Cox model versus generalized logrank test for time–to–event data with ties

Catalina Stefanescu*
London Business School, Regent’s Park, London NW1 4SA, UK

and

Devan Mehrotra
Merck Research Laboratories, Blue Bell, PA 19422, USA

Summary

The Cox proportional hazards model is the standard tool for relative risk estimation and inference for survival studies. The generalized logrank (GLR) approach (Mehrotra and Roth, 2001) has been proposed as a more efficient alternative to the Cox model when the number of subjects is small and there are no ties in the data. However, ties will result if continuous data are grouped by rounding to the nearest day, week, etc. This paper investigates the use of bootstrap techniques to improve the efficiency of relative risk estimators and the accuracy of confidence intervals within the GLR framework in the presence of ties. Results from a simulation study are used to compare the Efron

Key words: Cox regression; Generalized logrank test; Bootstrap; Tied failures
tie handling method for the Cox model with bootstrap and non-bootstrap versions of $GLR^E$, a GLR–based extension of the Efron approximation.

1. Introduction

Clinical studies often involve the comparison of two survival distributions. For example, subjects are randomized to two treatment groups, the time from randomization until the occurrence of a prespecified event (death, stroke, etc.) is recorded, then the survival distributions in the two groups are compared to determine the existence of a treatment effect. In this framework, the proportional hazards model (Cox, 1972) is the standard tool for relative risk estimation and inference. However, the performance of the Cox model can be less than optimal when the samples in the study are small. This is likely to be the case, for example, in phase II clinical trials where estimates of the relative risk are obtained from relatively small samples (20–30 subjects) and then are used to design larger phase III trials. In these settings, the point and interval estimates of relative risk based on the Cox model are not very efficient.

Mehrotra and Roth (2001) proposed the generalized logrank (GLR) approach as a more efficient alternative to the Cox model when the number of subjects is small. The GLR approach is asymptotically equivalent to the Cox model approach and can be used to test for departures from any hypothesized value of the relative risk. The method has been developed for continuous data. However, in practice, tied observations will occur if continuous data are grouped by rounding to the nearest day, week, etc, and in these cases the GLR approach is not recommended.
Several methods have been proposed for dealing with ties within the Cox model framework. Cox (1972), Kalbfleisch and Prentice (1973), Breslow (1974) and Efron (1977) have developed tie–handling approximations of the partial likelihood which have been implemented in standard statistical software. While all methods will generally yield similar results when the number of ties is small, their performance varies greatly when there are many tied event times arising from a coarse grouping of continuous data. This could be the case, for example, in clinical trials where the study population is large and the outcome common, or large cohort studies where time is measured in months. Hertz–Picciotto and Rockhill (1997) compared the performance of three tie–handling approximations under no censoring through a simulation study, and found that the Efron approximation performed far better than the Breslow and the Kalbfleisch and Prentice approaches.

This paper proposes the use of bootstrap techniques to improve the efficiency of relative risk estimators and the accuracy of confidence intervals within the GLR framework when ties are present. We investigate $GLR^E$, an extension of the GLR statistic for grouped data, which in special cases reduces to the score statistic based on the Efron partial likelihood. This will generally be biased in the presence of heavy ties. We use nonparametric bootstrap techniques to estimate the bias, then we adjust the point estimate and the confidence intervals computed from $GLR^E$.

The paper is structured as follows: Section 2 briefly reviews the Cox partial likelihood and the Efron tie–handling approximation. We describe the GLR statistic in Section 3, then we discuss $GLR^E$ and its bootstrap adjusted version $GLR_{B}^E$. Section 4 presents an application to the analysis of data.
from a clinical trial, as well as the results of a simulation study conducted to compare the performance of the Efron tie handling approximation with that of $GLR^E$ and $GLR_B^E$. Section 5 concludes the paper with a discussion of the results and an outline of several directions for future investigation.

2. The Cox Partial Likelihood

Let $N_A$ and $N_B$ individuals be randomized to treatment groups $A$ and $B$, respectively. Let the ordered (distinct) failure times for the combined data be $t_1 < t_2 < \ldots < t_k$. Denote by $D(t_i)$ the set of subjects who fail at time $t_i$, and let $d_i = d_{iA} + d_{iB}$ be the size of $D(t_i)$, where $d_{iA}$ and $d_{iB}$ subjects fail from treatment group $A$, respectively $B$. Denote by $R(t_i)$ the set of subjects at risk at time $t_i$, and let $n_i = n_{iA} + n_{iB}$ be the size of $R(t_i)$, where $n_{iA}$ and $n_{iB}$ subjects are at risk in treatment group $A$, respectively $B$. Let $Z_{ij}$ with $i = 1, \ldots, k$, $j = 1, \ldots, d_i$ be the group indicators for the individuals who fail at $t_i$, where $Z_{ij} = 1$ if the individual belongs to treatment group $A$ and $Z_{ij} = 0$ otherwise. Then $d_{iA} = \sum_{j=1}^{d_i} Z_{ij}$, for $i = 1, \ldots, k$. We assume throughout that censoring is non–informative and that the survival distributions in the two treatment groups have proportional hazard functions. Let $\lambda_A(t)$ and $\lambda_B(t)$ denote the hazard rates at time $t$ for treatments $A$ and $B$ respectively. The hazard ratio $\theta = \lambda_A(t)/\lambda_B(t)$ is constant over time and is commonly referred to as the relative risk. Let $\beta = \ln(\theta)$.

The data at time $t_i$ can be summarized in a $2 \times 2$ table as follows:

<table>
<thead>
<tr>
<th></th>
<th>Failed</th>
<th>Survived</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>$d_{iA}$</td>
<td>$n_{iA} - d_{iA}$</td>
<td>$n_{iA}$</td>
</tr>
<tr>
<td>Group B</td>
<td>$d_{iB}$</td>
<td>$n_{iB} - d_{iB}$</td>
<td>$n_{iB}$</td>
</tr>
<tr>
<td>Total</td>
<td>$d_i$</td>
<td>$n_i - d_i$</td>
<td>$n_i$</td>
</tr>
</tbody>
</table>
If there are no tied event times \((d_i = 1 \text{ for } i = 1, \ldots, k)\), the Cox partial likelihood function (Cox, 1972) is given by

\[
L(\beta) = \prod_{i=1}^{k} \frac{e^{Z_i \beta}}{\sum_{j \in R(t_i)} e^{Z_j \beta}} = \prod_{i=1}^{k} \frac{e^{d_i A \beta}}{n_i A e^\beta + n_i B}.
\]

When ties are present, the approximation to the likelihood proposed by Efron (1977) is given by

\[
L^E(\beta) = \prod_{i=1}^{k} L^E_i(\beta),
\]

where

\[
L^E_i(\beta) = \frac{d_i ! e^{d_i A \beta}}{\prod_{m=1}^{d_i} \left[ \sum_{l \in R(t_i)} e^{d_l A \beta} - \frac{m-1}{d_i} \cdot \sum_{l \in D(t_i)} e^{Z_l \beta} \right]}.
\]

This is the default approximation implemented in the S–Plus software.

3. The Generalized Logrank Statistic

3.1 Continuous Data (No Ties)

The generalized logrank statistic for continuous data has been proposed by Mehrotra and Roth (2001). Its derivation is based on the observation that at each time \(t_i\), if the number of failures \(d_{iB}\) in treatment group \(B\) arises from a binomial distribution \(B(n_{iB}, p_i)\), then under the proportional hazards assumption the number of failures \(d_{iA}\) in treatment group \(A\) arises from a binomial \(B(n_{iA}, \theta p_i)\) distribution, with \(0 \leq p_i \leq \min(1, \theta^{-1})\).

Let \(D_{iA}\) be the random variable denoting the number of events in group \(A\) at \(t_i\). The conditional distribution of \(D_{iA}\) given \(\Theta = (n_{iA}, n_{iB}, d_i, \theta, p_i)\) is non–central hypergeometric with

\[
P(D_{iA} = d_{iA} \mid \Theta) = \frac{n_{iA} \choose d_{iA}}{n_{iB} \choose d_{iB}} \frac{n_{iB} \choose d_{iB}}{n_{iA} \choose d_{iA}} \theta^{d_{iA}} (1 - \theta p_i)^{n_{iA} - d_{iA}} (1 - p_i)^{n_{iB} - d_{iB}} (1 + d_{iA}) \sum_{j \in G_i} \frac{n_{iA} \choose j}{n_{iB} \choose d_{iB} - j} \theta^j (1 - \theta p_i)^{n_{iA} - j} (1 - p_i)^{n_{iB} - d_{iB} + j}.
\]
where \( G_i = \{ j | \max(0, d_i - n_{iB}) \leq j \leq \min(d_i, n_{iA}) \} \).

Let \( E_{iA}(n_{iA}, n_{iB}, \theta, p_i) \) and \( V_{iA}(n_{iA}, n_{iB}, \theta, p_i) \) be the mean and variance of \( D_{iA} \) based on the conditional distribution (1). With continuous data we obtain

\[
E_{iA}(n_{iA}, n_{iB}, \theta, p_i) = \frac{n_{iA} \theta (1 - p_i)}{n_{iA} \theta (1 - p_i) + n_{iB} (1 - \theta p_i)}
\]

\[
V_{iA}(n_{iA}, n_{iB}, \theta, p_i) = \frac{n_{iA} n_{iB} \theta (1 - p_i)(1 - \theta p_i)}{[n_{iA} \theta (1 - p_i) + n_{iB} (1 - \theta p_i)]^2}.
\]

Then the GLR statistic for testing \( H_0 : \theta = \theta_0 \) versus \( H_A : \theta \neq \theta_0 \) is given by

\[
GLR(\theta_0, \tilde{p}) = \frac{\sum_{i=1}^{k} [d_{iA} - E_{iA}(n_{iA}, n_{iB}, \theta_0, \tilde{p}_i)]^2}{\sum_{i=1}^{k} V_{iA}(n_{iA}, n_{iB}, \theta_0, \tilde{p}_i)},
\]

where \( \tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_k) \) and \( \tilde{p}_i \) is an estimate of \( p_i \). When \( \theta_0 = 1 \), (2) reduces to Mantel’s logrank statistic (Mantel, 1966).

There are several ways of estimating the nuisance parameters \( p_i \). One possibility is a conditional approach which chooses the values of \( p_i \) that maximize the ”partial likelihood”

\[
\prod_i P(D_{iA} = d_{iA} | d_i, n_{iA}, n_{iB}, p_i, \theta).
\]

However, \( d_i \) may contain information about \( p_i \) that is neglected by conditioning on \( d_i \). Therefore, it may be preferable to take an unconditional approach and choose \( p_i \) that maximizes \( \prod_i L(p_i|\theta) \), where \( L(p_i|\theta) \) is the numerator of (1). Note that \( L(p_i|\theta) \) is proportional to the product of the two unconditional binomial likelihoods \( B(R_{iA}, \theta p_i) \) and \( B(R_{iB}, p_i) \). The estimates of \( p_i \) in the domain \( 0 \leq p_i \leq \min(1, \theta^{-1}) \) are then given by

\[
\tilde{p}_i = \tilde{p}_{i,\theta} = \frac{x_i - \sqrt{x_i^2 - 4(n_{iA} + n_{iB})d_i\theta}}{2(n_{iA} + n_{iB})\theta}
\]
where \( x_i = \theta (n_{iA} + d_i - d_{iA}) + n_{iB} + d_{iA} \). Small values of \( GLR[\theta_0, \tilde{p}(\theta_0)] \) support the null hypothesis \( H_0 : \theta = \theta_0 \). Therefore, the GLR estimator of \( \theta \) is the value \( \tilde{\theta}_{GLR} \) which satisfies

\[
GLR[\tilde{\theta}_{GLR}, \tilde{p}(\tilde{\theta}_{GLR})] = \inf_{\theta} GLR[\theta, \tilde{p}(\theta)].
\]

Let \( k^* = \sum_{i=1}^{k} \min(d_i, n_i - d_i, n_{iA}, n_{iB}) \). Mehrotra and Roth (2001) propose using the \( F(1, k^*) \) distribution as an approximation to the reference distribution of the GLR statistic (2) under the null hypothesis. When there are no ties in the data, \( k^* = \) is the number of 'informative' tables. Note also that \( F(1, k^*) \to \chi_1^2 \) as \( k^* \to \infty \), and that \( \chi_1^2 \) is the reference distribution for Mantel’s logrank statistic corresponding to the case \( \theta_0 = 1 \). With this approximation, a 100(1 − \( \alpha \))% confidence interval for \( \theta \) is given by

\[
\theta_{GLR}^L = \inf_{\theta} \{ \theta | GLR[\theta, \tilde{p}(\theta)] \leq F_\alpha(1, k^*) \}
\]

\[
\theta_{GLR}^U = \sup_{\theta} \{ \theta | GLR[\theta, \tilde{p}(\theta)] \leq F_\alpha(1, k^*) \}.
\]

### 3.2 Tied Observations

When there are many ties in the data, the GLR approach will not work well because the approximation of the likelihood using two independent binomial distributions for the two groups is no longer reasonable. A simple extension of the GLR statistic for grouped data can be obtained by replacing \( E_{iA} \) and \( V_{iA} \) in (2) with their means over all possible orderings of events between \( t_{i-1} \) and \( t_i \), for \( i = 1, \ldots, k \). Note that the "average" table at event time \( t_{ij} \) is given by Table 1, for \( i = 1, \ldots, k \) and \( j = 1, \ldots, d_i \). Accordingly, let
## Table 1

*“Average 2 × 2 Table” at $t_{i,j}$*

<table>
<thead>
<tr>
<th>Group</th>
<th>Failed</th>
<th>Survived</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>$\frac{d_{iA}}{d_i}$</td>
<td>$n_{iA} - j \cdot \frac{d_{iA}}{d_i}$</td>
<td>$n_{iA} - (j - 1) \cdot \frac{d_{iA}}{d_i}$</td>
</tr>
<tr>
<td>Group B</td>
<td>$\frac{d_{iB}}{d_i}$</td>
<td>$n_{iB} - j \cdot \frac{d_{iB}}{d_i}$</td>
<td>$n_{iB} - (j - 1) \cdot \frac{d_{iB}}{d_i}$</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>$n_i - j$</td>
<td>$n_i - (j - 1)$</td>
</tr>
</tbody>
</table>

$\bar{E}_{iA} = \sum_{j=1}^{d_i} E_{iA} \left( n_{iA} - (j - 1) \frac{d_{iA}}{d_i}, n_{iB} - (j - 1) \frac{d_{iB}}{d_i}, \theta, \bar{p}_{i,j} \right)$

and

$\bar{V}_{iA} = \sum_{j=1}^{d_i} V_{iA} \left( n_{iA} - (j - 1) \frac{d_{iA}}{d_i}, n_{iB} - (j - 1) \frac{d_{iB}}{d_i}, \theta, \bar{p}_{i,j} \right)$.

Then the GLR statistic for grouped data is

$$GLR^E(\theta_0, \bar{p}) = \frac{\{\sum_{i=1}^{k} [d_{iA} - \bar{E}_{iA}^E(n_{iA}, n_{iB}, \theta_0, \bar{p}_{i,j})]\}^2}{\sum_{i=1}^{k} \bar{V}_{iA}^E(n_{iA}, n_{iB}, \theta_0, \bar{p}_{i,j})}.$$ (3)

The parameters $p_{i,j}$ may be estimated with the values $\bar{p}_{i,j}$ that maximize

$$\Pi_{i,j} B(n_{iA}, \theta p_{i,j}) \cdot B(n_{iB}, p_{i,j})$$

and are given by

$$\bar{p}_{i,j} = \frac{h_i - \sqrt{h_i^2 - 4(n_{iA} + n_{iB} - j + 1)\theta}}{2(n_{iA} + n_{iB} - j + 1)\theta}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, d_i$$

where

$$h_i = \theta \left[ n_{iA} + \frac{d_{iB}}{d_i} - (j - 1) \frac{d_{iA}}{d_i} \right] + n_{iB} + \frac{d_{iA}}{d_i} - (j - 1) \frac{d_{iB}}{d_i}, \quad i = 1, \ldots, k.$$
Note that if $p_{i,j} = 0$ for all $(i, j)$, the $GLR^E$ statistic is the score statistic based on the Efron partial likelihood. Also, when all the risk sets (i.e. $n_i$’s) are very large, $GLR$ is approximately equivalent to Cox, and $GLR^E$ is equivalent to the Efron approximation, since in this case $\bar{p} \simeq 0$.

In order to derive confidence intervals and to conduct hypothesis tests, we need to approximate the reference distribution of the $GLR$–based statistic $GLR^E(\theta_0, \bar{p})$ under the null hypothesis $H_0 : \theta = \theta_0$. One possibility is to use the $F(1, k^*)$ distribution with $k^* = \sum_{i=1}^{k} \min(d_i, n_i - d_i, n_iA, n_iB)$, as described in Section 3.1. However, while the $F$ distribution was a reasonable approximation in the case when there were no ties in the data, in practice this approximation does not seem to work very well in the case of grouped data.

3.3 Bootstrap–GLR for Tied Data

We propose using bootstrap techniques to estimate the null distribution of the $GLR^E(\theta_0, \bar{p})$ statistic, as an alternative to choosing the $F(1, k^*)$ approximation. The bootstrap estimated bias may then be used to adjust the confidence intervals and the point estimate $\tilde{\theta}_{GLR^E}$ from the initial data.

Specifically, we’ve implemented the following nonparametric bootstrap procedure:

1) Resample with replacement $N$ times from the original data in each treatment group.

2) Calculate $\tilde{\theta}_{GLR^E}^*$, the bootstrap version of the $GLR^E$ statistic based on the resampled data sets.

3) Repeat steps 1) and 2) $B$ times in order to obtain an estimate of the
bootstrap distribution.

Denote by \( \tilde{\theta}^*_{GLR^E,1}, \ldots, \tilde{\theta}^*_{GLR^E,B} \) the sequence of bootstrap estimates resulting from the previous algorithm and let \( \tilde{\theta}^*_{GLR^E} \) be their median. An estimate of the bias of the \( GLR^E \) statistic is given by \( b_1 = \tilde{\theta}^*_{GLR^E} - \bar{\theta}_{GLR^E} \).

Then the bootstrap–GLR estimator of the relative risk \( \theta \) is the bias corrected \( GLR^E \) estimate:

\[
\tilde{\theta}_{GLR^E}^B = \tilde{\theta}_{GLR^E} - b_1.
\]

In order to derive the \( 100(1 - \alpha)\% \) bias corrected confidence intervals we shall follow the methodology of Carpenter and Bithell (2000). Let

\[
b_2 = \Phi^{-1}\left[ \frac{\#\{\tilde{\theta}^*_{GLR^E} < \tilde{\theta}_{GLR^E}\}}{B} \right].
\]

Let \( Q_L \) and \( Q_U \) be the percentiles of the bootstrap distribution required for the lower and upper endpoints of the bias corrected confidence interval. These are given by

\[
Q_L = (B + 1)\Phi(2b_2 + z_{\alpha/2}), \quad Q_U = (B + 1)\Phi(2b_2 + z_{1-\alpha/2}).
\]

Then the \( 100(1 - \alpha)\% \) bias corrected confidence interval has endpoints

\[
\theta_{GLR^E}^L = \tilde{\theta}_{QL}, \quad \theta_{GLR^E}^U = \tilde{\theta}_{QU}.
\]

4. Examples

4.1 Illustrative Example

To illustrate the bootstrap–GLR approach described in Section 3.3, we analyze the data from a clinical trial reported in Table 2. The data are survival times for patients in two treatment groups with \( n = 20 \) subjects.
Table 2

Time (in weeks) to a particular event. + denotes a censored observation.

<table>
<thead>
<tr>
<th>Group</th>
<th>Time (in weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A (old therapy, $N_A = 20$)</td>
<td>2  2  4+  8  8+  12  12  12  12  12+  12+  16+  16+  20+  24+  24+  28+  28+  36  36</td>
</tr>
<tr>
<td>Group B (new therapy, $N_B = 20$)</td>
<td>4+  4+  4+  4+  8  12+  12+  16+  16+  16+  16+  20+  20+  24+  28+  28+  32+  32+  36  36+</td>
</tr>
</tbody>
</table>

in each group. Post–randomization visits were scheduled at weeks 2 and 4 and every 4 weeks afterwards. There are many tied observations due to the coarse grouping, and there is also heavy censoring — 72.5% of the survival times are censored.

Table 3 reports the estimated relative risks and 95% confidence intervals using the Cox model based tie–handling approaches and the GLR based estimators $GLRE$ and $GLR_E^B$. The point estimates vary widely and the confidence intervals are quite large for all the classic approximations to the Cox likelihood. The confidence interval is still large for the $GLRE$ estimator which is biased due to the heavy ties. After adjusting for bias using the bootstrap–GLR approach, the confidence interval becomes much narrower.

4.2 Simulation Results

We conducted an extensive simulation study in order to compare the performance of the proposed GLR based estimators $GLRE$ and $GLR_E^B$ with that of the Efron tie handling approximation to the Cox partial likelihood.

We simulated a clinical trial study with $N$ subjects in each treatment
### Table 3
Estimated Relative Risks (95% confidence intervals).

<table>
<thead>
<tr>
<th>Cox Model–Based Approaches</th>
<th>GLR Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breslow 4.45 (0.96, 20.60)</td>
<td>GLR&lt;sup&gt;E&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cox 8.71 (1.06, 71.60)</td>
<td>GLR&lt;sup&gt;E&lt;/sup&gt;</td>
</tr>
<tr>
<td>Efron 5.12 (1.10, 23.86)</td>
<td>3.76 (1.03, 18.01)</td>
</tr>
<tr>
<td>K &amp; P 7.95 (1.04, 60.85)</td>
<td>2.58 (2.03, 4.76)</td>
</tr>
</tbody>
</table>

The analysis is done at time $T$, when patients are censored if they had not yet experienced the event of interest. For fixed $T$, the entry times into trial were generated as independent draws $\{e_{ij}\}$ from a uniform $(0, T)$ distribution, where $i = 1, \ldots, N$ and $j = A, B$. The survival times $\{s_{ij}\}$ were generated from Weibull distributions, independent of the entry times. For group $A$ subjects the survival distribution was $\text{Weibull}(\sqrt{0.5} \theta, 2)$ and for group $B$ subjects the survival distribution was $\text{Weibull}(\sqrt{0.5}, 2)$. Therefore the hazard rates for treatments $A$ and $B$ are $\lambda_A(t) = t\theta$ and $\lambda_B(t) = t$, and the relative risk is $\theta$. The observed survival times at the end of the study are then $y_{ij} = \min(s_{ij}, T - e_{ij})$, $i = 1, \ldots, N$ and $j = A, B$. Let $\pi_A(T)$ and $\pi_B(T)$ be the probabilities that any given subject will be censored in treatment groups $A$, respectively $B$, for a given $T$ and hazard rates $\lambda_A(t)$ and $\lambda_B(t)$. Since the number of subjects in the two groups is the same, the average probability of censoring is $\pi(T) = (\pi_A(T) + \pi_B(T))/2$.

We investigated combinations of $\beta = \ln(\theta) = 0.0$, 0.6, and 1.6 with $N = 10$, 20, and 40 patients in each group. We used two values $T_1$ and $T_2$ for $T$, chosen such that $\pi(T_1) = 0$ and $\pi(T_2) = 0.50$, corresponding to 0% and
50% censoring. Ties were produced by rounding the data to the nearest 0.02 and 0.10. We used $B = 2000$ nonparametric bootstrap samples to estimate the bias of $GLR^E$ as described in Section 3.3.

In some replications, the sampled data sets resulted in monotone likelihoods (Bryson and Johnson, 1981) where the Cox model estimate of $\beta$ cannot be calculated because the likelihood does not have a global maximum. This is the case if the highest failure time in one group precedes the smallest failure time in the other group, and thus estimation of the proportionality of hazard functions cannot be achieved. Generally, we discarded the corresponding data sets from computing simulation summaries. However, in order to avoid selection biases we report simulation results only for those cases where at most 1% of the replications resulted in a monotone likelihood.

The empirical mean squared error (MSE) and coverage of the 95% confidence interval for $\beta = \ln \theta$ were evaluated for the Cox model with Efron's approximation, as well as for the $GLR^E$ and $GLR^E_B$ approaches based on 1000 replications. Table 4 reports the empirical mean squared errors and the relative efficiency of the $GLR^E_B$ estimator computed as

$$\%RE = 100 \times \frac{(MSE_{Cox}^E)}{(MSE_{GLR^E_B})}.$$ 

The relative efficiency was uniformly greater than 100 per cent, ranging from 108% to 151%. The efficiency of the bootstrap–GLR approach is especially large for trials with small number of patients on each treatment group ($N = 10$ or $N = 20$).

Table 5 reports the empirical coverage of the 95% confidence intervals. The bootstrap–GLR approach provided adequate coverage in all cases, and
the width of the $\text{GLR}_B^E$ intervals was typically within ±5% of the width of the Cox$^E$ intervals.

5. Discussion and Future Research

We have proposed a GLR–based approach for relative risk estimation and inference from survival studies with ties in the data, under the proportional hazards assumption. The bootstrap–GLR approach is easy to implement, and the resulting estimator is consistently more efficient in terms of mean squared error than the classic estimator based on the tie–handling method of Efron. The gains in efficiency relative to the Cox model are particularly large in experiments with small to moderate size samples. We conjecture that the $\text{GLR}_B^E$, GLR and Cox model approaches are asymptotically equivalent.

The bootstrap–GLR methodology developed here can be extended to handle stratification as well as multiple groups. Also, it would be interesting to study the GLR version of the Kalbfleisch–Prentice and Breslow approximations to the Cox partial likelihood. These and other generalizations are the subject of further research.

References


Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when,
Table 4
Empirical mean squared error and relative efficiency for $\beta = \ln \theta$ using Cox model with Efron’s approximation ($Cox^E$), and the $GLR^E$ and $GLR^E_B$ approaches. (NC: not calculated because more than 1% of the simulated data sets resulted in a monotone likelihood.)

<table>
<thead>
<tr>
<th>N</th>
<th>$\beta$</th>
<th>$Cox^E$</th>
<th>$GLR^E$</th>
<th>$GLR^E_B$</th>
<th>% RE</th>
<th>$Cox^E$</th>
<th>$GLR^E$</th>
<th>$GLR^E_B$</th>
<th>% RE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rounding at 0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>0.252</td>
<td>0.187</td>
<td>0.168</td>
<td>150</td>
<td>NC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>0.111</td>
<td>0.093</td>
<td>0.086</td>
<td>129</td>
<td>0.209</td>
<td>0.184</td>
<td>0.169</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.117</td>
<td>0.103</td>
<td>0.099</td>
<td>118</td>
<td>0.216</td>
<td>0.186</td>
<td>0.170</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>0.185</td>
<td>0.156</td>
<td>0.157</td>
<td>118</td>
<td>0.334</td>
<td>0.269</td>
<td>0.258</td>
<td>130</td>
</tr>
<tr>
<td>40</td>
<td>0.0</td>
<td>0.055</td>
<td>0.050</td>
<td>0.047</td>
<td>117</td>
<td>0.104</td>
<td>0.097</td>
<td>0.096</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>0.083</td>
<td>0.077</td>
<td>0.077</td>
<td>108</td>
<td>0.149</td>
<td>0.135</td>
<td>0.134</td>
<td>112</td>
</tr>
<tr>
<td>Rounding at 0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>0.245</td>
<td>0.178</td>
<td>0.162</td>
<td>151</td>
<td>NC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>0.110</td>
<td>0.092</td>
<td>0.085</td>
<td>130</td>
<td>0.207</td>
<td>0.183</td>
<td>0.169</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>0.173</td>
<td>0.150</td>
<td>0.154</td>
<td>113</td>
<td>NC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.0</td>
<td>0.054</td>
<td>0.049</td>
<td>0.046</td>
<td>117</td>
<td>0.102</td>
<td>0.096</td>
<td>0.093</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>Pending</td>
<td></td>
<td></td>
<td></td>
<td>0.148</td>
<td>0.135</td>
<td>0.134</td>
<td>110</td>
</tr>
</tbody>
</table>
Table 5

Empirical coverage of the 95% confidence interval for $\beta = \ln \theta$ using Cox model with Efron’s approximation ($\text{Cox}^E$) and the $\text{GLR}^E$ and $\text{GLR}^E_B$ approaches. (NC: not calculated because more than 1% of the simulated data sets resulted in a monotone likelihood.)

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\beta$</th>
<th>0% censoring</th>
<th>50% censoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\text{Cox}^E$</td>
<td>$\text{GLR}^E$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rounding at 0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>94.4</td>
<td>95.3</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>95.0</td>
<td>95.2</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>95.7</td>
<td>95.1</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>96.5</td>
<td>94.1</td>
</tr>
<tr>
<td>40</td>
<td>0.0</td>
<td>94.0</td>
<td>94.1</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>96.2</td>
<td>95.7</td>
</tr>
<tr>
<td>Rounding at 0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>94.5</td>
<td>95.9</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>95.0</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>97.0</td>
<td>93.4</td>
</tr>
<tr>
<td>40</td>
<td>0.0</td>
<td>93.9</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>1.6</td>
<td>Pending</td>
<td></td>
</tr>
</tbody>
</table>


