Jenny J Zhang, Zhuoxin Sun, Han Yuan and Molin Wang*

Alternatives to the Kaplan–Meier estimator of progression-free survival

https://doi.org/10.1515/ijb-2019-0095
Received August 21, 2019; accepted June 22, 2020; published online September 16, 2020

Abstract: Progression-free survival (PFS), defined as the time from randomization to progression of disease or death, has been indicated as an endpoint to support accelerated approval of certain cancer drugs by the U.S. FDA. The standard Kaplan–Meier (KM) estimator of PFS, however, can result in significantly biased estimates. A major source for the bias results from the substitution of censored progression times with death times. Currently, to ameliorate this bias, several sensitivity analyses based on rather arbitrary definitions of PFS censoring are usually conducted. In addition, especially in the advanced cancer setting, patients with censored progression and observed death times have the potential to experience disease progression between those two times, in which case their true PFS time is actually between those times. In this paper, we present two alternative nonparametric estimators of PFS, which statistically incorporate survival data often available for those patients who are censored with respect to progression to obtain less biased estimates. Through extensive simulations, we show that these estimators greatly reduce the bias of the standard KM estimator and can also be utilized as alternative sensitivity analyses with a solid statistical basis in lieu of the arbitrarily defined analyses currently used. An example is also given using an ECOG-ACRIN Cancer Research Group advanced breast cancer study.

Keywords: bias reduction; progression-free survival; Kaplan–Meier estimator; nonparametric.

1 Introduction

The U.S. Food and Drug Administration (FDA) may approve a drug to treat serious or life-threatening diseases, where either no therapy currently exists or the drug shows significant improvement over current therapies, based on a surrogate endpoint that is reasonably likely to predict a clinical benefit (e.g., prolong survival). Such a process is known as accelerated approval. Progression-free survival (PFS), defined as the time from randomization to disease progression or death, has become increasingly more accepted as a cancer drug accelerated approval endpoint. The advantages that are central to PFS being such an accepted surrogate endpoint include (1) accessibility prior to observation of a survival benefit; (2) reflection of tumor growth, which could be highly correlated with cancer-related morbidity and death; and (3) non-subjectiveness to potential confounding introduced by subsequent therapies usually administered after tumor progression or termination of study therapy (https://www.fda.gov/media/71195/download).

However, missing progression data is a common problem with many PFS analyses, especially in trials with higher non-compliance rates (possibly due to higher toxicities or a lower quality of trial conduct). In addition, the assessment of disease progression can be time-consuming as it requires documentation of tumor

Jenny J Zhang and Zhuoxin Sun contributed equally to this work.

***Corresponding author: Molin Wang**, Departments of Biostatistics and Epidemiology, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Building 2, Boston, MA 02115, USA; and Harvard Medical School, and Brigham Women's Hospital, Boston, MA 02115, USA, E-mail: stmow@channing.harvard.edu
Jenny J Zhang: Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; Johnson and Johnson, Shanghai, China
Zhuoxin Sun: ECOG-ACRIN Statistical Center, Frontier Science and Technology Research Foundation, Brookline, MA 02446, USA; and Department of Mathematical Science, University of Arkansas, Fayetteville, AR 72701, USA
Han Yuan: Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

measurements determined by various imaging methods, consultation with the treating physicians, and/or evaluation by a central review committee. Thus, it is generally true that survival data, which are relatively straightforward to assess, are more updated and complete than progression data at a given time during data collection. The presence of a substantial proportion of death–PFS events, defined as those events where patients are censored with respect to progression but have an observed death time, can greatly complicate the analysis of PFS data and induce bias.

Many patients with death–PFS events, especially in the advanced cancer setting, have the potential to experience undocumented disease progression after their progression censoring time and before death. For such patients, the standard definition of PFS would replace their censored progression times with their observed death times, thus, biasing the Kaplan–Meier (KM) PFS estimator upwards.

In practice, to mitigate the bias induced by death–PFS events, modified definitions of PFS are often used to conduct sensitivity analyses. Several sensitivity analyses are usually required for regulatory submissions where PFS is used as the primary endpoint of approval. However, these sensitivity analyses are usually based on rather arbitrary definitions of censoring times. One such definition that is commonly used is to censor the PFS time on the date of the last radiological tumor assessment if a death is observed after more than one missed visit, where the length of a visit varies depending on the specific disease. To date, there has been no established standard definition of PFS actually used in practice; even the FDA's guidance for cancer drug approval gives several definitions of PFS to be used in sensitivity analyses to determine the robustness of a primary PFS analysis (https://www.fda.gov/media/71195/download).

In this paper, we develop two alternative nonparametric PFS estimators that *statistically* incorporate available survival data for patients with censored progression times, and evaluate their performances with respect to each other and compared to the standard KM estimator through extensive simulations. Ruan and Gray [1] considered such an incorporation of survival information for censored progression patients in their PFS sensitivity analysis method with dependent censoring. However, their focus was not on estimation of the PFS function and their method required the computationally intensive EM algorithm. The estimators presented here are straightforward to calculate and can greatly reduce the bias in the KM estimator; in addition, they can be utilized as alternative sensitivity analyses.

In Section 2, we describe in detail the alternative PFS estimators: (1) a generalized KM estimator using survival time as the upper bound for PFS for those patients with death–PFS events; and (2) a nonparametric empirical estimator assuming that the true progression times for those patients with death–PFS events are similar to the progression times of the subset of patients with similar survival and observed progression. In Section 3, we compare the performance of the estimators through extensive simulation studies, and give an example using an ECOG-ACRIN metastatic breast cancer clinical trial (E2100) in Section 4. We close with some discussion in Section 5.

2 PFS estimators

2.1 Notation

We first introduce notation that will be used throughout the paper; other more specific notation will be defined as needed in individual sections.

Let n denote the total number of patients. Let T^p and T^d denote the times to progression and death, respectively, with corresponding censoring times $U^p \geq 0$ and $U^d \geq 0$. Thus, we observe $X^p = \min(T^p, U^p)$ and $X^d = \min(T^d, U^d)$ with corresponding censoring indicators $\delta^p = 1(T^p \leq U^p)$ and $\delta^d = 1(T^d \leq U^d)$. Let $T = \min(T^p, T^d)$ represent the composite endpoint time to PFS with corresponding survival function $S_T(.)$. Our goal in this paper is to estimate $S_T(.)$. Under the standard (naive) PFS definition, the censoring indicator is now $\max(\delta^p, \delta^d)$, denoted by δ^s. Note that $\delta^s = 0$ if and only if $\delta^p = \delta^d = 0$, in which case the corresponding censoring time U^s is U^p and $U^p \leq U^d$. We observe $X^s = \min(T, U^s)$ under this standard PFS definition. When $\delta^p = 1$ and $\delta^d = 0$ or 1, we refer to those patients with observed progression times as having progression-PFS events ($X^s = T^p$). When $\delta^p = 0$ and $\delta^d = 1$, we refer to those patients as having a death–PFS event ($X^s = T^d$). The true PFS time T is actually between U^p

and T^d in the presence of death–PFS events. Thus, there exists a clear systematic bias, where X^s will generally overestimate T. This systematic bias is what our proposed PFS estimators will attempt to correct.

2.2 Kaplan–Meier estimator

We assume the reader is familiar with the standard KM estimator [2] and will only give its definition here without any derivational details. This is calculated directly based on (X^s, δ^s) in the standard PFS definition. Let $t_1 < \ldots < t_k$ denote the k distinct observed PFS times in the standard PFS definition, d_j be the total number of progression-PFS events and death-PFS events at time t_j, and p_j represent the total number of patients at risk at time t_j. The KM estimator of PFS can be written as

$$\widehat{S}_T(t) = \prod_{m=1}^{j}\left(1 - \frac{d_m}{p_m}\right).$$

2.3 Generalized KM estimator

We assume, for those patients with death–PFS events ($\delta^p = 0$ and $\delta^d = 1$), that their PFS times are between their progression censoring and survival times (i.e., $U^p < T \le T^d$). We also assume independent censoring. Let $t_1 < \ldots < t_k$ denote the k distinct observed PFS times in the standard PFS definition, d_j be the total number of progression–PFS events and death–PFS events at time t_j, and m_j be the number of censored patients ($\delta^s = 0$) in (t_j, t_{j+1}). Then, the total number of patients at risk for PFS at time t_j is $n_j = (m_j + d_j) + \ldots + (m_k + d_k)$. Let h_j denote the hazard for PFS at time t_j, then $S_T(t_j) = \prod_{q=1}^{j}(1 - h_q)$.

To estimate the parameter vector $\mathbf{h} = (h_1, \ldots, h_k)$, we let A_j and B_j denote the set of all patients who experienced, respectively, progression–PFS and death–PFS events at time t_j, and consider the following likelihood:

$$\begin{aligned} L^*(\mathbf{h}) &= \prod_{j=1}^{k}\left[\left\{\prod_{l\in A_j} f_T(t_l)\prod_{l\in B_j} P\left(U^p = u_l^p, T^d = t_l^d, U^p < T \le T^d\right)\right\} S_T(t_j)^{m_j}\right] \\ &= \prod_{j=1}^{k}\left[\left\{\prod_{l\in A_j} f_T(t_l)\prod_{l\in B_j} P\left(u_l^p < T \le t_l^d \middle| U^p = u_l^p, T^d = t_l^d\right)\right.\right. \\ &\qquad \left.\left. \star P\left(U^p = u_l^p \middle| T^d = t_l^d\right) P\left(T^d = t_l^d\right)\right\} S_T(t_j)^{m_j}\right]. \end{aligned} \tag{1}$$

We assume that U^p is independent of (T, T^d). This is similar to the non-informative censoring assumption used in standard survival data analysis methods. This assumption is reasonable in the scenario of administrative censoring, but may not hold if the censoring is due to worsening of disease. Violation of this assumption could cause bias in the proposed KM estimator. Under this assumption, $L^*(\mathbf{h})$ can be written as $L^*(\mathbf{h}) = L(\mathbf{h}) \star \prod_{j=1}^{k}\left[\prod_{l\in B_j} P(T^d = t_l^d)\right]$, where

$$L(\mathbf{h}) = \prod_{j=1}^{k}\left[\left\{\prod_{l\in A_j} f_T(t_l)\prod_{l\in B_j} P\left(u_l^p < T \le t_l^d \middle| T^d = t_l^d\right)\right\} S_T(t_j)^{m_j}\right] \tag{2}$$

is a partial likelihood. Furthmore, we apply the following approximation

$$P\left(u_l^p < T \le t_l^d \middle| T^d = t_l^d\right) \approx P\left(u_l^p < T \le t_l^d\right) \tag{3}$$

to (2), where equality is achieved when T is independent of T^d. The resulting partial likelihood can be expressed as in (4). To model $P(u_l^p < T \le t_l^d | T^d = t_l^d)$ directly, e.g., using a semiparametric dependence model, would require complicated estimation methods that would severely limit the applicability/practicality of the

estimator. We expect some loss of efficiency likely due to using the partial likelihood instead of the full likelihood; however, we will show through extensive simulation studies (Section 3) that this estimator has minimal bias.

$$
\begin{aligned}
L(\mathbf{h}) &= \prod_{j=1}^{k}\left[\left\{\prod_{l\in A_j} f_T(t_l)\prod_{l\in B_j} P\left(u_l^p < T \le t_l^d\right)\right\} S_T(t_j)^{m_j}\right] \\
&= \prod_{j=1}^{k}\left[\prod_{l=1}^{d_j}\left\{S_T\left(t^{\star}_{(j,l)}\right) - S_T(t_j)\right\}\left\{\prod_{q=1}^{j}(1-h_q)\right\}^{m_j}\right] \\
&= \prod_{j=1}^{k}\left[\prod_{l=1}^{d_j}\left\{S_T\left(t_{r(j,l)}\right) - S_T(t_j)\right\}\left\{\prod_{q=1}^{j}(1-h_q)\right\}^{m_j}\right] \\
&= \prod_{j=1}^{k}\left[\prod_{l=1}^{d_j}\left\{\prod_{q=1}^{r(j,l)}(1-h_q) - \prod_{q=1}^{j}(1-h_q)\right\}\left\{\prod_{q=1}^{j}(1-h_q)\right\}^{m_j}\right] \\
&= \prod_{j=1}^{k}\left[\prod_{l=1}^{d_j}\left\{\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)} - 1\right]\prod_{q=1}^{j}(1-h_q)\right\}\left\{\prod_{q=1}^{j}(1-h_q)\right\}^{m_j}\right] \\
&= \prod_{j=1}^{k}\left[\prod_{l=1}^{d_j}\left\{\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)} - 1\right](1-h_j)\right\}(1-h_j)^{n_j-d_j}\right] \\
&= \prod_{j=1}^{k}\left\{\prod_{l=1}^{d_j}\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)} - 1\right](1-h_j)^{n_j}\right\}, \qquad (4)
\end{aligned}
$$

where

$$
t^{*}_{(j,l)} = \begin{cases} t_{j-1} & \text{if progression} - \text{PFS event} \\ \text{progression censoring time corresponding to death time, } t_j, \text{ for individual } \ l & \text{if death} - \text{PFS event} \end{cases}
$$

and

$$
r(j,l) = \begin{cases} j-1 & \text{if progression} - \text{PFS event} \\ \text{number of } t_g\text{'s} \le t^{\star}_{(j,l)} \quad \text{for } g = 1,\ldots,k & \text{if death} - \text{PFS event} \end{cases}.
$$

It is worthwhile to note that when all the observed PFS events are progression–PFS events (i.e., $r(j,l) = j-1$ for all $j = 1, \ldots, k$), the likelihood in (4) reduces to the standard KM likelihood:

$$
\begin{aligned}
L(\mathbf{h}) &= \prod_{j=1}^{k}\left\{\prod_{l=1}^{d_j}\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)} - 1\right](1-h_j)^{n_j}\right\} \\
&= \prod_{j=1}^{k}\left[\frac{1}{1-h_j} - 1\right]^{d_j}(1-h_j)^{n_j} \\
&= \prod_{j=1}^{k}\left[\frac{h_j}{1-h_j}\right]^{d_j}(1-h_j)^{n_j} \\
&= \prod_{j=1}^{k} h_j^{d_j}(1-h_j)^{n_j-d_j}.
\end{aligned}
$$

We call this estimator the "generalized"-KM.

The corresponding log-likelihood to (4) is

$$l(\mathbf{h}) = \sum_{j=1}^{k} \left\{ \sum_{l=1}^{d_j} \log\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)} - 1 \right] + n_j \log(1-h_j) \right\}. \tag{5}$$

Closed-form solutions of $\mathbf{h}$ remain to be discovered, thus, a quasi-Newton method is used to maximize (5) and obtain estimates of $\mathbf{h}$. The resulting nonparametric generalized-KM (g-KM) estimator of PFS is

$$\widehat{S}_T(t_j) = \prod_{q=1}^{j}\left(1-\widehat{h}_q\right).$$

The variance-covariance matrix for $\mathbf{h}$, denoted $\mathbf{\Sigma}$, can be approximated by the inverse of the observed information matrix, $\frac{-\partial l(\mathbf{h})}{\partial \mathbf{h}}$; details are given in Appendix. Let $\mathbf{h}_j = (h_1, \ldots, h_j)$ and $\mathbf{\Sigma}_j$ denote the top $(j \times j)$ portion of $\mathbf{\Sigma}$, for $j = 1, \ldots, k$. The variance of $\widehat{S}_r(t_j)$ can then be obtained using the following formula, similar to Greenwood [3]:

$$\mathrm{Var}\left[\widehat{S}_T(t_j)\right] = \widehat{S}_T(t_j)^2 (1-\mathbf{h}_j)^{-1} \sum_j \left[(1-\mathbf{h}_j)^{-1} \right]',$$

where a prime, $'$, denotes the transpose.

2.4 Nonparametric empirical estimator

Let $v_1 < \ldots < v_k$ denote the k distinct observed failure times (i.e., times to progression or death). Also, let $d_j = \sum_{i=1}^{n} 1(X_i^d = v_j, \delta_i^d = 1)$ and $s_j = \sum_{i=1}^{n} 1(X_i^d \geq v_j)$ denote, respectively, the number of deaths at v_j and the number of patients at risk for death at v_j, $j = 1, \ldots, k$. Furthermore, we let $e_{mj} = \sum_{i=1}^{n} 1(X_i^p = v_m, \delta_i^p = 1, X_i^d > v_j)$ denote the number of patients who progressed at v_m and were alive at v_j, and $r_{mj} = \sum_{i=1}^{n} 1(X_i^p \geq v_m, X_i^d > v_j)$ denote the number of patients at risk for progression at v_m and alive at v_j, where $m \leq j$.

It is then straightforward to show that, for $t \in (v_j, v_{j+1}]$,

$$\begin{aligned} S_T(t) &= P(T > t) \\ &= P\left(T^p > t, T^d > t\right) \\ &= P\left(T^p > v_j, T^d > v_j\right) \\ &= P\left(T^p > v_j \middle| T^d > v_j\right) P\left(T^d > v_j\right) \\ &= S^p(t) S^d(t), \end{aligned} \tag{6}$$

where

$$\begin{aligned} S^p(t) &= P\left(T^p > v_j \middle| T^d > v_j\right) \\ &= P\left(T^p > v_j \middle| T^p > v_{j-1}, T^d > v_j\right) P\left(T^p > v_{j-1} \middle| T^p > v_{j-2}, T^d > v_j\right) \ldots \\ &\quad P\left(T^p > v_2 | T^p > v_1, T^d > v_j\right) P\left(T^p > v_1 | T^d > v_j\right). \end{aligned}$$

It follows that

$$\widehat{S}^p(t) = \prod_{m=1}^{j}\left(1 - \frac{e_{mj}}{r_{mj}}\right).$$

The second term in (6), $S^d(t) = P(T^d > v_j)$, can be estimated using the standard KM estimator, $\widehat{S}^d(t) = \prod_{m=1}^{j}\left(1 - \frac{d_m}{s_m}\right)$. Thus, the final nonparametric empirical PFS estimator is

$$\widehat{S}_T(t) = \prod_{m=1}^{j}\left(1 - \frac{e_{mj}}{r_{mj}}\right)\left(1 - \frac{d_m}{s_m}\right). \tag{7}$$

In (7), d_m is equal to zero if there is no death event at v_m. Thus, if all the observed PFS events are progression–PFS events and no death–PFS events, the empirical estimator becomes the standard KM estimator, similar to the g-KM approach. The above empirical estimator uses the survival data of those patients with censored

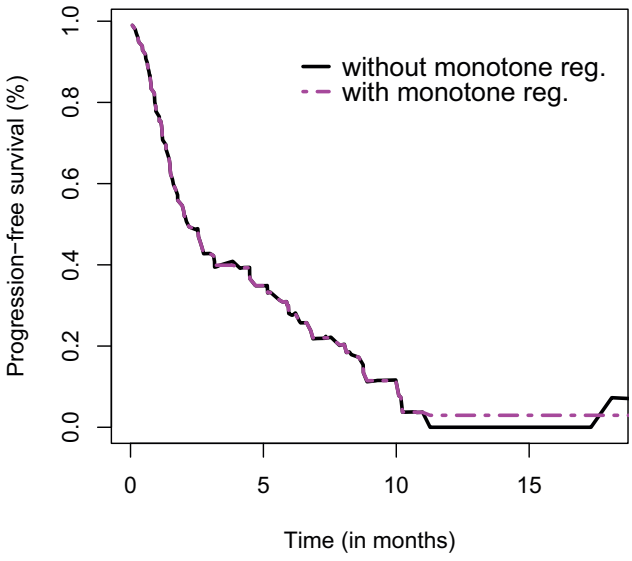

Figure 1: Example of estimated PFS curves using empirical estimator with/without monotone regression for a single sample (n = 100).

progression to obtain a more precise estimate of their PFS times. Specifically, this empirical estimator assumes that the progression times for those patients with death–PFS events are similar to the progression times of the subset of patients with similar survival and observed progression. Thus, it also imposes the standard KM assumption of independent censoring. If there is a smaller number of death events, the proposed empirical estimates will be close to the standard KM estimates. An analytical form for the variance of (7) still remains to be found, thus, we consider the bootstrap method [4] for variance estimation.

Given the small number of patients in the risk set at later times, the PFS function estimated using the empirical estimator in (7) can become unstable in the tail and is generally not monotonically decreasing. Thus, we employ monotone regression using the pool adjacent violators (PAV) algorithm to restrict the estimator to be monotonically decreasing [5]. The general idea behind monotone regression is to find a weighted least-squares fit $\mathbf{x}$ to a vector $\mathbf{y}$ with weights $\mathbf{w}$ subject to monotonicity constraints $x_1 \geq x_2 \geq \ldots \geq x_k$, i.e.,

$$\min \sum_{j=1}^{k} w_j \left(y_j - x_j\right)^2 \quad \text{subject to } x_1 \geq x_2 \geq \ldots x_k.$$

The PAV algorithm works as follows: if $y_{j-1} \leq y_j$, then $w^*_{j-1} = w_{j-1} + w_j$ and

$$x_{j-1} = y_{j-1} + \left(\frac{w_j}{w^*_{j-1}}\right)\left(y_j - y_{j-1}\right),$$

for $j = 1, \ldots, k$.

For our purposes, it is sufficient to let $\mathbf{w} = 1$. Figure 1 gives an example of the empirical estimator with and without monotone regression for a single simulated sample of size 100. As can be seen, the non-monotonicity is generally very mild, especially for earlier times and reasonably large sample sizes. Note that when we refer to the empirical estimator from now on, we are indicating the empirical estimator after applying monotone regression.

3 Simulation studies

In this section, we investigate the performance of the two PFS estimators presented in Section 2 with respect to bias and efficiency through an extensive simulation study by varying (a) the proportion of progression censored patients; (b) the relative difference between the true progression and death times; and (c) the correlation between progression and survival times. In a factorial simulation study, where factors (a) and (b)

each have two levels and factor (c) has four levels, we investigate a total of 16 different scenarios for each PFS estimator. The specific factor levels are chosen to reflect situations that are relevant in practice.

We vary factor (a) by generating data with 20 and 30% censoring with respect to progression. For factor (b), we vary the ratio of the mean time to progression ($\overline{T^p}$) and mean time to death after progression ($\overline{T^{d|p}}$), where the mean death time is then defined as $\overline{T^d} = \overline{T^p} + \overline{T^{d|p}}$ (note that $T^{d|p}$ is not a conditional distribution). Specifically, we consider ($\overline{T^p}$:$\overline{T^{d|p}}$) ratios of (1:1) and (1:2). Lastly, we allow correlations between T^p and T^d of 0.8, 0.7, 0.6, and 0.5. The proportion of patients censored with respect to death is fixed for all scenarios at 20%. The (censored progression %, censored death %) pairs (20%, 20%) and (30%, 20%) correspond to death–PFS proportions of around 10% and 20%, respectively.

Technical details of data generation are as follows. We first generate two independent variables, X_1 and X_2, from standard exponential distributions. Let $Y_1 = X_1$ and $Y_2 = X_1 + \omega X_2$, where

$$\omega = \sqrt{\frac{1 - r^2}{r^2}}$$

and r denotes the correlation between T^p and T^d. To generate data in accordance with the two levels of factor (b), we consider ($\overline{T^p}$, $\overline{T^d}$) pairs of (4, 8) and (4, 12) months, which correspond respectively to ($\overline{T^p}$:$\overline{T^{d|p}}$) ratios of (1:1) and (1:2). It follows that, $T^p = \overline{T^p} \times Y_1$ and $T^d = (\overline{T^d} \times Y_2)/(1 + \omega)$. Finally, both progression and death censoring times are generated from uniform distributions with appropriate parameters such that the respective censoring percentages of {20%, 30%} and 20% are expected. We generate 300 replicates for all scenarios and estimators.

For comparison purposes, we also include a KM estimator that uses an alternative definition of PFS, denoted KM*, where later death times are censored. Specifically, all patients with deaths beyond 3 months of their censored progression times have their PFS times censored at their progression censoring times. We use the fact that most advanced breast cancer trials have tumor assessments every 3 months. Although generally accepted, such cutoffs for censoring are rather arbitrarily determined with no statistical theory for support. Also, when a censoring cutoff of this kind is applied, a substantial proportion of the survival information is often ignored and the endpoint becomes more similar to a time to progression (TTP) rather than the preferred PFS endpoint. TTP analyses are usually only acceptable in situations where we can assume that the majority of deaths are unrelated to the disease, which is generally an unreasonable assumption in the advanced cancer setting (https://www.fda.gov/media/71195/download). In contrast to the KM* estimator, our proposed estimators attempt to statistically incorporate more survival information into the estimation of the PFS function. However, since the KM* estimator is often used in practice, we include it for completeness of comparison.

Tables 1 and 2 give the relative bias and standard errors (SE) of estimated PFS rates at 1, 3, 6, 9, and 12 months averaged across all replicates for 20% and 30% progression censoring, respectively. Relative bias is defined as (estimated PFS rates – true PFS rates)/true PFS rates. Figures 2–5 plot the average estimated PFS curves averaged across all replicates for respective progression censoring percentages 20 and 30% corresponding to Tables 1 and 2. Table 3 tabulates the absolute area between the average estimated and true PFS curves in Figures 2–5 standardized over time, and Table 4 calculates the percentage improvement of the proposed estimators and the KM* estimator over the KM estimator with respect to area between the average estimated and true PFS curves given in Table 3.

From all tables and figures, we can see some general trends in the bias of the KM estimator; the bias increases with increasing proportion of progression censored patients, increasing relative difference between the true progression and death times (or decreasing ($\overline{T^p}$:$\overline{T^{d|p}}$) ratio), and increasing correlation between T^p and T^d, where each trend is such that all other factors are held constant. The first two trends are quite intuitive and expected. Given our simulation data generation scheme, a patient with longer T^p is more likely to be censored with respect to progression. If, in addition, T^p and T^d are highly correlated, then that patient is also more likely to have longer T^d (compared to a patient from a population where T^p and T^d are less correlated), which implies that they are more likely to have longer post-progression survival, $T^{d|p}$. Since the KM definition of PFS replaces censored progression times with death times, the bias of the KM estimator is expected to increase with correlation (holding constant all other factors).

Table 1: Relative bias (SE) of estimated PFS rates for 20% progression censoring (averaged across 300 replicates).

Corr	Estimator	1 month	3 months	6 months	9 months	12 months	Median PFS
	True	0.779	0.472	0.223	0.105	0.050	2.773
Mean (T^p,T^d) ratio of (1:1)							
0.8	KM	0.003 (0.041)	0.030 (0.051)	0.112 (0.045)	0.257 (0.036)	0.480 (0.028)	0.045
	KM*	0.000 (0.042)	−0.004 (0.052)	−0.027 (0.046)	−0.048 (0.035)	0.160 (0.030)	0.001
	Empirical	−0.009 (0.042)	0.004 (0.052)	−0.009 (0.047)	−0.029 (0.037)	0.000 (0.027)	0.002
	g-KM	−0.003 (0.043)	−0.021 (0.065)	−0.081 (0.065)	−0.143 (0.049)	−0.100 (0.037)	−0.021
0.7	KM	0.003 (0.041)	0.028 (0.051)	0.103 (0.045)	0.229 (0.036)	0.360 (0.028)	0.043
	KM*	0.000 (0.042)	−0.004 (0.052)	−0.031 (0.045)	−0.057 (0.035)	0.120 (0.029)	−0.001
	Empirical	−0.009 (0.042)	0.004 (0.052)	−0.009 (0.047)	−0.029 (0.037)	−0.020 (0.026)	0.002
	g-KM	−0.003 (0.043)	−0.023 (0.066)	−0.090 (0.066)	−0.171 (0.049)	−0.140 (0.037)	−0.024
0.6	KM	0.003 (0.041)	−0.002 (0.051)	−0.022 (0.043)	−0.019 (0.033)	0.020 (0.024)	−0.005
	KM*	−0.001 (0.042)	−0.006 (0.051)	−0.031 (0.043)	−0.067 (0.032)	0.000 (0.024)	−0.008
	Empirical	−0.009 (0.042)	−0.004 (0.051)	−0.036 (0.044)	−0.057 (0.034)	−0.100 (0.023)	−0.007
	g-KM	0.001 (0.042)	−0.006 (0.053)	−0.036 (0.045)	−0.038 (0.034)	0.000 (0.025)	−0.011
0.5	KM	0.001 (0.041)	−0.017 (0.051)	−0.058 (0.042)	−0.086 (0.031)	−0.140 (0.022)	−0.022
	KM*	−0.004 (0.042)	−0.019 (0.051)	−0.076 (0.042)	−0.152 (0.031)	−0.160 (0.022)	−0.025
	Empirical	−0.010 (0.042)	−0.017 (0.051)	−0.081 (0.043)	−0.133 (0.032)	−0.220 (0.021)	−0.019
	g-KM	0.000 (0.042)	−0.021 (0.053)	−0.072 (0.044)	−0.114 (0.033)	−0.160 (0.023)	−0.028
Mean (T^p,T^d) ratio of (1:2)							
0.8	KM	0.003 (0.041)	0.044 (0.051)	0.188 (0.046)	0.552 (0.039)	1.220 (0.034)	0.069
	KM*	0.000 (0.042)	0.000 (0.052)	−0.013 (0.046)	−0.010 (0.036)	0.240 (0.031)	0.010
	Empirical	−0.009 (0.042)	0.000 (0.052)	−0.018 (0.047)	−0.038 (0.037)	0.000 (0.027)	−0.001
	g-KM	−0.003 (0.042)	−0.011 (0.057)	−0.049 (0.061)	−0.105 (0.051)	0.000 (0.045)	−0.014
0.7	KM	0.003 (0.041)	0.044 (0.051)	0.184 (0.046)	0.533 (0.039)	1.160 (0.033)	0.065
	KM*	0.000 (0.042)	0.002 (0.052)	−0.018 (0.046)	−0.019 (0.036)	0.240 (0.031)	0.010
	Empirical	−0.009 (0.042)	0.000 (0.052)	−0.013 (0.047)	−0.048 (0.037)	0.000 (0.027)	−0.002
	g-KM	−0.003 (0.042)	−0.008 (0.058)	−0.049 (0.061)	−0.105 (0.049)	0.000 (0.04)	−0.009
0.6	KM	0.003 (0.041)	0.042 (0.051)	0.184 (0.046)	0.495 (0.038)	1.080 (0.032)	0.063
	KM*	0.000 (0.042)	0.000 (0.052)	−0.018 (0.046)	−0.019 (0.036)	0.220 (0.030)	0.010
	Empirical	−0.009 (0.042)	0.000 (0.052)	−0.013 (0.047)	−0.048 (0.037)	0.000 (0.027)	−0.002
	g-KM	−0.003 (0.042)	−0.013 (0.059)	−0.067 (0.060)	−0.143 (0.049)	−0.080 (0.039)	−0.013
0.5	KM	0.003 (0.041)	0.042 (0.051)	0.179 (0.046)	0.467 (0.038)	1.000 (0.032)	0.063
	KM*	0.000 (0.042)	0.000 (0.052)	−0.022 (0.046)	−0.038 (0.036)	0.180 (0.030)	0.009
	Empirical	−0.009 (0.042)	0.000 (0.052)	−0.013 (0.047)	−0.048 (0.037)	−0.020 (0.027)	0.000
	g-KM	−0.003 (0.042)	−0.017 (0.059)	−0.072 (0.061)	−0.124 (0.049)	−0.080 (0.038)	−0.015

Relative bias = (estimated PFS rates − true PFS rates)/true PFS rates. SE denotes the estimated standard error; corr denotes (T^p, T^d) correlation; KM* is the KM estimator with alternative PFS definition; g-KM is the generalized-KM; all scenarios have 20% death censoring.

From Figures 2–5 and Table 3, we see that our proposed estimators and the KM* estimator are all very close to the true PFS function for all scenarios, while the KM estimator is generally biased upwards following the trends described above. Given that there are very few censored events early on, all estimated PFS curves are very similar for earlier times. Table 4 shows that the proposed empirical estimator improves the most upon the KM estimator in terms of decreasing the area between the estimated and true PFS curves; the average percentage of improvement across all scenarios is about 60%. The g-KM estimator comes in second, and the KM* estimator follows closely behind. All the estimators seem to be quite robust to varying (T^p, T^d) correlations, except for the KM estimator for ratio (1:1).

The result that the KM* estimator is very close to the true PFS function is expected given the nature of our simulated true samples, where there is no censoring and most PFS events are progressions, thus, there is little involvement of survival information in the true PFS function. Moreover, the fact that our proposed estimators perform comparably to the KM* estimator is a reassuring result; it supports that the proposed estimators can be used as alternative sensitivity analysis tools with a solid statistical basis, in contrast to the rather arbitrarily defined KM* estimator.

Table 2: Relative bias (SE) of estimated PFS rates for 30% progression censoring (averaged across 300 replicates).

Corr	Estimator	1 month	3 months	6 months	9 months	12 months	Median PFS
	True	0.779	0.472	0.223	0.105	0.050	2.773
Mean (T^p,T^d) ratio of (1:1)							
0.8	KM	0.006 (0.041)	0.057 (0.051)	0.224 (0.047)	0.495 (0.040)	0.780 (0.032)	0.086
	KM*	0.001 (0.042)	−0.002 (0.053)	−0.040 (0.048)	−0.019 (0.040)	0.300 (0.036)	0.003
	Empirical	−0.010 (0.042)	0.002 (0.054)	0.000 (0.050)	−0.019 (0.042)	0.020 (0.032)	−0.003
	g-KM	−0.006 (0.045)	−0.034 (0.070)	−0.139 (0.077)	−0.238 (0.058)	−0.180 (0.046)	−0.043
0.7	KM	0.006 (0.041)	0.055 (0.051)	0.206 (0.047)	0.438 (0.039)	0.640 (0.030)	0.081
	KM*	0.001 (0.042)	−0.006 (0.053)	−0.049 (0.048)	−0.067 (0.039)	0.220 (0.034)	−0.001
	Empirical	−0.010 (0.042)	0.002 (0.054)	−0.004 (0.049)	−0.019 (0.041)	0.000 (0.031)	−0.002
	g-KM	−0.009 (0.045)	−0.040 (0.072)	−0.161 (0.074)	−0.314 (0.055)	−0.340 (0.039)	−0.038
0.6	KM	0.003 (0.041)	0.028 (0.051)	0.085 (0.045)	0.200 (0.036)	0.340 (0.027)	0.037
	KM*	0.000 (0.042)	−0.011 (0.052)	−0.067 (0.045)	−0.114 (0.035)	0.020 (0.028)	−0.011
	Empirical	−0.010 (0.042)	−0.002 (0.053)	−0.036 (0.047)	−0.067 (0.037)	−0.060 (0.026)	−0.003
	g-KM	−0.005 (0.043)	−0.032 (0.068)	−0.143 (0.067)	−0.248 (0.049)	−0.240 (0.036)	−0.036
0.5	KM	0.003 (0.041)	0.002 (0.051)	0.000 (0.044)	0.010 (0.033)	0.020 (0.024)	0.004
	KM*	−0.003 (0.042)	−0.021 (0.052)	−0.094 (0.043)	−0.190 (0.032)	−0.180 (0.024)	−0.024
	Empirical	−0.012 (0.042)	−0.015 (0.052)	−0.081 (0.045)	−0.152 (0.034)	−0.200 (0.023)	−0.016
	g-KM	−0.001 (0.042)	−0.034 (0.063)	−0.126 (0.06)	−0.248 (0.042)	−0.320 (0.027)	−0.039
Mean (T^p,T^d) ratio of (1:2)							
0.8	KM	0.006 (0.041)	0.076 (0.051)	0.341 (0.048)	0.933 (0.043)	1.980 (0.039)	0.107
	KM*	0.000 (0.042)	0.004 (0.053)	−0.013 (0.049)	0.029 (0.042)	0.400 (0.037)	0.013
	Empirical	−0.009 (0.042)	0.000 (0.054)	−0.004 (0.050)	−0.010 (0.042)	0.100 (0.036)	−0.006
	g-KM	−0.001 (0.043)	−0.006 (0.061)	−0.072 (0.070)	−0.152 (0.064)	−0.060 (0.068)	−0.006
0.7	KM	0.008 (0.041)	0.076 (0.051)	0.350 (0.048)	0.943 (0.043)	1.980 (0.038)	0.113
	KM*	0.000 (0.042)	0.004 (0.053)	−0.013 (0.049)	0.057 (0.042)	0.500 (0.039)	0.013
	Empirical	−0.009 (0.042)	0.000 (0.054)	−0.004 (0.050)	−0.029 (0.042)	0.080 (0.035)	−0.004
	g-KM	−0.005 (0.043)	−0.013 (0.062)	−0.090 (0.07)	−0.171 (0.062)	−0.060 (0.056)	−0.021
0.6	KM	0.006 (0.041)	0.072 (0.051)	0.327 (0.048)	0.848 (0.042)	1.780 (0.037)	0.104
	KM*	0.001 (0.042)	0.004 (0.053)	−0.018 (0.049)	0.010 (0.041)	0.360 (0.036)	0.012
	Empirical	−0.009 (0.042)	0.000 (0.053)	−0.004 (0.050)	−0.019 (0.042)	0.080 (0.034)	−0.006
	g-KM	0.001 (0.042)	−0.008 (0.062)	−0.076 (0.071)	−0.200 (0.058)	−0.120 (0.050)	−0.017
0.5	KM	0.006 (0.041)	0.070 (0.051)	0.314 (0.047)	0.800 (0.041)	1.620 (0.036)	0.103
	KM*	0.001 (0.042)	0.002 (0.053)	−0.027 (0.048)	−0.010 (0.040)	0.320 (0.035)	0.012
	Empirical	−0.009 (0.042)	0.000 (0.053)	−0.004 (0.050)	−0.029 (0.041)	0.080 (0.034)	−0.004
	g-KM	−0.005 (0.043)	−0.025 (0.065)	−0.126 (0.070)	−0.229 (0.057)	−0.160 (0.051)	−0.022

Relative bias = (estimated PFS rates − true PFS rates)/true PFS rates. SE denotes the estimated standard error; corr denotes (T^p, T^d) correlation; KM* is the KM estimator with alternative PFS definition; g-KM is the generalized-KM; all scenarios have 20% death censoring.

With respect to efficiency, we see that the g-KM estimator is the least efficient while the other estimators have better, similar efficiency. Generally, lower efficiency is considered an undesirable property; however, we argue that at least part of the loss in efficiency of the g-KM estimator is actually a better reflection of the true variability. To see this, let's focus on those patients with death–PFS events. The KM and KM* estimators assume, respectively, that T^d and T^d within 3 months of U^p are the definitive PFS times, while the g-KM estimator assumes, more accurately, that $U^p < T \le T^d$, which induces some natural variability. Thus, part of the loss in efficiency is actually a more accurate reflection of the true variability, while the other part is admittedly due to the use of the partial likelihood instead of the full likelihood. However, the simulations show that the g-KM estimator performs quite well with respect to bias and could still be a useful alternative.

Given the assumptions for the empirical estimator, we should also expect some loss in efficiency due to conditioning on the subset of patients with similar death times and observed progression times, which results in very small risk sets at later times. However, we see that the efficiency of the empirical estimator is actually

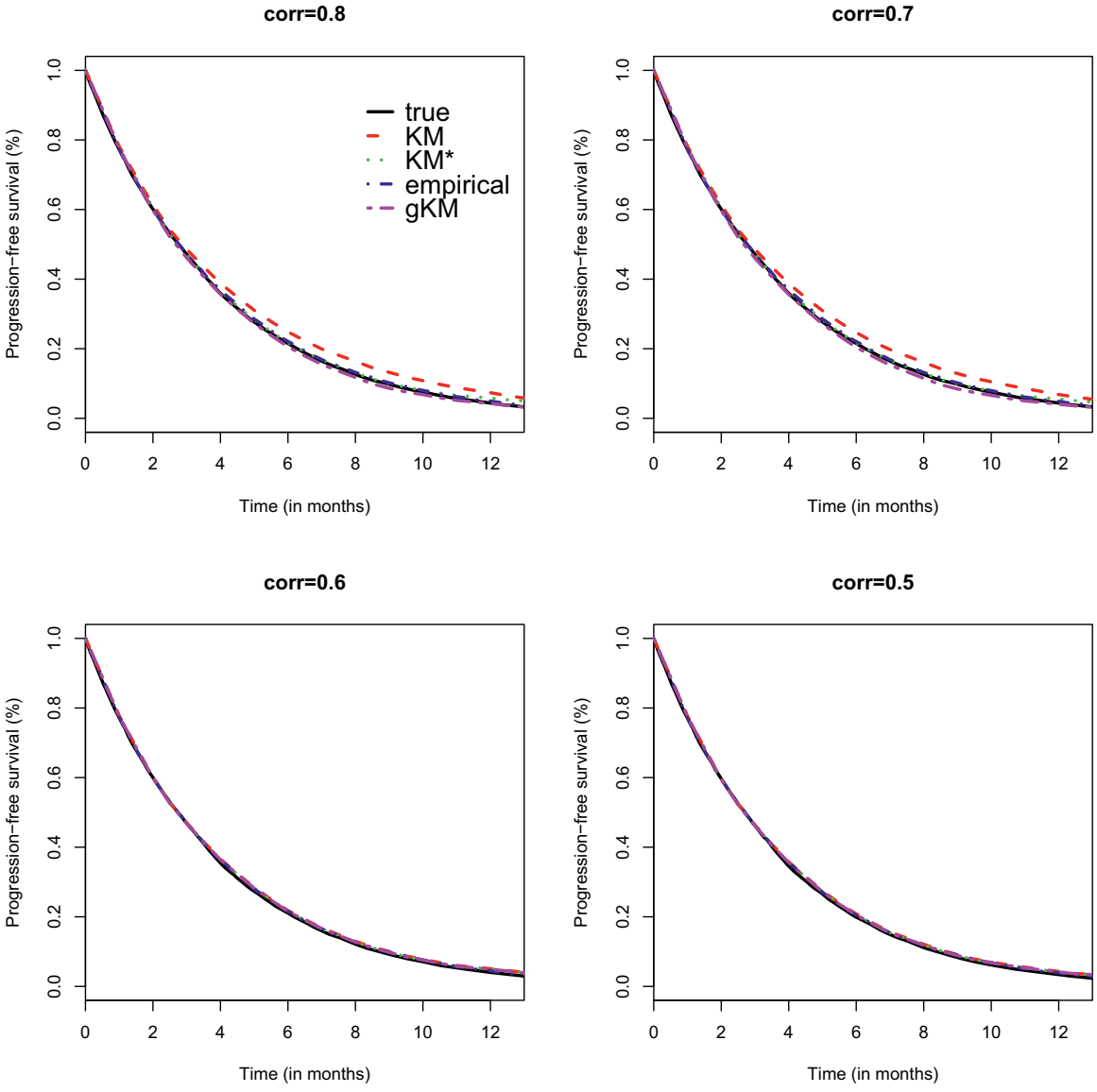

Figure 2: Average estimated PFS curves for $(\overline{T^p} : \overline{T^d}) = (1:1)$ and 20% progression censoring (300 replicates).

comparable to the KM and KM* estimators. Brief simulations (not shown) confirm that this is due to the use of monotone regression, which arbitrarily reduces the variability.

4 A data example

In the previous sections, we presented the details for the two proposed PFS estimators and evaluated their performance against the KM estimator through simulation studies. Now, we use an ECOG-ACRIN metastatic breast cancer study (E2100) to illustrate the application of the PFS estimators to real clinical data. E2100 was an open-label, randomized, phase III trial whose primary objective was to compare the efficacy of paclitaxel with the combination paclitaxel plus bevacizumab as first-line treatment for metastatic breast cancer [6]. The primary endpoint was PFS with overall survival as a secondary endpoint.

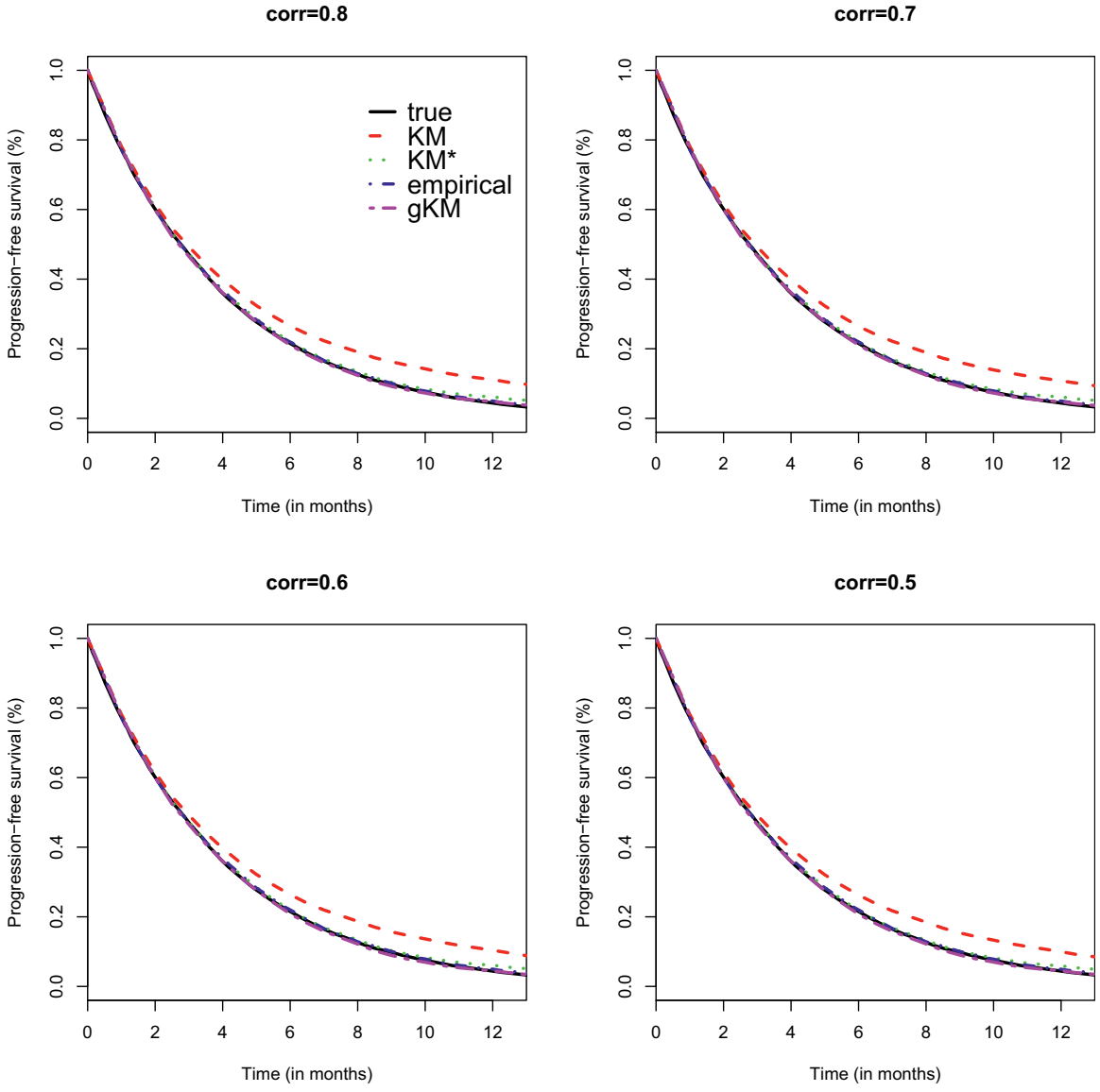

Figure 3: Average estimated PFS curves for $(\overline{T^p} : \overline{T^d}) = (1:2)$ and 20% progression censoring (300 replicates).

Table 5 summarizes the case status for the E2100 final analysis, where we censor for non-protocol treatment. There are a respective total of 329 and 349 patients on the paclitaxel and combination arms eligible for analysis. About 20% of patients on both arms experience death–PFS events. The $(\overline{T^p}:\overline{T^{d|p}})$ ratios for the paclitaxel and combination arms are about (1:2) and (1:1), respectively, with corresponding (T^p, T^d) correlations around 0.5 and 0.6. Both arms have about 30% progression censoring. Given these factors and our simulation results, we should expect the KM estimator to be biased upward for the paclitaxel arm and the other estimators (i.e., KM*, empirical, and g-KM) to be very close to the true PFS function. For the combination arm, less of a difference between the estimators is expected; however, the KM estimator should still be a slight overestimate of the truth.

Table 6 tabulates the estimated PFS rates for 3, 6, 12, 18, and 24 months for both arms with estimated standard errors (SE) and median PFS times. Figure 6 plots the corresponding estimated PFS curves. Indeed, the results show the KM estimator PFS curve to be above the other estimators, particularly for the paclitaxel arm, while less

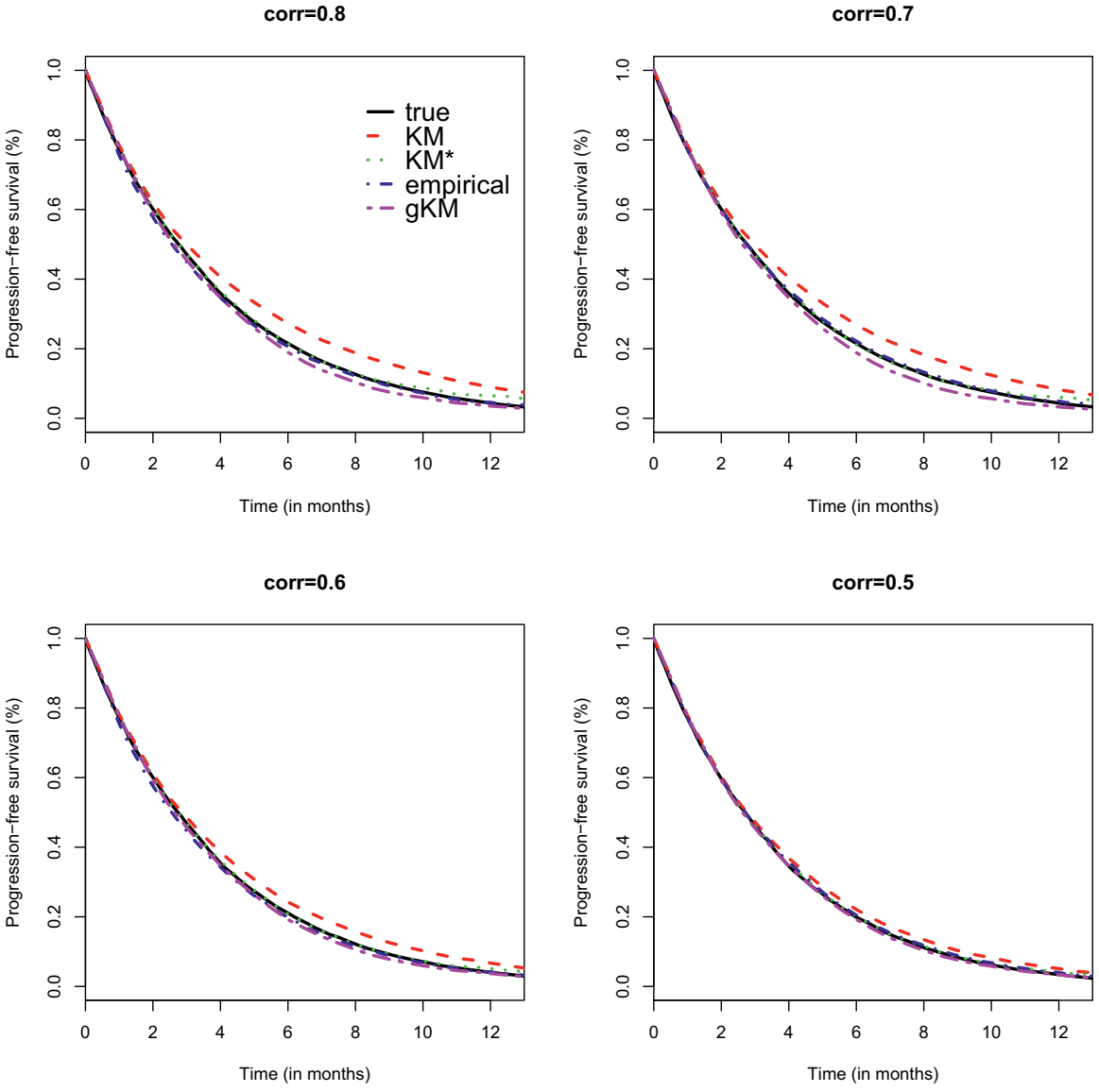

Figure 4: Average estimated PFS curves for $(\overline{T^p} : \overline{T^d}) = (1:1)$ and 30% progression censoring (300 replicates).

of a distinction can be made between the estimators for the combination therapy arm. Also as expected, the g-KM estimator is somewhat less efficient than the other estimators. These results imply that the survival probability for the PFS event in the paclitaxel arm could have been worse than that reported previously [6]. This does not change the final conclusion about the beneficial effect of the combination arm in this clinical paper.

5 Discussion

In this paper, we develop two alternative nonparametric PFS estimators that *statistically* incorporate the survival information that is often available for those patients with censored progression times. We use simulation studies to investigate the performance of the alternative estimators compared to the standard KM

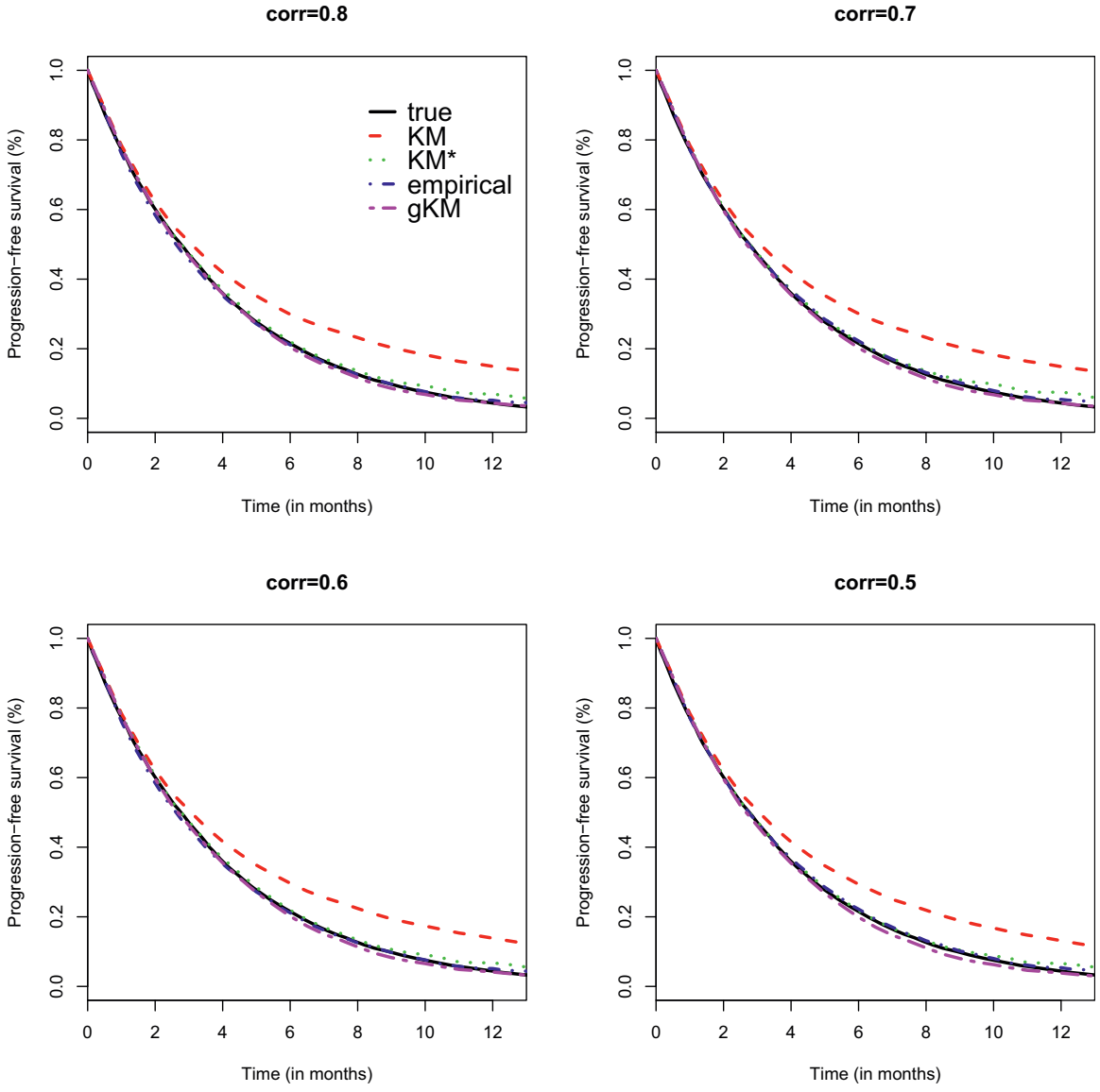

Figure 5: Average estimated PFS curves for $(\overline{T^p} : \overline{T^d}) = (1:2)$ and 30% progression censoring (300 replicates).

estimator. Our results show that when (a) the proportion of censored progression data is larger and survival data is fairly complete; (b) the correlation between time to progression and time to death is higher; and/or (c) the relative difference between the true progression and death times is larger, the KM estimator of PFS is generally positively biased. In such situations and others involving factors that were not considered here, current PFS analysis methods depend mostly on sensitivity analyses based on somewhat arbitrary and ad-hoc definitions of censoring to evaluate the robustness of the analyses. The estimators we develop here attempt to reduce the bias of the KM estimator through the implementation of a statistical correction for the set of patients with censored progression and observed death times. Both estimators are nonparametric and assume independent censoring, just as the standard KM estimator does. In the presence of death–PFS events, our proposed

Table 3: Absolute area between average estimated and true PFS curves standardized over time (300 replicates).

Ratio	Corr	20% progression censoring				30% progression censoring			
		KM	KM*	Empirical	g-KM	KM	KM*	Empirical	g-KM
(1:1)	0.8	0.045	0.024	0.015	0.021	0.062	0.025	0.016	0.025
	0.7	0.043	0.023	0.015	0.021	0.058	0.023	0.015	0.026
	0.6	0.026	0.025	0.015	0.024	0.044	0.022	0.017	0.022
	0.5	0.026	0.025	0.015	0.024	0.035	0.023	0.015	0.021
(1:2)	0.8	0.064	0.026	0.014	0.020	0.091	0.029	0.014	0.021
	0.7	0.062	0.026	0.014	0.021	0.092	0.030	0.016	0.021
	0.6	0.060	0.025	0.014	0.021	0.086	0.027	0.014	0.022
	0.5	0.059	0.025	0.014	0.020	0.082	0.026	0.015	0.023

Ratio denotes $(\overline{T^p} : \overline{T^d})$; corr denotes (T^p, T^d) correlation; KM* is the KM estimator with alternative PFS definition; g-KM is the generalized-KM; all scenarios have 20% death censoring.

Table 4: Percentage (%) improvement of proposed estimators over KM estimator w.r.t. area between average estimated and true PFS curves corresponding to Table 2.

Ratio	Corr	20% Progression censoring			30% Progression censoring		
		KM*	Empirical	g-KM	KM*	Empirical	g-KM
(1:1)	0.8	46.06	65.75	53.12	59.24	73.43	59.37
	0.7	45.44	64.26	49.71	59.89	73.72	54.67
	0.6	4.29	41.20	5.93	49.74	61.20	49.79
	0.5	4.39	41.58	6.91	34.94	56.50	41.33
(1:2)	0.8	59.00	77.26	67.92	68.42	84.73	76.73
	0.7	58.53	76.75	66.84	67.35	83.02	77.09
	0.6	57.98	76.11	65.90	68.16	83.87	74.53
	0.5	58.02	75.32	65.21	68.37	81.44	72.07

Ratio denotes $(\overline{T^p} : \overline{T^d})$; corr denotes (T^p, T^d) correlation; KM* is the KM estimator with alternative PFS definition; g-KM is the generalized-KM; all scenarios have 20% death censoring.

Table 5: Case status for E2100 analysis.

Arm	Eligible cases	PFS	Progression-PFS	Death-PFS	Progression-censored	Total deaths
Paclitaxel + bevacizumab	349	298	230	68	119	247
Paclitaxel	329	297	234	63	95	243
Total	678	595	464	131	214	490

methods can be generalized to estimate the baseline survival function and cumulative hazard in a Cox model. Estimation of baseline survival function or cumulative hazard is often necessary in risk prediction analyses.

As mentioned in Section 1, Ruan and Gray [1] also considered incorporating survival information for progression censored patients in their PFS sensitivity analysis method with dependent censoring. Although their focus was not on estimation of the PFS function, we modified and implemented their method for independent censoring and conducted a small simulation study to evaluate its performance compared to the KM and the two proposed estimators. The implementation is quite computationally intensive given the use of the EM algorithm and our simulation results (not shown) showed that our proposed estimators performed comparably in terms of bias and efficiency.

The generalized-KM estimator, which utilizes the fact that the true PFS time for a patient with a death–PFS event is somewhere between their progression censoring and death times, is monotonically decreasing and

Table 6: Estimated PFS rates (SE) and median PFS for E2100 analysis.

Estimator	3 months	6 months	12 months	18 months	24 months	Median PFS
Paclitaxel + bevacizumab						
KM	0.885 (0.017)	0.758 (0.023)	0.531 (0.027)	0.326 (0.026)	0.185 (0.022)	13.240
KM*	0.882 (0.018)	0.755 (0.024)	0.516 (0.029)	0.308 (0.028)	0.174 (0.024)	12.550
Empirical	0.882 (0.018)	0.749 (0.023)	0.512 (0.027)	0.301 (0.023)	0.167 (0.021)	12.518
g-KM	0.863 (0.031)	0.727 (0.050)	0.481 (0.059)	0.272 (0.058)	0.149 (0.059)	11.203
Paclitaxel						
KM	0.749 (0.024)	0.522 (0.028)	0.326 (0.027)	0.204 (0.024)	0.132 (0.020)	6.702
KM*	0.738 (0.025)	0.496 (0.029)	0.273 (0.028)	0.152 (0.025)	0.052 (0.019)	5.815
Empirical	0.737 (0.022)	0.497 (0.026)	0.275 (0.029)	0.148 (0.025)	0.055 (0.021)	5.914
g-KM	0.726 (0.042)	0.482 (0.059)	0.258 (0.049)	0.139 (0.062)	0.065 (0.057)	5.684

SE denotes the estimated standard error; KM* is the KM estimator with alternative PFS definition; g-KM is the generalized-KM.

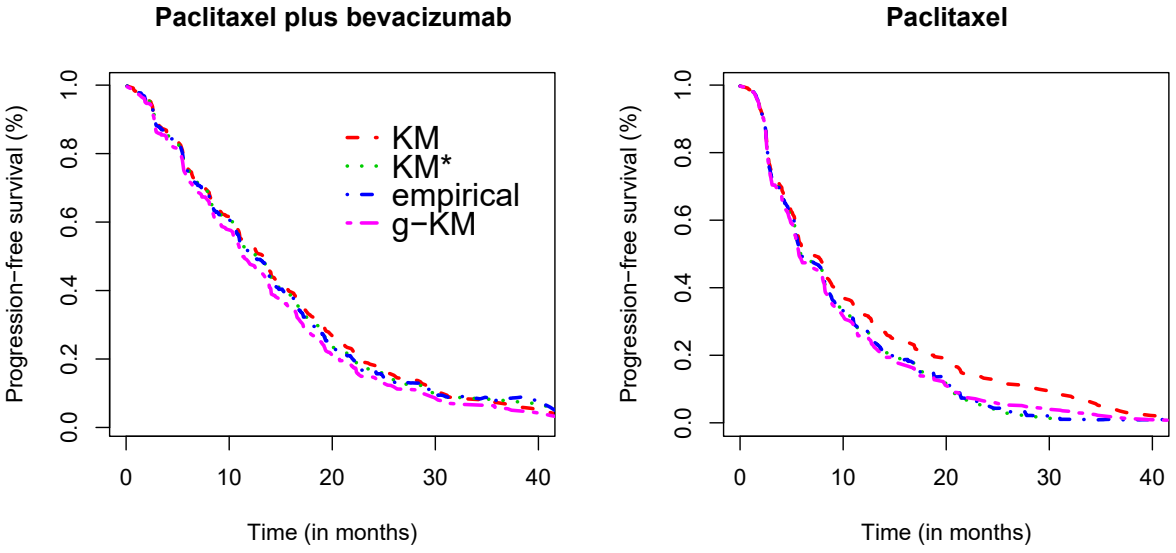

Figure 6: Estimated PFS curves for E2100 analysis.

has no obvious behavioral problems. As a result of using the partial likelihood instead of the full likelihood and the approximation in (2) though, the g-KM estimator is somewhat less efficient and may not have the desired asymptotic properties (e.g., consistency). However, our simulations show that this estimator still has minimal bias in practical situations, and it does so in a manner that is in line with the intention of the PFS endpoint (i.e., through statistical incorporation of available survival data to obtain an improved estimator). Compared to the KM* estimator, a major advantage of the g-KM estimator is that, in the death-PFS scenarios, it always uses death time as the upper limit of the PFS time. In contrast, in the KM* method, a patients censoring status in the death–PFS scenarios depends on whether death occurs before or after a pre-defined cutoff time, creating a potential problem of dependent censoring. Thus, we assert that the g-KM estimator can still be useful and may serve as an alternative tool for PFS sensitivity analyses. A possible extension would be to model $P(u_l^p < T \le t_l^d | T^d = t_l^d)$ directly (as mentioned in Section 2.3), using a semiparametric dependence model, and assess the trade-off between performance improvement and robustness to the semiparametric assumptions of the estimation method.

The empirical estimator assumes that the progression times for those patients with death–PFS events are similar to the progression times of the subset of patients with similar survival and observed progression. This estimator is generally not monotonically decreasing, thus, we apply monotone regression to restrict the estimated function to be monotone; as long as the sample size is reasonably large, the non-monotonicity is usually very mild. The efficiency of the estimator with monotone regression is comparable to the KM and KM* estimators. With respect to bias, our simulations show that the empirical estimator performs the best out of all the estimators evaluated for all scenarios considered. We recommend the empirical estimator as a bias-reducing alternative to the KM and KM* estimators for PFS analysis, particularly for data with a substantial proportion of death–PFS events. The empirical estimator may be used as an alternative sensitive analysis for PFS as well. A more in-depth study of the estimator, including a formal establishment of its asymptotic properties, would be a topic for future research.

In summary, we have proposed two novel PFS estimators in this paper. They are particularly useful in sensitivity analyses for data with a substantial proportion of death–PFS events. The nonparametric empirical estimator typically has less bias than all the other estimators considered in this paper. The g-KM estimator uses a likelihood approach where the likelihood is constructed based on the reality that there could be an unreported disease progression for a patient with a death–PFS event. Although the g-KM estimator typically has a larger variance than the other estimators, it performed reasonably well with respective to the areas between the estimated and the true PFS curves as shown in our simulation studies.

Acknowledgments: The authors thank the patients, physicians, nurses, and data managers who participated in the ECOG-ACRIN trial E2100. They further acknowledge support from the United States National Cancer Institute (CA-75362) and the United States National Institute of Health Cancer Training Grant (Jenny J. Zhang). They express their gratitude to Richard Gelber, Robert Gray, and Ann Partridge for their assistance throughout. They also thank the editor and referees for their insightful comments and suggestions.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix

Details for variance estimation of $\mathbf{h}$ in Section 2.3

Recall that $l(\mathbf{h}) = \sum_{j=1}^{k}\left\{\sum_{l=1}^{d_j}\log\left[\frac{1}{\prod_{q=r(j,l)+1}^{j}(1-h_q)}-1\right]+n_j\log(1-h_j)\right\}$, then

$$\begin{aligned}\frac{\partial l(\mathbf{h})}{\partial h_v} &= \sum_{j=v}^{k}\left[\sum_{l=1}^{d_j}1(r(j,l)+1\le v\le j)\right.\\ &\quad\left.\times\left\{\frac{\prod_{q=r(j,l)+1}^{j}(1-h_q)}{1-\prod_{q=r(j,l)+1}^{j}(1-h_q)}\frac{\prod_{q\in[r(j,l)+1,j]:q\neq v}(1-h_q)}{\left[\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]^2}\right\}\right]-\frac{n_v}{1-h_v}\\ &= \sum_{j=v}^{k}\left[\sum_{l=1}^{d_j}1(r(j,l)+1\le v\le j)\right.\\ &\quad\left.\times\left\{\frac{\prod_{q\in[r(j,l)+1,j]:q\neq v}(1-h_q)}{\prod_{q=r(j,l)+1}^{j}(1-h_q)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]}\right\}\right]-\frac{n_v}{1-h_v}\\ &= \sum_{j=v}^{k}\left[\sum_{l=1}^{d_j}1(r(j,l)+1\le v\le j)\left\{\frac{1}{(1-h_v)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]}\right\}\right]-\frac{n_v}{1-h_v}\end{aligned}$$

$$\frac{\partial^2 l(\mathbf{h})}{\partial h_\upsilon^2} = \sum_{j=\upsilon}^{k}\left\{\sum_{l-1}^{d_j} 1(r(j,\,l)+1 \le \upsilon \le j)\right.$$

$$\left.\times\left[\frac{-\left\{(1-h_\upsilon)\prod_{q\in[r(j,l)+1,j]:q\neq\upsilon}(1-h_q)-\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]\right\}}{\left\{(1-h_\upsilon)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]\right\}^2}\right]\right\} - \frac{n_\upsilon}{(1-h_\upsilon)}$$

$$= \sum_{j=\upsilon}^{k}\left\{\sum_{l-1}^{d_j} 1(r(j,\,l)+1 \le \upsilon \le j)\right.$$

$$\left.\times\left[\frac{1-\prod_{q=r(j,l)+1}^{j}(1-h_q)-\prod_{q=r(j,l)+1}^{j}(1-h_q)}{\left\{(1-h_\upsilon)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]\right\}^2}\right]\right\} - \frac{n_\upsilon}{(1-h_\upsilon)}$$

$$= \sum_{j=\upsilon}^{k}\left\{\sum_{l-1}^{d_j} 1(r(j,\,l)+1 \le \upsilon \le j)\right.$$

$$\left.\times\left[\frac{1-2\prod_{q=r(j,l)+1}^{j}(1-h_q)}{\left\{(1-h_\upsilon)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]\right\}^2}\right]\right\} - \frac{n_\upsilon}{(1-h_\upsilon)}$$

$$\frac{\partial^2 l(\mathbf{h})}{\partial h_\upsilon\,\partial\,h_p} = \sum_{j=\upsilon}^{k}\left\{\sum_{l-1}^{d_j} 1(r(j,\,l)+1 \le \upsilon,\, p \le j)\left[\frac{-(1-h_\upsilon)\prod_{q\in[r(j,l)+1,j]:q\neq p}(1-h_q)}{\left\{(1-h_\upsilon)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]\right\}^2}\right]\right\}$$

$$= \sum_{j=\upsilon}^{k}\left\{\sum_{l-1}^{d_j} 1(r(j,\,l)+1 \le \upsilon,\, p \le j)\left[\frac{-\prod_{q=r(j,l)+1}^{j}(1-h_q)}{(1-h_\upsilon)(1-h_p)\left[1-\prod_{q=r(j,l)+1}^{j}(1-h_q)\right]^2}\right]\right\}$$

References

1. Ruan PK, Gray RJ. Sensitivity analysis of progression-free survival with dependent withdrawal. Stat Med 2008; 27: 1180–98.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457–81.
3. Greenwood M. The natural duration of cancer. Reports on Public Health and Medical Subjects, 33. London: Her Majesty's Stationery Office; 1926: 1–26 pp.
4. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
5. Robertson T, Wright FT, Dykstra RL. Order restricted statistical inference. New York: Wiley; 1988.
6. Miller K, Wang M, Gralow J, Dickler M, Cobliegh M, Perez EA, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med 2007; 357: 2666–76.