

Article

A Brief Overview of Restricted Mean Survival Time Estimators and Associated Variances

Szilárd Nemes ^{1,2,*} , Erik Bülow ^{2,3}  and Andreas Gustavsson ¹

¹ BioPharma Early Biometrics and Statistical Innovation, Data Science & AI, BioPharmaceuticals R&D, AstraZeneca, 43183 Gothenburg, Sweden; andreas.gustavsson@astrazeneca.com

² Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, 40530 Gothenburg, Sweden; erik.bulow@vgregion.se

³ The Swedish Hip Arthroplasty Register, Registercentrum Västra Götaland, 41390 Gothenburg, Sweden

* Correspondence: szilard.nemes@astrazeneca.com

Received: 1 April 2020; Accepted: 22 May 2020; Published: 26 May 2020



Abstract: Restricted Mean Survival Time (*RMST*) experiences a renaissance and is advocated as a model-free, easy to interpret alternative to proportional hazards regression and hazard rates with implication in causal inference. Estimation of *RMST* and associated variance is mainly done by numerical integration of Kaplan–Meier curves. In this paper we briefly review the two main alternatives to the Kaplan–Meier method; analysis based on pseudo-observations, and the flexible parametric survival method. Using computer simulations, we assess the efficacy of the three methods compared to a fully parametric approach where the distribution of survival times is known. Thereafter, the three methods are directly compared without any distributional assumption for the survival data. Generally, flexible parametric survival methods outperform both competitors, however the differences are small.

Keywords: *RMST*; censoring; variance estimator; efficacy; Kaplan–Meier; flexible-survival methods; pseudo-observation

1. Introduction

In 1958, Kaplan and Meier [1] published their paper on the product limit estimator for survival probabilities, an estimator that can handle censored and/or truncated survival data. The easy to use and interpretable nature of the method, made it an instant success. The Kaplan–Meier estimator have since then become the dominant method to summarize survival data and parametric methods have lost relevance. This prompted Miller [2] to criticize this over-reliance of applied scientist on the Kaplan–Meier curve. Miller established, both theoretically and numerically, the loss of efficacy of the Kaplan–Meier curve compared to parametric methods. Jullum and Hjort [3] further explored this aspect and corroborated Miller’s findings. Meier et al. [4] have addressed some of Miller’s critique, by highlighting the difficulties in choosing the right parametric form. In addition to the Kaplan–Meier curve, Meier et al. [4] have addressed the efficacy of a functional of the Kaplan–Meier curve: the Restricted Mean Survival Time (*RMST*), the average survival time up to a given time point. Theoretical aspects of *RMST* (or Kaplan–Meier integrals) are well studied [5–7] and currently *RMST* is experiencing a renaissance [8–10] being hailed as a model-free, easy to interpret statistic with implications in causal inference [11]. Not needing to establish the proper parametric model has certain advantages. However, generally non-parametric methods have lower efficacy than parametric ones. Generally, practical applications do not consider parametric estimators for *RMST* and use either numerical integration of the Kaplan–Meier curve [12], pseudo-observations [13–15] or flexible parametric survival methods [16]. In this paper, we briefly summarize parametric and non-parametric

estimators for RMST and associated variance estimators. We will use Monte Carlo simulations to examine their efficacy.

2. Restricted Survival Times

We assume that survival times to an event of interest X_1, \dots, X_n are independently and identically distributed (*iid*) according to the cumulative distribution function $F(x)$ and survival function $S(x)$. Similarly we assume C_1, \dots, C_n to be *iid* censoring times according to the distribution function $G(c)$. Additionally, we assume that follow-up times are restricted, by an upper threshold τ set by the researcher. Thus the actual observed time for subject j is $T_j = \min(X_j, C_j, \tau)$. Additionally, we have $\delta_j = I\{X_j \leq \min(C_j, \tau)\}$ as event indicator that takes a value of 1 if the event of interest is recorded at the given time, 0 otherwise. We assume independence between failure and censoring time. We let $t_{(1)} \leq \dots \leq t_{(n)}$ denote the ordered observed survival times and $\delta_{(1)}, \dots, \delta_{(n)}$ their associated indicator values. Survival times are summarized with the help of the survival function $S(t) = P(T > t)$ and the τ -restricted mean survival times are given by $\int_0^\tau S(t) dt$ [12].

From Andersen et al. [14] we know that $\sqrt{n} \left(\int_0^\tau \hat{S}(t) dt - \int_0^\tau S(t) dt \right) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$ for some variance $\sigma^2 > 0$. In the next sections we review, then evaluate, different estimators for $\int_0^\tau S(t) dt$ and σ^2 .

3. Estimators of RMST and Associated Variance

The statistical literature describes several parametric and non-parametric ways of estimating $\int_0^\tau S(t) dt$ and its associated variance.

3.1. Parametric Methods

For parametric estimation, we need to make assumptions about the distribution of the unobserved survival times, X . Once the distribution function is established, the Restricted Mean Survival Time (μ_τ) is estimated as

$$\mu_\tau = \int_0^\tau S(\theta, t) dt.$$

where θ is the parameter vector for the assumed distribution function.

In the following, we outline three possibilities for parametric estimation

1. Likelihood and δ -method based approach.

Estimation of $\int_0^\tau S(\theta, t) dt$ requires knowledge of θ . Estimates for θ under censoring can be obtained with maximum likelihood. Given the censoring indicator $I_{\{X_j \leq C_j\}}$ and $Y_j = \min(X_j, C_j)$ we can define the likelihood function for θ as

$$L(\theta|Y, \delta) = \prod_{j=1}^n f(Y_j, \theta)^{I_{\{X_j \leq C_j\}}} [1 - F(Y_j, \theta)]^{1 - I_{\{X_j \leq C_j\}}}.$$

For most cases there are no closed-form solutions for $\hat{\theta}$ and $Var(\hat{\theta})$, however, numerical estimates are enough. Once estimates for θ are available, $\int_0^\tau S(\theta, t) dt$ can be estimated either by using $\hat{\theta}$ in closed form equations, or as plug-ins in numerical integration. The variance for $\int_0^\tau S(\theta, t) dt$ by the δ -method is given by

$$Var(\mu_\tau) = \sum_i \left(\frac{\delta\mu_\tau}{\delta\theta_i} \right)^2 Var(\theta_i) + \sum_i \sum_{i \neq j} \frac{\delta\mu_\tau}{\delta\theta_i} \frac{\delta\mu_\tau}{\delta\theta_j} Cov(\theta_i, \theta_j)$$

where θ_i is the i^{th} element of the population parameter.

2. M-estimation (or estimating equation)

M-estimation seeks solution to the vector equation $\sum_{i=1}^n \psi(T_i, \hat{\theta}) = 0$. Here, T_i are independently and identically distributed restricted survival times, $\hat{\theta}$ is a p -dimensional parameter and ψ is a known $(p \times 1)$ -function that does not depend on i or n [17]. M-estimates are asymptotically normal with variance $n^{-1}V(\theta_0)$. $V(\theta_0)$ is a sandwich variance estimator given by

$$V(\theta_0) = A(\theta_0)^{-1} B(\theta_0) \{A(\theta_0)^{-1}\}^T. \quad (1)$$

where,

$$A(\theta_0) = \mathbb{E} \left[- \frac{\partial}{\partial \theta^T} \psi(T_i, \theta) \right] \quad (2)$$

and

$$B(\theta_0) = \mathbb{E} \left[\psi(T_i, \hat{\theta}) \psi(T_i, \theta)^T \right]. \quad (3)$$

Wang [18] discusses alternative formulations for ψ with regards to the strictly non-negative and often skewed nature of survival data. For an in-depth discussion on the M-estimators the reader is referred to [17,19].

3. The second cumulant

The variance equivalent to the second cumulant of a probability distribution of T_j . If there is no censoring present $Var(T) = \mathbb{E}_\tau[T^2] - \mathbb{E}_\tau[T]^2$ and can be estimated as

$$Var(T) = 2 \int_0^\tau tS(t) dt - \left[\int_0^\tau tS(t) dt \right]^2$$

leading to $Var(\mu_t) = n^{-1}Var(T)$.

This estimator is not practical due to censoring. When we have censoring in the data Rosyton & Parmar [20] suggested multiplying $n^{-1}Var(T)$ with a positive scaling factor ϕ , so that $\phi = 1$ if no censoring and $\phi > 1$ otherwise. The scaling factor ϕ can be estimated with help of Monte Carlo simulation.

Alternatively, the variance can be estimated with the Stute estimator [6]. Adopting the notation of Stute [21], (i.e., $\int x dF(x) = \int x f(x) dx$) and $(1 - H) = (1 - F)(1 - G)$ and φ a transformation of X so that $n^{-1} \sum_i \varphi(X_i) \rightarrow \int \varphi dF \equiv S^\varphi = \mu_\tau$. Here, $Var[\int \varphi d\hat{F}_n] = n^{-1}\sigma_1^2$ where

$$\sigma_1^2 = \int_0^\tau \frac{\varphi^2}{1 - G} dF - \left[\int_0^\tau \varphi dF \right]^2 - \int_0^\tau \left[\int_x^\tau \varphi dF \right]^2 \frac{1 - F(x)}{[1 - H(x)]^2} G(dx).$$

Both approaches require assumptions about the distribution of the survival times and censoring times. These two estimators of variance will not be further evaluated, however they do represent an important aspect of *RMST*.

3.2. Flexible Parametric Survival Methods

In 2002, Royston and Parmar [22] introduced flexible parametric survival models as a simple way to deal with non-proportional hazards, and as a way of easy visualization of the hazard function. In a later paper the authors exemplified the utility of the model in estimation of *RMST* [16]. The model is defined through the log cumulative hazard function as

$$\ln H(t, z) = \ln H_0(t) + z^T(\beta), \quad (4)$$

where the baseline (H_0) is modeled as a restricted cubic spline in log time, $\ln H_0(t) = s(\ln t)$, and where z is a covariate vector. The splines $s(\ln t)$ are linear combination of basis functions and regression parameters. The model requires a number of knots and a probability distribution function. This later is often assumed to be the Weibull distribution, which we used in this article, with 2 knots. *RMST* is estimated by predicting the log cumulative hazard function using Equation (4) at a suitable grid of time points equally spaced between 0 and τ . Thereafter the cumulative hazard is transformed into the survival function $\hat{S}_{RP}(t)$, and $\mu_\tau = \int_0^\tau \hat{S}_{RP}(t) dt$. The variance is estimated with the δ -method or by bootstrapping.

While this approach is fully parametric and the survival function is completely specified, the model is ought to be robust against distribution misspecification. Integration proceeds with the formula of the survival function of the assumed distribution, but with numerical integration for the area under $\hat{S}_{RP}(t)$ between 0 and τ . Additionally, the splines and the piece-wise modeling between the knots offer great flexibility to the model.

3.3. Non-Parametric Methods

3.3.1. The Kaplan–Meier-Method

The Kaplan–Meier method is perhaps the most well-known and used method for estimation of *RMST*. In 1975, Meier [12] proved that the substitution of the Kaplan–Meier estimator in $\int_0^\tau S(t) dt$ provides an (approximately) unbiased estimate that is asymptotically normal.

The Kaplan–Meier estimator is given by

$$\hat{S}_{KM}(t) = \prod_{T_i \leq t} \left[1 - \frac{\delta_i}{R_i} \right]$$

where T_i is the ordered follow up time, R_i is the number of subjects at risk prior to T_i and δ_i the number of events that happened at time T_i . The *RMST* is estimated as

$$\mu_\tau = \int_0^\tau \hat{S}_{KM}(t) dt.$$

As far as we know, Irwin in 1949 [23], was the first to provide an estimator for the variance based on life tables. Later, Meier [12] in 1975 proved that reducing the data to the life table format was unnecessary and established the following variance estimator

$$V(\mu_\tau) \approx 2 \int_0^\tau \int_u^\tau S(u)S(v)C(u) dv du,$$

where $C(u) = \sum_i [(1 - S(u_i))/nS(u_i)]$ is the sum of the squared coefficient of variation for $S(u)$. Klein & Moeschberger [24] provided a variant with the Greenwood variance estimator of survival probabilities as plug-ins and popularized Meier's variance estimator in the following form:

$$\hat{V}(\mu_\tau) = \sum_{i=1}^D \left[\int_{t_i}^{\tau} \hat{S}_{KM}(t) dt \right]^2 \frac{\delta_i}{Y_i(Y_i - \delta_i)},$$

where $D = \sum_i \delta_i$. The Aalen - Johansen variance estimator of survival probabilities could also serve as plug-in in Meier's estimator, or re-sampling methods can be used as well. However, as Eaton et al. [25] noted based on Monte Carlo simulations, the estimator with the Greenwood plug-in give variance estimates closest to empirical and asymptotic variances.

3.3.2. Pseudo-observations

Pseudo-observations is a method that is based on the pseudo-values from a jackknife statistic constructed from simple summary statistic estimates which are then used in a generalized estimating equation to obtain estimates of the model parameters [13,15].

The parameter of interest is μ_τ . If there is no censoring, then $\mu_\tau = n^{-1} \sum_i T_i$. Under censorship, an approximately unbiased estimator for μ_τ is $\int_0^\tau \hat{S}(t) dt$. The pseudo-observation T_j^* associated with the j^{th} observed time is estimated as

$$T_j^* = n \int_0^\tau \hat{S}(t) dt - (n-1) \int_0^\tau \hat{S}^{-j}(t) dt$$

where \hat{S}^{-j} is the survival curve estimated without the j^{th} observation. The assumption is that $\mu(T^*) \approx \mu(T)$ and $\sigma^2(T^*) \approx \sigma^2(T)$. The most straightforward estimators are $\hat{\mu}(T^*) = n^{-1} \sum_j T_j^*$ and $\sigma^2(T^*) = (n-1)^{-1} \sum_j (T_j^* - \bar{T}^*)^2$. The approach preferred by Andersen et al. [13] is based on M-estimators. The usual assumption is that the underlying distribution is symmetric about the location parameter and the corresponding function ψ is an odd function for the location parameter (mean in our case, θ_1) and an even function for the targeted scale parameter (variance, θ_2) defined as

$$\sum_{i=1}^n \psi(T_j^*, \hat{\theta}_1, \hat{\theta}_2) = \begin{pmatrix} (T_j^* - \theta_1) \\ (T_j^* - \theta_1)^2 - \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Solving for $V(\theta_0)$ (Equation (1)) starts with solving $A(\theta_0)$ (Equation (2)) and $B(\theta_0)$ (Equation (3)). In this case $A(\theta_0) = I(\theta_0)$, the identity matrix. Solving for $B(\theta_0)$ gives

$$B(\theta_0) = \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}$$

where, μ_k is the k^{th} central moment. The variance estimator $V(\theta_0) = I(\theta_0)^{-1} B(\theta_0) \{I(\theta_0)^{-1}\}^T = B(\theta_0)$. The variability estimates by $V(\theta_0)$ are robust to model misspecification. For symmetric distributions (i.e., $f(\mu+x) = f(\mu-x)$) the off-diagonal elements of $V(\theta_0)$ are zero. For asymmetric distributions (i.e., $\mu_3 > 0$) there is an asymptotic variance inflation due to estimating variance in addition to the mean [19].

4. Simulation Studies

We used Monte Carlo simulations to gain insight of the small sample proprieties of the estimators. We used the R computing environment (R 3.6.0) [26] for calculations and illustrations. Codes are available from the corresponding author.

4.1. Relative Efficacy under Parametric Assumption

A first set of simulation studies was used to assess the relative efficacy of non-parametric methods and flexible parametric survival methods compared to the parametric estimates of variance. The relative efficiency (*eff*) of the estimators was assessed as the ratio between the estimated variances of the different estimators and the estimator with the minimum variance

$$eff = \frac{\min \langle \sigma_M^2 \rangle}{\sigma_M^2},$$

where M is the model considered for estimation. We were interested in seeing how relative efficacy varies as a function of (i) sample size, (ii) censoring percentage and (iii) restriction time. To facilitate calculation and possibility of obtaining closed form solutions for the parametric variance estimates, we assumed that survival times were exponentially distributed with rate, $\lambda = 1/365$. Similarly, we assumed that censoring times were exponentially distributed with rate γ .

In the first set of simulations we set the sample size to 50 and 100 with censoring percentages at 50% and 75%. Given the assumed censoring percentages (cp) we estimated γ , the rate parameter for the censoring distribution, as the solution for the equation

$$cp = 1 - \lambda(\lambda + \gamma)^{-1} \quad (5)$$

where $\lambda(\lambda + \gamma)^{-1}$ is an estimate for $P(X < C) = \int_0^t C(t) f(t) dt$ [27]. Every setting was simulated 1000 times and we estimated μ_τ and $Var(\mu_\tau)$ with the δ -method, M-estimation, the Kaplan–Meier method, flexible parametric survival and pseudo-observations with $\tau = 365$.

Table 1 presents the results of the simulation study. The expected $\mu_{365} = 230$ days. Generally, all methods provided unbiased estimates of μ_{365} , with small deviations.

Table 1. Efficacy (*eff*) and 95% confidence interval coverage probabilities of the different Restricted Mean Survival Time (RMST) estimators as a function of sample size (n) and censoring percentage (cp) at restriction time of 365 days, with the expected μ_τ of 230 days.

	δ -Method	M-Estimator	Kaplan–Meier	Flexible Parametric	Pseudo-obs
n = 50, CP = 50%					
$\widehat{\mu}_\tau$	230.1	230.1	230.7	231.2	230.7
$Var(\widehat{\mu}_\tau)$	368.0	362.1	437.5	429.3	446.9
<i>eff</i>	0.98	-	0.82	0.84	0.81
Coverage	0.94	0.94	0.93	0.93	0.94
n = 50, CP = 75%					
$\widehat{\mu}_\tau$	230.2	230.2	230.2	228.0	228.4
$Var(\widehat{\mu}_\tau)$	724.6	712.9	792.1	794.6	1068.7
<i>eff</i>	0.98	-	0.90	0.90	0.66
Coverage	0.93	0.92	0.89	0.89	0.90
n = 100, CP = 50%					
$\widehat{\mu}_\tau$	230.9	230.9	231.4	231.6	231.4
$Var(\widehat{\mu}_\tau)$	185.2	182.2	221.2	217.0	223.4
<i>eff</i>	0.98	-	0.83	0.84	0.81
Coverage	0.96	0.95	0.95	0.95	0.95
n = 100, CP = 75%					
$\widehat{\mu}_\tau$	229.9	229.9	229.4	229.1	229.1
$Var(\widehat{\mu}_\tau)$	367.0	365.1	421.8	406.3	485.6
<i>eff</i>	0.99	-	0.86	0.89	0.75
Coverage	0.95	0.94	0.92	0.94	0.94

Of the 5 tested variance estimators, M-estimation proved to have the smallest estimates. The δ -method variance differed only marginally from the M-estimation. Of the three practically usable methods, the Kaplan–Meier method, pseudo-observations and flexible parametric survival, the later had the highest relative efficacy, irrespective of the setting. To assess if the lower variance estimate of the flexible parametric survival, compared with the Kaplan–Meier method and pseudo-observation affects negatively statistical inference, we assessed the coverage portability of the approximate 95% Wald-type confidence intervals. There is no indication that the lower variance estimate of the flexible parametric survival method negatively affects statistical inference compared to Kaplan–Meier method and pseudo-observation. Although this simulation is not comprehensive we can conclude that the used Wald-type statistics in statistical inference for RMST should be used with caution when sample sizes are small and censoring percentage is high.

To gain further insights of the relative efficacy of the variance estimator a second set of simulation studies was run. First we simulated survival times with $\lambda = 1/365$ and $\gamma = 1/365$ with a sample size of 100 and a censoring percentage of 50%. We varied τ between 50 and 650 days and estimated μ_τ and $Var(\mu_\tau)$ with M-estimation, the Kaplan–Meier method, flexible parametric survival and pseudo-observations.

All estimators returned unbiased μ_τ estimates. As Figure 1a shows, the Kaplan–Meier method, flexible parametric survival and pseudo-observations have lower efficacy compared to parametric estimate of the variance. The estimated increases with longer restriction times and declines after a maximum is reached approximately at 600 days, when around 80 % of the subjects should have experienced the event of interest.

Similarly, we have assessed how the censoring percentage affects efficacy when everything else is kept constant. We simulated survival times with $\lambda = 1/365$, sample size of 100 and set the restriction time at 365 days.

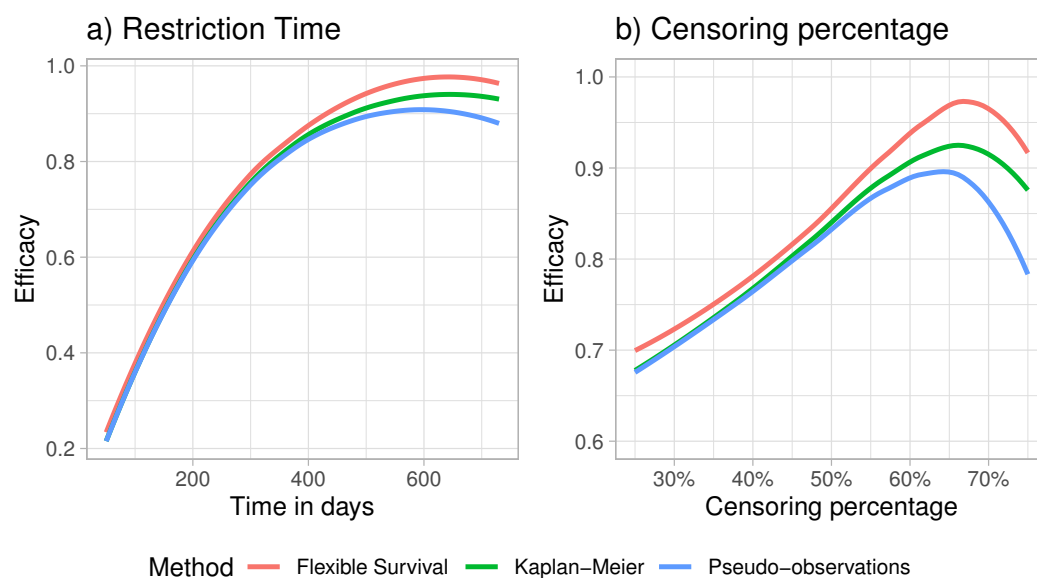


Figure 1. The effect of restriction time (a) and censoring percentage (b) on the efficacy of the three commonly used RMST estimation methods.

We estimated γ , the censoring rate with Equation (5). The pattern was similar to the previous setting. The efficacy increased up to around 65 % censoring rate, thereafter declined again.

4.2. Estimation with Unknown Distribution Function

Practitioners rarely, if ever, know the distribution of survival times, which makes parametric estimation difficult, if not impossible. Here, we aim to assess the precision, efficacy and confidence interval coverage, for the Kaplan–Meier method, pseudo-observations and flexible parametric survival methods. For the former two, distributional assumptions are less important. For flexible parametric survival methods with an assumed Weibull survival function, deviations from the assumptions might affect the outcome. In a first set of simulations, we simulated survival times with the log-logistic distribution, a distribution used in accelerated failure models. The hazard of the log-logistic distribution can take non-monotonic forms depending on the shape parameter. We assumed shape of 1.5 and a scale of 151, giving a mean survival time of approximately 365 days. Thereafter, we simulated survival times with a mixture distribution with $f(t) = 0.5 \exp(1/265) + 0.5 \exp(1/465)$. In both cases we assumed exponential censoring with a rate of $\gamma = 1/365$. Table 2 summarizes the results of the simulation studies with 1000 runs. Generally, flexible parametric survival methods outperform both the Kaplan–Meier method and pseudo-observations without compromising the coverage probability of the 95 % confidence intervals.

Table 2. Efficacy and 95 % confidence interval probabilities of the different RMST estimators at the mean survival time of 365 days, with the expected μ_τ of 182 days for the Log-logistic distribution and 225 days for the exponential distribution.

	Kaplan–Meier	Flexible Parametric	Pseudo-obs
Log-logistic, n = 50			
$\widehat{\mu}_\tau$	183.4	183.9	183.4
$\widehat{Var}(\mu_\tau)$	402.6	397.6	413.9
Coverage	0.93	0.94	0.93
Log-logistic, n = 100			
$\widehat{\mu}_\tau$	182.2	182.7	182.2
$\widehat{Var}(\mu_\tau)$	201.7	199.4	204.2
Coverage	0.94	0.94	0.94
Exp-mix, n = 50			
$\widehat{\mu}_\tau$	224.5	224.6	224.5
$\widehat{Var}(\mu_\tau)$	444.4	435.4	454.3
Coverage	0.95	0.95	0.95
Exp-mix, n = 100			
$\widehat{\mu}_\tau$	225.5	225.7	225.5
$\widehat{Var}(\mu_\tau)$	224.8	220.5	227.1
Coverage	0.96	0.95	0.96

4.3. Parametric Estimation under Model Misspecification

As we have seen in the previous subsection, parametric estimation outperforms non-parametric estimation, at least when the correct model is specified. However, this is rarely possible. In this subsection, we assess the parametric estimators and associated variances under model misspecification. In a first set of simulations, we simulated survival times with the log-logistic distribution as described above (Section 4.2). When the distribution of survival times is unknown, it is common to assume an Exponential distribution. While the Kaplan–Meier method was practically unbiased, the parametric estimator overestimated μ_τ . Of the two variance estimator the δ -method had the smallest estimates. This is expected as sandwich variance estimates are robust against model misspecification at the expense of efficacy (Table 3).

Table 3. Bias and accuracy of the parametric estimation with variance estimated by the δ -method and M-estimation, together with comparative numbers for the Kaplan–Meier method under model misspecification.

	$\widehat{\mu}_\tau$	$\widehat{Var}(\mu_\tau)$	Bias	MSE
$\tau = 50$				
Parametric(δ -met)	45.11	0.35	−1.48	2.54
Parametric (M-est)	45.11	0.39	−1.48	2.58
Kaplan–Meier	46.63	0.91	0.04	0.91
$\tau = 365$				
Parametric (δ -met)	188.24	191.42	6.04	227.96
Parametric (M-est)	188.24	213.10	6.04	249.63
Kaplan–Meier	182.54	203.06	0.35	203.18

See Table 3. In a second set of simulation, we simulated survival times according to a Weibull distribution with shape (k) and scale (λ) parameters. We assumed exponential censoring with rate $1/365$. We varied the scale parameter between 0.5 and 3. The Shape parameter was set so that the mean of Weibull survival times should equal 365, the mean of exponential survival time with rate $1/365$. We estimated λ with numerical root finding, by solving the equation $\lambda\Gamma(1 + k^{-1}) = 365$. At $k = 1$, $\lambda = 365$ and the Weibull distribution simplifies to an Exponential distribution with rate $1/365$. At these values parametric *RMST* estimation assuming Exponential survival distribution returned, as expected, nearly unbiased estimates. However, as k deviated from 1, the bias increased. The sign and magnitude of bias is dependent on the direction and magnitude of deviation from $k = 1$ (Figure 2).

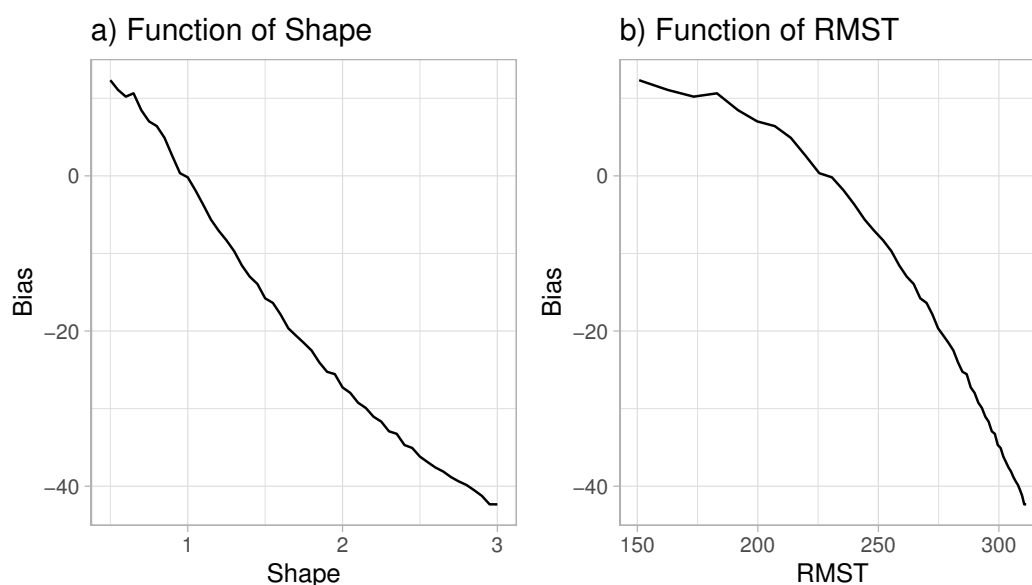


Figure 2. Bias under model misspecification of the parametric estimator as a function of restriction time (a) and censoring percentage (b).

5. Discussion

In this paper we presented the main estimators for *RMST* and associated variances. Additionally, with the help of simulation studies, we examined their small sample properties. As expected, the results of our simulations corroborated that parametric methods of estimation are more effective than non-parametric estimation. This is in line with previous findings about the Kaplan–Meier curve [2] or semi-parametric Cox-regression [3]. Naturally, just as Meier et al. [4] concluded, establishing the proper parametric form is difficult. Since the work of Meier et al., important advancements have been made in the field. The Focused Information Criterion (*FIC*) have been proposed for comparing general non-nested parametric models with a non-parametric alternative [28]. Application of *FIC* requires

advanced statistical/technical knowledge and it is possible that survival times do not follow a known and well-established distribution. Thus, it is likely that in practical applications, parametric estimators for *RMST* will be rare.

The numerical results and the patterns of efficacy as a function of censoring percentage are similar to the ones reported by Meier et al. [4] and by Jullum and Hjort [3]. Independently of the employed method, we could see that a small sample size and a high percentage of censoring, result in genuine departure from nominal coverage values of the approximate Wald-type confidence interval. This departure from the nominal coverage of the 95 % confidence intervals is not a fallacy of the *RMST* and associated variance estimators, more it is a sign of the inappropriateness of Wald-type intervals and statistics. Lawrence et al. [29] described this deviation from the normal distribution in small samples and proposed an approximation to the correct null distribution using the Cornish–Fisher expansion. In medium to large samples the normal approximation is expected to hold. Our simulation suggests that in samples ≥ 100 statistical inference will be correct. This value of ≥ 100 should not be taken as a proposed threshold. This is data dependent. Schall [30] proposed a bootstrap based framework to assess if the coverage probability of confidence intervals approximates the intended nominal coverage. Alternatively, one could either correct null distribution as Lawrence et al. [29] proposed or employ computer intensive methods. Jackknife estimate, that is behind numerous important theoretical results about *RMST* [21,31,32] could be a natural option or bootstrapping that has a straightforward application in survival analysis [33,34]. Nevertheless, bootstrapping *RMST* requires extra care. The parameter space for *RMST* is $(0, \tau]$ and *RMST* can be close to τ . Bootstrapping parameters on the boundary of the parameter space can return inconsistent results [35,36]. This is not a surprising result. While we assume normal distribution for μ_τ , its parameter space is restricted to $(0, \tau]$. A Wald-type 95 % confidence interval would require that 2.5 % of the distribution would be at either side of the interval. As $\mu_\tau \rightarrow \tau$ the right tail of the assumed normal distribution will include more and more values that are larger than τ . Thus the confidence interval might include values outside the parameter space. If $\mu_\tau \sim N(\widehat{\mu}_\tau, \widehat{Var}(\mu_\tau))$, then $P(\mu_\tau \geq \tau)$ determines the values of the μ_τ under the assumed distribution that falls outside of $(0, \tau]$. $P(\mu_\tau \geq \tau)$ is directly influenced by $\widehat{Var}(\mu_\tau)$ and decreases with rate \sqrt{n} as the sample size increases. In either case, further studies are needed to establish routines of statistical inference. Extra care is also needed in inferring from *RMST* based studies with Wald-type confidence intervals or Z-scores. Applied studies, while they do present *RMST* and associated confidence intervals, are interested in comparing different treatment regimens. This is done by calculating difference between two *RMST* estimates \mathcal{D} , or the ratio between them \mathcal{R} . Neither \mathcal{D} or \mathcal{R} is affected by the boundary value problems in realistic situations. The parameter space for \mathcal{D} is $(-\tau, \tau)$ and the \mathcal{D} approaches the boundaries if and only if *RMST* for one arm is close to zero and the *RMST* for the competing treatment should be close to τ . Under similar circumstances \mathcal{R} approaches to its boundary values of $(0, \infty)$.

One aspect that practitioners face when planing a clinical trial with *RMST* as their method of choice is deciding upon τ . The choice of the time window $(0, t]$ may be prespecified at the design stage of the study based on clinical considerations or the choice of time point τ could be data-dependent after the data is collected [37]. The choice of τ has direct implications on statistical power [38] and as we highlighted statistical power can be maximized by considering the intricate relationship between the choice of τ and the censoring percentage.

Although we observed differences in the magnitude of the variance between the Kaplan - Meier method, pseudo-observations and flexible parametric survival methods the practical implications are likely limited. The differences between the variance estimators might be numerically large however when compared to the magnitude of the *RMST* estimate, these differences will have limited practical implications. Thus, we refrain from suggesting one or the other method as the preferred one. All three methods have arguments that spoke for them. The Kaplan- Meier method is probably the most used and best known. Its application is simple and straightforward and require limited programming knowledge to implement on the top of Kaplan–Meier curve estimator, if not readily available in the

preferred software. The flexible parametric survival method takes the idea behind the Kaplan–Meier method one step further by assuming a parametric form and fully define a smoothed version of the survival curve. Estimation assumes definition of the survival distribution. Additionally, to increase flexibility and model fit, the survival function can be modeled as a natural cubic spline function of log time.

Standard statistical software provide readily available routines for estimation [39,40]. An approach similar in philosophy was recently proposed by Luo et al. [41] based on piecewise exponential survival and censoring times. This approach does not require smoothing of the survival curve, however the exponential assumption facilitates design and monitoring of survival trials based on RMST. Both estimation with the Kaplan–Meier method and flexible parametric survival method allows adjustment of the survival curves for covariates. However, interpretation for continuous variables is difficult. The third approach, pseudo-observations, is also readily available in most statistical software [42–44]. With pseudo-observation covariates can be regressed on restricted survival times with linear regression. While the estimates of the Cox-regression or flexible parametric survival regression are hazard rates, regression on pseudo-observation returns regression coefficients on the time scale, i.e. they directly quantify the change in restricted survival times as a function of change in the covariates.

In conclusion, we suggest researchers interested in using RMST to analyse survival data to adopt the method that best fit their purpose. If ease of implementation is paramount, then the Kaplan–Meier method is preferred. To increase statistical power, flexible parametric survival models could be the proper option as they show superior efficacy compared to the other practical applicable methods. Clinical studies that aim to assess the effect of treatments/exposures adjusted for relevant covariates could take benefit from considering pseudo-observations as their tool of preference. Naturally, parametric method of estimation and inference can be considered. However researchers should be extra cautious about violation of assumptions.

Author Contributions: Conceptualization, S.N., E.B. and A.G.; methodology, S.N.; writing—original draft preparation, S.N., E.B. and A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: The authors would like to thank the anonymous referees for their constructive comments and suggestions, that led to improvements in the manuscript.

References

1. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [\[CrossRef\]](#)
2. Miller, R.G., Jr. What price kaplan-meier? *Biometrics* **1983**, 1077–1081. [\[CrossRef\]](#)
3. Jullum, M.; Hjort, N.L. What price semiparametric Cox regression? *Lifetime Data Anal.* **2019**, *25*, 406–438. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Meier, P.; Karrison, T.; Chappell, R.; Xie, H. The price of Kaplan–Meier. *J. Am. Stat. Assoc.* **2004**, *99*, 890–896. [\[CrossRef\]](#)
5. Stute, W. The central limit theorem under random censorship. *Ann. Stat.* **1995**, *23*, 422–439. [\[CrossRef\]](#)
6. Stute, W. Kaplan–meier integrals. *Handbook Stat.* **2003**, *23*, 87–104.
7. Akritas, M.G. The central limit theorem under censoring. *Bernoulli* **2000**, *6*, 1109–1120. [\[CrossRef\]](#)
8. Uno, H.; Wittes, J.; Fu, H.; Solomon, S.D.; Claggett, B.; Tian, L.; Cai, T.; Pfeffer, M.A.; Evans, S.R.; Wei, L.J. Alternatives to hazard ratios for comparing efficacy or safety of therapies in noninferiority studies. *Ann. Intern. Med.* **2015**, *163*, 127. [\[CrossRef\]](#)
9. Uno, H.; Claggett, B.; Tian, L.; Inoue, E.; Gallo, P.; Miyata, T.; Schrag, D.; Takeuchi, M.; Uyama, Y.; Zhao, L.; et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J. Clin. Oncol.* **2014**, *32*, 2380. [\[CrossRef\]](#)

10. Hasegawa, T.; Misawa, S.; Nakagawa, S.; Tanaka, S.; Tanase, T.; Ugai, H.; Wakana, A.; Yodo, Y.; Tsuchiya, S.; Suganami, H.; et al. Restricted mean survival time as a summary measure of time-to-event outcome. *Pharm. Stat.* **2020**. [[CrossRef](#)]
11. Stensrud, M.J.; Aalen, J.M.; Aalen, O.O.; Valberg, M. Limitations of hazard ratios in clinical trials. *Eur. Heart J.* **2019**. [[CrossRef](#)] [[PubMed](#)]
12. Meier, P. Estimation of a distribution function from incomplete observations. *J. Appl. Probab.* **1975**, *12*, 67–87. [[CrossRef](#)]
13. Andersen, P.K.; Klein, J.P.; Rosthøj, S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **2003**, *90*, 15–27. [[CrossRef](#)]
14. Andersen, P.K.; Borgan, O.; Gill, R.D.; Keiding, N. *Statistical Models Based on Counting Processes*; Springer Science & Business Media: Berlin, Germany, 2012.
15. Andersen, P.K.; Pohar Perme, M. Pseudo-observations in survival analysis. *Stat. Methods Med. Res.* **2010**, *19*, 71–99. [[CrossRef](#)]
16. Royston, P.; Parmar, M.K. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat. Med.* **2011**, *30*, 2409–2421. [[CrossRef](#)]
17. Stefanski, L.A.; Boos, D.D. The calculus of M-estimation. *Am. Stat.* **2002**, *56*, 29–38. [[CrossRef](#)]
18. Wang, J.L. Asymptotic Properties of M-estimators Based on Estimating Equations and Censored Data. *Scand. Stat. Theory Appl.* **1999**, *26*, 297–318. [[CrossRef](#)]
19. Boos, D.D.; Stefanski, L. Likelihood Construction and Estimation. In *Essential Statistical Inference*; Springer: Berlin, Germany, 2013; pp. 27–124.
20. Royston, P.; Parmar, M.K. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med. Res. Methodol.* **2013**, *13*, 152. [[CrossRef](#)]
21. Stute, W. The jackknife estimate of variance of a Kaplan–Meier integral. *Ann. Stat.* **1996**, *24*, 2679–2704. [[CrossRef](#)]
22. Royston, P.; Parmar, M.K. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* **2002**, *21*, 2175–2197. [[CrossRef](#)]
23. Irwin, J. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Epidemiol. Infect.* **1949**, *47*, 188–189. [[CrossRef](#)] [[PubMed](#)]
24. Klein, J.P.; Moeschberger, M.L. *Survival Analysis: Techniques for Censored and Truncated Data*; Springer Science & Business Media: Berlin, Germany, 2006.
25. Eaton, A.; Therneau, T.; Le-Rademacher, J. Designing clinical trials with (restricted) mean survival time endpoint: Practical considerations. *Clin. Trials.* **2020**, 1740774520905563. [[CrossRef](#)] [[PubMed](#)]
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
27. Nadarajah, S. Reliability for lifetime distributions. *Math. Comput. Model.* **2003**, *37*, 683–688. [[CrossRef](#)]
28. Jullum, M.; Hjort, N.L. Parametric or nonparametric: The FIC approach. *Stat. Sin.* **2017**, 951–981. [[CrossRef](#)]
29. Lawrence, J.; Qiu, J.; Bai, S.; Hung, H.J. Difference in Restricted Mean Survival Time: Small Sample Distribution and Asymptotic Relative Efficiency. *Stat. Biopharm. Res.* **2019**, *11*, 61–66. [[CrossRef](#)]
30. Schall, R. The empirical coverage of confidence intervals: Point estimates and confidence intervals for confidence levels. *Biom. J.* **2012**, *54*, 537–551. [[CrossRef](#)] [[PubMed](#)]
31. Stute, W.; Wang, J.L. The jackknife estimate of a Kaplan–Meier integral. *Biometrika* **1994**, *81*, 602–606.
32. Azarang, L.; de Uña-Álvarez, J.; Stute, W. The Jackknife estimate of covariance of two Kaplan–Meier integrals with covariables. *Stats* **2015**, *49*, 1005–1025. [[CrossRef](#)]
33. Akritas, M.G. Bootstrapping the Kaplan–Meier Estimator. *J. Am. Stat. Assoc.* **1986**, *81*, 1032–1038.
34. Burr, D. A comparison of certain bootstrap confidence intervals in the Cox model. *J. Am. Stat. Assoc.* **1994**, *89*, 1290–1302. [[CrossRef](#)]
35. Deheuvels, P.; Mason, D.M.; Shorack, G.R. Some results on the influence of extremes on the bootstrap. *Ann. I.H.P. Prob. Stat.* **1993**, *29*, 83–103.
36. Andrews, D.W. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **2000**, *68*, 399–405. [[CrossRef](#)]

37. Tian, L.; Jin, H.; Uno, H.; Lu, Y.; Huang, B.; Anderson, K.M.; Wei, L. On the empirical choice of the time window for restricted mean survival time. *Biometrics* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Huang, B.; Kuan, P.F. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point. *Pharm. Stat.* **2018**, *17*, 202–213. [\[CrossRef\]](#)
39. Jackson, C. flexsurv: A Platform for Parametric Survival Modeling in R. *J. Stat. Softw.* **2016**, *70*, 1–33. doi:10.18637/jss.v070.i08. [\[CrossRef\]](#)
40. Lambert, P.C.; Royston, P. Further development of flexible parametric models for survival analysis. *Stata J.* **2009**, *9*, 265–290. [\[CrossRef\]](#)
41. Luo, X.; Huang, B.; Quan, H. Design and monitoring of survival trials based on restricted mean survival times. *Clin. Trials* **2019**, *16*, 616–625. [\[CrossRef\]](#)
42. Parner, E.T.; Andersen, P.K. Regression analysis of censored data using pseudo-observations. *Stata J.* **2010**, *10*, 408–422. [\[CrossRef\]](#)
43. Overgaard, M.; Andersen, P.K.; Parner, E.T. Regression analysis of censored data using pseudo-observations: An update. *Stata J.* **2015**, *15*, 809–821. [\[CrossRef\]](#)
44. Klein, J.P.; Gerster, M.; Andersen, P.K.; Tarima, S.; Perme, M.P. SAS and R functions to compute pseudo-values for censored data regression. *Comput. Methods Programs Biomed.* **2008**, *89*, 289–300. [\[CrossRef\]](#)



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).