STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **10**, June 2022, pp 815–828.
Published online in International Academic Press (www.IAPress.org)

On Accelerated Failure Time Models Performance Under Progressive Type-II Censoring

Amal Helu [1,*], Hani Samawi [2], Majd Alslman [1]

[1] *Department of mathematics, The University of Jordan, Jordan*
[2] *Jiann-Ping Hsu College of Public Health, Georgia Southern University, USA*

Abstract Accelerated failure time (AFT) models have intensive applications in many research areas, including but not limited to behavioral, chronic (e.g., cancer), and infectious diseases (e.g., HIV) research. In this paper, we investigate the performance of the AFT models when Progressive Type-II censoring schemes are performed. We demonstrate the usefulness of using these schemes. We discuss their testing procedure power, $Bias$, and MSE of the hazard ratio estimates compared to the same sample size of the uncensored data. Theoretically, we derive the models, the MLE scores, and the Fisher information matrix. A comparison between these estimators is provided by using extensive simulation. A real-life data example is provided to illustrate our proposed estimators.

Keywords Accelerated failure time model; hazard ratio; Progressive Type-II censoring; survival analysis.

AMS 2010 subject classifications 62E10, 62N01, 62N02, 62G30

DOI: 10.19139/soic-2310-5070-1446

1. Introduction

Clinical researchers often encounter time-to-events (survival time), including but not limited to time-to-death, time-to-first diagnostics of disease, etc. Survival data are often analyzed using the proportional hazard model (PH). The PH model is used in multivariate survival analysis to analyze p-covariates and/or risk factors' consequences on survival time. However, the semi-parametric PH model requires a constant hazard ratio over time. This assumption's violation may reduce the corresponding tests' power and misconception of parameters estimation results (Schemper, 1992).

Lee and Wang (2003) modified the PH model to make it a parametric model by assuming that the baseline hazard function follows some statistical distributions. However, this revised model is available for a few regression models: exponential, Gompertz, and Weibull. In addition, the new PH model can only be used with a proportional hazard assumption. On the other hand, the accelerated failure time (AFT), which is a parametric survival model, can be used as an alternative to the PH model, mainly to outdo the statistical problem due to the violation of the PH assumption (Wei,1992). The AFT model also accounts for covariates' effects through the log of survival times instead of the PH model's hazard rate. Also, the interpretation of the AFT results is more accessible than the results of the PH model.

The regression parameters are more intuitively interpreted concerning the change in the survival time median. Since we used the log-linear formulation to formulate the AFT models, they provided independent parameter estimates of the regression coefficients and the random frailty effects (Keiding et al., 1997). Therefore, the

*Correspondence to: Amal Helu (Email: a.helu@ju.edu.jo)

ISSN 2310-5070 (online) ISSN 2311-004X (print)

misspecification of the family for the frailty distribution may not affect the analysis severely. Lambert et al. (2004) confirmed the robustness of regression parameter estimates despite any misspecification of the frailty distribution. Applications of PH and AFT models traditionally depend on the assumption that the responses are independent of the units subject to failure.

Several researchers used AFT models to analyze survival data. For example, Chapman et al. (1992) studied the prognostic effects of breast cancer survival factors using four models: exponential, Weibull, Log-Logistic, and Log-Normal, and concluded that the Log-Normal model would provide the most acceptable fit to the data. Royston (2001) showed the Log-Normal AFT model's practical value in analyzing breast and ovarian cancer patients' survival times. Komarek et al. (2005) used the AFT model to analyze AIDS onset in the Women's Interagency HIV Study. Additionally, Lambert et al. (2004) implemented the AFT models with shared frailty to locate the prognostic factors for a kidney graft's survival time in patients from 31 transplant centers in the UK. Samawi et al. (2019) used Double Extreme Ranked Set Sampling to improve the AFT models' performance.

In survival analysis, a sample of size n objects is exposed to a life test to observe their failure times. The recorded data, is then used to model a time-to-failure distribution. This strategy may be unreasonable, expensive, and may take longer to observe all failure times. Due to some restrictions, the researcher may need to stop the study before recording all the subjects' failure times under consideration. Additionally, some test subjects may be removed from the experiment to be used in another study, which happens in cases of costly test subjects, such as clinical equipment. Moreover, in some cases, the failure is deliberate and expected; however, it does not happen due to operator flaws, equipment malfunction, test irregularity, etc. Samples that result from such situations are called censored samples.

The two most well-known censoring schemes are Type-I and Type-II. With Type-I, we stop the experiment after a predetermined time. While, with a Type-II, the experiment is terminated after a predetermined number of failures. However, these two censoring schemes do not allow intermediate elimination of active units throughout the experiment other than at the final termination point. Therefore, the focus has been on progressive censoring in the last few years.

A Progressive Type-II censoring is a generalization of Type-II censoring (P-II). It allows researchers to remove subjects before the final termination point due to specific situations, such as the loss of contact with individuals studied. Under this type of censoring, we replace n independent units simultaneously on a life testing experiment to observe their failure times; however, we only observe $m(< n)$ failure times. The censored times occurred progressively in m stages: After the first failure is observed, randomly remove R_1 units immediately from the $(n-1)$ survivors, leaving $(n-1-R_1)$ survival items. Then, after the failure of the second unit, remove another R_2 units randomly from the remaining survival items. This process continues until m failures are observed, and all the remaining $(n-m-R_1-R_2\ldots,R_{m-1}=R_m)$ survived units are removed from the experiment. It is assumed that these n units' lifetimes are independent and identically distributed with a common distribution function F. Moreover, n, m, and the censoring scheme $R=\{R_1,R_2,\ldots,R_m\}$ are all pre-specified. Note that if $R_1=R_2=\ldots=R_{m-1}=0$, then $R_m=n-m$, which corresponds to Type-II censoring. If $R_1=R_2=\ldots=R_m=0$, then $(m=n)$, which represents the complete data set. Readers may refer to Balakrishnan and Cramer (2014) for a comprehensive literature review on progressive censoring.

There has been considerable attention paid to Progressive Type-II censoring due to its potential applications in reducing sample sizes required for lifetime experiments and time-to-event trials. Moreover, the availability of high-speed computing resources enhances the focus on progressive censoring.

Numerous authors have discussed inferences under progressive censoring by using different lifetime distributions. Some of these include Viveros and Balakrishnan (1994), Alvarez-Andrade et al. (2007), Saracoglu (2012), Musleh and Helu (2014), Helu and Samawi (2017) & (2019), Helu et al. (2020), Helu and Samawi (2021) and Pushkarna et al. (2020). An overview of progressive censorship can be found in Balakrishnan and Cramer (2014).

To our knowledge, no one had studied how the AFT model would operate when the data are Progressively Type-II censoring. The main goal of this study is to develop a general framework for simulating progressive Type-II censored survival times with predefined censoring schemes and censoring rates.

2. Preliminaries

Assuming the random variable of failure time of an event denoted by T. The distribution of T is usually identified using three functions: the survival function $S(t)$, the hazard rate function or risk function, indicated by $h(t)$ and the probability density (or probability mass) function, represented by $f(t)$. The concept of censoring data is a unique feature of survival data. Censoring issues arise when the exact time to event for a subject is unknown. In this article, our focus will be on Progressive Type-II censoring.

Given, $t_1, t_2, ..., t_n$, $K = n(n - R_1 - 1)(n - R_1 - R_2)...(R_{m-1} - m + 1)$, with censoring scheme $R = \{R_1, R_2, ..., R_m\}$, the likelihood function is given by

$$l(\boldsymbol{\beta}) = K \prod_{i=1}^{m} f(t_\mathbf{i}) S(t_i)^{R_i}, \tag{1}$$

where, $\boldsymbol{\beta}$ is the set of parameters.

2.1. AFT model

Similar to Liu (2012) and Samawi et al. (2019), the log-linear form of the AFT model, $logT_i$ concerning, is given by

$$logT_i = \beta_0 + x_{i1}\beta_1 + ... + x_{ip}\beta_p + \sigma\varepsilon_i,\ i = 1, 2, ..., m. \tag{2}$$

Note that σ is a scale parameter, and ϵ_i, the random error term, which is assumed to have a specific distribution. Adopting the same notation as in Samawi et al. (2019), the survival function at the time t_i is as follows:

$$\begin{aligned} S_i(t) &= P(logT_i \geq logt_i) \\ &= P\left(\varepsilon_i \geq \frac{logt_i - \beta_0 - \mathbf{w}_i'\boldsymbol{\beta}}{\sigma}\right) \\ &= S_0\left(\frac{logt_i - \beta_0 - \mathbf{w}_i'\boldsymbol{\beta}}{\sigma}\right),\ 0 < t_i < \infty \end{aligned} \tag{3}$$

where, $\mathbf{w}_i = (x_{i1}, \ldots, x_{ip})'$, $i = 1, 2, \ldots, m$, assumed to be observed, and $\boldsymbol{\beta} = (\beta_1, \beta_1, \ldots, \beta_p)'$ are the coefficients of the predictors. Furthermore, the hazard function for T_i at t is

$$h_i(t|\mathbf{w}_i, \boldsymbol{\beta}) = \frac{1}{t_i\sigma} h_0\left(\frac{logt_i - \beta_0 - \mathbf{w}_i'\boldsymbol{\beta}}{\sigma}\right),\ i = 1, 2, ..., m,$$

where, $h_0(t)$ is the baseline hazard function at survival time t. As in Samawi et al. (2019), the covariates $\{X_1, X_2, ..., X_p\}$ are assumed to have a multiplicative effect on the hazard function. Therefore, the predicted value of the hazard function, given $\{x_{i1}, ..., x_{ip}\}$, which is denoted by $\hat{h}(t_i|x_{i1}, ..., x_{ip})$, can take values in the range $(0, \infty)$.

In this paper, we discuss the performance of the AFT model using Progressive Type-II censoring. The article unfolds as follows: The AFT regression model and its properties using Progressive Type-II censoring are discussed in Section 3. The simulation study in Section 4 compares the performance of Progressive Type-II censoring with complete uncensored data under the AFT model. In Section 5, all methods are demonstrated using real-life data from the University of Chicago's Billings hospital study, which took place between 1958 and 1970. Finally, we conclude with closing remarks in Section 6.

3. AFT models based on Progressively Type-II Censoring

3.1. AFT- Exponential regression model

For the exponential AFT model, the hazard function is constant over time. Therefore, the hazard function can be written as

$$h_i(t|\mathbf{w}_i, \boldsymbol{\beta}) = \lambda \, exp\left(-\mathbf{w}_i'\boldsymbol{\beta}\right), \; i = 1, 2, ..., m. \tag{4}$$

If we view $log\lambda$ as a coefficient and place it into the regression coefficients vector $\boldsymbol{\beta}$, then (4) can be simplified to

$$h_i(t|\mathbf{w}_i, \boldsymbol{\beta}) = exp\left(-\mathbf{w}_i'\boldsymbol{\beta}\right), \; i = 1, 2, ..., m. \tag{5}$$

Thus, the survival function given the exponential distribution of the event time T is given by

$$S_i(t) = exp\left[-exp\left(logt_i - \mathbf{w}_i'\boldsymbol{\beta}\right)\right], -\infty < logt_i < \infty. \tag{6}$$

The probability density (pdf) and the log-likelihood functions are given by (7) and (8), respectively

$$f(t_{\mathbf{i}}) = exp\left[\left(logt_i - \mathbf{w}_i'\boldsymbol{\beta}\right) - exp\left(logt_i - \mathbf{w}_i'\boldsymbol{\beta}\right)\right], -\infty < logt_i < \infty. \tag{7}$$

$$L(\boldsymbol{\beta}) = log[l(\boldsymbol{\beta})] \propto \sum_{i=1}^{m} \left[\left(logt_i - \mathbf{w}_i'\boldsymbol{\beta}\right) - (R_i + 1)exp\left(logt_i - \mathbf{w}_i'\boldsymbol{\beta}\right)\right] \tag{8}$$

For the jth covariate, the MLE of β_j is the solution to the following equation

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{m} \left[-x_{ij}\left\{1 - (R_i + 1)exp\left(logt_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}\right)\right\}\right] = 0.$$

On the other hand, the second partial derivative of the log-likelihood function is used to obtain the Fisher information matrix, and it is given by

$$-\left[\left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_{j'}}\right)\right]_{(p+1)\times(p+1)} = \sum_{i=1}^{m} \left[x_{ij}x_{ij'}(R_i + 1)t_i \, exp\left(-\mathbf{w}_i'\boldsymbol{\beta}\right)\right]. \tag{9}$$

3.2. AFT-Weibull regression model

The Weibull distribution function $(W(\lambda, \delta))$ is usually formulated as an extreme value distribution since $log(T)$ can be expressed as a function of the Weibull parameters and follows extreme value distribution. Assume T is distributed as $W(\lambda, \delta)$, where λ is the scale parameter and δ is the shape parameter. Then the hazard function is given by

$$h_i(t|\mathbf{w}_i, \boldsymbol{\beta}) = (\delta^*)^{-1} e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*}\right)}, \; i = 1, 2, ..., m. \tag{10}$$

where, $\delta^* = \frac{1}{\delta}$ and $\boldsymbol{\beta}$ is the regression coefficient vector. Therefore, the survival function of T is

$$S_i(t) = exp\left[-e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*}\right)}\right], -\infty < logt_i < \infty, \tag{11}$$

with pdf

$$f(t_i) = (\delta^*)^{-1} exp\left[\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*} - e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*}\right)}\right], -\infty < logt_i < \infty. \tag{12}$$

Using the likelihood function equation in (1), the log-likelihood function based on P-II can be written as

$$L(\boldsymbol{\beta}) = log[l(\boldsymbol{\beta})] \propto \sum_{i=1}^{m} \left[(-log\delta^*) + \left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*} \right) - (R_i + 1)e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*} \right)} \right] \tag{13}$$

The MLE approach for the jth covariate parameter is obtained by solving

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{j=1}^{m} \left[\frac{-x_{ij}}{\delta^*} \left\{ 1 - (R_i + 1)e^{\left(-\frac{logt_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\delta^*} \right)} \right\} \right] = 0. \tag{14}$$

In a similar manner to the exponential AFT model, the Fisher information matrix can also be derived using the second partial derivative,

$$-E[(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_{j'}})]_{(p+1)\times(p+1)} = E\left(\sum_{i=1}^{m} \left[\frac{x_{ij}x_{ij'}}{\delta^*}(R_i + 1)e^{\left(\frac{logT_i - \mathbf{w}_i'\boldsymbol{\beta}}{\delta^*} \right)} \right] \right). \tag{15}$$

3.3. AFT Log-Logistic regression model

The hazard function for the Log-Logistic AFT model is given by

$$h_i(t|\mathbf{w}_i, \boldsymbol{\beta}) = \frac{e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right)}}{\lambda \left(1 + \frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right)}, \; i = 1, 2, ..., m, \tag{16}$$

where λ is the scale parameter for the Log-Logistic distribution. Thus, the survival function of the Log-Logistic survival time T is given by:

$$S_i(t) = \left[1 + e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right)} \right]^{-1}, -\infty < logt_i < \infty. \tag{17}$$

The log-likelihood function based on (1) simplify to

$$L(\boldsymbol{\beta}) = log[l(\boldsymbol{\beta})] \propto \sum_{i=1}^{m} \left\{ \begin{array}{c} (-log\lambda) + \left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right) - log\left[1 + \left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right) \right] \\ -(R_i + 1)log\left[1 + e^{\left(\frac{logt_i - \mathbf{w}_i'\boldsymbol{\beta}}{\lambda} \right)} \right]. \end{array} \right\} \tag{18}$$

The MLE estimate of the jth covariate's parameter is obtained by solving

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{m} \left[-\frac{x_{ij}}{\lambda} \left\{ 1 - \frac{1}{1 + \left(\frac{logt_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right)} - \frac{(R_i + 1)e^{\left(\frac{logt_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right)}}{1 + e^{\left(\frac{logt_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right)}} \right\} \right] = 0. \tag{19}$$

In addition, Fisher's information matrix is as follows:

$$-E\left[\left(\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_{j'}} \right) \right]_{(p+1)\times(p+1)} = E\left(\sum_{i=1}^{m} \frac{x_{ji}x_{j'i}}{\lambda^2} \left[\left[1 + \left(\frac{logT_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right) \right]^{-2} - \frac{(1 + R_i)e^{\left(\frac{logT_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right)}}{\left[1 + e^{\left(\frac{logT_i - \mathbf{w}_i'\hat{\boldsymbol{\beta}}}{\lambda} \right)} \right]^2} \right] \right). \tag{20}$$

Finally, the other AFT regression models for the Log-Normal and the Gamma distributions have similar derivations, and their performance will be discussed in the simulation section.

4. Simulation Studies

We conducted a simulation study based on 10,000 samples of size n to investigate the performance of the AFT models based on P-II censoring schemes. We calculated the $Bias$, MSE, test power of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$, and 95% confidence interval coverage for the hazard ratio (HR) for each set of the simulated samples. We consider, for $n = 300$, six values of m namely; $m = 180, 200, 230, 250$ and 300, providing failure information percentage: $\{(\frac{m}{n} x100\%)$; 60, 66.7, 76.7, 83.3, 93.3, and $100\%\}$.

Four sampling schemes are considered, namely, "censoring left" when $(n - m)$ items are removed at the time of the first failure and "censoring right" when $(n - m)$ items are removed at the time of the mth failure. And when the items are removed uniformly throughout the experiment, we call them censoring uniformly. Finally, the scheme with $n = m$, $R = (0, \ldots, 0)$ denotes the complete sample, represented in all Figures by a blue horizontal line. Note that, when $n = 15, m = 10$, then the censoring scheme $R = (2, 3, 0^{*8})$ means that after the first failure, two items are removed at random from the remaining 14 items, then after the second failure, three items are removed at random from the remaining 11 items. The next eight failure times are observed. For simplicity of notations, $R = (0^{*4})$ indicates $R = (0, 0, 0, 0)$. The table below provides more details on all schemes used in this paper.

Table 1. Censoring schemes and percentages

% of censored data	(n, m)	censoring left	censoring right	censoring uniformly
40%	$(300, 180)$	$(120, 0, \ldots, 0)$	$(0, 0, \ldots, 120)$	$(0, 1, 1, 0, \ldots, 0, 1, 1)$
33.3%	$(300, 200)$	$(100, 0, \ldots, 0)$	$(0, 0, \ldots, 100)$	$(1, 0, 1, 0, \ldots, 1, 0)$
23.3%	$(300, 230)$	$(70, 0, \ldots, 0)$	$(0, 0, \ldots, 70)$	$\left(7, 0^{*22}, \ldots, 7, 0^{*22}\right)$
16.67%	$(300, 250)$	$(50, 0, \ldots, 0)$	$(0, 0, \ldots, 50)$	$\left(1, 0^{*4}, 1, 0^{*4}, \ldots, 1, 0^{*4}\right)$
6.67%	$(300, 280)$	$(20, 0, \ldots, 0)$	$(0, 0, \ldots, 20)$	$\left(1, 0^{*13}, 1, 0^{*13}, \ldots, 1, 0^{*13}\right)$

Ten thousand replicates of Progressively Type-II censored samples are generated from the following AFT models: Weibull, Gamma, Log-Normal, and Log-Logistic. In our simulation, we take $\sigma = 1$, $\beta(= 0,\ 0.05,$ and $0.2)$, where, β represents the coefficient associated with the covariates. Due to the article's length, we present only some of the simulation results. The other results are similar.

We computed the $Bias$ and $MSEs$ of the estimated conditional hazard ratios from the simulated data. The values are provided in Figures 1 and 8. It is observed from Figure 1-4 that the MSE values of $\hat{HR}$ are consistently small. Moreover, as the failure rate $\frac{m}{n}$ increases, we notice that not only do the MSE values decrease, but they also become quite close to the MSEs of the complete sample. Furthermore, the censoring scheme does not affect the MSEs when data are simulated from Gamma and Log-Normal AFT models (Figure 1 and 2). However, when data are simulated from the Log-Logistic and Weibull AFT models, censoring from the right outperforms the other schemes (see Figure 1 and 2). It is easy to notice from Figure 5-8 that the $Bias$ values under all schemes are significantly small. Furthermore, the $Bias$ based on Progressively Type-II censoring is competitive with the $Bias$ based on the complete sample but is even smaller in many cases (see Figure 5-8).

The coverage probabilities are shown in tables 2 and 4. In general, the coverage probability based on Progressively Type-II censoring performs fairly similar to the coverage probability based on the uncensored data ($n = 300$) for all proposed AFT models. As for the Weibull and Log-Normal AFT models, the coverage probability based on left censoring fiercely competes with all other censoring schemes, including the complete uncensored data. It almost always reaches the desired level of 0.95, even when the failure rate is as low as 60%. From tables 3 and 4, the test power increases as the failure rate $\frac{m}{n}$ increases.

Moreover, Progressive Type-II censoring schemes in the Gamma and the Log-Normal AFT models behave nearly identical to each other. This result indicates that censoring schemes have no severe impact on the test's power when data are from the models mentioned above. It is worth noticing that the tests based on Progressive Type-II censoring are as powerful as those based on the complete uncensored sample. We can easily see from tables 3 and 4 that they all achieve close estimation to the test nominal value of 0.05 under the null hypothesis in all cases. Furthermore, the power of the test increases as m/n increases. A greater power of more than 0.9 is achieved under Weibull and Log-Normal AFT models.

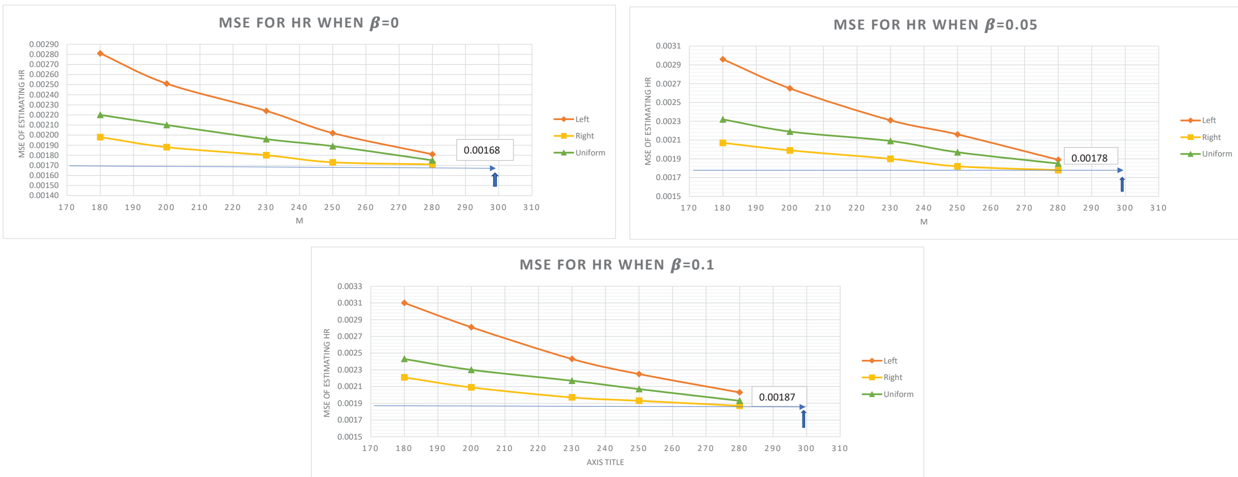

Figure 1. Graphical representation of MSE values for estimates of the Hazard Ration (HR) when data are progressively type-II censoring from the Weibull AFT model, with $n = 300$ and different m and β. The blue horizontal line represents the MSE value for the complete data.

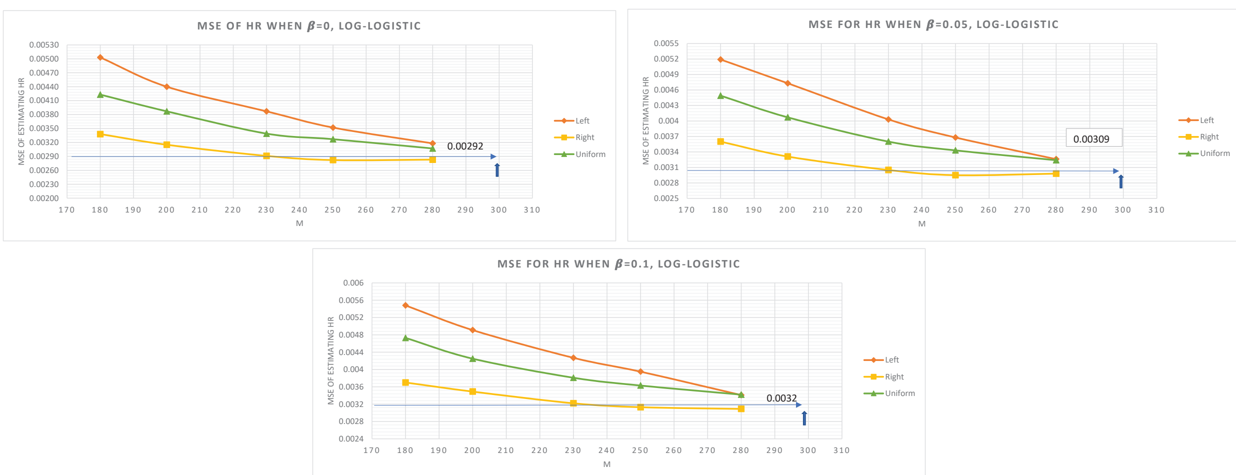

Figure 2. Graphical representation of the MSE values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Log-Logistic AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the MSE values for the complete data.

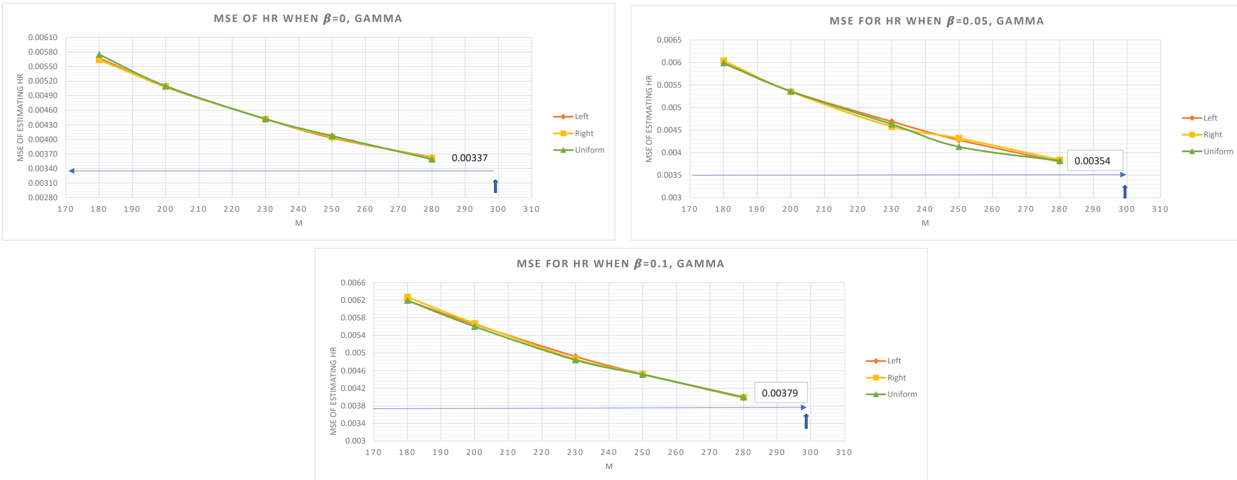

Figure 3. Graphical representation of the MSE values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Gamma AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the MSE values for the complete data.

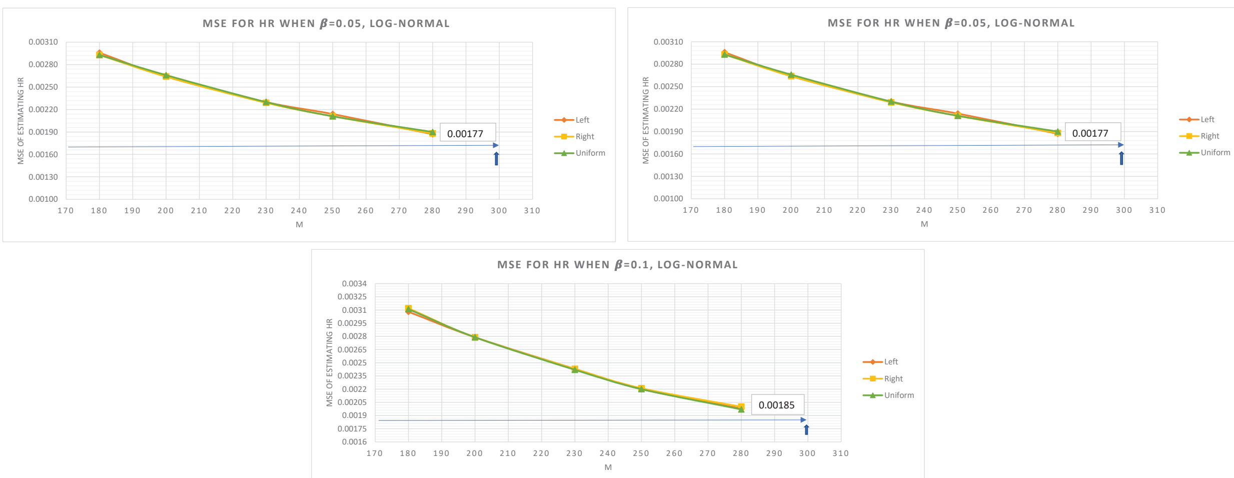

Figure 4. Graphical representation of the MSE values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Log-Normal AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the MSE values for the complete data.

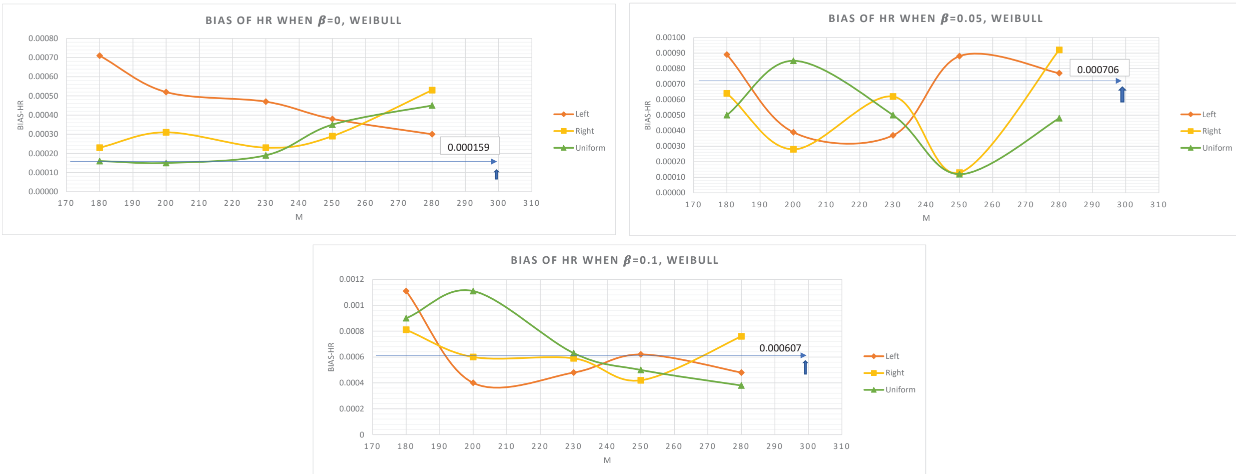

Figure 5. Graphical representation of $Bias$ values for estimates of the Hazard Ration (HR) when data are progressively type-II censoring from the Weibull AFT model, with $n = 300$ and different m and β. The blue horizontal line represents the $Bias$ value for the complete data.

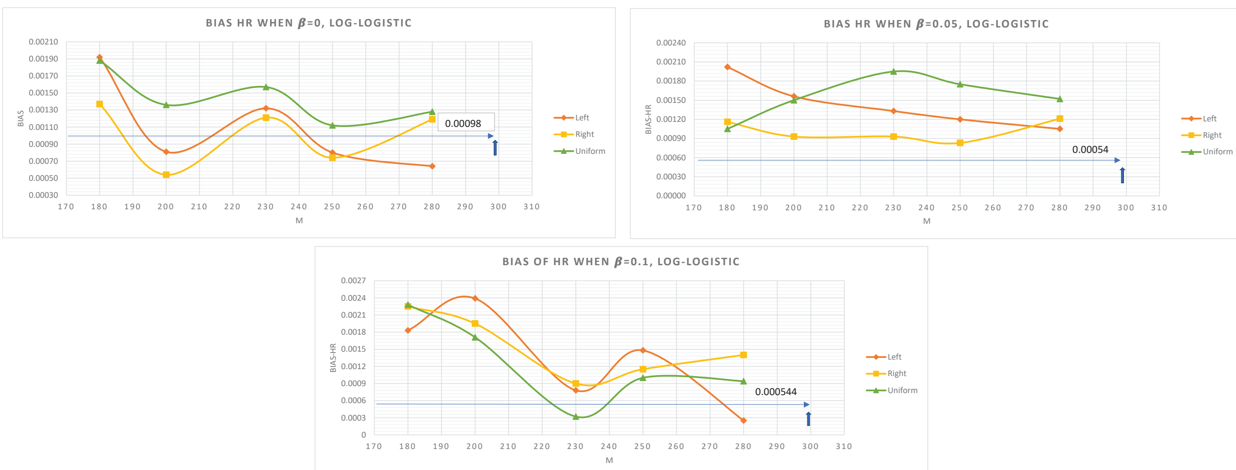

Figure 6. Graphical representation of the $Bias$ values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Log-Logistic AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the $Bias$ values for the complete data.

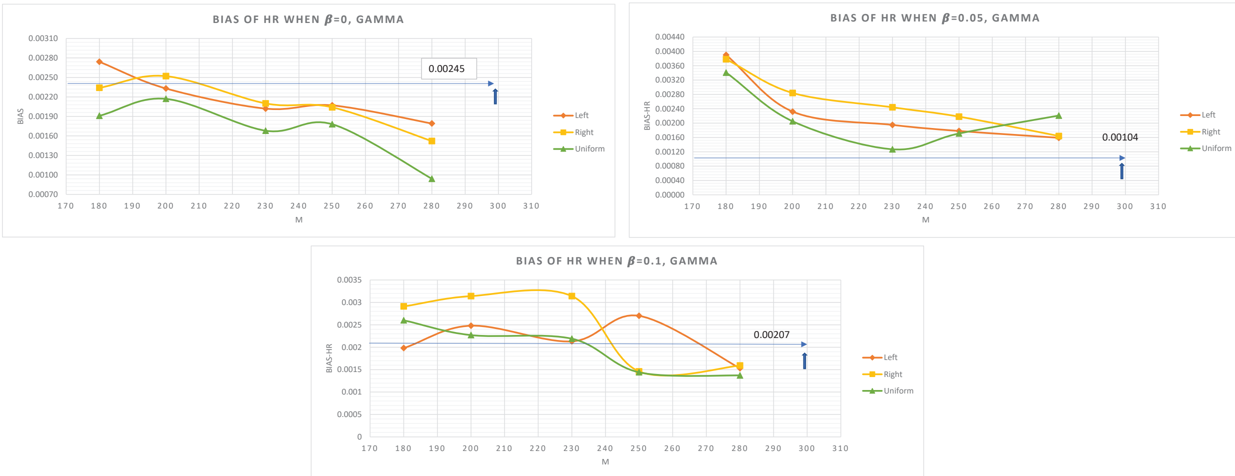

Figure 7. Graphical representation of the $Bias$ values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Gamma AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the $Bias$ values for the complete data.

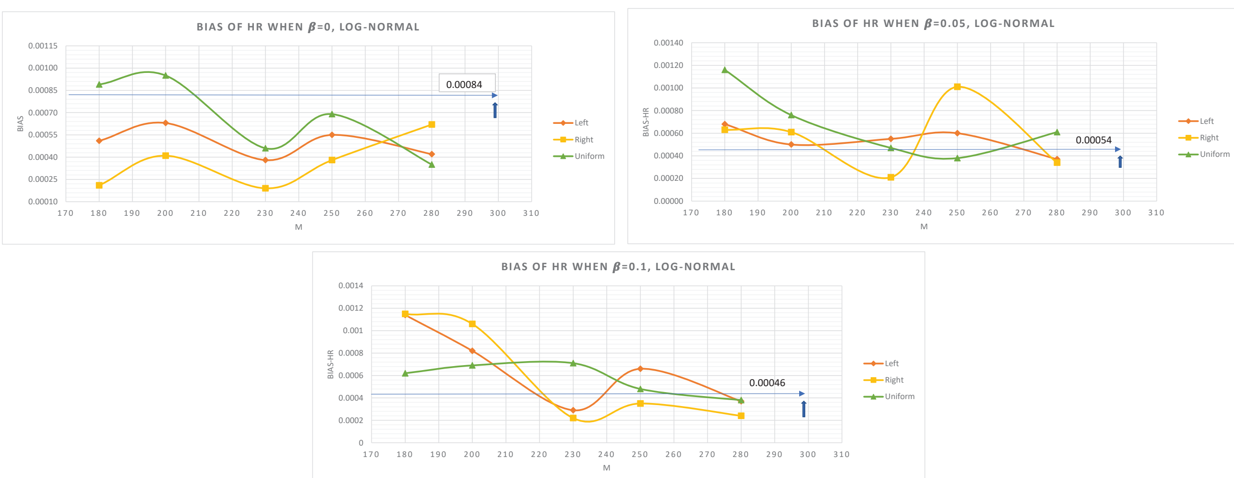

Figure 8. Graphical representation of the $Bias$ values for the estimates of the Hazard Ratio (HR) when data are Progressively Type-II censoring data from the Log-Normal AFT models, with $n = 300$ and for different values of m and β. The blue horizontal line represents the $Bias$ values for the complete data.

Table 2. Estimating the 95% confidence interval coverage for Weibull, Gamma, Log-Normal, and Log-Logistic distributions

β_1	m	Weibul Model			Gamma Model			Log-normal Model			Log-logistic Model		
		Left	Right	Uniform	Left	Right	Uniform	Left	Right	Uniform	Left	Right	Uniform
0	180	0.9471	0.8972	0.9391	0.94830	0.94800	0.94370	0.9503	0.9484	0.9471	0.9478	0.9443	0.9489
	200	0.9515	0.9004	0.9384	0.94520	0.94770	0.94520	0.9476	0.9519	0.9502	0.9491	0.9427	0.9488
	230	0.9494	0.9056	0.9457	0.94490	0.94740	0.94890	0.9521	0.9494	0.9527	0.9494	0.9455	0.9511
	250	0.9456	0.9174	0.9434	0.94750	0.94670	0.94470	0.9470	0.9508	0.9460	0.9486	0.9445	0.9500
	300	0.9482	0.9274	0.9455	0.94910	0.94700	0.94870	0.9491	0.9486	0.9469	0.9474	0.9420	0.9475
0.05	180	0.9490	0.8963	0.9400	0.9496	0.9468	0.9433	0.9483	0.9487	0.9484	0.9503	0.9418	0.9502
	200	0.9494	0.8968	0.9374	0.9505	0.9461	0.9461	0.9488	0.9508	0.9458	0.9500	0.9464	0.9467
	230	0.9469	0.9038	0.9416	0.9494	0.9496	0.9495	0.9492	0.9486	0.9496	0.9507	0.9434	0.9476
	250	0.9436	0.9133	0.9415	0.9451	0.9439	0.9436	0.9466	0.9519	0.9496	0.9504	0.9441	0.9501
	300	0.9479	0.9334	0.9507	0.9491	0.9458	0.9467	0.9510	0.9496	0.9478	0.9509	0.9360	0.9485
0.1	180	0.9486	0.8893	0.9397	0.9470	0.9496	0.9486	0.9494	0.9491	0.9471	0.9491	0.9452	0.9441
	200	0.9480	0.9017	0.9400	0.9452	0.9449	0.9471	0.9480	0.9511	0.9477	0.9504	0.9426	0.9483
	230	0.9482	0.9072	0.9456	0.9461	0.9475	0.9499	0.9471	0.9473	0.9482	0.9499	0.9449	0.9526
	250	0.9500	0.9157	0.9427	0.9436	0.9462	0.9462	0.9490	0.9482	0.9520	0.9496	0.9440	0.9484
	300	0.9462	0.9325	0.9497	0.9487	0.9495	0.9481	0.9477	0.9470	0.9491	0.9505	0.9450	0.9484

Table 3. Estimating the power of testing $HR = 1$ for Weibull, Gamma, Log-Normal, and Log-Logistic distributions, when $\alpha = 0.05$

β_1	m	Weibul Model			Gamma Model			Log-normal Model			Log-logistic Model		
		Left	Right	Uniform	Left	Right	Uniform	Left	Right	Uniform	Left	Right	Uniform
0	180	0.0529	0.1028	0.0609	0.05170	0.05200	0.05300	0.0497	0.0516	0.0529	0.0522	0.0557	0.0511
	200	0.0485	0.0996	0.0616	0.0527	0.05240	0.05250	0.0524	0.0481	0.0498	0.0509	0.0573	0.0512
	230	0.0545	0.0944	0.0543	0.05120	0.05130	0.05110	0.0466	0.0506	0.0473	0.0506	0.0545	0.0489
	250	0.0544	0.0826	0.0566	0.05230	0.05330	0.05310	0.0536	0.0504	0.0543	0.0514	0.0555	0.0502
	300	0.0518	0.0726	0.0546	0.05090	0.05300	0.05130	0.0509	0.0514	0.0531	0.0526	0.0580	0.0525
0.05	180	0.2688	0.6068	0.4271	0.1101	0.1129	0.1085	0.2709	0.2661	0.2693	0.1225	0.2196	0.1490
	200	0.2889	0.5844	0.4399	0.1069	0.1135	0.1099	0.2952	0.3084	0.2999	0.1299	0.2168	0.1526
	230	0.3256	0.5721	0.4047	0.1204	0.1147	0.1195	0.3295	0.3344	0.3282	0.1447	0.2217	0.1772
	250	0.3566	0.5366	0.4094	0.1307	0.1352	0.1364	0.3535	0.3573	0.3516	0.1488	0.2149	0.1648
	300	0.3898	0.4861	0.4138	0.1419	0.1354	0.1406	0.3885	0.3897	0.3896	0.1616	0.2042	0.1709
0.1	180	0.7621	0.9815	0.9277	0.2644	0.2737	0.2716	0.7644	0.7618	0.7713	0.3344	0.6339	0.4367
	200	0.7939	0.9813	0.9352	0.2932	0.2957	0.2913	0.8026	0.8070	0.8020	0.3721	0.6449	0.4622
	230	0.8474	0.9738	0.9226	0.3295	0.3337	0.3342	0.8485	0.8583	0.8577	0.4202	0.6354	0.4908
	250	0.8780	0.9674	0.9284	0.3610	0.3509	0.3460	0.8851	0.8851	0.8836	0.4586	0.6272	0.4908
	300	0.9284	0.9562	0.9263	0.3874	0.3859	0.3852	0.9170	0.9132	0.9136	0.4793	0.5924	0.5036

5. Real data illustration

We illustrate our proposed approach using data from the University of Chicago's Billings Hospital's study conducted between 1958 and 1970 on the survival of patients who had undergone surgery for breast cancer (see Haberman, 1976, Landwehr et al., 1984). The number of cases in the study is $n = 306$ (the number of events, $m = 81$). The data set contains four variables, Age of the patient at the time of operation, patient's year of operation (year - 1900), number of positive axillary nodes detected, and survival status. We fitted the data using Weibull, Gamma, Log-Logistic, and Log-Normal AFT models. Tables 4 & 5 present the two best-fit models to the data, namely Log-Logistic and Log-Normal based on Progressive Type-II censoring. (https://archive.ics.uci.edu/ml/datasets/Haberman's+Survival).

Table 4. The results of the AFT Log-Logistic fitted model, using Type-II right censoring

Fit Statistics (Unlogged Response)								
-2Log Likelihood	626.154							
LLogistic AIC(smaller is better)	632.154							
LLogistic AICC(smaller is better)	632.233							
LLogistic BIC(smaller is better)	643.324							
Analysis of maximum likelihood parameter estimate								
Parameter	**df**	**Estimate**	**Standard error**	**95% confidence limits**		**Chi-square**	**p-value**	
Intercept	1	4.2248	0.008	4.2092	4.2404	280843	< 0.0001	
Number of nodes Coefficient	1	-0.0027	0.0006	-0.0040	-0.0015	18.25	< 0.0001	
Scale parameter	1	0.0401	0.0035	0.0338	0.0476			

Table 5. The results of the AFT Log-Normal fitted model, using Type-II right censoring

Fit Statistics (Unlogged Response)							
-2Log Likelihood	626.154						
LLogistic AIC(smaller is better)	632.154						
LLogistic AICC(smaller is better)	632.233						
LLogistic BIC(smaller is better)	643.324						
Analysis of maximum likelihood parameter estimate							
Parameter	**df.**	**Estimate**	**Standard error**	**95% confidence limits**		**Chi-square**	**p-value**
Intercept	1	4.2262	0.0089	4.2092	4.2087	224686	< 0.0001
Number of nodes Coefficient	1	-0.0026	0.0007	-0.0040	-0.0040	13.88	0.0002
Scale parameter	1	0.0733	0.0058	0.0338	0.0627		

Tables 4 and 5 indicate that the number of nodes has a statistically significant reverse relationship to the survival time of patients' undergone surgery for breast cancer. Both Log-Logistic and Log-Normal models provide similar results.

6. Final remarks

Recently, progressive censoring has received substantial attention from many researchers. It is due to its advantages in reducing the cost and time of the tests. Moreover, the availability of high-speed computing resources enhances the focus on progressive censoring. This paper proposed a more efficient survival regression analysis method for AFT models based on Progressive Type-II censoring with four censoring schemes: censoring from right, left, uniform, and uncensored samples. First, we studied parameter estimation based on the maximum likelihood approach. Then, based on the inverse information matrix, we provided an expression for the estimated variance-covariance matrix. The results showed that despite reducing the amount of information available for estimating the hazards ratios, the Progressive Type-II censoring performance is still outstanding and even better than the estimates based on complete data. Furthermore, Progressive Type-II censoring can significantly increase power when implemented on the AFT models. The simulation study showed that, in general, the power of the test increases as the failure rate m/n increases. Moreover, Progressively Type-II censoring proves to be reasonably accurate (all $Bias$ values are less than 0.0008) with smaller MSEs and broader coverage than those under the complete sample.

REFERENCES

1. Alvarez-Andrade, S., Balakrishnan, N., & Bordes, L. (2007). Homogeneity tests based on several progressively Type-II censored samples. Journal of multivariate analysis, 98(6), 1195-1213.
2. Balakrishnan, N., & Cramer, E. (2014). The art of progressive censoring. Statistics for industry and technology.
3. Chapman, J., Trudeau, M., Pritchard, K., Sawka, C., Mobbs, B., Hanna, W., Kahn, H., McCready, D., & Lickley, L. (1992). A comparison of all-subset Cox and accelerated failure time models with Cox step-wise regression for node-positive breast cancer. Breast cancer research and treatment, 22(3), pp.263-272.
4. Haberman, S. J. (1976). Generalized residuals for log-linear models, proceedings of the 9th International Biometrics Conference.
5. Helu, A., & Samawi, H. (2017). On Marginal Distributions under Progressive Type II Censoring: Similarity/Dissimilarity Properties. Open Journal of Statistics, 7(4), 633-644.
6. Helu, A., & Samawi, H. (2019). On estimation of overlapping measures for exponential populations under progressive first failure censoring. Quality Technology & Quantitative Management, 16(5), 560-574.
7. Helu, A., & Samawi, H. (2021). Statistical Analysis Based on Adaptive Progressive Hybrid Censored Data From Lomax Distribution. Statistics, Optimization & Information Computing, 9(4), 789-808.
8. Helu, A., Samawi, H., Rochani, H., Yin, J., & Vogel, R. (2020). Kernel density estimation based on progressive type-II censoring. Journal of the Korean Statistical Society, 49(2), 475-498.
9. Keiding, N., Andersen, P. K., & Klein, J. P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. Statistics in medicine, 16(2), 215-224.
10. Komarek, A., Lesaffre, E., & Hilton, J. F. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. Journal of Computational and Graphical Statistics, 14(3), 726-745.
11. Lambert, P., Collett, D., Kimber, A., & Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. Statistics in medicine, 23(20), 3177-3192.
12. Landwehr, J. M., Pregibon, D., & Shoemaker, A. C. (1984). Graphical methods for assessing logistic regression models. Journal of the American Statistical Association, 79(385), 61-71.
13. Lee, E. T., & Wang, J. (2003). Statistical methods for survival data analysis (Vol. 476). John Wiley & Sons.
14. Liu, X. (2012). Survival analysis: models and applications. John Wiley & Sons.
15. Musleh, R. M., & Helu, A. (2014). Estimation of the inverse Weibull distribution based on progressively censored data: Comparative study. Reliability Engineering & System Safety, 131, 216-227.
16. Pushkarna, N., Saran, J., & Verma, K. (2020). Progressively Type-II right censored order statistics from Hjorth distribution and related inference. Statistics, Optimization & Information Computing, 8(2), 481-498.
17. Royston, P. (2001). The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. Statistica Neerlandica, 55(1), 89-104.
18. Samawi, H., Helu, A., & Rochani, H. (2019). Further Increasing Fisher's Information for Parameters of Association in Accelerated Failure Time Models via Double Extreme Ranks. Pakistan Journal of Statistics and Operation Research, 679-700.
19. Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. Journal of the Royal Statistical Society: Series D (The Statistician), 41(4), 455-465.

20. Saracoglu, B., Kinaci, I., & Kundu, D. (2012). On estimation of $R = P(Y < X)$ for exponential distribution under progressive type-II censoring. Journal of Statistical Computation and Simulation, 82(5), 729-744.
21. Viveros, R., & Balakrishnan, N. (1994). Interval estimation of parameters of life from progressively censored data. Technometrics, 36(1), 84-91.
22. Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Statistics in medicine, 11(14-15), 1871-1879.