



Sudan University of Science and Technology



College of Graduate Studies

Research title:

**Survival Analysis for Epidemic Diseases of Children under Five Years
in Dr. Jafar Ibn Auf Pediatric Specialized Hospital 2012–2016**

تحليل البقاء على قيد الحياة للأمراض الوبائية للأطفال دون سن الخامسة
بمستشفى الدكتور جعفر بن عوف التخصصي للأطفال 2012 – 2016

A Dissertation

Submitted to Department of Statistics – College of Science Sudan University of
Science and Technology for Requirements the Degree of Ph.D. in Statistics

Prepared By:

Khalid Mustafa Shoaib Mustafa

Supervisor:

Dr. Ahmed Mohammed Abdallah Hamdi

Co-supervisor:

Dr. Altaiyb Omer Ahmed Mohammed

Oct 2018

الآية

قال تعالى:

(إِنَّا نَحْنُ نُحْيِي الْمَوْتَى وَنَكْتُبُ مَا قَدَّمُوا وَآثَارَهُمْ وَكُلَّ شَيْءٍ أَحْصَيْنَاهُ فِي إِمَامٍ مُّبِينٍ)

صدق الله لعظيم

سورة يس الآية (12)

DEDICATION

TO my parents

TO my wife and twin

TO my family

&

TO my teachers and friends

TO all Children and parents with love and respect

Acknowledgement

I would like to express my special appreciation and thanks to my Supervisor *Dr. Ahmed Mohammed Abdallah Hamdei* and Co-Supervisor *Dr. Altaiyb Omer Ahmed Mohammed*, both of you have been a tremendous mentor for me. I would like to thank you for encouraging my research and allowing me to grow as a research scientist. I would also like to thank the dissertation committee members, *Dr. Dr. Ahmed Mohammed Abdallah Hamdei* as chairperson, *Prof: Amin Ibrahim Mohammedas* external examiner; *Prof: Afra Hashim Abdullatif Mohammed* as internal examiner, for serving as my committee members even at hardship. I also want to thank you for letting my defence be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you. I therefore, would also immensely grateful thanks to Children parents, physicians, nurses, and nurse aids in the paderaita units at Sudan University of Science and Technology as well Khartoum University, addition to that the department of epidemiology at *Jafar Ibn Auf Pediatric Hospital* in Khartoum, whom have constructing the human model.

At the end, I would like to express special thanks to my parent. Words cannot express how grateful for all the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. I would also like to thank all my friends who supported me in writing and invited me to strive towards my goal.

Abstract

This study aims to use the survival methods in large censored data, sample of 1098 Sudanese children under five years from both categories, affected with different five diseases are acute renal failure, congenital deformity heart, leukemia, septicemia and sickle cell disease) at *Jafar Ibn Auf Pediatric Hospital in Khartoum/ Sudan*, from 2012-2016. The purpose is to estimate the accurate probability of survival time for children in existence heavy censored data. The two methods used are Modified Weighted Kaplan-Meier (MWKM) and Accelerated Failure Time (AFT) model using SPSS, STATA, NCSS and XLstat. The main hypotheses tested are that no differences between probability survival and hazard rates, Cox proportional hazard model and accelerated failure time model in estimating the probability survival time. The results obtained 235 (21%) of children died during the study with median's survival time of 16 day per disease and 863 (79%) are censored till the study end with survival rate (0.97). Modified Weighted Kaplan Meier estimator gives accurate survival probability of 100% to the last censored child, and when the survival of Child (j) is p_j equal 0.1 to 0.9, the estimator gives the accurate probability survival time to the last censored children from 0.2167 to 0.9002, respectively. Moreover, AFT model is better than Proportional Hazard model in estimating the large censored data. This is proved through goodness-of-fit test (AIC&BIC), which obtained Weibull AFT model has fitted better, more valuable and realistic predicted than the survival and hazard functions than Proportional Hazard model. The study recommended using MWKM and AFT model in estimating the probability survival time for such dataset.

المستخلص

تهدف هذه الدراسة إلى استخدام طرق البقاء في رقابة البيانات الضخمة لعينة مكونة من 1098 طفلاً سودانياً دون سن الخامسة من الجنسين، مصابين بخمسة أمراض مختلفة هي الفشل الكلوي الحاد، تشوه القلب الخلقي، اللوكيميا، سرطان الدم وفقر الدم المنجلي بمستشفى جعفر بن عوف للأطفال / الخرطوم، السودان من 2012-2016م. الهدف هو تقدير الإحتمال الدقيق لوقت البقاء على قيد الحياة للأطفال في ظل رقابة البيانات الضخمة. الطريقتان المستخدمتان هما كابلان-ماير المعدل والمرجح و نموذج وقت الفشل المتسارع باستخدام SPSS ، STATA ، NCSS ، XLstat. الفرضيات الرئيسية التي تم إختبارها أنه لا توجد فروق بين معدلات البقاء على قيد الحياة وخطر الوفاة، وبين نموذج المخاطر النسبية لكوكس ونموذج وقت الفشل المتسارع. النتائج التي توصلت إليها الدراسة هي 235 (21%) حالة وفاة الأطفال اثناء الدراسة بوسيط زمن البقاء حتى الوفاة 16 يوماً لكل مرض و 863 (79%) من الأطفال تمت مراقبتهم حتى نهاية الدراسة بمعدل بقاء على قيد الحياة 0.97؛ كابلان-ماير المعدل والمرجح قدر إحتتمالية البقاء على قيد الحياة بدقة 100% لآخر طفل يخضع للرقابة، فإذا كان بقاء الطفل z هو الزمن t بإحتمال (0.1 الى 0.9) فإن كابلان-ماير المعدل والمرجح يعطى احتمال البقاء على قيد الحياة لآخر طفل يخضع للرقابة بدقة 0.2167 إلى 0.9002، وايضاً نموذج وقت الفشل المتسارع أفضل من نموذج المخاطرة النسبية في تقدير البيانات الكبيرة الخاضعة للرقابة، و من خلال إختبار مدى ملاءمتها خلصنا بأن نموذج وقت الفشل المتسارع وبيبل هو الأكثر قيمة وواقعية في تقدير دالتي البقاء والمخاطرة. توصي الدراسة بإستخدام كابلان-ماير المعدل والمرجح ونموذج وقت الفشل المتسارع لتقدير إحتتمال وقت البقاء على قيد الحياة لمثل هذه البيانات.

TABLE OF CONTENTS

Subject	Page
Quranic verse	I
Dedication	II
Acknowledgment	III
Abstract in English	IV
Abstract in Arabic	V
Table of Contents	VI
List of Tables	IX
List of Figures	XIII

TABLE OF CONTENTS

N	Chapter One: Introduction	Page
1.0	Preface	1
1.1	Research Problem	2
1.2	Research Objectives	3
1.3	Research Importance	4
1.4	Research Hypothesis	5
1.5	Research Materials and Methods	5
1.6	Previous studies	6
1.6.1	Current study vs. the previous studies	15
1.7	Organization of the Dissertation	16
Chapter Two: Literature Review		
2.0	Preface	17
2.1	Survival Analysis	18

2.1.1	Survival time	20
2.1.2	Survival function	20
2.1.3	Lifetime's distribution function	25
2.1.5	Hazard function and cumulative hazard function	26
2.1.6	Comparison of survival distributions	26
2.1.7	Censoring data types	30
2.2	Nonparametric, Semiparametric and Parametric Approaches	35
2.2.1	Nonparametric approach	36
2.2.2	Semiparametric Approach	42
2.2.3	Parametric Approach	54
2.2.3.1	Parameters and statistics	55
2.2.3.2	Parametric proportional hazards model	57
2.2.3.3	Accelerated failure time model	61
2.2.3.4	Estimation of AFT models	65
Chapter Three: Epidemic Diseases for Children under fifth		
3.0	Preface	76
3.1	Health of Sudanese Children	76
3.2	Epidemic Diseases of Sudanese Children under 5 Years	80
3.2.1	Research and global health	82
3.2.2	Sample of non communicable Children diseases	83
3.3	Sudanese Children Mortality	89
3.4	Jafar Ibn Auf Paediatric Hospital	90
Chapter Four: Analysis of Data		
4.1	Methodology	91
4.1.1	Description of Enhanced Selective Acknowledgements	91

4.2	Statistical Analysis	94
4.3	Survival Analysis	95
Chapter Five: Result and Recommendation		
5.1	Result	157
5.1.1	Finding	158
5.2	Recommendation	160
	References	162
	Appendix	170

LIST of TABLES

N	Title of Table	Page
(2.1)	Observed data and corresponding rank	38
(2.2)	Mean for Two or more equal ranks	38
(2.3)	Summary of parametric AFT models	64
(2.4)	Comparison of Cox PH model and AFT model	74
(3.1)	Mortality rates by age group in Sudan in 2014	78
(3.2)	Early Childhood Mortality by Regional in 2014	78
(3.3)	Early Childhood Mortality per Sudanese States in 2014	78
(3.4)	Trend in under5 and neonatal mortality among Sudanese Children 1990-2015	82
(3.5)	Causes of Children's Acute Renal Failure in Developing Countries	85
(3.6)	Leukaemia in Sudanese Children under Five Years in Oct 2017	87
(3.7)	Distribution of signs of sepsis of the study population	88
(3.8)	Relation between the blood culture and Creative protein	88
(3.9)	Infant mortality (deaths/1,000 live births), 2016-2017 estimation	90
(4.1)	Quartile Statistics	92
(4.2)	Diseases Rank	92
(4.3)	Rank test of group variables using Kruskal Wallis	92
(4.4a)	Median Test Frequency	93
(4.4b)	Test Statistics for disease type	93
(4.4c)	Test of variables median	93
(4.5)	% of Survival Time of Children under 5 years	98
(4.6)	ROC analysis of sensitivity and specificity for the disease	100

(4.7a)	Area under the curve (AUC)	101
(4.7b)	AUC null hypothesis test	102
(4.8)	Skewness/Kurtosis tests for Normality	105
(4.9)	Survival rate estimation and 95% C.I by K.M and W.K.M	106
(4.10)	Description of censored types and death	109
(4.11)	Diseases type and affected children details	110
(4.12)	Diseases type and affected children details per gender	110
(4.13)	Odds Status (case control studies)	111
(4.14)	Median survival time test of enumerating sample-space	111
(4.15)	Declaration of survival data	113
(4.16)	Median days of survival time for Children under Five Yrs.	113
(4.17a)	Exposed statusdeath and unexposed censored	114
(4.17b)	Exposed gender female and unexposed male	114
(4.17c)	Exposed disease with and without symptoms	114
(4.18)	Children times at risk	115
(4.19)	Number of events per disease	115
(4.20a)	Summary statistics of Acute Renal Failure	115
(4.20b)	Survival probability estimation by W.K.M vs. M.W.K.M in ARF	115
(4.20c)	Survival probability estimation by M.W.K.M in ARF	117
(4.21a)	Summary statistics of Congenital Deformity Heart (CDH)	119
(4.21b)	Survival probability estimation by W.K.M vs. M.W.K.Min CDH	119
(4.21c)	Survival probability estimation by M.W.K.M in CDH	121
(4.22a)	Summary statistics of Leukaemia	122

(4.22b)	Survival probability estimation of W.K.M vs. M.W.K.M for Leukaemia	122
(4.22c)	Survival probability estimation of M.W.K.M for CDH	124
(4.23a)	Summary statistics of Septicaemia	126
(4.23b)	Survival probability estimation of W.K.M vs. M.W.K.M for Septicemia	126
(4.23c)	Survival probability estimation of M.W.K.M for Septicaemia	128
(4.24a)	Survival probability estimation of W.K.M vs. M.W.K.M for SCD	131
(4.24b)	Survival probability estimation of M.W.K.M for SCD	131
(4.24c)	Survival probability estimation of M.W.K.M for Sickle Cell Disease	136
(4.25)	MWKM survival probability of last 2 Children for 5 diseases	141
(4.26)	Test of equality of the survival distribution functions (DF = 4)	141
(4.27)	MWKM result for the last 10 censored Children ($p_j = 0.5$)	142
(4.28a)	Variable Summary Report and Break per Gender= Male	143
(4.28b)	Variable Summary Report, Break per Gender= Female	143
(4.29)	Kruskal-Wallis equality of populations rank test for survival time	145
(4.30)	Events observed and expected per diseases	146
(4.31)	K.M estimate the survival function (events) vs. types of diseases	147
(4.32)	Gender qualitative variable	147
(4.33)	Weibull Accelerated Failure Time Model	153
(4.34)	Akaike's information criterion and Bayesian information criterion	153
(4.35)	Reliability of Parametric Distributions per disease	154

LIST of FIGURES

N	Title of Figures	Page
2.1	Characteristics of the Survival	25
2.2	Survival curve of disease since last hospital visit	46
2.3	Survival probability at mean of covariates	49
2.4	Deviance and Martingale residuals against the risk score for Cox PH model	51
3.1	Sudanese Children under five mortality rates for May 2011.2016	77
3.2a	Number of Sudanese birth from 2015 to 2017	77
3.2b	Sudanese Child mortality per 1000 live birth	79
4.1	Receiver Operating Characteristic Sensitivity/Specificity	97
4.2	The negative specificity of the survival time	98
4.3	The sensitivity of the diseases type	99
4.4	The probabilities of the Survival time	100
4.5	Sensitivity and specificity of survival time	101
4.6	Life tables for the Survival data	102
4.7	Survival Distribution Function by K-M	103
4.8	Survival Distribution Function by Adjusted K-M	103
4.9a	Cumulative hazard function versus survival time	104
4.9b	log cumulative hazard function versus log survival time	104
4.10	Survival distribution function	105
4.11a	Survival distribution estimation used K-M	107
4.11b	Survival distribution estimation used W.K.M	108
4.12a	Survival probability used K-M, WKM and MWKM	108
4.12b	Survival probability used MWKM if $P_j =$ Survival rate value or 0.5	108

4.12c	Survival probability used K-M and WKM	111
4.13	Distribution of symptom severity in 5 diseases sample	112
4.14	Survival probability estimation for the last 10 censored Children	142
4.15	Survival probability for 5 diseases used MWKM	142
4.16	Kaplan-Meier survival curve	146
4.17a	The probability of sample size power	147
4.17b	Weibull comparison survival curve for 5 diseases	148
4.17c	Weighted Weibull comparison survival curve for 5 diseases	149
4.18	lognormal survival curve (a) and weighted lognormal survival curve (b) comparison of 5 diseases	150
4.19	Survivor time and Children<5 at risk per disease	151
4.20	Comparison of AFT models for time of diseases progression vs. percentage of the targeted children	152