Research Article

Statistics in Medicine

Received 28 January 2014, Accepted 23 December 2014 Published online 28 January 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6431

Summarising censored survival data using the mean residual life function

Alberto Alvarez-Iglesias,[a*†] John Newell,[a] Carl Scarrott[b] and John Hinde[a,c]

The mean residual life function provides a clear and simple summary of the effect of a treatment or a risk factor in units of time, avoiding hazard ratios or probability scales, which require careful interpretation. Estimation of the mean residual life is complicated by the upper tail of the survival distribution not being observed as, for example, patients may still be alive at the end of the follow-up period. Various approaches have been developed to estimate the mean residual life in the presence of such right censoring. In this work, a novel semi-parametric method that combines existing non-parametric methods and an extreme value tail model is presented, where the limited sample information in the tail (prior to study termination) is used to estimate the upper tail behaviour. This approach will be demonstrated with simulated and real-life examples. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: extreme value theory; generalised Pareto distribution; mean residual life; survival analysis

1. Introduction

Time to event (survival) data are found in many different disciplines such as medicine, biology and epidemiology and in reliability studies in engineering. The survivor function plays a key role in the summary and analysis of survival data. This function gives the probability of the event of interest happening beyond any particular time point, and estimates of this function have been used as graphical summaries for decades, especially in the analysis of medical data. Less known in the biomedical context, however, is the use of the mean residual life (MRL) function, which has been used traditionally for reliability problems. The main advantage of this function is that it summarises the information in units of time (and not as a probability), which can be more relevant if one wants to convey the survival information to a patient. Although estimates of the survivor function can be easily obtained even in the presence of censoring, the same is not true for estimates of the MRL function if the censoring is due to the termination of the study (type I censoring).

In this paper, a new method is proposed for the estimation of the MRL function $m(t)$ under random right censoring, which can include type I censoring. We propose to split the estimation of $m(t)$ into two parts. The first part is based on the estimation of the restricted mean residual lifetime from time t to a threshold u that is close to the maximum observed event time T^*. The estimate of the restricted MRL function is obtained by ignoring the area under the Kaplan–Meier estimate of the survivor function from u to ∞. In the second part, the estimate of $m(u)$ uses extreme value theory, in particular the generalized Pareto distribution (GPD), to model the upper tail arising from longer-term survivors. The generalised Pareto distribution is justified by results from extreme value theory and includes the upper tail behaviour of all the usual parametric models used for lifetimes as a special case, for example, exponential, Weibull and negative Weibull, gamma, chi-squared, normal and Frèchet.

[a] *HRB Clinical Research Facility, National University of Ireland Galway, Galway, Ireland*
[b] *School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
[c] *School of Mathematics, Statistics and Applied Mathematics, NUI Galway, Galway, Ireland*
**Correspondence to: Alberto Alvarez-Iglesias, HRB Clinical Research Facility, National University of Ireland Galway, University Road, Galway, Ireland.*
[†] *E-mail: alberto.alvarez-iglesias@nuigalway.ie*

The remainder of this paper is organised as follows: in Section 2, the MRL function is introduced, and a review of some of the methods to estimate the MRL function is given. In Section 3, the proposed method to estimate the MRL is introduced and the estimation procedure detailed. In Section 4, the results of a simulation study to assess the performance of the new estimator are presented. Finally, in Section 5, the proposed estimator is applied to real-life data.

2. Mean residual life function estimation

The MRL function at time t is defined as

$$m(t) = E(T - t \mid T > t) = \frac{1}{S(t)} \int_t^{\infty} S(s)ds \tag{1}$$

where the random variable T is the time until the event of interest occurs, and $S(t) = P(T > t)$ is the survivor function. The MRL function can be interpreted as the remaining expected lifetime given survival up to time t.

In Figure 1, an illustrative example is given for a hypothetical intervention and control population under different survival modes, namely stable (e.g. constant hazard), recovery (e.g. hazard decreasing over time), terminal (e.g. increasing hazard) and infectious (initial high hazard decreasing over time). For each mode, the corresponding survivor, hazard and mean residual functions are displayed. The benefit of the MRL function is that, in addition to the function uniquely determining the distribution of the survival times, it can also be used as an alternative to the survivor or hazard functions to provide a summary of the effectiveness of the intervention in units of time, rather than as a difference (or ratio) in probabilities or hazards.

One of the challenges in survival analysis is the estimation of the MRL function, under non-informative right censoring. Estimates can be obtained in different ways, mostly by substituting appropriate estimates of the survivor function in (1). Yang [1] proposed a method for the estimation of the MRL function involving an empirical distribution function related to the product-limit formula of [2]. Gill [3] studied the asymptotic properties of the Kaplan–Meier estimator of the survivor function and applied the results to the estimation of the mean lifetime. Ruiz and Guillamón [4] derived non-parametric estimates of the numerator in (1) based on kernel density estimators and estimated the denominator using the empirical

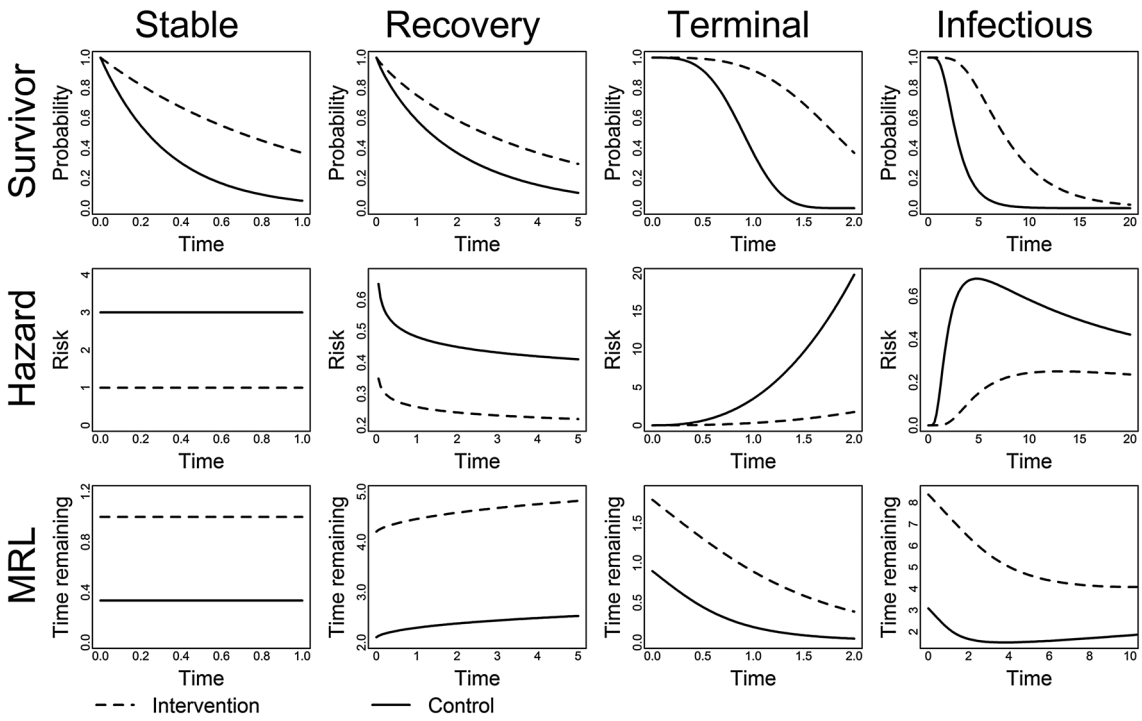

Figure 1. Graphical summaries of different hypothetical survival modes. MRL, mean residual life.

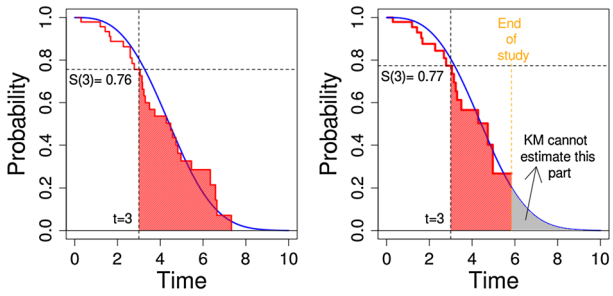

Figure 2. Example of data simulated under right random censoring (left) and under type I censoring in addition to right random censoring (right). KM, Kaplan–Meier.

survival function for the complete data case. Chaubey and Sen [5] considered a different smoothing approach for the complete data case based on the classical Hille theorem [6], which can be used to obtain smooth estimators of $S(t)$ in (1). They used this approach to obtain a smooth estimator of the MRL function and studied its asymptotic properties. More recently, Chaubey and Sen [7] have extended these results to the case of right censored data. Zhou and Jeong [8] estimated the MRL function at time t using the non-parametric maximum likelihood estimator based on the Kaplan–Meier estimator and derived confidence intervals using the general empirical likelihood ratio test.

The methods described above do not provide appropriate estimates of the MRL function when observations are censored at the maximum observed time. This is the case in many medical studies that are carried out for a limited period of time, where for some patients the event of interest does not occur before the end of the follow-up period. This type of censoring is known as type I censoring in reliability studies. Figure 2 illustrates an example in which data were simulated under right random censoring (left) and under right random censoring in addition to type I censoring (right).

As Shen, Xie and Tang [9] note, only a few papers focus on the estimation of the MRL function when this type I censoring is present. Guess and Park [10] proposed the use of conservative non-parametric confidence bounds for the MRL function based on the formula $m'(t) = m(t)\lambda(t)-1$ where $\lambda(t)$ is the hazard function (this formula can be easily derived from (1)). Other approaches consist of the extrapolation of the survival function beyond the maximum observed event time T^*. Moeschberger and Klein [11] proposed several methods (including the use of a parametric model) for '*completing*' the Kaplan–Meier estimator of the survivor function. Klein, Lee and Moeschberger [12] suggested treating uncensored observations nonparametrically and using a parametric model for the censored observations. However, as pointed out by Su and Fang [13], a potential problem in the use of a parametric approach is that the estimated function tends to fit poorly in the tail of the distribution.

To overcome this problem, a few authors have proposed a hybrid estimator of the mean survival time, combining the Kaplan–Meier curve and a parametric model that can be estimated locally, using the available data in the tail of the distribution. For instance, Su and Fang [13] proposed the use of the exponential distribution for the estimation of the tail. This parametric model however has a prescribed tail behaviour that is insufficiently flexible to apply when the tail behaviour is unknown a priori. In particular, it has infinite support, which is at best a crude approximation for human studies. A flexible model that provides for an upper bound on the support is desirable. In some studies, a heavier tail than an exponential may also be of relevance or may at least provide a suitable approximation over the scales of interest. Some authors, such as Gelber, Goldhirsch and Cole [14] or Gong and Fang [15], have proposed the use of goodness-of-fit procedures in order to choose the appropriate parametric model for the tail of the hybrid estimator. However, the selection of candidate distributions is arbitrary, and the large variability associated with the Kaplan–Meier curve could compromise the selection of the appropriate model.

In the following section, we present a new hybrid semi-parametric estimator that differentiates from other hybrid estimators in that the parametric model for the tail of the distribution is based on the GPD, a classic extreme value model that has rigorous mathematical justification for approximating the upper tail of almost all population distributions [16]. It is a challenging problem to approximate the upper tail of any distribution because of the inherent lack of sample data in the tails. Extreme value theory is used to justify appropriate models for the upper tail behaviour, along with the inference tools needed to provide reliable parameter estimation and uncertainty estimates. The theory of extremes also specifically focuses on extrapolation of models past the range of the observed data, which is a key issue when the upper tail is

censored, so the tools developed from such extremal theory are ideal in this application. The advantage of using the GPD is that it provides a flexible model. Such flexibility avoids the key problems with model mis-specification that are inherent when particular models, for part or the entire lifetime distribution, are assumed a priori, which can constrain the upper tail behaviour that is of primary interest. The available data themselves will therefore determine the tail behaviour, rather than this being prescribed a priori by the choice of lifetime distribution.

3. Semi-parametric estimate

The method proposed in this paper is based on the assumption that the conditional distribution of $T_u = T - u \mid T > u$, the excesses above a threshold u, can be approximated by the GPD when u is sufficiently large. This assumption can be derived from a core result in extreme value theory due to Leadbetter [17] that says that if there exists sequences of constants $a_n > 0$ and b_n, such that $(M_n - b_n)/a_n$ converges to G in distribution, where G is a non-degenerate distribution and M_n is the sample maximum of a sequence of i.i.d. observations from a population, then G is of the same type (up to location and scale parameters) as the generalized extreme value (GEV) distribution. The asymptotic arguments for this result are similar to those for the central limit theorem, where the distribution of a suitably linearly renormalised sample mean of i.i.d. observations from a population with finite variance converges to the normal distribution.

The distribution function of the GEV is

$$G(x) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}$$

where $[x]_+ = \max(x, 0)$ and μ and $\sigma > 0$ are the location and scale parameters, respectively. The shape parameter ξ determines the behaviour of the tail of the distribution:

- $\xi > 0$ heavy upper tail,
- $\xi = 0$ exponential upper tail (defined in the limit $\xi \to 0$),
- $\xi < 0$ short tail with finite upper end-point given by $-1/\xi$.

The preceding result provides an asymptotically motivated model for the block maximum of a random sample. A tail expansion of the GEV distribution can be used to motivate the GPD as a conditional distribution of T_u the excesses above some high threshold u. We assume that the limit representation holds for some large n and therefore the distribution function of the sample maximum approximately follows a GEV distribution, that is, $\left[F_T(a_n x + b_n)\right]^n \approx G(x) = \exp\left\{-[1+\xi x]_+^{-1/\xi}\right\}$ for all x such that $a_n x + b_n \geqslant u$. Thus, $F_T(y) \approx G^{1/n}\left(\frac{y-b_n}{a_n}\right)$ for $y \geqslant u$. Due to the max-stability property of the GEV distribution [16], the right-hand side of this upper tail approximation for F_T is the upper tail of a GEV distribution of the same type. Let $S_{T_u}(x)$ be the survivor function of the excess T_u given by

$$S_{T_u}(x) = P(T - u > x \mid T > u) = P(T > u + x \mid T > u) = \frac{S_T(u+x)}{S_T(u)} = \frac{1 - F_T(u+x)}{1 - F_T(u)}$$

where the excesses $x \geqslant 0$ and $S_T(x)$ are the survival function of T. Using the tail expansion of the GEV gives the upper tail approximation:

$$S_{T_u}(x) \approx \frac{1 - G(u+x)}{1 - G(u)} = \frac{1 - \exp\left\{-\left[1+\xi\left(\frac{u+x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}}{1 - \exp\left\{-\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}}$$

where the location μ and scale σ parameters subsume the normalisation constants. A first-order approximation then gives

$$S_{T_u}(x) \approx \frac{\left[1+\xi\left(\frac{u+x-\mu}{\sigma}\right)\right]_+^{-1/\xi}}{\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-1/\xi}} = \left[1+\frac{\xi x}{\sigma_u}\right]_+^{-1/\xi} \tag{2}$$

where $\sigma_u = \sigma + \xi(u - \mu)$. The right-hand side in formula (2) corresponds to the survivor function of the GPD with scale parameter σ_u (which depends on u) and shape parameter ξ. As with the GEV, the shape parameter ξ determines the behaviour of the upper tail.

- $\xi > 0$: heavy tailed.
- $\xi = 0$: Exponential tail with mean σ_u.
- $\xi < 0$: short tailed with upper bound $-\sigma_u/\xi$ for excesses and $u - \sigma_u/\xi$ for the maximum lifetime.

Therefore, the conditional distribution of the excesses $T_u = T - u \mid T > u$ can be modelled by the GPD when the threshold u is sufficiently 'large' for the preceding asymptotic approximation to be appropriate. The MRL at time $t = u$ can then be estimated by the expected value of the estimated conditional distribution, which in the case of the GPD is

$$E(T_u = T - u \mid T > u) = \frac{\sigma_u}{1 - \xi}.$$

The selection of the threshold u depends on a trade-off between bias and variance. For high values of u, the asymptotic approximation will be good, but parameter estimates will have a high variance due to the limited sample of excesses. On the other hand, a low value of u will provide more sample excesses thus reducing the estimation variance, but the asymptotic approximation may be poorer, leading to bias. In the context of this application, it might happen that the optimal threshold lies beyond the maximum observed time T^*, in which case, bias in the estimates may be unavoidable. All the results presented in this paper assume a moderate amount of type I censoring (up to 20%).

3.1. Description of the method

Let T be the true survival time and let C be the censoring time, and assume that T and C are independent. For an individual i, the observed pair is (X_i, δ_i), where $X_i = \min(T_i, C_i)$ and $\delta_i = I_{\{T_i < C_i\}}$. Let $\{(x_1, \delta_1), (x_2, \delta_2), \ldots, (x_n, \delta_n)\}$ be a random sample of observed times and censoring indicators. Let T^* be the maximum observed time. The estimate of the MRL function at time t, $\hat{m}(t)$, is obtained as follows:

(1) Choose the threshold u and select the observations x_i that are between u and T^*. Let m denote the number of these observations. There is much ongoing research into automated approaches for the choice of threshold [18]. The performance of such approaches in the presence of censoring is unknown, so we used traditional model fit diagnostics [16] to determine the threshold. A reasonable threshold choice, according to the standard diagnostics in all the simulations and applications in the succeeding text, was found to be the 80% quantile of the observed event times (uncensored observed times), meaning that between u and T^*, there are 20% of the uncensored observed times.

(2) Obtain the maximum likelihood estimates of the parameters ξ and σ_u of the GPD using the m observations that are between u and T^*. The censored likelihood function is

$$L(\xi, \sigma_u; t_i, \delta_i) = \prod_{i=1}^{m} \left(\frac{1}{\sigma_u} \left[1 + \frac{\xi t_i}{\sigma_u} \right]_+^{-\frac{1}{\xi} - 1} \right)^{\delta_i} \left(\left[1 + \frac{\xi t_i}{\sigma_u} \right]_+^{-1/\xi} \right)^{1 - \delta_i},$$

where t_i are the excesses above the threshold u, that is, $t_i = x_i - u$. This function can be maximised using standard optimization methods.

(3) Estimate the restricted mean residual lifetime from time t to u. The estimation of the restricted MRL function is obtained by ignoring the area under the Kaplan–Meier estimate of the survivor function from u to ∞. This estimate will be called $\hat{m}_{KM}(t)$.

(4) Estimate the MRL at time u for the fitted GPD as

$$\hat{m}(u) = \frac{\hat{\sigma}_u}{1 - \hat{\xi}}$$

where $\hat{\xi}$ and $\hat{\sigma}_u$ are the maximum likelihood estimates from step 2.

(5) Finally, estimate the MRL at time $t < u$ as

$$\hat{m}(t) = \hat{m}_{KM}(t) + \frac{\hat{m}(u) S_{KM}(u)}{S_{KM}(t)}$$

In step 5, $\hat{m}_{KM}(t)S_{KM}(t)$ is the area under the Kaplan–Meier (KM) estimate of the survivor function from t to u. On the other hand, $\hat{m}(u)S_{KM}(u)$ is the estimate of the area under the survivor function from u to ∞. Therefore, the estimate of the MRL is just the sum of these two areas divided by the estimate of the survivor function at time t.

Note that a constraint is imposed on the shape parameter in the likelihood in step 2 to ensure that the distribution has a finite mean and variance (in general, the rth moment of the GPD only exists for $\xi < 1/r$). In the simulations, the shape parameter was rarely found to be close to the constraint of $\xi \leqslant 0.5$, so this bound is not influential on the simulation results.

4. Simulation study

The main goal of this simulation study was to assess whether or not the proposed method produced adequate estimates of the MRL function. Three different survival distributions were used (exponential, gamma and log-normal) and were chosen to mimic the different survival experiences represented in Figure 1. These also represent a range of tail behaviours with the log-normal providing heavy tail behaviour.

In order to investigate the effect of the presence of type I censoring on the estimation of the MRL values, a slightly modified version of the usual theoretical setting for the generation of a random sample of survival times had to be considered. Let T be the true survival times and let C be the censored times, and assume that T and C are independent. For observation i, the observed tuple is (X_i, δ_i) where $X_i = \min(T_i, C_i, T^*)$, $\delta_i = 1_{\{T_i<C_i \,\cap\, T_i<T^*\}}$ and T^* is the time of the termination of the study. Note that, under this setting, an observation i will be censored not just when $T_i > C_i$ but also when $T_i > T^*$ (it could happen that $T_i < C_i$ but $T_i > T^*$, in which case observation i would still be censored).

The simulations were carried out with different combinations of the percentage of right random censoring (0%, 10% and 20% of censored observations between 0 and T^*) and type I censoring (0%, 10% and 20% of censored observations at time T^*). The simulations are executed as follows:

- Draw a random sample of size n, $(t_1, t_2, \ldots, t_n)$, from the distribution of T.
- Draw a random sample of n censoring times $(c_1, c_2, \ldots, c_n)$ from the uniform $(0, M)$.
- Obtain the observed values as $x_i = \min(t_i, c_i, T^*)$ for $i = 1, 2, \ldots, n$.
- Obtain the censoring indicators as $\delta_i = 1_{\{t_i \leqslant c_i \,\cap\, t_i \leqslant T^*\}}$ for $i = 1, 2, \ldots, n$.
- Using the observed values (x_i, δ_i), $i = 1, 2, \ldots, n$, obtain an estimate of the MRL function at time $t = 0$ using the proposed method.
- Calculate the difference between the true and estimated values of $m(0)$.
- Repeat these steps 500 times. Use values to assess the bias of the estimator and its standard deviation.

The values of M and T^* have to be calculated in order to achieve the desired percentages of censoring. To show how this was done, let B be the proportion of observations that are censored between 0 and T^*. Then,

$$P(\{C < T\} \cap \{C < T^*\}) = \frac{1}{M}\int_0^{T^*} S_T(c)dc = B. \tag{3}$$

On the other hand, let A be the proportion of type I censoring. Then,

$$\begin{aligned} P(\,(\{C < T\} \cap \{C > T^*\}) \cup (\{T < C\} \cap \{T > T^*\})\,) = \\ = \frac{1}{M} S_T(T^*)(M - T^*) = A. \end{aligned} \tag{4}$$

From equation (3),

$$M = \frac{1}{B}\int_0^{T^*} S_T(c)dc \tag{5}$$

and substituting M in equation (4), we obtain

$$S(T^*)\left(\frac{1}{B}\int_0^{T^*} S_T(c)\,dc - T^*\right) - \frac{A}{B}\int_0^{T^*} S_T(c)dc = 0$$

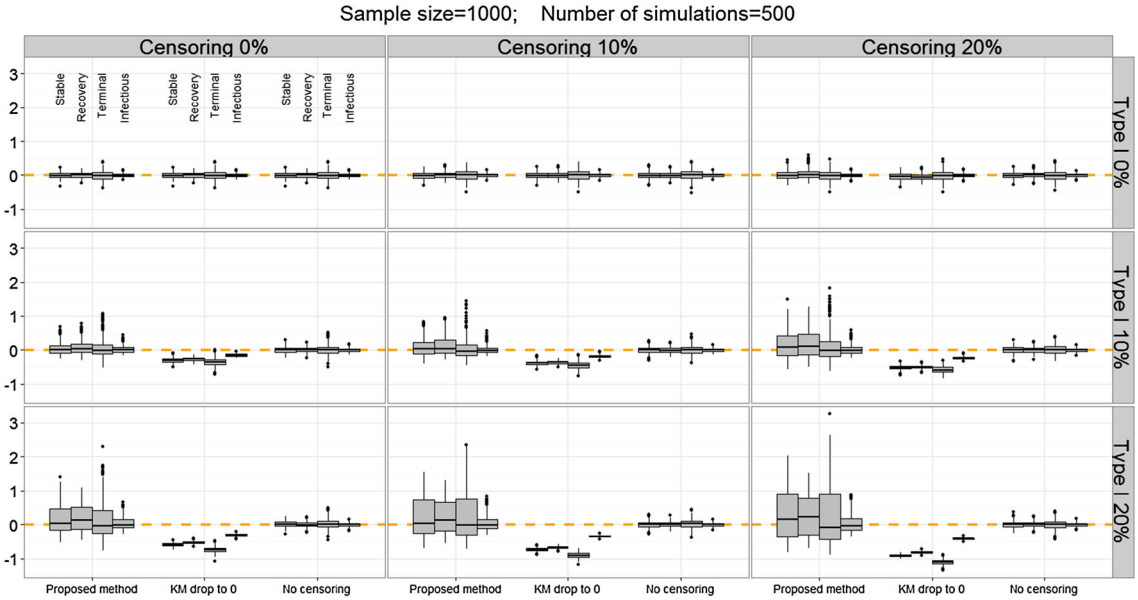

Figure 3. Deviations between the estimates of the mean residual life function at time 0 and the true value for different combinations of random right and type I censoring.

This is an implicit equation that can be solved for T^* using Newton–Raphson. Once T^* has been calculated, M can be obtained from Equation (5).

Figure 3 shows the results of simulating samples under different combinations of censoring and different distributions. The sample size was $n = 1000$, and the box plots presented are based on 500 simulations. The vertical axis represents the deviations between the true value, $m(0)$, and the estimated value. Three different methods were considered: firstly, the method proposed in this paper, which uses the GPD to estimate the MRL values ('Proposed method'); secondly, the method based on the restricted mean lifetime, which consists of dropping to 0 the Kaplan–Meier estimate of the survivor function at time T^*, that is, ignoring the area under the survivor function from u to ∞ ('KM drop to 0'); and finally, the last method evaluates the estimates one would get if all of the observations were uncensored ('No censoring'). This last method serves as a reference to establish how much variability is actually present for the estimation. The columns in the plot represents different percentages of random right censoring whereas the rows represent different percentages of type I censoring, that is, censoring due to the termination of the study. In addition to this, four different distributions were considered:

- Exponential ($\lambda = 1/3$), constant MRL 'Stable');
- Gamma ($k = 0.7, \lambda = 3$), increasing MRL 'Recovery');
- Gamma ($k = 2, \lambda = 3$), decreasing MRL ('Terminal'); and
- Log-normal ($\mu = 1, \sigma = 0.5$), decreasing MRL first and increasing after ('Infectious').

As can be seen in the figure, the proposed method seems to perform well compared with the method based on the restricted MRL. The latter method only gives reasonable results when the percentage of censoring due to the termination of the study is 0. The variability of estimates of the MRL function at time 0 of the log-normal distribution appears to be lower than that of the other distributions, which may explain the reduction in variability of the proposed method when compared with the other distributions. The worst results are produced for 20% random right censoring and 20% type I censoring, where there is some evidence of biassed results, especially for the exponential distribution. Apart from this, the proposed method produces very reasonable results for all of the different distributions and all the different combinations of censoring.

The simulations were also performed at different time points. The time points were chosen to be 1/3 and 2/3 of T^*. The results obtained were similar to the ones obtained at time 0 with a corresponding increase of variability (Figures 4 and 5).

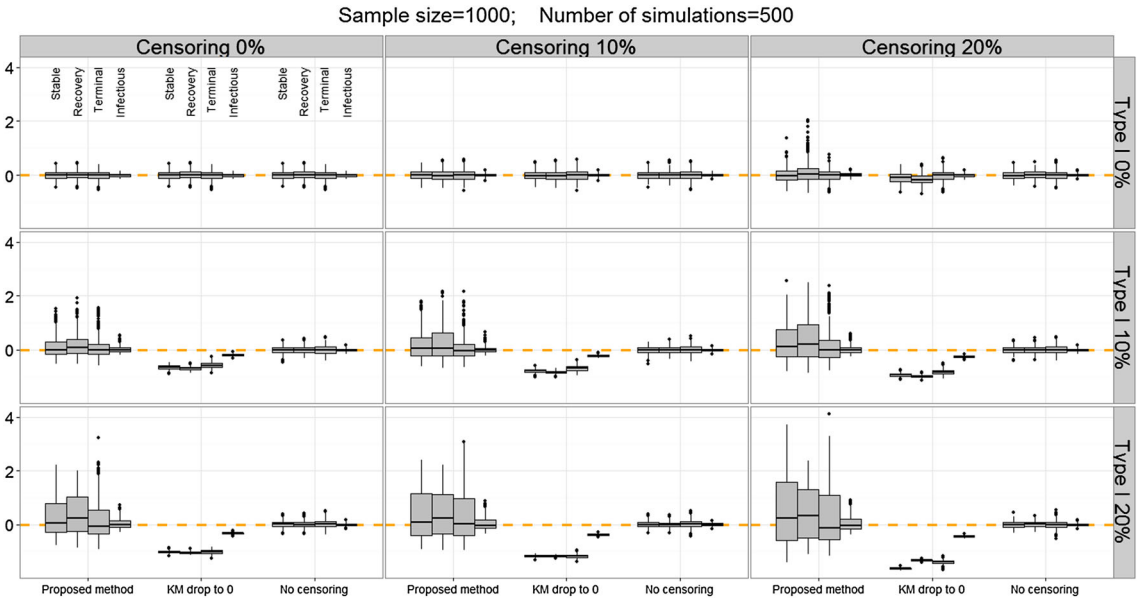

Figure 4. Deviations between the estimates of the mean residual life function at time 1/3 of T^* and the true value for different combinations of random right and type I censoring.

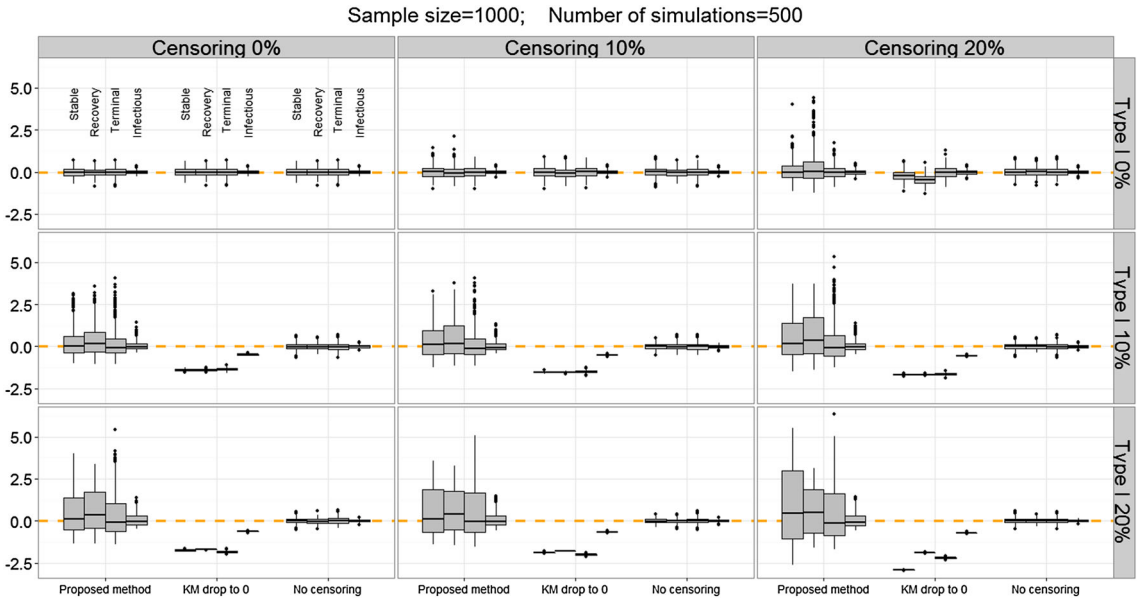

Figure 5. Deviations between the estimates of the mean residual life function at time 2/3 of T^* and the true value for different combinations of random right and type I censoring.

5. Applications

We illustrate the proposed method using two datasets. The first one is based on the historical 6-MP data [19], the first randomised, double-blind, placebo-controlled sequential study in acute leukaemia. This dataset has been used extensively in the survival analysis literature as an illustrative example in theoretical and applied work, but not to date in the context of MRL. Figure 6 shows the Kaplan–Meier estimates of the survivor functions (left) and the estimates of the MRL applying the proposed method. There is evidence of a significant treatment effect ($p < 0.001$) on the basis of a log-rank test. The MRL function provides a useful summary of the effectiveness of this treatment in units of time rather than on a probabilistic scale or through a comparison of a fixed time point, for example, median survival (7 weeks for

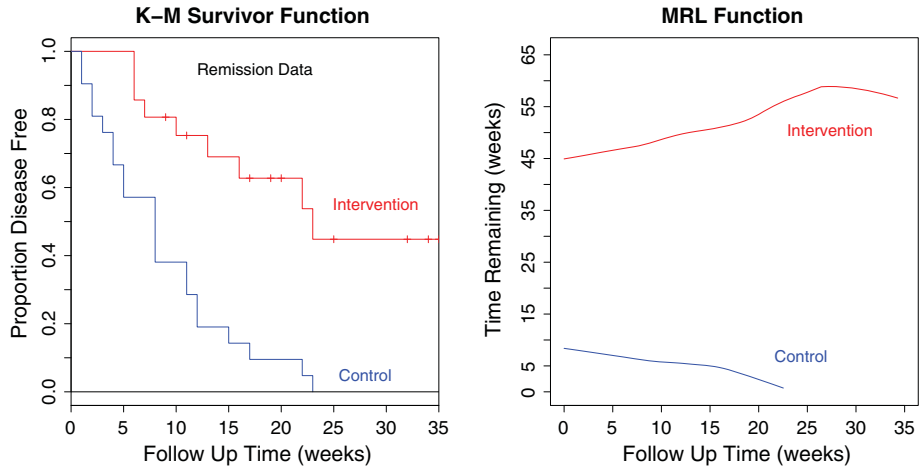

Figure 6. Historical 6-MP data. Left: Kaplan–Meier (K–M) estimates of the survivor functions. Right: estimates of the mean residual life (MRL) functions applying the proposed method.

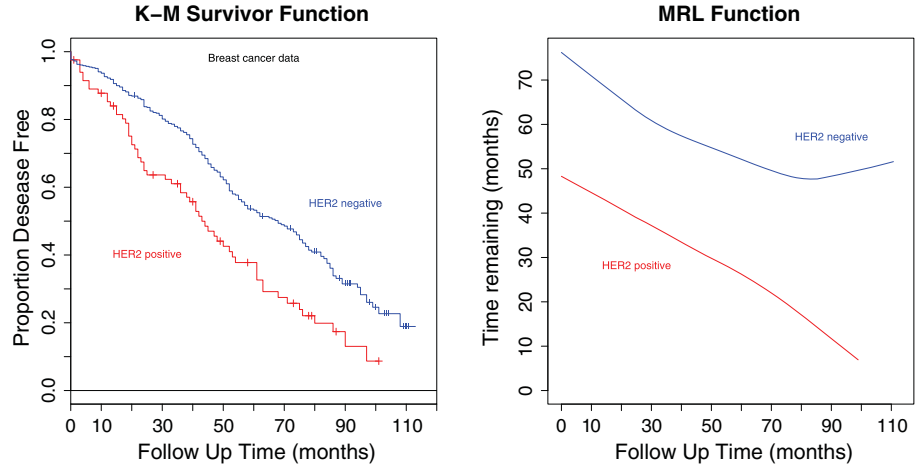

Figure 7. Women with breast cancer from the west of Ireland. Left: Kaplan–Meier (K–M) estimate of the survivor function. Right: estimate of the mean residual life (MRL) function using the proposed method.

the control group and 23 weeks for the intervention group). For example, it is estimated that the intervention improves survival by, on average, 35 weeks upon initiation of the treatment, which increases slightly across time.

The second dataset is based on a sample of women with breast cancer from the west of Ireland. The event of interest was the recurrence of the disease. Figure 7 (left) shows the KM estimate of the survivor function by HER2 status. HER2 is a protein that promotes the growth of cancer cells, and it is generally accepted that women who test positive for this protein have worse prognosis, as evidenced by the estimated survivor functions for those with and without the risk factor (log-rank test $p < 0.001$). What is unclear is the effect the presence of this risk factor has on time to recurrence in units of time. The estimate of the MRL suggests that the mean time to recurrence for HER2-positive patients is, on average, approximately 20 months earlier compared with HER2-negative patients.

6. Discussion

There is a need to present the results from survival analysis studies in ways that provide clear and simple interpretations for both clinicians and patients. The MRL function is a very promising way of doing this with a natural answer to the question 'how long do I have left?' in units of time. Comparing these functions for different treatments, or the presence of certain risk factors, also gives an easily understood summary,

avoiding hazard ratios or probability scales that require careful interpretation. One reason that the MRL function has perhaps been overlooked is the problem of its estimation from censored data, particularly where part of the censoring process is due to study termination. Paradoxically, this type of censoring is very common in the biomedical context, where the use of this type of summary is needed, and very few papers have focused on estimating the mean residual lifetime in this situation. Although the use of the restricted MRL can be used as an alternative in some situations, as in the case of group/treatment comparisons, there is still a need for reliable estimates in some applications as in the long-term treatment effects on quality-adjusted survival [20, 21].

The method presented here aims to obtain estimates of the MRL function in the presence of moderate type I censoring, occurring in studies with a fixed follow-up time period in which the event of interest does not occur for all of the subjects before the maximum time under study T*. In the absence of this type of censoring, when all the individuals experience the event of interest before study termination, a simple non-parametric approach (using the KM estimate of the survivor function) would be sufficient to obtain the desired estimates of the MRL function. We consider censoring at a fixed maximum time under study, as this provides a worst case scenario. Under a more liberal random censoring regime, the impact on MRL estimation will be reduced. We have shown through simulations and illustrative applications that, if a moderate proportion of patients is censored at T*, it is still possible to obtain reasonable estimates of the MRL function.

The critical issue with estimation of the MRL is that it is dependent on a reliable tail approximation. A common approach in the literature is to assume a parametric population distribution for the lifetimes that is then extrapolated to the missing tail. This approach is very restrictive and may lead to biassed inference, as models that work well for the bulk of the lifetimes (mode of the lifetime distribution) may not approximate the tail very well. The inherent lack of tail data, due to the censoring, may also mean any poor tail fit is not identifiable in standard goodness-of-fit diagnostics. Nonparametric alternatives for tail approximation are also extremely challenging, because of this inherent lack of tail data.

It is well known in the field of extreme value modelling [16] that if one is interested in tail approximation, then one should focus the model, its subsequent inference and goodness-of-fit validation on the tail observations only. Following this argument, some authors have proposed the use of hybrid estimators, where the parametric model is fitted locally using the information contained in the tail of the distribution. However, the selection of the tail parametric model is only justified based on goodness-of-fit arguments, which might be compromised because of the large variability. The parametric model we are proposing for the tail, the GPD, is a classic extreme value model for approximating the tail of the distribution, beyond some suitably high thresholds. The upper (and lower) tail of almost all population distributions can be approximated by the GPD. There are certain mild regularity conditions that guarantee this tail convergence [16], but for practical purposes in such survival studies, the GPD will provide a reliable tail approximation. As previously mentioned, the GPD includes the upper tail of all the common parametric survival models (e.g. exponential, gamma and log-normal) as special cases.

In this scenario, inferences are required well beyond the observed tail of the data. An assumption of stability is required such that regularities in the unobservable tail of the distribution are assumed to reach far enough back into the observable region that extrapolation may be based on a model fitted to the observed events [22]. This is precisely the motivation behind extreme value theory, its resultant models and inferential approaches that underpin this extrapolation. The guiding principle is to use the limited data that are available to provide reliable tail inferences, including realistically accounting for the uncertainties involved.

Based on our results, the proposed method can be used both in the absence of type I censoring, giving equivalent results to the simple non-parametric approach and in the presence of moderate type I censoring, outperforming the existing approaches considered. If the percentage of type I censoring is more than moderate (e.g. 20% or more), the task of estimating the MRL would be much more challenging, because of there being even less information in the tail. In this situation, we could end up with there being no observations beyond the threshold at which the GPD provides a reliable tail approximation, and this is certainly the main limitation of the proposed estimator. In such situations, one could use the restricted MRL function as an alternative; however, the estimates obtained would be negatively biassed. Further, the interpretation of restricted MRL is also much more challenging for practitioners and patients, when individual level summary predictions are desirable, and the selection of the restriction time is often hard to justify and explain. It is the desire to provide a more interpretable measure of survival that is the motivation behind this research.

It is important to point out that the estimator presented here assumes that the population under investigation (censored or not) is homogeneous. If the data collected comes from a mixture where the patients that have censored observations at T* (i.e. long-term survivors) behave differently from patients with uncensored observations, even with a small proportion of type I censoring, the method we are proposing could be compromised (as could almost all the other approaches in the literature, which also commonly assume such homogeneity). A careful discussion with the clinicians would probably be necessary in order to determine if the population of long-term survivors differs in any way from those with observed survival times.

The methods presented here, using results from extreme value theory, are a promising approach to the general estimation of MRL functions and seem to outperform previously proposed solutions. We see the presentation of the MRL as a useful addition to complement formal inferential techniques in survival analysis and, if so desired, confidence regions for the MRL function, or indeed the difference in MRL between groups, or strata, could be generated using bootstrap methods. But, perhaps more importantly, we believe that the MRL could become a useful and routine communication aid for summarising survival studies.

Acknowledgements

We would like to thank the Associate Editor and referees for their helpful comments from which the presentation of the paper has improved greatly.

The authors would also like to thank Prof. Michael Kerin (National Breast Cancer Research Institute, NUI Galway, Ireland) and Prof. Grace Callagy (Discipline of Pathology, NUI Galway, Ireland) for providing the Breast cancer dataset used in this paper. This study was initiated whilst Carl Scarrott was on sabbatical funded by the University of Canterbury and hosted by NUI Galway, for which their support is gratefully acknowledged.

This work was supported by the Irish Research Council for Science, Engineering and Technology (IRCSET) and the Science Foundation Ireland Award 07/MI/012.

References

1. Yang G. Life expectancy under random censorship. *Stochastic Processes and their Applications* 1977; **6**(1):33–39.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**(282):457–481.
3. Gill R. Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics* 1983; **11**(1):49–58.
4. Ruiz JM, Guillamón A. Nonparametrie recursive estimator for mean residual life and vitality function under dependence conditions. *Communications in Statistics - Theory and Methods* 1996; **25**(9):1997–2011.
5. Chaubey YP, Sen PK. On smooth estimation of mean residual life. *Journal of Statistical Planning and Inference* 1999; **75**(2):223–236.
6. Hille E. *Functional Analysis and Semi-groups.* American Mathematical Society Colloquium Publications. 31. New York: American Mathematical Society (AMS). XI, 528 p., 1948.
7. Chaubey YP, Sen A. Smooth estimation of mean residual life under random censoring. *IMS Collections* 2008; **1**:35–49.
8. Zhou M, Jeong JH. Empirical likelihood ratio test for median and mean residual lifetime. *Statistics in Medicine* 2011; **30**(2):152–159.
9. Shen Y, Xie M, Tang LC. Nonparametric estimation of decreasing mean residual life with type II censored data. *IEEE Transactions on Reliability* 2010; **59**(1):38–44.
10. Guess F, Park DH. Nonparametric confidence bounds, using censored data, on the mean residual life. *IEEE Transactions on Reliability* 1991; **40**(1):78–80.
11. Moeschberger ML, Klein JP. A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics* 1985; **41**(1):253–259.
12. Klein JP, Lee SC, Moeschberger ML. A partially parametric estimator of survival in the presence of randomly censored data. *Biometrics* 1990; **46**(3):795–811.
13. Su Z, Fang L. A novel method to calculate mean survival time for time-to-event data. *Communications in Statistics - Simulation and Computation* 2012; **41**(5):611–620.
14. Gelber RD, Goldhirsch A, Cole BF. Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. International Breast Cancer Study Group. *Controlled Clinical Trials* 1993; **14**(6):485–499.
15. Gong Q, Fang L. Asymptotic properties of mean survival estimate based on the Kaplan–Meier curve with an extrapolated tail. *Pharmaceutical Statistics* 2012; **11**:135–140.
16. Coles S. *An Introduction to Statistical Modeling of Extreme Values* (1st edn.). Springer Series in Statistics. Springer: London, 2001.
17. Leadbetter MR. Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields* 1983; **65**:291–306.
18. Scarrott C, MacDonald A. A review of extreme value threshold estimation and uncertainty quantification. *Revstat-Statistical Journal* 2012; **10**(1):33–60.

19. Freireich EJ, Gehan E, Frei E, Schroeder LR. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia. *Blood* 1963; **21**(6):699–716.
20. Cole BF, Gelber RD, Anderson KM. Parametric approaches to quality-adjusted survival analysis. International Breast Cancer Study Group. *Biometrics* 1994; **50**(3):621–631.
21. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the Q-TWiST method. *The American Statistician* 1995; **49**(2):161–169.
22. Davison AC, Padoan SA, Ribatet M. Statistical modeling of spatial extremes. *Statistical Science* 2012; **27**(2):161–186.