Efficient Estimation of the Cox Model
With Auxiliary Subgroup Survival Information

Chiung-Yu Huang

Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
(Joint work with Jing Qin and Huei-Ting Tsai)

SRCOS Summer Research Conference
June 8, 2016
Outline

- Motivation
- Summarizing auxiliary survival information
- A double empirical likelihood approach for synthesizing information
- Simulations
- Data example: a pancreatic cancer study
- Remarks
Motivation

- Demand for statistical methods that better utilize data made available in the public domains has been increasingly recognized.

- To guide clinical diagnostic and management decisions, meta-analysis is often carried out to summarize results from relevant clinical studies.

- Information made available for meta-analysis:
  - Summary statistics or regression coefficients associated with the covariate of interest from each study.
  - Individual-level data from all clinical studies.
  - In between, a mixture of individual-level data and summary statistics.
Motivation

- Individual-level data from clinical studies provide estimates of the treatment-covariate interactions or effects of biomarkers that may not be reported in existing publications;

  Sample size or number of events may be too small to provide accurate estimates

- 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.

- Properly combining individual-level data with aggregate data from external sources such as disease registries can yield more efficient estimation and better prediction.
Motivation

- There has been a rising interest in better exploiting the population-based cancer survival statistics made available by the Surveillance, Epidemiology and End Results (SEER) program of NCI.

  e.g. Liu, Zheng, Prentice & Hsu (JASA 2014) developed risk prediction model for colorectal cancer using individual-level data from Women’s Health Initiative (WHI) Study and overall cancer incidence from SEER.
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.

The SEER Cancer Statistics Review 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.
Ovarian cancer: 5-year relative and period survival by race, diagnosis year, stage and age.

### Table 21.8

**Cancer of the Ovary (Invasive)**

**5-Year Relative and Period Survival (Percent) by Race, Diagnosis Year, Stage and Age**

<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>All Races, Females</th>
<th>White Females</th>
<th>Black Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;65</td>
<td>≥65</td>
<td></td>
</tr>
<tr>
<td>1970-1973&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>32</td>
</tr>
</tbody>
</table>
| 1975-1977<sup>b</sup> | 36.0 | 44.2 | 21.4 | 35.3 | 43.5 | 21.1 | 41.8 | 49.0 | 24.9<sup>g</sup>
| 1978-1980<sup>b</sup> | 37.7 | 47.3 | 22.1 | 36.8 | 46.5 | 22.3 | 38.8 | 49.2 | 22.8<sup>g</sup>
| 1981-1983<sup>b</sup> | 38.9 | 50.2 | 23.1 | 39.0 | 50.2 | 23.1 | 37.5 | 45.4 | 22.6<sup>g</sup>
| 1984-1986<sup>b</sup> | 38.4 | 49.4 | 24.1 | 37.5 | 49.5 | 23.7 | 39.4 | 52.2 | 21.2<sup>g</sup>
| 1987-1989<sup>b</sup> | 36.2 | 51.1 | 23.6 | 38.1 | 51.2 | 23.9 | 33.7 | 46.9 | 17.1<sup>g</sup>
| 1990-1992<sup>b</sup> | 40.5 | 55.3 | 24.2 | 40.3 | 55.9 | 24.1 | 35.6 | 49.1 | 16.9<sup>g</sup>
| 1993-1995<sup>b</sup> | 41.4 | 55.7 | 26.5 | 40.9 | 55.4 | 25.9 | 41.5 | 54.8 | 26.9<sup>g</sup>
| 1996-1998<sup>b</sup> | 43.8 | 57.0 | 28.4 | 43.2 | 56.5 | 28.4 | 39.1 | 53.1 | 20.6<sup>g</sup>
| 1999-2003<sup>b</sup> | 43.7 | 56.4 | 27.9 | 43.3 | 56.5 | 28.0 | 35.7 | 46.0 | 19.9<sup>g</sup>
| 2004-2010<sup>b</sup> | 44.6<sup>c</sup> | 57.4<sup>c</sup> | 27.5<sup>c</sup> | 44.3<sup>c</sup> | 57.7<sup>c</sup> | 27.9<sup>c</sup> | 36.4 | 45.3 | 21.6<sup>g</sup>

**5-Year Period Survival (Percent)**

<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>All Races, Females</th>
<th>White Females</th>
<th>Black Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;65</td>
<td>≥65</td>
<td></td>
</tr>
</tbody>
</table>
| 2010              | 45.3 | 58.6 | 27.7 | 45.2 | 59.0 | 28.1 | 35.0 | 45.6 | 20.4

**Stage Distribution (%) 2004-2010<sup>ce</sup>**

<table>
<thead>
<tr>
<th>Stage</th>
<th>All Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>33,841</td>
</tr>
<tr>
<td>Percent</td>
<td>100%</td>
</tr>
<tr>
<td>Localized</td>
<td>15</td>
</tr>
<tr>
<td>Regional</td>
<td>18</td>
</tr>
<tr>
<td>Distant</td>
<td>61</td>
</tr>
<tr>
<td>Unstaged</td>
<td>6</td>
</tr>
</tbody>
</table>

**5-Year Relative Survival (Percent), 2004-2010<sup>e</sup>**

<table>
<thead>
<tr>
<th>Age at Diagnosis</th>
<th>All Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages &lt;65</td>
<td>74.9</td>
</tr>
<tr>
<td>Ages 65-74</td>
<td>58.0</td>
</tr>
<tr>
<td>Ages 75+</td>
<td>47.4</td>
</tr>
<tr>
<td>Ages ≥85</td>
<td>37.1</td>
</tr>
<tr>
<td>Ages ≥95</td>
<td>19.1</td>
</tr>
<tr>
<td>Ages ≥100</td>
<td>57.5</td>
</tr>
<tr>
<td>Ages ≥110</td>
<td>27.3</td>
</tr>
</tbody>
</table>

Stage<sup>f</sup>:

<table>
<thead>
<tr>
<th>Stage</th>
<th>All Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Localized</td>
<td>92.3</td>
</tr>
<tr>
<td>Regional</td>
<td>71.7</td>
</tr>
<tr>
<td>Distant</td>
<td>27.4</td>
</tr>
<tr>
<td>Unstaged</td>
<td>21.0</td>
</tr>
</tbody>
</table>
Table 21.8

Cancer

<table>
<thead>
<tr>
<th>Year of Diagnosis:</th>
<th>All</th>
<th>&lt;65</th>
<th>65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960-1963^a</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1970-1973^a</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1975-1977^b</td>
<td>36.0</td>
<td>44.2</td>
<td>21.4</td>
</tr>
<tr>
<td>1978-1980^b</td>
<td>37.7</td>
<td>47.3</td>
<td>22.1</td>
</tr>
<tr>
<td>1981-1983^b</td>
<td>38.8</td>
<td>50.2</td>
<td>23.1</td>
</tr>
<tr>
<td>1984-1986^b</td>
<td>38.4</td>
<td>49.4</td>
<td>24.1</td>
</tr>
<tr>
<td>1987-1989^b</td>
<td>38.2</td>
<td>51.1</td>
<td>23.6</td>
</tr>
<tr>
<td>1990-1992^b</td>
<td>40.5</td>
<td>55.3</td>
<td>24.2</td>
</tr>
<tr>
<td>1993-1995^b</td>
<td>41.4</td>
<td>55.7</td>
<td>26.5</td>
</tr>
<tr>
<td>1996-1998^b</td>
<td>43.8</td>
<td>57.0</td>
<td>28.4</td>
</tr>
<tr>
<td>1999-2003^b</td>
<td>43.7</td>
<td>56.4</td>
<td>27.9</td>
</tr>
<tr>
<td>2004-2010^b</td>
<td>44.6</td>
<td>57.4</td>
<td>27.5</td>
</tr>
</tbody>
</table>

5-Year Period Survival (Percent)\[^cd\]

<table>
<thead>
<tr>
<th>Year</th>
<th>All</th>
<th>&lt;65</th>
<th>65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>45.3</td>
<td>58.6</td>
<td>27.7</td>
</tr>
</tbody>
</table>

5-year survival probabilities among ovarian cancer patients diagnosed before and after age 65 are 58.6% and 27.7%.
Notation and Model Setup

Notation

- **\(X\)**: a \(p \times 1\) vector of baseline covariates. \(X \sim G\).
- **\(T\)**: time to event; \(T \mid X \sim \lambda(t \mid X), S(t \mid X)\)
- **\(C\)**: censoring time.
- **\(Y = \min(T, C)\)** and **\(\Delta = I(T \leq C)\)**.

Assume that \(T\) follows the proportional hazards model

\[
\lambda(t \mid \mathbf{x}) = \lambda(t) \exp(\beta' \mathbf{x}),
\]

and \(C\) is independent of \(T\) given \(X\).
5-year survival probabilities among ovarian cancer patients diagnosed before and after age 65 are 58.6% and 27.7%.
(SEER Cancer Statistics Review 1973-2011)
5-year survival probabilities among ovarian cancer patients diagnosed before and after age 65 are 58.6% and 27.7%.
(SEER Cancer Statistics Review 1973-2011)

Our approach to combine information: empirical likelihood methods.

- likelihood function under Cox model based on individual-level data.
- constraints based on external subgroup information.
- maximize the likelihood function w.r.t. the constraints.

References:
Thomas and Grunkemeier (1975)
Owen (1988) and Qin & Lawless (1994)
Li, Li & Zhou (2005), Zhou (2006)
5-year survival probabilities among ovarian cancer patients diagnosed before and after age 65 are 58.6% and 27.7%.
(SEER Cancer Statistics Review 1973-2011)

Write $X = (Z, W)$.

- $Z$: age of diagnosis
- $W$: biomarkers and other risk factors of ovarian cancer.

Let $\Omega_k$ denote the $k$th subgroup whose 5-year survival is available:

$\Omega_1 = \{(Z, W) : Z < 65\}$ and $\Omega_2 = \{(Z, W) : Z \geq 65\}$

5-year survival probabilities are expressed as

$\Pr(T > 5 \mid X \in \Omega_1) = 0.586$ and $\Pr(T > 5 \mid X \in \Omega_2) = 0.277$
A general expression of the auxiliary survival information at $t^*$ is

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

or, equivalently,

$$\text{pr}(T > t^*, X \in \Omega_k) - \phi_k \times \text{pr}(X \in \Omega_k) = 0.$$

By double expectation and under the assumed Cox model,

$$E \left[ E \{ I(T > t^*) I(X \in \Omega_k) - \phi_k I(X \in \Omega_k) \mid X \} \right]$$

$$= E \left[ I(X \in \Omega_k) \{ S(t^* \mid X) - \phi_k \} \right]$$

$$= E \left\{ I(X \in \Omega_k) \{ \exp \{ -\Lambda(t^*) \exp(\beta' X) \} - \phi_k \} \right\} = 0.$$
Define the estimating function

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

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

$$E\{\Psi_k(X, \beta, \Lambda)\} = 0, \quad k = 1, \ldots, K. \quad (1)$$

This only involves $\beta$ and $\Lambda(t^*)$. Setting $\alpha = \Lambda(t^*)$, (1) can be reexpressed as

$$E\{\Psi_k(X, \beta, \alpha)\} = 0, \quad k = 1, \ldots, K.$$
Synthesizing Information Using Empirical Likelihood Method

- The log full likelihood of the observed data $(\Delta_i, Y_i, X_i), i = 1, \ldots, n$

\[ \ell_F = \sum_{i=1}^{n} \Delta_i \{ \beta' X_i + \log \lambda(Y_i) \} - \Lambda(Y_i) \exp(\beta' X_i) + \log \{ dG(X_i) \} \]

- Denote by $\lambda_i$ the jump size of $\Lambda$ at $Y_i$
  and by $p_i$ the jump size of $G$ at $X_i$.

Define $S^{(k)}(t, \beta) = n^{-1} \sum_{j=1}^{n} I(Y_j \geq t) \exp(\beta' X_j) X_j^{\otimes k}, k = 0, 1, 2$. 
Empirical Likelihood for Right-Censored Data

The log full empirical likelihood can then be decomposed as

$$\ell_F = \ell_C + \ell_M$$

where the log conditional likelihood of \((Y, \Delta)\) given \(X\)

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

and the log marginal likelihood of \(X\)

$$\ell_M = \sum_{i=1}^{n} \log(p_i),$$
For a fixed $\beta$, differentiating the log conditional likelihood $\ell_C$ w.r.t. $\lambda_i$ 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-type estimator

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

Another well-known result is that replacing $\Lambda$ with $\hat{\Lambda}_B$ in $\ell_C$ yields the log partial likelihood function.
We first assume that the auxiliary survival information is consistent with the individual-level data.

In other words, individuals in the clinical study are a representative sample of the population from which the aggregate survival information is derived.

This assumption will be relaxed later.
A simple idea to combine auxiliary information is maximizing $\ell_F$ w.r.t.

$$p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \text{and} \quad \sum_{i=1}^{n} p_i \Psi_k(X_i; \beta, \Lambda) = 0, \quad k = 1, \ldots, K.$$

Note: the last constraint is derived from $E\{\Psi_k(X; \beta, \Lambda)\} = 0$

Because $\Psi_k$ only involves the value of $\Lambda(t)$ at $t = t^*$, intuitively, one may replace $\Lambda(t^*)$ in $\Psi_k$ 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 $\beta$. 
Key: treat $\alpha = \Lambda(t^*)$ as a nuisance parameter

$$\alpha = \Lambda(t^*) = \int_0^\infty I(u \leq t^*) d\Lambda(u)$$

Maximizing

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

w.r.t. the constraint

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

leads to the maximizer

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

where $\nu$ is the Lagrange multiplier.
Replacing $\lambda_i$ in $\ell_C$ with

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

(2)

gives the constrained log conditional likelihood

$$\sum_{i=1}^{n} \Delta_i \left[ \beta' X_i - \log\{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)\} \right] + n\nu\alpha.$$
Double Empirical Likelihood Approach

Next, we maximize

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

w.r.t. the constraints

$$p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \Psi_k(X_i, \beta, \alpha) = 0, \ k = 1, \ldots, K.$$

Applying the classic empirical likelihood argument, we have

$$p_i = \frac{1}{n} \times \frac{1}{1 + \xi' \Psi(X_i, \beta, \alpha)},$$

where $\Psi(x, \beta, \alpha) = \{\Psi_1(x, \beta, \alpha), \ldots, \Psi_K(x, \beta, \alpha)\}'$ and $\xi = (\xi_1, \ldots, \xi_K)'$ are the Lagrange multipliers.
Double Empirical Likelihood Approach

Replacing $p_i$ in $\ell_M$ with

$$p_i = \frac{1}{n} \times \frac{1}{1 + \xi' \Psi(X_i, \beta, \alpha)}$$

gives the constrained log marginal likelihood

$$- \sum_{i=1}^{n} \log \{1 + \xi' \Psi(X_i, \beta, \alpha)\}$$
Combing the two constrained log likelihoods yields the constrained log full likelihood function $\ell$, where, up to a constant,

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

The procedure allows us to change an infinite dimension problem to a finite-dimension problem at the expense of introducing an additional $(K + 2)$-dimensional parameters.
Double Empirical Likelihood Approach

To estimate $\theta = (\beta, \xi, \nu, \alpha)$, we solve a system of empirical score equations:

\[
U_1(\beta, \xi, \nu, \alpha) = \sum_{i=1}^{n} \Delta_i \left\{ X_i - \frac{S^{(1)}(Y_i, \beta)}{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)} \right\}
\]

\[
- \sum_{i=1}^{n} \frac{\xi' \frac{\partial \Psi}{\partial \beta}(X_i, \beta, \alpha)}{1 + \xi' \Psi(X_i, \beta, \alpha)}
\]

\[
U_2(\beta, \xi, \nu, \alpha) = \sum_{i=1}^{n} \frac{\Psi(X_i, \beta, \alpha)}{1 + \xi' \Psi(X_i, \beta, \alpha)},
\]

\[
U_3(\beta, \xi, \nu, \alpha) = \sum_{i=1}^{n} \left\{ \frac{\Delta_i I(Y_i \leq t^*)}{S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*)} - \alpha \right\}
\]

\[
U_4(\beta, \xi, \nu, \alpha) = \sum_{i=1}^{n} \left\{ \frac{\xi' \frac{\partial \Psi}{\partial \alpha}(X_i, \beta, \alpha)}{1 + \xi' \Psi(X_i, \beta, \alpha)} - \nu \right\}
\]
Properties

Denote by \( \hat{\theta} = (\hat{\beta}, \hat{\xi}, \hat{\nu}, \hat{\alpha}) \) the solution to \( U(\theta) = 0 \).

**Theorem**

Assume that \( X \) is bounded, the true regression parameter \( \beta_0 \) lies in a compact set, and both \( T \) and \( C \) are absolutely continuous. Moreover, assume that \( E\{\Psi(X, \beta_0, \alpha_0)\Psi(X, \beta_0, \alpha_0)\}' \) 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

\[
\Gamma^{-1} = (\Sigma + BQ^{-1}B')^{-1}.
\]

Note: \( \Sigma^{-1} \) is the asymptotic covariance-covariance matrix of the maximum partial likelihood estimator.
To estimate $\Lambda(t)$, we consider the following empirical likelihood-based estimator to incorporate the auxiliary survival information:

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

We can show that $n^{1/2} \{\hat{\Lambda}_{EL}(t) - \Lambda(t)\}$ converges to a zero-mean Gaussian process on $[0, \tau]$. 
Test for Conformity

To test the conformity of the auxiliary survival information, that is, $\xi = 0$, we consider an empirical likelihood ratio test statistic

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

Note that under $\xi = 0$, the likelihood $\ell(\beta, \alpha, 0, \nu)$ is maximized by $(\beta, \nu, \alpha) = (\hat{\beta}_{PL}, 0, \hat{\alpha}_{PL})$, 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**

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

See Qin and Lawless (1994).
An Extension

- It is desired to allow the aggregate data to have a different survival time model.

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

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

  so that the potential differences in the two data sources are characterized by a scale factor $\rho$.

- The auxiliary survival information $\Pr(T > t^* \mid X \in \Omega_k) = \phi_k$ can be summarized by

  $$E\{\tilde{\Psi}_k(X, \beta, \alpha, \rho)\} = 0$$

  where $\tilde{\Psi}_k(X, \beta, \alpha, \rho) = I(X \in \Omega_k)[\exp\{-\rho \alpha \exp(\beta' X)\} - \phi_k]$.
The constrained log full likelihood function is

\[ \tilde{\ell}(\beta, \xi, \nu, \alpha, \rho) = \sum_{i=1}^{n} \Delta_i \left[ \beta' X_i - \log\left\{ S^{(0)}(Y_i, \beta) + \nu I(Y_i \leq t^*) \right\} \right] + n\nu\alpha - \sum_{i=1}^{n} \log\left\{ 1 + \xi' \tilde{\Psi}(X_i, \beta, \alpha, \rho) \right\}. \]

Maximize \( \tilde{\ell} \) to estimate the parameters.
An Extension

To test if the same baseline hazard function is shared by the individual-level data and the aggregate data, that is, \( \rho = 1 \), 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, the empirical log-likelihood ratio \( \tilde{R} \) converges in distribution to a \( \chi^2 \) random variable with 1 degrees of freedom as \( n \to \infty \).
Simulations

- $X_1 \sim N(0,1)$
  
  $X_2$ is a Bernoulli r.v. with $\text{pr}(X_2 = 1) = 0.5$.

- $T$ in the individual-level data was generated from PHM
  
  (A) $\lambda(t \mid X_1, X_2) = \lambda(t) \exp(\beta_1 X_1 + \beta_2 X_2)$

  (B) $\lambda(t \mid X_1, X_2) = \lambda(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)$

- $C$ is an uniform r.v. so that the censoring rate was approximately 0%, 30%, and 50%.

- $n = 100$. 
Simulations

- Auxiliary survival information is derived from the Cox model with a different baseline hazard function $\lambda^*(t) = 1.5 \lambda(t)$, that is, $\rho = 1.5$.

- Auxiliary survival information at $t = 0.5$ for subgroups

$$\Omega_1 = \{(X_1, X_2) : X_1 \leq 0, X_2 = 0\}$$

and

$$\Omega_2 = \{(X_1, X_2) : X_1 > 0, X_2 = 0\}$$

are approximately 0.57 and 0.77 under both Model (A) and Model (B).
Table 1: Summary statistics for the estimation of Model (A) with $\rho = 1.5$.

<table>
<thead>
<tr>
<th>% cens</th>
<th>Coef</th>
<th>$\beta_1$</th>
<th>Bias</th>
<th>ESD</th>
<th>RE</th>
<th>Bias</th>
<th>ESD</th>
<th>RE</th>
<th>Bias</th>
<th>ESD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>PL</td>
<td>-1</td>
<td>12</td>
<td>-</td>
<td>2</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>0</td>
<td>5</td>
<td>5.99</td>
<td>-22</td>
<td>15</td>
<td>1.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\text{DEL}_\rho$</td>
<td>-1</td>
<td>5</td>
<td>4.71</td>
<td>1</td>
<td>22</td>
<td>1.02</td>
<td>8</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>PL</td>
<td>-2</td>
<td>14</td>
<td>-</td>
<td>2</td>
<td>27</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>1</td>
<td>5</td>
<td>8.39</td>
<td>-25</td>
<td>19</td>
<td>1.99</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\text{DEL}_\rho$</td>
<td>0</td>
<td>2</td>
<td>5.20</td>
<td>0</td>
<td>11</td>
<td>1.01</td>
<td>2</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>PL</td>
<td>-2</td>
<td>16</td>
<td>-</td>
<td>2</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>1</td>
<td>5</td>
<td>8.84</td>
<td>-28</td>
<td>22</td>
<td>2.09</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>$\text{DEL}_\rho$</td>
<td>0</td>
<td>6</td>
<td>8.29</td>
<td>1</td>
<td>31</td>
<td>1.04</td>
<td>12</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The true parameter values for $(\beta_1, \beta_2)$ are $(-0.5, 0.5)$; $\Lambda(t) = t^2$; PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; DEL, the double empirical likelihood estimator $\hat{\beta}$; $\text{DEL}_\rho$, the extended double empirical likelihood estimator $\hat{\beta}_\rho$ that allows for a different baseline hazard function for the aggregate data; Bias and ESD, 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 1: Summary statistics for the estimation of Model (B) with $\rho = 1.5$.

<table>
<thead>
<tr>
<th>%cens</th>
<th>Coef</th>
<th>$\beta_1$</th>
<th>Bias</th>
<th>ESD</th>
<th>RE</th>
<th>Bias</th>
<th>$\beta_2$</th>
<th>ESD</th>
<th>RE</th>
<th>Bias</th>
<th>$\beta_3$</th>
<th>ESD</th>
<th>RE</th>
<th>Bias</th>
<th>ESD</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>PL</td>
<td>-1</td>
<td>17</td>
<td>–</td>
<td>2</td>
<td>24</td>
<td>–</td>
<td>-2</td>
<td>24</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>0</td>
<td>5</td>
<td>9.70</td>
<td>-24</td>
<td>16</td>
<td>2.10</td>
<td>6</td>
<td>17</td>
<td>2.04</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL$\rho$</td>
<td>0</td>
<td>6</td>
<td>9.38</td>
<td>2</td>
<td>23</td>
<td>1.02</td>
<td>-2</td>
<td>18</td>
<td>1.73</td>
<td>9</td>
<td>40</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>30%</td>
<td>PL</td>
<td>-2</td>
<td>21</td>
<td>–</td>
<td>3</td>
<td>29</td>
<td>–</td>
<td>-2</td>
<td>28</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>0</td>
<td>6</td>
<td>14.1</td>
<td>-27</td>
<td>19</td>
<td>2.20</td>
<td>4</td>
<td>20</td>
<td>2.09</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL$\rho$</td>
<td>-1</td>
<td>6</td>
<td>13.6</td>
<td>2</td>
<td>28</td>
<td>1.04</td>
<td>-2</td>
<td>21</td>
<td>1.84</td>
<td>12</td>
<td>46</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>50%</td>
<td>PL</td>
<td>-2</td>
<td>27</td>
<td>–</td>
<td>4</td>
<td>35</td>
<td>–</td>
<td>-2</td>
<td>36</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL</td>
<td>0</td>
<td>6</td>
<td>23.5</td>
<td>-30</td>
<td>23</td>
<td>2.36</td>
<td>0</td>
<td>24</td>
<td>2.23</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>DEL$\rho$</td>
<td>0</td>
<td>6</td>
<td>22.6</td>
<td>2</td>
<td>33</td>
<td>1.09</td>
<td>-4</td>
<td>25</td>
<td>2.11</td>
<td>14</td>
<td>53</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

NOTE: The true parameter values for $(\beta_1, \beta_2, \beta_3)$ are $(-0.5, 1, -0.5)$; $\Lambda(t) = t^2$;
PL, the maximum partial likelihood estimator $\hat{\beta}_{PL}$; 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 aggregate data; Bias and ESD, 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.
Data Example

- A retrospective cohort study of 209 consecutive patients who had surgical resection of pancreatic cancer and follow-up at the Johns Hopkins Hospital from between 1998 and 2007.

- We aim to evaluate the effects of presence of lymph nodes, positive resection margins, presence of perineural invasion (PNI), age at surgery ($\leq 65$ and $> 65$), and gender on overall survival.
Three sets of auxiliary survival information reported in Cameron et al. (Annals of Surgery, 2006) estimated from 1000 consecutive pancreatectomies performed by a single surgeon between March 1969 and May 2003:

(I) the three-year survival probabilities for node-negative and node-positive patients were 40% and 26%.
(II) the three-year survival probabilities for margin-negative and margin-positive patients were 35% and 20%.
(III) All the four survival probabilities given in (I) and (II).
Table: Estimated regression coefficients of the Cox model for the pancreatic cancer study.

<table>
<thead>
<tr>
<th></th>
<th>Nodes</th>
<th>Margins</th>
<th>PNI</th>
<th>&gt; 65 years</th>
<th>Male</th>
<th>ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>SE</td>
<td>Coef</td>
<td>SE</td>
<td>Coef</td>
<td>SE</td>
</tr>
<tr>
<td>PL</td>
<td>0.37</td>
<td>0.22</td>
<td>0.41</td>
<td>0.17</td>
<td>1.09</td>
<td>0.42</td>
</tr>
<tr>
<td>DEL₁ρ</td>
<td>0.25</td>
<td>0.10</td>
<td>0.42</td>
<td>0.17</td>
<td>1.09</td>
<td>0.40</td>
</tr>
<tr>
<td>DEL₂ρ</td>
<td>0.37</td>
<td>0.23</td>
<td>0.34</td>
<td>0.06</td>
<td>1.10</td>
<td>0.40</td>
</tr>
<tr>
<td>DEL₃ρ</td>
<td>0.27</td>
<td>0.08</td>
<td>0.36</td>
<td>0.05</td>
<td>1.10</td>
<td>0.41</td>
</tr>
</tbody>
</table>

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. DEL₁ρ, the extended double empirical likelihood estimator \( \hat{\beta}_ρ \) that allows for a different baseline hazard function for the aggregate data.
Remarks

- More flexible than the conventional meta-analysis; automatically combine survival information for different subgroups; the information may be derived from different studies.

E.g., one study may publish survival probabilities for different age groups, while the other study may publish survival probabilities for different disease stages.

- 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.

- 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.