stats

MDPI

Article

Asymptotic Relative Efficiency of Parametric and Nonparametric Survival Estimators

Szilárd Nemes

BioPharma Early Biometrics and Statistical Innovation, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, 43183 Gothenburg, Sweden; szilard.nemes@astrazeneca.com

Abstract: The dominance of non- and semi-parametric methods in survival analysis is not without criticism. Several studies have highlighted the decrease in efficiency compared to parametric methods. We revisit the problem of Asymptotic Relative Efficiency (ARE) of the Kaplan–Meier survival estimator compared to parametric survival estimators. We begin by generalizing Miller's approach and presenting a formula that enables the estimation (numerical or exact) of ARE for various survival distributions and types of censoring. We examine the effect of follow-up time and censoring on ARE. The article concludes with a discussion about the reasons behind the lower and time-dependent ARE of the Kaplan–Meier survival estimator.

Keywords: survival estimates; efficiency; parametric survival; censoring

Citation: Nemes, S. Asymptotic Relative Efficiency of Parametric and Nonparametric Survival Estimators. *Stats* **2023**, *6*, 1147–1159. https://doi.org/10.3390/stats6040072

Academic Editor: Wei Zhu

Received: 13 September 2023
Revised: 20 October 2023
Accepted: 21 October 2023
Published: 25 October 2023

1. Introduction

In medical statistics, survival analysis is often performed using the nonparametric Kaplan–Meier estimator [1] together with the semiparametric Proportional Hazards Model by Cox [2]. Although the theory of parametric estimators of survival is well-developed [3], their application has been limited in biomedical research. Miller [4] was one of the first to criticize this practice and highlight the loss of efficiency compared to parametric methods. Miller's work was continued by Klein and Moeschberger [5] and Aranda-Ordaz [6] and was partly rebutted by Meier and collaborators [7] and recently touched upon by Jullum and Hjort [8]. Cheng and Lin [9] proposed a new maximum likelihood estimator under the competing risks model with proportional hazards and assessed its efficacy against the Kaplan–Meier estimator.

In this paper, we revisit the problem of the Asymptotic Relative Efficiency (ARE) of the Kaplan–Meier estimator compared to parametric estimators of survival. We begin by extending Miller's method for calculating the Asymptotic Relative Efficiency (ARE) and present a formula that enables the estimation (numerical or exact) of ARE for various survival and censoring distributions, as well as different censoring types. We examine how censoring and the type of censoring affect ARE and discuss the reasons behind the lower and time-dependent ARE of the Kaplan–Meier estimator compared to parametric estimators of survival.

2. Notation and Estimators

We assume that the survival times for an event of interest, $X_1, \ldots, X_n$, are distributed independently and identically according to the distribution function $F(x)$ and the survival function $S(x) = 1 - F(x)$. We do not always have complete information for all subjects due to loss to follow-up or insufficient follow-up time. The times to censoring are denoted by $C_1, \ldots C_n$, and are assumed to be independent and identical, and survival function $\bar{G}(c) = 1 - G(c)$. Further censoring occurs when the study period ends at the predetermined time τ; all remaining subjects at risk are censored. Thus, the actual observed time for the subject j is $T_j = \min\left(X_j, C_j \wedge \tau\right)$. We assume independence between failure

and censoring time. We define $\delta_j = I\{X_j \leq C_j \wedge \tau\}$ as an event indicator, taking values 1 if an event is recorded prior to censoring, otherwise 0. We consider two competing estimators for the survival function $S(t)$, the nonparametric Kaplan–Meier estimator, and the parametric maximum likelihood estimator, noted hereafter with the subscripts $_{km}$ and $_{ml}$. Their respective variances will be noted as v_{km} and v_{ml}. In addition, we define $\mathcal{N}(t) = \sum_j^n I(T_j \leq t, \delta_j = 1)$, the process that counts the number of events in $(0, t]$ and $\mathcal{R}(t) = \sum_j^n I(T_j > t)$ the number at risk at time t. Although the index is not used to simplify notation through the whole of the text, we generally deal with sequences of estimators indexed by the sample size n.

2.1. The Kaplan–Meier Estimator

The naïve nonparametric survival function estimator $S(t) = n^{-1}\mathcal{N}(t)$ is only feasible when there is no censoring in the data, i.e., $P(X \leq C) = 1$. In this case, the survival at $\forall t$ is a binomial probability with variance $n^{-1}S(t)(1 - S(t))$.

The nonparametric product limit or Kaplan–Meier estimator [1] that can incorporate censored data is given by

$$S_{km}(t) = \prod_{T_i \leq t} \left[1 - \frac{\delta_i}{\mathcal{R}(T_i)}\right], \tag{1}$$

where T_i is the ordered sequence of observed event times and t is the absolute time. The Kaplan–Meier is asymptotically unbiased and has as a limiting distribution

$$\sqrt{n}\{\hat{S}_{km}(t) - S(t)\} \xrightarrow{D} N(0, v_{km}), \tag{2}$$

with asymptotic variance

$$v_{km}(t) = \frac{1}{n}\{1 - F(t)\}^2 \int_0^t \frac{f(u)}{\{1 - F(u)\}^2\{1 - G(u)\}}\, dt. \tag{3}$$

This requires knowledge of the survival and censoring distribution, thus is not applicable in real-life situations and its nonparametric variant, the Greenwood estimator, is used:

$$v_{km}(t) = S_{km}^2(t) \sum_{T_i \leq t} \frac{\delta_i}{\mathcal{R}(T_i)[\mathcal{R}(T_i) - \delta_i]}. \tag{4}$$

There are alternatives to the Greenwood variance estimator, but as Klein [10] concluded, the variance-bias trade-off favours the Greenwood estimator.

The Parametric Survival Estimator

Next, we establish the notation for the maximum likelihood estimators. The parametric survival function is given by $\hat{S}_{ml}(t, \hat{\theta})$, where $\hat{\theta} = \arg\max_\theta l_n(\theta)$ with $l_n(\theta) = \sum_{i=1}^n \log f(X_i, \theta)$ being the log-likelihood function of the data

$$l(\theta \mid T_i, \delta_i) = \prod_{i=1}^n \{f(T_i)\bar{G}(T_i)\}^{\delta_i}\{(1 - F(T_i))g(T_i)\}^{1-\delta_i}. \tag{5}$$

As the censoring distribution G does not depend on θ, censoring does not contribute to the estimation of θ_{ml} and it is factored out during maximization of the likelihood. On the other hand, the variance and Fisher information of θ_{ml} does account for censoring [11,12], with

$$J^T(\theta) = \int_0^\infty \left\{\frac{\partial}{\partial\theta}\log f\right\}^2 f\bar{G}\, dx = J^X(\theta)P(X < C), \tag{6}$$

with the superscript ^{T} indicating that calculation refers to the observed follow-up time, while the superscript ^{X} is the possibly unobserved survival times. Here, $P(X < C) = \int_0^\infty F(u)g(u)\,du$ [13]. In addition we define the score function $U(t,\theta)$ as the first derivative of $\log f(X_i,\theta)$ with regard to θ.

Under mild regularity conditions, the maximum likelihood estimator is consistent and converges in probability $\hat{\theta}_{ml} \xrightarrow{\mathcal{P}} \theta$ as the sample size $n \to \infty$ and is asymptotically normally distributed,

$$\sqrt{n}\{\hat{S}_{ml}(\hat{\theta};t) - S(\theta_0;t)\} \xrightarrow{D} N(0, v_{ml}), \tag{7}$$

with variance v_{ml}

$$v_{ml}(t) = \frac{1}{n}\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}^t J^{-1}\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}, \tag{8}$$

where J is the Fisher information matrix.

3. Asymptotic Relative Efficiency

For the statistic of interest $S(t)$ and for consistent estimators $S_{km}(t)$ and $S_{ml}(t)$, the Asymptotic Relative Efficiency is given by

$$ARE(S_{km}(t), S_{ml}(t), F) = \frac{v_{ml}(t)}{v_{km}(t)}. \tag{9}$$

From the Cramér–Rao theorem, we know that $ARE(S_{km}S_{ml}, F) \leq 1$, and ARE are the needed factor change in the sample size so that the competing estimators perform equally [14].

Theorem 1. *Under the assumption that $S_{km}(t) \to S(t)$ and $S_{ml}(t) \to S(t)$ as $n \to \infty$, Equation (9) with censoring considered can be defined as*

$$ARE(S_{km}(t), S_{ml}(t), F) =$$
$$\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}^t \left\{J^X(\theta)P(X < C)S(t)(1 - S(t))\varphi(t)\right\}^{-1}\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}.$$

Proof. The naïve nonparametric survival function estimator $\hat{S}(t) = n^{-1}\mathcal{N}(t)$ it is feasible only when there is no censoring in the data , i.e., $P(X \leq C) = 1$. In this case, the survival at $\forall t$ is a binomial probability with variance $n^{-1}S(t)(1 - S(t))$. As information is lost due to censoring of the binomial variance being inflated by factor $\varphi(t)$ [15], where

$$\varphi(t) = \frac{S(t)}{1 - S(t)}\int_0^t \frac{\alpha(u)}{S(u)\bar{G}(u)}du \tag{10}$$

and $n^{-1}\hat{S}(t)\left(1 - \hat{S}(t)\right)\varphi(t)$ is estimated by the Greenwood variance estimator. Censoring causes a loss of information with $J^T(\theta) = J^X(\theta)P(X < C)$ [11,12], where $P(X < C) = \int_0^\infty F(u)g(u)\,du$ [13]. The Delta-method variance of the maximum likelihood survival estimate, considering the proportion of censoring, is

$$\frac{1}{nJ^X(\theta)P(X < C)}\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}^2, \tag{11}$$

while the variance of the nonparametric Kaplan–Meier estimate with variance inflation due censoring is given by

$$\frac{1}{n} S_{km}(t)(1 - S_{km}(t))\varphi(t). \tag{12}$$

By plugging these two estimators into Equation for ARE, we conclude the proof. □

4. ARE as Correlation between Estimates

Crámer [16] showed that ARE equals the square of the correlation between the competing estimates, ρ^2, with

$$\hat{\rho}(S_{km}(t), S_{ml}(t)) = \frac{v_c(t)}{\sqrt{v_{km}(t)v_{ml}(t)}}, \tag{13}$$

where $v_c = Cov(S_{km}(t), S_{ml}(t))$. A necessary and sufficient condition for Crámer's ARE to equal ARE as defined in Equation (9) is that v_c equals the variance of the most efficient estimator. As both $S_{km}(t)$ and $S_{ml}(t)$ are consistent estimates of $S(t)$, then for any constant α, a linear combination of the two in the form of $\alpha S_{km}(t)(1 - \alpha)S_{ml}(t)$ is also unbiased and has variance

$$\{1 - \alpha\}^2 v_{km}(t) + 2\alpha\{1 - \alpha\}v_c(t) + \alpha^2 v_{ml}(t). \tag{14}$$

This variance is minimum at $\alpha = 0$ and the derivative at $\alpha = 0$, thus $v_c(t) = v_{ml}(t)$. This of course does not imply that $v_c(t)$ and $v_{ml}(t)$ will return equal estimates in finite samples but they have the same limit as $n \to \infty$. The estimator for $v_c(t)$ is presented in the Appendix.

5. ARE for Exponential Survival and Exponential Censoring

In this section, we provide closed form solutions based on Theorem 1 for exponential survival and censoring times. Besides the existence of a closed form solution for ARE, the ARE of the exponential distribution can be considered as a lower bound to many distributions used in survival analysis such as Weibull, Gamma, or Generalized Gamma, or any other parametric extension of the exponential distribution. The addition of extra parameters to the model results in the diagonal elements of the inverse of the information matrix being always greater than the corresponding elements of the simpler model. In the following, we assume exponential survival times with hazard λ and exponential censoring times with hazard γ. The survival function for exponential survival times is $S_\lambda(t) = e^{-t\lambda}$ with variance $Var(S_\lambda(t)) = t^2(e^{-t\lambda})^2\lambda^2 n^{-1}$. Moreover, as F is known to be an exponential distribution, the event probability for the binomial distribution is $p(t) = e^{-\lambda t}$. In this case, we have a closed form solution for Theorem 1,

$$ARE(t) = \frac{(t\lambda)^2}{(e^{\lambda t} - 1)P(X < C)\varphi(t)}. \tag{15}$$

We will now consider four scenarios: (1) no censoring, (2) random censoring, (3) type I censoring with fixed censoring times, and (4) hybrid, type I, and random censoring times. For the sake of simplicity, we will denote ARE with subscripts 1, 2, 3, and 4 for each scenario.

5.1. No-Censoring

If the survival time is recorded for every study participant, then $\varphi(t) = P(x < Y) = 1$ giving

$$ARE_1(t) = \frac{(t\lambda)^2}{(e^{\lambda t} - 1)}. \tag{16}$$

In this case, the maximum efficacy of the nonparametric estimator is 64.76% and happens at $t = 1.593\lambda^{-1}$, i.e., 1.6 times the mean survival time.

5.1.1. Random Censoring Times

In this scenario, the observed time for subject j is $T_j = \min\left(X_j, C_j\right)$ and $\delta_j = I\{X_j \leq C_j\}$. Additionally, we have

$$P(X < C) = \frac{\lambda}{\lambda + \gamma} \quad \text{and} \quad \varphi(t) = \frac{\lambda}{\lambda + \gamma} \frac{\left(e^{t(\lambda+\gamma)} - 1\right)}{\left(e^{t\lambda} - 1\right)}. \tag{17}$$

Combing the above with Equation (15), the asymptotic relative efficiency is

$$ARE_2(t) = \frac{\{t(\lambda + \gamma)\}^2}{(e^{t(\lambda+\gamma)} - 1)}. \tag{18}$$

The maximum efficacy of the nonparametric estimator is observed at $t = 1.593(\lambda + \gamma)^{-1}$, and as for the non-censored case, it equals 64.76%. As $1.593(\lambda + \gamma)^{-1} \leq 1.593(\lambda)^{-1}$, this maximum is reached earlier, and in the beginning of the follow-up period, censoring increases the efficiency of the Kaplan–Meier estimator (Figure 1).

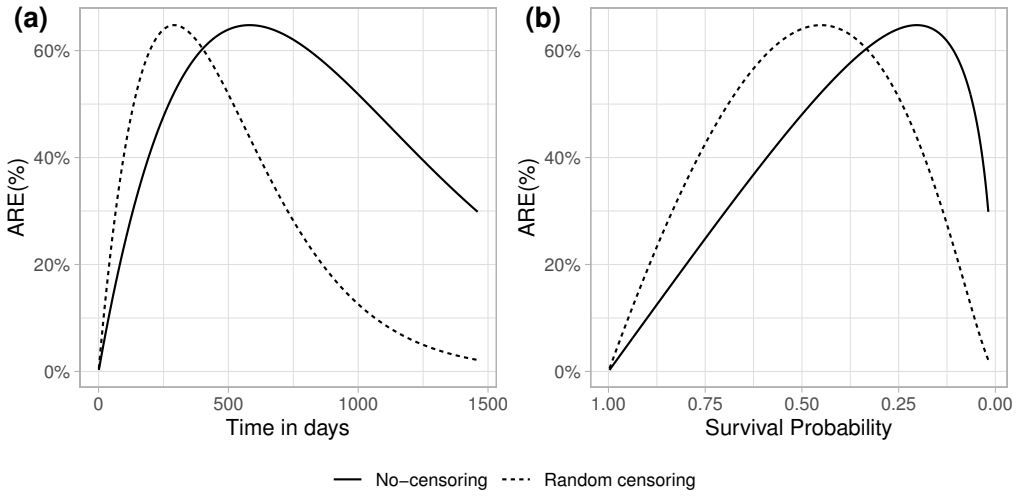

Figure 1. Numeric illustration of the Asymptotic Relative Efficiency (ARE) of the Kaplan–Meier estimator relative to the maximum likelihood estimator with a known exponential distribution with a rate of 1/365; ARE as a function of time (**a**) and as a function of survival probabilities (**b**).

5.1.2. Type I or Fixed Censoring

Here, the observed time for subject j is $T_j = \min\left(X_j, \tau\right)$and $\delta_j = I\{X_j \leq \tau\}$. In this case, $\forall t < \tau\ \varphi(t) = 1$ and $P(X < \tau) = (1 - e^{-\lambda\tau})$, resulting in

$$ARE_3(t) = \frac{(t\lambda)^2}{(e^{t\lambda} - 1)(1 - e^{-\lambda\tau})}. \tag{19}$$

As $(1 - e^{-\lambda\tau}) \leq 1$, the ARE of the nonparametric variance estimator exceeds the ARE of the nonparametric variance estimator when there is no censoring and $\lim_{\tau\to\infty} ARE_2(t) = ARE_1(t)$.

5.1.3. Hybrid Type I and Random Censoring

Here, just as for Type I censoring, follow-up is stopped at a pre-assigned time point τ, but in addition participants might leave the study prior to τ for reasons unrelated to the studied disease. Thus, $T_j = \min\left(X_j, C_j \wedge \tau\right)$. The minimum of two exponentially dis-

tributed variables with parameters λ and γ is also exponentially distributed with parameter $\lambda + \gamma$, thus

$$ARE_4(t) = \frac{(t(\lambda+\gamma))^2}{(e^{t(\lambda+\gamma)} - 1)(1 - e^{-(\lambda+\gamma)\tau})}. \quad (20)$$

Type I (or fixed) censoring has no impact on the Greenwood variance estimator, but it does reduce the information available in the data for the estimation of the parametric variance with $P(X < C \wedge \tau)$ and improves the ARE of the Kaplan–Meier estimator up τ by $P(X < C \wedge \tau)^{-1}$. At relatively short follow-up times (i.e., less than the average survival time) and high survival rates, the point-wise ARE of the Kaplan–Meier estimator at τ and close to τ can exceed 64% (Figure 2). Of course, if there is additional random censoring, the heavier the censoring the greater the asymptotic relative efficiency of the parametric survival estimator (see Appendix B). For the effect of additional dimensions to the parametric estimators, see Appendix C.

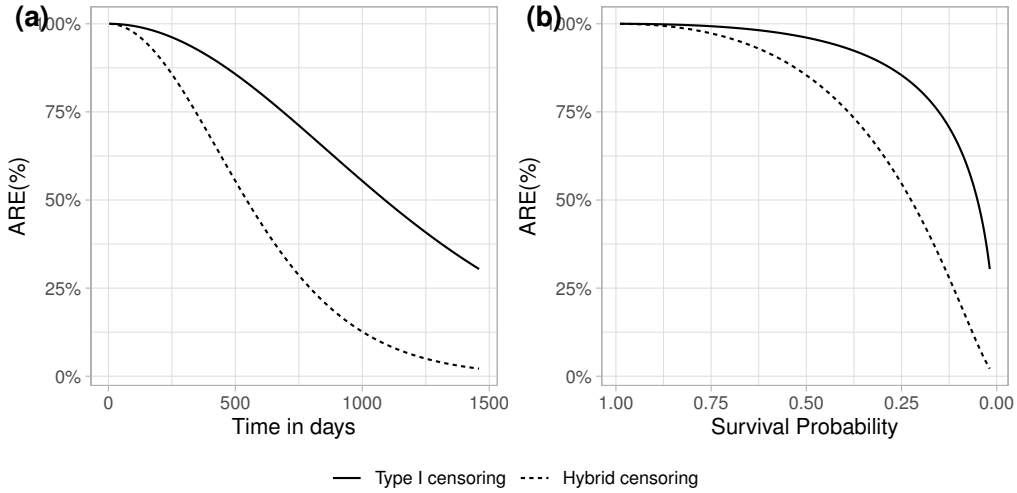

Figure 2. (**a**) Numeric illustration of the Asymptotic Relative Efficiency (ARE) of the Kaplan–Meier estimator relative to the maximum likelihood estimator with known exponential distribution with rate of 1/365 at time of the Type I censoring; ARE as a function of time (**a**) and as a function of survival probabilities (**b**).

5.2. Efficiency in Finite Samples

While comparing the asymptotic variances of different estimates is essentially comparing confidence intervals and is invariant under nonlinear transformation of the target parameter, Cox [17] highlighted that such stability of interpretation does not always hold true for small-sample comparisons. Serfling [14] noted that asymptotic relative efficiencies pertain to large sample comparisons and need not reliably indicate small sample performance, and exemplified with the relative efficiency of median vs mean as an estimator of location for normal distribution. Identical trends can be seen in the survival analysis as well. For the closely related Nelson–Allen estimator, Peña and Rohatgi [18] concluded that, for sample sizes under 30 asymptotic approximations can be misleading, specially at the extremes of the follow-up times. With the use of a simulation study, we looked at the ratio of variances' small sample characteristics. We assumed survival times are exponential with a rate of 1/365 and with sample sizes ranging from 25 to 1000. We have fitted a parametric and the non-parametric Kaplan–Meier estimator to the data and survival times of 50, 180, and 365. Table 1 summarises the results of 1000 such simulations. The findings imply that the Kaplan–Meier estimator's small sample relative effectiveness depends on both the sample size and the survival time. In finite samples, the ratio of variances typically surpasses ARE, and this is particularly clear at the start of the follow-up when ARE is low.

This can be somewhat explained by how reliable the estimation is. Outliers in small samples can inflate variances and skew parametric survival estimates [6]. The finite sample bias of the maximum likelihood parameter [19,20] and the Kaplan–Meier estimator complicates comparisons, and the likely comparison of Mean Squared Errors should be preferred over a comparison of variances [21].

Table 1. Ratio of variances for Kaplan–Meier and parametric survival estimates for exponential survival times with a median survival time of 365 as a function of sample size and the procentual deviation for ARE in parenthesis.

N	t = 50	t = 180	t = 365
ARE	12.78	38.15	58.20
25	16.35 (27.92%)	40.63 (19.4%)	59.15 (7.45%)
50	14.89 (16.49%)	39.32 (9.15%)	58.67 (3.70%)
75	14.03 (9.77%)	39.09 (7.35%)	58.48 (2.21%)
100	13.65 (6.79%)	38.65 (3.91%)	58.31 (0.88%)
125	13.57 (6.17%)	38.69 (4.22%)	58.51 (2.44%)
150	13.35 (4.45%)	38.61 (3.59%)	58.41 (1.66%)
200	13.22 (3.43%)	38.44 (2.26%)	58.34 (1.11%)
250	13.11 (2.57%)	38.27 (0.93%)	58.30 (0.37%)
1000	12.89 (0.85%)	38.23 (0.62%)	58.23 (0.25%)

6. Discussion

With a focus on the impact of censoring and different forms of censoring on ARE, we compared the asymptotic relative efficiency of the Kaplan–Meier survival estimator vs. parametric survival estimators in this work. Due to the existence of closed-form solutions as well as historical considerations, we have used the exponential distribution as an example. Previous papers on the subject noted that the maximum ARE of the Kaplan–Meier estimator in the case of exponential survival times is 64% [4,8,22]. However, the maximum ARE can be close to 1, as we have demonstrated in studies using type I censoring. This is particularly clear when the average survival time is longer than the follow-up time. Heuristically, the amount of information available for estimation and the way in which this information is used are the two main factors that account for the Kaplan–Meier estimator's relatively low ARE in comparison to parametric estimators. The Kaplan–Meier survival and the Greenwood variance estimator only use data up to time t and censor anything after that for any given time-point t. All available data are utilized by parametric estimators. The way in which these data are used is the second factor that needs to be taken into account. The parameter vector of the assumed distribution is estimated by parametric models. Since the parameters of the distribution and time are mapped one-to-one to parametric survival probabilities, the correlation between survival estimates at any given two time points is 1. Naturally, knowledge of the survival probabilities at t_1 also provides knowledge of the Kaplan–Meier estimates at t_2. If $t_1 < t_2$, then the two survival probabilities are correlated and, by definition, $S(t_1) > S(t_2)$. However, as the time interval between t_1 and t_2 increases and as the percentage of censored observations increases, the correlation between $S(t_1)$ and $S(t_2)$ decreases (see Appendix D). Asymptotic Relative Efficiency was long associated with significance testing, and the alternative formulation as the limiting correlation coefficient between estimators was first applied in [23,24]. With the help of simulation, we have illustrated that the realizations, on average, are equal between the ratio of variances and the Crámer correlation estimate of ARE. From the perspective of correlation, we can look at ARE as the amount of (linear) information that the Kaplan–Meier estimate contains about the form of the parametric model. Near zero ARE or correlation at the beginning of the follow-up likely suggests that the Kaplan–Meier estimate confers limited knowledge about possible parametric estimators, which might complicate the data driven choice of parametric estimators [8,25,26]. While theoretically this is a problem at follow-up times

where the survival probability approaches zero, in practice, at that time, survival estimates are rarely meaningful from an interpretation point of view [27].

In this paper, we offered a more nuanced view of the question raised by Miller [4], the price (i.e., loss of efficiency) of the routine use of the Kaplan–Meier estimator. We looked into the different censoring types and their effect of the ARE. Contrary to previous studies, we suggest that the price of the Kaplan–Meier estimator is acceptable. As we have shown, studies with Type I censoring have relatively high efficiency. We have also seen that adding extra dimensions to parametric estimators increases the efficiency of the Kaplan–Meier estimator. While the exponential distribution is widely used in theoretical works and research planning phases, its relatively rigid form makes it an unlikely candidate for a parametric model. Looking at ARE from the perspective of correlation suggests that applying the framework of Jullum and Hjort [8,25] to select parametric models for studies with follow-up times shorter than the mean survival time could be difficult. It is likely that, for the foreseeable future, the Kaplan–Meier estimator will be the method of choice in survival analysis.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Covariance and Correlation between Estimators of Survival

Approximation by Averages theorem [28] states that for W_i that are *iid* the statistic $\mathcal{T}(W_1, ..., W_n)$ has the following approximation

$$\mathcal{T} - \mathcal{T}_\infty = \frac{1}{n}\sum_{j=1}^{n} h(X_j) + R_n \tag{A1}$$

where $\sqrt{n}R_n \xrightarrow{P} 0$ as $n \to \infty$. For finding h we define the survival function as a functional statistic $\mathcal{T}$ represented in terms of the empirical distribution function F_n, indexed by the sample size n. For functional statistics h is represented by influence curve (or function)

$$h_{\mathcal{T}}(w) = \lim_{\epsilon \to 0} \frac{\{\mathcal{T}[(1-\epsilon)F + \epsilon\Delta_w] - \mathcal{T}(F)\}}{\epsilon}. \tag{A2}$$

where Δ_w is a a distribution function that puts all probability mass in the point w.

Further, for afunctional $\mathcal{T}(F) = \int_R a(x)dF(x)$ we denote the influence components

$$h_F(w_j) = a(w_j) - \mathcal{T}(F), \tag{A3}$$

a measure the effects of individual data point j on the estimand. Using the functional Taylor's theorem (von Mises expansion) we can expand the statistical functional of interest as

$$\mathcal{T}(\hat{F}) = \mathcal{T}(F) + \frac{1}{n}\sum_{j} h_F(x_j) + o_p(1). \tag{A4}$$

which leads to

$$E\{h_F(x)\} = 0 \quad \text{and} \quad Var(\mathcal{T}(F)) = \frac{E\{h_F^2(x)\}}{n}. \tag{A5}$$

This is true irrespective if we consider the nonparametric (h_{km}) or parametric (h_{ml}) influence components.

Reid [29] derived the influence function for censored data and showed that variance based on h is equivalent to the Greenwood estimator.

For a parametric family with parameter θ Equation (A1) can be rewritten as

$$\theta_{ml} - \theta_{\infty} = \frac{1}{n}\sum_{j=1}^{n} J^{-1}(\theta)\left\{\frac{\partial}{\partial\theta}\log f(x_j,\theta)\right\} + R_n \tag{A6}$$

thus influence function and the score function of the likelihood function are scalar multiples of each other, with Fisher information as multiplication factor [3]. With all this in place, by using the characteristics of the covariance

$$Cov\{\hat{S}_{km}(t) - S(t), \hat{S}_{ml}(t) - S(t)\} = Cov\{\hat{S}_{km}(t), \hat{S}_{km}(t)\} \tag{A7}$$

and noting that the

$$E\{\hat{S}_{km}(t) - S(t)\} = E\{h_{km}\} = 0 \tag{A8}$$

and

$$E\{\hat{S}_{ml}(t) - S(t)\} = E\{h_{ml}\} = 0 \tag{A9}$$

we have

$$v_c = \frac{1}{n}\left\{\frac{\partial S_{ml}(t,\theta)}{\partial\theta}\right\}\sum_{j} U_{\theta}(x_j)h_{km}(x_j) \tag{A10}$$

With the help of a simulation study we compare the long-run properties of the two competing estimators: the ratio of variances and Crámer correlation estimate of ARE. We assumed exponential survival and censoring times, both with rate parameters of 1/365, i.e., a 50% censoring rate. The sample size was set to 1000, to assure that we could obtain nonparametric estimates of survival and associated variability up to twice the mean survival time. We run 1000 simulations and used as reference the analytically derived ARE (Equation (18)). As Figure A1 illustrates the the choice of estimator between the ratio and Crámer correlation estimate of ARE it is a choice of convenience, and no appreciable differences between the two were observed. True, simulation studies falls short of analytical proof of equivalence. However, changing the parameters of the simulation led to similar results.

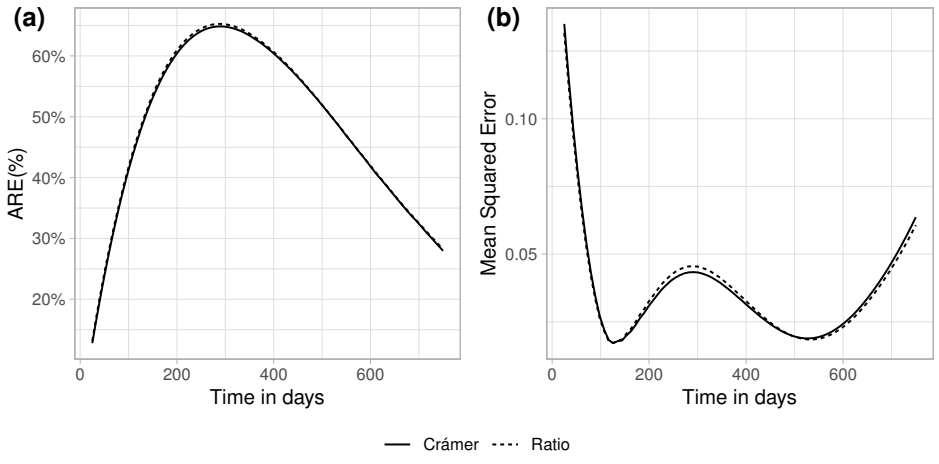

Figure A1. (**a**) The ratio of variances and Crámer correlation estimate of ARE of the Kaplan–Meier estimator relative to the maximum likelihood estimator for exponential survival and censoring times (rate of 1/365) and (**b**) the Mean Squared Error of the tow estimators as function of follow-up time.

Appendix B. ARE under Hybrid Censoring

As we noted in the main text, Type I fixed censoring increases the ARE of the Kaplan –Meier survival estimator with factor $P(X < C \wedge \tau)$. As Figure A2 illustrates at relatively short Type I censoring times the ARE of the Kaplan–Meier estimator increases and at any time point up to τ, and this increase $\{P(X < C \wedge \tau)\}^{-1}$. As $\tau \to \infty$ this effect diminishes.

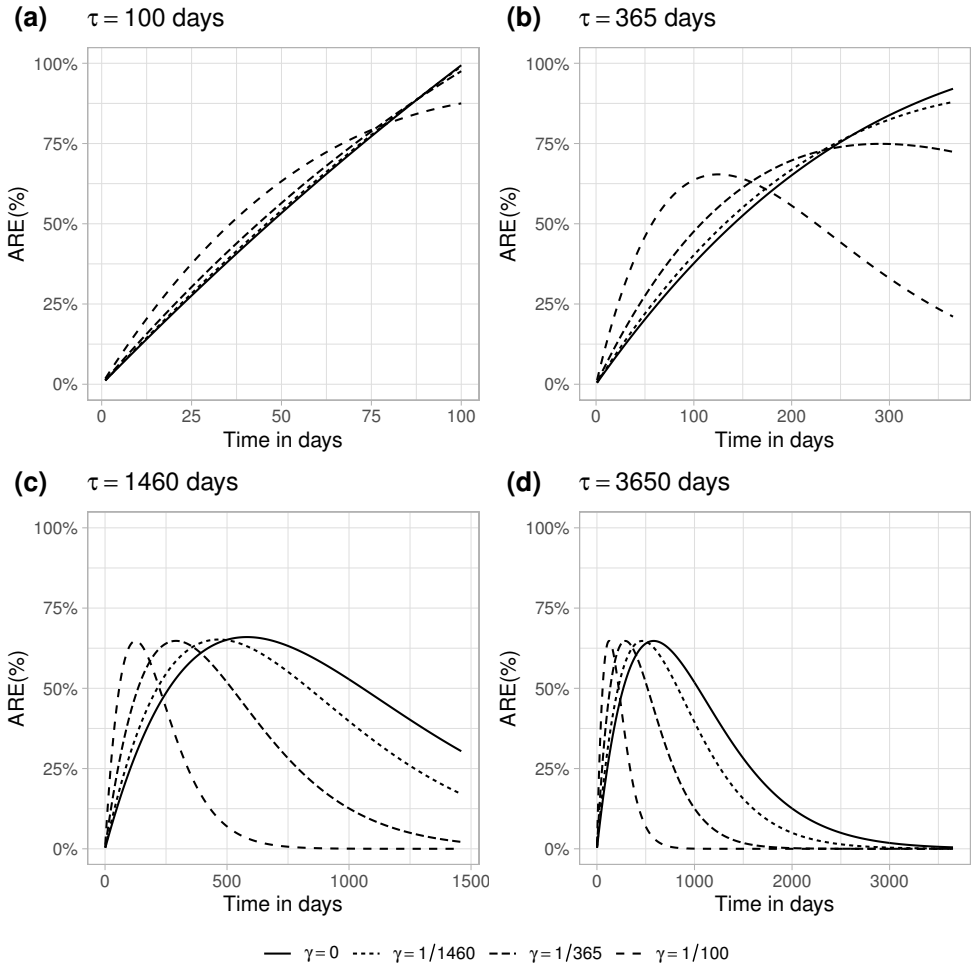

Figure A2. Asymptotic Relative Efficiency as a function of random censoring and the timing of the Type I censoring.

Appendix C. Multi-Dimensional Parametric Models

A set of simulations was run to illustrate the effect of adding additional parameters to the parametric models. As previously, we have simulated 1000 observation with exponential survival times (rate of 1/365) and exponential censoring times (rate of 1/365 leading to 50% and 1/1460 leading to 20% cenoring). In addition to the Exponential distribution, we have used the Weibull, Gamma and Generalized Gamma distributions for estimation of survival probabilities and associated variances. These later three distributions are extensions of the Exponential distribution with the addition of a power parameter (Weibull), convolution parameter (Gamma) or both (Generalized Gamma) [30], and if these additional parameters equal 1 then they simplify to the Exponential distribution. Thus, all three are consistent estimators of the survival probability. Figure A3 summarises the results of 1000 simulations. As expected the ARE for Kaplan–Meier estimator increases as additional parameters added to the parametric models. While ARE against the Exponential model followed what Equation (18) showed, the Kaplan–Meier estimator had better efficiency when compared against the Gamma, Weibull, and especially the three-parameter generalized Gamma distribution. This results are expected and the loss in efficacy generally becomes small as the number of

parameters in the parametric model increases [31]. In addition one can see that increase in censoring reduced the ARE of the Kaplan–Meier estimator.

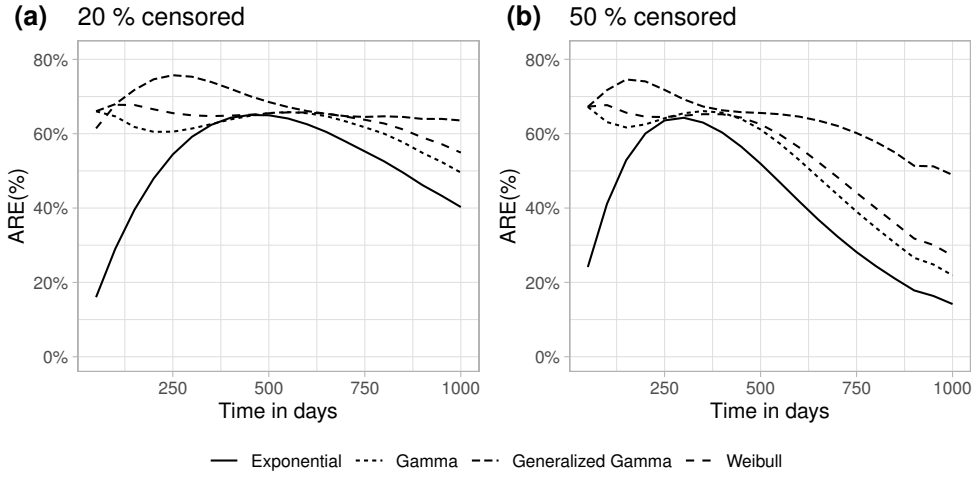

Figure A3. Asymptotic Relative Efficiency of the Kaplan–Meier estimator against parametric models with moderate (**a**) and high censoring precentage (**b**).

In addition to the three above mentioned distributions related to the Exponential distribution, we have tested the Log-logistic distribution. The log-logistic distribution is frequently used for events whose rate increases initially and decreases later, as, for example, mortality rate from cancer following diagnosis or treatment. The survival function is given by $S(t) = \left\{1 + (t/\alpha)^{\beta}\right\}^{-1}$, β is the shape parameter and α is the scale. Both parameters are strictly positive and α is the median of the survival times and with increasing β the probability distribution of the survival a times are increasingly concentrated around α. In a simulation study we have set $\alpha = 365$ and $\beta = 1, 2, 3, \& 4$.

As it can be seen on Figure A4 a with increasing β the events are concentrated more and more around the median survival time. In addition with accelerated event rate the ARE of the Kaplan–Meier estimator rapidly approaches zero at the extremes of the survival time. At $\beta = 1$ the ARE of the Kaplan–Meier estimator stays in the vicinity of the maximum ARE of 75%.

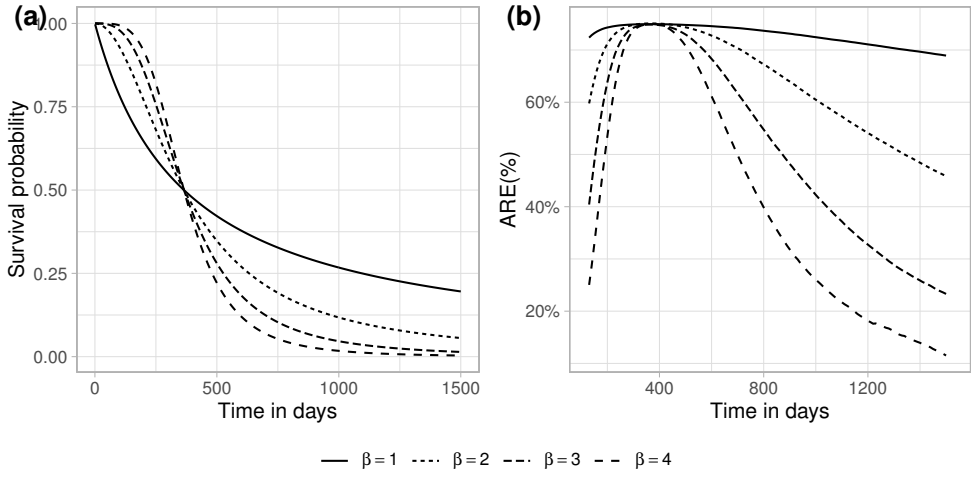

Figure A4. Survival (**a**) and Asymptotic Relative Efficiency of the Kaplan–Meier estimator for Log-logistic survival times (**b**).

Appendix D. Correlation between Kaplan–Meier Survival Estimates at Different Time Points

Derivation of the correlation coefficient for Kaplan–Meier survival estimates at different time points largely follows Kaplan and Meier [1] (see Section 6.2 and Equations (2h) and (2g)), however with the notation updated according to Meier [22].

For survival times $t_1 \leq t_2$ the correlation (ρ) between the Kaplan–Meier survival estimates is

$$\rho\{S_{km}(t_1), S_{km}(t_2)\} = \sqrt{\frac{1 - S_{km}(t_1)}{S_{km}(t_1)} \frac{S_{km}(t_2)}{1 - S_{km}(t_2)} \frac{\varphi(t_1)}{\varphi(t_2)}} \tag{A11}$$

where $\varphi(t)$ is the variance inflation factor of the binomial variance due to censoring. For any given t_1 the correlation between $S_{km}(t_1)$ and $S_{km}(t_2)$ decreasing with distance between t_1 and t_2. This is easy to see, as with t_2 increasing $S_{km}(t_2)$ decreasing and $S_{km}(t_2)\{1 - S_{km}(t_2)\}^{-1}$ is getting smaller. For survival times $t_1 \leq t_2$ we have that $\varphi(t_2) \leq \varphi(t_2)$ so its easy to see that censoring further reduces correlation.

References

1. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481.
2. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B (Methodol.)* **1972**, *34*, 187–202. [CrossRef]
3. Hjort, N.L. On inference in parametric survival data models. *Int. Stat. Rev./Rev. Int. Stat.* **1992**, *60*, 355–387. [CrossRef]
4. Miller, R.G., Jr. What price Kaplan–Meier? *Biometrics* **1983**, *39*, 1077–1081. [CrossRef]
5. Klein, J.P.; Moeschberger, M.L. The robustness of several estimators of the survivorship function with randomly censored data. *Commun. Stat.-Simul. Comput.* **1989**, *18*, 1087–1112. [CrossRef]
6. Aranda-Ordaz, F.J. Relative efficiency of the Kaplan–Meier estimator under contamination: Relative efficiency of the kaplan-meier. *Commun. Stat.-Simul. Comput.* **1987**, *16*, 987–997. [CrossRef]
7. Meier, P.; Karrison, T.; Chappell, R.; Xie, H. The price of Kaplan–Meier. *J. Am. Stat. Assoc.* **2004**, *99*, 890–896. [CrossRef]
8. Jullum, M.; Hjort, N.L. What price semiparametric Cox regression? *Lifetime Data Anal.* **2019**, *25*, 406–438.
9. Cheng, P.E.; Lin, G.D. Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Stat. Probab. Lett.* **1987**, *5*, 75–80.
10. Klein, J.P. Small sample moments of some estimators of the variance of the Kaplan–Meier and Nelson-Aalen estimators. *Scand. J. Stat.* **1991**, *18*, 333–340.
11. Zheng, G.; Gastwirth, J.L. On the Fisher information in randomly censored data. *Stat. Probab. Lett.* **2001**, *52*, 421–426. [CrossRef]
12. Zheng, G.; Balakrishnan, N. Fisher Information: Nonstandard Situations. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.
13. Nadarajah, S. Reliability for lifetime distributions. *Math. Comput. Model.* **2003**, *37*, 683–688. [CrossRef]
14. Serfling, R. Asymptotic Relative Efficiency in Estimation. *Int. Encycl. Stat. Sci.* **2011**, *23*, 68–72.
15. Nemes, S.; Gustavsson, A.; Taib, Z. Variance Inflation Due to Censoring in Survival Probability Estimates. *Statistica* **2020**, *80*, 395–412.
16. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
17. Cox, D.R. *Principles of Statistical Inference*; Cambridge University Press: Cambridge, UK, 2006.
18. Peña, E.A.; Rohatgi, V.K. Small sample and efficiency results for the Nelson-Aalen estimator. *J. Stat. Plan. Inference* **1993**, *37*, 193–202. [CrossRef]
19. Mardia, K.V.; Southworth, H.R.; Taylor, C.C. On bias in maximum likelihood estimators. *J. Stat. Plan. Inference* **1999**, *76*, 31–39. [CrossRef]
20. Xuemao Zhang, S.P.; Wang, Y.G. Small sample bias correction or bias reduction? *Commun. Stat.-Simul. Comput.* **2021**, *50*, 1165–1177. [CrossRef]
21. Geurts, J. Some Small–Sample Non–Proportional Hazards Resulis For The Kaplan–Meier Estimator. *Stat. Neerl.* **1985**, *39*, 1–13. [CrossRef]
22. Meier, P. Estimation of a distribution function from incomplete observations. *J. Appl. Probab.* **1975**, *12*, 67–87. [CrossRef]
23. Noether, G.E. On a theorem of Pitman. *Ann. Math. Stat.* **1955**, *26*, 64–68. [CrossRef]
24. Van Eeden, C. The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *Ann. Math. Stat.* **1963**, *34*, 1442–1451. [CrossRef]
25. Jullum, M.; Hjort, N.L. Parametric or nonparametric: The FIC approach. *Stat. Sin.* **2017**, *27*, 951–981. [CrossRef]
26. Nemes, S.; Gustavsson, A.; Jauhiainen, A. Focused Information Criterion for Restricted Mean Survival Times: Non-Parametric or Parametric Estimators. *Entropy* **2022**, *24*, 713.

27. Gebski, V.; Garès, V.; Gibbs, E.; Byth, K. Data maturity and follow-up in time-to-event analyses. *Int. J. Epidemiol.* **2018**, *47*, 850–859. [CrossRef] [PubMed]
28. Boos, D.; Stefanski, L. *Essential Statistical Inference: Theory and Methods*; Springer: New York, NY, USA, 2013.
29. Reid, N. Influence Functions for Censored Data. *Ann. Stat.* **1981**, *9*, 78–92. [CrossRef]
30. Marshall, A.W.; Olkin, I. *Life Distributions*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 13.
31. Efron, B. Logistic regression, survival analysis, and the Kaplan–Meier curve. *J. Am. Stat. Assoc.* **1988**, *83*, 414–425. [CrossRef]