

Imputation Methods for a Bathtub Hazard Model with Right-Censored and Interval-Censored Data

Ismail I.^{1,2}, Arasan J.¹, Safari M. A. M.¹, Mustafa M. S.¹

¹*Department of Mathematics and Statistics, Faculty of Science,
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

²*College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, UiTM Cawangan Kelantan,
Kampus Machang Bukit, Ilmu, 18500 Machang, Kelantan, Malaysia*

(Received 10 February 2025; Revised 28 September 2025; Accepted 29 September 2025)

In this study, we extended a two-parameter bathtub hazard model that exhibits either an increasing or a bathtub-shaped failure rate depending on its shape parameter. The model was expanded to include covariates in the presence of right- and interval-censored data. To handle interval-censored data, we employed two imputation techniques: midpoint and random imputation. A simulation study was conducted to evaluate and compare the performance of these imputation methods using standard error (SE) and root mean square error (RMSE) values. The findings indicated that random imputation outperforms midpoint imputation in most instances, achieving lower SE and RMSE across different censoring proportions and sample sizes. To demonstrate the practical applicability of the extended model, a real dataset on the occurrence of oral lesions in children after liver transplantation was used for illustrative purposes.

Keywords: *bathtub-shaped; right-censored; interval-censored; imputation; midpoint; random.*

2010 MSC: 62N02

DOI: 10.23939/mmc2025.04.1087

1. Introduction

The failure rate function is an important characteristic in modelling lifetime data. Among the failure rate functions, common hazard functions, such as constant, increasing, or decreasing, are typically modeled using well-known distributions – such as Weibull, gamma, or exponential. However, these models are limited in their ability and flexibility to represent non-monotonic failure rates, such as bathtub-shaped failure rates. This hazard function is characterized by three main phases: an initial period of decreasing hazard, followed by a constant hazard phase, and finally an increasing hazard phase [1, 2]. Additionally, this hazard function is relevant in contexts such as human mortality and newly launched products, where high initial failure rates can be due to infant mortality, design defects, production errors, or inexperienced maintenance, which are then followed by a constant failure rate during the useful life and eventually by high failure rates due to aging or wear-out – a pattern that resembles the shape of a bathtub [3]. Therefore, there are many life distributions developed to model data with bathtub-shaped distribution over the last few decades. Most of them are generalizations, modifications, and extensions of existing distributions. For instance, one important example is the extension of the Weibull distribution. Early work by [4] proposed a three-parameter exponentiated Weibull (EW) distribution that can exhibit a bathtub-shaped failure rate. Other extensions of the Weibull distribution have been discussed in [5] and [6], among others.

An interesting two-parameter lifetime distribution that can exhibit either monotonically increasing or bathtub-shaped failure rates was proposed by [7]. He pointed out that the confidence intervals for its shape parameter and the joint confidence regions for both parameters have a closed form, which makes this model the only known two-parameter distribution with bathtub-shaped failure rates that provides exact joint confidence regions for its parameters. The Chen distribution is referred to as the bathtub

hazard model throughout this paper. Due to its useful properties, the bathtub hazard model was extensively studied by other researchers. For example, [8] proposed a simple exact statistical test for the shape parameter of the bathtub distribution and found that the new test was more powerful than the one originally proposed by Chen [7]. [9] discussed exact confidence intervals and exact joint confidence regions for the parameters of the bathtub hazard model based on progressively type-II censored samples. [11] investigated interval estimation methods for the parameters of the bathtub hazard model using type-II censored samples. [10] derived maximum likelihood and Bayesian estimators for the bathtub hazard model under progressively type-II censored samples. However, despite certain useful properties, the model lacks practical applications because it does not possess a scale parameter [12, 13]. Hence, researchers were motivated to extend the bathtub hazard model. In this regard, [12] extended the model to a three-parameter distribution and named it the extended-Weibull distribution due to its relation to the Weibull distribution. [13] subsequently studied the properties of the extended model, including the shapes of the density and failure rate functions. They also obtained confidence intervals for the parameters and carried out a likelihood ratio test to evaluate the extended model. Several other studies have contributed to extending the bathtub hazard model, as reported in [14–17].

Several studies have been conducted to discuss and modify the bathtub hazard model. However, few studies have extended the model to accommodate covariates, whether fixed or time-dependent. Therefore, this paper aims to present an extended version of the bathtub hazard model, which incorporates a fixed covariate in the presence of right- and interval-censored data. To handle interval-censored data, two imputation methods will be employed: midpoint imputation and multiple imputation. The methods will then be compared based on the standard error (SE) and root mean square error (RMSE).

The midpoint imputation approach has gathered significant attention from researchers in studies related to AIDS. Several studies, including those by [18–20], have employed similar methodologies to estimate the year of HIV-1 seroconversion by using the midpoint between the dates of a negative and a subsequent positive HIV-1 antibody test. Seroconversion is the period during which an individual's blood serum changes from HIV-negative to HIV-positive due to the presence of the virus. Consistent with the findings of [19, 21], it was observed that when the seroconversion interval is relatively short (less than two years), the midpoint closely aligns with estimates based on different assumptions about the underlying distribution. In addition to midpoint imputation, several other studies employed right-point or left-point imputation. In addition, a probability imputation method, such as random and multiple imputation techniques, was commonly used to impute the event times for interval-censored observations. In a study by [22], several imputation methods, including midpoint, left-point and right-point, random, mean, and median imputation, were compared for a parametric Cox model with partly interval-censored data. The results showed that random imputation, as well as mean and median imputation, yielded better performance compared to other imputation techniques. In a recent study conducted by [23], it was found that the accuracy of midpoint imputation was influenced by sample size; as the sample size increased, the corresponding value decreased. Conversely, the multiple imputation method consistently produced low residuals, irrespective of sample size.

The paper is organized as follows: the next section presents the bathtub hazard model and its distributional properties, followed by a brief discussion of imputation methods. A subsequent section provides a simulation study that compares the performance of parameter estimates with regard to each imputation method. The extended model with the best-performing imputation method is then illustrated using an oral lesion occurrence dataset. Finally, the paper concludes with a summary of findings.

2. Bathtub hazard model

The cumulative density function of the bathtub distribution is given by

$$F(t) = 1 - e^{\lambda(1-e^{t^\gamma})}, \quad t \geq 0. \quad (1)$$

The probability density function of the model can be defined as

$$f(t; \lambda, \alpha) = \lambda \gamma t^{\gamma-1} e^{t\gamma} e^{\lambda(1-e^{t\gamma})}. \quad (2)$$

The survival and failure rate functions are given, respectively, by

$$S(t; \lambda, \gamma) = e^{\lambda(1-e^{t\gamma})}, \quad (3)$$

$$h(t; \lambda, \gamma) = \lambda \gamma t^{\lambda-1} e^{t\gamma}, \quad (4)$$

where λ and γ are the parameters. The parameter γ can determine the shape of the hazard function. The failure rate is increasing when $\gamma < 1$ and is bathtub-shaped otherwise.

In order to accommodate covariate effects to the hazard function of the bathtub hazard model, we allow the parameter to be a function of the covariate. The function can be written as

$$\lambda_i = e^{-\beta_0 - \beta_1 x_i}. \quad (5)$$

Let $\theta = (\gamma, \beta_0, \beta_1)$ be the vector of the model parameters. Maximum likelihood estimation (MLE) method was used to estimate the three parameters.

Suppose we have interval-censored and right-censored lifetimes for $i = 1, 2, \dots, n$ observations. Let t_i denotes the failure time of the i^{th} observation. Let L_i and R_i be the left and right endpoints of the intervals for the i^{th} subject, respectively. The following censoring indicator is defined to incorporate right and interval censored data to the likelihood function,

δ_{I_i} = if the subject is interval-censored, 0 otherwise;

δ_{R_i} = if subject is right censored, 0 otherwise;

δ_{E_i} = if subject is not censored (exact survival time is observed), 0.

The likelihood function for right-, interval censored or uncensored without any imputation is,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i)]^{\delta_{E_i}} [S(t_{R_i})]^{\delta_{R_i}} [S(t_{L_i}) - S(t_{R_i})]^{\delta_{I_i}} \\ &= \prod_{i=1}^n \left[e^{(-\beta_0 - \beta_1 x_i)} \gamma t_i^{\gamma-1} e^{t_i \gamma} e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_i \gamma})} \right]^{\delta_{E_i}} \\ &\quad \times \left[e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_{R_i} \gamma})} \right]^{\delta_{R_i}} \left[e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_{L_i} \gamma})} - e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_{R_i} \gamma})} \right]^{\delta_{I_i}}. \end{aligned} \quad (6)$$

On the other hand, the likelihood function for uncensored, right- and interval-censored with imputation method is,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(t_i)]^{\delta_{E_i}} [S(t_{R_i})]^{\delta_{R_i}} [f(\tilde{t}_i)]^{\delta_{I_i}} \\ &= \prod_{i=1}^n \left[e^{(-\beta_0 - \beta_1 x_i)} \gamma t_i^{\gamma-1} e^{t_i \gamma} e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_i \gamma})} \right]^{\delta_{E_i}} \\ &\quad \times \left[e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_{R_i} \gamma})} \right]^{\delta_{R_i}} \left[e^{e^{(-\beta_0 - \beta_1 x_i)} \gamma \tilde{t}_i^{\gamma-1} e^{\tilde{t}_i \gamma} e^{e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{\tilde{t}_i \gamma})}} \right]^{\delta_{I_i}}, \end{aligned} \quad (7)$$

where $\tilde{t}_i$ is the imputed time, which can be the midpoint, right point, or left point of the interval or the randomly imputed value.

Taking the logarithm of (7) we get the log-likelihood function as follows

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n \left[\delta_{E_i} \left(-\beta_0 - \beta_1 x_i + \ln(\gamma) + (\gamma - 1) \ln(t_i) + t_i \gamma + e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_i \gamma}) \right) \right] \\ &\quad + \delta_{R_i} \left[e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{t_{R_i} \gamma}) \right] \\ &\quad + \delta_{I_i} \left[-\beta_0 - \beta_1 x_i + \ln(\gamma) + (\gamma - 1) \ln(\tilde{t}_i) + \tilde{t}_i \gamma + e^{(-\beta_0 - \beta_1 x_i)} (1 - e^{\tilde{t}_i \gamma}) \right]. \end{aligned} \quad (8)$$

3. Imputation method

This study employed two imputation techniques, which will be compared in a simulation study as discussed in the first section. In midpoint imputation (MI) method, the event time is imputed by the midpoint of the censoring interval, $[l_i, r_i]$ calculated as $(l_i + r_i)/2$.

In the random imputation (RI) technique, a single imputation is performed on interval-censored event time. A modification of the existing multiple method was done where the imputed value is randomly selected from the distribution of the event time. Specifically, following the method outlined by [24], it is assumed that the event time T follows a known continuous distribution. Let t denote an interval-censored event time observed within the interval $(l, r]$, where $l \leq r$. Then, the survival functions at the endpoints of the interval are defined as: $S(l) = P(T > l)$ and $S(r) = P(T > r)$. A random number is then simulated from a uniform distribution over the range $[S(r), S(l)]$. The resulting uniform variate is transformed using the inverse transform method to estimate the imputed observation time, t , within the interval.

4. Simulation study

A simulation study was conducted to evaluate the performance of the bathtub hazard model with right-censored and interval-censored data, incorporating fixed covariate. The study was carried out with $N = 1000$ replications across varying sample sizes, $n = 50, 100, 150, 200, 250$ and different levels of censoring proportion (CP): 0%, 10%, 20%, 30%, 40% and 50%. Specifically, the parameter of interest, parameters γ , β_0 and β_1 and were set at 0.4, 3.3 and 0.9, respectively. The simulation was done using the R programming language, following the algorithm as outlined below:

1. Generate covariate values x_i from a standard normal distribution.
2. Generate a sequence of random numbers u_i from a standard uniform distribution within the interval $(0, 1)$ to find the event times, t_i for $i = 1, 2, \dots, n$.
3. Generate censoring times, c_i from an exponential distribution, with the adjustment on the value of μ to obtain the desired censoring proportion.
4. Calculate the survival time t_i by:

$$t_i = \left(\ln \left(1 - \frac{\ln(1 - U_i)}{e^{(-\beta_0 - \beta_1 x_i)}} \right) \right)^{1/\gamma}. \quad (9)$$

To assess the performance of the parameter estimates generated for each imputation method across varying sample sizes and censoring proportions, the SE and RMSE values are computed as follows:

$$\widehat{SE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\hat{\theta}_i - \frac{1}{N} \sum_{j=1}^N \hat{\theta}_j \right)^2}, \quad (10)$$

$$RMSE = \sqrt{(SE^2 + bias^2)}, \quad (11)$$

where the bias value can be obtained as

$$bias = E(\hat{\theta}) - \theta. \quad (12)$$

To evaluate which imputation method produces better estimates, SE and RMSE values were compared. In Tables 1 and 2, the SE and RMSE values for each parameter using the MI method are presented, while Tables 3 and 4 depict these values for RI. It can be observed that both SE and RMSE values increase as the censoring proportion increases, while they decrease with larger sample sizes. This suggests that better estimates can be obtained with larger sample sizes and smaller values of censoring proportion. In addition, line plots are generated to visually illustrate the performance and provide a clearer comparison between the two imputation methods based on SE and RMSE values, see Figure 1-8. Instead of considering all combinations of the five censoring proportions and sample sizes, the line plots in this paper highlight a subset of combinations: two censoring proportion values ($CP = 10\%$ to

reflect lower censoring and $CP = 50\%$ for higher censoring) and two sample sizes ($n = 50$ to represent smaller datasets and $n = 250$ for larger datasets).

Table 1. Standard error (SE) values for parameter estimates using midpoint imputation.

Estimates	CP(%)	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = 250$
$\hat{\gamma}$	0	0.0231	0.0163	0.0133	0.0111	0.0099
	10	0.0252	0.0169	0.0140	0.0121	0.0109
	20	0.0263	0.0184	0.0147	0.0127	0.0114
	30	0.0273	0.0197	0.0154	0.0135	0.0121
	40	0.0284	0.0207	0.0158	0.0140	0.0130
	50	0.0288	0.0223	0.0184	0.0156	0.0137
$\hat{\beta}_0$	0	0.3440	0.2367	0.1953	0.1618	0.1474
	10	0.3612	0.2381	0.1981	0.1712	0.1511
	20	0.3613	0.2492	0.1983	0.1708	0.1523
	30	0.3679	0.2577	0.1986	0.1769	0.1557
	40	0.3643	0.2571	0.2046	0.1780	0.1616
	50	0.3749	0.2728	0.2244	0.1885	0.1676
$\hat{\beta}_1$	0	0.1976	0.1337	0.1064	0.0922	0.0793
	10	0.1985	0.1342	0.1075	0.0948	0.0869
	20	0.2125	0.1414	0.1145	0.0953	0.0863
	30	0.2086	0.1452	0.1143	0.1009	0.0898
	40	0.2241	0.1502	0.1145	0.1011	0.0941
	50	0.2201	0.1613	0.1283	0.1090	0.0984

Table 2. Root mean square error (RMSE) values for parameter estimates using midpoint imputation.

Estimates	CP (%)	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = 250$
$\hat{\gamma}$	0	0.0239	0.0166	0.0135	0.0111	0.0100
	10	0.0252	0.0178	0.0157	0.0150	0.0148
	20	0.0285	0.0241	0.0216	0.0233	0.0228
	30	0.0322	0.0303	0.0323	0.0306	0.0314
	40	0.0405	0.0398	0.0346	0.0381	0.0417
	50	0.0424	0.0506	0.0491	0.0510	0.0491
$\hat{\beta}_0$	0	0.3518	0.2391	0.1967	0.1629	0.1490
	10	0.3663	0.2381	0.1990	0.1723	0.1525
	20	0.3613	0.2504	0.2013	0.1803	0.1624
	30	0.3679	0.2605	0.2089	0.1861	0.1704
	40	0.3647	0.2637	0.2138	0.1937	0.1780
	50	0.3751	0.2771	0.2312	0.1989	0.1821
$\hat{\beta}_1$	0	0.2013	0.1347	0.1070	0.0928	0.0798
	10	0.2031	0.1369	0.1090	0.0956	0.0886
	20	0.2192	0.1437	0.1164	0.0966	0.0874
	30	0.2146	0.1483	0.1163	0.1031	0.0917
	40	0.2317	0.1534	0.1177	0.1036	0.0963
	50	0.2267	0.1649	0.1315	0.1111	0.1004

As illustrated in Figures 1–4, the SE and RMSE values generated from both MI and RI method were plotted for two different censoring proportions of 10% and 50%. For the lower censoring proportion of 10%, there is not much difference in the SE values between MI and RI (see Figure 1). However, at $CP = 50\%$, there is a clear observation that RI produces smaller SE values (see Figure 2). Similarly, for RMSE values at the lower censoring proportion of 10%, RI exhibits lower RMSE values than MI as depicted in Figure 3. In Figure 4, a similar trend is observed at $CP = 50\%$, except for the parameter β_0 , where MI shows smaller RMSE values. For all cases, both the SE and RMSE values decrease as the sample size increases.

Table 3. Standard Error (SE) Values for Parameter Estimates Using Random Imputation.

Estimates	CP(%)	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = 250$
$\hat{\gamma}$	0	0.0230	0.0162	0.0131	0.0110	0.0104
	10	0.0238	0.0158	0.0134	0.0116	0.0103
	20	0.0240	0.0169	0.0136	0.0116	0.0105
	30	0.0248	0.0169	0.0139	0.0120	0.0107
	40	0.0249	0.0180	0.0143	0.0123	0.0112
	50	0.0278	0.0178	0.0144	0.0127	0.0110
$\hat{\beta}_0$	0	0.3355	0.2354	0.1916	0.1644	0.1490
	10	0.3373	0.2301	0.1896	0.1674	0.1456
	20	0.3367	0.2288	0.1839	0.1564	0.1427
	30	0.3237	0.2266	0.1884	0.1592	0.1421
	40	0.3256	0.2314	0.1837	0.1596	0.1413
	50	0.3134	0.2261	0.1858	0.1591	0.1433
$\hat{\beta}_1$	0	0.1958	0.1372	0.1064	0.0935	0.0820
	10	0.1948	0.1258	0.1068	0.0913	0.0819
	20	0.1936	0.1303	0.1116	0.0926	0.0840
	30	0.1995	0.1331	0.1086	0.0933	0.0826
	40	0.1956	0.1318	0.1049	0.0955	0.0823
	50	0.2068	0.1340	0.1083	0.0963	0.0841

Table 4. Root Mean Square Error (RMSE) Values for Parameter Estimates Using Random Imputation.

Estimates	CP(%)	$n = 50$	$n = 100$	$n = 150$	$n = 200$	$n = 250$
$\hat{\gamma}$	0	0.0238	0.0165	0.0134	0.0111	0.0105
	10	0.0239	0.0160	0.0140	0.0126	0.0120
	20	0.0246	0.0191	0.0162	0.0157	0.0160
	30	0.0267	0.0221	0.0212	0.0206	0.0201
	40	0.0295	0.0278	0.0261	0.0253	0.0265
	50	0.0329	0.0317	0.0299	0.0304	0.0296
$\hat{\beta}_0$	0	0.3425	0.2370	0.1938	0.1652	0.1499
	10	0.3378	0.2384	0.2046	0.1906	0.1828
	20	0.3575	0.2821	0.2417	0.2365	0.2413
	30	0.3817	0.3268	0.3182	0.3063	0.3041
	40	0.4228	0.3940	0.3785	0.3658	0.3798
	50	0.4604	0.4394	0.4143	0.4185	0.4081
$\hat{\beta}_1$	0	0.1988	0.1381	0.1074	0.0938	0.0826
	10	0.1955	0.1258	0.1073	0.0920	0.0834
	20	0.1937	0.1310	0.1130	0.0953	0.0887
	30	0.2045	0.1355	0.1125	0.0999	0.0902
	40	0.1996	0.1356	0.1107	0.1046	0.0925
	50	0.2068	0.1372	0.1123	0.1022	0.0943

On the other hand, Figures 5–8 illustrate a comparison of the SE and RMSE values for MI and RI specifically at two sample sizes, $n = 50$ and 250 . At both sample sizes, RI consistently produced smaller SE values compared to MI (see Figures 5 and 6). For RMSE, RI also yielded smaller values except for the parameter, β_0 , where MI showed a smaller RMSE values as the censoring proportion increased, as presented in Figures 7 and 8. In summary, RI demonstrates better performance, as it produces smaller values of SE and RMSE in most instances, particularly for parameters γ and β_1 . However, for β_0 , there is slight inconsistency, with MI showing better RMSE values at $CP = 50\%$, also for $n = 50$ and $n = 250$.

This indicates that RI generally provides more accurate and precise parameter estimates in the bathtub hazard model with interval-censored data compared to MI at various sample sizes and censoring proportions.

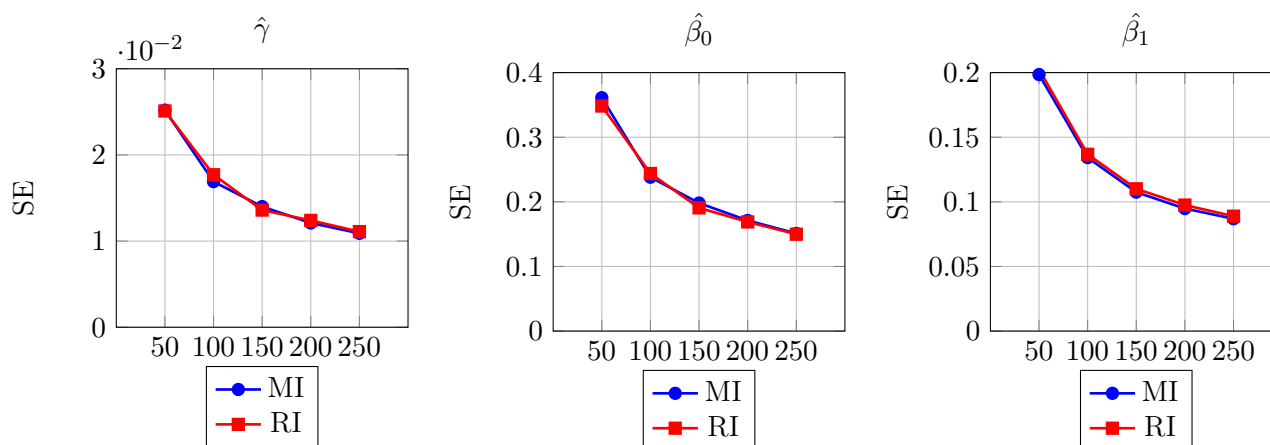


Fig. 1. Line plots comparing SE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $CP = 10\%$.

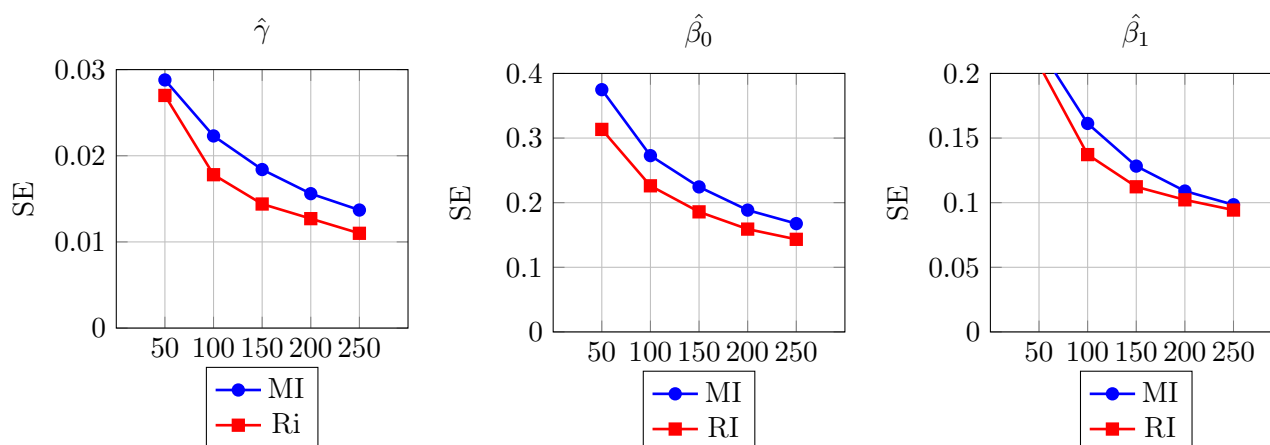


Fig. 2. Line plots comparing SE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $CP = 50\%$.

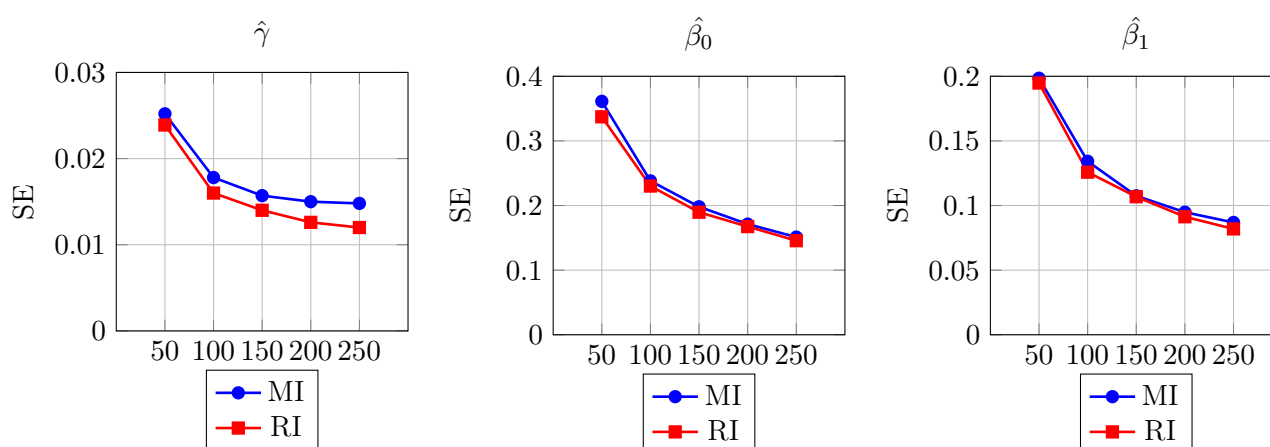


Fig. 3. Line plots comparing RMSE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $CP = 10\%$.

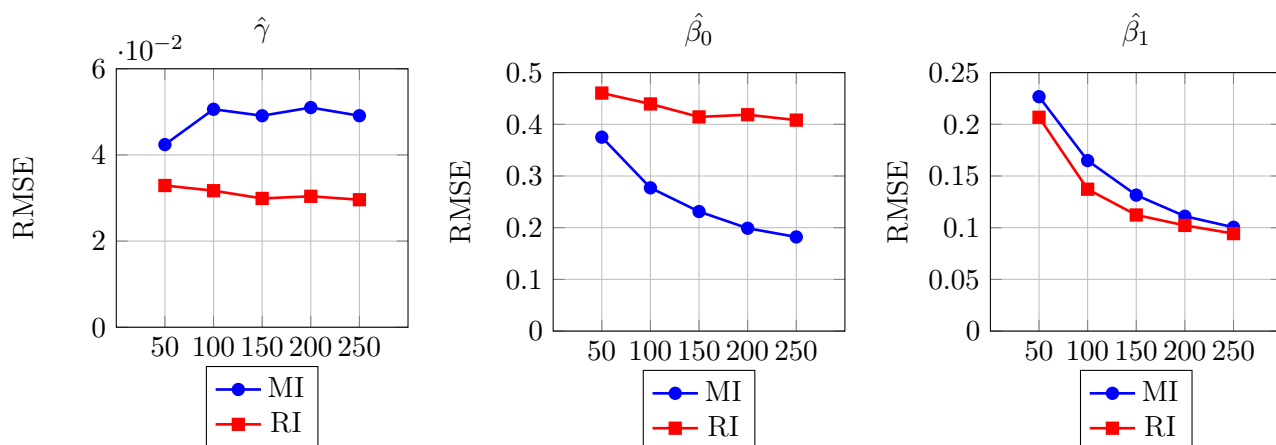


Fig. 4. Line plots comparing RMSE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $CP = 50\%$.

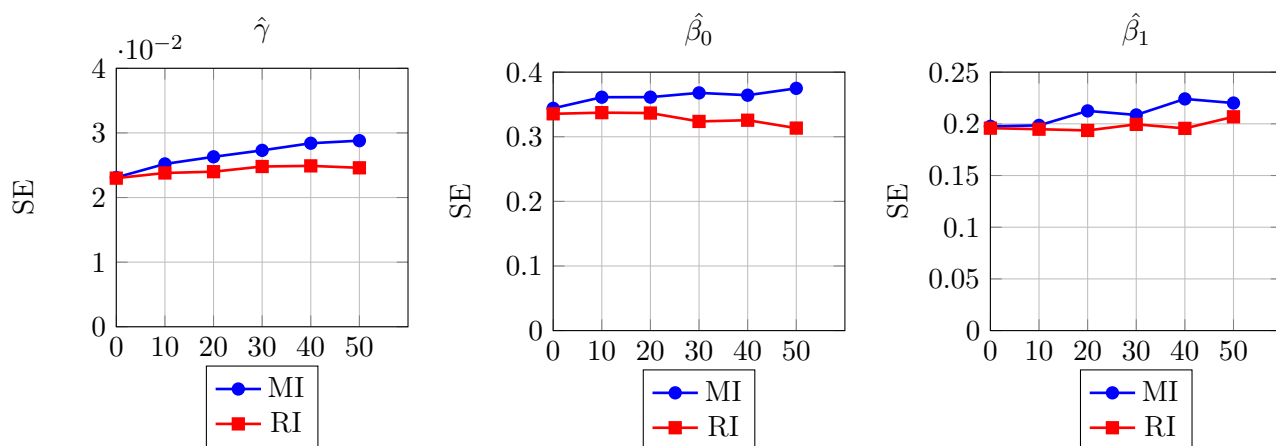


Fig. 5. Line plots comparing SE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $n = 50$.

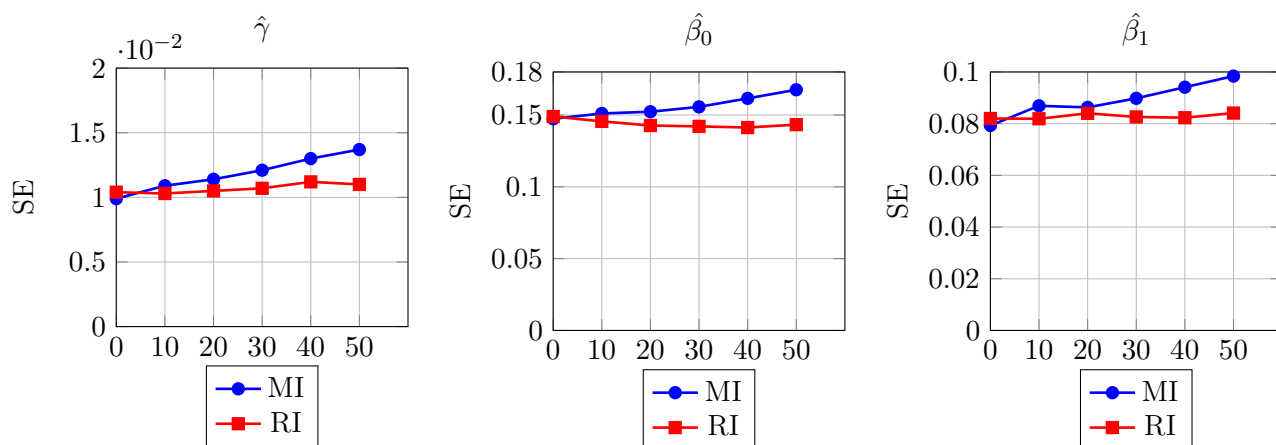


Fig. 6. Line plots comparing SE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $n = 250$.

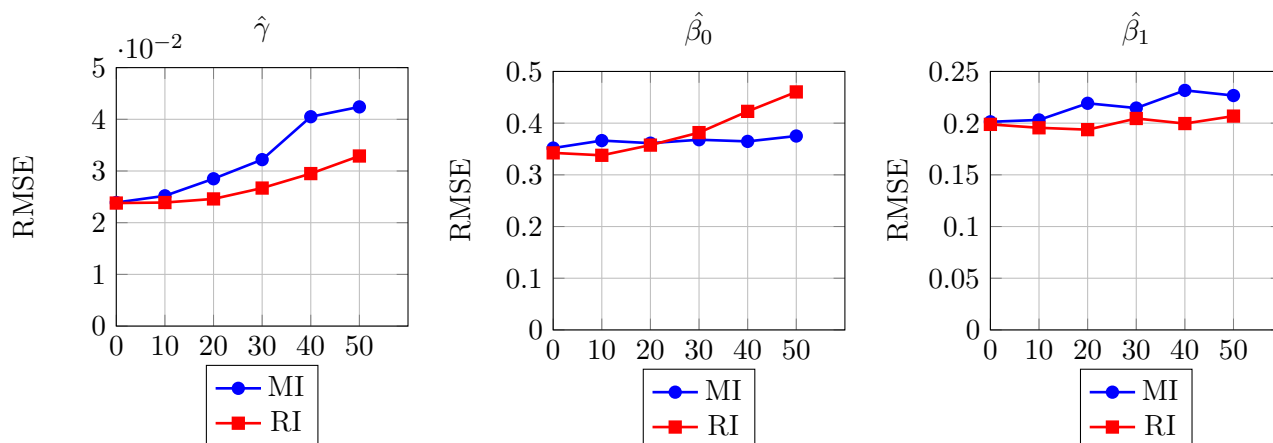


Fig. 7. Line plots comparing RMSE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $n = 50$.

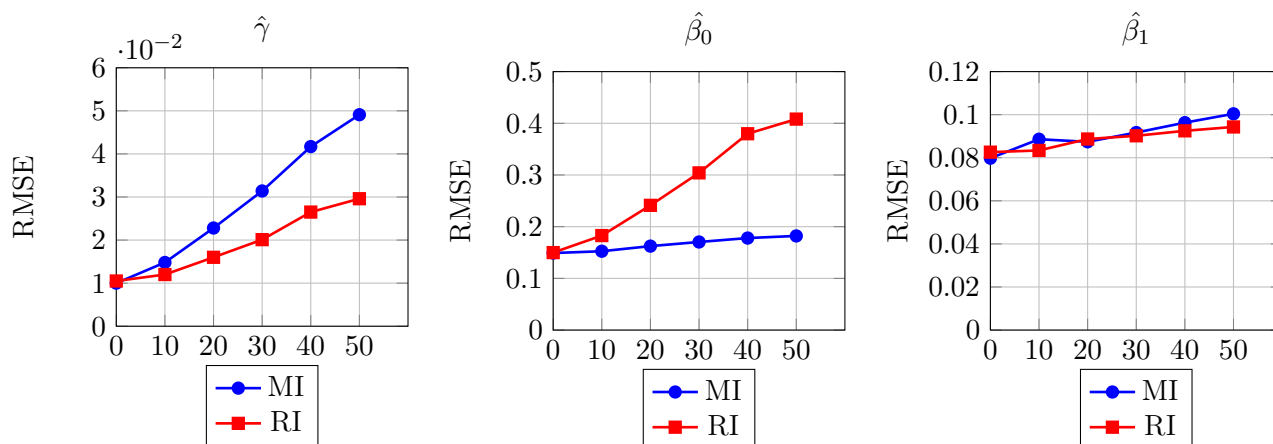


Fig. 8. Line plots comparing RMSE values for $\hat{\gamma}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ using midpoint and random imputation at $n = 250$.

5. Real data application

To show the applicability of bathtub hazard model in real data sets with interval-censored data, here we consider a data set on oral lesions after liver transplantation in children aged between 0 to 18 years taken from a study conducted A.C. Camargo Cancer Center in São Paulo, Brazil from 2013 to 2016, as discussed in Calsavara [25]. Random imputation was used in dealing the interval-censored data in the data set. Figure 9 illustrates the survival probabilities obtained using a non-parametric Kaplan–Meier (solid line) and based on the bathtub hazard model (blue dotted line).

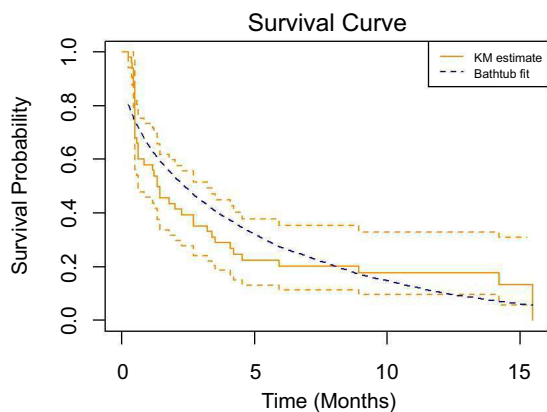


Fig. 9. Plot of estimated survival probabilities based on the fitted distribution and Kaplan–Meier estimator.

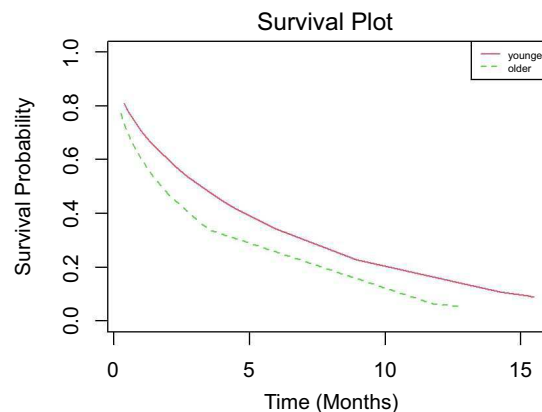


Fig. 10. Survival plot.

From Figure 9, the bathtub estimates based on the bathtub hazard model appear to be roughly close to Kaplan–Meier estimates suggesting the bathtub hazard distribution is appropriate for the oral lesion data. Next, we obtained the bathtub hazard model with covariate, to examine the significance of the covariate (age) on the time for oral lesions to develop after liver transplantation surgery. The covariate of interest is the age variable which denotes patients is younger (age ≤ 1 year, $n = 23$ patients) or older ($1 < \text{age} < 18$ year, $n = 27$ patients) at the time of transplantation. Based on the results presented in Table 5, age group is not statistically significant (p-value = 0.2174 > 0.05). Although the p-value suggests that there is no significant difference between the younger and older groups, the survival plot (Figure 10) shows a slight visual difference between these groups. This difference indicates that the analysis may require larger sample sizes or the inclusion of additional covariates to detect an effect.

Table 5. Maximum likelihood estimation (MLE) of oral lesion data.

Parameters	Estimates (Standard Error)	t-value	p-value
$\hat{\gamma}$	0.3439 (0.0316)	10.893	$< 2 \times 10^{-16}$
$\hat{\beta}_0$	1.9878 (0.5289)	3.758	0.000171
$\hat{\beta}_1$	-0.3809 (0.3088)	-1.223	0.2174

6. Conclusion

In this study, we proposed an extension of the bathtub hazard model by incorporating a fixed covariate in the presence of right- and interval-censored data. We compared two imputation methods for handling interval-censored data: midpoint and random imputation. The simulation results indicated that random imputation outperformed midpoint imputation, yielding smaller SE and RMSE values at most levels of censoring proportion and across different sample sizes. Finally, we analyzed a real dataset on oral lesion development after liver transplantation for illustrative purposes. The results demonstrated a good fit of the bathtub hazard model to the data, highlighting its potential relevance in medical research and clinical practice. Future research may explore other imputation approaches, such as conditional mean, conditional median, or multiple imputation. Furthermore, we expect that the extended model will find broader applications in medical research and related fields.

-
- [1] Lai C. D., Xie M., Murthy D. N. P. B. Ch. 3. Bathtub-shaped failure rate life distributions. *Handbook of Statistics*. **20**, 69–104 (2001).
 - [2] Rajarshi S., Rajarshi M. B. Bathtub distributions: a review. *Communications in Statistics – Theory and Methods*. **17** (8), 2597–2621 (1988).
 - [3] Thach T. T., Briš R. An additive Chen-Weibull distribution and its applications in reliability modeling. *Quality and Reliability Engineering International*. **37** (1), 352–373 (2021).
 - [4] Mudholkar G. S., Srivastava D. K. Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data. *IEEE Transactions on Reliability*. **42** (2), 299–302 (1993).
 - [5] Bebbington M., Lai C.-D., Zitikis R. A flexible Weibull extension. *Reliability Engineering & System Safety*. **92** (6), 719–726 (2007).
 - [6] Liao Q., Ahmad Z., Mahmoudi E., Hamedani G. G. A New Flexible Bathtub-Shaped Modification of the Weibull Model: Properties and Applications. *Mathematical Problems in Engineering*. **2020**, 3206257 (2020).
 - [7] Chen Z. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*. **49** (2), 155–161 (2000).
 - [8] Wu J.-W., Lu H.-L., Chen C.-H., Wu C.-H. Statistical inference about the shape parameter of the new two-parameter bathtub-shaped lifetime distribution. *Quality and Reliability Engineering International*. **20** (6), 607–616 (2004).
 - [9] Wu S.-J. Estimation of the two-parameter bathtub-shaped lifetime distribution with progressive censoring. *Journal of Applied Statistics*. **35** (10), 1139–1150 (2008).

- [10] Rastogi M. K., Tripathi Y. M., Wu S.-J. Estimating the parameters of a bathtub-shaped distribution under progressive type-II censoring. *Journal of Applied Statistics*. **39** (11), 2389–2411 (2012).
- [11] Wang R., Sha N., Gu B., Xu X. Statistical Analysis of a Weibull Extension with Bathtub-Shaped Failure Rate Function. *Advances in Statistics*. **2014**, 304724 (2014).
- [12] Xie M., Tang Y., Goh T. N. A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*. **76** (3), 279–285 (2002).
- [13] Tang Y., Xie M., Goh T. N. Statistical Analysis of a Weibull Extension Model. *Communications in Statistics – Theory and Methods*. **32** (5), 913–928 (2003).
- [14] Chaubey Y. P., Zhang R. An extension of Chen’s family of survival distributions with bathtub shape or increasing hazard rate function. *Communications in Statistics – Theory and Methods*. **44** (19), 4049–4064 (2015).
- [15] Bhatti F. A., Hamedani G. G., Najibi S. M., Ahmad M. On the Extended Chen Distribution: Development, Properties, Characterizations and Applications. *Annals of Data Science*. **8** (1), 159–180 (2021).
- [16] Tarvirdizade B., Ahmadpour M. A New Extension of Chen Distribution with Applications to Lifetime Data. *Communications in Mathematics and Statistics*. **9** (1), 23–38 (2021).
- [17] Sarhan A. M., Mustafa A. A new extension of the two-parameter bathtub hazard shaped distribution. *Scientific African*. **17**, e01345 (2022).
- [18] Lui K.-J., Darrow W. W., Rutherford G. W. A model-based estimate of the mean incubation period for AIDS in homosexual men. *Science*. **240** (4857), 1333–1335 (1988).
- [19] Mariotto A. B., Mariotti S., Pezzotti P., Rezza G., Verdecchia A. Estimation of the acquired immunodeficiency syndrome incubation period in intravenous drug users: A comparison with male homosexuals. *American Journal of Epidemiology*. **135** (4), 428–437 (1992).
- [20] Williams C. F., Klinzman D., Yamashita T. E., Xiang J., Polgreen P. M., Rinaldo C., Liu C., Phair J., Margolick J. B., Zdunek D., Hess G., Stapleton J. T. Persistent GB Virus C Infection and Survival in HIV-Infected Men. *The New England Journal of Medicine*. **350** (10), 981–990 (2004).
- [21] Law C. G., Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine*. **11** (12), 1569–1578 (1992).
- [22] Zyoud A. M. H., Elfaki F. A. M., Hrairi M. Parametric model based on imputation techniques for partly interval censored data. *Journal of Physics: Conference Series*. **949**, 012002 (2017).
- [23] Kim H. J., Kim S., Lee E. Cox proportional hazards regression for interval-censored data with an application to college entrance and parental job loss. *Economies*. **10** (9), 218 (2022).
- [24] Arasan J., Midi H. Bootstrap based diagnostics for survival regression model with interval and right-censored data. *Austrian Journal of Statistics*. **52** (2), 66–85 (2023).
- [25] Calsavara V. F., Rodrigues A. S., Rocha R., Tomazella V., Louzada F. Defective regression models for cure rate modeling with interval-censored data. *Biometrical Journal*. **61** (4), 841–859 (2019).

Методи імпутації для моделі ризику “ванна” з даними з правосторонньою та інтервальною цензурою

Ісмаїл І.^{1,2}, Арасан Дж.¹, Сафарі М. А. М.¹, Мустафа М. С.¹

¹*Кафедра математики та статистики, Факультет природничих наук, Університет Путра Малайзії 43400 UPM Серданг, Селангор, Малайзія*

²*Коледж обчислювальної техніки, інформатики та математики, Університет технологій MARA, UiTM Каванган Келантан, Кампус Мачанг Букіт, Ілму, 18500 Мачанг, Келантан, Малайзія*

У цьому дослідженні розширено двопараметричну модель ризику типу “ванна”, яка демонструє або зростаючий, або ванноподібний коефіцієнт відмов залежно від параметра форми. Модель доповнено включенням коваріат за умов наявності даних, цензурованих правосторонньо та інтервально. Для обробки інтервально цензурованих даних застосовано два методи імпутації: серединну та випадкову. Було проведено імітаційне моделювання для оцінки та порівняння ефективності цих методів із використанням стандартної помилки (SE) та середньоквадратичного відхилення (RMSE). Результати показали, що випадкова імпутація в більшості випадків переверщує серединну, забезпечуючи нижчі значення SE та RMSE за різних рівнів цензурування та розмірів вибірки. Щоб продемонструвати практичну застосовність розширеної моделі, було використано реальний набір даних про виникнення уражень ротової порожнини у дітей після трансплантації печінки.

Ключові слова: ванноподібний; правосторонньо цензурований; інтервально цензурований; імпутація; середня точка; випадковий.