | 0.90 | 16 | 4 | 1.96 |
| | | DEL _{ρ} | 1 | 2 | 9.39 | 0 | 15 | 1.02 | 2 | 21 | 0 | 2 | 1.05 | 0 | 6 | 1.04 | |

NOTE: β_1 and β_2 are the regression coefficients, where the true parameter values are $(-0.5, 0.5)$; $\Lambda(t) = \int_0^t$ is the baseline cumulative hazard function evaluated at t ; PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; DEL, the double empirical likelihood estimator $\hat{\beta}_{DEL}$; DEL _{ρ} , the extended double empirical likelihood estimator $\hat{\beta}_{DEL\rho}$ that allows for a different baseline hazard function for the aggregate data; Bias and ES, empirical bias ($\times 100$) and empirical standard deviation ($\times 100$) of 1,000 regression parameter estimates; RE, the empirical variance of the maximum partial likelihood estimator divided by that of the double empirical likelihood estimators.

Summary statistics for the estimation of Model (B) $\lambda(t | X_1, X_2) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$ when $\rho = 1.5$.

Table 4

| Proportion censored | n | Coef | β_1 | | | β_2 | | | β_3 | | | ρ | | | $\Lambda(0.3)$ | | | $\Lambda(0.7)$ | | | | |
|---------------------|-----|---------------|-----------|----|------|-----------|----|------|-----------|----|------|--------|----|----|----------------|------|----|----------------|------|----|----|------|
| | | | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | Bias | ES | RE | | |
| 0% | 100 | PL | -1 | 17 | - | 2 | 24 | - | -2 | 24 | - | - | - | - | - | - | 0 | 3 | - | - | 10 | - |
| | | DEL | 0 | 5 | 9.70 | -24 | 16 | 2.10 | 6 | 17 | 2.04 | - | - | - | - | - | 4 | 3 | 0.88 | 15 | 7 | 2.02 |
| | | DEL $_{\rho}$ | 0 | 6 | 9.38 | 2 | 23 | 1.02 | -2 | 18 | 1.73 | 9 | 40 | 0 | 3 | 1.01 | 0 | 3 | 1.01 | 0 | 10 | 1.04 |
| 30% | 400 | PL | -1 | 8 | - | 1 | 12 | - | 0 | 11 | - | - | - | - | - | 0 | 1 | - | - | 0 | 5 | - |
| | | DEL | 2 | 2 | 10.1 | -24 | 8 | 2.08 | 6 | 8 | 1.82 | - | - | - | - | - | 4 | 1 | 0.95 | 15 | 4 | 1.90 |
| | | DEL $_{\rho}$ | 1 | 2 | 9.75 | 1 | 12 | 1.02 | -1 | 9 | 1.51 | 2 | 19 | 0 | 1 | 1.02 | 0 | 1 | 1.02 | 0 | 5 | 1.03 |
| 50% | 100 | PL | -2 | 21 | - | 3 | 29 | - | -2 | 28 | - | - | - | - | - | 0 | 3 | - | - | 0 | 12 | - |
| | | DEL | 0 | 6 | 14.1 | -27 | 19 | 2.20 | 4 | 20 | 2.09 | - | - | - | - | - | 4 | 3 | 0.95 | 16 | 9 | 1.93 |
| | | DEL $_{\rho}$ | -1 | 6 | 13.6 | 2 | 28 | 1.04 | -2 | 21 | 1.84 | 12 | 46 | 0 | 3 | 1.03 | 0 | 3 | 1.03 | 0 | 12 | 1.05 |
| 50% | 400 | PL | -1 | 9 | - | 1 | 14 | - | 0 | 13 | - | - | - | - | - | 0 | 2 | - | - | 0 | 6 | - |
| | | DEL | 1 | 3 | 14.2 | -26 | 9 | 2.17 | 5 | 10 | 1.92 | - | - | - | - | - | 4 | 2 | 0.98 | 15 | 4 | 1.94 |
| | | DEL $_{\rho}$ | 0 | 3 | 13.7 | 1 | 13 | 1.04 | -1 | 10 | 1.70 | 3 | 21 | 0 | 1 | 1.04 | 0 | 1 | 1.04 | 0 | 6 | 1.05 |
| 50% | 100 | PL | -2 | 27 | - | 4 | 35 | - | -2 | 36 | - | - | - | - | - | 0 | 3 | - | - | 0 | 14 | - |
| | | DEL | 0 | 6 | 23.5 | -30 | 23 | 2.36 | 0 | 24 | 2.23 | - | - | - | - | - | 4 | 3 | 1.00 | 16 | 10 | 1.93 |
| | | DEL $_{\rho}$ | 0 | 6 | 22.6 | 2 | 33 | 1.09 | -4 | 25 | 2.11 | 14 | 53 | 0 | 3 | 1.06 | 0 | 3 | 1.06 | 0 | 14 | 1.08 |
| 50% | 400 | PL | -1 | 12 | - | 1 | 16 | - | 0 | 16 | - | - | - | - | - | 0 | 0 | - | - | 0 | 7 | - |
| | | DEL | 2 | 3 | 21.3 | -29 | 11 | 2.19 | 2 | 11 | 2.05 | - | - | - | - | - | 4 | 2 | 0.99 | 16 | 5 | 1.88 |
| | | DEL $_{\rho}$ | 1 | 3 | 20.6 | 0 | 16 | 1.03 | -2 | 11 | 1.90 | 3 | 23 | 0 | 2 | 1.04 | 0 | 2 | 1.04 | 0 | 6 | 1.06 |

NOTE: β_1 , β_2 , and β_3 are the regression coefficients, where the true parameter values are $(-0.5, 1, -0.5)$. $\Lambda(t) = \rho^t$ is the baseline cumulative hazard function evaluated at t . PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; DEL, the double empirical likelihood estimator $\hat{\beta}_{DEL}$; DEL $_{\rho}$, the extended double empirical likelihood estimator $\hat{\beta}_{DEL_{\rho}}$ that allows for a different baseline hazard function for the aggregate data; Bias and ES, empirical bias ($\times 100$) and empirical standard deviation ($\times 100$) of 1,000 regression parameter estimates; RE, the empirical variance of the maximum partial likelihood estimator divided by that of the double empirical likelihood estimators.

Table 5

Estimated regression coefficients of the Cox model for the prostate cancer study.

| | cADT | | PSA | | cADT*PSA | | 76-80 years | | >80 years | | ρ | |
|-------------------|-------|-------|-------|-------|----------|-------|-------------|-------|-----------|-------|--------|-------|
| | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
| Complete | 0.193 | 0.066 | 0.905 | 0.084 | 0.359 | 0.099 | 0.316 | 0.058 | 0.764 | 0.051 | - | - |
| Subset | 0.433 | 0.251 | 1.121 | 0.298 | 0.117 | 0.377 | 0.268 | 0.254 | 0.822 | 0.205 | - | - |
| DEL I | 0.309 | 0.156 | 1.076 | 0.083 | 0.158 | 0.251 | 0.266 | 0.248 | 0.804 | 0.198 | - | - |
| DEL $_{\rho}$ I | 0.414 | 0.213 | 1.081 | 0.084 | 0.155 | 0.252 | 0.270 | 0.251 | 0.823 | 0.204 | 1.132 | 0.212 |
| DEL II | 0.382 | 0.220 | 1.125 | 0.300 | 0.193 | 0.309 | 0.259 | 0.251 | 0.812 | 0.199 | - | - |
| DEL $_{\rho}$ II | 0.398 | 0.229 | 1.125 | 0.299 | 0.192 | 0.308 | 0.255 | 0.249 | 0.805 | 0.199 | 0.977 | 0.115 |
| DEL III | 0.236 | 0.059 | 1.076 | 0.083 | 0.240 | 0.106 | 0.259 | 0.248 | 0.797 | 0.194 | - | - |
| DEL $_{\rho}$ III | 0.237 | 0.060 | 1.077 | 0.084 | 0.241 | 0.107 | 0.263 | 0.249 | 0.806 | 0.199 | 1.022 | 0.111 |

NOTE: Coef, the estimated coefficient; SE, the bootstrap standard error given by the standard deviation of the 1000 estimates. SE for the complete dataset with 4548 cases is the asymptotic standard error estimates. DEL, the double empirical likelihood estimator $\hat{\beta}$. DEL $_{\rho}$, the extended double empirical likelihood estimator $\hat{\beta}_{\rho}$ that allows for a different baseline hazard function for the aggregated data.

Table 6

Estimated regression coefficients of the Cox model for the pancreatic cancer study.

| | Nodes | | Margins | | PNI | | > 65 years | | Male | | ρ | |
|-------------------|-------|------|---------|------|------|------|------------|------|-------|------|--------|------|
| | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE | Coef | SE |
| PL | 0.37 | 0.22 | 0.41 | 0.17 | 1.09 | 0.42 | 0.28 | 0.16 | -0.29 | 0.15 | - | - |
| DEL $_{\rho}$ I | 0.25 | 0.10 | 0.42 | 0.17 | 1.09 | 0.40 | 0.29 | 0.16 | -0.29 | 0.16 | 0.80 | 0.08 |
| DEL $_{\rho}$ I | 0.37 | 0.23 | 0.34 | 0.06 | 1.10 | 0.40 | 0.27 | 0.16 | -0.30 | 0.16 | 0.80 | 0.08 |
| DEL $_{\rho}$ III | 0.27 | 0.08 | 0.36 | 0.05 | 1.10 | 0.41 | 0.28 | 0.16 | -0.30 | 0.15 | 0.80 | 0.08 |

NOTE: Coef, the estimated coefficient; SE, the bootstrap standard error given by the standard deviation of the 1000 estimates. SE for the maximum partial likelihood estimator is the asymptotic standard error estimates. DEL $_{\rho}$, the extended double empirical likelihood estimator $\hat{\beta}_{\rho}$ that allows for a different baseline hazard function for the aggregate data.