THE APPLICATION OF HAZARD FUNCTION IN THE CASE STUDY OF THE USE OF AN IUD FOR WOMEN

NOR AFIAH BINTI NOR HANAFI

UNIVERSITI TEKNOLOGI MALAYSIA

THE APPLICATION OF HAZARD FUNCTION IN THE CASE STUDY OF THE USE OF AN IUD FOR WOMAN

NOR AFIAH BINTI NOR HANAFI

A report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science Mathematics

Faculty of Science
Universiti Teknologi Malaysia

NOVEMBER 2009

To my beloved mother and father

ACKNOWLEDGEMENTS

First and foremost, all praise be to Allah S.W.T., the Almighty, the Benevolent for His blessing and guidance for giving me the inspiration to complete this dissertation.

A lot of thanks to Dr. Zarina Bt Mohd Khalid for her guidance and support for me during the whole semester to complete this dissertation.

Furthermore, I would also like to dedicate my appreciation to my fellow friends for their motivation throughout this semester. Without their endless support, this dissertation will not complete as presented here.

Last but not least, I also extend my appreciation to all my colleagues and others who have provided assistance in various occasions. To me, their help in giving views and tips had provided me with useful aid in completing this project.

ABSTRACT

This dissertation presents the application of hazard function in the case study of the use of an IUD for women. The data in this case study is the secondary data that had been retrieved from Modeling Survival Data in Medical Research Second Edition by Collet, 1952. The Kernel nonparametric method and parametric Weibull hazard estimation methods, such as Maximum Likelihood and Least Square method in estimating hazard function are being compared with Kaplan-Meier estimate of hazard function. These comparisons are done through graphical comparison and analysis of standard error. The analysis of these results is done by using Mathcad and Minitab.

ABSTRAK

Disertasi ini memaparkan aplikasi fungsi risiko di dalam kajian kes terhadap pengunaan IUD di kalangan wanita. Data di dalam kajian kes ini merupakan data darjah kedua yang diambil dari *Modelling Survival Data in Medical Research Second Edition* oleh Collet, 1952. Teknik tidak berparameter Kernel dan teknik berparameter Weibull seperti teknik kebarangkalian maksimum dan segi empat sama terkecil dalam menganggar fungsi risiko telah dibandingkan dengan teknik menganngarkan fungsi risiko Kaplan-Meier. Perbandingan ini dilakukan melalui perbandingan graf dan analisis ralat am. Analisis ini dijalankan menerusi penggunaan Mathcad dan Minitab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

b	-	Bandwidth
c	-	Censoring time
d_j	-	Death at time *j*th death time
$F_T(\,.\,)$	-	Cumulative density function
$f_T(\,.\,)$	-	Probability density function
$h_T(t)$	-	Failure rate at time *t*
$H(\,.)$	-	Integrated hazard function
$h(t), \lambda(t)$	-	Hazard function at time *t*
$\hat{h}(t),\ h^*(t),\ \hat{\lambda}(t)$	-	Estimated hazard function
$K(x)$	-	Kernel function of variable *x*
k	-	Shape parameter of Weibull hazard function
$L(\theta)$	-	Likelihood function of random variable θ
L_n	-	Modified empirical distribution of *L*
n	-	Number of observation
n_j	-	Observation at risk at *j*th death time
$S(t)$	-	Survival function at time *t*
$\hat{S}(t)$	-	Estimated survival function
$se(\theta)$	-	Standard error of θ
$t_{(j)}$	-	*j*th death time
$W(\lambda,k)$	-	Weibull hazard distribution
λ	-	Scale parameter of Weibull hazard function
$\hat{\sigma}^2$	-	Estimate of variance

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the Problem

The beginning of survival analysis may be traced back to early work on mortality in the seventeenth century as Graunt published the first Weekly Bill of Mortality in London and Healey published the first lifetable. Ever since then, the lifetable method has been used regularly by actuaries, statisticians, and biomedical researchers in governmental and private agencies. During World War II, reliability of military equipment became a significant issue. This directed to the study of the durability or the "lifetime" of industrial devices.

After the war, the techniques used to analyze the reliability of industrial devices were further expanded and pertained to the study of survival time of cancer patients. The phrase "lifetime analysis" used by industrial reliability engineers was altered to "survival analysis" by cancer researchers. During four decades ago, survival analysis has turn into one of the most regularly used techniques for analyzing data in disciplines ranging from medicine, epidemiology, and environmental health, to criminology, marketing, and astronomy.

The primary variable in survival analysis is survival time, no longer limited to mean the time to death. The term “survival time” is used loosely for the time period from a starting time point to the occurrence of a certain event. Examples of survival time are: the time to the development of diabetic retinopathy from the time of diagnosis of diabetes, the time to parole of a prisoner, the duration of first marriage, workman’s compensation or other insurance claims and lifetime of electronic devices and computer components.

In summarizing survival data, there are two functions of central interest. The two functions are the survival function and the hazard functions. There are many ways to estimate these functions. For this dissertation, the focused is on estimating the hazard function instead of the survival function. The hazard function will be applied in the case study of the use of an IUD for woman. The Kernel and parametric Weibull hazard functions estimates, which are Maximum Likelihood and Least Square methods will be compared with Kaplan-Meier estimate in order to find the most suitable approach that can be used to estimate the hazard function.

1.2 Statement of the Problem

There are two types of nonparametric estimation of hazard function that will be discussed in this dissertation such as Kaplan-Meier estimate and Kernel Method estimate. While for the parametric estimate of Weibull hazard function, Maximum Likelihood and Least Square Method will be conclude. From these two approaches, there will be a most suitable approach to estimate the hazard function of a finite data compared with others. The problem arises when we want to determine the suitable approach that can be used to estimate the hazard function. In order to achieve a solution

to this, the comparison of all these approaches need to be done and once again we will have some problem on how to estimate the hazard function using the Kaplan-Meier, Kernel method, Maximum Likelihood Estimate and Least Square Method, how to find the standard error of the estimated hazard function and how to find the most appropriate approach among these two approaches to estimate hazard function.

1.3 Objectives of the Study

The objectives of this study are:

1.3.1 To study the basic concept of Kaplan-Meier, Kernel method, Maximum Likelihood Estimate and Least Square Method in order to estimate the hazard function of the time to discontinuation of the use of an IUD for 18 woman.

1.3.2 To compare the nonparametric and parametric estimates using Kaplan-Meier, Kernel method, Maximum Likelihood Estimate and Least Square Method in order to find their advantages and disadvantages.

1.3.3 To determine the most suitable approach that can be used to estimate the hazard function through graphical comparison and the analysis of error.

1.4 Scope of the Study

This study will discuss the application of two types of approaches of hazard function which are nonparametric and parametric approaches in the case study of the use of an IUD for woman.

For nonparametric approaches, Kaplan-Meier estimate and kernel method estimate will be discussed and for parametric approaches, Maximum Likelihood Estimation and Least Square Method will be discussed.

In addition, this study will also determine the most precise approach to estimate the hazard function.

1.5 Significance of the Study

The outcomes of this study will give advantages to statistical and medical fields. In statistics, this study will offer advance knowledge on the use of suitable approaches in analyzing non-negative random variable for instance time to certain event. While in medical, this research will advanced the knowledge to medical practitioners on the risk of a person at certain times after being given some kind of treatment so that they can improve their treatment to cure the patients.

1.6 Organization of the Report

This report includes five chapters which are Introduction, Survival Analysis, The Estimation of Hazard Function Procedures, Results and Discussion and Conclusions and Suggestions.

The introduction of the study is stated in Chapter 1 of the report. This chapter contains six subtopics such as background of the problem, statement of the problem, objectives of the study, scope of the study, significance of the study and organization of the report.

Chapter 2 contains the discussion about survival analysis. In this chapter, survival analysis is being introduced by the failure time and the hazard function. After that, the following subtopics, which are the nonparametric hazard function estimation and parametric Weibull hazard function estimation, are discussed.

Next, Chapter 3 explains about the estimation of hazard function procedures. This chapter stated five interesting subtopics which are introduction, the medical data, the nonparametric estimation of the hazard function and the parametric estimation of Weibull hazard function.

Then, Chapter 4 includes the results and discussions of the report. The main things that being stated in this chapters are introduction, the nonparametric estimation of hazard function for the time to discontinuation of the use of an IUD for 18 women, the parametric estimate of Weibull hazard function for the data of time to discontinuation of

the use of an IUD for 18 Women and the graphical comparison of the nonparametric and parametric approaches in estimating the hazard function

.

Lastly, Chapter 5 includes the conclusions and suggestions. The Conclusions and suggestions are brought upon based on findings as preparation for future research.

REFERENCES

Agnew, D.C., Constable, C. (2004). *Least Square Estimation.*

Collet, D. (1952). *"Modeling Survival Data in Medical Research"*. Boca Rotan, FL: Chapman & Hall.

Collet D (1994). *"Modeling Survival Data in Medical Research".* Chapman & Hall.

Commenges, D., Huber, C. Nikulin, M.S., (2003). *"Probability,Statistics and Modeling In Public Health".* New York, NY: Springer Science + Business Media.

Cook, A. (2008). *Survival and Hazard Functions*. Introduction to Survival Analysis, National University of Singapore.

Cox, D.R., Hinkley, D.V., Reid, N., Snell, E.J. (1991). *"Statistical Theory and Modeling: In Honour of Sir David Cox, FRS. Edited By D.V. Hinkley, N. Reid and E.J. Snell".* London, New York : Chapman & Hall.

Cox, D. R., Oakes, D. (1984). "*Analysis of Survuval Data*". London, New York, Tokyo, Melbourne, Madras: Chapman & Hall.

Franklin, C. H. (2003). *Maximum Likelihood Estimation.* Duration Models: Exponential and Weibull Likelihoods, University of Wisconsin-Madison.

Grimshaw, S. D., McDonald, J., McQueen, G. R., Thorley, S. (2003). *Estimating Hazard Functions for Discrete Lifetimes.* Radical Eye Software.

Horová, I., Pospísil, Z., Zelinka, J. (2009). Hazard Function for Cancer Patients and Cancer Cell Dynamics. *Journal of Theoretical Biology.* 258, 437-443.

Horová, I., Zelinka, J. (2006). *Kernel Estimates of Hazard Functions for Biomedical Data Sets*.

Huang, B. (2005). *Nonparametric Estimation of Hazard Function from Censored Data by Kernel Method.*

Ives, M., Funk, R., Dennis, M. (2000). *LI Analysis Training Series.* Survival Analysis / Life Tables, Chesnut Health System.

Kim, C., Bae, W., Park, B.U. (2005). Nonparametric Hazard Function Estimation Using the Kaplan-Meier Estimate. *Nonparametric Statistics, Taylor and Francis Group.* 17(8), 937 - 948.

Lee, E. T., Go, O. T. (1997). Survival Analysis in Public Health. *Annual Review Public Health.* 18, 105-134.

Müller, H. G., Wang, J. L. (2007). *Density and Failure Rate Estimation with Application to Reliability.* Encyclopedia of Statistics in Quality and Reliability.

Nochai, T., Bodhisuwan, W. (2000). Statistical Reliability Analysis of Sometypes of Two-Parameter Life Time Distributions. *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications.* June, 13-15. Universiti Sains Malaysia, Penang.

Pérez, G. E., Cimadevila, H. L., Río, A. Q. D. (2002). Nonparametric Analysis of the Time Structure of Seismicity in a Geographic Region. *Annals of Geophysics.* Vol. 45(3/4).

Razali, A. M., Salih, A. A., Mahdi, A. A. (2009). Estimation Accuracy of Weibull Distribution Parameters. *Journal of Applied Sciences Research, INSInet Publication.* 5(7), 790-795.

Rodríguez, G. (2005). *Nonparametric Estimation in Survival Models*. Princeton.

Subramanian, S., Bean, D. (2008). Hazard Function Estimation from Homogeneous Right Censored Data with Missing Censoring Indicators. *Statistical Methodology, ELSEVIER, Science Direct.* 5, 515-527.

Tutz, G., Pritscher, L. (1996). Nonparametric Estimation of Discrete Hazard Functions. *Lifetime Data Analysis, Kluwer Academic Publishers.* 2, 291-308.

Vaal, V. A., Koshkin, G. M. (1999). Kernel Nonparametric Estimation of the Hazard Rate Function and Its Derivatives. *KORUS'99, Mathematics, IEEE Xplore.* 496-500.

Wang, J. L. (2003). Smoothing Hazard Rates. *Encyclopedia of Biostatistics.*

Wang, Q. H. (2008). Some Bounds for the Error of an Estimator of the Hazard Function With Censored Data. *Statistics and Probability Letters.* 44, 319-326.

Xie, Z., Yan, J. (2008). *Kernel Density Estimation of Traffic Accidents in a Network Space.* Geography / Geology Faculty Publications, Western Kentucky University.