Biometrics -, 1–21 DOI: -
- -

Bayesian Nonparametric Analysis of Restricted Mean Survival Time

Chenyang Zhang and Guosheng Yin*

Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong

**email:* gyin@hku.hk

Summary: The restricted mean survival time (RMST) evaluates the expectation of survival time truncated by a prespecified time point, because the mean survival time in presence of censoring is typically not estimable. The frequentist inference procedure for RMST has been widely advocated for comparison of two survival curves, while research from the Bayesian perspective is rather limited. For the RMST of both right- and interval-censored data, we propose Bayesian nonparametric estimation and inference procedures. By assigning a mixture of Dirichlet processes (MDP) prior to the distribution function, we can estimate the posterior distribution of RMST. We also explore another Bayesian nonparametric approach using the Dirichlet process mixture model and make comparisons with the frequentist nonparametric method. Simulation studies demonstrate that the Bayesian nonparametric RMST under diffuse MDP priors leads to robust estimation and under informative priors it can incorporate prior knowledge into the nonparametric estimator. Analysis of real trial examples demonstrates the flexibility and interpretability of the Bayesian nonparametric RMST for both right- and interval-censored data.

Key words: Bayesian hypothesis testing; Clinical trials; Interval censoring; Mixture of Dirichlet processes; Survival analysis.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Survival data often appear in phase II and phase III clinical trials, where the primary focus is to evaluate the treatment effect of a new therapy in comparison with the standard of care. Model-based approaches have been widely used for quantifying survival benefit due to their capability of incorporating covariate information. However, parametric or semi-parametric models might be problematic if the underlying model assumptions are incorrect. For example, the hazard ratio (HR) is a common choice to assess the survival difference, which relies upon the proportional hazards (PH) assumption. If the PH assumption is violated, the estimated HR may not have a clinically meaningful interpretation (Uno et al., 2014). To mitigate the influence of invalid model assumptions, nonparametric (model-free) estimators have been proposed, such as the t-year survival rate and percentiles of the survival function. However, these estimators mainly focus on local survival information at a particular time point of the survival curve, rather than providing a global summary of survival over time.

Recently, an alternative measure, called the restricted mean survival time (RMST), has attracted much research attention (Royston and Parmar, 2013; Uno et al., 2014). The RMST is computed as the area under the survival curve up to a prespecified time point τ. For right-censored data, RMST can be easily estimated by plugging in the Kaplan–Meier (KM) curve for the survival function. The RMST inherits the robustness property from the KM estimator which is the nonparametric maximum likelihood estimator (NPMLE) for the survival function, and is thus free from any model assumption. More importantly, RMST evaluated at τ has a clinically meaningful interpretation, which is the expected survival time during a τ-period of follow-up. Zhao et al. (2016) proved the asymptotic properties of the estimated RMST and developed the frequentist inference procedure. However, research on the RMST from the Bayesian nonparametric perspective is rather limited.

To fill in the gap, we develop a Bayesian nonparametric approach to estimating the RMST

with either right- or interval-censored data. In particular, we adopt a Bayesian nonparametric model by assigning a mixture of Dirichlet processes (MDP) prior to the distribution function F (Antoniak, 1974), and construct a Gibbs sampler to generate the posterior samples of RMST. The proposed Bayesian MDP-based RMST can be viewed as an interpolation between a frequentist nonparametric estimator and a parametric estimator, which provides a useful tool for Bayesian nonparametric survival analysis.

The rest of this paper is organized as follows. Section 2 reviews the related literature on RMST and Bayesian nonparametric approaches, and Section 3 presents the posterior sampling algorithm for RMST using an MDP prior on F. Section 4 includes simulation studies to compare the performances of the Bayesian MDP-based approach with other existing methods. Two real clinical trial examples are provided in Section 5 to illustrate the applications of the Bayesian nonparametric RMST. Section 6 concludes this paper with some remarks.

2. Background

2.1 *Restricted Mean Survival Time*

The mean survival time is of great interest in clinical studies, while it cannot be estimated accurately due to the presence of censored observations. The restricted mean survival time, RMST(τ), which is calculated as the area under the survival curve up to a certain time point τ (Royston and Parmar, 2013), can be interpreted as the average survival (or event-free) time for patients during the τ-period of follow-up. For right-censored data, we can easily obtain the estimator of RMST(τ) by plugging in the KM curve, $\hat{S}(\cdot)$, which is the NPMLE of the survival function $S(\cdot)$, i.e., $\widehat{\text{RMST}}(\tau) = \int_0^\tau \hat{S}(u)du$. Through integration of the KM estimator over time, RMST can capture the global profile of survival. In contrast, the t-year survival rate or median survival time mainly depends a particular time point or a probability with

no regard on how the survival path reaches that point or how the survival curve behaves past that point.

[Figure 1 about here.]

The RMST can circumvent the weaknesses of the commonly used clinical measures for survival data. The HR and other parametric estimators often rely upon certain model assumptions and they may be inaccurate if the model is misspecified. As shown in Figure 1 which exhibits KM curves for two treatment groups in the CheckMate-057 trial (Borghaei et al., 2015), the median survival time fails to detect clinical effectiveness for treatment with delayed effect, because it ignores the survival information below the median. Sometimes, the percentile survival time is not even available when the event rate is low, i.e., the KM curve cannot drop to the percentile of interest. Computed as integration of the survival curve from 0 to τ, RMST(τ) is always estimable, and it yields a nonparametric and global summary for treatment effect with a transparent clinical interpretation.

Due to the dependence of RMST(τ) on τ, τ should be prespecified with caution, and different choices of τ may lead to different conclusions when comparing two treatments. To maintain the asymptotic validity of the RMST estimation and capture survival information as much as possible, a commonly suggested choice for τ is the largest follow-up time among all right-censored observations (Tian et al., 2020).

Based on the asymptotic properties of $\widehat{\text{RMST}}(\tau)$, Zhao et al. (2016) developed the frequentist inference procedure for RMST with right-censored data. Nowadays, RMST has been widely used in clinical data analysis (Saluja et al., 2019; Angelucci et al., 2020) and trial design (Guimarães et al., 2021). It has also been used for developing the number needed to treat (NNT) for survival endpoints, which has better interpretations than the traditional NNT (Yang and Yin, 2019).

2.2 *Interval Censoring*

Interval censoring often appears when the event time T cannot be exactly observed, but is known to fall inside an interval $(L, R]$. The interval-censored observations are typically represented as a collection of intervals $\{(L_i, R_i]\}_{i=1}^n$. Right censoring can be viewed as a special case of interval censoring if either $(L_i, R_i]$ shrinks to one value with $L_i = R_i$ or $R_i = \infty$.

[Figure 2 about here.]

For interval-censored data, the NPMLE $\hat{S}$ of the survival function S takes the form of probabilities assigned on disjoint intervals, and the behaviors of the survival curve within these disjoint intervals are versatile (Turnbull, 1976). Thus, $\hat{S}$ for interval-censored data is not unique, which only provides the lower and upper bounds displayed as rectangles as shown in Figure 2. A naive approach is to treat interval-censored observations $(L_i, R_i]$ as right-censored (X_i, Δ_i), i.e., $(X_i = R_i, \Delta_i = 1)$ if $R_i < \infty$, and $(X_i = L_i, \Delta_i = 0)$ if $R_i = \infty$. Turnbull (1976) suggested a nonincreasing step function which drops at the left endpoints of those rectangles of $\hat{S}$, i.e., the most pessimistic scenario for the survival probability. Linear smoothing estimation for the survival function has also been considered for interval-censored data (Pan, 2000; Zhang et al., 2020).

2.3 *Bayesian Nonparametric Analysis*

To mitigate limitations of parametric modelling, Ferguson (1973) defined the Dirichlet process (DP) and discussed DP priors on the posterior estimation for nonparametric models. Let $DP(\alpha G_0)$ denote a DP with parameter αG_0 where α is a precision parameter and G_0 is a base probability distribution. Ferguson (1973) showed that if a DP prior is assigned to the distribution function F, $F \sim DP(\alpha G_0)$, and given the observed data $T_1, \ldots, T_n \overset{\text{i.i.d.}}{\sim} F$, then the posterior $F|\{T_1, \ldots, T_n\}$ is again a DP with parameter $\alpha G_0 + \sum_{i=1}^n \delta_{T_i}$, where $\delta_{T_i} = 1$ at point T_i and 0 otherwise (i.e., δ_{T_i} has a point mass at T_i). Therefore, the DP can be utilized

as a conjugate prior on the distribution function F for Bayesian nonparametric problems. Instead of fixing αG_0 in the DP, Antoniak (1974) further proposed a mixture of Dirichlet processes (MDP) by taking αG_0 to be random. Let $\boldsymbol{\Theta}$ be a parameter space, and for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $M_{\boldsymbol{\theta}}$ is a positive real number, $G_{\boldsymbol{\theta}}$ is a probability distribution and $M_{\boldsymbol{\theta}} G_{\boldsymbol{\theta}}$ is the parameter of a Dirichlet process $DP(M_{\boldsymbol{\theta}} G_{\boldsymbol{\theta}})$. Let H be a probability distribution on $\boldsymbol{\Theta}$, and then $\int_{\boldsymbol{\Theta}} DP(M_{\boldsymbol{\theta}} G_{\boldsymbol{\theta}}) dH(\boldsymbol{\theta})$ is an MDP with a mixing distribution H.

The DP prior and its variants have been widely used in Bayesian nonparametric survival analysis. For right-censored data, Susarla and Van Ryzin (1976) assigned a DP prior on the distribution function F and derived the posterior moment estimate of F under the squared error loss. The posterior distribution of F under a DP prior in right-censoring cases was shown to be an MDP (Blum and Susarla, 1977). Johnson and Christensen (1986) generalized the model in Susarla and Van Ryzin (1976) to nested interval-censored data and derived an explicit formula for the posterior estimate of the survival function. Kuo et al. (1992) constructed a Gibbs sampler which iteratively samples the exact event time under the interval-censored constraint and the parameters of interest. Calle and Gómez (2001) obtained the posterior estimate of the survival function for interval-censored data by using a DP prior on F and showed that the Bayesian estimator shrinks the nonparametric estimator to a parametric one. Zhou (2004) derived an explicit formula for the nonparametric Bayesian estimator of the survival function with interval-censored data under a DP prior. Instead of assigning a DP prior to F, Doss (1994) obtained posterior samples for the density function under an MDP prior on F by successive substitution sampling with interval-censored data. Doss and Huffer (2003) constructed Gibbs samplers to approximate the posterior distribution of F under MDP priors, which significantly improved computational efficiency.

Another important tool in Bayesian nonparametric analysis is the Dirichlet process mixture (DPM) model. Given observations $T_1, \ldots, T_n$, the DPM model assumes that $T_i | \boldsymbol{\theta}_i \sim$

$f(\cdot|\boldsymbol{\theta}_i),\ \boldsymbol{\theta}_i \sim G, i = 1, \ldots, n$, and $G \sim DP(\alpha H)$, where α is a precision parameter and H is a base probability distribution of the DP. Escobar (1994) applied the DPM model to estimate a vector of normal means. Escobar and West (1995) illustrated Bayesian density estimation under a DPM model via the predictive distribution of a future observation. Kottas and Gelfand (2001) employed the DPM to model the error distribution under median regression. De Iorio et al. (2009) developed a Bayesian semiparametric survival model based on DPM without the PH assumption. MacEachern and Müller (1998) and Neal (2000) summarized existing Gibbs sampling techniques and proposed new algorithms for nonconjugate DPM models. Gelfand and Kottas (2002) proposed a fully nonparametric approach to sampling from the posterior distribution of general functionals of the mixture distribution.

Under the DPM model, we can obtain the posterior estimates of the survival curve and RMST based on the predictive distribution of a future observation T_{n+1} and its latent parameter $\boldsymbol{\theta}_{n+1}$ (Gelfand and Mukhopadhyay, 1995). Given the k-th posterior sample $(\boldsymbol{\theta}_1^{(k)}, \ldots, \boldsymbol{\theta}_n^{(k)})$, one can draw a posterior sample of G from $DP(\sum_{i=1}^{n} \delta_{\boldsymbol{\theta}_i^{(k)}} + \alpha H)$. A posterior sample of RMST is a linear functional of $F(\cdot|G)$, which can be obtained using the parametric model $f(\cdot|\boldsymbol{\theta})$ and a realization of $G|$Data (Gelfand and Kottas, 2002). Compared with the MDP approach, DPM provides a smoother distribution estimate because its posterior estimator has a form of a distribution over some mixture models, which has been widely used in density estimation and clustering problems (Escobar, 1994; Escobar and West, 1995; MacEachern and Müller, 1998). Nevertheless, the DPM estimator is relatively sensitive to the specification of the parametric model $f(\cdot|\boldsymbol{\theta})$, which may compromise its performance in modeling nonparametric measures from the Bayesian perspective.

By assigning an MDP prior to the distribution function F, we can directly obtain a realization from the posterior distribution of F at each iteration of the Gibbs sampler without extra operations (Doss and Huffer, 2003). The distribution function F under an MDP prior

centers on a family of parametric distribution $G_{\boldsymbol{\theta}}$ with a prior $\boldsymbol{\theta} \sim H(\boldsymbol{\theta})$. When $M_{\boldsymbol{\theta}} \to \infty$ for all $\boldsymbol{\theta}$, the base measure $G_{\boldsymbol{\theta}}$ of MDP dominates the model and the posterior estimate of F converges to the Bayesian parametric model under the distribution $G_{\boldsymbol{\theta}}$ and the prior $\boldsymbol{\theta} \sim H(\boldsymbol{\theta})$. When $M_{\boldsymbol{\theta}} \to 0$ for all $\boldsymbol{\theta}$, the nonparametric component dominates the model, and the posterior of F converges to the NPMLE. As a result, the Bayesian nonparametric method with an MDP prior on F takes a middle-ground approach between parametric and nonparametric models.

3. Methodology

For interval-censored data, we introduce a sampling algorithm for drawing posterior samples of RMST. At each iteration, a posterior sample of the distribution function F is generated by a Gibbs sampler (Doss and Huffer, 2003), and then the posterior sample of RMST(τ) can be calculated as the area under the survival curve from 0 to τ using the trapezoid rule.

For simplicity, we assume $M_{\boldsymbol{\theta}} \equiv M$ and $F \sim \int_{\boldsymbol{\Theta}} DP(MG_{\boldsymbol{\theta}})dH(\boldsymbol{\theta})$. Given an MDP prior on F and the observations $\boldsymbol{T} = (T_1, \ldots, T_n)$, Antoniak (1974) proved that the posterior distribution of F has the form,

$$F|\boldsymbol{T} \sim \int_{\boldsymbol{\Theta}} DP\left(MG_{\boldsymbol{\theta}} + \sum_{i=1}^{n} \delta_{T_i}\right) dH(\boldsymbol{\theta}|\boldsymbol{T}),$$

$$dH(\boldsymbol{\theta}|\boldsymbol{T}) = c(\boldsymbol{T}) \prod_{i=1}^{m(\boldsymbol{T})} dG_{\boldsymbol{\theta}}(Y_i)\; dH(\boldsymbol{\theta}), \tag{3.1}$$

where $m(\boldsymbol{T})$ is the number of distinct values of $\boldsymbol{T}$, $\boldsymbol{Y} = \{Y_1, \ldots, Y_{m(\boldsymbol{T})}\}$ are distinct values of $\boldsymbol{T}$, and $c(\boldsymbol{T})$ is a normalization constant.

Due to interval censoring, the exact event time T_i is typically unavailable and we only know that T_i falls inside an interval $A_i = (L_i, R_i]$. Thus, the observed data are composed of $\{A_i : T_i \in A_i = (L_i, R_i], i = 1, \ldots, n\}$. When a DP prior is assigned to F, the Pólya urn scheme (Blackwell and MacQueen, 1973) can be used to sample from the conditional

distribution $\boldsymbol{T}|\boldsymbol{\theta}, \{A_i\}_{i=1}^n$. Based on the posterior distribution $\boldsymbol{\theta}|\boldsymbol{T}$ in (3.1) and the Pólya urn scheme, a Gibbs sampler can be constructed, for which the k-th iteration proceeds as follows:

$$
\begin{aligned}
\boldsymbol{\theta}^{(k)} \sim & H(\boldsymbol{\theta}|\boldsymbol{T}^{(k-1)}), \\
T_i^{(k)} \sim & \left(MG_{\boldsymbol{\theta}^{(k)}} + \sum_{j<i} \delta_{T_j^{(k)}} + \sum_{j>i} \delta_{T_j^{(k-1)}} \right)_{A_i}, \quad i = 1, \ldots, n, \quad \text{(Pólya urn scheme)} \\
F^{(k)} \sim & DP\left(MG_{\boldsymbol{\theta}^{(k)}} + \sum_{i=1}^n \delta_{T_i^{(k)}} \right),
\end{aligned} \tag{3.2}
$$

where $(\alpha)_A$ denotes a truncated measure by restricting α in the set A. We can further compute RMST(τ) by the trapezoid rule,

$$
\begin{aligned}
\mathrm{RMST}^{(k)}(\tau) \approx & \sum_{j=1}^{L} \frac{S^{(k)}(\tau_j) + S^{(k)}(\tau_{j-1})}{2}(\tau_j - \tau_{j-1}), \\
S^{(k)}(\tau_j) = & 1 - \sum_{l=1}^{j} F^{(k)}((\tau_{l-1}, \tau_l]), \\
\left\{ F^{(k)}((\tau_{j-1}, \tau_j]) \right\}_{j=1}^{L+1} \sim & \text{Dirichlet}\left(p_1^{(k)}, \ldots, p_{L+1}^{(k)} \right), \\
p_j^{(k)} = & MG_{\boldsymbol{\theta}^{(k)}}((\tau_{j-1}, \tau_j]) + \sum_{i=1}^{n} \delta_{T_i^{(k)}}((\tau_{j-1}, \tau_j]), \ j = 1, \ldots, L+1,
\end{aligned}
$$

where $0 = \tau_0 < \tau_1 < \cdots < \tau_L = \tau < \tau_{L+1} = \infty$ are time grids for the trapezoid method, $F((a, b]) = P_{X \sim F}(a < X \leqslant b)$ and $\delta_x((a, b]) = I(a < x \leqslant b)$.

We conduct an extra step to reallocate cluster locations every n_{extra} iterations in order to improve the mixing of the chain and the exploration of the posterior cluster structure (Doss and Huffer, 2003; MacEachern, 1994; Bush and MacEachern, 1996). In practical applications, we set $n_{\text{extra}} = n_{\text{thin}}$, the thinning interval. The entire sampling procedure is detailed in Algorithm 1.

Bayesian nonparametric estimation for the distribution function F based on an MDP prior can be viewed as an interpolation between the nonparametric and parametric estimates (Doss, 1994; Doss and Huffer, 2003). The parametric part is formulated by $H(\boldsymbol{\theta})$ and $G_{\boldsymbol{\theta}}$, and

Algorithm 1 Gibbs sampler for Bayesian nonparametric MDP-based RMST

Prior: $M, G_{\boldsymbol{\theta}}$ and $H(\boldsymbol{\theta})$ for the mixture of Dirichlet processes prior on F.
Observations: $\{A_i = (L_i, R_i],\ \text{for } i = 1, \ldots, n\}$.
Settings: N (number of posterior samples); n_{extra} (reallocation frequency to update cluster centers); n_{thin} (thinning interval); L (number of points for the trapezoid rule).
Initialization: $\boldsymbol{T}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$.

$\tau_j \leftarrow j\tau/L\ , j = 1, \ldots, L$; $\tau_{L+1} \leftarrow \infty$
for $k = 1, \ldots, Nn_{\text{thin}}$ **do**
  Generate $\boldsymbol{\theta}^{(k)} \sim H(\boldsymbol{\theta}|\boldsymbol{T}^{(k-1)})$
  Generate $T_i^{(k)} \sim \left(MG_{\boldsymbol{\theta}^{(k)}} + \sum_{j<i}\delta_{T_j^{(k)}} + \sum_{j>i}\delta_{T_j^{(k-1)}}\right)_{A_i},\ i = 1, \ldots, n$
  if (k mod n_{extra}) $= 0$ (modulo operation) **then**
    $\{Y_j^{(k)}\}_{j=1}^{m(\boldsymbol{T}^{(k)})} \leftarrow$ distinct values of $\boldsymbol{T}^{(k)}$
    Obtain the subindices $(c_1, \ldots, c_n)$ so that $T_i^{(k)} = Y_{c_i}^{(k)}$
    for $j = 1, \ldots, m(\boldsymbol{T}^{(k)})$ **do**
      $B_j \leftarrow \cap_{i:c_i=j} A_i$
      Generate $Y_j^{(k)} \sim \left(MG_{\boldsymbol{\theta}^{(k)}}\right)_{B_j}$
    end for
    $T_i^{(k)} \leftarrow Y_{c_i}^{(k)},\ i = 1, \ldots, n$
  end if } Extra step
  $p_j^{(k)} \leftarrow MG_{\boldsymbol{\theta}^{(k)}}((\tau_{j-1}, \tau_j]) + \sum_{i=1}^{n}\delta_{T_i^{(k)}}((\tau_{j-1}, \tau_j]),\ j = 1, \ldots, L+1$
  Generate $\{F^{(k)}((\tau_{j-1}, \tau_j])\}_{j=1}^{L+1} \sim \text{Dirichlet}(p_1^{(k)}, \ldots, p_{L+1}^{(k)})$
  Calculate $\text{RMST}^{(k)}$ by the trapezoid rule
end for
Output: $\text{RMST}^{(k)},\ k = mn_{\text{thin}},\ m = 1, \ldots, N$

M controls the proportions of nonparametric and parametric contributions in the posterior estimation. One can specify $G_{\boldsymbol{\theta}}$ as some parametric survival model and $H(\boldsymbol{\theta})$ as a conjugate prior, e.g., for a scalar parameter θ we take G_θ to be $\text{Exp}(\theta)$ and $H(\theta)$ to be $\text{Gamma}(a_0, b_0)$. To incorporate prior information, $H(\boldsymbol{\theta})$ and $G_{\boldsymbol{\theta}}$ can be estimated from historical data or elicited from experts' opinion.

4. Simulation Studies

In the simulation studies, we compare the proposed Bayesian nonparametric RMST using the MDP prior with that using the DPM prior, as well as Bayesian parametric and frequentist nonparametric methods. We investigate RMST estimation for both right- and interval-

censored data by evaluating the posterior mean of RMST and coverage probability (CP) of the equal-tailed credible interval (CrI) for Bayesian approaches.

4.1 *Simulation Settings*

We consider both exponential and log-normal distributions for the base measure to examine the influence of model flexibility on RMST estimation. For the Bayesian MDP approach with an exponential base distribution, we assume that G_θ is a family of exponential distributions with rate parameter θ and the prior H for θ is Gamma(a_0, b_0). Under the log-normal base distribution, we set $G_{(\mu,\xi)}$ to be Log-Normal(μ, ξ) where ξ is a precision parameter, and take a normal-gamma mixing distribution for $H(\mu, \xi)$, i.e., $(\mu, \xi) \sim$ Normal-Gamma$(\mu_0, \lambda_0, a_0, b_0)$.

For the DPM model, we consider (i) the exponential mixture model, $T_i|\theta_i \sim \text{Exp}(\theta_i), \theta_i \sim G, G \sim DP(\alpha H)$, where H is a Gamma(a_0, b_0) distribution; and (ii) the log-normal mixture model, $T_i|\mu_i, \xi_i \sim$ Log-Normal$(\mu_i, \xi_i), (\mu_i, \xi_i) \sim G, G \sim DP(\alpha H)$, where H is a Normal-Gamma$(\mu_0, \lambda_0, c_0, d_0)$ distribution. We construct a Gibbs sampler to obtain the posterior samples of RMST(τ) by iteratively generating the exact event time T_i given the interval constraint A_i and sampling the parameter of interest (Kuo et al., 1992). The detailed sampling algorithm for the DPM approach is given in Web Appendix A.

Under the Bayesian parametric approach, we assume $T_i|\theta \sim \text{Exp}(\theta), i = 1, \ldots, n$ with a gamma prior $\theta \sim$ Gamma(a_0, b_0), as well as $T_i|\mu, \xi \sim$ Log-Normal(μ, ξ) with a normal-gamma prior $(\mu, \xi) \sim$ Normal-Gamma$(\mu_0, \lambda_0, c_0, d_0)$. For interval-censored data $\{A_i\}_{i=1}^n$, we apply a Gibbs sampler (Kuo et al., 1992) to obtain the posterior samples of θ, which iterates between sampling from $T_i|A_i, \theta$ and $\theta|\{T_i, i = 1, \ldots, n\}$. The posterior samples of RMST(τ) can be computed by RMST$(\tau|\theta) = (1 - e^{-\tau\theta})/\theta$.

For both right- and interval-censored cases, observed data are respectively simulated from Weibull, piecewise exponential and log-normal distributions with moderate ($n = 100$) and small ($n = 20$) sample sizes. Two piecewise exponential distributions are considered: (A)

$\lambda(t) = 0.5, t \in [0,1]; \lambda(t) = 1, t \in (1,\infty)$; (B) $\lambda(t) = 1, t \in [0,1]; \lambda(t) = 0.5, t \in (1,\infty)$. The study end time is set as 2 for all simulations. We adopt diffuse priors with hyper-parameters $(a_0, b_0) = (0.01, 0.01)$ under the exponential base measure, $(\mu_0, \lambda_0, c_0, d_0) = (0, 0.01, 0.01, 0.01)$ under the log-normal base measure, and $M = 10^{-6}$ for the Bayesian MDP approach. For the Bayesian DPM approach, we take $\alpha \sim \text{Gamma}(0.01, 0.01)$.

For right-censored data, we assume that the censoring time follows an exponential distribution, under which the rate parameter is adjusted to yield a censoring rate of 40%. As suggested by Tian et al. (2020), τ is set to be the maximum observed time in each simulation.

For interval-censored data, each patient is monitored by every 0.2 time interval to mimic the pattern of real clinical trials in which patients are examined periodically. The first examination time follows Uniform$(0, 0.2)$, and patients may miss any of the examinations with probability $p_{\text{dropout}} = 0.2$ due to loss of the follow-up. The interval observation of a patient is recorded as the shortest interval covering the exact event time T_i. In each simulation, we set τ as the maximum value of $\{L_i, i = 1, \ldots, n\}$. We evaluate the performances of RMST estimation using the MDP, DPM, Bayesian parametric, naive (which treats interval-censored observations as right-censored ones) and linear smoothing methods (Zhang et al., 2020). The linear smoothing RMST can be viewed as a frequentist nonparametric estimator for interval-censored cases.

[Table 1 about here.]

[Table 2 about here.]

4.2 *One-sample Simulations*

Tables 1 and 2 present the simulation results based on 5000 replications for the right- and interval-censored RMST estimation under diffuse priors, respectively. With relatively large sample size $n = 100$, both frequentist and Bayesian MDP approaches can deliver accurate point estimates and CPs in both right- and interval-censoring cases. When sample size is as

small as $n = 20$, the Bayesian MDP-based RMST outperforms the frequentist counterpart in all scenarios in terms of CPs, because the frequentist confidence interval is constructed based on the asymptotic property. The posterior estimates of RMST under the Bayesian parametric and nonparametric DPM methods depend on the fit of the specified model. When the model is correct, both approaches can provide accurate RMST estimates and the CPs are close to the nominal level 95%. However, when the model is misspecified, the posterior RMST estimates under Bayesian parametric and DPM approaches with diffuse priors are biased with inaccurate CPs. The exponential base measure in the DPM is sensitive to model misspecification and thus the performances of Bayesian nonparametric DPM and parametric approaches are poor when the model is misspecified. The log-normal distribution has two parameters, which is thus more flexible and can mitigate the influence of model misspecification. Nevertheless, the Bayesian nonparametric DPM still yields inappropriate interval estimates with CPs larger than 0.96 when the assumed model does not match the true event time distribution and sample size is $n = 100$. With smaller sample size $n = 20$, CPs decrease under all approaches and the Bayesian nonparametric DPM method shows slightly better performances than the Bayesian MDP-based estimator. The RMST under the DPM model has the form $\text{RMST}(\tau|G) = \int \text{RMST}(\tau|\theta)dG(\theta)$, for which the specified distribution of the event time plays an important role. By taking the most optimistic treatment for interval-censored observations, the naive RMST estimator is substantially larger than the true value in all scenarios, and the coverage probabilities are far below the nominal level.

In addition, we evaluate the performances of the Bayesian nonparametric and parametric methods under informative priors to examine the influence of prior information on the corresponding RMST estimation. Simulation results summarized in Web Tables 1 to 4 of Web Appendix B demonstrate that the Bayesian MDP-based RMST is more robust to misspecified prior information and all Bayesian methods would benefit from appropriate

priors and yield smaller mean squared errors (MSEs) in comparison with the frequentist NPMLE.

4.3 *Two-sample Simulations*

In the comparison of two survival curves, a one-sided hypothesis test can be constructed using RMST,

$$H_0 : \text{RMST}_1(\tau) \leqslant \text{RMST}_2(\tau) \quad \text{versus} \quad H_1 : \text{RMST}_1(\tau) > \text{RMST}_2(\tau).$$

Under Bayesian approaches, we calculate the posterior probability of the null hypothesis, $P(H_0|\text{Data})$, which can be approximated using the posterior samples of $\text{RMST}_j(\tau),\ j = 1, 2$. The null hypothesis would be rejected under the significance level α if $P(H_0|\text{Data}) \leqslant \alpha$ (Shi and Yin, 2020).

[Table 3 about here.]

[Table 4 about here.]

Simulation results on two-sample comparison under diffuse priors based on 1000 replications are given in Tables 3 and 4 for right- and interval-censored data, respectively. Clearly, the Bayesian MDP and frequentist nonparametric approaches can maintain the test size at the nominal level 5% in all scenarios, while the Bayesian DPM and parametric methods tend to be more conservative. In terms of power, with both medium ($n = 100$ per group) and small ($n = 20$ per group) sample sizes, the Bayesian MDP and frequentist nonparametric approaches deliver similar performances, while the Bayesian DPM and parametric methods are less likely to reject H_0 under H_1, especially when the model is misspecified.

5. Examples

Two real clinical studies with right- and interval-censored data respectively are used to illustrate the application of our proposed Bayesian nonparametric RMST. We compare the

estimated survival curves using the Bayesian MDP and DPM methods under the log-normal model using diffuse priors with the frequentist nonparametric counterparts. The Bayesian MDP-based RMST is evaluated at several time points to examine the influence of time τ on RMST estimation and two-sample Bayesian hypothesis tests. We assign the same MDP prior (M, H, G) to the two treatment groups and set the prior probabilities of H_0 and H_1 to be equal, i.e., $P(H_0) = P(H_1) = 1/2$. The Bayes factor in favor of H_1 over H_0 is used for Bayesian hypothesis testing, which is equal to the posterior odds $P(H_1|\text{Data})/P(H_0|\text{Data})$.

5.1 *CheckMate-057 Trial*

The CheckMate-057 trial (Borghaei et al., 2015) was conducted to evaluate the treatment efficacy of nivolumab versus docetaxel for patients with advanced nonsquamous non-small-cell lung cancer. The sample sizes for the nivolumab and docetaxel groups are 292 and 290, respectively. Figure 1(a) shows the overall survival (OS) curves for the two arms. Borghaei et al. (2015) reported the superiority of nivolumab over docetaxel in OS with an HR of 0.73 (95% CI [0.59, 0.89]). During the first seven months, patients receiving docetaxel had higher survival rates than those receiving nivolumab. After the crossing of survival curves at month 7, the situation reversed till the end of the study. We plot the estimated log(HR) over time in Figure 1(b). The PH test based on weighted residuals (Grambsch and Therneau, 1994) indicated a violation of the PH assumption ($p = 0.001$). Several Bayesian nonparametric methods have been proposed to deal with the non-PH cases (De Iorio et al., 2009).

We take $\tau = 24$ months to investigate the treatment effect on preventing the occurrence of death during the two-year follow-up. Based on the posterior samples obtained by the Bayesian MDP-based approach, the posterior estimate of 24-month RMST for the nivolumab and docetaxel groups are 12.91 (95% CrI [11.90, 13.92]) and 11.17 (95% CrI [10.30, 12.07]) months, respectively. We observe an RMST difference of 1.74 months with the 95% CrI [0.39, 3.10] between the two groups, which indicates that during the two-year follow-up

period, patients receiving nivolumab on average enjoyed an extra survival gain of 1.74 months compared with those receiving docetaxel. With the nivolumab and docetaxel arms as groups 1 and 2 in the one-sided hypothesis test, the estimated $P(H_0|\text{Data})$ is 0.006, which is identical to the one-sided frequentist p-value. The posterior odds favoring H_1 over H_0 is approximately 171, which provides decisive evidence against H_0.

To evaluate the influence of the choice of τ on the RMST estimation, we also examine the RMST at month 6. During the first six months, the survival curve of the docetaxel group is above that of the nivolumab group and the estimated posterior mean of the RMST difference is -0.18 (95% CrI [-0.44, 0.09]) months between the nivolumab and docetaxel groups. Furthermore, we obtain $P(H_0|\text{Data}) = 0.9$ and a Bayes factor of 0.11 favoring H_1 over H_0.

5.2 *Breast Cosmesis Study*

The Breast Cosmesis Study (BCOS) dataset (Finkelstein and Wolfe, 1985) is collected from a retrospective study on the time to cosmetic deterioration of the breast to compare the treatment benefit between the addition of adjuvant chemotherapy to the radiation therapy (RCT, 48 patients) and the radiation therapy (RT, 46 patients) alone. Patients were required to visit the clinic every 4 to 6 months, leading to interval-censored observations. The NPMLE of survival curves for the two groups are exhibited in Figure 2(a) with solid lines, which are displayed in rectangles within some intervals due to the ambiguities on the estimation of survival curves caused by interval censoring. We also plot the posterior means of survival probabilities using the Bayesian nonparametric MDP and DPM approaches under diffuse priors.

For the BCOS dataset, we choose $\tau = 46$ months which is the minimum value of the largest left endpoints for the interval censored observations in the RT and RCT groups. The Bayesian nonparametric MDP approach yields the posterior mean of RMST(46)=32.7 for the RT group

with the 95% CrI $[27.9, 37.1]$ and 24.1 for the RCT group with the 95% CrI $[20.6, 27.9]$. The difference in RMST(46) between the two groups is 8.6 months, which indicates that patients in the RT group enjoyed on average additional 8.6 months free of cosmetic deterioration during the 46-month follow-up compared with those in the RCT group. As shown in Figure 2(b), the two histograms of RMST are quite separated, indicating overwhelming evidence that RT provided more treatment benefit than RCT for patients. For the hypothesis test of $H_0 : \text{RMST}_{\text{RT}}(46) \leqslant \text{RMST}_{\text{RCT}}(46)$ versus $H_1 : \text{RMST}_{\text{RT}}(46) > \text{RMST}_{\text{RCT}}(46)$, the posterior probability of the null hypothesis is 0.003, which is close to the one-sided p-value $= 0.001$ obtained by the linear smoothing RMST inference (Zhang et al., 2020). The Bayes factor of 341 for H_1 over H_0 provides decisive evidence in favor of the alternative hypothesis.

We also consider two additional time points $\tau = 35$ and 25 months, under which histograms for the posterior RMST samples are shown in Figure 2(c) and (d) respectively. With a smaller value of τ, the histograms of the posterior RMST samples for two treatments are largely overlapped. From Figure 2(a), two survival curves cross at 17 months, and the survival curve of the RT group stays above that of the RCT group during the remaining follow-up till 46 months. Therefore, the RMST, computed as the integration of the survival function from 0 to τ, would accumulate more evidence in favor of H_1 as τ increases.

6. Discussion

As a nonparametric metric for quantifying treatment effect, the RMST is model-free and robust, and further it has a clinically meaningful interpretation. We develop the estimation and inference procedure for the RMST in the Bayesian nonparametric framework using an MDP prior on the distribution function F. A Gibbs sampler is constructed to generate posterior samples of the RMST. Not only does the proposed method perform well for right-censored data but it also fits to interval-censored scenarios. Simulation results show that under diffuse priors our Bayesian MDP-based RMST can produce consistent estimates even

with small sample size and the corresponding hypothesis test can maintain the test size and yield higher power than that using the Bayesian DPM method. Compared with the Bayesian DPM approach, the MDP-based RMST is less sensitive to the parametric model setting. Additional studies shown in the Web Appendix B demonstrate that the Bayesian nonparametric MDP approach can provide better RMST estimation with smaller MSEs by incorporating accurate prior information and it can mitigate the influence of misinformative prior beliefs due to the contribution from the nonparametric component.

Selection of the time point τ is critical in RMST estimation. In the frequentist paradigm, the largest follow-up time is a common choice for τ (Tian et al., 2020), while such τ is data-dependent. Under the Bayesian nonparametric MDP approach, we can obtain posterior samples of RMST for any τ, because the parametric model component can provide 'imputation' beyond the largest follow-up time. One should specify τ from a clinical or scientific perspective according to the requirements of the trial. Moreover, our Bayesian RMST can also be utilized as a tool in clinical trial design to substitute for the t-year survival rate or the median survival time.

Acknowledgements

We thank the three referees, the Associate Editor and Editor for their many constructive and insightful comments that have led to significant improvements in this paper. This research was partially supported by funding from the Research Grants Council of Hong Kong (17308420).

Data Availability Statement

The data used in this paper are available from the authors upon request.

References

Angelucci, E., Li, J., Greenberg, P., Wu, D., Hou, M., Montaňo Figueroa, E. H., Rodriguez, M. G., Dong, X., Ghosh, J., Izquierdo, M., et al. (2020). Iron chelation in transfusion-dependent patients with low-to intermediate-1–risk myelodysplastic syndromes: A randomized trial. *Annals of Internal Medicine* **172,** 513–522.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* **2,** 1152–1174.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1,** 353–355.

Blum, J. and Susarla, V. (1977). On the posterior distribution of a Dirichlet process given randomly right censored observations. *Stochastic Processes and their Applications* **5,** 207–211.

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E., Chow, L. Q., Vokes, E. E., Felip, E., Holgado, E., et al. (2015). Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *New England Journal of Medicine* **373,** 1627–1639.

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83,** 275–285.

Calle, M. L. and Gómez, G. (2001). Nonparametric Bayesian estimation from interval-censored data using Monte Carlo methods. *Journal of Statistical Planning and Inference* **98,** 73–87.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* **65,** 762–771.

Doss, H. (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *The Annals of Statistics* **22,** 1763–1786.

Doss, H. and Huffer, F. W. (2003). Monte Carlo methods for Bayesian analysis of survival data using mixtures of Dirichlet process priors. *Journal of Computational and Graphical Statistics* **12,** 282–307.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89,** 268–277.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90,** 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1,** 209–230.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41,** 933–945.

Gelfand, A. E. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **11,** 289–305.

Gelfand, A. E. and Mukhopadhyay, S. (1995). On nonparametric Bayesian inference for the distribution of a random sample. *Canadian Journal of Statistics* **23,** 411–420.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81,** 515–526.

Guimarães, H. P., e Silva, P. G. d. B., Liporace, I. L., Sampaio, R. O., Tarasoutchi, F., Paixão, M., Hoffmann-Filho, C. R., Patriota, R., Leiria, T. L., Lamprea, D., et al. (2021). A randomized clinical trial to evaluate the efficacy and safety of rivaroxaban in patients with bioprosthetic mitral valve and atrial fibrillation or flutter: Rationale and design of the RIVER trial. *American Heart Journal* **231,** 128–136.

Johnson, W. and Christensen, R. (1986). Bayesian nonparametric survival analysis for grouped data. *Canadian Journal of Statistics* **14,** 307–314.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* **96,** 1458–1468.

Kuo, L., Smith, A. F., MacEachern, S., and West, M. (1992). Bayesian computations in survival models via the Gibbs sampler. In *Survival Analysis: State of the Art*, pages 11–24. Springer.

MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics-Simulation and Computation* **23,** 727–741.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7,** 223–238.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9,** 249–265.

Pan, W. (2000). Smooth estimation of the survival function for interval censored data. *Statistics in Medicine* **19,** 2611–2624.

Royston, P. and Parmar, M. K. (2013). 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 Medical Research Methodology* **13,** 152.

Saluja, R., Everest, L., Cheng, S., Cheung, M., and Chan, K. K. (2019). Assessment of whether the american society of clinical oncologys value framework and the european society for medical oncologys magnitude of clinical benefit scale measure absolute or relative clinical survival benefit: An analysis of randomized clinical trials. *JAMA Oncology* **5,** 1188–1194.

Shi, H. and Yin, G. (2020). Reconnecting p-value and posterior probability under one- and two-sided tests. *The American Statistician* pages 1–11.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association* **71,** 897–

902.

Tian, L., Jin, H., Uno, H., Lu, Y., Huang, B., Anderson, K. M., and Wei, L. (2020). On the empirical choice of the time window for restricted mean survival time. *Biometrics* doi:10.1111/biom.1323.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)* **38,** 290–295.

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L., et al. (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32,** 2380–2385.

Yang, Z. and Yin, G. (2019). An alternative approach for estimating the number needed to treat for survival endpoints. *PLOS ONE* **14,** 1–14.

Zhang, C., Wu, Y., and Yin, G. (2020). Restricted mean survival time for interval-censored data. *Statistics in Medicine* **39,** 3879–3895.

Zhao, L., Claggett, B., Tian, L., Uno, H., Pfeffer, M. A., Solomon, S. D., Trippa, L., and Wei, L. (2016). On the restricted mean survival time curve in survival analysis. *Biometrics* **72,** 215–221.

Zhou, M. (2004). Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Statistica Sinica* **14,** 533–546.

Supporting Information

Web Appendices, Tables referenced in Section 4, and R codes to reproduce numerical results are available with this paper at the Biometrics website on Wiley Online Library.

Received . *Revised* . *Accepted* .

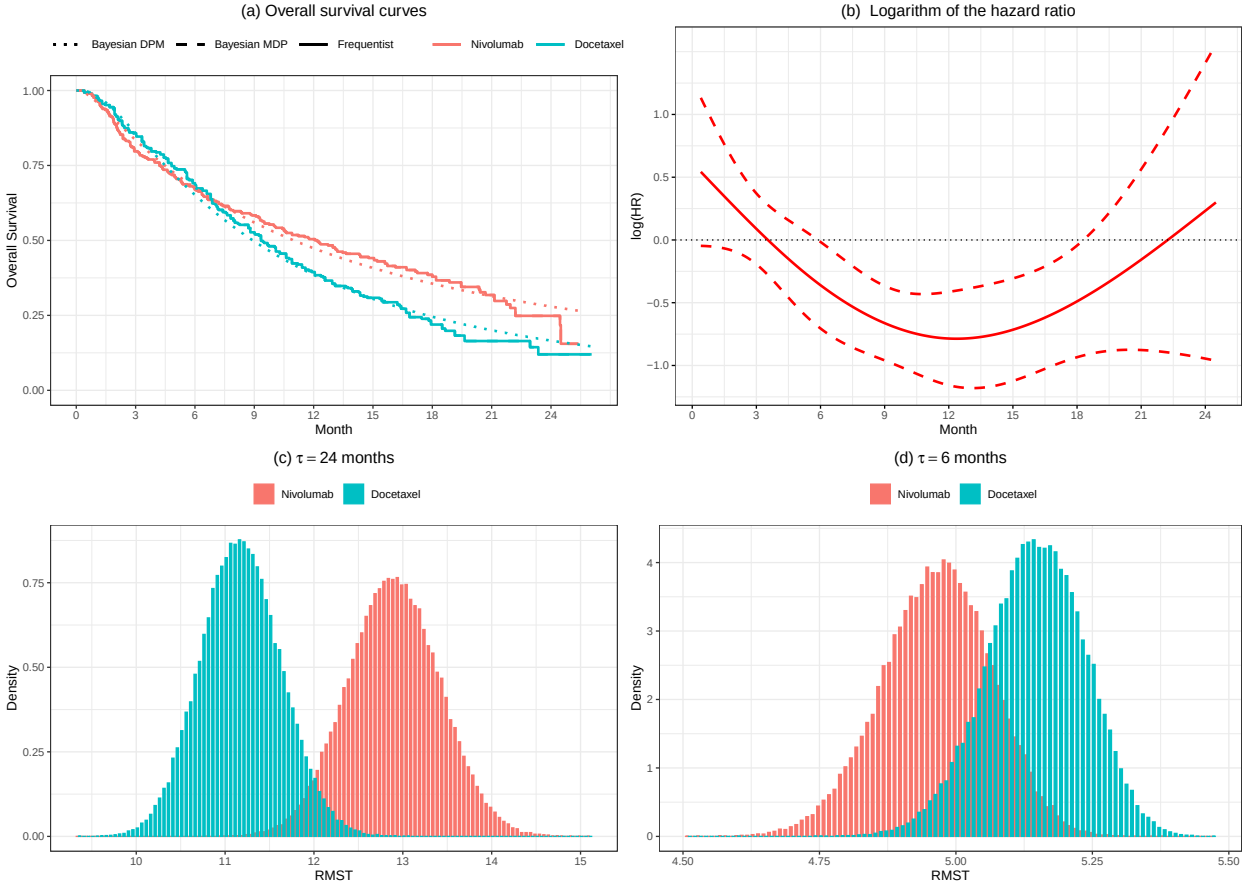

Figure 1: (a) Kaplan–Meier estimators (solid lines), posterior means using Bayesian MDP (dashed lines, but completely overlapped with Kaplan–Meier curves) and DPM (dotted lines) for overall survival of the CheckMate-057 trial (Borghaei et al., 2015). (b) Logarithm of the hazard ratio (HR) (nivolumab versus docetaxel) for overall survival. Histograms of the posterior RMST samples for the docetaxel (green) and nivolumab (red) groups under a diffuse MDP prior evaluated at (c) $\tau = 24$; (d) $\tau = 6$ months.

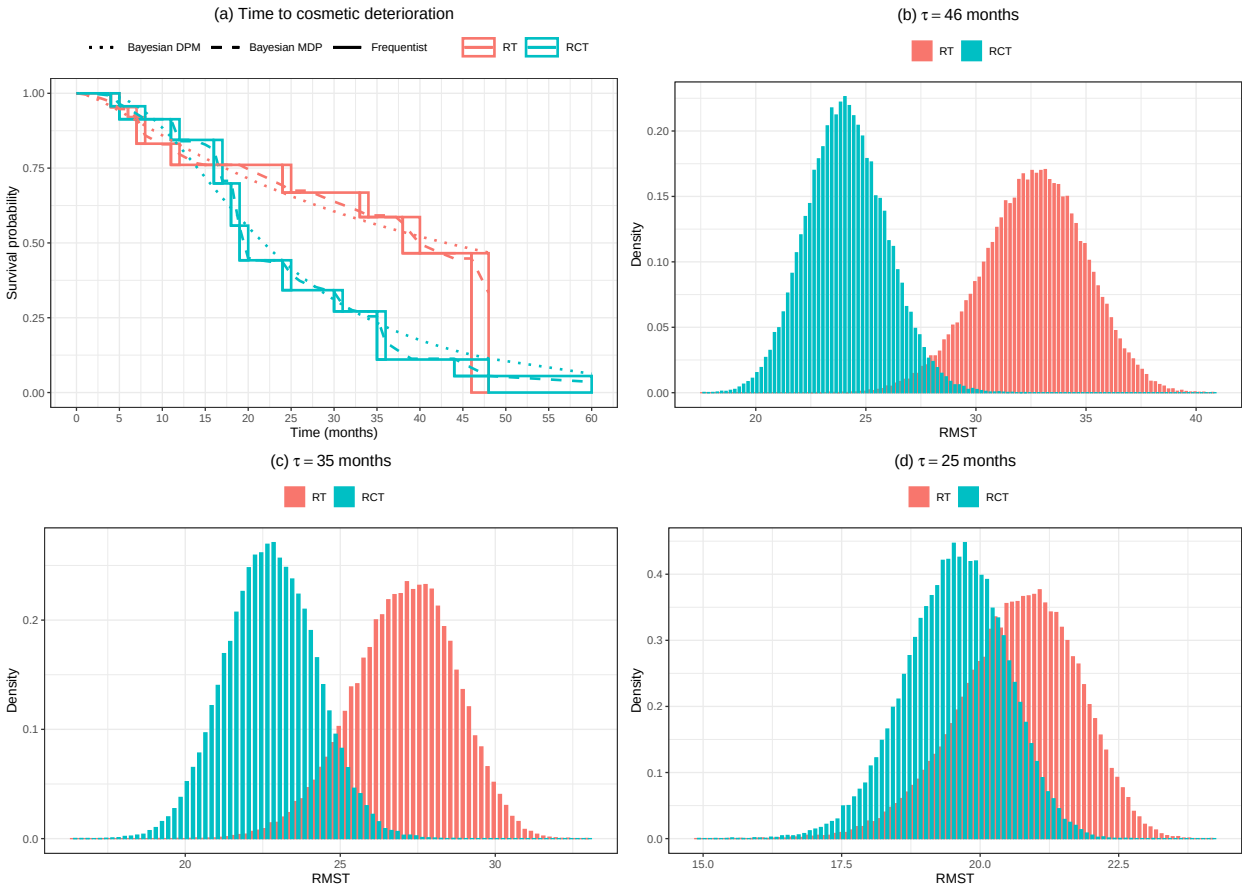

Figure 2: (a) Survival probabilities for the time to cosmetic deterioration of the BCOS study (Finkelstein and Wolfe, 1985) estimated by the frequentist NPMLE (solid lines), Bayesian MDP (dashed lines) and DPM (dotted lines). Histograms of the sampled posterior RMSTs under a diffuse MDP prior for the radiation therapy (RT, red) and the radiation + chemotherapy therapy (RCT, green) groups evaluated at (b) $\tau = 46$; (c) $\tau = 35$; (d) $\tau = 25$ months.

Table 1: One-sample simulation results on RMST estimation and 95% coverage probability (CP) with right-censored data using Bayesian nonparametric MDP and DPM as well as Bayesian parametric and frequentist nonparametric approaches.

				Bayesian Nonparameteric				Bayesian		Frequentist	
			True	MDP		DPM		Parametric		Nonparametric	
Model	n	Distribution	RMST	Est.	CP	Est.	CP	Est.	CP	Est.	CP
Exp*	100	Weibull(1,0.5)	0.83	0.83	0.950	0.84	0.950	0.89	0.801	0.83	0.944
		Exp(1)	0.86	0.86	0.943	0.87	0.947	0.87	0.947	0.86	0.941
		Weibull(1.25,2)	1.08	1.08	0.942	0.99	0.820	0.99	0.821	1.08	0.940
		Pw-Exp(A)‡	1.17	1.17	0.945	1.12	0.920	1.14	0.956	1.17	0.943
		Pw-Exp(B)	0.92	0.92	0.946	0.94	0.936	0.95	0.914	0.92	0.942
		LN(0,1)	1.11	1.11	0.948	1.10	0.962	1.10	0.961	1.11	0.946
		LN(-0.5,1)	0.81	0.81	0.945	0.82	0.960	0.82	0.956	0.81	0.940
		LN(0.5,1)	1.42	1.42	0.946	1.37	0.912	1.37	0.912	1.42	0.945
	20	Weibull(1,0.5)	0.82	0.83	0.934	0.90	0.919	0.89	0.834	0.83	0.916
		Exp(1)	0.85	0.86	0.940	0.88	0.947	0.88	0.943	0.86	0.927
		Weibull(1.25,2)	1.07	1.08	0.939	0.99	0.954	0.99	0.954	1.08	0.933
		Pw-Exp(A)	1.17	1.17	0.931	1.13	0.961	1.15	0.963	1.17	0.918
		Pw-Exp(B)	0.92	0.92	0.936	0.97	0.928	0.96	0.921	0.92	0.918
		LN(0,1)	1.11	1.11	0.937	1.11	0.963	1.11	0.961	1.11	0.925
		LN(-0.5,1)	0.79	0.81	0.948	0.83	0.953	0.83	0.953	0.81	0.939
		LN(0.5,1)	1.42	1.42	0.941	1.38	0.957	1.38	0.958	1.42	0.927
LN†	100	LN(0,1)	1.11	1.11	0.950	1.11	0.956	1.11	0.951	1.11	0.946
		LN(-0.5,1)	0.81	0.81	0.945	0.81	0.952	0.81	0.946	0.81	0.940
		LN(0.5,1)	1.42	1.42	0.947	1.41	0.953	1.42	0.947	1.42	0.944
		Pw-Exp(A)	1.17	1.17	0.946	1.15	0.981	1.13	0.936	1.17	0.946
		Pw-Exp(B)	0.92	0.92	0.947	0.94	0.960	0.94	0.955	0.92	0.946
		Weibull(1,0.5)	0.83	0.83	0.949	0.83	0.970	0.84	0.962	0.83	0.944
		Exp(1)	0.86	0.86	0.945	0.87	0.970	0.87	0.964	0.86	0.940
		Weibull(1.25,2)	1.08	1.08	0.943	1.07	0.972	1.07	0.963	1.08	0.941
	20	LN(0,1)	1.11	1.11	0.938	1.12	0.949	1.12	0.947	1.11	0.926
		LN(-0.5,1)	0.79	0.81	0.947	0.81	0.946	0.81	0.944	0.81	0.939
		LN(0.5,1)	1.42	1.42	0.940	1.41	0.949	1.42	0.947	1.42	0.926
		Pw-Exp(A)	1.17	1.17	0.931	1.14	0.950	1.14	0.947	1.17	0.918
		Pw-Exp(B)	0.92	0.92	0.932	0.94	0.946	0.94	0.943	0.92	0.917
		Weibull(1,0.5)	0.82	0.82	0.932	0.84	0.957	0.84	0.957	0.83	0.914
		Exp(1)	0.85	0.85	0.940	0.87	0.954	0.87	0.955	0.86	0.927
		Weibull(1.25,2)	1.07	1.08	0.940	1.07	0.949	1.07	0.947	1.08	0.933

Note: *Exponential, †Log-Normal, ‡Piecewise-Exponential.

Table 2: One-sample simulation results on RMST estimation and 95% coverage probability (CP) with interval-censored data using Bayesian nonparametric MDP and DPM as well as Bayesian parametric, naive (which treats interval-censored data as right-censored) and linear smoothing approaches.

				Bayesian Nonparametric				Bayesian		Naive		Linear	
			True	MDP		DPM		Parametric		Right-censored		Smoothing	
Model	n	Distribution	RMST	Est.	CP	Est.	CP	Est.	CP	Est.	CP	Est.	CP
Exp*	100	Weibull(1,0.5)	0.83	0.83	0.950	0.84	0.951	0.92	0.672	0.92	0.761	0.83	0.947
		Exp(1)	0.86	0.86	0.950	0.87	0.955	0.87	0.954	0.99	0.527	0.86	0.948
		Weibull(1.25,2)	1.08	1.08	0.953	0.96	0.552	0.96	0.553	1.21	0.257	1.08	0.948
		Pw-Exp(A)‡	1.17	1.17	0.953	1.11	0.884	1.11	0.883	1.28	0.579	1.17	0.951
		Pw-Exp(B)	0.92	0.92	0.949	0.94	0.932	0.97	0.875	1.03	0.646	0.92	0.947
		LN(0,1)	1.11	1.11	0.942	1.10	0.951	1.10	0.949	1.22	0.587	1.11	0.941
		LN(-0.5,1)	0.81	0.81	0.946	0.81	0.957	0.81	0.954	0.94	0.417	0.81	0.944
		LN(0.5,1)	1.42	1.42	0.942	1.37	0.907	1.37	0.902	1.50	0.701	1.42	0.939
	20	Weibull(1,0.5)	0.83	0.83	0.937	0.90	0.913	0.93	0.809	0.92	0.894	0.83	0.921
		Exp(1)	0.86	0.87	0.938	0.88	0.954	0.88	0.946	0.99	0.845	0.86	0.923
		Weibull(1.25,2)	1.08	1.08	0.932	0.97	0.955	0.97	0.953	1.21	0.742	1.08	0.922
		Pw-Exp(A)	1.17	1.17	0.930	1.12	0.964	1.12	0.961	1.28	0.833	1.17	0.918
		Pw-Exp(B)	0.92	0.92	0.939	0.97	0.930	0.98	0.903	1.03	0.868	0.92	0.920
		LN(0,1)	1.11	1.11	0.931	1.10	0.961	1.10	0.958	1.22	0.852	1.11	0.918
		LN(-0.5,1)	0.81	0.81	0.934	0.82	0.955	0.82	0.956	0.93	0.845	0.81	0.915
		LN(0.5,1)	1.42	1.42	0.932	1.38	0.950	1.38	0.949	1.50	0.845	1.42	0.916
LN†	100	LN(0,1)	1.11	1.11	0.943	1.11	0.951	1.11	0.941	1.22	0.587	1.11	0.938
		LN(-0.5,1)	0.81	0.81	0.945	0.81	0.954	0.81	0.948	0.94	0.418	0.81	0.944
		LN(0.5,1)	1.42	1.42	0.941	1.41	0.951	1.42	0.945	1.50	0.701	1.42	0.938
		Pw-Exp(A)	1.17	1.17	0.952	1.15	0.972	1.13	0.928	1.28	0.575	1.17	0.949
		Pw-Exp(B)	0.92	0.92	0.950	0.93	0.960	0.93	0.951	1.03	0.647	0.92	0.947
		Weibull(1,0.5)	0.83	0.83	0.951	0.82	0.967	0.82	0.955	0.92	0.763	0.83	0.949
		Exp(1)	0.86	0.86	0.951	0.85	0.971	0.85	0.961	0.99	0.528	0.86	0.948
		Weibull(1.25,2)	1.08	1.08	0.951	1.07	0.972	1.07	0.961	1.21	0.256	1.08	0.948
	20	LN(0,1)	1.11	1.11	0.931	1.12	0.947	1.12	0.944	1.22	0.852	1.11	0.918
		LN(-0.5,1)	0.81	0.81	0.934	0.82	0.949	0.82	0.947	0.93	0.844	0.81	0.913
		LN(0.5,1)	1.42	1.42	0.931	1.41	0.947	1.41	0.943	1.50	0.845	1.42	0.915
		Pw-Exp(A)	1.17	1.17	0.932	1.14	0.948	1.14	0.943	1.28	0.833	1.17	0.919
		Pw-Exp(B)	0.92	0.92	0.937	0.93	0.951	0.93	0.949	1.03	0.868	0.92	0.921
		Weibull(1,0.5)	0.83	0.83	0.939	0.83	0.955	0.83	0.953	0.92	0.894	0.83	0.920
		Exp(1)	0.86	0.87	0.937	0.86	0.959	0.86	0.959	0.99	0.847	0.86	0.924
		Weibull(1.25,2)	1.08	1.08	0.930	1.07	0.953	1.07	0.950	1.21	0.741	1.08	0.922

Note: *Exponential, †Log-Normal, ‡Piecewise-Exponential.

Table 3: Two-sample simulation results with right-censored data of sample sizes $n = 100$ and 20 per group using Bayesian nonparametric MDP and DPM as well as Bayesian parametric and frequentist nonparametric approaches. The column 'Prob' represents the probability of rejecting H_0.

				Bayesian Nonparametric						Bayesian			Frequentist		
	Fitted	True Distribution		MDP			DPM			Parametric			Nonparametric		
n	Model	Group 1	Group 2	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob
Under H_0: $\text{RMST}_1(\tau) \leqslant \text{RMST}_2(\tau)$															
100	Exp*	Exp(1)	Exp(1)	0.86	0.87	0.046	0.87	0.87	0.039	0.87	0.87	0.046	0.86	0.87	0.047
		LN(0,1)	LN(0,1)	1.11	1.12	0.052	1.10	1.10	0.040	1.10	1.10	0.039	1.11	1.12	0.051
		Pw-Exp‡(B)	Pw-Exp(B)	0.92	0.92	0.050	0.95	0.95	0.033	0.95	0.95	0.058	0.92	0.92	0.052
	LN†	Exp(1)	Exp(1)	0.86	0.87	0.046	0.87	0.87	0.024	0.87	0.88	0.029	0.86	0.87	0.046
		LN(0,1)	LN(0,1)	1.11	1.12	0.054	1.11	1.12	0.043	1.11	1.12	0.049	1.11	1.12	0.052
		Pw-Exp(B)	Pw-Exp(B)	0.92	0.92	0.050	0.93	0.93	0.031	0.93	0.94	0.034	0.92	0.92	0.051
20	Exp	Exp(1)	Exp(1)	0.85	0.86	0.045	0.87	0.87	0.048	0.87	0.87	0.049	0.85	0.86	0.047
		LN(0,1)	LN(0,1)	1.11	1.11	0.067	1.11	1.11	0.050	1.11	1.11	0.050	1.11	1.11	0.070
		Pw-Exp(B)	Pw-Exp(B)	0.91	0.92	0.053	0.95	0.96	0.067	0.95	0.96	0.070	0.91	0.92	0.054
	LN	Exp(1)	Exp(1)	0.85	0.86	0.046	0.86	0.86	0.035	0.85	0.86	0.034	0.85	0.86	0.047
		LN(0,1)	LN(0,1)	1.11	1.11	0.067	1.11	1.12	0.059	1.11	1.12	0.065	1.11	1.11	0.070
		Pw-Exp(B)	Pw-Exp(B)	0.91	0.92	0.053	0.93	0.93	0.043	0.93	0.93	0.042	0.91	0.92	0.055
Under H_1: $\text{RMST}_1(\tau) > \text{RMST}_2(\tau)$															
100	Exp	LN(-0.5,1)	LN(0,1)	0.81	1.12	0.942	0.82	1.10	0.875	0.82	1.10	0.901	0.81	1.12	0.942
		LN(0,1)	LN(0.5,1)	1.12	1.42	0.933	1.10	1.37	0.864	1.10	1.37	0.897	1.12	1.42	0.937
		Pw-Exp(B)	Pw-Exp(A)	0.92	1.17	0.784	0.95	1.12	0.416	0.95	1.12	0.512	0.92	1.17	0.793
		Exp(1)	Weibull(1.25,2)	0.86	1.08	0.797	0.87	0.99	0.345	0.87	0.99	0.352	0.86	1.08	0.800
		Weibull(1,0.5)	Exp(1)	0.83	0.86	0.096	0.85	0.87	0.000	0.87	0.89	0.094	0.83	0.86	0.094
	LN	LN(-0.5,1)	LN(0,1)	0.81	1.12	0.943	0.81	1.12	0.886	0.81	1.12	0.944	0.81	1.12	0.943
		LN(0,1)	LN(0.5,1)	1.12	1.42	0.937	1.12	1.42	0.865	1.12	1.42	0.940	1.12	1.42	0.935
		Pw-Exp(B)	Pw-Exp(A)	0.92	1.17	0.789	0.93	1.15	0.298	0.94	1.13	0.582	0.92	1.17	0.793
		Exp(1)	Weibull(1.25,2)	0.86	1.08	0.794	0.87	1.07	0.542	0.87	1.07	0.685	0.86	1.08	0.802
		Weibull(1,0.5)	Exp(1)	0.83	0.86	0.093	0.84	0.87	0.057	0.84	0.87	0.070	0.83	0.86	0.095
20	Exp	LN(-0.5,1)	LN(0,1)	0.80	1.07	0.405	0.82	1.06	0.343	0.82	1.06	0.342	0.80	1.07	0.418
		LN(0,1)	LN(0.5,1)	1.11	1.42	0.460	1.11	1.38	0.367	1.11	1.38	0.369	1.11	1.42	0.473
		Pw-Exp(B)	Pw-Exp(A)	0.92	1.16	0.320	0.96	1.12	0.187	0.96	1.12	0.195	0.92	1.16	0.327
		Exp(1)	Weibull(1.25,2)	0.86	1.07	0.312	0.87	0.97	0.108	0.87	0.98	0.112	0.86	1.07	0.324
		Weibull(1,0.5)	Exp(1)	0.81	0.87	0.083	0.87	0.88	0.040	0.88	0.88	0.077	0.81	0.87	0.087
	LN	LN(-0.5,1)	LN(0,1)	0.80	1.07	0.402	0.81	1.07	0.397	0.81	1.07	0.406	0.80	1.07	0.419
		LN(0,1)	LN(0.5,1)	1.11	1.42	0.460	1.11	1.41	0.430	1.11	1.41	0.449	1.11	1.42	0.473
		Pw-Exp(B)	Pw-Exp(A)	0.92	1.16	0.320	0.93	1.13	0.211	0.93	1.13	0.222	0.92	1.16	0.332
		Exp(1)	Weibull(1.25,2)	0.86	1.07	0.314	0.86	1.05	0.243	0.86	1.05	0.249	0.86	1.07	0.322
		Weibull(1,0.5)	Exp(1)	0.81	0.87	0.085	0.82	0.87	0.057	0.82	0.87	0.056	0.81	0.87	0.086

Note: *Exponential, †Log-Normal, ‡Piecewise-Exponential.

Table 4: Two-sample simulation results with interval-censored data of sample sizes $n = 100$ and 20 per group using Bayesian nonparametric MDP and DPM as well as Bayesian parametric and linear smoothing approaches. The column 'Prob' represents the probability of rejecting H_0.

				Bayesian Nonparametric						Bayesian			Linear		
	Fitted	True Distribution		MDP			DPM			Parametric			Smoothing		
n	Model	Group 1	Group 2	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob	$\widehat{\text{RMST}}_1$	$\widehat{\text{RMST}}_2$	Prob
				Under H_0: $\text{RMST}_1(\tau) \leqslant \text{RMST}_2(\tau)$											
100	Exp*	Exp(1)	Exp(1)	0.86	0.86	0.054	0.87	0.87	0.047	0.87	0.87	0.053	0.86	0.86	0.052
		LN(0,1)	LN(0,1)	1.12	1.12	0.052	1.10	1.10	0.043	1.10	1.10	0.043	1.12	1.12	0.053
		Pw-Exp‡(B)	Pw-Exp(B)	0.92	0.92	0.050	0.96	0.96	0.025	0.97	0.97	0.066	0.92	0.92	0.055
	LN†	Exp(1)	Exp(1)	1.12	1.12	0.051	1.12	1.12	0.044	1.12	1.12	0.055	1.12	1.12	0.052
		LN(0,1)	LN(0,1)	0.86	0.86	0.052	0.85	0.85	0.028	0.85	0.85	0.044	0.86	0.86	0.052
		Pw-Exp(B)	Pw-Exp(B)	0.92	0.92	0.051	0.93	0.93	0.029	0.93	0.93	0.046	0.92	0.92	0.053
20	Exp	Exp(1)	Exp(1)	0.86	0.86	0.065	0.87	0.87	0.052	0.87	0.87	0.051	0.86	0.86	0.064
		LN(0,1)	LN(0,1)	1.11	1.11	0.050	1.10	1.10	0.031	1.10	1.10	0.031	1.11	1.11	0.057
		Pw-Exp(B)	Pw-Exp(B)	0.92	0.93	0.049	0.98	0.98	0.052	0.98	0.98	0.061	0.92	0.93	0.061
	LN	Exp(1)	Exp(1)	0.86	0.86	0.062	0.85	0.85	0.044	0.85	0.85	0.047	0.86	0.86	0.066
		LN(0,1)	LN(0,1)	1.11	1.11	0.054	1.11	1.12	0.042	1.12	1.12	0.043	1.11	1.11	0.057
		Pw-Exp(B)	Pw-Exp(B)	0.92	0.93	0.054	0.93	0.94	0.042	0.93	0.94	0.042	0.92	0.93	0.059
				Under H_1: $\text{RMST}_1(\tau) > \text{RMST}_2(\tau)$											
100	Exp	LN(0,1)	LN(-0.5,1)	1.12	0.81	0.947	1.10	0.81	0.907	1.10	0.82	0.930	1.12	0.81	0.949
		LN(0.5,1)	LN(0,1)	1.42	1.12	0.941	1.37	1.10	0.882	1.37	1.10	0.913	1.42	1.12	0.941
		Pw-Exp(A)	Pw-Exp(B)	1.17	0.92	0.802	1.11	0.96	0.327	1.11	0.97	0.434	1.17	0.92	0.802
		Weibull(1.25,2)	Exp(1)	1.08	0.86	0.809	0.96	0.87	0.250	0.96	0.87	0.252	1.08	0.86	0.815
		Exp(1)	Weibull(1,0.5)	0.86	0.83	0.115	0.87	0.84	0.000	0.92	0.87	0.183	0.86	0.83	0.112
	LN	LN(0,1)	LN(-0.5,1)	1.12	0.81	0.949	1.12	0.81	0.893	1.12	0.81	0.956	1.12	0.81	0.950
		LN(0.5,1)	LN(0,1)	1.42	1.12	0.941	1.42	1.12	0.836	1.42	1.12	0.948	1.42	1.12	0.940
		Pw-Exp(A)	Pw-Exp(B)	1.17	0.92	0.805	1.15	0.93	0.351	1.13	0.93	0.664	1.17	0.92	0.804
		Weibull(1.25,2)	Exp(1)	1.08	0.86	0.814	1.07	0.85	0.603	1.07	0.85	0.810	1.08	0.86	0.817
		Exp(1)	Weibull(1,0.5)	0.86	0.83	0.111	0.85	0.82	0.069	0.85	0.82	0.088	0.86	0.83	0.110
20	Exp	LN(0,1)	LN(-0.5,1)	1.11	0.81	0.477	1.10	0.82	0.405	1.10	0.82	0.411	1.11	0.81	0.488
		LN(0.5,1)	LN(0,1)	1.42	1.11	0.469	1.37	1.10	0.358	1.37	1.10	0.363	1.42	1.11	0.473
		Pw-Exp(A)	Pw-Exp(B)	1.17	0.93	0.308	1.12	0.98	0.162	1.12	0.98	0.170	1.17	0.93	0.319
		Weibull(1.25,2)	Exp(1)	1.08	0.86	0.360	0.97	0.87	0.095	0.97	0.87	0.096	1.08	0.86	0.376
		Exp(1)	Weibull(1,0.5)	0.86	0.82	0.085	0.91	0.87	0.068	0.92	0.87	0.107	0.86	0.82	0.088
	LN	LN(0,1)	LN(-0.5,1)	1.11	0.81	0.481	1.11	0.81	0.459	1.11	0.81	0.465	1.11	0.81	0.486
		LN(0.5,1)	LN(0,1)	1.42	1.11	0.469	1.41	1.12	0.433	1.41	1.12	0.443	1.42	1.11	0.479
		Pw-Exp(A)	Pw-Exp(B)	1.17	0.93	0.307	1.14	0.94	0.229	1.14	0.94	0.241	1.17	0.93	0.322
		Weibull(1.25,2)	Exp(1)	1.08	0.86	0.358	1.07	0.85	0.311	1.06	0.85	0.321	1.08	0.86	0.374
		Exp(1)	Weibull(1,0.5)	0.86	0.82	0.084	0.85	0.82	0.062	0.85	0.82	0.067	0.86	0.82	0.088

Note: *Exponential, †Log-Normal, ‡Piecewise-Exponential.