On Nelson–Aalen type estimation in the partial Koziol–Green model

By: Haimeng Zhang, M. Bhaskara Rao

Zhang, H. and Rao, M.B. (2010). On Nelson - Aalen type estimation in the partial Koziol-Green model. *Statistics*, *44(5)*, 455 - 465.

Abstract:

In this paper, a Nelson–Aalen (NA) type estimator is derived and its sample properties are compared with the partial Abdushukurov–Cheng–Lin (PACL), generalized maximum likelihood (GMLE), and Kaplan–Meier (KM) estimators under the partial Koziol–Green model. These comparisons are made through Monto Carlo simulations under various sample sizes. The results indicate that the NA estimator always performs better than the KM estimator and is competitive with other estimators. Moreover, the PACL, GMLE, and NA estimators are shown to be asymptotically equivalent.

Keywords: Koziol–Green model | partially informative censoring | partial Abdushukurov–Cheng–Lin estimator | generalized maximum likelihood estimator | Nelson–Aalen type estimator | 62N01 | 62G05 | 62G20

Article:

1. Introduction

Let X, Y, and Z be three independent random variables with unknown continuous cumulative distributions $F(t)$, $G_1(t)$, and $G_2(t)$ for $t \in \mathbb{R}^+ = [0, \infty)$, respectively. Assume $\{X_i\}_{1\leq i\leq n}$, $\{Y_i\}_{1\leq i\leq n}$, and $\{Z_i\}_{1\leq i\leq n}$ are three independent sequences of n *i.i.d.* copies of X, Y, and Z, respectively. Let $\bar{F}(t) = 1 - F(t)$ be the survival function corresponding to F. The Koziol–Green with partially informative censoring model, or PKG for short, is based on the following assumption:

$$\bar{G}_1(t) = \bar{F}(t)^{\beta}, \quad t \geq 0, \tag{1}$$

for some unknown $\beta>0$. Under the PKG model, we observe two *i.i.d.* random sequences $\{U_i\}_{1\le i\le n}, \{\Delta_i\}_{1\le i\le n}$, where $U_i = \min\{X_i, Y_i, Z_i\} = X_i \wedge Y_i \wedge Z_i$ and Δ_i given by

$$\Delta_i = \begin{cases} 1 & \text{if } X_i \le Y_i \wedge Z_i, \\ 0 & \text{if } Y_i \le X_i \wedge Z_i, \\ -1 & \text{if } Z_i \le X_i \wedge Y_i. \end{cases} \quad (2)$$

Here *X* could be censored by *Y* on the right, which is the case of informative censoring in view of Equation (1), or be censored by *Z*on the right, which is the case of uninformative censoring. This model was first proposed and studied by Gather and Pawlitschko [1].

To estimate the survival function F^{-}, a number of estimators have been proposed in the literature. These include a Kaplan–Meier (KM) estimator by ignoring the partially informative censorship of *Y*, a partial Abdushukurov–Cheng–Lin estimator (PACL) by Gather and Pawlitschko [1], and a generalized maximum likelihood (GMLE) estimator by Zhang and Rao [2]. In this paper, we consider a Nelson–Aalen type estimator (NA for short; see [3, 4]) for the estimation of the survival function F^{-}.

Our paper is organized as follows. In Section 2, we briefly introduce the KM, PACL, and GMLE estimators. The NA estimator is derived in Section 3. Small and large sample comparisons on the performance of KM, PACL, GMLE, and NA estimators via simulations are presented in Section 4.1. In Section 4.2, we investigate how these estimators work out graphically for a real data set. The asymptotic equivalence of the NA estimator and the PACL estimator is provided in Section 5. Some concluding remarks are given in Section 6.

2. KM, PACL, and GMLE estimators

Let $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$ be the ordered statistics of $U_1, U_2, \ldots, U_n$ and $\Delta_{[1]}, \Delta_{[2]}, \ldots, \Delta_{[n]}$ the concomitant Δ values of the ordered U_i s. First we consider the KM estimator. Let $W_i = \min\{Y_i, Z_i\}$ and $\eta_i = I[\Delta_i = 1]$ so that $U_i = \min\{X_i, W_i\}$, i.e. one ignores the partially informative censorship of $\{Y_i\}$ and combines Y_i and Z_i to yield a general random right censorship model with given observations $\{U_i\}_{1\le i\le n}$ and $\{\eta_i\}_{1\le i\le n}$. Here $I[A]$ represents the indicator function of an event *A*. Consequently, the KM estimator 5 can be exploited to estimate $F^{-}(t)$,

$$\hat{\bar{F}}_{\mathrm{KM}}(t) = \prod_{i=1}^{n}\left(1 - \frac{\eta_{[i]}}{n-i+1}\right)^{I[U_{(i)}\le t]}, \quad t \ge 0, \quad (3)$$

where $\eta_{[1]}, \ldots, \eta_{[n]}$ are the concomitant η values corresponding to the ordered U_i s. The subscript KM indicates the method of estimation used.

The PACL estimator, proposed by Gather and Pawlitschko [1], is an analogue of Abdushukurov–Cheng–Lin (ACL) estimator [6,7]. It is given by

$$\hat{\hat{F}}_{\text{PACL}}(t) = (\hat{\hat{K}}(t))^{\hat{p}_{\text{PACL}}}, \quad t \geq 0, \tag{4}$$

where

$$\hat{\hat{K}}(t) = \prod_{i=1}^{n} \left(1 - \frac{\varepsilon_{[i]}}{n-i+1}\right)^{I[U_{(i)} \leq t]}, \quad t \geq 0, \quad \text{and}$$

$$\hat{p}_{\text{PACL}} = \frac{\sum_{i=1}^{n} I[\Delta_i = 1]}{\sum_{i=1}^{n} I[\Delta_i \neq -1]}, \tag{5}$$

$$= 0 \quad \text{if the denominator above is 0.}$$

Here $\hat{p}_{\text{PACL}}$ is the estimator of $p = \mathbb{P}(\Delta = 1 | \Delta \neq -1) = 1/(1+\beta)$, which gives

$$\hat{\beta}_{\text{PACL}} = \frac{1 - \hat{p}_{\text{PACL}}}{\hat{p}_{\text{PACL}}}, \tag{6}$$

and $\varepsilon_{[1]}, \ldots, \varepsilon_{[n]}$ denote the concomitant ϵ-values of the ordered U_i s with $\varepsilon_i = I[\Delta_i \neq -1]$.

Zhang and Rao[2] derived a GMLE estimator involving the maximization of the generalized likelihood of the data over all distributions F of X and G_2 of Z. In view of the fact that asymptotics of the GMLE are derived under the assumption of continuous F and G_2, we assume that $0 = U_{(0)} < U_{(1)} < U_{(2)} < \cdots < U_{(n)}$. For notational simplicity, we define $D_j = I[\Delta_{[j]} = 1]$, $C_j = I[\Delta_{[j]} = 0]$, and $E_j = I[\Delta_{[j]} = -1]$. Obviously, $C_j + D_j + E_j = 1$, $j = 1, 2, \ldots, n$. Therefore, the generalized likelihood of the data is given by

$$L(F, G_1, G_2) = L = \prod_{i=1}^{n} [p_i \mathbb{P}(Y \geq U_{(i)}) \mathbb{P}(Z \geq U_{(i)})]^{D_i} [q_i \mathbb{P}(X > U_{(i)}) \mathbb{P}(Z \geq U_{(i)})]^{C_i}$$
$$\times [r_i \mathbb{P}(X > U_{(i)}) \mathbb{P}(Y > U_{(i)})]^{E_i}, \tag{7}$$

where $p_i = \mathbb{P}(X = U_{(i)})$, $q_i = \mathbb{P}(Y = U_{(i)})$, $r_i = \mathbb{P}(Z = U_{(i)})$, for i=1, 2, …, n. Therefore, maximizing the likelihood L over all distributions F and G_2 satisfying Equation (1) yields the GMLE given by 2

$$\hat{\hat{F}}_{\mathrm{GMLE}}(t) = \begin{cases} \prod_{j=1}^{i-1}(1-\hat{a}_j) & \text{if } U_{(i-1)} \le t < U_{(i)}, \quad i = 1, 2, \ldots, n, \\ \prod_{j=1}^{n}(1-\hat{a}_j) & \text{if } t \ge U_{(n)} \end{cases} \tag{8}$$

with the convention that the empty product is equal to 1. Here $\hat{a}_i$ is given by

$$\hat{a}_i = \begin{cases} (1+(n-i)(1+\hat{\beta}_{\mathrm{GMLE}}))^{-1} & \text{if } D_i = 1, \quad C_i = E_i = 0, \\ 1 - \left(\frac{1+(n-i)(1+\hat{\beta}_{\mathrm{GMLE}})}{(n-i+1)(1+\hat{\beta}_{\mathrm{GMLE}})}\right)^{1/\hat{\beta}_{\mathrm{GMLE}}} & \text{if } C_i = 1, \quad D_i = E_i = 0, \\ 0 & \text{if } E_i = 1, \quad D_i = C_i = 0 \end{cases} \tag{9}$$

and the GMLE estimator $\hat{\beta}_{\mathrm{GMLE}}$ of β is the solution of the following estimating equation

$$\sum_{C_i+D_i=1}\left[(n-i)D_i \log\frac{(n-i)(1+\beta)}{1+(n-i)(1+\beta)} - C_i\frac{(n-i)+1}{\beta^2}\log\frac{1+(n-i)(1+\beta)}{(n-i+1)(1+\beta)}\right] = 0. \tag{10}$$

Another version of the GMLE estimator (denoted as the GMLE1 estimator from now on) was also considered by Zhang and Rao[2]. It was obtained by substituting $\hat{\beta}_{\mathrm{GMLE}}$ with $\hat{\beta}_{\mathrm{PACL}}$ of Equation (6) in Equation (9). Note that both GMLE estimators are proper, i.e. $\prod_{j=1}^{n}(1-\hat{a}_j) = 0$ only if the last observation $U_{(n)}$ is uncensored.

It has been shown that the PACL estimator (4) is more efficient than the KM estimator (3) [1]. Simulations based on small and large samples have suggested that the PACL estimator and the GMLEs are asymptotically equivalent [2,8].

3. NA type estimator

We now derive the NA type estimator of β and *F*. Under the provision when $0 = U_{(0)} < U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ are distinct, the total likelihood, from Equation (7), can be rewritten as

$$L \equiv \prod_{i=1}^{n} [\bar{G}_1(U_i)\bar{G}_2(U_i)\mathrm{d}F(U_i)]^{D_i} [\bar{F}(U_i)\bar{G}_2(U_i)\mathrm{d}G_1(U_i)]^{C_i} [\bar{F}(U_i)\bar{G}_1(U_i)\mathrm{d}G_2(U_i)]^{E_i}$$

$$= \prod_{i=1}^{n} \bar{F}^{1+\beta}(U_i)\beta^{C_i} \left[\frac{\mathrm{d}F(U_i)}{\bar{F}(U_i)}\right]^{(C_i+D_i)} \quad \text{(up to a constant not related to unknown parameters).}$$

Let $\lambda(t)\mathrm{d}t = \mathrm{d}F(t)/\bar{F}(t)$ and $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$ be the hazard density and the cumulative hazard function, respectively. Rewrite the likelihood to obtain

$$L = \prod_{i=1}^{n} \exp\{C_{[i]} \log \beta - (1+\beta)\Lambda(U_{(i)}) + (C_{[i]} + D_{[i]}) \log \lambda(U_{(i)})\}.$$

Here $C_{[i]}$ and $D_{[i]}$ are the censoring indicators corresponding to $U_{(i)}, i = 1, 2, \ldots, n$. Letting the hazard mass at each observation $U_{(i)}$ be λ_i, we have $\Lambda(U_{(i)}) = \sum_{j=1}^{i} \lambda_j$. Hence, the likelihood is

$$L = \prod_{i=1}^{n} \exp\{C_{[i]} \log \beta - (1+\beta) \sum_{j=1}^{i} \lambda_j + (C_{[i]} + D_{[i]}) \log \lambda_i\}.$$

The log-likelihood will then be

$$\log L = \sum_{i=1}^{n} \left[C_{[i]} \log \beta - (1+\beta) \sum_{j=1}^{i} \lambda_j + (C_{[i]} + D_{[i]}) \log \lambda_i \right]$$

$$= \sum_{i=1}^{n} C_{[i]} \log \beta - (1+\beta) \sum_{i=1}^{n} \sum_{j=1}^{i} \lambda_j + (C_{[i]} + D_{[i]}) \sum_{i=1}^{n} \log \lambda_i$$

$$= \sum_{i=1}^{n} C_{[i]} \log \beta - (1+\beta) \sum_{i=1}^{n} (n-i+1)\lambda_i + \sum_{i=1}^{n} (C_{[i]} + D_{[i]}) \log \lambda_i.$$

Taking the derivatives of log L over β and λ_i s, respectively, to yield

$$\beta = \frac{\sum_{i=1}^{n} C_{[i]}}{\sum_{i=1}^{n} (n-i+1)\lambda_i}, \quad \lambda_i = \left(\frac{1}{1+\beta}\right) \frac{C_{[i]} + D_{[i]}}{n-i+1},$$

which give $\hat{\beta}_{\mathrm{NA}} = \hat{\beta}_{\mathrm{PACL}} = (1 - \hat{p}_{\mathrm{PACL}})/\hat{p}_{\mathrm{PACL}}$, and $\hat{\lambda}_i = \hat{p}_{\mathrm{PACL}}(C_{[i]} + D_{[i]})/(n - i + 1)$. Therefore, the survival function $\bar{F}(t)$ can be estimated by

$$\hat{\bar{F}}_{\mathrm{NA}}(t) = \prod_{\{i|U_i \le t\}} (1 - \hat{\lambda}_i) = \prod_{\{i|U_i \le t\}} \left(1 - \frac{\hat{p}_{\mathrm{PACL}}(C_{[i]} + D_{[i]})}{n - i + 1}\right), \quad t \ge 0. \tag{11}$$

It is worthwhile to notice that under general random censorship model, the KM estimator and the NA estimator are the same. However, this is not the case for the PKG model. Under the PKG model, one should first notice that if $\hat{\beta}_{\mathrm{PACL}} \neq 1$, the NA estimator $\hat{\bar{F}}_{\mathrm{NA}}(t)$ is always an improper survival function even if the last observation is a failure, which is not true for the PACL and GMLE estimators. As a simple example, we consider the following data set given by 2.

It is easy to see that $\hat{\beta}_{\mathrm{PACL}} = 1$ from Equation (6) and $\hat{\beta}_{\mathrm{GMLE}} = 2.423$ from estimating Equation (10). Simple calculation gives the following (Table 1).

Table 1. Estimates and likelihoods of PACL, GMLEs, and NA estimators.

Time	**PACL**	**GMLE**	**GMLE1**	**NA**
$0 \le t < 2$	1	1	1	1
$2 \le t < 5$	0.894	0.939	0.9	0.9
$5 \le t < 8$	0.775	0.856	0.771	0.788
$8 \le t < 10$	0.548	0.714	0.578	0.591
$t \ge 10$	0	0	0	0.295
Likelihood	0.0002262	0.0003057	0.0002315	0.0001146

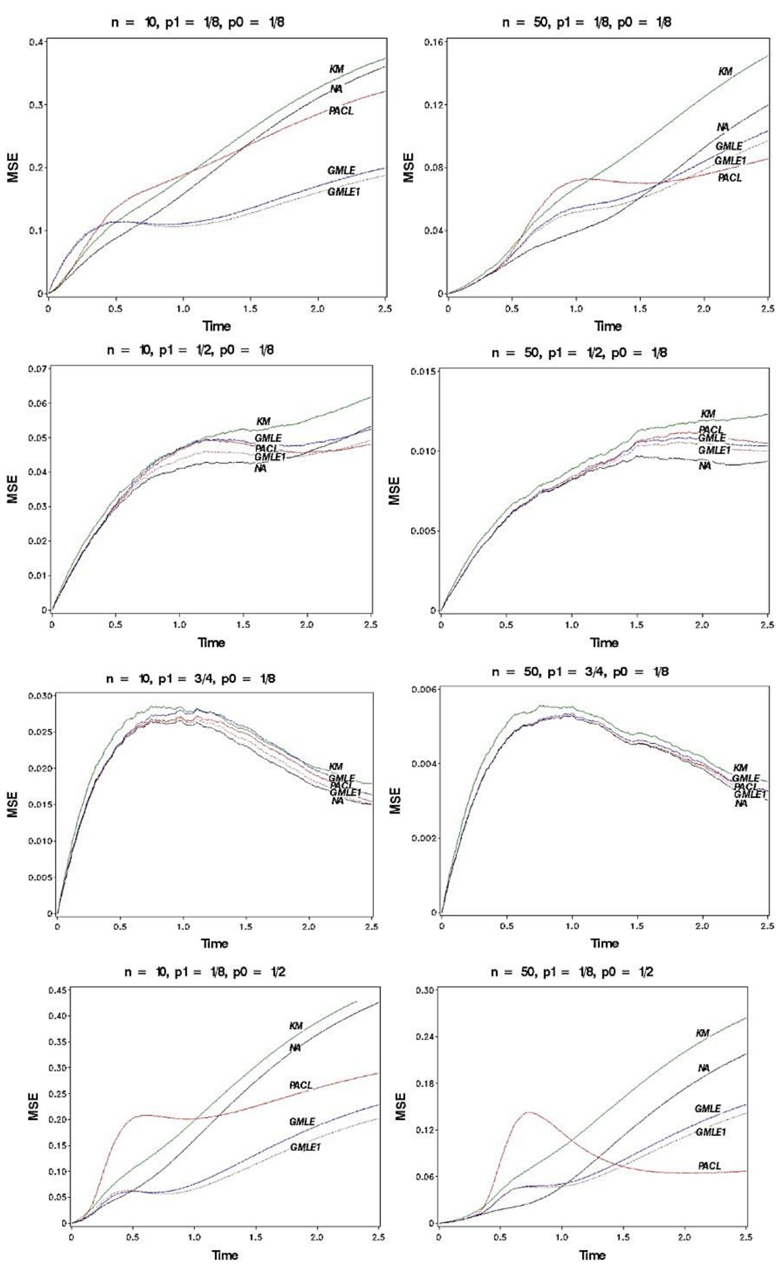
Obviously, the GMLE estimator yields the largest likelihood value as expected, but only the NA estimator gives an improper estimate.

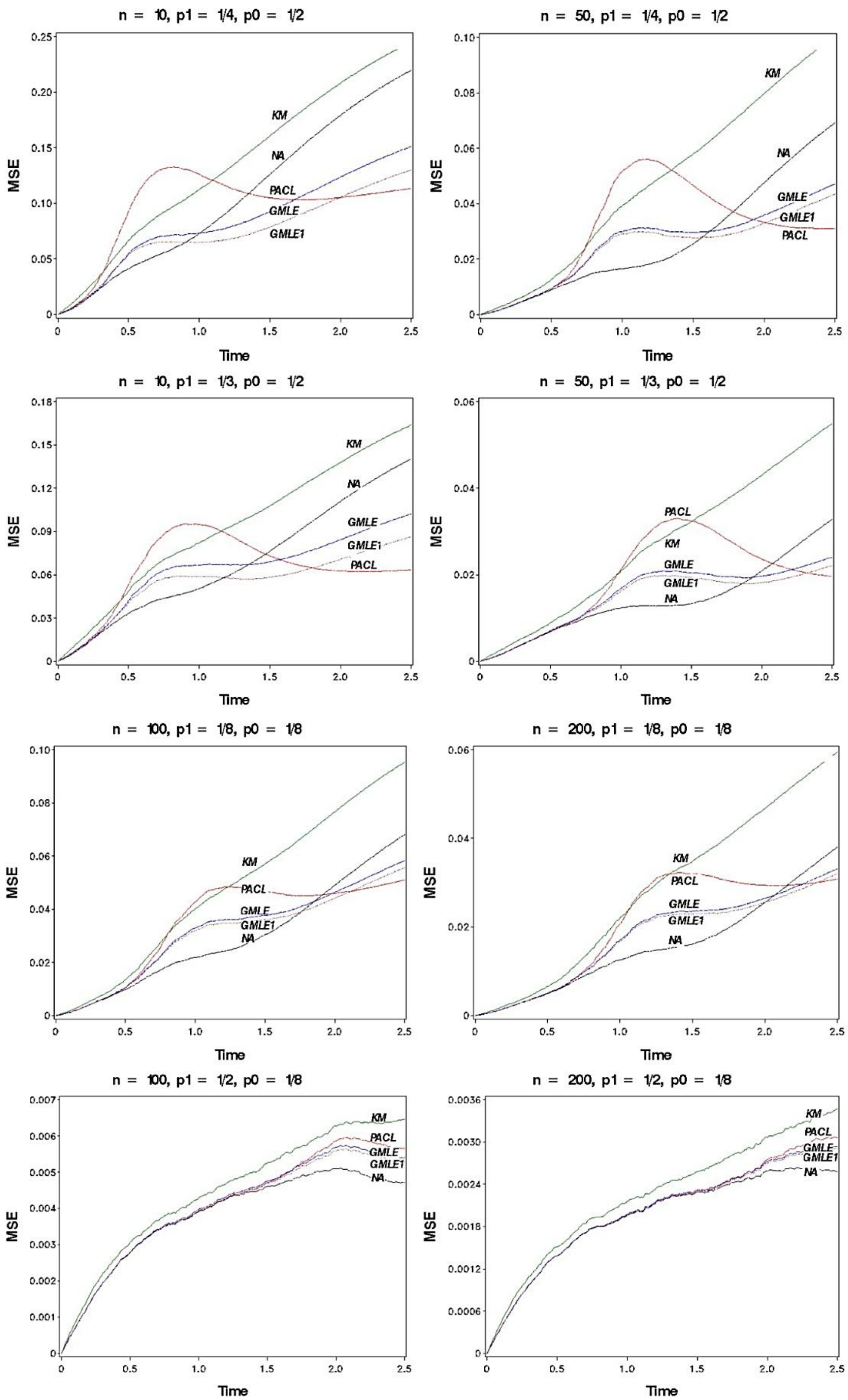
4. Simulation and real data results

4.1. Simulations

In this section, we investigate small and large sample properties of PACL, KM, GMLE, GMLE1, and NA estimators through a simulation study. Following Gather and Pawlitschko [1], we fix $p_i = \mathbb{P}(\Delta = i), i = 0, 1$, and let $\bar{F}(t) = \exp(-t), t \ge 0$ so that Y has the survival function $\bar{G}_1(t) = \exp(-\beta t), t \ge 0$, with $\beta = p_0/p_1$. We choose the non-informative censoring time Z to follow a gamma distribution with scale parameter equal to one which leads to the shape parameter $\gamma = -\ln(1 - p_0 - p_1)/\ln((p_0/p_1) + 2)$.

Figure 1 presents the simulated mean squared errors (MSE) curves for each of the estimators: KM estimator (3), PACL estimator(4), GMLE and GMLE1 estimators (8), and NA estimator (11). These curves are generated with the same selected combinations of p_0 and p_1 as reported in 1, but with the sample sizes of n=10, 50, 100, and 200. Simulations for n=100 and 200 are added to examine patterns of MSE curves under the large sample sizes. The MSE values are calculated, following 1, at 250 time points between 0 and 2.5, and each value is based on 5000 replications.




n = 10, p1 = 1/4, p0 = 1/2
KM
NA
PACL
GMLE
GMLE1
MSE
Time
n = 50, p1 = 1/4, p0 = 1/2
KM
NA
GMLE
GMLE1
PACL
MSE
Time
n = 10, p1 = 1/3, p0 = 1/2
KM
NA
GMLE
GMLE1
PACL
MSE
Time
n = 50, p1 = 1/3, p0 = 1/2
PACL
KM
GMLE
GMLE1
NA
MSE
Time
n = 100, p1 = 1/8, p0 = 1/8
KM
PACL
GMLE
GMLE1
NA
MSE
Time
n = 200, p1 = 1/8, p0 = 1/8
KM
PACL
GMLE
GMLE1
NA
MSE
Time
n = 100, p1 = 1/2, p0 = 1/8
KM
PACL
GMLE
GMLE1
NA
MSE
Time
n = 200, p1 = 1/2, p0 = 1/8
KM
PACL
GMLE
GMLE1
NA
MSE
Time

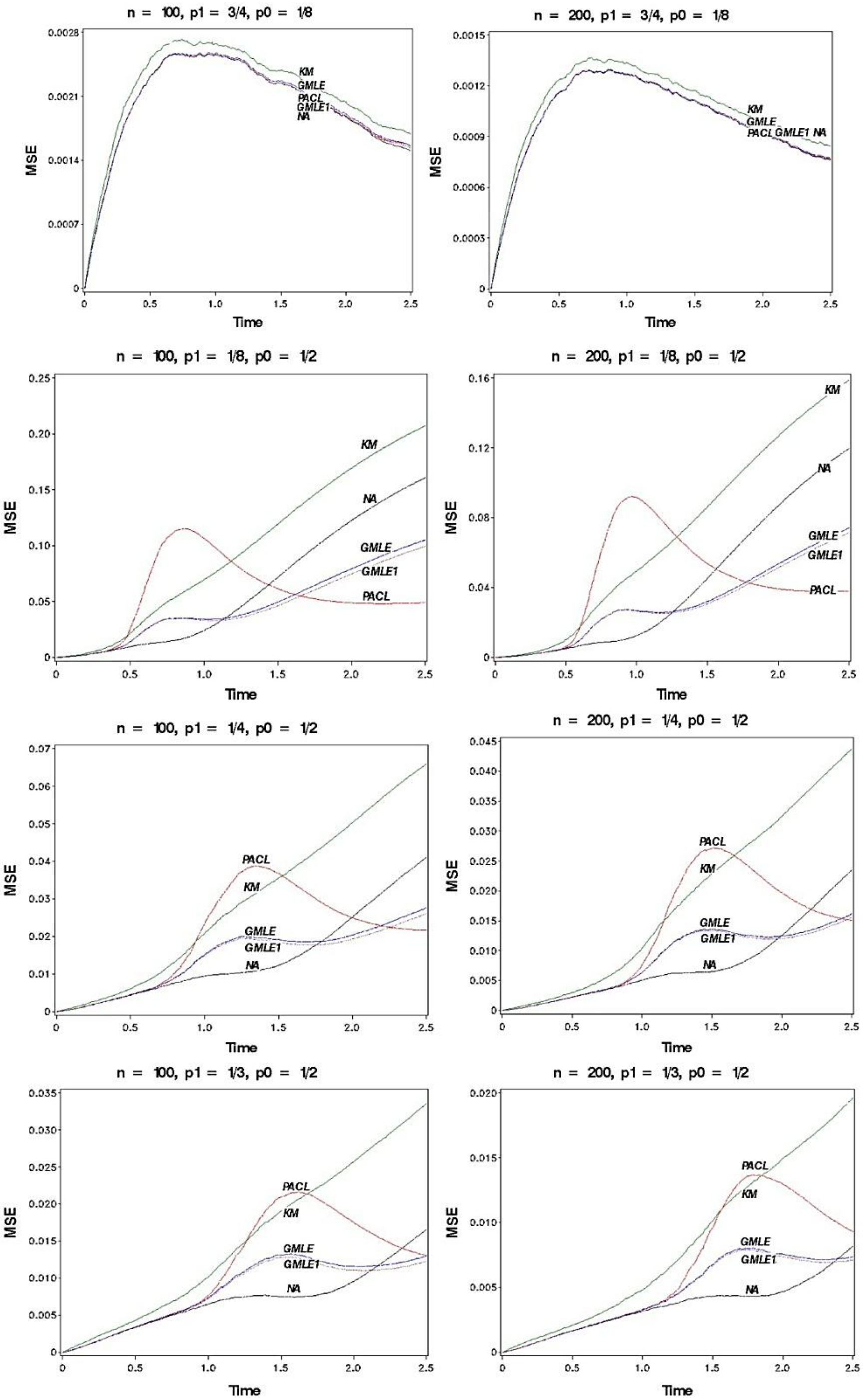


Figure 1. Simulated MSE curves.

From our simulation study, the PACL, GMLE, GMLE1, and NA estimators generally outperform the KM estimator in the sense of smaller MSEs, which is consistent with the results from Gather and Pawlitschko[1] and sample sizes *n*=10 and 50. Note that the NA estimator has the smaller MSEs than the KM estimator out of all selected combinations and sample sizes.
When *t* values are small or $p_1 = \mathbb{P}(\Delta = 1) \geq 1/2$, the NA estimator is the best out of all

proposed estimators. In all, the MSE curves from the NA estimator are moving smoothly along with the *t* values.

It should be noted that there are some regions of values of *t*, especially when the sample size is large and $\beta = p_0/p_1 \geq 1$, where the PACL and the GMLE estimators exhibit some abnormal behaviour (big humps). However, the MSE curves from the NA estimator always lie below those from proposed estimators. In summary, the NA estimator performs the best for small *t* values, while the PACL and GMLE estimators outperform the rest for large *t* values.

4.2. Real data example

We consider a real data set as reported in [9, Chapter 1.3] to investigate the survival time to the recurrence of the disease after a bone marrow transplantation for leukemia conducted at various hospitals in the USA and Australia. The data set consists of 38 patients who had been diagnosed with an acute lymphoblastic leukaemia, out of those there are 13 cases exhibiting uncensored times to relapse, 14 non-recurrence of leukaemia until the end of the study (observed up to seven years) (uninformative censoring), and 11 deaths before relapse during the period of follow-up (informative censoring). It has been tested that Assumption (1) holds, and in fact the PKG model provides a good fit to the data. In addition, the computed PACL estimator and the computed KM estimator for the time to relapse are very close to each other and the pointwise confidence intervals over a wide range of time based on the PACL estimator are smaller, which also indicates in favour of the PKG model[1]. The present paper provides a new estimate of the survival curves in addition to the ones provided by Gather and Pawlitschko[1] and Zhang and Rao[2].

Figure 2 (below) shows the survival curves of the KM estimator, PACL estimator, GMLE estimators, and NA estimator for the time to relapse. One should notice that the PACL estimator, GMLE estimators, and NA estimator jump at any observation that is uncensored or informatively censored while the KM estimator jumps only at the uncensored observations since it ignores the informative censoring *Y*, the same phenomena as noted in Gather and Pawlitshcko[1]. In addition, these estimators are very close to each other, agreeing up to three decimals.

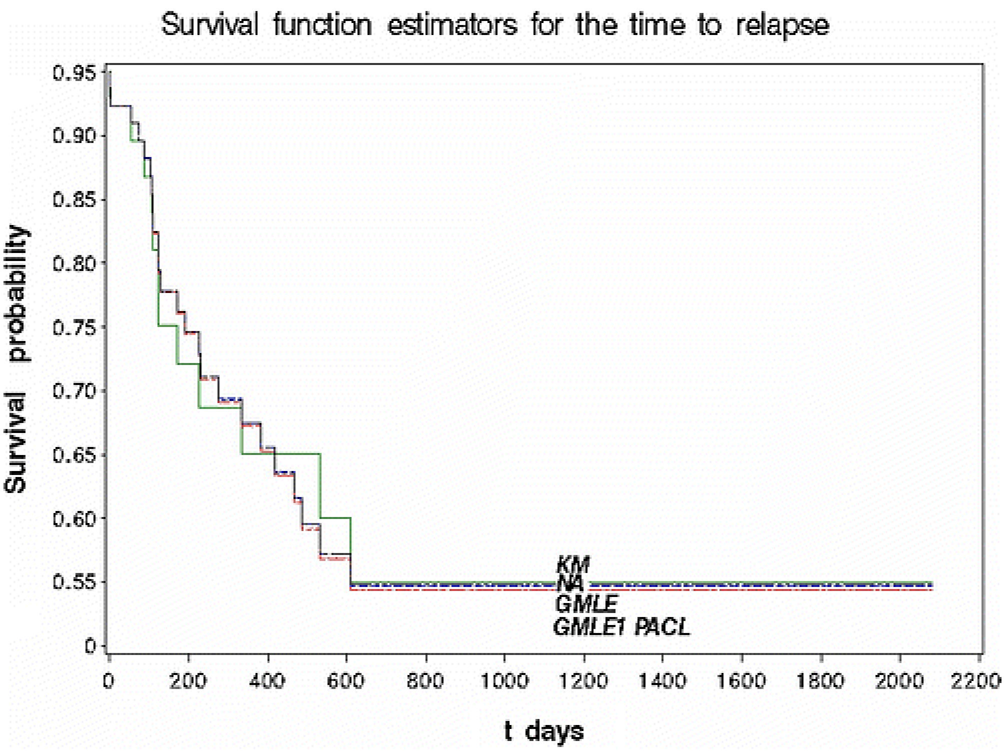


Figure 2. Survival function estimators for the time to relapse.

5. Asymptotic normality of NA estimator

It has been shown that the PACL estimator (4) is more efficient than the KM estimator (3)[1]. Simulations based on small and large samples have suggested that the PACL estimator and the GMLE are asymptotically equivalent[2,8]. In this section, we provide an asymptotic equivalence of the NA estimator with the PACL estimator (and therefore with the GMLE estimators from Zhang and Rao[8]).

Let $F(t)$ and $G_2(t)$ be fixed continuous distribution functions and $\beta = \beta_0$ be fixed. Let $p_i = \mathbb{P}(\Delta = i), i = 0, 1$ and $\alpha_0 = 1/(1+\beta_0) = p_1/(p_1 + p_0)$. We also let $\tau_1 = \sup\{t, \bar{F}(t) > 0\}$ and $\tau_2 = \sup\{t, \bar{G}_2(t) > 0\}$. We focus on the time interval $[0, T_0]$, where $T_0 < \min\{\tau_1, \tau_2\}$. Our main results are given by the following theorems.

Theorem 1

(Asymptotic equivalence of estimators of $\bar{F}(t)$) $\hat{\bar{F}}_{\mathrm{NA}}(t)$ *and* $\hat{\bar{F}}_{\mathrm{PACL}}(t)$ *are asymptotically equivalent, i.e.* $\sup_{0 \le t \le T_0} n^{1/2} |\hat{\bar{F}}_{\mathrm{NA}}(t) - \hat{\bar{F}}_{\mathrm{PACL}}(t)| \to_{\mathrm{a.s.}} 0.$

From Theorems 4.3 and 4.4 of Gather and Pawlitschko[1] and Theorem 1, the following asymptotic results for the NA estimator are immediate.

Theorem 2

1. (Law of the iterated logarithm) *With probability one,*

$$\sup_{0\le t\le T_0} |\hat{\hat{\bar{F}}}_{\mathrm{NA}}(t) - \bar{F}(t)| = O(n^{-1/2}(\log\log n)^{1/2}).$$

2. (Weak convergence) *The sequence of random processes* $\{n^{1/2}(\hat{\hat{\bar{F}}}_{\mathrm{NA}}(t) - \bar{F}(t)), 0 \le t \le T_0\}$ *converges weakly to the Gaussian process* $W(t)$ *with mean* $\mathbb{E}W(t) = 0$, *and for* $s, t \in [0, T_0]$,

$$\mathrm{Cov}(W(s), W(t)) = \alpha_0 \bar{F}(s)\bar{F}(t)\left(C(s,t) + \frac{1-\alpha_0}{p_0+p_1}\log(\bar{F}(t))\log(\bar{F}(s))\right)$$

with $C(s,t) = \int_0^{s\wedge t}(1/(\bar{F}(u))^{1/\alpha_0}\bar{G}_2(u))(\mathrm{d}F(u)/\bar{F}(u))$.

Proof of Theorem 1

Obviously, it is sufficient to prove

$$\sup_{0\le t\le T_0} n^{1/2}|\hat{\Lambda}_{\mathrm{NA}}(t) - \hat{\Lambda}_{\mathrm{PACL}}(t)| \longrightarrow_{\mathrm{a.s.}} 0,$$

where $\hat{\Lambda}_{\mathrm{NA}}(t) = -\log \hat{\hat{\bar{F}}}_{\mathrm{NA}}(t)$ and $\hat{\Lambda}_{\mathrm{PACL}}(t) = -\log \hat{\hat{\bar{F}}}_{\mathrm{PACL}}(t)$. First, let $nK_n(t) = \sum_{i=1}^{n} I[U_i \le t]$ denote the number of observations of Us up to time t. Then Equations (11) and (4) can be rewritten as, respectively, for $t \le T_0$,

$$\hat{\hat{\bar{F}}}_{\mathrm{NA}}(t) = \prod_{i=1}^{nK_n(t)}\left(1 - \frac{\hat{p}_{\mathrm{PACL}}(C[i]+D[i])}{n-i+1}\right), \tag{12}$$

$$\hat{\hat{\bar{F}}}_{\mathrm{PACL}}(t) = \prod_{i=1}^{nK_n(t)}\left(1 - \frac{\varepsilon_{[i]}}{n-i+1}\right)^{\hat{p}_{\mathrm{PACL}}}. \tag{13}$$

Therefore,

$$\hat{\Lambda}_{\mathrm{NA}}(t) = -\sum_{i=1}^{nK_n(t)}\left(1 - \frac{\hat{p}_{\mathrm{PACL}}(C[i]+D[i])}{n-i+1}\right) \cong \hat{p}_{\mathrm{PACL}}\sum_{i=1}^{nK_n(t)}\frac{C[i]+D[i]}{n-i+1} + O(n^{-1}),$$

$$\hat{\Lambda}_{\mathrm{PACL}}(t) = -\hat{p}_{\mathrm{PACL}}\sum_{i=1}^{nK_n(t)}\log\left(1 - \frac{\varepsilon_{[i]}}{n-i+1}\right) \cong \hat{p}_{\mathrm{PACL}}\sum_{i=1}^{nK_n(t)}\frac{\varepsilon_{[i]}}{n-i+1} + O(n^{-1}),$$

where, the asymptotic equivalence ≅ is in the almost sure sense and uniformly over $[0, T_0]$. Theorem 1 follows. ▪

6. Concluding remarks

In this paper, we proposed a NA type's estimator in the Koziol–Green model with partial informative censoring. We compare the performance of the estimator with other existing estimators, including the KM estimator, PACL estimator, and GMLE estimators. It has been shown that in a substantial range of t values, the NA estimator outperforms all other proposed estimators in the sense of smallest MSE values.

It should be noted that the partial Koziol–Green model is a semiparametric model, where β is the parametric component while $\bar{F}$ is the non-parametric component of interest. It is always of interest to know if a proposed estimator of $\bar{F}$ uses the available data efficiently. Understandably, under the PKG model, the GMLE, PACL, and NA estimators do and are asymptotically efficient due to their asymptotic equivalence and the asymptotic efficiency of the PACL estimator[10].

References

1. Gather, U. and Pawlitschko, J. 1998. Estimating the survival function under a generalized Koziol–Green model with partially informative censoring. *Metrika*, 48: 189–207.

2. Zhang, H. and Rao, M. B. 2004. On generalized maximum likelihood estimation in the proportional hazards model with partially informative censoring. *Metrika*, 59: 125–136.

3. Nelson, W. 1969. Hazard plotting for incomplete failure data. *J. Qual. Technol.*, 1: 27–52.

4. Nelson, W. 1972. Theory and application of hazard plotting for censored failure data. *Technometrics*, 14: 945–966.

5. Kaplan, E. L. and Meier, P. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, 53: 457–481.

6. A. Abdushukurov, *On some estimates of the distribution function under random censorship*, Conference of Young Scientists, VINITI No. 8756-V, Math. Inst. Acad. Sci., Uzbek SSR, Tashkent, 1984

7. Cheng, P. E. and Lin, G. D. 1987. Maximum likelihood estimator of a survival function under the Koziol–Green proportional hazards model. *Statist. Probab. Lett.*, 5: 75–80.

8. Zhang, H. and Rao, M. B. 2006. A note on the generalized maximum likelihood estimator in partial Koziol–Green model. *Statist. Probab. Lett.*, 76: 813–820.

9. Klein, J. P. and Moeschberger, M. L. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer.

10. Zhang, H. 2004. Asymptotic efficiency of estimation in the partial Koziol–Green model. *J. Statist. Plann. Inference*, 124: 399–408.