COMPARISON OF PARAMETRIC MODELS USING RIGHT CENSORED DATA FOR BREAST CANCER PATIENTS

SYAHILA ENERA BINTI AMRAN

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Science

Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia

SEPTEMBER 2018

To my parents

Mr. Amran bin Kandati and Mrs Rosnah binti Manap

and to my siblings

Yuslafadzly and Syahrul Azlan

This humble work is a sign of my love to you.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to my supervisor, Dr. Mohd Asrul Affendi Abdullah for his continuous support and motivation, given throughout the duration of this research.

I am very grateful to the Ministry of Higher Education and the University of Tun Hussein Onn Malaysia on all financial assistance for the thesis. Besides, I also like to express my gratitude to my parents, Amran bin Kandati and Rosnah binti Manap also all my family members for untiring love and support me in doing this thesis.

Last but not least, my sincere appreciation also extends to all my fellow colleagues for sharing their knowledge and idea in helping in any situation during finishing this research. Appreciation also goes to everyone involved directly or indirectly towards the compilation of this thesis. INSYA ALLAH.

ABSTRACT

In medical research, time-to-event data commonly happen as it reflects the time until an individual has an event of interest. The event of interest can be the occurence of disease, death or the side effect of the treatment given. However, right censoring is often arising when studying the time-to-event data. The data are said to be censored when some individuals are still alive at the end of the study or lost to follow up at a certain time. One of the methods to handle the censored observation is the survival analysis. Hence, this study was carried out to analyze the right censoring survival data by using three different parametric models; exponential model, Weibull model, and log-logistic model. Data of breast cancer patients from general hospital in Johor Bahru were used to illustrate the right censoring data. When analyzing the breast cancer data, all three distributions were shown the consistency of data with the line graph of cumulative hazard function resembles a straight line going through the origin. Its show that the parametric models used in this study were appropriate to analyze the survival data. In order to determine the best parametric model in analyzing the survival of breast cancer patients, the performance of each model was compared based on the value obtained from corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC) and mean square error (MSE). Based on the model selections, the log-logistic model found to be the best model with smallest value in AICc, BIC, and MSE. Besides that, a simulation study was also carried out to see the performance of parametric models with a different number of sample sizes. The coverage probability was carried out to determine the accuracy of the simulations study. As the result, the log-logistic model was the best fitted parametric model for the survival analysis of breast cancer compared with the exponential and Weibull model.

ABSTRAK

Dalam penyelidikan perubatan, data masa yang menunjukkan tempoh masa berlakunya sesuatu peristiwa terhadap individu sering kali berlaku. Peristiwa yang sering berlaku kebiasaannya adalah ketika individu tersebut menghidap penyakit, kematian ataupun mengalami masalah selepas mendapatkan rawatan. Namun, penapisan kanan seringkali berlaku ketika mengkaji data tersebut di mana beberapa individu masih hidup pada akhir kajian atau hilang pada masa tertentu ketika kajian sedang dilakukan. Salah satu kaedah untuk mengendalikannya adalah analisis kelangsungan hidup. Oleh itu, kajian ini dijalankan untuk menganalisis data kelangsungan hidup yang mengandungi penapisan kanan dengan menggunakan tiga model parametrik yang berbeza; model eksponen, model Weibull dan model log-logistik. Data pesakit kanser payudara dari hospital di Johor Bahru telah digunakan bagi menggambarkan jenis data tersebut. Apabila menganalisis data kanser payudara, ketiga-tiga model parametrik tersebut menunjukkan konsistensi data dengan graf garis lurus untuk fungsi kumulatif bahaya. Ia menunjukkan bahawa model parametrik yang digunakan di dalam kajian ini adalah sesuai untuk menganalisis data kelangsungan hidup. Bagi menentukan model parametrik yang terbaik dalam menganalisis pesakit kanser payudara, prestasi setiap model telah dibandingkan berdasarkan pembetulan Kriteria Informasi Akaike (AICc), Kriteria Informasi Bayesian (BIC) dan purata ralat kuasa (MSE). Berdasarkan perbandingan model, model loglogistik merupakan model yang terbaik kerana mempunyai nilai yang paling kecil di AICc, BIC dan MSE. Selain daripada itu, kajian simulasi juga dijalankan untuk melihat prestasi ketiga-tiga model parametrik dengan bilangan sampel yang berbeza. Liputan kebarangkalian dikira bagi menentukan ketepatan kajian simulasi tersebut. Hasilnya, kedua-dua analisis menunjukkan bahawa model log-logistik adalah model parametrik yang terbaik untuk menganalisis kelangsungan hidup bagi kanser payudara berbanding dengan model eksponen dan model Weibull.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

β	- Beta
α	- Alpha
θ	- Theta
λ	- Lambda
χ	- Chi
Δ	- Delta
Σ	- Sum
WHO	- World Health Organisation
NBCF	- National Breast Cancer Foundation
MLE	- Maximum Likelihood Estimation
AIC_C	- Corrected Akaike Information
BIC	- Bayesian Information Criterion
MSE	- Mean square error
MREC	- Medical Research & Ethics Committee

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

In this chapter, a brief explanation of the definition of survival analysis and the breast cancer will be explained. This chapter will also consist of nine other subchapters which are the research background, problem statements, research questions, research objectives, research hypotheses, scope of the study, significance of the study and summary for this thesis. The structure of subsequent chapters in the thesis is then clearly laid out.

1.1 Background of study

Survival analysis describes the analysis of data where the outcome of survival time is the occurrence of some particular event or the end-point. The occurrence of an event can be the death from diseased, occurrence of the new disease, divorce or any experience of interest that may happen to an individual. There are many researchers interested in studying in the medical field. For example Xiong & Ji (2004), Guo & Zeng (2014), Pan *et al.* (2014) and Gong & Fang (2013).

Survival methods are usually involved with both censored and uncensored observations in order to estimate the model parameters. The survival and hazard functions are one of the key concepts for describing the distribution of the event times. The survival functions defined as the probability of the individual survive up to the event time while the hazard function defines as the potential of the event will occur given that the individuals have survived until the specific of time. In survival analysis, one of the most important aspects is to estimate the relationship between risk factors and time-to-event (Guo & Zeng, 2014). Unfortunately, the survival times

are frequently censored in medical and reliability studies (Huang, 1999). The survival time of an individual is said to be censored when there has some missing information regarding with the survival time of the individual or the subject's study. Guo & Zeng (2014) said that instead of the accurate measurement, only partial information of survival time is observed. There are generally three reasons for censoring may occur. The first one is when the person does not experience the event before study time ends, second is when the individual loss to follow-up during the study and lastly when the person withdraws from the study because of death or some other reasons. Kleinbaum & Klein (2006) illustrated the situation graphically as shown in Figure 1.1.

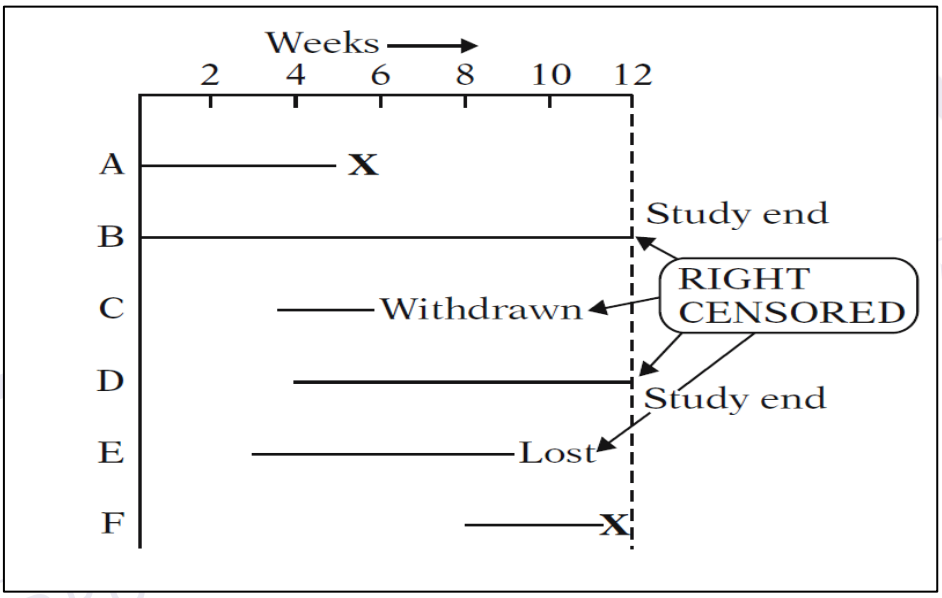

Figure 1.1: Graphically illustrated of censoring (Kleinbaum & Klein, 2006)

Figure 1.1 illustrates the censoring situation where X is the occurrence of the event. A and F are not censored since both of them have experienced the event before the study ends. The survival time for A is 5 weeks and for F, the survival is 3.5 weeks since the person enters the study at week 8 until getting the event at week 11.5. Compare with B and D, both are censored because they do not experience the event before the study ends. The survival time for B is at least 12 weeks while D has 8 weeks of survival time since the person enters the study at week 4. C enters the study between second and third weeks and after that, he is withdrawn from the study at week 6. It means the survival time is censored after 3.5 weeks and E is lost to

follow up after 6 weeks he enters the study which means censored time for E is 6 weeks .

There are three types of censoring in analyzing the survival observations which are the right censoring, left censoring and interval censoring. In survival analysis, the right censoring is the most commonly encountered form when the information of the survival time is incomplete. Right censoring occurred when a subjects stop to follow-up before observing the event. For example, suppose the patients are followed in a study for 32 weeks. However, the patients do not experience the event of interest along the duration of the study. The survival time for these patients are considered to be at least as long the duration of the study.

Another example of right censoring in clinical trials is when each patient is observed through several examinations. However, a patient may skip one or more preappointed visits and lead to the incomplete information regarding with the time of the survival. In these situations, the event time lies on an interval of the form (L, R], where L is the last time the patient is seen without having the disease, and R is the first time the patient appeared with the disease. Thus, the patient without any disease should have $R = \infty$ is said to be right censored and the patient with $L = 0$ corresponds to the left-censored (Kim *et al*., 2016).

Hence, the purpose of this study is to analyze the survival of right censoring data by using parametric models. There are three different parametric models used in this study which are the exponential model, Weibull model, and log-logistic model. In order to determine the best parametric model, the performance of each model will be compared based on the mean square error (MSE), corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC). Since censored data often occur in medical studies, it is an important issue to choose the correct or best fitting model for a given data set (Elsherpieny *et al*, 2013). Moreover, data set from breast cancer patients will be used to assess the proposed model and to illustrate the difference between these models using statistical R software. A simulation study will be conducted in order to evaluate the performance of each parametric when using different number of sample sizes.

1.2 Research motivation

There are many methods used in survival analysis. Survival methods include parametric, nonparametric and semiparametric approaches are available to analyze the relationship of a set of predictor variables with the survival time (Zhinbiou, 2008). Nonparametric models assume that the data distribution cannot be defined in the of such a finite set of parameters as it does not require any assumptions of the distribution population. The disadvantage of nonparametric models imposes few restrictions in the form of the joint distribution of data observations and the precision of estimators is often poor compared with parametric modeling (Powell, 1994).

The semiparametric model is a statistical model that includes both parametric and nonparametric models (Powell, 1994). The Cox proportional hazards regression model is the most frequently used of a semiparametric model for the analysis of survival data. It is because the Cox model allows testing the differences of survival times between two or more groups of interest. As a hybrid of the parametric and nonparametric approaches, the semiparametric models allows the estimators of the parameters are consistent under a broader range of conditions and more precise than nonparametric approach.

Parametric methods generally assume that the survival times of observations follow certain known probability distributions. Although the survival models in parametric analysis can easily be interpreted and estimated accurately, the parametric estimates can be inconsistent and give inaccurate if the assumptions of parametric distributions are violated (Hardle et al., 2012). Model parameters in these settings are usually estimated using an appropriate modification of maximum likelihood. In this study, maximum likelihood estimations are used to determine the parameter of exponential, Weibull and log-logistic distributions for the breast cancer survival analysis. Maximum likelihood estimation finds the value of the parameter that maximizes the likelihood of the current data observations (Sharif-Razavian & Zollmann, 2008).

One of the distributions that involve in this study is the exponential distributions. In survival analysis, exponential distribution is often used to model the time elapsed between events. Unlike other distributions, exponential distribution only has one parameter called rate parameter. One of the properties of the exponential

distribution is the memoryless (Meintanis & Iliopoulos, 2003). The distribution of waiting time until a certain event occurs has no relation whether or not it has happened in the past. Moreover, the hazard function of the exponential distribution is constant when the survival time is exponentially distributed.

Weibull distribution is widely used in reliability and survival analysis. It used to describe various types of failure components due to its versatility and relative simplicity. Unlike exponential, Weibull distribution has three types of parameters which are the shape, scale and location parameter. However, the value of location parameter frequently set to be zero and form as the two-parameterWeibull distribution (Lai *et al.*, 2006). Besides, the Weibull distribution includes the exponential when the shape parameter equal to one and the Rayleigh distribution when the value of shape parameter is two. The Weibull distribution also will transform into the Gumbel distribution with a simple log transformation. The shape parameter of Weibull distribution also gives an effect to the hazard function. The failure rate function of Weibull is a decreasing function when the shape parameter is less than 1, constant when the value equal to one and increasing function when the value of shape parameter is more than one.

Other than exponential and Weibull distributions, the log-logistic distribution also used in this study to analyze the survival of breast cancer patients. The log-logistic distribution can be used to model the lifetime an object or the organism. Same with Weibull distribution, the log-logistic also have a continuous shape and scale parameter. The advantages of the log-logistic distribution are the simpler hazard function and survival function. Besides, the log-logistic distribution is a mixture of Gompertz distribution and Gamma distribution with the same value of the mean and the variance equal to one (Al-Shomrani *et al*., 2016). The log-logistic distribution is closely being similar to the normal distribution and hence, the hazard function is expected to look like the hazard function of log-normal distribution. The hazard rate of this distribution initially increases and then decreases and at times can be a hump-shaped hazard.

Besides, parametric model selection criterions will be used to determine the best parametric model in this study. Model selection is an important part of any statistical analysis, and many approaches have been proposed over the years for dealing with this issue. The most common criteria used for model selections are AIC and BIC as both criteria are used for judging the quality of a model (Kadane and

Lazar, 2004). However, AIC does not appropriate in the analysis when the number of observations is small. Hence, this study is used corrected AICc, BIC and MSE as the criterion of parametric model selection. The best model is the one that provides the minimum value of the approach.

When analyzing the survival data, the number of covariates varies depending on the requirements of the researchers. Covariate is an independent variable that affects the outcome of a study. Adding a covariate to a model can increase the accuracy of the result of the study. In this study, Chi-square tests for independence and correlation coefficients are used to determine the significant relationship between two different variables. The Chi-squared test involves with hypothesis testing method where the null hypothesis is that there is no association between the two variables and the alternative hypothesis is that there is an association between the two variables in a contingency table. If the *P*-value of the table is less than 0.05, it means that the null hypothesis is rejected and accepts the alternative hypothesis (McHugh, 2013). As for correlation coefficient, if the value is close to 1, it indicates that the variables are positively linearly related to each other.

In this study, breast cancer data is used to analyze the survival of each patient by using parametric models. Breast cancer is the most common cancer among women in Malaysia. Each Malaysian woman has 1 out of 20 chance to develop the breast cancer in her lifetime (Leong *et al*., 2007). According to theWorld Health Organization (WHO), there are estimated 1.6 million new cases are diagnosed worldwide and in 2015, 560 thousand women died caused by breast cancer. Breast cancer is a disease in which the cancer cell or malignant cell was formed in the tissues of the breast. The incidence rate of breast cancer continues to show an increasing trend worldwide. This may be due to the increase in risk factors such as a number of lifestyle and genetic factors. Well-known risk factors such as nulliparity, not breastfeeding, family history and use of oral contraceptives are observed to be associated with increasing number of breast cancer cases in Malaysian women (Yip et al., 2014). Breast Cancer Welfare Association Malaysia is an organization that helps individuals with breast cancer by providing emotional, social and material support as well as education on breast health issues including early detection of breast cancer.

There are several ways to treat the breast cancer depending on its type and stage. Early breast cancer also called as stage I is the abnormal cells that begin to

develop with no longer than two centimeters and has not spread to the lymph nodes in the armpit. In this stage, it is highly treatable, however, it requires immediate treatment which the patients usually undergoes surgery or radiation. Surgery is the main treatment for stage I breast cancer. This cancer can be treated with either breast-conserving surgery (BCS) or mastectomy. According to Breast Cancer Welfare Association Malaysia, if the patients have undergone BCS, radiation therapy is usually given after surgery to lower the chance of cancer coming back in the breast. If a mastectomy is done, radiation therapy is less likely to be needed, but it might be given depending on the condition of the patients. However, chemotherapy is often not a necessary treatment for earlier stages of cancer.

As for the stage II and III or known as locally advanced breast cancer, the tumor is between two and five centimeters in size. It may have spread to the lymph nodes under the arm or at the surrounding of breast tissue. The treatment of locally advanced breast cancer requires a combination of systemic chemotherapy, surgery, and radiotherapy to optimize the chance of cure (Giordano, 2003). However, the treatment often starts with chemotherapy. The purpose of giving chemotherapy first is to make the breast cancer smaller and to destroy any cancer cells that may be spread elsewhere in the body, but cannot be detected using routine tests.

Stage IV or known as metastatic breast cancer is the last stage in the breast cancer. Metastatic breast cancer cannot be cured as cancer has spread beyond the breast to other organs in the body, most often the bones, lungs, liver or, less commonly, brain (Yip *et al*., 2006). Generally, breast cancer treatment can be divided into two parts which are local therapy and systemic therapy. Local therapy includes surgery and radiation therapy for the breast area while systemic therapy includes chemotherapy, hormone therapy, and targeted therapy. Most women with stage IV breast cancer are treated with systemic therapy while local treatments sometimes might also be used to help shrink tumors or treat symptoms. According to National Breast Cancer Foundation (NBCF), the symptoms of patients at stage IV depends on the condition in which cancer has spread in the body. Based on the previous study, breast cancer patients can be treated combining the local therapy with systematic therapy (Yip *et al*., 2014).

1.3 Problem statement

There are lots of methods used to analyze the survival data which include parametric, nonparametric and semiparametric models. However, censored data are usually observed in studying time-to-event, where the observation is only partial instead of the precise information. Besides, the efficient estimation procedure and the corresponding computation will not be developed. Because of that, it is an important issue to handle the censoring data in order to prevent any bias information in the analysis.

Analyzing censoring data either right censored, left censored or interval censored data has been a major challenge in medical research. Even though many nonparametric and semi-parametric statistical methods were developed over two decades ago, these methods were not well understood by the medical community due to their complexity. Parametric methods are easy to implement, unfortunately, this method received little attention in practice. It is because the impact of model misspecification on the estimation results were not well understood (Gong & Fang, 2013). Moreover, it is an important issue in choosing the correct distribution for a given data set. Many researchers are faced with the problem of choosing the correct model (Elsherpieny et al., 2013). If the wrong model is chosen, the subsequent calculations will be incorrect, and that will certainly result in wrong decisions.

Breast cancer remains the most common cancer among women in the most of developed countries. The incidence of the breast cancer in most of the countries is still on the increase. In Brazil, breast cancer is the second cause of women dying because women are still afraid to search for medical help and only tend to do it after the cancer is the advanced stage (Santos *et al*., 2010). Two main factors that cause the increasing number of breast cancers in Malaysia are lack of awareness of breast cancer among women and the strong influence of traditional and cultural belief. Because of the strong traditional influence, many women in rural areas will initially seek traditional or alternative treatment such as faith healing or known as bomoh before they go to the hospital after the initial treatment has failed (Leong *et al*,, 2007). Although there are activities to increase the awareness of breast cancer such as talk, seminar, and education on the prevention and treatment, it only effective among women living in urban areas. Hence, this study will conduct an analysis by

using the breast cancer data to improve the survival rate of the breast cancer incidence.

Sample size plays an important role in survival analysis. The results of the analysis may differ when performing on a limited number of specimens or patients and on a large sample size. A study with small sample size may produce inconclusive results, biased outcomes and does not represent the population of the study. Sample size should be considered because it allows the researcher to estimate the precision of the experiment to yield or to control for the risk of reporting a false-negative finding (Biau *et al*., 2008). Hence, this study is interested to analyze the effect of different sample size on the right censored data through simulation.

1.4 Research objectives

This study embarks on the following objectives which are:

1. To analyze the survival of breast cancer data using exponential,Weibull and loglogistic model.
2. To measure the performance of the parametric model based on the corrected Akaike Information Bayesian (AICc), Bayesian Information Criterion (BIC), mean square error (MSE) and mean bias (MB).
3. To validate the performance of the parametric model with different sample sizes through a simulation study.

1.5 Scope of study

Generally, the aim of this study was to determine the best parametric model in right censored data by comparing the performance of each parametric model in term of MSE, AIC_c, BIC, and mean bias. The breast cancer patients data was used in this study and will be collected at general hospital in Johor Bahru. It was chosen as the location of this study since there was one of the currently reference hospitals for the breast cancer treatment. The following data were collected using a secondary data and it consist of 38 female infected patients. The duration of follow up period was 9 years where it started from 30 Disember 2008 until 15 February 2017. This study involves the survival time of the breast cancer patient, t as the dependent variable while the independent variables are the age in years, X_1 and treatment given to the

patients X_2. For the treatment of breast cancer, the study indicates "1" for the local therapy treatment, "2" indicates the systematic therapy treatment and "3" indicates the patients receive both local and systematic therapy treatment. By using statistical R software, the right censored data will be analyzed in order to determine the best parametric model. Right censored data with different sample of size also will be analyzed to evaluate the performance of each parametric model through simulation study.

1.6 Significance of study

Nowadays, there are many problems that arise in medical studies due to the increasing in the number of cases diseased women in Malaysia and breast cancer is one of the diseases that is gained more attention in the community. As there are increasing number of breast cancer diseases, an appropriate statistical tool can be used in order to improve the solving of these problems. This situation has lead this study to analyze the breast cancer patients with right censored data in medical studies. Hence, the significant contribution of this study is to improve the methods in solving the breast cancer problems with right censored data by comparing the performance of parametric models and help to reduce the number of breast cancer cases in this country.

Medical sector has a large impact on the economy of this country by providing adequate facilities and services to treat their patients. Since the number of breast cancer diseases increases day by day, the service provided by the health care become limited and lead to the poor management system. The system will be missed out some important information such as type of treatment given or stage of breast cancer on the patients body. This situation can lead the miscalculation on the survival rate of the patients. Since this study will provide the best-fitted model in survival analysis, it can help the management system to handle the missing information effectively The identification of significant treatments also can reduce the mortality rates of breast cancer disease.

Besides that, this study also can improve the selection of sample size in right censored data. Most of previous studies compared between two different methods or models without considering the precision of the anticipated estimation. Moreover, sample size should be considered when planning a study because it allows the

researcher to control for the risk of reporting a false-negative finding. Hence, by constructing a simulation study with different number of sample size, researcher can find out the effect of sample size to the performance of the parametric models. This study also provide the best error measurement in parametric model selection when using right censored data in which future researches can use it in their studies.

1.7 Structure of thesis

In Chapter 2, a brief review of survival analysis literature is discussed. In particular, the different but equivalent ways of studying parametric modelling in survival analysis are distinguished. Furthermore, a literature review of statistical modelling for breast cancer survival is presented and analysed.

A specific data set consisting of 38 female infected patients with different age is presented in Chapter 3. This data originates from Hospital Sultan Ismail, Johor Bahru and will be used as the principal application throughout the thesis. Three different parametric models which are Weibull, exponential and log-logistic distributions are formulated. The parametric model selection is implemented using AIC_c, BIC, MSE and mean bias approach. Then, a simulation with different number of sample size is applied to the breast cancer data set. Hence, in Chapter 4 explaining the results were analysed, table and figure were discussing in this chapter.

Finally, Chapter 5 provides conclusions and further considerations for modeling the survival analysis with other covariates and distributions. Potential areas for future work are also highlighted.

1.8 Summary

This chapter has introduced the underlying motivations behind this thesis. The study of statistical models to analyze the survival of breast cancer patients, which are general enough to apply in other applications. In addition, the aims and structure of the thesis were clearly outlined. In the subsequent Chapter 2, statistical models for survival analysis are reviewed and a literature review of breast cancer cases are provided.

CHAPTER 2

LITERATURE REVIEW

In this chapter, previous findings by other researchers related to this study have been read through before they are used as references to support our findings. Based on the previous study, there were not much previous findings on right censored data with parametric model. Most of the past studies are interested semi and nonparametric models. This chapter is divided into four different subtopics which are the breast cancer, survival analysis, right censored data and lastly the model selection.

2.1 Breast cancer

Breast cancer has become one of the common diseases among women in Malaysia as well as worldwide. According to Malaysian Oncological Society, about one in 19 women in this country is at risk to suffer from breast cancer compared to one in eight in Europe and United States. There are some previous researches study the breast cancer incidence in Malaysia and summarise the impact of the research findings to clinical practice. Approximately 50 percent of Malaysian women have the cancer at earlier age which are before the age of 50 years compared to women in Western countries, 20 percent are diagnosed before the age of 50 years (Yip *et al*., 2006). Hisham & Yip (2004) also found the same finding as the prevalent age group is 40 to 49 years old for women to diagnose with the breast cancer. In fact, both studies also found out that Malay ethnicity is significantly associated with the large tumor in breast cancer when compared with Chinese and Indians ethnicities (Yip *et al*., 2006; Hisham & Yip, 2004).

Most of the previous researchers believe that the factors contribute to the increasing number of breast cancer are the lack of awareness of breast cancers

REFERENCES

Abadi, A., Yavari, P., Dehghani-Arani, M., Alavi-Majd, H., Ghasemi, E., Amanpour, F. & Bajdik, C. (2014). Cox models survival analysis based on breast cancer treatments. *Iranian journal of cancer prevention*, 7(3), p. pp. 124.

Abdullah, N.A., Wan Mahiyuddin, W., Muhammad, N.A., Ali, Z.M., Ibrahim, L., Ibrahim Tamim, N. et al. (2013). Survival rate of breast cancer patients in malaysia: a population-based study. *Asian Pac J Cancer Prev*, 14(8), pp. pp. 4591–4621.

Al-Shomrani, A.A., Shawky, A., Arif, O.H. & Aslam, M. (2016). Log-logistic distribution for survival data analysis using mcmc. SpringerPlus, 5(1), p. pp. 1774.

Alharpy, A.M. & Ibrahim, N.A. (2013). Parametric tests for partly interval-censored failure time data under weibull distribution via multiple imputation. *Journal of Applied Sciences*, 13(4), p. pp. 621.

Allison, P.D. (2010). Survival analysis using SAS: a practical guide. *Sas Institute*.

Aron, A., Guo, H., Mettas, A. & Ogden, D. (2009). Improving the 1-parameter weibull: A bayesian approach. In: Reliability and Maintainability Symposium, 2009., IEEE, pp. pp. 25–30.

Baghestani, A.R., Moghaddam, S.S., Majd, H.A., Akbari, M.E., Nafissi, N. & Gohari, K. (2014). Survival analysis of patients with breast cancer using weibull parametric model. *Asian Pacific journal of cancer prevention*, 16(18), pp. pp. 8567–8571.

Beadle, B.M., Woodward, W.A. & Buchholz, T.A. (2011). The impact of age on outcome in early-stage breast cancer. *In: Seminars in Radiation Oncology*, 21(1), pp. pp. 26–34.

Besjakov, J. & Akesson, K. (2008). Long-term survival and fracture risk after hip fracture: a 22-year follow-up in women. *Journal of Bone and Mineral*

Research, 23(11), pp. pp. 1832–1841.

Bhattacharya, P. & Bhattacharjee, R. (2010). A study on weibull distribution for estimating the parameters. *Journal of Applied Quantitative Methods*, 5(2), pp. pp. 234–241.

Biau, D.J., Kerneis, S. & Porcher, R. (2008). Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clinical orthopaedics and related research*, 466(9), pp. pp. 2282–2288.

Buis, M.L. (2006). An introduction to survival analysis. *Department of Social Research Methodology Vrije Universiteit Amsterdam [Online]*.

Burnham, K.P. & Anderson, D.R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. *Springer Science & Business Media.*

Chaouch, M., Laib, N. & Said, E.O. (2016). Nonparametric m-estimation for right censored regression model with stationary ergodic data. *Statistical Methodology*, 33, pp. pp. 234–255.

Cheng, S.H.C., Yu, B.L., Horng, C.F., Tsai, S.Y., Chen, C.M., Chu, N.M., Tsou, M.H., Lin, C.K., Shih, L.S. & Liu, M.C. (2016). Long-term survival and stage i breast cancer subtypes. *Journal of Cancer Research and Practice*, 3(1), pp. pp. 1–8.

Choi, I., Wells, B.J., Yu, C. & Kattan, M.W. (2011). An empirical approach to model selection through validation for censored survival data. *Journal of biomedical informatics*, 44(4), pp. pp. 595–606.

Cohen, A.C. (1965). Maximum likelihood estimation in the weibull distribution based on complete and on censored samples. *Journal of Technometrics*, 7(4), pp. pp. 579–588.

Collett, D. (2015). Modelling survival data in medical research. *CRC press*.

Cong, C. & Sambandham, S. (2009). Survival and relapse time among different histology types of breast cancer. *Neural, Parallel and Scientific Computations*, 17(3), p. pp. 281.

Cong, C. & Tsokos, C.P. (2010). Parametric and nonparametric analysis of breast cancer treatments. *International Journal of Biomedical Sciences*, 5(2), p. pp. 1.

Conkin, J. (2001). A log logistic survival model applied to hypobaric decompression sickness. *Fifty-first Workshop of the Undersea and Hyperbaric Medical*

Society, Seattle, pp. pp. 75–93.

Corradini, M.G. & Peleg, M. (2004). Demonstration of the applicability of the weibull–log-logistic survival model to the isothermal and nonisothermal inactivation of escherichia coli k-12 mg1655. *Journal of food protection*, 67(11), pp. pp. 2617–2621.

Cousineau, D. & Helie, S. (2013). Improving maximum likelihood estimation using prior probabilities: A tutorial on maximum a posteriori estimation and an examination of the weibull distribution. *Tutorials in Quantitative Methods for Psychology*, 9(2), pp. pp. 61–71.

Crowther, M.J. & Lambert, P.C. (2014). A general framework for parametric survival analysis. *Statistics in medicine*, 33(30), pp. pp. 5280–5297.

Crumer, A.M. (2011). comparison between weibull and cox proportional hazards models. *Kansas State University*: Ph.D. Thesis.

Dalgaard, P. (2008). *Introductory statistics with R.* Springer Science & Business Media.

Dey, A.K. & Kundu, D. (2009). Discriminating between the log-normal and loglogistic distributions. *Communications in Statistics-Theory and Methods*, 39(2), pp. pp. 280–292.

Dziak, J.J., Coffman, D.L., Lanza, S.T. & Li, R. (2012). Sensitivity and specificity of information criteria. *The Methodology Center and Department of Statistics, Penn State, The Pennsylvania State University*, pp. pp. 1–10.

Elfaki, F.A.M., Azram, M. & Usman, M. (2012). Parametric cox's model for partly interval-censored data with application to aids studies. *International Journal of Applied Physics and Mathematics*, 2(5), p. pp. 352.

Elsherpieny, E.A., Ibrahim, S.A. & Radwan, N. (2013). Discriminating between weibull and log-logistic distributions. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(8), pp. pp. 3358–3371.

Gamel, J.W., Vogel, R.L., Valagussa, P. & Bonadonna, G. (1994). Parametric survival analysis of adjuvant therapy for stage ii breast cancer. *Cancer*, 74(9), pp. pp. 2483–2490.

Giordano, S.H. (2003). Update on locally advanced breast cancer. *The oncologist*, 8(6), pp. pp. 521–530.

Gong, Q. & Fang, L. (2013). Comparison of different parametric proportional hazards models for interval-censored data: A simulation study.

Contemporary clinical trials, 36(1), pp. pp. 276–283.

Grosbois, V., Gimenez, O., Gaillard, J.M., Pradel, R., Barbraud, C., Clobert, J., Moller, A. & Weimerskirch, H. (2008). Assessing the impact of climate variation on survival in vertebrate populations. *Biological Reviews*, 83(3), pp. pp. 357–399.

Grzenda, W. (2012). Bayesian exponential survival model in the analysis of unemployment duration determinants. *Acta Universitatis Lodziensis. Folia Oeconomica*, (Acta Universitatis Lodziensis, Folia Oeconomica nr 269/2012).

Guo, S. & Zeng, D. (2014). An overview of semiparametric models in survival analysis. *Journal of Statistical Planning and Inference*, 151, pp. pp. 1–16.

Gupta, R.C., Akman, O. & Lvin, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal*, 41(4), pp. pp. 431–443.

Haberman, S.J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, pp. pp. 815–841.

Han, G., Schell, M.J. & Kim, J. (2014). Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Statistics in medicine*, 33(1), pp. pp. 59–73.

Hardle, W.K., M¨uller, M., Sperlich, S. & Werwatz, A. (2012). *Nonparametric and semiparametric models*. Springer Science & Business Media.

Hassan, A.S., Hemeda, S.E., Maiti, S.S. & Pramanik, S. (2017). The generalized additive weibull-g family of distributions. *International Journal of Statistics and Probability*, 6(5), p. pp. 65.

Hirose, K., Kawano, S., Konishi, S. & Ichikawa, M. (2011). Bayesian information criterion and selection of the number of factors in factor analysis models. *Journal of Data Science*, 9(2), pp. pp. 243–259.

Hisham, A.N. & Yip, C.H. (2004). Overview of breast cancer in malaysian women: a problem with late diagnosis. *Asian Journal of Surgery*, 27(2), pp. pp. 130–133.

Hu, T. & Xiang, L. (2016). Partially linear transformation cure models for interval censored data. *Computational Statistics & Data Analysis*, 93, pp. pp. 257–269.

Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica*, pp. pp. 501–519.

Jackson, C.H. (2016). flexsurv: a platform for parametric survival modeling in r. *Journal of Statistical Software*, 70.

Jafari, A.A. & Zakerzadeh, H. (2015). Inference on the parameters of the weibull distribution using records. *arXiv preprint arXiv*:1501.02201.

Johnson, J.B. & Omland, K.S. (2004). Model selection in ecology and evolution. *Trends in ecology & evolution*, 19(2), pp. pp. 101–108.

Kartsonaki, C. (2016). Survival analysis. *Diagnostic Histopathology*, 22(7), pp. pp. 263–270.

Khan, S.A. & Khosa, S.K. (2016). Generalized log-logistic proportional hazard model with applications in survival analysis. *Journal of Statistical Distributions and Applications*, 3(1), p. pp. 16.

Kim, J., Kim, Y.N. & Kim, S.W. (2016). Frailty model approach for the clustered interval-censored data with informative censoring. *Journal of the Korean Statistical Society*, 45(1), pp. pp. 156–165.

Kleinbaum, D. & Klein, M. (2006). Survival analysis: a self-learning text, springer science & business media. *Journal of New York*.

Kundu, D. & Howlader, H. (2010). Bayesian inference and prediction of the inverse weibull distribution for type-ii censored data. *Computational Statistics & Data Analysis*, 54(6), pp. pp. 1547–1558.

Kundu, D. & Mitra, D. (2016). Bayesian inference of weibull distribution based on left truncated and right censored data. *Computational Statistics & Data Analysis*, 99, pp. pp. 38–50.

Kundu, D., Mitra, D. & Ganguly, A. (2017). Analysis of left truncated and right censored competing risks data. *Computational Statistics & Data Analysis*, 108, pp. pp. 12–26.

Kysely, J. (2010). Coverage probability of bootstrap confidence intervals in heavy tailed frequency models, with application to precipitation data. *Theoretical and applied climatology*, 101(3-4), pp. pp. 345–361.

Lai, C., Xie, M. & Murthy, D. (2003). A modified weibull distribution. *IEEE Transactions on reliability*, 52(1), pp. pp. 33–37.

Lai, C.D., Murthy, D. & Xie, M. (2006). Weibull distributions and their applications. *In: Springer handbook of engineering statistics, Springer London*, pp. pp. 63– 78.

Leong, B., Chuah, J.A., Kumar, V.M. & Yip, C.H. (2007). Breast cancer in sabah,

malaysia: a two year prospective study. *Asian Pac J Cancer Prev*, 8(4), pp. pp. 5–9.

Lin, C. & Zhou, Y. (2016). Semiparametric varying-coefficient model with right censored and length-biased data. *Journal of Multivariate Analysis*, 152, pp. pp. 119–144.

Lyles, R.H., Guo, Y. & Greenland, S. (2012). Reducing bias and mean squared error associated with regression-based odds ratio estimators. *Journal of statistical planning and inference*, 142(12), pp. pp. 3235–3241.

Ma, J., Heritier, S. & Lo, S.N. (2014). On the maximum penalized likelihood approach for proportional hazard models with right censored survival data. *Computational Statistics & Data Analysis*, 74, pp. pp. 142–156.

Ma, L., Hu, T. & Sun, J. (2016). Cox regression analysis of dependent interval censored failure time data. *Computational Statistics & Data Analysis*, 103, pp. pp. 79–90.

McCloskey, D.N. & Ziliak, S.T. (1996). The standard error of regressions. *Journal of economic literature*, 34(1), pp. 97–114.

McHugh, M.L. (2013). The chi-square test of independence. *Biochemia medica: Biochemia medica*, 23(2), pp. pp. 143–149.

Meintanis, S.G. & Iliopoulos, G. (2003). Characterizations of the exponential distribution based on certain properties of its characteristic function. *Kybernetika*, 39(3), pp. pp. 295–298.

Montaseri, M., Charati, J.Y. & Espahbodi, F. (2016). Application of parametric models to a survival analysis of hemodialysis patients. *Nephro-urology monthly*, 8(6).

Naik, P.A., Shi, P. & Tsai, C.L. (2007). Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association*, 102(477), pp. pp. 244–254.

Norsaadah, B., Rusli, B., Imran, A., Naing, I. &Winn, T. (2005). Risk factors of breast cancer in women in kelantan, malaysia. *Singapore medical journal*, 46(12), p. pp. 698.

Okello, J.O. & Ka, D.A. (2014). Parametric models and future event prediction base on right censored data. *American Journal of Mathematics and Statistics*, 4(5), pp. pp. 205–213.

Osman, M. & Ghosh, S.K. (2012). Nonparametric regression models for right

censored data using bernstein polynomials. *Computational Statistics & Data Analysis*, 56(3), pp. pp. 559–573.

Pan, C., Cai, B., Wang, L. & Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Computational Statistics & Data Analysis*, 74, pp. pp. 198–208.

Pant, M.D. (2017). Characterizing log-logistic (ll) distributions through methods of percentiles and l-moments. *Department of Curriculum and Instruction, University of Texas at Arlington.*

Pantazis, N., Kenward, M.G. & Touloumi, G. (2013). Performance of parametric survival models under non-random interval censoring: A simulation study. *Computational Statistics & Data Analysis*, 63, pp. pp. 16–30.

Pearson, K. (1909). Determination of the coefficient of correlation. *Journal of Science*, 30(757), pp. pp. 23–25.

Pham, M.H. (2014). Survival analysis-breast cancer. *Undergraduate Journal of Mathematical Modeling*: One+ Two, 6(1), p. pp. 4.

Pourhoseingholi, M.A., Vahedi, M. & Rahimzadeh, M. (2013). Sample size calculation in medical studies. *Gastroenterology and Hepatology from bed to bench*, 6(1), p. 14.

Powell, J.L. (1994). Estimation of semiparametric models. *Handbook of econometrics*, 4, pp. pp. 2443–2521.

Reath, J., Dong, J. & Wang, M. (2016). Improved parameter estimation of the loglogistic distribution with applications. *Computational Statistics*, pp. pp. 1–18.

Reed,W.J. (2011). A flexible parametric survival model which allows a bathtub-shaped hazard rate function. *Journal of Applied Statistics*, 38(8), pp. pp. 1665–1680.

Rezaianzadeh, A., Peacock, J., Reidpath, D., Talei, A., Hosseini, S.V. & Mehrabani, D. (2009). Survival analysis of 1148 women diagnosed with breast cancer in southern iran. *BMC cancer*, 9(1), p. pp. 1.

Santos, L., Lima, R., Bezerra, L., Lyra, P. & Conci, A. (2010). Parametric analysis on the influences of tumor position and size in breast temperature profile. *In: Proceedings of 17th International Conference on Systems, Signals and Image Processing*, pp. pp. 478–481.

Schulz, C., Hafner, H.M., Breuninger, H. & Leiter, U. (2014). Local recurrence and

survival in acral lentiginous melanoma comparing 3d histology and conventional histology. JDDG: *Journal der Deutschen Dermatologischen Gesellschaft*, 12(10), pp. pp. 881–889.

Scott, S., Jones, R. & Williams, W. (1984). Review of data analysis methods for seed germination. *Crop science*, 24(6), pp. pp. 1192–1199.

Seco, G.V., Herrero, E.T., Perez, J.C.N. & Rosario, P. (2014). Performance evaluation of recent information criteria for selecting multilevel models in behavioral and social sciences. *International Journal of Clinical and Health Psychology*, 14(1), pp. pp. 48–57.

Sen, B. & Xu, G. (2015). Model based bootstrap methods for interval censored data. *Computational Statistics & Data Analysis*, 81, pp. pp. 121–129.

Sharif-Razavian, N. & Zollmann, A. (2008). An overview of nonparametric bayesian models and applications to natural language processing. *Science*, pp. pp. 71–93.

Shen, P.s. (2015). Conditional mle for the proportional hazards model with left truncated and interval-censored data. *Statistics & Probability Letters*, 100, pp. pp. 164–171.

Shi, J., Chen, X. & Zhou, Y. (2015). The strong representation for the nonparametric estimator of length-biased and right-censored data. *Statistics & Probability Letters*, 104, pp. pp. 49–57.

Siannis, F., Copas, J. & Lu, G. (2005). Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6(1), pp. pp. 77–91.

Singh, H.P. & Mehta, V. (2017). Improved estimation of the scale parameter for loglogistic distribution using balanced ranked set sampling. *Journal of Statistics*, 53.

Sonderby, S. & Winther, O. (2014). *Non-parametric survival analysis in breast cancer using clinical and genomic markers*.

Su, P.F., Chi, Y., Li, C.I., Shyr, Y. & Liao, Y.D. (2011). Analyzing survival curves at a fixed point in time for paired and clustered right-censored data. *Computational Statistics & Data Analysis*, 55(4), pp. pp. 1617–1628.

Surendran, S. & Tota-Maharaj, K. (2015). Log logistic distribution to model water demand data. *Procedia Engineering*, 119, pp. pp. 798–802.

Tuan, H.W., Lin, S.C. & Julian, P. (2017). *Improvement for amelioration inventory model with weibull distribution. Mathematical Problems in Engineering*,

2017.

Vahdaninia, M. & Montazeri, A. (2004). Breast cancer in iran: a survival analysis. *Asian pacific journal of cancer prevention*, 5(2), pp. pp. 223–225.

Xiao, X., Hu, Q., Yu, D. & Xie, M. (2014). Study of an imputation algorithm for the analysis of interval-censored data. *Journal of Statistical Computation and Simulation*, 84(3), pp. pp. 477–490.

Xiong, C. & Ji, M. (2004). Analysis of grouped and censored data from step-stress life test. *IEEE Transactions on Reliability*, 53(1), pp. pp. 22–28.

Yang, B., Xiao, C. & Zhao, X. (2010). Interval-censored data analysis of gaps between recurrent events with cure fraction. *In: Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on, IEEE*, pp. pp. 1–4.

Yip, C., Bhoo, P.N. & Teo, S. (2014). A review of breast cancer research in malaysia. *Medical Journal of Malaysia*, 69, pp. pp. 8–12.

Yip, C.H., Taib, N. & Mohamed, I. (2006). Epidemiology of breast cancer in malaysia. *Asian Pacific Journal of Cancer Prevention*, 7(3), p. pp. 369.

Zhang, D. (2008). Effects of model selection on the coverage probability of confidence intervals in binary-response logistic regression. *University of Maryland, College Park*.

Zhao, G. (2008). Nonparametric and parametric survival analysis of censored data with possible violation of method assumptions. *ProQuest*.

Zhao, M., Zhao, Y. & McKeague, I.W. (2015). Empirical likelihood inference for the odds ratio of two survival functions under right censoring. *Statistics & Probability Letters*, 107, pp. pp. 304–312.

Zyoud, A., Elfaki, F. & Hrairi, M. (2016). Non-parametric estimate based in imputations tech-niques for interval and partly interval censored data. *Science International (Lahore)*, 28(2), pp. pp. 879-884.