UNIVERSITI TEKNOLOGI MARA
Cawangan Kedah
Kampus Sungai Petani

2021 iCMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

# e-PROCEEDINGS

## of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

**4-5 August 2021**

**Driving Research Towards Excellence**

# e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

*Driving Research Towards Excellence*

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:
Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:



Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

## TABLE OF CONTENT

## PART 1: MATHEMATICS

## PART 2: STATISTICS

## PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

## PART 4: OTHERS

# EXPONENTIAL MODEL FOR SIMULATION DATA VIA MULTIPLE IMPUTATION IN THE PRESENT OF PARTLY INTERVAL-CENSORED DATA

**Salman Umer[1] and Faiz Elfaki[2,*]**
[1,2]Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, P. O. Box 2713, Doha, Qatar

([1] su1305814@qu.edu.qa, [2] felfaki@qu.edu.qa)

Survival analysis or time-to-event analysis refers to a set of methods to analyze the time between entry to a study and a subsequent event where time to failure of an experimental unit and that could be one of the main types of censored such as Partly Interval-Censored (PIC). In this paper, the likelihoods are applied to estimate the function of survival and parameters in the exponential model when imputation methods such as Multiple Imputation (MI) and Left Imputation (LI) in the present of PIC data. The numerical evidence via simulated breast cancer data suggests that the estimates from MI are more accurate than the LI in the present of PIC data. In additional to that the patient who received chemotherapy and hormone treatment has greater survival rate than a patient who did not receive both treatments.

***Keywords***: Partly interval-censored; Exponential model; multiple imputations.

## 1. Introduction

Survival analysis is one of the important statistical methods used to analyzes time-to-event type of data which helps with decision making. Different outcomes can be generated by analyzing the data. Then, useful information and accurate inferences can be drawn. Survival analysis is highly valued because it places a high emphasis on death and component failures (Singh and Totawattage, 2013). As a general term, survival time is used, although it applies to the period of time it took for a patient to pass from remission to relapse and from diagnosis to death (Clark et al 2003). Moreover, survival analysis is utilized in a number of sectors including applied engineering, medical research, and education. Many methods have been used in the analysis of survival such as Weibull distribution. Weibull distribution is used in survival analysis to model the distribution is adequately flexible despite having only two parameters.

Furthermore, Cox's proportional hazards regression model is often used in medical articles. Additionally, for analyzing and modeling survival data, exponential distributions are one of the most convenient and useful distributions.

Sukhateme 1937 discussed the use of exponential distributions instead of normal distributions, eliminating many forms of variation. On another note, to examine failure data, Davis (1952) used exponential distributions and compared that analysis to a normal distribution. Staplin et al (2015) used a survival model based on piecewise exponential for sensitivity analysis. While, Jung et al (2018) used exponential distribution survival model based on cancer clinical trial.

A subject is left censored when they have failed before the study. While it's right censored when lost to follow before study ended or the event occur (Kleinbaum and Klein, 2005). Additionally, within an interval of time if the event of interest is occurs, the subject is interval censored. Furthermore, a study is PIC when it consists of interval-censored data and exact data (Kim, 2003); Alharpy and Ibrahim (2013); Zyoud et al (2016); Gao et al., (2017) and Saeed and Elfaki (2020). Based on the simulated medical PIC data, an exponential distribution model will be used in this study via MI and LI methods

## 2. Methodology

The most useful and convenient model in survival analysis is exponential model. The exponential distribution has been applied to biological, engineering, and actuarial problems by Teissier (1934), Weibull (1939) and Steffensen (1930). Moreover, Davis (1952) compared the analysis of failure data with the normal distribution using exponential distribution.

The general likelihood function for exponential model in the presence of censoring is;

$$L(\lambda; t) = \prod_{i=1}^{n} [(f(t_i))^{\delta_i} (S(t_i))^{1-\delta_i}] \quad (1)$$

where $f(t_i) = \lambda \exp(-\lambda t_i)$ if $\delta_i = 1$ is the density function, $S(t_i) = \exp(-\lambda t_i)$ if $\delta_i = 0$ is the function of survival, $t > 0$ and $\lambda$ ($\lambda > 0$) is the hazard rate. Then equation (1) became as;

$$L(\lambda; t) = \prod_{i=1}^{n} [(\lambda \exp(-\lambda t_i))^{\delta_i} (\exp(-\lambda t_i))^{1-\delta_i}] \quad (2)$$

The log-likelihood of the sample in equation (2) will be as follows;

$$\log[L(\beta_0, \beta; t, z)] = \prod_{i=1}^{kn} [(\delta_i (log \mu_i^{-1} - \mu_i^{-1} t_i)) + (1 - \delta_i)(-\mu_i^{-1} t_i)] \quad (3)$$

where $\lambda_i^{-1} = \beta_0 + \beta' z_i = \mu_i$ as suggested by Marubini and Valsecchi (1995), $\beta_0$ is the regression coefficient of $z_{0i}$.

Next, we will take the first and second derivation with respect to $\beta_0$ and $\beta$ and then numerical method will be used for estimating our parameters. In additional to that the Likelihood Ratio Test will be used to perform hypotheses test for the difference between patients who received the treatment and those who did via our estimations.

## 3. Simulation data

Using simulation, one can model random events, in a way in which the results are analogous to their real-world counterparts. By analyzing the simulated outcomes, researchers gain a deeper understanding and knowledge of the real world. For the purpose of evaluating alternative methods and comparing them, the simulation technique is an extremely valuable tool.

In order to examine the influence of our model, as well as compare the exploratory variable in the breast cancer data set, a simulation study is conducted. The simulation data is generated using a normal distribution, since the normal is deemed more reasonable based on the data for real breast cancer (real data is not addressed here in this paper, reader may refer to Umer, (2021)) when compared to log-logistic or Weibull distributions. Furthermore, a sample of 20000 simulation samples were used for each treatment; chemotherapy and hormone. The percentages of exact observation was generate for each treatment with 0% and 25% for PIC data (Kim 2003; Zyoud et al. (2016); and Saeed and Elfaki (2020)). Moreover, two failure types that is with treatment and without treatment were obtained from each simulation study. Next section we present the results for the simulation data.

## 4. Simulation Results

Table 1 shows the results computed by our model via MI with 0 % and 25% of exact observation of the PIC data for hormone and chemotherapy treatment. The results indicate that both treatments are significant with respect to their Standard Error (SE), and p-value.

Based on the multiple imputation technique, Figures 1 and 2 show the estimation of the survival function computed with the exponential model based on hormone treatment. This result indicate that the curves for survival function estimated from exact 0% and 25% observations, are very similar.

Similarly, the two failure types that is with chemo treatment and without chemo treatment based on exact observation of 0% and 25% through multiple imputation are shown in Figures 3 and 4 respectively. The result based on 25% of exact observation showed that both types of failure are similar which indicate that multiple imputation showed better result for high exact observation compared to lower observation as in 0%.

Figure 5 and 6 show that the estimated survival function based on chemo and hormone respectively via left point. This result showed there is different in the survival curve obtained by left point compare with exact observation.

Additionally, the results discuss that the patient who received chemotherapy and hormone treatment has greater survival rate than a patient who did not receive both treatments based on the above-mentioned results and with respect to the standard error and p-value for the estimations.

Furthermore, the null hypothesis is rejected for no difference between patients who received the treatment and those who did not based on LRT which is -151719 and -152199 for hormone and chemo respectively.

**Table 1**: Results from Hormone and Chemotherapy based on the simulation data, obtained by exponential model using MI.

| Type of Parameter | % Exact observation | Estimate | CI of 95% | SE | P-value |
|---|---|---|---|---|---|
| **Hormone** | | | | | |
| Coefficient | 0% | -0.28934 | (-0.3175, -0.2612) | 0.0143600 | 2e-16 |
| Rate | | 0.001264 | (0.00123, 0.00129) | 0.0000128 | 2e-16 |
| Coefficient | 25% | -0.28952 | (-0.3177, -0.2614) | 0.0143600 | 2e-16 |
| Rate | | 0.001264 | (0.00124, 0.00129) | 0.0000128 | 2e-16 |
| **Chemotherapy** | | | | | |
| Coefficient | 0% | -0.11309 | (-0.14121, -0.0849) | 0.0143500 | 3.29e-15 |
| Rate | | 0.001139 | (0.00112, 0.00116) | 0.0000115 | 3.29e-15 |
| Coefficient | 25% | -0.11319 | (-0.14131, -0.0851) | 0.0143500 | 3.1e-15 |
| Rate | | 0.001139 | (0.00112, 0.00116) | 0.0000115 | 3.1e-15 |

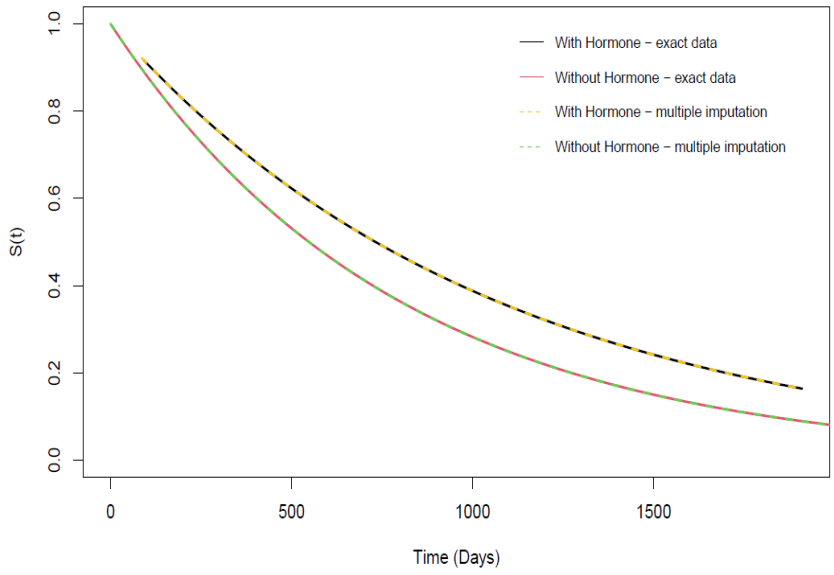


**Figure 1:** Survival function S(t) for exact data of 0% based on Hormone treatment.

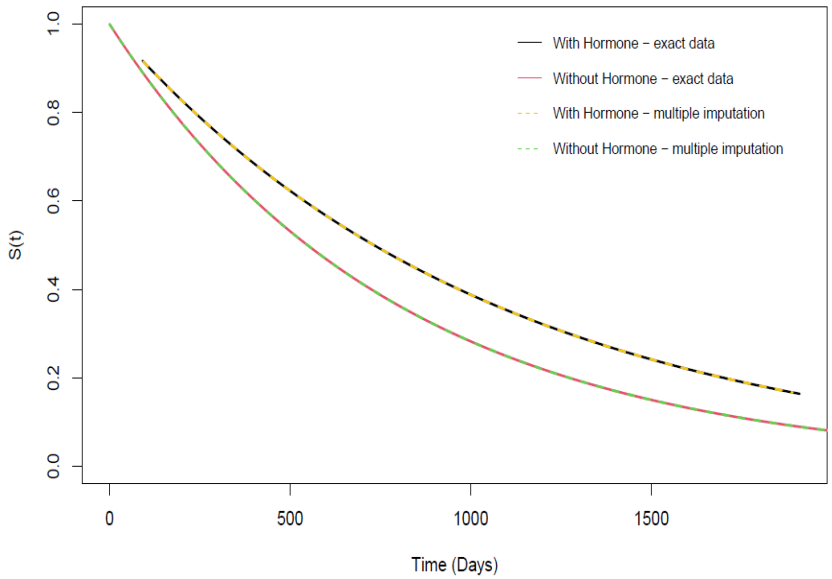


**Figure 2:** Survival function S(t) for exact data of 25% based on Hormone treatment.

.

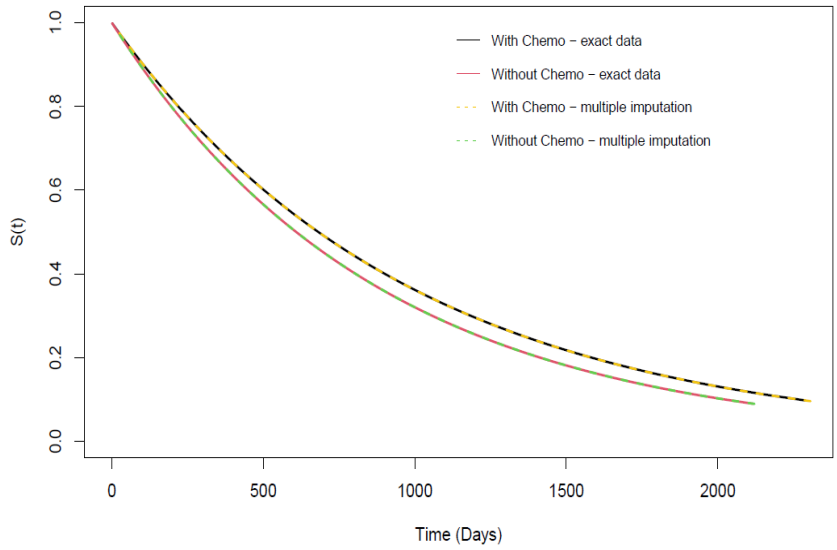


**Figure 3:** Survival function S(t) for exact data of 0% based on Chemotherapy treatment.

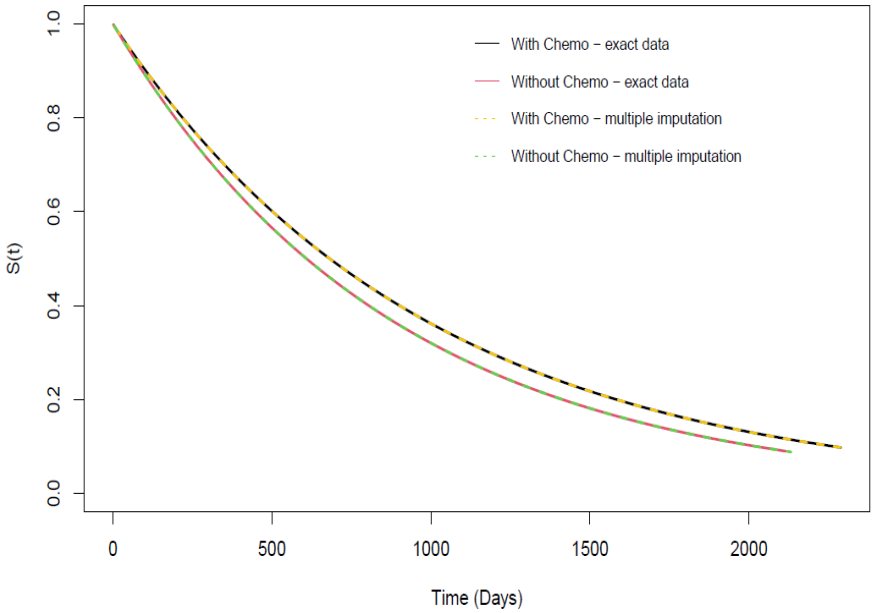


**Figure 4:** Survival function S(t) for exact data of 25% based on Chemotherapy treatment.

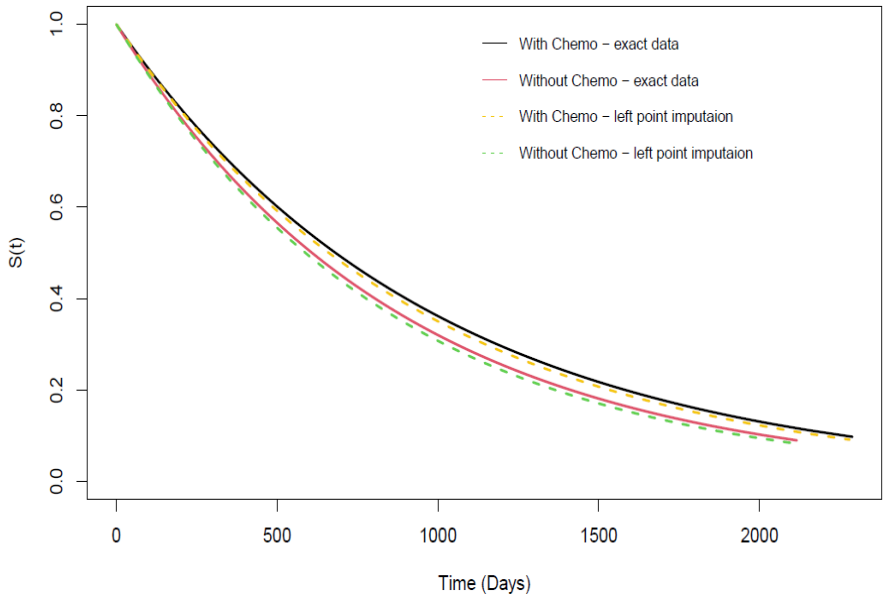


**Figure 5:** Survival function S(t) for exact data of 0% based on Chemotherapy treatment.

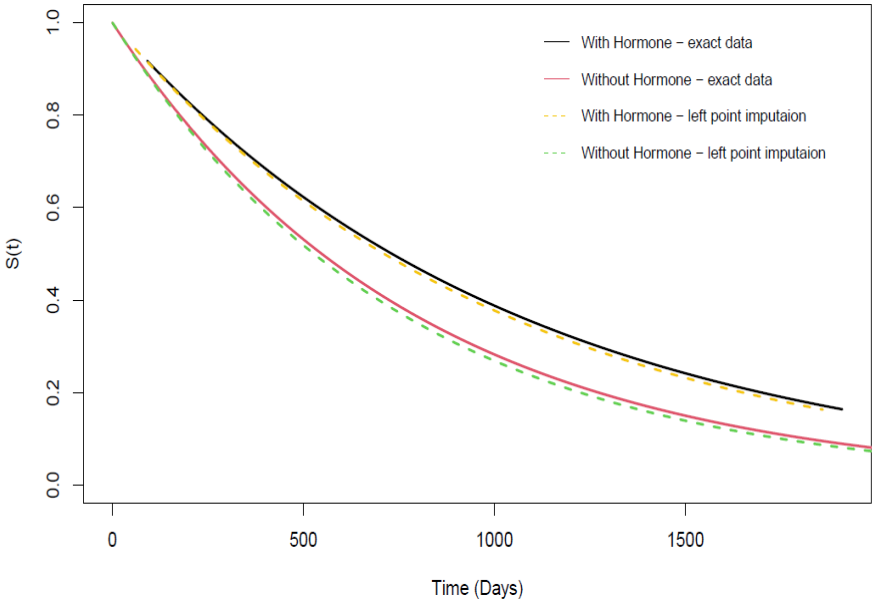


**Figure 6:** Survival function S(t) for exact data of 0% based on Hormone treatment.

## 5. Conclusion

Survival model via MI technique which assists in the process of PIC data being simplified are used based on exponential distribution. A two- month interval period is used to format the data as Partly Interval-Censored and the exact value is generated using MI technique based on simulation data. Furthermore, the models fit well with the simulated data for different percentages with exact observations of 0% and 25%.

R software was used for computation in the study. The simulation study assists in evaluating the exponential model's influence. Additionally, simulation helps to compare the covariates in the breast cancer data set. We conclude that via MI, the results from simulation are suitable for partly interval censored.

The present study focuses on the treatments in the dataset, upcoming research in this field can be stretched to include studying the properties of other parameters and different factors in the model.

**References**


Alharphy and Ibrahim (2013)."Parametric Test for Partly Interval-censored Failure Time Data under Weibull Distribution via Multiple Imputation." *Journal of Applied Science*, 13(4), pp 621-626.

Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003). Survival analysis part I: basic concepts and first analysis. *British journal of cancer*, *89*(2), 232-238.

Davis, D. J. F. (1952). An analysis of some failure data. *Journal of the American Statistical Association*, *47*(258), 113-150.

Gao, F., Zeng, D., and Lin, D. Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*, *73*(4), 1161-1168.

Jung, I., Ko, H. J., Rha, S. Y., and Nam, C. M. (2018). Reconsideration of sample size and power calculation for overall survival in cancer clinical trials. *Contemporary clinical trials communications*, *12*, 90-91.

Kim. J. S. (2003). Maximum Likelihood Estimation for the Proportional Hazards Model with Party Interval-Censored Data. *J. R. Statist. Soc.* Series B 65: 489-502.

Kleinbaum, D. G., and Klein, M. (2005). Competing risks survival analysis. Survival Analysis: A self-learning text, 391-461.

Marubini E., and Valsecchi M.G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*, 267–292, John Wiley: London.

Saeed N. M, and Elfaki F. A (2020). "Parametric Weibull Model Based on Imputations Techniques for Partly Interval-Censored Data." *Austrian Journal of Statistics*, *73*(49): 30-37.

Singh, R. S., and Totawattage, D. P. (2013). The statistical analysis of interval-censored failure time data with applications. *Open Journal of Statistics*, (3):155-166.

Staplin, N. D., Kimber, A. C., Collett, D., and Roderick, P. J. (2015). Dependent censoring in piecewise exponential survival models. *Statistical Methods in Medical Research*, *24*(3), 325-341.

Steffensen, J. F. (1930). Infantile mortality from an actuarial point of view. *Scandinavian Actuarial Journal*, *1930*(2), 272-286.

Sukhatme, P. V. (1937). Test of significance for samples of the x2-population with two degrees of freedom. *Annals of Eugenics*, *8*(1), 52-56.

Teissier, G. (1934). *Dysharmonies et discontinuités dans la Croissance.* Paris, Hermann & cie (No. 591.13 T4).

Umer, S. (2021). Exponential model for breast cancer Partly Interval-Censored data via multiple imputation. Master Theses, Qatar University.
https://qspace.qu.edu.qa/handle/10576/33/browse?type=author&value=Umer%2C+Salman.

Weibull, W. (1939). *A statistical theory of the strength of materials*. Generalstabens Litografiska Anstalts Förlag, Stockholm.

Zyoud, Elfaki and Hrairi (2016). “Non-parametric estimate based in imputations techniques for interval and Partly Interval-Censored data.” *Science International (Lahore), 28, pp 879-884.*

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4