Published in final edited form as:
Stat Med. 2014 January 15; 33(1): 59–73. doi:10.1002/sim.5915.

Improved Survival Modeling in Cancer Research Using a Reduced Piecewise Exponential Approach

Gang Han[1], **Michael J. Schell**[2,3], and **Jongphil Kim**[2,3]

Gang Han: Gang.Han@yale.edu

[1]Department of Biostatistics, Yale University School of Public Health, 60 College Street, New Haven, CT 06520

[2]Department of Biostatistics, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL, 33612

[3]Oncologic Sciences, University of South Florida, 4202 E. Fowler Ave Tampa, FL, 33620

Abstract

Statistical models for survival data are typically nonparametric, e.g., the Kaplan-Meier curve. Parametric survival modeling, such as exponential modeling, however, can reveal additional insights and be more efficient than nonparametric alternatives. A major constraint of the existing exponential models is the lack of flexibility due to distribution assumptions. A flexible and parsimonious piecewise exponential model is presented to best use the exponential models for arbitrary survival data. This model identifies shifts in the failure rate over time based on an exact likelihood ratio test, a backward elimination procedure, and an optional presumed order restriction on the hazard rate. Such modeling provides a descriptive tool in understanding the patient survival in addition to the Kaplan-Meier curve. This approach is compared with alternative survival models in simulation examples and illustrated in clinical studies.

Keywords

Survival analysis; exponential survival; non-small-cell lung cancer; median survival

1. Introduction

Time-to-event data are commonly encountered in life science applications for which proper and efficient statistical modeling is becoming more important in the past decade. This is mainly because that new molecular information, which can be from gene sequencing, identifying biomarkers in translational research, or analyzing protein expression data. Having sufficient numbers of patients to demonstrate the effectiveness of novel targeted therapies for each disease subtype has therefore become a challenge and a necessity in the design of clinical trials. Nonparametric models, such as the Kaplan-Meier curve, can graphically demonstrate the distribution, but the stepwise local estimates involved render an inefficient formulaic summary of the survival distribution. Parametric models, e.g., exponential models, can routinely reduce the study sample size. Miller [1] showed that the asymptotic efficiencies of the Kaplan-Meier estimator (KME) relative to the exponential estimate are 0.64, 0.48, and 0.59 at survival fractions of 0.25, 0.5, and 0.1 under the

*Correspondence to: Department of Biostatistics, Yale University School of Public Health, 60 College Street, New Haven, CT 06520; Gang.Han@yale.edu.

exponential assumption (with concurrence from Meier *et al.* [2]). Thus, a properly fitted exponential model could achieve similar power as Kaplan-Meier based fitting but with one third to half of the patients saved. In clinical trials with survival outcomes, correct use of exponential models can greatly reduce cost, patient size, follow-up time, which makes the proposed trial more ethically attractive and more achievable. These benefits are especially significant for studies of rare diseases and those require data from detailed chart review. Single-arm phase II oncogenic clinical trials with survivorship at some fixed time as the endpoint provides one such opportunity for improving on either the sample size or the precision of the estimate. An example of such a trial was given in Seymour *et al.* [3].

Although the exponential modeling has tremendous advantage in saving financial/patient resources and accelerating the approval of agents, its application is limited due to the fact that exponential models can be applied only if the distribution assumption is valid. Little statistical methodology has ever been developed to take into account the exponential saving under non-exponential survival distributions. Motivated by this fact, we propose in this article a modeling approach to take advantage of the exponential saving for any survival distribution by developing a piecewise exponential distribution with as few exponential pieces as possible at a given significant level.

Piecewise exponential distribution has been recognized as a simple and flexible tool in survival analysis. For example, Berry *et al.* [4] built a piecewise baseline hazard function in their Bayesian model, allowing the hazard rate to vary in each of the follow-up years; Berry *et al.* [5] emphasized that these piecewise exponential models are valuable in cancer immunotherapy trials discussed in Hoos *et al.* [6]; Edwards *et al.* [7] identified a sudden reduction in the mortality rates for prostate cancer, which implied that a model with piecewise constant hazard assumption would be particularly useful for interpreting such cancer survival and to facilitate the treatments and diagnoses [8]. The use of piecewise exponential models has been advocated in the literature for more than 35 years [9–11]. The major challenge in using the piecewise exponential model is to identify significant change-points in the failure rate over time. A number of articles have been published on detecting a single change-point with piecewise constant failure assumption presumed to be true [12–20]. Some articles were published in the past 3–5 years for detecting multiple change-points in failure rate. For example, Demarqui *et al.* [21] introduced a full Bayesian approach to determine the number of change-points. Goodman *et al.* [8] proposed detecting multiple change-points in piecewise constant hazard function using a Wald type test based on maximum likelihood estimates and a forward selection sequential testing procedure. Both approaches yielded reasonable solutions in the simulation examples and real examples shown in the two articles. However, we believe that there is a gap between these methodologies and medical research/practice because of the following reasons. First, the number of patients in a trial can be too small for the asymptotic test (e.g., Wald type tests) to hold. An exact test is desirable but has not been proposed yet. Secondly, although the Bayesian methods in Demarqui *et al.* [21] are promising, scant prior information may limit its use in practice. Thirdly, the forward selection approach may not be able to identify all significant change-points given that the model election algorithm terminates if the sequential testing fails to reject the null hypothesis for the first time [22]. Lastly, none of the existing methods has been combined with an order restriction in the failure rate, which is not uncommon in cancer studies [7,23].

In this article, we develop a novel approach to identify all statistically significant change points in failure rate based on a set of time-to-event data. This approach involves three key components: an exact likelihood ratio test, a backward elimination procedure, and an optional order restriction. We refer to the corresponding piecewise exponential estimate as “the reduced piecewise exponential estimate” or “RPEXE” due to the fact that it represents a

reduction in the number of failure rate changes in the piecewise exponential estimate (PEXE) of Kim *et al.* [24]. Rest of the article is organized as follows. In Section 2, we show a motivating example of RPEXE in a lung cancer clinical trial. In Section 3 we introduce the piecewise exponential distribution and derive the proposed RPEXE approach. In Section 4 we compare the proposed approach with existing parametric and non-parametric modeling methods in simulation examples. We demonstrate the use of RPEXE in real examples in Section 5.

2. A Motivating Example

The idea of the proposed approach is to describe the survival using a piecewise exponential distribution with changepoints that indicate statistically significant changes of failure rates. Location of the changepoints is determined by the survival data and failure rate order restrictions. As a result, the fitted model is not subjected to distribution assumptions, but is still more parsimonious than nonparametric alternatives. Built on an exact test [25] and a backward elimination procedure [26], the proposed approach is able to identify significant changepoints and provide reasonable survival estimates without requiring a large number of events. As a motivating example, we demonstrate RPEXE and alternative parametric and nonparametric approaches in a cancer trial in the rest of this section before introducing the technical details of RPEXE in Section 3.

Lung cancer has the highest mortality among all types of cancer. About 85% of the newly diagnosed lung cancer patients have non-small cell lung cancer (NSCLC). Simon *et al.* [27] and Simon *et al.* [28] reported that the median overall survival of NSCLC patients is about 8 to 10 months and the 1 year survival rate is about 31% to 36%. In recent years, customized chemotherapy has been considered as a promising method to improve the longevity and possible cure of NSCLC patients. The failure rate of NSCLC patients is expected to decrease over time. In this motivating example we apply RPEXE to investigate the survival data in a study for NSCLC patients using customized therapy conducted at H. Lee Moffitt Cancer Center and Research Institute. The data set contains 178 patients who were treated on institutional clinical trials for newly diagnosed late-stage NSCLC. Of these patients 154 died and 24 were censored with the last event occurring around month 51. The solid curve in Figure 1 shows the Kaplan-Meier estimate (KME) of the survival function.

Prior to implementing RPEXE, we modeled this data using distributions in SAS 9.3, PROC LIFEREG. The most flexible model we have found is the generalized gamma. When there is no covariate, the generalized gamma distribution has three model parameters: the intercept, β, the shape parameter, δ, and the scale parameter σ. As specified in Cox *et al.* [29], the probability density function (p.d.f.) of the generalized gamma distribution can be written as

$$f(t|\beta,\delta,\sigma)=\frac{|\delta|}{\sigma t\Gamma(\delta^{-2})}\times\left[\delta^{-2}\left(e^{-\beta}t\right)^{\delta/\sigma}\right]^{\delta^{-2}}\times\exp\left[-\delta^{-2}\left(e^{-\beta}t\right)^{\delta/\sigma}\right]. \quad (1)$$

The generalized gamma distribution becomes exponential if $\delta = 1$ and $\sigma = 1$, Weibull if $\delta = 1$, gamma if $\delta = \sigma$, and lognormal if $\delta = 0$. The generalized gamma model gives the estimate of β: 2.405, with 95% confidence interval (CI): (2.086, 2.725), the estimate of δ: 0.339, with 95% CI: (−0.068, 0.747), and the estimate of σ: 1.337, with 95% CI:(1.163, 1.538). The confidence limits indicate that the distribution is not exponential, gamma, or Weibull with the lognormal distribution also being rather unlikely. Figure 1 plots the estimated survival function from a nonparametric estimate (KME) and GGE. The estimated $S(t)$ from GGE is consistently higher than KME when $t \in (15, 40)$. This can cause incorrect interpretation of the survivorship. We will implement the proposed approach to demonstrate that it can

address this issue and improve the estimation efficiency compared with KME in Section 5.1 after introducing the method in Section 3 and conducting simulation studies in Section 4.

3. The Reduced Piecewise Exponential Model

We let $[Z]$ denote the distribution of a generic random variable Z and z denote a realization from $[Z]$. We let "log" denote the natural logarithm. Suppose $X_e \in (0,\infty)$ is a random variable with the exponential distribution having parameter λ. Suppose $X_{pe} \in (0,\infty)$ is a random variable with a piecewise exponential distribution divided by change-points $t_{(1)} < t_{(2)} < \cdots < t_{(p)}$ at which the failure rate varies.

The p.d.f. $f(X_e|\lambda)$, survival function $S(X_e|\lambda)$, and hazard function $h(X_e|\lambda)$ are $f(X_e|\lambda) = \exp\{-X_e/\lambda\}/\lambda$, $S(X_e|\lambda) = \exp\{-X_e/\lambda\}$, and $h(X_e|\lambda) = f(X_e|\lambda)/S(X_e|\lambda) = 1/\lambda$. Further, the mean and median of X_e are $E(X_e|\lambda) = \lambda$ and $M(X_e|\lambda) = \lambda\log(2)$. Note that the exponential parameter λ can be written as e^{β} with the notation of the generalized gamma distribution in (1).

Define $t_{(0)} = 0$ and $t_{(p+1)} = \infty$. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{p+1})$ denote the vector of the unknown model parameters, where $1/\lambda_j$ is the instantaneous failure rate in $(t_{(j-1)}, t_j]$ for $j = 1, \ldots, p+1$. Define $E_0 = 1$ and

$$E_j = \prod_{k=1}^{j} \exp\left\{-\frac{t_{(k)} - t_{(k-1)}}{\lambda_k}\right\} = E_{j-1}\exp\left\{-\frac{t_{(j)} - t_{(j-1)}}{\lambda_j}\right\}.$$

Note that $(t_{(j)}, E_j)$ is the point where the j^{th} piece of the $S(X_{pe}|\boldsymbol{\lambda})$ starts in the plot of $S(X_{pe}|\boldsymbol{\lambda})$ against X_{pe}. So $[(X_{pe} - t_{(j-1)})/E_{j-1}] \sim \text{Gamma}(1, \lambda_j)$ for $X_{pe} \in [t_{(j-1)}, t_{(j)})$, where Gamma$(1, \lambda_j)$ denotes the gamma distribution with location parameter 1 and scale parameter λ_j. Thus, $S(X_{pe}|\boldsymbol{\lambda})$, $f(X_{pe}|\boldsymbol{\lambda})$, $h(X_{pe}|\boldsymbol{\lambda})$, and the cumulative hazard function $H(X_{pe}|\boldsymbol{\lambda})$ are

$$\begin{aligned} S(X_{pe}|\boldsymbol{\lambda}) &= E_{j-1}\exp\left\{-\frac{X_{pe} - t_{(j-1)}}{\lambda_j}\right\}, \\ f(X_{pe}|\boldsymbol{\lambda}) &= E_{j-1}\frac{1}{\lambda_j}\exp\left\{-\frac{X_{pe} - t_{(j-1)}}{\lambda_j}\right\} \\ h(X_{pe}|\boldsymbol{\lambda}) &= \frac{1}{\lambda_j}, \\ H(X_{pe}|\boldsymbol{\lambda}) &= -\log(E_{j-1}) + \frac{X_{pe} - t_{(j-1)}}{\lambda_j}, \end{aligned} \quad (2)$$

for $X_{pe} \in (t_{(j-1)}, t_{(j)}]$ and $j = 1, \ldots, p+1$. The mean and median survival times are $E(X_{pe}|\boldsymbol{\lambda}) = \lambda_1 + \sum_{i=1}^{p}(\lambda_{i+1} - \lambda_i)E_i$ and $\mu_{1/2}(X_{pe}|\boldsymbol{\lambda}) = t_{(m)} + \lambda_{m+1}\log(2E_m)$, respectively, where m is in $\{1, \ldots, p\}$ such that $E_m \geq 0.5$ and $E_{m+1} < 0.5$.

Suppose a study involves N patients, where D of them failed at time points $t_1 \leq t_2 \leq \cdots \leq t_D$, and the other $(N - D)$ subjects were censored at $t_{D+1}, \ldots, t_N$. Let $t_1\star < t_2\star < \ldots < t_{D^\star}\star$ denote all the distinct values in $\{t_1, \ldots, t_D\}$ where $D^\star \leq D$. We assume that the censoring mechanism is independent with the distribution of the failure time. Define $t_0 = 0$. The total-time-on-test (TTOT) between t_A and t_B for $t_A < t_B$ is defined to be the sum of all subjects' times in $(t_A, t_B]$ for all $t_A, t_B > 0$; i.e.,

$$\mathrm{TTOT}(t_A, t_B) = \sum_{j=1}^{N} \max[0, \min(t_j, t_B) - t_A].$$

The normalized spacing between t_A and t_B is defined as TTOT(t_A, t_B) /d_{AB}, where $d_{AB} > 0$ is the number of events occurred in (t_A, t_B].

For exponentially distributed data, the maximum likelihood estimate (MLE) of λ is $\hat{\lambda} = \sum_{i=1}^{N} t_i / D$([30], page 41) and the 100(1 − α)% equal-tailed confidence interval of λ is

$$\lambda \in \left[\frac{2D\hat{\lambda}}{\chi^2_{2D,1-\alpha/2}}, \frac{2D\hat{\lambda}}{\chi^2_{2D,\alpha/2}} \right],$$

where $\chi^2_{2D,\alpha/2}$ is the 100 × α/2% lower quantile at of the central chi-square distribution with 2D degree of freedom [31]. Similarly, the first and second order derivatives of the likelihood function of the piecewise exponential distribution log $L(\boldsymbol{\lambda} | t_1, \ldots, t_N)$ lead to the MLE of λ_j , i.e.,

$$\hat{\lambda}_j = \frac{\mathrm{TTOT}(t_{(j-1)}, t_{(j)})}{d_{(j)}}$$

and confidence interval of λ_j is

$$\lambda_j \in \left[\frac{2d_{(j)}\hat{\lambda}_j}{\chi^2_{2d_{(j)},1-\alpha/2}}, \frac{2d_{(j)}\hat{\lambda}_j}{\chi^2_{2d_{(j)},\alpha/2}} \right],$$

for all j = 1, … , p + 1. The MLEs of $S(X_{pe}|\boldsymbol{\lambda})$, $h(X_{pe}|\boldsymbol{\lambda})$, and $M(X_{pe}|\boldsymbol{\lambda})$ are $S(X_{pe}|\hat{\boldsymbol{\lambda}})$, $h(X_{pe}|\hat{\boldsymbol{\lambda}})$, and $M(X_{pe}|\hat{\boldsymbol{\lambda}})$, respectively. The confidence intervals of $\{S(X_{pe}|\boldsymbol{\lambda}), h(X_{pe}|\boldsymbol{\lambda}), M(X_{pe}|\boldsymbol{\lambda})\}$ can be computed by plugging in the lower and upper bounds of $\boldsymbol{\lambda}$. Multiple testing adjustments, e.g., Bonferroni approach, can be implemented to control the level of a confidence interval when multiple model parameters are involved.

The key to the proposed RPEXE model is to set change-points $t_{(1)}, \ldots, t_{(p)}$ given the data $\{t_1, \ldots, t_N\}$. Our approach builds on three components. The first component is a likelihood ratio test. Let x_1 and x_2 denote realizations from Gamma(n_1, λ_1) and Gamma(n_2, λ_2), respectively. We suppose that n_1 and n_2 are *known* positive integers since the numbers of events will be known positive integers. The scale parameters λ_1 and λ_2 are *unknown*. The null and alternative hypotheses are

$$\mathrm{H}_0: \lambda_1 = \lambda_2 \text{ vs.} \mathrm{H}_1: \lambda_1 \neq \lambda_2.$$

Under the null hypothesis, we let $\lambda_1 = \lambda_2 = \lambda$. The likelihood ratio test (LRT) statistic can be derived as

$$\varphi(x_1, x_2) = \left(\frac{x_1}{x_1+x_2}\right)^{n_1} \left(\frac{x_2}{x_1+x_2}\right)^{n_2}.$$

The level α likelihood ratio test rejects H_0 if $\varphi(x_1, x_2) \le C_\alpha$, where C_α is a real value such that $P(\varphi(X_1,X_2) \le C_\alpha) = \alpha$ under H_0. [25] quantified the p-value and proved that this exact likelihood ratio test is the uniformly most powerful unbiased (UMPU) test of H_0 vs. H_1. This test is useful for detecting failure rate change in a single patient group and for testing the equivalence of the failure rates of two exponentially distributed groups.

The second component is a backward elimination procedure, which is used with the LRT to detect significant changes in the failure rate over time. Specifically, given $t_1\star < t_2\star < \cdots < t_{D^\star}\star$, it is possible to model the data with piecewise exponential distributions having 1 to $D^\star$ pieces. We let $\{T_1, T_2, \ldots, T_{D^\star}\star\}$ denote $D^\star$ TTOTs in $(0, t_1\star)$, $(t_1\star, t_2\star), \ldots, (t_{D^\star-1}\star, t_{D^\star}\star)$ and $\{d_1, d_2, \ldots, d_{D^\star}\star\}$ denote the corresponding numbers of events. Thus $D = \sum_{i=1}^{D^\star} d_i$. The backward elimination (BE) procedure has four steps:

Step 1. Let $\ell = D^\star$. Compute $\ell - 1$ pairs of TTOTs, $\{(T_1, T_2), \ldots, (T_{\ell-1}, T_\ell)\}$, and the corresponding pairs of events, $\{(d_1, d_2), \ldots, (d_{\ell-1}, d_\ell)\}$

Step 2. Compute the $\ell - 1$ p-values using the LRT for the $\ell - 1$ pairs of TTOTs.

Step 3. If the largest p-value exceeds a critical value $\alpha^\star$ and $\ell > 1$, add up the corresponding pair of TTOTs and pair of events (so that there become $\ell - 2$ pairs), and let $\ell = \ell - 1$.

Step 4. Repeat steps 2 and 3 until the largest p-value is smaller than a critical value $\alpha^\star$ or $\ell = 0$.

The third component is an optional order restriction. Expert knowledge may suggest that the change in failure rate follows a pattern. For example, Simon *et al.* [27] and Simon *et al.* [28] indicated that the failure rate of NSCLC patients should decrease over time. Bray *et al.* [23] studied U-shaped failure rate functions. In such situations, integrating appropriate order restrictions with a statistical model may be necessary. To implement simple order restriction, we use the pool-adjacent-violators-algorithm (PAVA) [32]. The RPEXE with non-increasing and non-decreasing failure rates, which we will call isotonic RPEXE and antitonic RPEXE, are computed by first implementing the PAVA to the $D^\star$ normalized spacings $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{D^\star}\star\}$, and then running the LRT test of H_0 vs. H_1. We let $\underline{L}$ denote the number of level sets after the PAVA operation. The resulting L estimates $\hat{\lambda}_1, \ldots, \hat{\lambda}_L$ will satisfy the assumed order restriction, and will be used in the BE procedure. Other less restrictive order restrictions may be favorable in applications, e.g., the monotonic order restriction where the failure rate can be either non-increasing or non-decreasing. Isotonic, antitonic, and monotonic order restrictions have been constructed for piecewise linear regression [26]. Another example is the IDFR (or DIFR) that denotes the increasing-decreasing (or decreasing-increasing) failure rate. Such order restrictions are often referred to as umbrella alternatives when dealing with non-survival data. When the partial order restriction is composed of multiple orderings (e.g., IDFR order where the shift time from an increasing to decreasing failure rate is not known), we use the ordering that maximizes the likelihood. See Robertson *et al.* [33] (page 34–36) for details about the order restricted maximum likelihood value and estimate.

Integrating the three components, we choose the change-points by first implementing an order restriction on the failure rate estimates, and then eliminate all insignificant change-points detected by the likelihood ratio test using the BE procedure. The elimination procedure stops when the largest p-value of all possible change-points is less than a critical value $\alpha^\star$, which controls the type I error of the simultaneous inference of all change-points. Given a confidence level α, we estimate the critical value $\alpha^\star$ by $\hat{\alpha^\star}$ using a 3-stage Monte Carlo (MC) procedure:

Stage 1. Generate an i.i.d. sample $\{x_1, \ldots, x_D\}$ from Exponential(1) (or Gamma(1, 1)). Let $D^\star$ denote the number of distinct values in $\{x_1, \ldots, x_D\}$.

Stage 2. If an order restriction is imposed, using PAVA to combine the values that violate the restriction to obtain L distinct level sets.

Stage 3. Initialize $\ell = D^\star$, or $\ell = L$ if there is an order restriction. Repeat the steps 2–3 in the BE procedure $(\ell - 1)$ times. Save the minimum of the p-values as $p_{\min}$.

Repeat the above 3-stage MC procedure N_s times. Let $\hat{\alpha^\star}$ denote the αth lower quantile among the N_s values. The 95% asymptotic lower and upper bounds of $\alpha^\star$ are the N_s^lth and N_s^uth smallest p-values among all the N_s $p_{\min}$ values, where $N_s^l = N_s \times \left(\alpha - z^{0.025}\sqrt{\alpha(1-\alpha)/N_s}\right)$, $N_s^u = N_s \times \left(\alpha + z^{0.025}\sqrt{\alpha(1-\alpha)/N_s}\right)$, and $z^{0.025}$ is the upper 0.025 quantile of the standard normal distribution.

Following the aforementioned algorithm, we have numerically estimated some critical values for practical use. Table 1 summarizes the estimates and 95% confidence intervals of $\alpha^\star$ corresponding to $\alpha = 0.05$ and $\alpha = 0.1$ from isotonic RPEXE, monotonic RPEXE, and BE for $D = 20$, 50, 200, and 800. The critical values isotonic RPEXE also apply to antitonic RPEXE. The log of $\hat{\alpha^\star}$ values are found to be linearly related with the log of the event number D with $R^2 > 0.9$. Thus we fit two regression lines of $\log(\hat{\alpha^\star})$ on $\log(D)$ for $\alpha = 0.05$ and $\alpha = 0.2$, respectively, which can be used as simple formulas to generate $\hat{\alpha^\star}$ given α and $D \in [20, 800]$. Table 2 shows the estimated regression coefficients $(\hat{\beta_0}, \hat{\beta_1})$ in regression lines of the form

$$\log(\hat{\alpha}^\star) = \hat{\beta}_0 + \hat{\beta}_1 + log(D)$$

for isotonic RPEXE, monotonic RPEXE, and BE with $\alpha = 0.05$ or $\alpha = 0.1$, and $D \in [20, 800]$. In practice, $\alpha^\star$ can simply be estimated as

$$\hat{\alpha}^\star = \exp(\hat{\beta}_0) \times D^{\hat{\beta}_1},$$

given that the number of event D is between 20 and 800.

Using a regression line $\log(\hat{\alpha^\star}) = \hat{\beta_0} + \hat{\beta_1} \times log(D)$, $\alpha^\star$ can be estimated for any $D \in [20, 800]$. The estimates in Table 1 lead to regression lines for isotonic RPEXE, monotonic RPEXE, umbrella alternative RPEXE, and BE characterized by $(\hat{\beta_0}, \hat{\beta_1})$ for $\alpha = 0.05$ and $\alpha = 0.1$ in Table 2. The simulation results in Tables 1 and 2 can be used to estimate the critical value $\alpha^\star$ at significance levels $\alpha = 0.05$ and $\alpha = 0.1$ in practice.

4. Method Comparison

In this section, we compare RPEXE with other predictors to show the major benefits and limitations of each survival model, and provide a guide for their practical usage. Before demonstrating the results, we list the survival models used in the simulation, describe the data, and discuss how to quantify the predictive accuracy.

We classify the survival models into three groups. The first group is non-parametric. A typical nonparametric estimator is the Kaplan-Meier estimate. The second group is parametric: exponential, Weibull, lognormal, gamma, and the exponential estimate with type II censoring at the median survival (we use "exp q0.5" to denote this predictor in Figures 3.). The third group is the piecewise exponential estimate including PEXE, monotonic RPEXE (MonoRPEXE), and umbrella alternative RPEXE (UmbRPEXE) of different significant levels $a \in \{0.1, 0.25, 0.5, 0.75, 1\}$.

A useful comparison between different models is based on data sets having known survival probabilities. The tool we adopt to generate survival data sets is the taxonomy of hazard functions in Cox *et al.* [29]. They classified any trend of the failure rate of survival distributions using the scale parameter σ and shape parameter λ of the generalized gamma distribution. Note that "λ" is used to denote the exponential parameter in Section 3, but is used to denote the shape parameter of the generalized gamma distribution in Section 4 (to be consistent with the notation in Cox *et al.* [29]). The failure rate is increasing if $\lambda > \sigma$ and $\lambda < 1/\sigma$, decreasing if $\lambda < \sigma$ and $\lambda > 1/\sigma$, decreasing then increasing if $\lambda > \sigma$ and $\lambda > 1/\sigma$, and increasing then decreasing if $\lambda < \sigma$ and $\lambda < 1/\sigma$. Figure 2 (a) plots the two curves of the functional relations $\lambda = \sigma$ (the gamma line) and $\lambda = 1/\sigma$ (the ammag line) in the two dimensional space of λ vs. σ, with the exponential distribution being their intersection. Using this tool, we can summarize the survivorship of a data set with a point estimate in this space. Recall that Weibull distributions lie on the dot-dash line ($\lambda = 1$) while lognormal distributions lie on the dotted line ($\lambda = 0$).

We generated our data sets with estimates of (σ, λ) from the Surveillance Epidemiology and End Results (SEER) cancer survival database, using data in lung, pancreas, stomach, breast, colon, and melanomas from three decades (i.e., 1975- 1984, 1985–1994, and 1995–2004), for localized, regional, and distant cancers. For each cancer type, decade, and disease extent condition (localized, regional, or distant), we estimated the survival distribution using the generalized gamma distribution. Figure 2 (b) and (c) show the estimated values of (σ, λ) for the six types of cancer. We can see that the estimates are located mainly in three regions. The first region has $\sigma \in [0.3, 0.7]$, $\lambda \in [2, 2.8]$, and close to the ammag line. The second region has λ and σ both close to 1 and $\lambda < 1/\sigma$. The third region has $\sigma \in [1, 2]$ and $\lambda \in [-0.5, 0.5]$, and $\lambda < 1/\sigma$. It is worth noting that none of the cancer survival distribution lies substantially in regions B or C of Figure 2 (a).

We pick four typical locations to compare the predictive accuracy. The locations and their corresponding diseases are 1: (σ, λ) = (0.27, 3.34) of 1995–2004 regional melanomas, 2: (σ, λ) = (1.03, 0.93) of 1995–2004 localized colon cancer, 3: (σ, λ) = (1.41, 0.41) of 1995–2004 distant colon cancer, and 4: (σ, λ) = (−0.34, 1.93) of 1975–1984 regional melanomas. The simulation data are generated from the generalized gamma distribution with sample size 50 using the estimated (σ, λ) and location parameter β fixed at 1. We predict the survival probabilities at times corresponding to the real probabilities at 0.1, 0.2, …, 0.9.

One thousand simulation samples were generated to evaluate the predictive accuracy. We estimated the survival probability using every model for each of the 1000 samples. The censored-exponential model ("exp q0.5" in Figures 3) is the exponential model fitted to data

censored at the sample median. We evaluated the pointwise predictive accuracy using the root mean squared predictive error (RMSPE). For example, suppose the true survival probability at t_0 is $S(t_0)$ and the 1000 predictions are $\{\hat{S}_i(t_0)\}_{i=1,\ldots,1000}$, the RMSPE at t_0 is computed as

$$\sqrt{\frac{\sum_{i=1}^{1000}\left(\widehat{S}_i(t_0)-S(t_0)\right)^2}{1000}}.$$

We illustrate in Figures 3 the RMSPEs from existing approaches (the upper row) and those from the RPEXE models (the lower row). The (σ, λ) pair (0.27, 3.34), associated with Figure 3 (a) and (b), is roughly on the ammag line. The distribution is not near the lognormal, gamma, or exponential distribution locations. Accordingly, their RMSPEs are higher than that of the KME, with the exponential, censored exponential, and lognormal models behaving worst. All RPEXE predictors show small prediction errors that are visually comparable to the Kaplan-Meier curve.

In Figure 3 (c) and (d), $(\sigma, \lambda) = (1.03, 0.93)$ is near (1, 1) implying that the survival distribution is close to exponential. As a result, the exponential distribution has the smallest predictive error among all left panel models. Notable, the censored exponential has much larger RMSPE than the exponential at all survival probabilities. The RPEXE estimates with relatively small a ($a = 0.1$ and 0.25) have the predictive errors similar to the exponential model, with both UmbRPEXE and MonoRPEXE behaving like an exponential distribution (with one piece).

In Figure 3 (e) and (f), $(\sigma, \lambda) = (1.41, 0.41)$ implies that the true distribution is between lognormal and Weibull, and both models have small predictive errors. The single exponential model, however, has very poor predictive error, while the censored-exponential is much better but still high for low survival probabilities. The KME and PEXE models generally have small predictive errors. Figure 3 (f) shows the flexibility of RPEXE in that the RPEXE predictors with moderate-to-large a level ($a \geq 0.5$) have comparable predictive accuracies of the lognormal and KME.

For Figure 3 (g) and (h), the point $(\sigma, \lambda) = (1.93, -0.34)$ is in the area D of Figure 2 (a) with the shape parameter less than 0. In this situation, the only parametric distribution near the point is the lognormal. As a result, in the left panel only the lognormal, KME, and PEXE models have small RMSPEs. In Figure 3 (h), the RPEXE models, however, generally provide small RMSPEs. Similar to the example in Figure 3 (e) and (f), for moderate to large a level ($a \geq 0.5$), the predictive errors from the RPEXE models are comparable to the best predictions of the models in the upper row.

To show flexibility and robustness of each predictor, we summarized for each predictor the ratio (R_{rmspe}) of its maximum predictive error (at all survival probabilities) to the smallest maximum predictive error among all predictors (including RPEXE estimators without order restriction shown as "BeRPEXE") in Table 3. The models are listed according to roughly increasing flexibility, although the monotonic and umbrella RPEXE models have been clustered together and their relative flexibility is somewhat complex. The highest ratio and the root mean squared ratio (of R_{rmspe}) are given in Table 4 with the rank among all models in the parenthesis. We expect a robust model to have a generally moderate ratio at all times. Thus, the three most robust models in Tables 3 and 4 are MonoRPEXE at $a = 0.5$, MonoRPEXE at $a = 0.25$, and UmbRPEXE at $a = 0.5$, and top six most robust models all belong to RPEXE followed by PEXE and KME. The high ranks of the parametric models

confirm that they are least robust and least flexible in modeling cancer survival. We also calculate the ratio from the absolute bias ($R_{|bias|}$) and ratio from the standard deviation of predictions (R_{std}) for each model in Table 3. In Table 5 we calculated the maximum ratio and root mean squared ratio for $R_{|bias|}$ and R_{std} to investigate the bias variance trade-off for each model. We can see that the Kaplan-Meier estimate is the most unbiased estimate because it obtains the smallest bias in Table 3 but has relatively large variation (rank 15 of the 22 models) for both maximum and RMSR of R_{std} in Table 4. Bias and variance of the parametric models strongly depend on whether the model assumption holds. If yes, the parametric models can have small bias and variance compared to other models and thus obtain good predictive accuracy, e.g., exponential and Weibull for $(\sigma, \lambda) = (1.03, 0.93)$ and lognormal for $(\sigma, \lambda) = (1.93, 0.34)$. Generally parametric models have smaller variance in prediction compared with Kaplan-Meier estimate in Table 3. However, bias and variance of a parametric model can be large if the model assumption fails. For example, it can be seen in Table 4 that parametric models have the largest biases and some have relatively large standard deviations. The RPEXE models obtain a balance between the bias and variance. The rankings in both bias and variance are moderate most of the times for RPEXE estimates. Roughly a higher level of a corresponds to less bias but bigger variance. Compared with other exponential or piecewise exponential models, the BeRPEXE with significance level 1.0 is PEXE. In contrast to the large predictive error, bias, and standard deviation of exponential estimate with data censored at median survival, the proposed models have much better performance given the fact that the changepoints are determined by the data and order restrictions.

Within the RPEXE models, the UmbRPEXE models are generally more flexible than the MonoRPEXE models, and the RPEXE with a higher level is more flexible. This property enables the RPEXE predictors to incorporate assumption of the flexibility and expert knowledge about the flexibility and trend of the failure rate. Our experience is that for $a \leq 0.1$, RPEXE can be comparable to the exponential fit when the exponential assumption roughly holds, but can improve the exponential fit if the assumption does not hold. The MonoRPEXE with $a = 0.50$ equals or exceeds KME in all 4 scenarios, with the biggest improvement occurring where the model was close to exponential. If one believes that the survival distribution is close to exponential, one could shift a down to 0.25.

We further compare the predictive accuracy of RPEXE and KME in Figure 4. The two panels in Figure 4 respectively depict the maximum ratio and root mean squared ratio of R_{rmspe} against the level a. The Kaplan-Meier estimator, shown as the dashed line, has a larger relative predictive error than the certain levels of the UmbRPEXE and BeRPEXE ($a = 0.5$ and 0.75), and almost all levels of the MonoRPEXE predictors.

5. Application

5.1. Applying RPEXE in the Lung Cancer Study

In this section we show how RPEXE was applied to the lung cancer clinical trial in Section 2. We implemented isotonic RPEXE because the failure rate was believed to decrease over time. According to the monotonic RPEXE regression line estimate in Table 2, for $a = 0.05$, $a^{\star}$ is about 0.004. Table 5 lists the time points after PAVA in the order of the backward elimination and their corresponding p-values. P-values and actions (second and third columns) in table 5 indicate that the piecewise exponential function has one change-point at $t_{(1)} = 28$. The two exponential parameters are 14.74 and 64.62 before and after 28 months, respectively. Twenty eight (16%) of the patients were alive at month 28.0. Because more than half of the patients died prior to 28 months, the estimated median survival from the fit is $\log(2) \times 14.74 = 10.22$ months.

This analysis provides valuable information for a lung cancer clinician. The failure rate per month for patients having the therapy is 0.068 = 1/14.74 before month 28 and 0.015 = 1/64.62 afterwards. This change suggests that the customized therapy may have substantially extended the survival in twenty eight (16%) of the patients. An alternative model can be a cure mixture model [34, 35]. However, the cure mixture models may suffer from a nonidentifiability issue because the estimated cure rate is strongly related with the choice of the parametric distribution and the estimated model parameters. The proposed RPEXE is free from this issue in that given the numbers of events before and after month 28, $\hat{\lambda_1}$ and $\hat{\lambda_2}$ are independent [30] (page 42).

We further compare RPEXE with alternative parametric and nonparametric models in this example. Figure 2 plots the estimated survival function from KME and from RPEXE. We can see that the RPEXE estimate is less biased than the generalized gamma estimate (in Figure 1) when $t \in (15, 40)$. In this lung cancer clinical trial, the primary endpoints were the median survival time, and one- and two-year survival rates, which could be computed using either KME or RPEXE. Table 6 provides the estimates and confidence intervals. The confidence intervals from KME are obtained by implementing the LIFETEST procedure in SAS, version 9.3 with a loglog transformation. We can see that the two estimates give similar estimates but RPEXE produces tighter confidence intervals for all three endpoints. For example, the improvement of RPEXE over KME in the accuracy of median survival estimation is 36.5% (= 1 − (11.764 − 8.965)/(12.184 − 7.778)).

5.2. Some Other Applications

With the following three examples, we show wide usage of the proposed RPEXE approach.

Example 1. Identifying failure rate change with a small number of events— Ewing Sarcoma/Primitive Neuroectodermal Tumor (EWS/PNET) is an aggressive sarcoma that commonly affects children and young adults. Bui *et al*. [36] recruited 36 EWS/PNET patients to test the association between the pathological interpretation of biomarker Cx43 and patient overall survival, where Cx43 is one of Connexins homogeneous protein that had been shown to be associated with lung, breast, prostate, liver, stomach, and colon cancers. Among the 36 patients, 8 had a negative Cx43 interpretation and the rest 28 had a positive interpretation. Only 2 of the 8 Cx43 negative patients were found dead. There were research interests in testing if the hazard rates changed in both groups. With such a small number of events in the Cx43 negative group, asymptotic tests can not be applied. The p-value (with monotonic order restriction) from the UMPU test was 0.045, which is significant ($\alpha = \alpha^{\star} =$ 0.05 for $d = 2$). Thus, even with a small number of data, we are able to detect violation of exponentiality.

Example 2. Improving estimation and test power in a Phase II study—Mahmood *et al*. [37] investigated the safety and efficacy of sunitinib (an inhibitor for soft tissue sarcomas) in 3 common histological subtypes: “Lipo” for liposarcoma, “Leio” for leiomyosarcoma, and “MFH” for malignant fibrous histocytoma, which had numbers of events 18, 15, and 14, respectively. They were interested in knowing whether patients in the three histological types have different survival times. After verifying the exponential distribution assumption in all three groups, point and interval estimates of the median survival were shown in Table 3 of Mahmood *et al*. [37]. The 95% confidence intervals of the exponential estimates were almost all tighter than those of the Kaplan-Meier estimates, which was consistent with the claim in Miller [1] that exponential estimates have better precision. We further compared the failure rates in the three groups using the likelihood ratio (UMPU) test. The p-values were computed as 0.77, 0.42, and 0.60 for Lipo vs. Leio, Lipo vs. MFH, and Leio vs. MFH comparisons, respectively. Comparing with p-values from the

log-rank test (0.74 for Lipo vs. Leio, 0.87 for Lipo vs. MFH, and 0.92 for Leio vs. MFH), we can see that the p-values from the UMPU test are either comparable or much smaller. Although both tests suggested that patients' survival in the three histological types were not significantly different, the discrepancy in p-values from the two tests may indicate that with the number of events around 15, log-rank test could be less powerful or too conservative compared with the UMPU test. Using UMPU test could achieve the same power with a smaller sample size.

Example 3. Engaging the exponential saving in single-arm Phase II trial designs—Schell *et al*. [38] compared exponential and Kaplan-Meier approaches in the power analysis of single arm clinical trial designs with a survival endpoint. Assuming that exponential distribution was valid up to a time, they computed sample sizes required for testing the survival probability with power 0.9, type I error 0.1, and presumed hazard ratios (which equals the ratio of the median survival under the alternative divided and the median survival under the null). Based on their computation, the sample size saving could be substantial (30% to 40%) even if the distribution was exponential through a time less or equal to twice of the median survival. The primary concern of using exponential estimates in practice has been the lack of theoretical justification of exponential assumption [2]. The proposed RPEXE approach is thus a favorable tool for validating the exponentiality through the study period or up to a certain time in the trial in that the corresponding exponential based power analysis could lead to tremendous sample size saving in a single-arm Phase II trial design.

6. Discussion

Landmark estimates, such as the median and 1-, 2- to 5-year estimates, seem to etch themselves into the minds of clinicians as summary measures of the effectiveness of a given therapy. Thus, it is important that they be as accurate as possible, both for treatment decision making and for possible planning for future trials with survival as a key endpoint. The nonparametric Kaplan-Meier method retains full flexibility, and in an unspecified context, such an estimator seems ideal. However, total adherence to the KME can come at a cost, as Ruppert Miller aptly said in Miller [1] and responded by Meier *et al*. [2]. Statisticians typically pull the estimates from the fitted relationship rather than from individual independent variable value when fitting a linear relationship. Similarly, we believe that it is important to identify and take advantage of the the parametric models in survival analysis. We note that the proposed RPEXE fit accomplishes these desiderata by being built on the three components: an exact likelihood ratio test, a backward elimination procedure, and an optional order restriction. Based on our simulation studies and real examples, the RPEXE approach can provide a continuous, parsimonious, and data-adaptive estimate of the survival function. It can guide practitioner to fully take into account exponential fitting's benefits for any survival data, i.e., 1, saving financial and patient resources, and 2, accelerating the approval of agents to make the proposed trial design ethically attractive. As showed in Sections 4 and 5, the proposed method routinely matched or exceeded the estimation performance of the Kaplan-Meier estimator and other parametric fits for some SEER survival data and clinical trial data.

In cancer research practice, we believe that RPEXE with an order restriction will be more useful than the unrestricted model. One issue with the unrestricted model sometimes is the identification of small time intervals where multiple failures occur in situations where short-lived hazard spikes are not thought to be scientifically reasonable. For example, in the analysis of the NSCLC data in Section 5, when no order restriction is imposed, the change-point in this application was at month 18 rather than at month 28 largely because events occurred in two consecutive days in month 18 and thus the estimated failure rate increases

and then decreases dramatically there, showing a spike in the plot against time. Moreover, an order restricted RPEXE model imposes more structure, which can result in greater gains compared to the Kaplan-Meier estimator. The RPEXE with monotonic order restriction is the counterpart in survival analysis of the monotonic regression model proposed by Schell *et al*. [26]. Given the results presented here, we recommend that RPEXE with the monotonic order restriction and $\alpha = 0.50$ be used routinely for estimating the median survival time and other cancer survival landmark values, and that RPEXE with $\alpha = 0.10$ and a proper order restriction be used to validate a time point up to which the exponential assumption holds.

The proposed RPEXE approach has its scope of usage. First, unlike other flexible survival models, e.g., spline-based modeling [39, 40], in this paper we focus on summarizing and interpreting the survival distribution by estimating hazard rates and the intervals over which they obtain. Secondly, the RPEXE approach is not developed for modeling covariates. It can be used to describe, visualize, and interpret the survival of all patients or patient subgroups in a cancer clinical trial. Lastly, RPEXE is more suitable for analyzing cancer clinical trial data where the sample size is typically small, and in modeling all plausible change-points. For large data sets such as national registry data, Goodman *et al*. [8] is a proper approach if the goal is to build a parsimonious model. The RPEXE software with examples is available upon request.

Acknowledgments

The authors would like to thank Gerold Bepler and George R. Simon for providing the non-small cell lung cancer data. We thank the editor, associate editor, and two reviewers for their constructive comments that improved the quality of this article. This research was sponsored, in part, by the National Institute of Health and the National Cancer Institute, grant 1RC2CA14833201.

References

1. Miller RG. What price kaplan-meier? Biometrics. 1983; 39:1077–1081. [PubMed: 6671119]
2. Meier P, Karrison T, Chappell R, Xie H. The price of kaplan-meier. Journal of the American Statistical Association. 2004; 99:890–896.
3. Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, Ratain MJ, Blanc ML, Stewart D, Crowley J, et al. The design of phase ii clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. Clinical Cancer Research. 2010; 16:1764–1769. [PubMed: 20215557]
4. Berry SM, Berry DA, Natarajan K, Lin CS, Hennekens CH, Belder R. Bayesian survival analysis with nonproportional hazards: Meta-analysis of pravastatin-aspirin. Journal of the American Statistical Association. 2004; 99:36–44.
5. Berry DA. The hazards of endpoints. Journal of the National Cancer Institute. 2010; 102:1376–1377. [PubMed: 20826734]
6. Hoos A, Eggermont AMM, Janetzki S, Hodi FS, Ibrahim R, Anderson A, Humphrey R, Blumenstein B, Old L, Wolchok J. Improved endpoints for cancer immunotherapy trials. Journal of the National Cancer Institute. 2010; 102:1388–1397. [PubMed: 20826737]
7. Edwards BK, Brown ML, Wingo PA, Howe HL, Ward E, Ries LAG, Schrag D, Jamison PM, Jemel A, Wu XC, et al. Annual report to the nation on the status of cancer, 1975–2002, featuring population-based trends in cancer treatment. Journal of the National Cancer Institute. 2005; 97:1407–1427. [PubMed: 16204691]
8. Goodman MS, Lib Y, Tiwari RC. Detecting multiple change points in piecewise constant hazard functions. Journal of Applied Statistics. 2011; 38:2523–2532. [PubMed: 22707842]
9. Breslow N. Covariance analysis of censored survival data. Biometrics. 1974; 30:34–41.
10. Aitkin M, Laird N, Francis B. A reanalysis of the stanford heart transplant data. Journal of the American Statistical Association. 1983; 78:264–292.

11. Clark DE, Ryan LM. Concurrent prediction of hospital mortality and length of stay from risk factors on admission. Health Services Research. 2002; 37:631–645. [PubMed: 12132598]
12. Matthews DE, Farewell VT. On testing for a constant hazard against a change-point alternative. Biometrics. 1982; 38:463–468. [PubMed: 7052152]
13. Nguyen GS, Rogers GS, Walker EA. Estimation in change-point hazard rate models. Biometrika. 1984; 71:299–304.
14. Matthews DE, Farewell VT. On a singularity in the likelihood for a change-point hazard rate model. Biometrika. 1985; 72:703–704.
15. Yao YC. Maximum likelihood estimation in hazard rate models with a change-point. Communications in Statistics: Theory and Methods. 1986; 15:2455–2466.
16. Raftery A, Akman V. Bayesian analysis of a poisson process with a change-point. Biometrika. 1986; 73:85–89.
17. Müller HG, Wang JL. Nonparametric analysis of changes in hazard rates for censored survival data: an alternative to change-point problems. Biometrika. 1990; 77:305–314.
18. Loader CR. Inference for a hazard rate change point. Biometrika. 1991; 78:749–757.
19. Arani RB, Soong S, Weiss HL, Wood MJ, Fiddian PA, Gnann JW, Whitley R. Phase specific analysis of herpes zoster associated pain data: a new statistical approach. Statistics in Medicine. 2001; 20:2429–2439. [PubMed: 11512133]
20. Martinussen T. Aalen additive hazards change-point model. Biometrika. 2007; 94:861–872.
21. Demarqui FN, Loschi RH, Colosimo EA. Estimating the grid of time-points for the piecewise exponential model. Lifetime Data Analysis. 2008; 14:333–356. [PubMed: 18463801]
22. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models, issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996; 15:361–387. [PubMed: 8668867]
23. Bray, TA.; Crawford, GB.; Proschan, F. Technical Report 01-82-0660. NTIS; Springfield, Va: 1967. Maximum likelihood estimation of a u-shaped failure rate function. Mathematical Note No. 534, Mathematics Research Laboratory, Boeing Scientific Research Laboratories, Available fromAD66-3678
24. Kim JA, Proschan F. Piecewise exponential estimator of the survivor function. Institute of Electrical and Electronics Engineers Transactions on Reliability. 1991; 40:134–139.
25. Han G, Schell MJ, Kim J. Comparing two exponential distributions using the exact likelihood ratio test. Statistics in Biopharmaceutical Research. 2012; 4:348–356. [PubMed: 23814641]
26. Schell MJ, Singh B. The reduced monotonic regression method. Journal of the American Statistical Association. 1997; 92:123–137.
27. Simon GR, Turrissi A. Management of small cell lung cancer: Accp evidence-based clinical practice guidelines. Chest. 2007; 132:324S–339S. [PubMed: 17873178]
28. Simon GR, Extermann M, Chiappori A, Williams C, Begum M, Haura RKE, Ismail-Khan R, Schell M, Antonia SJ, Bepler G. Phase 2 trial of docetaxel and gefitinib in the first-line treatment of patients with advanced stage non-small cell lung cancer (nsclc) who are 70 years of age or older. Cancer. 2008; 112:2021–2029. [PubMed: 18300255]
29. Cox C, Chu H, Schneider MF, Munoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Statistics in Medicine. 2007; 26:4352–4374. [PubMed: 17342754]
30. Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. Wiley; New York: 1980.
31. Epstein B, Sobel M. Some theorems relevant to life testing from an exponential distribution. Annals of Mathematical Statistics. 1954; 25:373–381.
32. Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E. An empirical distribution function for sampling with incomplete information. The Annals of Mathematical Statistics. 1955; 26:641–647.
33. Robertson, T.; Wright, FT.; Dykstra, RL. Order-restricted statistical inference. Wiley; New York: 1988.
34. Chen MH, Ibrahim JG. Maximum likelihood methods for cure rate models with missing covariates. Biometrics. 2001; 57:43–52. [PubMed: 11252617]

35. Yu B, Peng Y. Mixture cure models for multivariate survival data. Computational Statistics & Data Analysis. 2008; 52:1524–1532.

36. Bui M, Han G, Acs G, Gonzalez R, Reed D, Pasha IL, Zhang P. Connexin 43 is a potential prognostic biomarker for ewing sarcoma (ews)/primitive neuroectodermal tumor (pnet). Sarcoma. 2011:Article ID 971050.10.1155/2011/971050.

37. Mahmood ST, Agresta S, Vigil C, Zhao X, Han G, D'Amato G, Calitri CE, Dean M, Garrett C, Schell MJ, et al. Phase ii study of sunitinib malate, a multi-targeted tyrosine kinase inhibitor in patients with relapsed or refractory soft tissue sarcomas. focus on 3 prevalent histologies: leiomyosarcoma, liposarcoma, and malignant fibrous histiocytoma. The International Journal of Cancer. 2011; 129:1963–1969.

38. Schell, MJ.; Haura, E.; Kim, J.; Han, G.; Simon, G. A novel approach to increase the efficiency of survival-endpoint phase 2 trials. Submitted

39. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. Journal of the American Statistical Association. 1992; 87:942–951.

40. Abrahamowicz M, MacKenzie T, Esdaile JM. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. Journal of the American Statistical Association. 1996; 91:1432–1439.

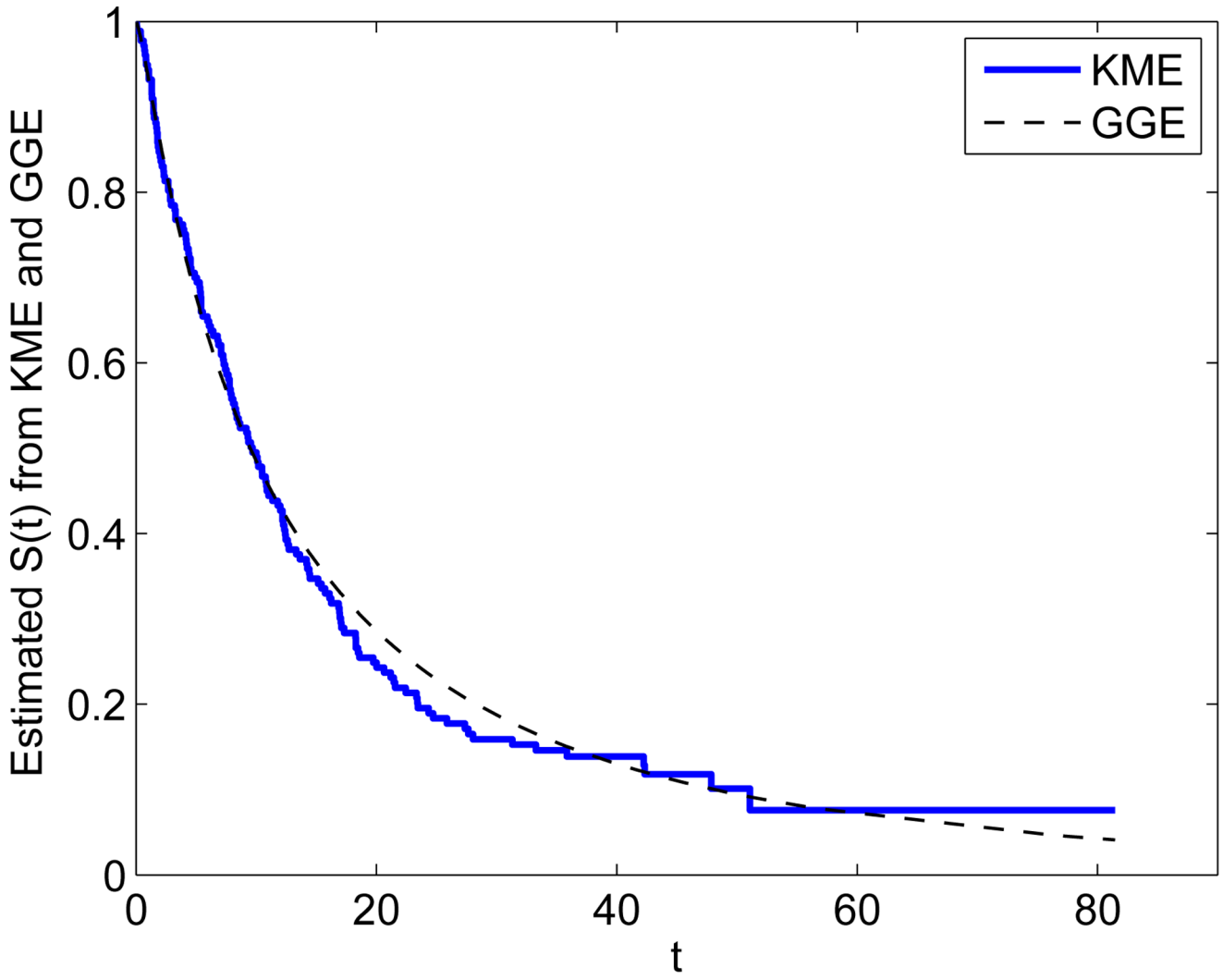

Figure 1.
Plots of estimated survival function from the KME (– – – – – – –) and GGE (- - - -).

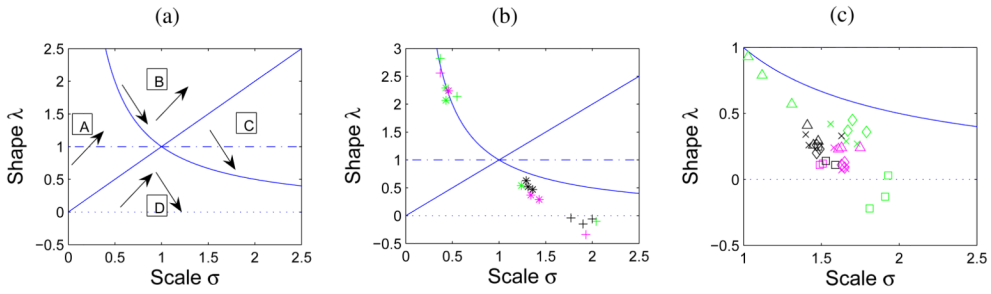

Figure 2.

(a) Taxonomy of the scale (σ) and shape (λ) parameters of the generalized gamma distribution (modified from Cox et al. (2007)). The arrows indicate trend of the failure rate in regions A–D. (b) Estimates of melanomas (+) and breast cancer (*). (c) Estimates of lung (×) stomach (◇), pancreas (□) and colon (△) cancers. Green color corresponds to localized cancer, pink color to regional cancer, and black color to distant cancer.

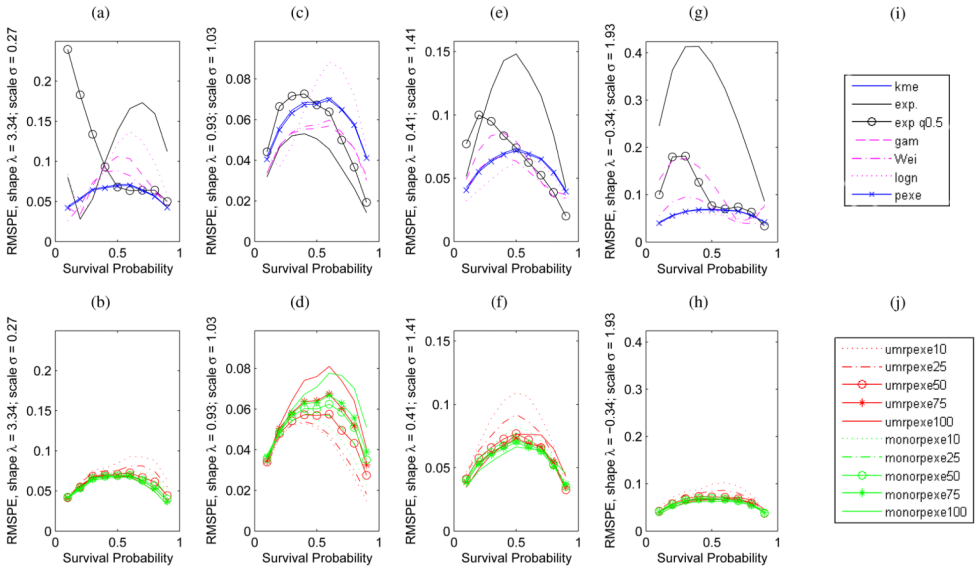

Figure 3.

(a)–(h): Plots of RMSPE vs. estimated survival probability at the four locations of (σ, λ) where the upper row shows the RMSPEs of existing estimates, and the lower row shows the RMSPEs of proposed estimates. (i) and (j) display symbols, where “kme” is Kaplan-Meier estimate; “exp.”, “exp q0.5”, “gam”, “Wei”, “logn”, and “pexe” are exponential, exponential with data censored at the median survival, gamma, Weibull, lognormal, and pexe estimates; “umrpexe10”-”umrpexe100” and “monorpexe10”-”monorpexe100” are umbrella alternative RPEXE and monotonic RPEXE estimates, respectively, with significant level α at 10%, 25%, 50%, 75%, and 100%.

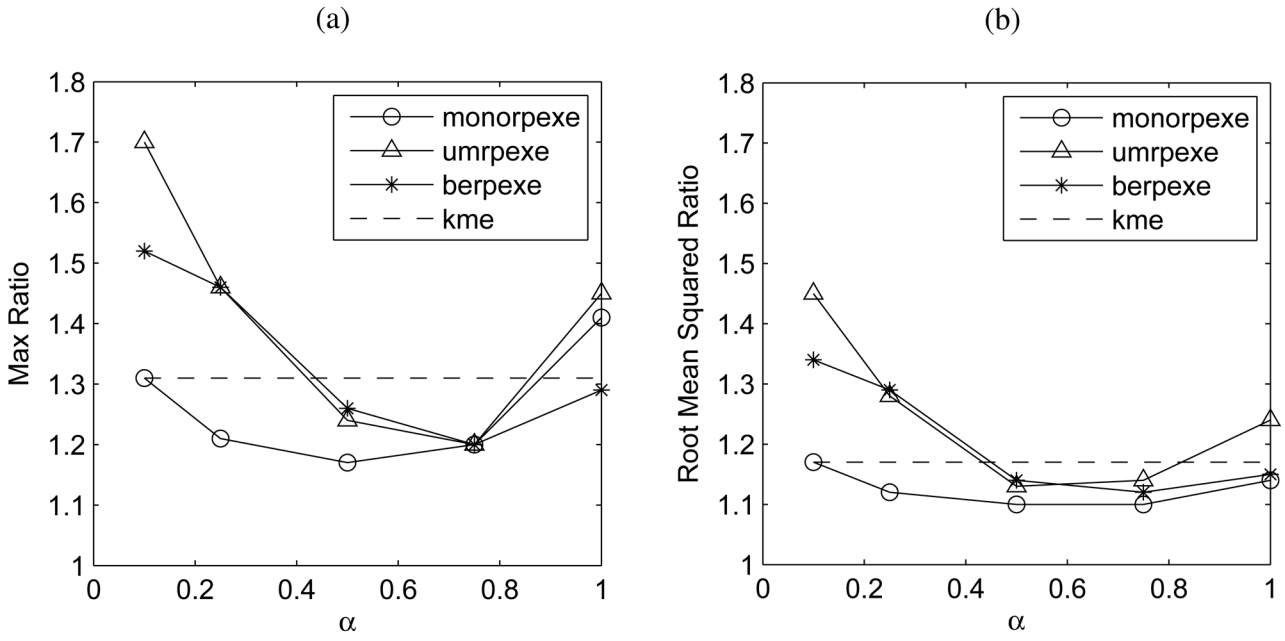

Figure 4.

(a) Max ratio (of R_{rmspe}) vs. α level of the RPEXE estimators and KME. (b) RMSR (of R_{rmspe}) vs. α level of the RPEXE estimators and KME.

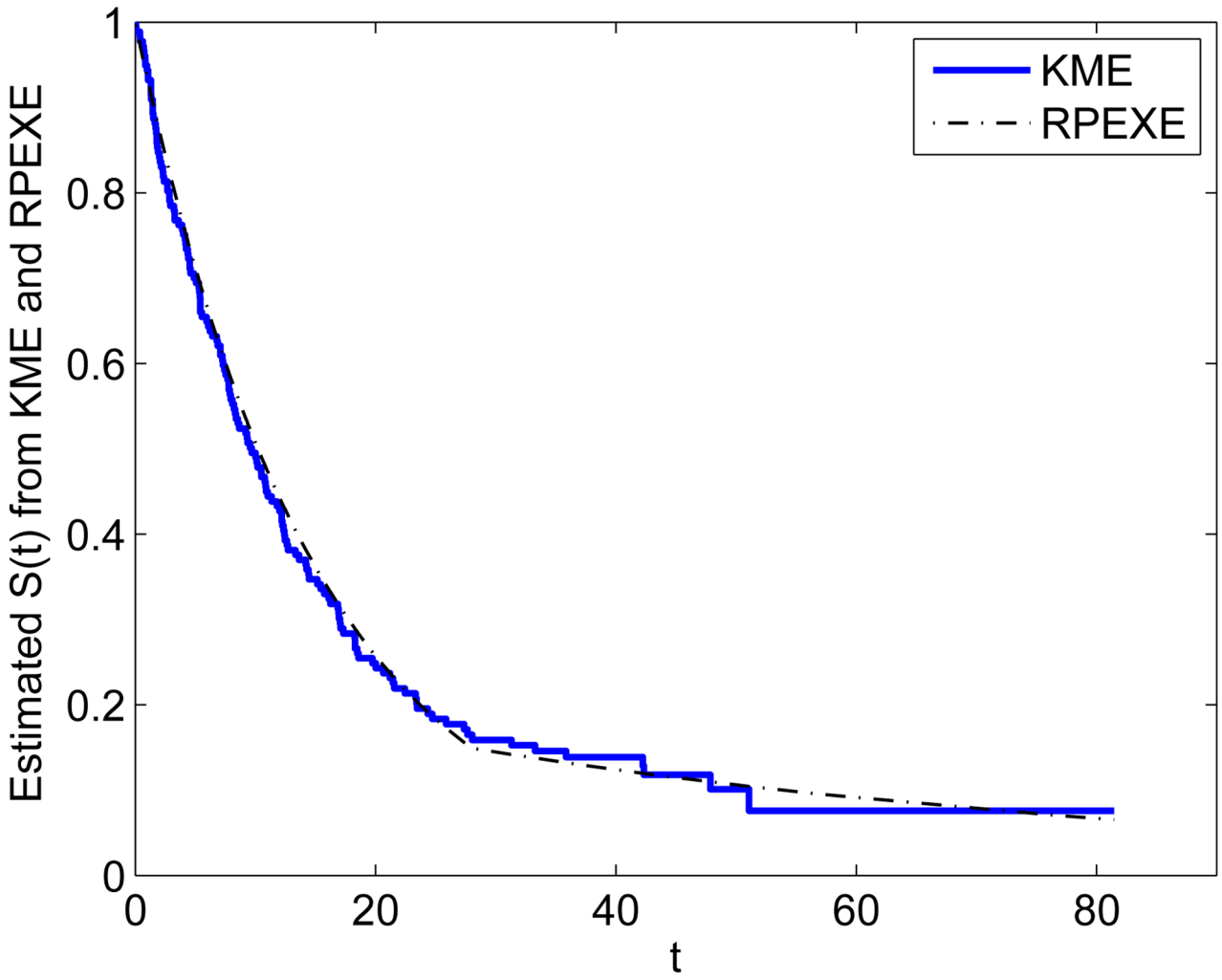

Figure 5.
Plots of estimated survival function from the KME (– – – – – – –) and RPEXE (–·–·–·).

Table 1

Estimates and 95% confidence limit differences (in the parentheses) for $\alpha^{\star}$ corresponding to $\alpha = 0.05$ and $\alpha = 0.1$ from isotonic RPEXE, monotonic RPEXE, and BE for $D = 20$, 50, 200, and 800 with $N_s = 100,000$. Adding and subtracting the difference from each estimate comprise at least a 95% confidence interval of $\alpha^{\star}$.

	Isotonic RPEXE	Monotonic RPEXE	Umbrella RPEXE	BE, no order restriction
$D = 20$, $\alpha = 0.05$	.0107 (.0003)	.0048 (.0002)	0.0017 (.0001)	.0016 (.0001)
$D = 50$, $\alpha = 0.05$	.0065 (.0002)	.0029 (.0001)	0.00055 (.00003)	.00044 (.00002)
$D = 200$, $\alpha = 0.05$	.0038 (.0001)	.0017 (.0001)	0.00010 (0.00001)	.00008 (< .00001)
$D = 800$, $\alpha = 0.05$	.0026 (.0001)	.0011 (.0001)	0.00002 (< 0.00001)	.00001 (< .00001)
$D = 20$, $\alpha = 0.1$	.0244 (.0006)	.0108 (.0002)	0.0037 (0.0002)	.0034 (.0001)
$D = 50$, $\alpha = 0.1$	.0152 (.0004)	.0067 (.0001)	0.0012 (.0001)	.0010 (.00002)
$D = 200$, $\alpha = 0.1$	.0090 (.0002)	.0038 (.0001)	0.00024 (.00001)	.00020 (< .00001)
$D = 800$, $\alpha = 0.1$	.0061 (.0001)	.0026 (.0001)	0.00005 (< .00001)	.00002 (< .00001)

Table 2

Estimated linear coefficients $(\hat{\beta_0}, \hat{\beta_1})$ of the isotonic RPEXE, monotonic RPEXE, umbrella alternative RPEXE, and BE for $\alpha = 0.05$ and $\alpha = 0.1$.

	Isotonic RPEXE	Monotonic RPEXE	Umbrella RPEXE	BE, no order restriction
$\alpha = 0.05$	(−3.483, −0.380)	(−4.233, −0.394)	(−1.223, −2.704)	(−2.356, −1.360)
$\alpha = 0.1$	(−2.670, −0.372)	(−3.448, −0.385)	(−1.195, −2.024)	(−1.511, −1.370)

Table 3

Ratios (denoted by R_{rmspe}, $R_{|bias|}$, and R_{std}) between each model's maximum RMSPE (or absolute bias or standard deviation) and the smallest maximum RMSPE (or absolute bias or standard deviation) among all the models for the four settings of (σ, λ) in Figure 3. The "BeRPEXE" denotes the RPEXE estimator with no order restrictions.

	Ratios for (σ, λ) = (0.27, 3.34)			**Ratios for (σ, λ) = (1.03, 0.93)**			**Ratios for (σ, λ) = (1.41, 0.41)**			**Ratios for (σ, λ) = (1.93, 0.34)**		
Model	R_{rmspe}	$R_{\|bias\|}$	R_{std}	R_{rmspe}	$R_{\|bias\|}$	R_{std}	R_{rmspe}	$R_{\|bias\|}$	R_{std}	R_{rmspe}	$R_{\|bias\|}$	R_{std}
exponential	2.59	43.60	1.00	1.00	1.88	1.00	2.23	35.45	1.44	6.95	106.81	3.77
exp. q0.5	3.54	57.71	2.39	1.36	1.16	1.37	1.55	23.71	1.26	3.04	47.69	1.33
gamma	1.57	23.83	2.22	1.02	2.09	1.02	1.31	17.58	1.09	2.96	43.43	1.84
Weibull	1.30	17.76	2.42	1.06	1.67	1.06	1.02	10.25	1.00	1.58	19.87	1.16
lognormal	2.03	31.38	2.72	1.58	19.38	1.22	1.00	7.87	1.01	1.00	5.22	1.00
MonoRPEXE 10	1.11	7.80	2.44	1.03	1.39	1.04	1.21	5.02	1.29	1.31	4.67	1.37
MonoRPEXE 25	1.06	5.30	2.39	1.09	2.14	1.09	1.13	2.65	1.22	1.21	2.52	1.27
MonoRPEXE 50	1.02	3.27	2.36	1.14	2.73	1.14	1.07	2.80	1.16	1.17	2.16	1.23
MonoRPEXE 75	1.01	1.57	2.34	1.20	1.38	1.21	1.04	4.38	1.13	1.14	3.75	1.19
MonoRPEXE 100	1.00	3.87	2.30	1.41	4.53	1.41	1.02	9.16	1.04	1.07	7.87	1.10
UmbRPEXE 10	1.36	14.53	2.70	1.00	1.80	1.00	1.64	18.88	1.43	1.70	14.06	1.61
UmbRPEXE 25	1.18	10.24	2.52	1.02	1.58	1.03	1.39	10.74	1.37	1.46	8.60	1.47
UmbRPEXE 50	1.07	5.14	2.42	1.06	1.60	1.07	1.14	3.35	1.23	1.24	3.39	1.30
UmbRPEXE 75	1.05	2.37	2.41	1.20	2.32	1.20	1.13	2.25	1.23	1.17	2.00	1.30
UmbRPEXE 100	1.05	3.82	2.42	1.45	4.90	1.07	1.23	6.13	1.23	1.19	5.40	1.30
BeRPEXE 10	1.26	13.31	2.51	1.01	1.70	1.02	1.49	14.54	1.38	1.52	10.96	1.50
BeRPEXE 25	1.21	12.26	2.46	1.02	1.66	1.02	1.43	12.49	1.37	1.46	8.92	1.46
BeRPEXE 50	1.09	6.52	2.41	1.08	1.92	1.08	1.14	3.88	1.22	1.26	3.68	1.32
BeRPEXE 75	1.02	3.65	2.34	1.20	2.90	1.21	1.08	1.89	1.17	1.16	1.67	1.22
BeRPEXE 100	1.03	2.12	2.39	1.29	3.10	1.29	1.09	2.46	1.18	1.16	2.44	1.21
pexe	1.03	2.12	2.39	1.29	3.10	1.29	1.09	2.46	1.18	1.16	2.44	1.21
Kaplan-Meier	1.06	1.00	2.44	1.31	1.00	1.32	1.11	1.00	1.21	1.17	1.00	1.23

Stat Med. Author manuscript; available in PMC 2015 January 15.

Table 4

Maximum ratios (L_∞ penalty) and root mean squared ratios (RMSR) (L_2 penalty) for R_{rmspe}, $R_{|bias|}$, and R_{std}) of each model with ranks from low to high.

	The Maximum and RMSR of R_{rmspe}		The Maximum and RMSR of $R_{\|bias\|}$		The Maximum and RMSR of R_{std}	
Model	**Max ratio (rank)**	**RMSR (rank)**	**Max ratio (rank)**	**RMSR (rank)**	**Max ratio (rank)**	**RMSR (rank)**
exponential	6.95 (22)	3.91 (22)	106.81 (22)	60.35 (22)	3.77 (22)	2.14 (22)
exp. q0.5	3.54 (21)	2.55 (21)	57.71 (21)	39.27 (21)	2.39 (9)	1.65 (17)
gamma	2.96 (20)	1.87 (20)	43.43 (20)	26.31 (20)	2.21 (1)	1.62 (13)
Weibull	1.58 (17)	1.26 (14)	19.87 (18)	14.30 (18)	2.42 (12)	1.53 (1)
lognormal	2.03 (19)	1.47 (19)	31.38 (19)	19.03 (19)	2.72 (21)	1.65 (16)
MonoRPEXE 10	1.31 (11)	1.17 (12)	7.80 (12)	5.24 (12)	2.44 (16)	1.62 (14)
MonoRPEXE 25	1.21 (5)	1.12 (4)	5.30 (9)	3.39 (8)	2.39 (8)	1.58 (6)
MonoRPEXE 50	1.17 (1)	1.10 (1)	3.27 (5)	2.77 (6)	2.36 (5)	1.56 (4)
MonoRPEXE 75	1.20 (2)	1.10 (2)	4.38 (7)	3.38 (7)	2.34 (3)	1.55 (3)
MonoRPEXE 100	1.41 (12)	1.14 (6)	9.16 (13)	6.73 (13)	2.30 (2)	1.55 (2)
UmbRPEXE 10	1.70 (18)	1.45 (18)	18.88 (17)	13.86 (17)	2.70 (20)	1.80 (21)
UmbRPEXE 25	1.46 (14)	1.28 (15)	10.74 (14)	8.61 (14)	2.52 (19)	1.69 (19)
UmbRPEXE 50	1.24 (6)	1.13 (5)	5.14 (8)	3.60 (9)	2.42 (13)	1.60 (8)
UmbRPEXE 75	1.20 (3)	1.14 (7)	2.37 (2)	2.24 (2)	2.41 (11)	1.60 (12)
UmbRPEXE 100	1.45 (13)	1.24 (13)	6.13 (10)	5.13 (11)	2.42 (13)	1.60 (8)
BeRPEXE 10	1.52 (16)	1.34 (17)	14.54 (16)	11.31 (16)	2.51 (18)	1.69 (20)
BeRPEXE 25	1.46 (15)	1.29 (16)	12.49 (15)	9.85 (15)	2.46 (17)	1.67 (18)
BeRPEXE 50	1.26 (7)	1.14 (8)	6.52 (11)	4.32 (10)	2.41 (10)	1.60 (7)
BeRPEXE 75	1.20 (4)	1.12 (3)	3.65 (6)	2.65 (5)	2.34 (4)	1.57 (5)
BeRPEXE 100	1.29 (8)	1.15 (9)	3.10 (3)	2.56 (3)	2.39 (6)	1.60 (10)
pexe	1.29 (8)	1.15 (9)	3.10 (3)	2.56 (3)	2.39 (6)	1.60 (10)
Kaplan-Meier	1.31 (10)	1.17 (11)	1.00 (1)	1.00 (1)	2.44 (15)	1.63 (15)

Stat Med. Author manuscript; available in PMC 2015 January 15.

Table 5

The time points in backward elimination of isotonic RPEXE, the corresponding p-values, and the actions.

Time	P-value	Action
1.9	0.978 (Not Significant)	Eliminate
23.5	0.961 (Not Significant)	Eliminate
47.9	0.754 (Not Significant)	Eliminate
12.7	0.705 (Not Significant)	Eliminate
24.7	0.688 (Not Significant)	Eliminate
0.1	0.601 (Not Significant)	Eliminate
2.4	0.306 (Not Significant)	Eliminate
18.6	0.085 (Not Significant)	Eliminate
28.0	< 0.001 (Significant)	Regarded as the change-point

Table 6

Estimates and the confidence intervals of median survival time and 1 and 2 year survival rates from KME and RPEXE.

	Median Survival Time		One Year Survival Prob.		Two Year Survival Prob.	
	Estimates	**Interval**	**Estimates**	**Interval**	**Estimates**	**Interval**
KME	9.685	(7.778, 12.184)	0.431	(0.357, 0.505)	0.195	(0.135, 0.255)
REPEXE	10.215	(8.965, 11.764)	0.443	(0.395, 0.493)	0.196	(0.156, 0.243)