DISTRIBUTION AND DETERMINANTS OF TIME-TO-DORMANCY OF MEDICAL RECORDS ATTHE UNIVERSITY COLLEGE HOSPITAL, IBADAN NIGERIA, 1990-2014

BY

BENSON MACAULAY OWEGHORO
(Matric No: 102296)
BHIM, PGDS (Ibadan), MSc Statistics (Ibadan)

A THESIS SUBMITTED

TO

DEPARTMENT OF EPIDEMIOLOGY AND MEDICAL STATISTICS
FACULTY OF PUBLI HEALTH
IN PARTIAL FUFILMENT OF THE REQUIREMENTS FOR THE DEGREEE OF THE

DOCTOR OF PHILOSOPHY
OF THE
UNIVERSITY OF IBADAN

December 2019

CERTIFICATION

I certified that this work was carried out by Mr. Benson Macaulay OWEGHORO in the Department of Epidemiology and Statistics, Faculty of Public Health, University of Ibadan, Ibadan.

……..….

Supervisor
Professor Sola Ayeni
B.Sc. M.Sc. (London) PhD (London)

……..

Co-Supervisor
Dr. Onoja Matthew Akpa
B.Sc., M.Sc. PhD (Ilorin)

DEDICATION

This work is dedicated to God and to my grandchildren, I love you all.

ACKNOWLEDGEMENTS

To God be the glory for his mercy and grace from the start to the end of this research work, may his name be praised for ever.

This has been a long journey and wouldn't have ended successfully without key individuals. My thanks goes to my mentor and Supervisor, Professor Olusola Ayeni, who "found me and led me through from nothing to something". I salutehis patience in reading through all the copious amounts of material I wrote including the content that never finallymade it into this thesis and being part of this long journey the last 6 years supervising the work. He showed the stuff it takes to be the first Medical Statistician in African. I cannot thank him enough for his encouragement, suggestions and contributions towards the success of this work. Your fatherly words of advice and your gentle approach to giving corrections provided me the tonic to keep going. To you I say thank you.

I cannot but recount the role of my Co-Supervisor, Dr. O. M. Akpa, whose office became my research secretariat; he was so tolerant and always ready to attend to my quarries despite his tight schedules. Your words of encouragement was in most time gave the push. I say thank you.

I thank the head of Department, Epidemiology and Medical Statistics, Dr. M. D. Dairo for providing the enabling environment at the crucial stage of the research work, your concern for my success was appreciated. My thanks also goes to Dr. Oyindamola Yusuf, the former head of Department, Epidemiology and Medical Statistics, I appreciate your contributions, the Almighty God will continue to bless you

I cannot forget the guiding role by Dr. J. O. Akinyemi, your suggestions were very useful. Thank you Sir. I also thank Dr. Tunde Adedokun for his encouragement and contributions to the work. I appreciate all the Lecturers in the Department of Epidemiology and Medical Statistics, Dr. A. S. Adebowale, Dr. R. F. Afolabi, and Dr. E. A. Bamigboye, I thank you all. I thank the non-academic staff for their support.

I appreciate my senior colleagues in the Department of Library, Archival and Information Studies. I appreciate Professor Iyabo Motolagbe Mabawonku for her encouragement and mentoring. Dr. A. A. Abioye, a wonderful friend and confidant, I thank you for your advice and encouragements, words cannot be enough to express my appreciation. I also appreciate Professor K. I. N. Nwalo, Professor S. O. Popoola, Professor G. O. Alegbeleye, Professor O. A. Okwilagbe, Dr. Airen Adetimehin, Dr. D. A. Akangbe, Dr. J. K. Akpotiade, Mr. O. I. Igudia, and Mr. O. O. Folorunso. The Almighty God will bless you all.

I wish to specially thank my friends, but for their encouragement and supports this opportunity would never had been. I am grateful to Dr. Adedayo Adepoju of the

Department of Statistics, University of Ibadan, who introduced me to medical statistics, encouraged and guided me, you will never lack God's favour. To Mrs. Elizabeth Tunrayo Lere Oluwole, I will ever be grateful, you were there at the beginning, stayed and stood by me till the end, your encouragement and assistance proved you are a true friend indeed, thank you. I appreciate all my friends too numerous to mention, my students in the Department of Library, Archival and Information Studies, University of Ibadan and all those who had contributed in one way or the other to this success story, God bless you.

I thanked the University of Ibadan, Ibadan for giving me the opportunity to pursue the programme. I also appreciate the support and encouragement of Mr. Ibrahim Mohammed Mami, Registrar, Health Records Officers Registration of Nigeria. I thank the Head of Department and Staff of the Health Records Management, University College Hospital, Ibadan for support during the collection of data for the research. I thank Mr. Wole Oyedele and Mr. Olaniyi Mathew Olutola for their roles at analysing the data for the study, thank you for your wonderful contributions.

Finally I want to appreciate my wife and the children for their encouragement and supports during the period of the research, you are a wonderful family. I also acknowledge my daughter-in-law, I appreciate you. God bless you all.

ABSTRACT

Medical Records are important in patient care, follow-up and clinical research. However medical records often become dormant due to cessation of patient-healthcare provider interaction. Retention of dormant record isinefficient, ineffective, wastes time and resources for storage and may hinder retrieval of active medical records. Knowledge of time-to-dormancy of these records is important to formulate retention and disposal policies for medical records management. However, there is paucity of information on time-to-dormancy of medical records in Nigeria. This study therefore was conducted to determine the statistical distribution, estimates of time-to-dormancy and predictors of record dormancy at the University College Hospital, Ibadan, Nigeria.

A review of medical records from 1990-2014 was conducted at University College Hospital, Ibadan. From 478,300 available records within the study period, systematic sampling technique was used to select 7,685 records. Information on patient's characteristics (date of first and last visits, gender, age, clinic attended, and other clinical and treatment-outcomes) were extracted from each record using a data extraction proforma. The outcome variablewas time-to-dormancymeasured as the period from creation of a record to the point at which the record becomesdormant. Data analyses were done using descriptive statistics and Kaplan-Meier Method. Estimatedhazard rates of dormancy were plotted against time, log of cumulative hazards*[log-log(S(t)]*were plotted on log of time (*log(t))*to determine the statistical distribution and its shape parameter wasestimated. Parametric hazard model was used to identify determinants of time-to-dormancy. Performance of model of choice was compared to a semi-parametric Cox Proportional Hazard(CPH)model. Log likelihood (*-2logL)*and Akaike Information Criterion (AIC) were used to evaluate CPH and Weibull models that best fitted time-to-dormancy data, while statistical significance was set at $\alpha = 0.05$.

Patientsage 31-60 years were 40.3%, male constituted 52.4%, and 55.4%resided in Oyo State. Hospital admission rate was 30.0%, while 98.8% patients were alive at the time of last entry. Records with ≥2 entries attained dormancy in 151.9 months (95% CI=128.7-

179.1). Hazard plots of time-to-dormancy exhibited a bathtub shape,*[log-log(S(t)]* on *log(t)* plots indicated a linear relationship,with estimated shape parameter of 0.6, suggesting Weibull distribution. Values of-*2logL* forCPH (11061.4) and Weibull (4371.9);and AIC for CPH (11075.4) and Weibull (4389.9). Weibull model indicated that being female (HR=1.1,CI=1.0-1.2); admitted-patient (HR=1.2,CI=1.0-1.4); attendance at Surgical Out-patient (HR=1.1,CI=0.9-1.3); discharged against medical advice (DAMA) (HR=9.0,CI=2.1-36.1) and death (HR=3.6,CI=0.5-25.9), were associated with dormancy. Similarly, CPH regression model indicated that female (HR=1.1,CI=1.0-1.3); admitted-patient (HR=1.2,CI=1.0-1.4); attendance at Surgical Out-patient (HR=1.0,CI=0.9-1.3); DAMA(HR=17.9,CI=4.3-74.9) and death (HR=3.1,CI=0.4-22.4), equally influenced dormancy.

Weibull model provided the best fit suggesting a minimum retention period of 151.9 months for medical records.Records of females, admitted-patients,those who attended surgical out-patient, patients discharged against medical advice and deadpatientsare more likely to become dormant earlier. A medical records retention policy should be formulated based on the estimated time-to-dormancy.

Keywords: Medical records, Dormancy,Records management, Records retention, Time-to-dormancy.

Word count: 467

TABLE OF CONTENT

List of Tables

List of Figures

LIST OF ABBREVIATIONS

DAMA	Discharge Against Medical Advice
RC	Reference category
MOP	Medical Out Patient
SOP	Surgical Out Patient
GYNAE	Gynaecology Clinic
CHOP	Children Out Patient
AHIMA	American Health Information Management Association
MST	Median Survival Time
MDT	Median Dormancy Time
TTD	Time-To-Dormancy
PH	Proportional Hazard
CPH	Cox Proportional Hazard
HR	Hazard Ratio
CI	Confidence Interval

CHAPTER ONE
INTRODUCTION

1.1 Background to the Study

Proper records management is an important function of every successful organization, healthcare organizations inclusive. Healthcare facilities specialise in providing patient care and not records management, however providing patient care is information-based. Volumes of information are created at every instance patients are seen and these become records that must be managed effectively and efficiently. Information is the life blood of patient care and the indicator on which the quality of patient care is measured. Records are documented account of activities and regardless of the format or medium in which they are held,it serves as a corporate memory and are required for legal or statutory compliance.

The art of keeping patient records is said to be as old as medicine itself (Huffman 2014) and arguably has been in existence since the evolution of medicine. In recent years medical records management practice has become more clearly defined and more widely recognised. According to Wissmann (2015), medical records management at its core represents all the activities associated with the collection and management of health information, in all settings across the healthcare spectrum, in relation to all recipients of healthcare and for multiple purposes to support the healthcare ecosystem. Quality medical records are critical to the provision of healthcare. Decisions about diagnoses, treatment, medications, preventive health, and all aspects of healthcare depend on accurate information being available at the right time to the right healthcare provider about the right patient or consumer of healthcare. A good medical records management system could mean the difference between life and death for some patients. It is the backbone of patient care and considered one of the important elements in patient care.

The term medical record refers to both the physical folder that exists for each patient and the body of information found therein. However, authors used the terms interchangeably Roach*et al.*(2006), others prefer to create a distinction between the physical folder and the information found therein (Skurka 1998;Galani and Nikiforou 2006; McWay 2008; Katuu 2015). For the purposes of this study, medical records is the preferred term and is defined as a patient record, containing information that may be described any

documentary material or information, oral or recorded in any form, that is created or received by a health care provider, and relates to the past, present or future, physical, social or mental health of an individual, or the past, present or future provision of health care to an individual. It includes all the documents health providers create or receive in the course of their encounter and transactions with the patient. These records are maintained by a group of professionals known as Medical Records Officers who keep the records for current and future use till the records may become uneconomical to be retained any further. In carrying out this function the Medical Records Officer plan, collect, aggregate, analyse, and disseminate individual patient and aggregate clinical data, making them experts in managing health data and processes in the healthcare information system.Medical records are important legal documents for both the healthcare provider and the patient.

Good record management practice involveshaving an organised approach to record-keeping, being able to locate and retrieve records when required; and keeping what is needed only for as long as is required,(University of Strathclyde, 2012),. In the opinion of the Rinchart-Thompson (2008), health organisations must be committed to ensuring that complete and accurate medical records are managed and disposed of in accordance with established records management policies.

The processes whereby records are created, managed, stored and disposed of are technically referred to as records management and described as the systematic administration of records through its entire life cycle, from creation, use, retention, to final disposition. A core and universally acceptable concept referred to as 'Records Life Cycle' theory, invented by Schellenberg and later developed by Penn, states that records management are in four phases which are creation,active, semi-active, and the inactive phase,(Penn, *et al* 1994, Shepherd and Yeo 2003). The concept of creation-to-disposition is analogous to biological birth-to-death. Records are therefore likened to organisms which are born, live and die (inactive) at an age, thereby obeying the "records life cycle" theory. This theory has become a basic and important concept in records management. According to Aduku and Abdul (2012) and Records Management Bulletin (2012), Records Life Cycle is based on the idea that records become less important as time passes and that 90% of active use of a record takes place during the first 90 days after it is

created. This short period of high use is followed by a longer period of low use where the records only need to be looked up occasionally. Eventually, even this limited use will end and the records will become inactive and have no further value to their creator. This assumption had been shown to be true for all records including patient medical records. Inactive records are regarded as dormant and should be disposed of.Dormancy is the state of the record becoming inactive with no further entries inserted. Such records can be safely weeded from the filing system creating space for new ones.

In records life cycle, organizations have to define in policy statement(s), how long a record is to be kept in each phase and how records are disposed of and archived. Tools, systems, and procedures are developed to manage each phase of the life cycle. In the view of Hoke (2011), all records are dynamic–never static. Even records as long-lived as a sequoia have a date of creation, a use/purpose, and a date of disposition or archiving. The times of a record's creation and disposition or archiving are the limits of its life cycle**.** Because records life cycle is not defined by national boundaries, record's management policies help to look at records' progressive stages. Laws, regulations, and customs cannot change the fact that records need management from creation to disposition. The policies in different stages of the life cycle may vary, but not the scope of governance. Not all records created deserve to be kept permanently or even for longer period, as significant costs usually associated with the creation, maintenance, distribution, and storage of records can be reduced with proper records management. According to Sullivan and Wyatt (2009), information exists only to support decisions and actions and if it fails to do this, it becomes irrelevant noise.

The life cycle is the starting point for creating a records management policy programme, regulations and guidelines. Without a record's management policies, patient records management would not be cost effective and the retention of inactive/dormant records in the filing system would be counterproductive.

At point of creating a record, organisations must consider how long a record should be retained and how would such a record be disposed at the end of its life. According to Hoke, (2011) the creation of a record and the disposition are at each end of the life cycle, with disposition as the point where information finally loses relevance and is removed from the current information governance programme.

Disposition may pose a serious risk factor because the thought of retaining everything is bad records management practice. Although storage may be considered cheap, costs of administration cumulated over time could be very high, and much higher when you are in litigation, emergency or there is a regulatory inquiry and you have to produce a record buried in a mountains of records. The cost in terms of human resources and time needed to search through can be much higher.

Experience had shown that most patients' records are active for a relatively short period of time and can be disposed of, while others may need to be preserved for longer period, and still some permanently archived for historical purpose. The active period for which patient's record needs be retained will need to be determined, depending on the time-to-dormancy (TTD) of such record which is the duration or length of time between first contact when a record is created and last contact with the patient's record, estimated as *"date of last contact – date of first contact;"* this is the survival time. Information need not be kept after it is no longer required otherwise valuable resources and unnecessary cost may be wasted or incurred. Retaining records in any form, paper or electronic, for an extended period of time has cost implications for the hospital organisation. Therefore, it is not only good practice but also important for hospitals that patients records are only kept for as long as they are required. It is a statutory requirement under medical records retention policyin some countries that patients' records be kept for as long as is required for the purpose for which it was created. Retaining such records indefinitely 'just in case' could then amount to institutions breach of statutory requirements. The challenge therefore is for a hospital to manage her patients' records, making sure that those records with active information are preserved till the end of the survival time, while dormant records are disposed of in accordance to records management policies.

According to the American Health Information Management Association (2008)in determining how long patients' records are retained, hospitals should consider applicable laws and regulations, administrative policies and medical practice. Though minimum standards are set by the statute of limitations, each institution must be guided in retention policy formulation by the institutional peculiarities such as the nature, purpose and use of the medical records, filing space, manpower recourses and other patients' characteristics.Hospitals may therefore develop policy(s) on records management, either

as a stand-alone policy or as an integrated part of a broader national suite of information or knowledge management policies. These policies will provide guidelines on how records are managed through their life cycle. Over the years archivists have researched into records management strategies; however no known effort had been made to estimate time-to-dormancy of records. A major problem facing medical records practice in Nigeria is the non-availability of a strategic policy on patients' records management. The implication is that the process of creation, maintenance, retention, disposal and archiving of patients' records are not standardised. Yet managing patients' records is intrinsic to the health information management (HIM) practice because it comes with a number of challenges bordering on policies and guidelines on retention period, mode of disposal, and archiving.

The level of medical record management in the University College Hospital, Ibadan and other Nigerian hospitals, as in other developing economies is still below the global acceptable standard. Preliminary survey of medical record departments revealed a common sight of patients' records filed on broken wooden shelves or records lying on the floor with both active and inactive records put together in confusion. Agbaje (1991) observed that the rate at which records' of patients seen in the University College Hospital, Ibadan was growing was creating problems of storage, retrieval and security. Aduge-Ani (2003), also reported "a crisis of confidence between patients and medical records personnel in the General hospital, Wuse, Abuja leading to patients taking their medical records home for safe keeping as a result of challenges of missing records. This situation would had been averted if a policy on patients' records management was put in place. According to Records Management University of Washington (2014), an important step in the maintenance of a successful filing system is the identifying and managing inactive records. The Ministry of Health, NSW Australia (2012) directed that health organizations must ensure high standards for management of patients' records are maintained consistent with policies on current best practice requirements. Hospitals are under moral, ethical and legal obligations to maintain and manage the records of patients so that patient information are timely, accurate, complete, accessible, cost-effective, and useable for patient care. Accurate and complete information, at the right time, makes a

better healthcare delivery and these can be made possible only when there are policies, on patient records management.

Tavakoli et al (2007) in one study on the retention and destruction process of medical records found that hospital management are confused about the required time for the retention of medical records, leading to lack of space due to long retention of inactive records while some are forced to destruct records prematurely. In another study, Ebadifar (2004) concluded that there is the lack of regular and united approach in Iran's hospitals on the important tasks of medical records retention and disposal.

The inactive phases of records life cycle, cannot be efficiently and effectively managed without the establishment of a policy that specified time-to-dormancy on the records. According to Howell, Jr. and Cogar (2003),there must be a well-defined method for managing records – retaining what is needed and eliminating what is not;with a standardized methodology developed for creating records retention programmes, with each retention policy created to the specifications of the individual hospitals. No organisation can keep all her records for ever no matter how important such records. Therefore it is important for efficient and effective patients' records management that a critical study to estimate the time-to-dormancy of records be carried out in each hospital.

1.2Statement of the Problem

One of the most significant challenges in health care is the ability to effectively manage patient information. Preliminary studies show that there are no documented retention or policy guidelines on medicalrecords management in Nigeria. Medical records in Nigerian hospitals are crisis managed due to the absence of policy guidelines on retention and when to dispose dormant records, (Agbaje 1991, Adgbe-Ani 2013; Oweghoro, 2015).The outcomes of this is the retention of inactive i.e. dormant records in the filing system longer than necessary and with a negative consequences on medical records management and hence poor patient care.

The patient'smedical record,a legal document, may be based in various storage mediumthoughtraditionally paper-based. With the advent of digitalisation some healthcare providers created hybrid patients' records, a medical record that is partly paper-based and partly electronic. As technology improved, some healthcare institutions

are moving to an entirely electronic health record system, however the developing countries including Nigeria had been slow in catching on this development, (Jamoom, *et al*, 2014; Shortlffe, 2016). Regardless of the medium, paper or electronic, in which the patient information resides, the concept of records management holds. Expectedly as health information management moves from the paper-based to electronic system, the complexity and need to develop institutional and national policy on patients' records management, and in particular retention, disposal and archiving policies become more important. These polices usually establish procedures, rules and regulations that set out frameworks to ensure that the creation, retention period and disposal and archiving of patients' medical records held within hospitals are managed in accordance with established policyguidelines. Typically space is too expensive for the storage of patients' records referred to infrequently. Universally, space for filing patient medical record is a major constrain in the hospital, hence records that have passed their active life (active period being an estimate of how long the records will be required "in-department" for patient care) are periodically purged from the filing system and relocated to inactive secondary storage before final disposal and archiving. Disposal is the term used to cover the final action taken on inactive or dormant patient medical records and this may range from preservation on storage media, archiving, recyclingto destruction. The disposal should be determined by an empiricallyappraised process of the retention period and time-to-dormancy.

Evidences abound that records cannot be retained forever hence there is need for policies on retention and disposal management, both of which are functions of time-to-dormancy. The destruction of records is an irreversible act but the physical space required make permanent retention of all records created by hospitals an impractical option. It is therefore mandatory for hospitals that time-to-dormancy and characteristics of dormancy for patient medical records areestimated empirically to ensure records that are required for medical, research or legal purposes are not inadvertently disposed or destroyed and at the same time dormant records are not retained beyond their economic values. Patients' health information containedin the medical records only exists to support clinical decisions and actions and if it fails to do this, it is irrelevant noise and should be disposed of.

In the University College Hospital, Ibadan patients' records grow at an astonishing rate, preliminary investigation showed that an average of 53.14 new records are created daily, a total of 1594.33 per month and 19,130 records a year. These records also grow in volume proportional to patient revisit rates. A major challenge is how to manage these volumes of patient medical records created daily. Personal observations by the researcher in the medical records department of the hospital revealed that the notion of records management do not go beyond the phases of creation and uses of the record life cycle, whileretention, disposaland archivingof inactive records are practically neglected. This neglect could be as a result of lack of policy guidelineson retention period,resulting in inefficient and poor management of inactive patient medical records.As a result, patient records are crowded into few available filing cabinets with accompanying on-the-floor filing with the resultant misfiling and mislaying of records and inactive/dormant records retainedin the filing system indefinitely. The outcome of this is longer retrieval time of medical records, longer patient waiting time, inadequate information for patient care management, inadequate research materials and non-availability of patient medical records when needed. Above all poor patient care management and waste in human and material resources.

The gap created by the absence of policies on medical records management can only be bridged with the knowledge of time-to-dormancy and if time for retention, disposal and archiving of patient records are specified. Only then can patient medical records management which is fundamental to quality healthcare services be strengthened. Toestimate time-to-dormancy for the formulation of policies on patient records management and to specify retention, disposaland archiving periods, require the knowledge of statistical distribution and their parameters of time-to-dormancy of medical records.Literature showed that records cannot be retained forever but failed to quantify time-to-dormancy.No known empirical studies had however been done to determine the time–to-dormancy of patient medical record or factors that may contribute to patient record dormancy in the University College Hospital, Ibadan in particular or in other Nigerian hospitals.

It is therefore presumed that determining the general characteristics of the statistical distribution, their parameters of the survival function,the hazard functions and factors that

increase the risk of dormancy of patient's medical record will fill this gap. The attending knowledge can then be used to promote the formulation of policies for medical records management policy guidelines for the retention, disposal and archiving of patient medical record, with the resultant best practices in patient records management in the University College Hospital and also serve as a guide to other health institutions in Nigeria.

1.3 Aims of the study

To determine the general characteristics of the statistical distribution of time-to-dormancy and the parameters of the survival function ofpatients records created between 1st January 1990 and 31st December 2014, at the UCH, Ibadan, and articulatetheir implications on medical recordsmanagement and archiving at the hospital.

1.4 Objectives of the Study:

i. determine the general statistical distribution and its survival functions of time-to-dormancy of medical records of patients seen at the UCH, Ibadan between1990 and 2014;

ii. determine the form and shape of the hazard rate of the medical records of patients in order to identify the appropriate statistical model(s) for the analysis of time-to-dormancy of medical records of patients seen in UCH between 1990-2014;

iii. estimate the percentiles and their *SEs*of the distribution of time-to-dormancy of medical records of patients seen at the UCH between 1990-2014;

iv. examine if the distribution and parameters are the same for medical records of patients seen at different periods of time between 1990 and 2014;

v. determine demographic and clinical factors associated with time-to-dormancy of medical records of patients seen in UCH between 1990-2014;

vi. highlight guidelines based on findings from the study that would enable the drafting of a policy on medical records management in UCH, Ibadan

1.5 Research Questions

The following research questions based on the research objectives will be guiding this study:

i. What are the characteristics of the distribution of time-to-dormancy of records of patients seen at the UCH, Ibadan, from 1990 and 2014?
ii. What are the survival functions of time-to-dormancy of records seen in UCH Ibadan, from 1990-2014?
iii. What is the form and shape of the hazard rate of records seen in UCH Ibadan, from 1990-2014?
iv. What is/are the suitable statistical model(s) that best fit time-to-dormancydata of medical records seen in UCH Ibadan, from 1990-2014?
v. What are the percentiles and their *SE* of cumulative distribution of TTD?
vi. Are the distributions and its parameters same for records of patients seen at different periods of time between 1990 and 2014?
vii. What are the demographic and clinical factors associated with the length/distribution of time-to-dormancy of records seen at the UCH Ibadan from 1990-2014?

1.6 Justification for the Study

In other parts of the world, there are institutional and nationalpolicies and regulations that set a time limit on the number of years records are to be retained based on statute of limitation and institutional policies. These policies,regulations and laws are developedand served as a guide for managing patient medical records. Retention period for medical records will vary from country to country and also with institutions. The practice is for a country to develop a retention,disposal and archiving policy for patient medical records. Each hospital can then take a cue from the national policy to develop its own medical records management policy.Benchmark for retention, disposal and archiving period should not be set arbitrarily, rather there must be an attempt to determine the most suitable time frame such that valuable information are not destroyed. Any health institution that has no policy and guidelines forpatient medical records management run a great risk of low quality patient care, in addition to violating the statute of limitation of the state.

Patient records management is about controlling records within a framework of policies, standard operating procedures, systems, processes and behaviors. Together they

ensure that reliable evidence of actions and decisions are kept and remained available for reference and used when needed, and that the organisation benefits from effective management of one of its key assets, its records. Patients' records contain vial information that could affect the survival of a patient hence the need to develop a record management policy that is based on empirical studies.

The assumption is that the value of any information (patient information inclusive) is determined by the use of such information over time. Where information is not used over time, it is assumed to be dormant. However the point of dormancy for a record should be quantified. A serious deficiency in the Record Life Cycle model is the failure to quantify the time between when a record is createdandwhen it becomes inactive, i.e. for how long does a medical record remain active before it becomes dormant and declaredfit for disposal. This study will address this important question which is an important parameter required for developing records management policy guidelines. However, this can only be done if the distribution and parameters of time-to-dormancy of patient medical records are known. In addition, there is the need to find out the individual and joint contributions of factors that could contribute to time-to-dormancy of the patient medical records.

1.7 Definition of operational terms

Age at Registration: Patient's age at registration was classified into:

<10	children
10-20	adolescent
21-30	youth
31-60	adult
61+	older Adult

Age at dormancy: the time (in months/years) from the creation of a record (indicated by the first entry)to the point where the record is declared inactive (indicated by last entry) and can be safely weeded from the filing system. This is the survival time of a record

Dormancy: Dormancy is the state of the record becoming inactive with no further entries inserted.

Hazard rate: the instantaneous rate of failure in a process or the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval.

Inactive records:a record that is no longer referenced on a regular basis and therefore needs to be stored in a less accessible place since they are not used frequently having reached their cut-off state as defined on a Records Retention Schedule;

Information governance (IG): the management of information to support an organization's present and future, keeping in mind the regulatory, legal, environmental, and operational requirements. Synonymous with records management.

Medical Records Management: All activities and processes involved in the planning, creation, organisation, use and dissemination, maintenance, disposition and evaluation of patients records in a health care facility.

Medical Record: The term is used for both the physical folder that exists for each individual patient and the body of information found therein.

Patients Medical Information: Any documentary material or information, oral or recorded in any form, that is created or received by a health care provider, and relates to the past, present, or future; physical, social or mental health of an individual, or the past, present or future provision of health care to an individual. It includes all the documents health providers create or receive in the course of their encounter and transactions with the patient.

Patient Records: this is synonymous with medical records and will be used as such in this work.

Penultimate appointment: The last but one appointment given to a patient after which the patient is discharged from all forms treatment for a condition

Record Archiving: Removing inactive/dormant medical record to a remote storage place.

Records Disposal: The process by which inactive/dormant records are either archived for secondary storage, transformed into another storage media, or destroyed; or the point where information finally loses relevance and is irretrievably removed from the current information governance programme or the disposition phase in records management, when records are assessed to determine their retention value using general disposal schedules or records disposal schedules leading to either the preservation or destruction of such record.

Records life Cycle: The concept in records management that records go through the phases of creation, active, inactive and final disposition.

Records Retention Period: The length of time over which patient records are kept for use having been regarded as still active and of value, defined as the time-to-dormancy of the records.

Retention Schedules: A retention guideline that indicates the shortest amount of time records are required to be retained

Survival time*:* is the time to the occurrence of a given event which can be the development of a disease, response to a treatment, relapse, deathor dormancy of records

Time-to-dormancy: The period from creation of a record (indicated by first entry) to the point at which the record attain inactivity/dormancy (indicated by date of last entry), or the point where information finally loses relevance. This is the period for which the record should be retained, (Same as dormancy time).

CHAPTER TWO
LITERATURE REVIEW

2.1 Literature for this study will be reviewed under the following subheadings:

- **2.2 An overview of patient medical records management;**
- **2.3 Policies and guidelines for patient medical records, retention, disposal and archiving;**
- **2.4 Concepts of survival analysis;**
- **2.5 Analysing Time-To-Event data,**
- **2.6 Theoretical framework for the study,**
- **2.7 Conceptual model, expected results and conclusions**

2.2 An overview of patient medical records management

Records contain information that are valuable resources to the delivery of high-quality evidence based patients care and many other key health service deliverables, and they have more values when it is accurate, up to date and accessible when it is needed.

According to a document accredited to the National Hospitals Office (NHO), 2007, an effective records management service ensures that information is properly managed, is available whenever and wherever there is a justified need for that information, and in whatever medium it is required and which is compliant with the relevant legislation",

In the hospital patient medical records are essential tools in the management of patient care, litigations, medical and epidemiological research and health care planning and administrations. According to Department of Health, Social Services and Public Safety (2004), an effective records management system ensures that information is properly managed and made available whenever and wherever there is a justified need for that information to:

- Support patient/client care and continuity of care;
- Support service provision;
- Support day-to-day business which underpins the delivery of care;
- Support evidence-based clinical practice;
- Support sound administrative and managerial decision making, as part of the knowledge base for Health and Social Care services;
- Meet legal requirements, including requests from patients/clients under subject access provisions of the DPA 1998 or the Freedom of Information (FOI) Act 2000;
- Assist clinical/professional and other types of audits;
- Support improvements in clinical/professional and service effectiveness through research and also to support archival functions by taking account of the historical importance of material and the needs of future research; or
- Support choice and control of patients and clients over treatment and services.

These multiple functions and users of medical records identified over the years as a result of development in hospital records management, brought about by dynamism in medical practice, must have resulted in the various names, such as medical records, hospital chart, outpatient record, clinical record, health record, patient hospital record, electronic health record, electronic medical record, and such descriptors for the basic records. These terms are used for both the physical folder that exists for each individual patient and for the body of information found therein, (Dana andMcWay 2010). The patient medical record is generally defined as a document that contains a complete and accurate description of a patient's history, condition, diagnostic and therapeutic treatment, and the results of treatment. It should include detailed personal, medical, financial, and social data about the patient.

In a hospital set-up the patient medical record contains evidence of activities by the care provider resulting from the interaction with the patient and these are often referred to as the patient's health information. The value of any information is in the content, context and structure rather than their physical format. Records are a valuable resource; they form what is commonly referred to as the "corporate memory"of an organization. Because of the information they contain, records are evidence of activities undertaken hence it is an institution's best ally in terms of protecting her rights and interests. High-quality information underpins the delivery of high-quality evidence-based health and social care, and many other key service deliverables in the hospital.

Every organisation including the hospital must meet the requirements of its regulatory environment and it is therefore important that they put in place record management programmes to control the quality and quantity of information created and received. The ISO 15489: (2001) standard defines records management as the field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records. In essence, records management is the management of information throughout the information life. Records and information management can therefore be described as the efficient and systematic control of all records from their creation or receipt, through their processing, distribution, organization, storage, and retrieval to their ultimate disposition at a point when they are no more useful for the purpose for which they were created. Records Management is a logical and organised approach to the creation, maintenance, use and disposition of records which ensures records can be easily retrieved when required and disposed of in accordance with policies, guidelines, laws and contracts.According to Akussah (1996), it is globally accepted among archivist and records management professionals that the records life cycle concept is the best approach to records management. And this probably explains the used as a base for developing frameworks for managing records' (Ngulube and Tafor 2006).

Historically, Theodore Schellenberg invented the records life-cycle concept while working in the National Archives of the USA in the 1930s (Shepherd and Yeo 2003:5). According to the life-cycle records management framework, records pass through four

basic conceptual stages during their life. Though different scholars had presented these stages differently, Charman (1984), Hardcastle (1989), Hare and McLeod (1997), and Penn, *et al* (1994) all have in common the view that records pass through creation phase, active phase to a semi-active and then to a non-active stage. The four phases of the life-cycle appears distinct from each other with the temptation to estimate time frames for each phase, this however contrary records continuum concept which seeall the four phases of records management as interrelated forming a continuum, (Atherton 1985;McKemmish, S., 1997; McKemmish, *et al* 2005; Society of American Archivists, 2016). Both concept however agreed that record passes through creation, active, semi-active phases until they eventually 'die',

Healthcare professionals appreciate the value of keeping accurate and detailed records for each patient in the hospital as a moral expectation, professional ethics and requirement by law. It is therefore a good practice for every healthcare organization to have in place a records management policy, guidelines and up-to-date legislative requirements on records standards and management. This is particularly important with respect to patient's medical records creation, maintenance, retention, dormancy and disposition. The development of such policies and standards ensure good quality and efficient patient medical records management which is an essential ingredient supporting high quality of patients care. Good records management is a precondition for continuity of patient care and can reduce the risk of adverse incidents through misplace for untraceable records. According to the Medical Protection Society, South Africa, (2012) adequate medical records that is properly managed provide physicians and other care providers information to document the essential parts of each patient contact without reference to memory. The medical records should therefore be comprehensive enough to allow a colleague to carry on where you left off. Poor-quality medical records are not only a major cause of iatrogenic injuries, they also make difficult to defend a clinical negligence claim or a disciplinary inquiry; it is axiomatic that poor note-keeping is evidence of poor clinical practice". Effective records management service ensures information are properly managed, made available when and where needed and in compliance with the relevant legislative policies", (NHO 2007, Sullivan and Wyatt, 2009, and University of Strathclyde, 2012)

Good patients' records management is one element of information governance, that can be described as a set of multi-disciplinary structures, policies, procedures, processes and controls implemented to manage information at an enterprise level, supporting an health institution's past, immediate and future activities, legal, risk, environmental and environmental and operational requirements.According to University of Strathclyde, (2012), records management best practice requires:

- an organised approach;
- that records are located and retrieved when required;
- provide evidence of activities, decisions and actions;
- that you keep what you need only for as long as is required; and
- ensure long-term preservation of records of archival value.

This practice can only be achieved through the establishment of institutional and national management policy guidelines.

2.3 Policy guidelines for records retention, disposal and archiving

Good records management starts with a policy and guidelines which reflects an organization's needs. A records management policy can be described as an authoritative statement of intent to manage records in an appropriate and suitable manner for as long as they are required for business purposes (The National Archives, 2012). It is intended to form the initial framework or principles which express how records should be managed within the organisation. The objective of the records management policy should be the creation and management of authentic, reliable, complete and usable records which are capable of supporting business functions and activities of the organization for as long as they are required. According to Archives and Records Management Association International (ARMA), **(**2016) business and government create enormous quantities of records each business day and to control the growth of these records, an organisation needs policies to help maintain and dispose of records that are no longer needed.

Records retention policies specify the length of time business records are to be retained. The retention policy is based on the concept that information has a life cycle, which is the time period from the creation of a record to its final disposition. And that record documents an organisation's business operations and are essential to effectively

managing that business. Organisations define what constitutes a business record as this will make operational recordkeeping decisions easier. Patients' records are created for a variety of reasons, including complying with government regulatory or statutory reporting requirements, documenting daily business activities, documenting research and development methods for possible patent applications, as well preserving the legal rights of the care business. For whatever reason a record is created, there is a useful active life of that record … a period of time when the record is important for business decisions. Policies and standards are vital items in any form of management without which, it is difficult to evaluate the effectiveness of any process being undertaken. Policies and standards are benchmarks and guidelines used to check on the quality of work being undertaken. According to Aduku and Abdul, (2012) "the fundamental concept behind records management is that each record has a life cycle and this is based on the ideal that most records become less activewith ageand that 90% of the active life of any record takes place during the first 90 days of creation." When the information contained in a record no longer has any immediate value, the record should be removed from active accessibility and depending on the nature of the record; it is either retained, transferred, archived or destroyed. All records regardless of storage media type (hardcopy or electronic) are dynamic and never static, they have a date of creation and disposition, when they become dormant. (Iron Mountain 2005, Hoke 2011). Whether a record is in paper or digital format does not determine its value or retention period; its content is the key factor and which records to keep and for how long will also vary from organization to organization. Each organisation will be guided by operational policies and regulations.

Literature has shown that most patients' records are useful for a relatively short period and can be destroyed, while others need to be preserved for years; and still some permanently. Hospitals therefore need policies to help maintain and dispose of patient medical records that are no longer active and had become dormant. Evidence abound in literature that records need not be kept indefinitely; and that there should be a policy guidelines to guide and regulate the retention, disposal and archiving of records. According to Howell and Cogar (2003) and Arruda, et al (2003) a good retention policy typically has two principal elements; a schedule identifying the retention periods (minimum and maximum) for all documents covered by the policy," anda "framework

for the administration of the policy…"Madu, (2004) observed that "records life-cycle management comprises of policies, processes, practices, services and tools used to align the value of information with the most appropriate and cost effective infrastructure from the time arecord is created through to its final disposition."In formulating records management policies and guidelines organisations should take cognizance ofrelated legal retention periods, consideration of national regulatory requirements, contractual obligations, intellectual property requirements and statutes of limitations. These various legal requirements must then be harmonised with organisation's operational considerations, which may extend the retention periods.

Rockefeller Archive Centre, (2008) had explained that in the United States both federal and state laws stipulates varying minimum retention periods for different documents created in differentorganisations. This could be as a result of varying degree of importance and uses. While records of accident reports and claims (settled cases), accounts receivable and payable ledgers and schedules are retained for 7 years, correspondencewith customers and vendors, and administrative records are kept for only 3 years. Understanding how records are managed therefore requires understanding the legal context in which such records can and should be created, managed, retained and disposed of or archived. The implication of this is that various organisations would need to use legislative policies and regulations to ensure that records are retained and disposed of or archivedat appropriate time. According to Chibambo (2003), it is not enough for an organisation to have a records good management policies framework that consists of information-related laws, policies and standards of practices, the necessary qualified human resources to implement and manage the systems must be in place. Supporting this ascension, Iron Mountain (2005), in their document titled best practices initiative, stated that regardless of media type (hardcopy or electronic) record retention periods are based on legal, regulatory, and operational requirements and that the development of a legally credible records retention schedule is broken down into four activities:

- Identify major record groups
- Create a universal classification scheme
- Perform legal research
- Overlay operational retention requirements.

According to literature setting retention periods had to be guided by laws, institutional practice, regulations relevant to practice settings, benefits and risks associated with retention among others operational requirements. (Iron Mountain 2005, Sturm 2012).
As part of records management policies, records retention scheduleshould support an organization's effort to manage and control the costs of information storage, locate and retrieve documents for legitimate use, and dispose of or archived records at the end of their life cycle. Instituting formal and legally credible records retention policies enables an organisation to meet both operational needs and the legal requirements of mandated retention periods.

It is therefore obvious that the foundation upon which any records management policy is developed is the "records life cycle theory" based on the assumption that records become less important as time passes. The short period of high use, followed by a longer period of low use when the records only need to be looked up occasionally. Eventually, even this limited use will end and the records will have no further value to the organisation in respective of the nature of its business. However for a records management policyto be effective and efficient, time frames should be estimated for the life cycle based on each organisations institutional practice.

2.3.1 Policies and guidelines for retention, disposal and archiving law practice records

Non-profit organisations, like for-profit ones, be it legal, financial, business or medical, may retain certain records created beyond current use needs, this may be according to regulatory, legal, financial, or operational requirements.In order to ensure that legal records are well managed, many governments and organisations including law organisations, are implementing records management plans. These plans ensure records are efficiently and adequately managedto meet legal and administrative requirements of the organisation or government. Law practice are document and information intensive and the advent of digital technology has increased the volume and complexity of records created, making it impossible to be adequately managed on an ad hoc basis, (professional counsel guide for lawyers and law firms, 2007).This had created a challenge for law firms to develop and follow a records management policy and procedures for managing records

of their practice. Like in other information intensive organisations lawyers and law firms require a comprehensive set of records management policies and procedures to address the entire lifecycle of records created. The guide further stated thatwhile all phases of the records life cycle are important, perhaps more attention should be focusedmore on managing the inactive records. Records management best practices suggest records management policy plansshould govern every stage of their lifecycle including, file creation, data privacy as properly managedrecords facilitate responses to client inquiries about the progress of matters, other issues, often allowing disputes to be resolved faster.

Organisations may vary in their goals andoperations but managing records retention and disposal approaches tend to have something in common in the way they are created managed, retained and disposed of or archived. Some records may have their retention period short, others may be long depending on their functionalities. In the business of lawHowell, Jr. and Cogar (2003)stated records are often the vehicle by which compliance is established therefore it is impossible for an organization to achieve acceptable legal compliance without an appropriate and functioning records retention policy.

A good and legally compliant records retention schedule, a disposal policy, and archiving plan would provide the foundation of a good records management programme. This is the platform for thorough protection from risk and litigation. A records retention schedule is a document that an organisation uses to ensure that records are kept only as long as legally and operationally required, and that inactive records are disposed of in a systematic and controlled manner.According toAmerican Records Managers Association (ARMA) International, (2016) an organization should retain herrecords for a specified time, considering operational, legal, regulatory and fiscal requirements, and those of all relevant binding authorities. At the heart of effective records management is the determination of the period of retention driven by legal mandate but driven the operational needs of the corporation. Analysing United States legal requirements relating to records retention in Law and Records Management,Skupsky (n.d.) found that there are laws which require records to be maintained, but do not specify a retention period and there are those situations where no requirements are found. Noting that there are basically four types of legal requirements for records retention generally encountered:

i. *Specific Requirement State* - Many federal and state requirements will indicate a specific retention period for records;

ii. *Limitations of Action:* Limitations of action are not records retention requirements; instead, they represent the period during which an organization may be involved in a legal action or litigation (either as plaintiff or defendant). Records may be useful during this period to pursue a legal course of action or to defend oneself. The appropriate records retention period for legal purposes, therefore, relates more to litigation strategy rather than to actual legal requirements.

iii. *No Retention Period Stated*. A large number of statutes and regulations contain phrases such as "the following records shall be maintained . . . " Although under this type of provision, records must be maintained, the organization is not provided sufficient information to determine how long the record must be maintained

iv. *No Records Maintenance or Retention Requirements Found After Research*.

Records managers often encounter statutes and regulations which state that certain records must be maintained, but fail to provide a specific retention period. This type of provision is very typical; in fact, most developing countries statutes and regulations do not state specific retention periods. In Nigeria the National Health Act,(2014) stipulates that healthcare providers will maintain health records for each patient but for how long such records are to be kept and how they are disposed were not indicated. The solution out of this uncertainty is each organisation to determine time-to-dormancy for records created and formulate policies and regulations for records management policies. This becomes justified in view of Rockefeller Archive Center, (2008) argument that most records managers seem to have difficulty in determining the legal requirements for records:

(1) when the law requires the maintenance of the record but does not state a specific retention period; or

(2) when no legal retention requirements have been identified related to a specific record, especially after extensive research.

Based upon the federal Paperwork Reduction Act as interpreted and implemented by regulations published by the U.S. Office of Management and Budget, there appears to be a presumption that no records required under federal regulations (not statutes) need be retained longer than three years unless the federal agency involved has stated (and justified) a longer retention period.

It is important to conduct legal research to determine what the retention period for each record class must be. This work often requires the assistance of legal counsel, consultants or external records management experts. At a minimum, these types of legal requirements must be considered.Federal, State, Local, and International(if relevant) and in addition to legal requirements, operational retention requirements must also be taken into account. This is the length of time that a record must be retained to meet departmental, operational or user group record needs. The final retention period should be the longer of the two, (Iron Mountain 2005).

An organization's records management programme should be supported by policies and procedures that address each component of the records management programme in accordance with operational and legal requirements. According to Iron Mountain (2005) though law practice records are classified into two basic categories as firm records and client records both require effective record retention and disposal policies. Not all information or data produced or used by lawyers is an actual record,while records need to be maintained and retained beyond the termination of a representation, non-records need not be. Record retention and disposal policies establish set periods for the initial retention of various identified classes of records andestablish separate retention schedules for firm records and client matter records, as it is customary to separatesuch records early in the retention process as the information they contain are of varying importance to the firm. According to the Professional counsel guide for lawyers and law firms, (2007), Lawyers like in any other system, choosing to keep everything forever is neither practical nor appropriate; generally, records should be kept long enough to preserve evidence in the event it is needed in defense of a professional liability claim.

In setting retention periods, lawyers therefore need to be realistic about the time and costs associated with implementing and maintaining multiple record retention periods. In most cases the costs of reviewing a closed client matter file several times due

to the existence of differing retention periods for various types of records far outweigh the costs associated with retaining some records longer than actually needed.

2.3.3 Policies guidelines for retention, disposal and archiving of financial records

Organisation must maintain book and records of accounting activities performed, ranging from audited financial report, a review, a tax return, or a specific management report, all these had to be done to summarize and analyse facts and figures to support reports, tax returns and conclusions. The important question then is for how long these records must be retained, (Federal Taxation Committee,2004). Organisations make retention decisions based on the content and purpose of records and retention periods are mostly determined by these requirements. In Nigeria, legislation requires that financial records be kept for an indefinite period and some for specific periods, (Financial Control and Management Act (1958) Revised Financial Regulations, 2000). The term "indefinite" is not defined in this legislation, but clearly requires that documents be retained for as long as the relevant entity exists. It is of note that once an entity ceases to exist, the obligation on that entity to retain documents "indefinitely" also ceases to exist. A record retention and disposal policies shouldindicate how long a record should be stored before it is destroyed or archived and in addition specify who takes responsibility.According to the International Records Management Trust, (2002) the Nigerian Financial Control and Management Actprovide for the retention periods for financial records, as yet, there are no standards and practices to control the retention and disposal of these records, though theprimary responsibility rests with the Accountant General and the Auditor General,there is no-one in either department who 'champions' record. In South Africa, according to the Institute of Chartered Accountants, (2013) "the general requirement, (as required by the Companies Act and other legislation), is that a company keep information and to retain such information for a period of at least seven years ora longer period than specified in the applicable legislation.

According to World Bank (2000), the establishment of effective records management system provide a cost effective deterrent to fraud and serve as an important tool in combating corruption. Corroborating this submission, International Records Management Trust, (2002) and Igbokwe-Ibeto, (2013) submitted that proper management

of financial records in the public sector is fundamental to the management of resources and the elimination of corruption., that an effective records management system is fundamental and crucial in combating corruption, explaining that in Nigeria, it is a known fact that corrupt officials often arrange for records to disappear to avoid prosecution but where there are good records management policies, loss of records would be prevented. Dearstyne (1985) and Shepherd (2006:10) in their opinion that appropriate records management programme will help organisations to conduct business in an efficient, accountable manner, deliver services consistently, support managerial decision making and transparent policy formation and ensure continuity in policy execution, management and administration.

2.3.3 Policy guidelines for retention, disposal and archiving of university records

Universities ashigher educational establishments act asgenerators and repositories of knowledge, and both these role are information driven. University record is any form of record created either in paper or digital format that provides evidence of the decisions and actions of the University while undertaking its business; thatmay take the forms teaching and learning, research, community service, organisational, commercial or cultural activities, (Griffith University, 2018). University are service delivery organisations therefore records created should be efficiently managed. The purpose of a university's records management policies is to provide a mechanism for retention and disposal of records created in accordance with its legal business obligations either as a private and apublicly-funded university. These policies guide users to those records that should be retained and for how long and to enable the universities to legally dispose of records that are no more needed. Iwhiwhu(2005) reveals that records management policy on records are not available in Nigerian universities; hence records are managed without recourse to the principles of records management. The absence of Records Manual, retention and disposal guidelines, trained personnel to man records sections, lack of facilities for storage, and retrieval of records, no filing manual, inadequate computers to manage the volume of records generated and the poor attitude of management towards records and records management constitute the problems of records management in Nigerian universities. Corroborating Iwhiwhu

2005),Ifedili and Agbaire (2011), found out that the general consensus in Nigerian universities was that record-keeping practice was below average, with records sections been manned by unqualified and unskilled personnel. This is also in line with the findings of Akor and Udensi, (2013) that though record keeping occupies a strategic position in the efficient and effective management of the university system, findings showed that records management is not receiving the attention it deserves at IBB University Lapai and Federal University of Technology, Minna. Abdulrahaman, (2015) had observed that university records are not properly managed because staff engaged in records management units in the universities in North Central Nigeria are not adequate in number and training, which is in line with the findings of Nworgu, (2005) andAbioye, (2006), that though records management is a specialised field, many organisation employees learn on the job.

In the University of Waterloo records are properly managed to ensures that records are available for University administration for as long as they are needed to meet statutory, regulatory, policy, contractual, and operational requirements, and are disposed of appropriately when they have reached the end of their retention period, (University of Waterloo, 2016)

University records may be classified into students, teaching, researches and management records. The University of Massachusetts (2009) records retention and disposition matrix span from few months to 6 years, and from accident reports to annual financial statements that are kept permanently. Like any other organisations universities develop their record retention policies to managerecords that are created. Formulating a records management guide, the Newman University (2005) assert that the principles of the data protection act directed that data should only be kept for as long as needed but considering increase in litigation, some records need to be kept carefully for a longer period. These dual statements, if somewhat opposing, mean that the retention and disposition of records is a complex operation which institutions have to consider with care. Newman University (2005) records management policyguidelinesrecommended retention period classified records as follows:

- academic 2 – 6 years
- Management 3 – 11 years

- Estate/ Health and Safety 3 – 40 years
- Library and IT 1 – 7 years
- Suppliers 2 – 7 years
- Student etc. 1 – 7 years

This varying period of retention establishes the need for the estimation of time-to-dormancy for records to guide the formulation of a retention policy.

2.3.4 Policy guidelines for retention, disposal and archiving of business records

Records document organisation's business operations and are essential for effective managing of a business. The ability to properly and consistently retain records is especially important as most businesses are required by law to retain confidential client information, along with employee or company data, for a minimal amount of time. Many types of documents eventually outlive their purpose, and holding onto such records for too long puts an organisation at risk of a security breach and non-compliance with privacy legislation.

Organisations are expected to make retention decisions based on the content and purpose of records. In view of American Records Management Association International, (2016) this retention periods are determined by following these requirements legal and regulatory, fiscal, operational, historicalfactors.

Once its records retention requirements are determined, such organisation must conduct a risk assessment to determine the appropriate retention period for each type of record. Retention decision makers must be aware that the presence or absence of records can be either helpful or harmful to the organisation. Therefore, to minimise risks and costs associated with records retention, it is essential to immediately dispose of records after their retention period expires.

How long business records should be determined by a retention schedule that balances each record's usefulness with the legal requirements. To some degree, this will depend on the type of business, and the lifecycle of specific documents. It would be necessary to determine a retention schedule for each type of document, and then create a secure destruction schedule for those documents to reduce risks associated with data breaches.

The Government of Canada published a guideline that provides guidance to institutions regarding the establishment of minimum retention periods for those common administrative records which support the General Administration Function of the Government of Canada. Government of Canada, (2011) when records are covered by an existing MIDA the retention information offered takes the form of retention guidelines expressed in months, calendar years and fiscal years. In the absence of specific retention guidance and unless specified otherwise, the five year retention period for policy and procedures and the two year retention period for routine records should be applied to similar records related to each sub-heading/activity listed in this function. An organization may have separate policies and procedures for records retention, active file management, inactive file management, vital records, e-mail management, and any other area of records management. Policies and procedures set standards and serve as evidence of management's support of and investment in a compliant records management programme.

Haphazard patterns of records disposal may appear suspicious and can suggest that unfavorable or embarrassing records were destroyed intentionally. Records disposition should be an inherent element of an organization's overall records management program and should cover both active and inactive records. Standard policies should be set at the corporate and not at department level and be reviewed by legal and compliance professionals. The implementation of the policies should be treated as a consistent process, not an event, because they will need to keep pace with organisation growth and regulatory changes. Upon expiration of a record's required retention period, all records identified as eligible should be approved for destruction unless there is a legitimate business reason to postpone that destruction. The official version or "record copy" of a particular record should be maintained for the longest approved retention period subscribed in the Records Retention Schedule.Consistent disposal practices provide retention and regulatory compliance and decrease corporate risk when conducted in accordance with an approved records retention schedule. An established pattern of systematic records retention and disposition serves as evidence of an organization's good faith in attempting to conform to the law. The need for compliant records management best practices need to be demonstrated daily in all businesses

2.3.5 Policyguidelines for retention, disposal and archiving of patient medical records, a global outlook

Keeping good quality medical records is essential yet most developing countriesoften neglected this aspect of a health-care practitioner's workload and most neglected is the management of the retention and disposal of inactive records which are component of the inactive records phase of the records life cycle.Despite the facts thatlaw and organisational needs recommend keeping of patients'records,medical records professionals frequently pose questions about how long should the patient record be kept.Unfortunately there is no universal answer to this question and multiple factors need to be consideredin determining a retention and disposal of patient records. (McWay 2002; Abdelhak, *et al*2012).

Literature has shown that many African countries including Nigeria lack functional policy guidelines on patient recordsmanagement,resulting inpoor patientrecord retention and disposals.Where there are policy guidelines for records and archives management, none arespecifically developed for patients' medical records management.In Kenya, patient records management systems and practices face serious challenges because there are no conventional policies and standards that govern medical records management, (Health Matrix Network, 2008, 2013). In South Africa healthcare facilities are expected to retain patients records for a minimum of 6 years after the cessation of a patient's treatment (Health Professionals Council of South Africa (HPCSA), 2008), the policy stated that health records should be retained for not less than six (6) years from the date of last contact and in the case of minors and mentally incompetent patients, the records should be kept for a longer period (HPCSA, 2008).Thoughthere are guidelines on retention, the mode of disposal and archiving were not specified creating confusion for medical records professionals. It is expected that at a point in time every records out-used its purposes and become inactive and need to be disposed,according to (Zegers et al., 2009; Raff and James 2003) the value of each records should be evaluated before theyaredisposed of as they are source of quantitative information. This is in line with (Mennillo 2006).that in Australia records management policy expects an objective assessment of individual patient records basis rather than adopting a broad-axe approach

based on the length of time for which a patient has not been seen. The result is that each state imposes its own specific legal requirements on the retention of medical records, subject to implementation by health-care facilities.

Epidemiological studies had shown that often there is a long period between exposure and onset of certain conditions,supported by HPCSAguidelines that certain health conditions take a long period to manifest themselves therefore certain records be kept for periodsnot less than 25 years yet a balance must be reached between the costs of (indefinite) retention of records (in terms of space, equipment, etc.). In determining the appropriate cut-off for a specific recordretention policies, records managers consider active life of records based on frequency of use, function, resources and operationalrequirements.In line with this policy Singh (2011) cautioned thatrecords retention policies must be reasonable, consistent, and uniform in the context of the facts and circumstances surrounding the relevant documents and reflects deadlines and requirements imposed by the applicable law or regulations.

In line with the Personal Information Act (2013), records of personal information must not be retained any longer than is necessary for achieving the purpose for which the information was collected and processed unless in terms of professional rules of practice or contractually obligation. This is supported byHSE (2013),policy that however desirable it is to keep in original format every single record forever, the reality is there is limitation to storage capacity and perpetual retention of all records will be a breach of the Data Protection Acts. Another factor to be considered in determining the retention policies of patient records is the statute of limitation, this is law that sets forth a fixed time frame in which a lawsuit must be brought as specified in the applicable statute of limitation (McWay, 2010).

In Australia the Medical Insurance Group (2009) “do not recommend the destruction of medical notes but cautionedin their retention policythat the notes are hospitals best defense in the event of a claim, at which time you will need to rely on them. The group however, accept that storage of records indefinitely is often impractical and if records are not to be kept indefinitely then a valid policy for the retention, disposal and archiving is mandatory.

According to Singapore Medical Council (2000), medical records of patients can be safely disposed of 10 years after patient's last contact, except a minor or maternity records, untilpatient is 25 yearsor in case of brain-damaged to the patientthen records should be kept for 10 years after death.The Council classify medical records retentionperiod into primary and secondary record, which agreed withthe classification of record management into records in-current use, semi-current and non-current use in line with the phases of record life cycle theory (Penn, *et al* 1994; Agere, *et al*. 1999; Shepherd and Yeo 2003).

When developing policy guidelines the cost and space implications of keeping records indefinitely must again be balanced as well as statutory obligations to keep certain types of records for specific periods. According to Arabzadeh, Azizi, and Alimadadi (1999)different countries adopt different strategies policies and laws,in the United States each state has minimum medical record retention and disposal periods that range between 5 – 7 years from date of last contact, (Davis and Lacour, 2002), the Medical Council of New South Wales (2010)however required that apatient medical record be kept for at least 7 years from date of last entry in the record, unless the patient is less than 18 years and in that eventuality the record should be kept until the patient attains the age of 25 years.

The life cycle of records management begins with creation and ends when the record become inactive and is disposed of. The goal for hospital should be the efficient managementat all stages of the record life cycle to ensure record availability and economic reality. The processes involved in the creation phase of the record life cycle is easy and most institutions do not show concerns record management policies at this level. However, during the active phase when records are used, issue of maintenance arises. Lack of space, labour, maintenance processes and retrieval issues are encountered. According to Yaya, Japheth Abdulazeez et al.(2015)the main problems being faced by hospital authorities in records management in most developing countries include shortage of experienced personnel; lack of planning in storage of active records and need for effective storage, control of inactive records and lack of determination of records retention period.

Tavakoli and Jahanbakhsh, (2013), I n a study of 30 hospital in Isfahan, Iran find out that only 53.8% of the hospitals had retention policies and 34.6% on disposition procedures of which only 50% of these policies were developed by hospitals; the study showed further that while inpatient records were kept for about 15 years outpatient records were retained for between 3 and 25 years. The study concluded that majority of hospitals have no written retention and disposal policy or guidelines on medical records and those that have were developed by each institution

In the United State, there is no uniform standard record retention policy for all hospitals and providersinstead, there areinstitutional policiesdeveloped to create a compliant retention programme. A survey of 250 hospitals had showed that every hospital in the USA have one form of medical record retention policies in place, (Rinehart-Thompson. 2008). Rinehart-Thompson. (2008), in another study to find out the views of 526 physicians about medical records retention period in the United State, 41% of the respondents were of the opinion that medical records should be retained for 7 years, 6% for 10 years, 15% for 15 years, 14% for 20 years, and 24% agreed with more than 20 years, with rural hospitals and general physicians suggesting a shorter retention period than the specialised hospitals

However the American Health Information Management Associationhad recommended 10 years after the most recent encounter as a guide (American Health Information Management Association, (World Health Organization 2001, Cunningham and Wiedemann 2011; AHIMA), 2013; Downing and Pye 2013). According to AHIMA (2011) the development of record retention policyshould ensure patient health informationis available to meet the needs of continued patient care, legal requirements, research, education, and other legitimate uses of the organization;include guidelines that specify what records to keep, the period for which it is kept, and the storage medium on which it will be maintained; and a clear disposal policies and procedures that include appropriate methods of destruction for each medium.

In the United Kingdom records management policies stipulates a minimum retention period of not more than 30 years from creation, (Department of Health, 2006) whereas in Scotland minimum retention period of 8 years after conclusion of treatment was recommended, (Medical and Dental Defence Union of Scotland, 2013)..Retention of

patient records are mostly influenced byinternal and external forces; internal storage constraints, and fiscal concernsand external forces will range from statute of limitations to new technologies which play major roles in formulating a decision. (McWay, 2002).

Dearstyne (1985), identified the benefits of records management to include discouraging the creation of records that really aren't needed, reduces future costs by ensuring that expensive new equipment, saves space by removing inactive records and by ensuring the timely disposal of records that are no longer needed, good governance and faster access to needed information, ensure administrative continuity, and make informed policy decisions, preserving important research records. Corroborating the submission of Dearstyne (1985), Shepherd (2006) states that when records are managed as part of an appropriate records management programme it help the organisation to conduct business in an efficient, accountable manner, deliver services consistently, support managerial decision making and transparent policy formation and ensure continuity in policy execution, management and administration. In summary, an effective records management will ensure that records are available for use when needed.

It therefore follows that each hospital have the responsibility to develop policies on patient records retention, disposal and archiving guided by operational requirement and the statute of limitation or any requirement in laws. Above all each hospital operational requirementsshould be considered by determining the period of active life of the patient records. The determination of the period of active life of the patient records requires a good understanding of the general characteristics of the statistical distribution of the survivorshipof time-to-dormancy of the patient records,their parameters and the hazard functions.Alsofactors that may contribute to patient records dormancy would have to be determined.Interestingly no empirical studies to estimate the dynamics of retention or dormancy or factors that predict dormancyhas ever been published.Rather all the estimates of retention periods used for setting the rules were based on hunches and intuitions of hospital administrators. Little is known about time-to-dormancy in literature on records management, particularly in managing patient medical records. However, formulating a patient medical records retention policy requires the knowledge of the pattern of the statistical distribution,their parameters and hazard rates of the time-to-dormancy of such records.

Hospitals in Nigeria, and particularly in UCH, Ibadan, management of medical records have no standard retention,disposal or archiving policy nor guideline regulating what should be kept, for what period and what should not. This is however not in line with the global best practice for records management. A hospitalshouldhave a management and archiving policiesthat set time frame for retention, disposaland destruction of medical records. The policies must bebased on empirical approach to the estimation of time-to-dormancy of the medical records. This can only be done if the statistical distribution of the survival functions, their parameters and the hazard functions of the time-to-dormancy of the medical record created in UCH, Ibadan is determined.

The gap created by lack of policy on retention, archiving and disposal of medical records in Nigeria and particularly in UCH, Ibadan, needs to be addressed to strengthen records management practice; which is fundamental to quality care.

2.4 A brief review of statistical methods of survival analysisused in the study

Record like biological organisms iscreated (born) and becomes inactive (die)at a specific age, hence in records managementthe event of interest is the length or duration of time from creation to the time of inactivity (point of dormancy). Time to occurrence of a particular event carries a great significance in epidemiology, medical or biological studies. In medical research the outcome variable (or) event of interest may be death of a patient, relief from pain, the recurrence of symptoms, disease incidence, relapse from remission, remission duration of certain disease in clinical trials, incubation time of certain diseases, (Venkatesan, 1990, 2003); and in industry, failure time of certain manufactured products (Cox and Snell 1968; Crowley and Hu,1977; Kalbfleisch and Prentice, 1980; Miller, 1981; Cox and Oakes, 1984; Clayton, 1978; Jenkins, 1997; Andersen ,1992). When the main outcome under assessment is the time to an event of interest like we have in records management, the generic name for the time is survival time. Survival data are rarely normally distributed but are skewed and usually comprise typically of many early events and relatively few late ones. It is these features of the data that make the special methods called survival analysis, a collection of statistical procedures used to study time-to-event analysis,(Ramadurai and Ponnuraja 2011, Singh and Mukhopadhyay 2011), necessary.

Kaplan-Meier Product Limit Method (K-M)had been found to be very effective and useful in fitting distribution'sgeneral characteristics and estimation of their parameters, survival functions, *S(t),* form and shape of the hazard rate, *λ(t),* to survival time data, (Lee and Wang, 2003; Kleinbaum and Klein, 2012).

2.4.1 Kaplan-Meier survival curves

Most survival analyses of time to events use some or all of Kaplan–Meier (K-M) plots, log-rank tests, and Cox (proportional hazards) regression.Kaplan-Meier estimator, anon-parametric technique is often used in clinical and epidemiologic research to model time at risk until event,(Zhao, 2008; Rich et al, 2010). According to Wang and Chow (2007) the statistical method for the analysis of time-to-event data is very different fromthose commonly used methods for othertypes of data.Kaplan and Meier (1958),Cox and Oakes (1984) and Kalbfleisch andPrentice (2002) presented a non-parametric approach toestimate survival function using standard Kaplan Meier (KM) technique.

The Kaplan-Meier (K-M) method also referred to as the Product-Limit Estimator of survival at time, t, has been used variously in studies to determine the distribution and its parameters of time-to-event data. This is as a result of the methods ability to estimate the probabilities of survival functions and summaries the survival data of time-to-event data, (Abeyseker and Sooriyarachchi, 2009).Suppose thatan event of interest,here thepatient medical records became dormant in the time, t, with $t_1 < t_2 .....to, d_i$ eventsoccurred at time t_i and$Y_i$were the number of medical recordsthat were at risk at time t_i. The KM estimator defined for all values of t in the range was definedas:

$$\hat{S}(t) = \begin{cases} 1 & if \quad t < t_1 \\ \prod_{t_i \le t} \left(1 - \frac{d_j}{n_j}\right) & if \quad t_1 < t \end{cases}$$

where t_i denotes the first observed the time, d_i represents the number of individuals at time t, and n_i indicates the number of individualsthat had not experienced the event, and

have also not been censored, by time *i*. It is obvious for $t < t_1, \hat{S}(t) = 1$ and when $n_i = d_i$, then $\hat{S}(t) = 0$, $S\hat{}(t) = 0$ for $t \geq t_i$.

Again the KM estimator consists of the product of a number of conditional probabilities resulting in an estimated survival function $S(t)$ in the form of a step function.(Smithand Smith, n.d.). The KM estimator of the survival function *S(t)* can be defined as:

$$\hat{S}(t) = \prod_{t_s \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

$$= \prod_{t_j} \frac{n_j - dj}{n_j} \qquad \text{for} \qquad 0 \leq t \leq \text{ t} \qquad \ldots 2.1$$

where d_j is the number of recordsthat experience the event at time $t_{(j)}$, and $n_{(j)}$ is the number of recordsthat had not yet experienced the event at that time and are therefore still at risk for experiencing it, (Akram, et al, 2007; Zhao, 2008).The Kaplan-Meier survival curve can then be defined as the probability of surviving in a given length of time while considering time in many small intervals(Altman, 1992, (Goel, Khanna, and Kishore 2010).

2.4.2 Estimating the median and percentiles of time-to-event from the Kaplan–Meier

The distribution of survival time always tends to be positively skewed, hence the median is usually preferred as a summary measure. The *p*-percentile of survival time is the analysis time at which *p%* of subjects have failed and *1−p%* have not. Hence the median survival time is the time beyond which 50% of the subjects in the population under observation are expected to survive, i.e., the value of:

$$t(50) \quad at \quad \hat{S}(t(50)) \quad = 0.5 \qquad \ldots 2.2$$

Other percentiles of survival times are obtainable from the Kaplan–Meier product-limit estimate of the survivor function,*S(t)*. Because the non-parametric estimates of *S(t)* are step-functions, it will not usually be possible to realise an estimate of survival time that makes the survivor function exactly equal to 0.5. Instead, the estimated median survival

time, $\hat{t}(50)$, is defined to be the smallest observed survival time for which the value of the estimated survivor function is less than 0.5 (Collett, 2003)

Based on this assertion estimated median survival time is given by:

$$\text{Estimate}\,\hat{t}(50) - \min\left\{t_i \,\middle|\, \hat{S}(t_i) < 0.5\right\} \qquad \ldots 2.3$$

where t_i is the observed survival time for the i^{th}subject, $i = 1, 2, \ldots, n$. In general, the estimate of the p^{th} percentile is:

$$\hat{t}(p) = \min\left\{\hat{t} \,\middle|\, \hat{S}(t_i) < 1 - \frac{p}{100}\right\} \qquad \ldots 2.4$$

The variance of the percentile is:

$$\text{var}\left[\hat{S}\{t(p)\}\right] = \left(\frac{d\hat{S}\{t(p)\}}{dt(p)}\right)^2 \text{var}\{t(p)\}, \qquad \ldots 2.5$$

Where *t(p)* is the p^{th}percentile of the distribution and $\hat{S}\{t(p)\}$ is the Kaplan Meier estimate of the survivor function at *t(p).* Now,

$$-\frac{d\hat{S}\{t(p)\}}{dt(p)} = \hat{f}\{(p)\}, \qquad \ldots 2.6$$

an estimate of the pdf of the survival time at *t(p*), and rearranging equation (2.4), we have

$$\text{var}\{t(p)\} = \left(\frac{1}{\hat{f}\{t(p)\}}\right)^2 \text{var}\left[\hat{S}\{t(p)\}\right] \qquad \ldots 2.7$$

The standard error of estimated *t(p),* the estimated p^{th}percentile is given by:

$$se\{\hat{t}(p)\} = \frac{1}{\hat{f}\{\hat{t}(p)\}} se\left[\hat{S}\{\hat{t}(p)\}\right], \qquad \ldots 2.8$$

and the estimated p^{th} percentiles 100(1 – α) confidence interval for *t(p)* has a limits of

$$\hat{t}(p) \pm z_{\alpha/2} se\{\hat{t}(p)\} \qquad \ldots 2.9$$

Where $z_{\alpha/2}$ is the upper (one sided) α/2 point of the standard normal distribution.

The interest in this study is onaverage cumulative dormancy at time, *t*, defined as the percentage of patient medical records that became dormant (inactive) at time, *t*.

If the estimated median survival time, $\hat{t}(50)$, is defined to be the smallest observed survival time for which the value of the estimated survivor function is **less** than 0.5, and the p^{th}percentile of survival time is the time at which p% of subjects in the population have failed and *(1−p)%* have not, then by extension it will be safe to say that the p^{th}percentile of dormancy time is the time at which *p%* of patient medical recordsbecome dormant (inactive) and *(1−p)%*are still active. Thus we substitute the Median Dormancy Time, *(MDT*) for theMedian Survival Time*, MST*. From this we can conveniently estimate the dormancy time for 25^{th}, 50^{th}, 75^{th} and the 95^{th} percentiles.

The quantile function Q(τ) is related to the cumulative distribution function F(t), where S(t) = 1−F(t), by the relationship:

$$F\left(Q_\tau(\tau\ )\right) = P\left(T \le Q_\tau(\tau)\right) = \tau \qquad \ldots 2.10$$

Bellavia,(2015), explained that there is a univocal correspondence between the quantile and the survival function. When T is continuous, *Q(τ) = t only if F(t) = τ*, that is, the quantile function is the minimum value of t below which a randomly selected individual from the population will fall (100·τ)% of the times,

2.4.3The hazard function, λ (t)

The primary focus of survival analysis is to model the hazard rate, which has the following relationship with the *f(t)* and *S(t):*

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} \qquad \ldots 2.11$$

Indicating a defined relationship between *S(t) and h(t),* which is given by

$$\lambda(t) = -\frac{d}{dt}[\log S(t)] \qquad \ldots 2.12$$

The focus of this study was to estimate the form and shape of the hazard rate, *λ(t),*anddetermine the distribution of time-to-dormancyof patient medical records.The hazard function gives the conditional failure rate, defined as the probability of failure during very small time interval, given that the individual having survived to the

beginning of the intervals or as the limit of the probability that an individual fails in a very short interval, $(t = \Delta t)$, given that the individual has survived to time t.

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\left[\begin{array}{l} an \quad individual \quad dying \quad in \quad the \quad time \text{ interval } (t+\Delta t) \\ given \quad the \quad individual has \ survived \ to \quad t \end{array} \right]}{\Delta t}$$

i.e. $\lambda(t) = \lim_{\Delta \to 0} \frac{P(t \le T < t + \Delta / T \ge t)}{\Delta t} = \frac{f(t)}{S(t)}$ … *2.13*

This can also be defined in terms of the cumulative distribution *F(t)*and probability density function f(t) as: $\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$ … *2.14*

The hazard function describes the relative likelihood of an event occurring at time, t, f(t) conditional on the subject's survival up to that time, t, S(t). The hazard rate thus describes the instantaneous rate of failure at time, t, and ignores the accumulation of hazard up to time, t, unlike F(t) and S(t) the hazard function.

It is of note that the derivative of the survival function $S(t)$ is equal to $f(t)$. Thedistribution of T is specified by its hazard function as well because the survivorfunction is determined by the hazard function:

$$\frac{d}{dt}\ln(S(t)) = \frac{-f(t)}{S(t)} = -\lambda(t) \quad \ldots 2.15$$

While the survivor function focuses on the probability of not failing, the hazard function focuses on failing, thus, in some sense, it can be considered as being the complement of the information provided by the survivor function. The greater the hazard function therefore, the shorter is the survival time, (Rao and Schoenfeld, 2007); the hazard function may increase, decrease, remain constant, or indicate a more complicated process such as the bathtub curve that describes the process of human life, which at infancy has an initial period of high risk of death, approximately constant at middle age and increases with old age, (Lee and Wang 2003,Hagar and Dukic 2015),

A commonapproach to estimate $\lambda(t)$, is to use the cumulative hazard, $H(t)$, This is defined as the integral of the hazard, or the area under the hazard function between times 0 and t, and differs from the log-survivor curve only by sign, that is

$$H(t) = -\log[S(t)] \qquad \ldots 2.16$$

The interpretation of $H(t)$ may be difficult but can be thought of as the cumulative force of mortality, or the number of events that would be expected for each individual by time t if the event were a repeatable process, (Clark, Bradburn, Love, and Altman 2003). It serves as an intermediary measure for estimating $\lambda(t)$, and also a diagnostic tool in assessing the validity of the Weibull model by plotting the log negative log of the Kaplan Meier survival estimates, *log(-log of S(t),* against the log of time, *log(t).* The slope of a line fitted to the plot can then be used to estimate the shape parameter of the distribution.

2.5 Modelling time-to-event data

Time-to-event data are modelled to explore how the survival experience of a group of subjects depends on the value of one or more explanatory variables, whose values have been recorded for subject at time origin. Two main reasons account for this, first is to determine which combination of potential explanatory variables affect the form of the hazard function, secondly is to obtain an estimate of the hazard of the hazard function itself for the subject, (Collett, 2003). Two common approach are the semi-parametric with the Cox PHmodel the most widely used, andthe Exponential and Weibullmodels as the most common distribution for parametric modelling of survival data, (Kleinbaum and Klein, 2012).

2.5.2 Semi-Parametric Survival Analysis Models

According to Buis, (2006) non-parametric, semi-parametric and parametric techniques are popularand often used in the analysis of time-to-event data. Cox Proportional Hazards Model,introduced by Cox (1972) do not impose a parametric form for the distribution of hazard of survival. Though the most frequently used, the Cox regression do have its limitations,this is especially so if we have additional information on the characteristics of each individual which may be affecting its survival.

2.5.2 Cox regression models

Cox proportional hazard model is semi-parametric to the extent that no assumptions are made about the form of the baseline hazard, except for a key assumption which is the proportional hazards. The cox model is of the form:

$$h(t;x) = h_0(t)\exp\{\beta_1 x_1 + ... + \beta_k x_k\} \quad \text{... 2.17}$$

Where h(t; x) is the hazard function at time t, for a subject with covariate value $x_{1, ...} x_k$ $h_0(t)$is the baseline hazard function, i.e., the hazard function when all covariates equal zero, expis the regression coefficient for the i^{th}covariate, x_i the i^{th}covariate in the model, and $\beta_i$is the regression coefficient for the i^{th}covariate

The Cox Model is different from ordinary regression in that the covariates are used to predict the hazard function, and not *Y* itself. The baseline hazard function can take any form, except that it cannot be negative. The exponential function of the covariates is used to insure that the hazard is positive. There is no intercept in the Cox Model as any intercept could be absorbed into the baseline hazard. The proportional hazards follows that the ratio of *h(t; x)* for two different covariate values are:

$$\frac{h(t;x)}{h(t;x)} = \frac{h_0(t)\exp\{\ \beta_1 x_{i1} + ... + \beta_k x_{ik}\}}{h_0(t)\exp\{\ \beta_1 x_{j1} + ... + \beta_1 x_{jk}\}} \quad \text{... 2.18}$$

$$= \exp\{\ \beta_1(x_{i1} - x\ ) = ... + \beta_k(x_{ik} - x_{jk})\} \quad \text{... 2.19}$$

h(t) cancels out => the ratio of those hazards is the same at all-time points and for a single dichotomous covariate, say with values 0 and 1, the hazard ratiois

$$\frac{h(t;\ x\ =\ 1)}{h(t;\ x\ =\ 0)} = \frac{h_0(t)e^{\beta * 1}}{h_o(t)e^{\beta * 0}} = \frac{e^{\beta}}{e^{0}} = e^{e} \quad \text{... 2.20}$$

Let T_i be the failure time for subject i, $i = 1,....,n$. If T_i follows the Cox proportional hazards regression model, then the hazard function for T_i at time *t > 0,* conditional on the *p x 1* covariate vector Zi , is

$$\lambda(t|Z_i) = \lambda_0(t)\exp(\beta' Z,) \quad \text{... 2.21}$$

where $\lambda_0\ (t)$ is the baseline hazard function (i.e. the hazard function when all covariates take value zero) and β is a $p \times 1$ vector of regression coefficients. Statistics are designed

to check whether interaction terms between elements of z_i or higher order terms in the elements of Z_i need to be added to $\beta' zi$.

Using counting process notation, the information in the data can be represented by

$$\{N_i(t), Y_i(t), Z_i : 0 < t < \infty\}$$

where $N_i\ (t)$ takes value one if subject *i* has been observed to fail prior to time t and takes value zero otherwise and Yi (t) takes value one if subject i is at risk at time *t* and takes value zero otherwise. Then the Cox partial likelihood score vector equals

$$u(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \{Z_i - \bar{Z}(s, \beta)\} dN,(s) \qquad \ldots 2.22$$

Where $Z(s, \beta) = \dfrac{\sum_{j=1}^{n} Z_j Y_j(s) e^{\beta z_i}}{\sum_{j=1}^{n} Y_j(s) e^{\beta z_j}}$ is a weighted average of the Z_i 's and $dN_i(s) = N_i(s) - N_i(\bar{s})$ is a binary random variable that equals one if subject *i* fails at time *s* and equals zero otherwise. The maximum partial likelihood estimate $\hat{\beta}$ is the solution to $u(\hat{\beta}) = 0$

2.5.3Test for Proportional Hazards (PH) Assumption

A key assumption of the Cox regression model is the proportional hazards assumption that the hazard ratio is constant over time, or that thehazard for an individual is proportional to the hazard for any other individual, (Therneau and Grambsch, 2000).

Let $x^* = (x_1^*, x_2^*, ..., x_p^*)$ and $x = (x_1, x_2, ..., x_p)$ be the covariates of two individuals.

The hazard ratio is given as follows:

$$\exp\left[\sum_{i=1}^{p} \beta_1 (x_i \hat{} = x_i)\right]. \qquad \ldots 2.23$$

Suppose two groups 1 and 2 (say, group 1 is receiving a new treatment and group 2 is receiving a standard treatment), are compared with respect to the hazard of each group. Let λ_1 (t | group 1) and λ_2 (t | group 2) be the hazard functions of group 1 and group 2 respectively, where *t > 0.* Then the two groups are said to have proportional hazard, when the hazard ratio $\Psi$is constant over time. That is,

$$\frac{\lambda_1(t|group\quad 1)}{\lambda_2(t|group\quad 2)} = \psi, for\quad all\quad t \qquad \ldots 2.24$$

Though the Cox PH model is the most popular method of examining the effect of explanatory variables on time-to-event data, it however requires that the assumption of proportional hazards be assessed when fitting a PH model and there are numerous methods in the literature (Cox and Snell, 1968; Moore and Spruill, 1975Hosmer. and Lemeshow, 1980; Schoenfeld, 1980; Schoenfeld, 1982; Moreau, O'Quigley and Lellouch, 1986; Parzen and Lipsitz, 1999)for checking the assumption of PHs

This assessment can be done by many numericalor graphical approaches, none of these approaches are known to be better than the others in finding out whether the hazards are proportional or not. However, Schoenfeld's global test and the graphical approach had been successfully used.

2.5.4The Schoenfeld's global test for Cox PH assumption

Schoenfeld (1980), Moreau, O'Quigley and Mesbah (1985), and Moreau, O'Quigley, and Lellouch (1986) have proposed goodness-of-fit statistics for the Cox proportional hazards models. These statistic are based on the notion of partitioning the subjects into mutually exclusive regions based on their covariate values.Abeyseker and Sooriyarachchi (2009), had shown the Schoenfeld's global goodness-of-fittest as the most objective among other methods.

Studies had shown that the global goodness-of-fit test proposed by (Schoenfeld, 1980) was considered useful for testing the Cox PH assumption of the time-to-event data, because of its power to detect the insufficiency of covariates in describing the relative risks and the assumption of PH, when applied to the fitted model.

With the global statistical significance of the model,output gives p-values for three alternative tests for overall significance of the model: The likelihood-ratio test, Wald test, and score log-rank statistics. These three methods are asymptotically equivalent such that for large enough N, they will give similar results and for small N, they may differ somewhat.

2.5.5The "log-log" plot for testing Cox PH assumption

In a graphical test an initial indication of failure of this assumption is when the survival curves under consideration cross and diverge. The most widely used approach is the so-called "log-log" plots, which are plots of $\log(-\log(S(t)))$ vs. $\log(t)$, where t = time. When these plots show a non-parallel pattern, the proportional hazards assumption is said to be violated, (Kleinbaum and Klein, 2012).The PH assumption implies that $S(t) = S_0(t)^{esp(\beta x)}$; thus, the survival curves are powers of one another. This observation are used as a check of the PH assumption through inspection of the Kaplan-Meier survival curve estimates. The PH assumption also implies that

$$H(t) = H_0(t)exp(\beta x), \quad \ldots 2.25$$

and, thus, the cumulative hazard curves have a constant ratio. Here again, crossing curves indicate violations of the PH assumption. Since

$$H(t) = -\log S(t), \quad \ldots 2.26$$

we used the - log transformation of the Kaplan-Meier estimate for this assessment. The PH assumption further implies that

$$\log H(t) = \log H_0(t) + \beta x; \quad \ldots 2.27$$

thus, the PH model can be rewritten as:

$$\log[-\log S(t)] = \log[-\log S_0(t)] + \beta x. \quad \ldots 2.28$$

Therefore, under PH, plots of $\log[-\log S_i(t)]$ (or equivalently, plots of $\log \hat{H}_i(t)$ are roughly parallel. It is possible to simply take *log[-log]* transformation of the Kaplan-Meier estimates and check for equidistance between the curves for single binary covariate, while for the two-sample case, several literature suggested plotting $H_1(t)$ *vs* $H_0(t)$. Under PH, $H_1(t) = \theta Ho(t)$ *where* $\theta = exp(\beta)$ is constant over t.

Plotting the log negative log Kaplan-Meier survival estimates [*log(-log of S(t)],*against the log of time,*log(t),* for two or more levels of covariates presents five possible results:

- Parallel straight lines implies that Weibull, Proportional Hazard and Accelerated Failure Time(AFT) assumptions hold
- Parallel straight lines with slope of 1 indicates Exponential. PH and AFT

- Parallel but not straight lines indicates PH but not Weibull or AFT however Cox model can used
- Not parallel and not straight indicate the distribution is not Weibull and the PH is violated
- Not parallel but straight lines is an indication that Weibull holds, but PH and AFT are violated, different p

The key points are that straight lines support the Weibull assumption and parallel curves support the PH assumption andif the plots are parallel but not straight then the PH assumption holds but not the Weibull, (Kleinbaum and Klein, 2012).

2.6 Parametric survival analysis modelsfor time-to-event data

Experience had shown that on some occasions pattern of survivorship data follows a predictable pattern and in such situations, parametric distributions can be used to describe time-to-event. Parametric models make assumptions about the distribution of failure times and the relationship between covariates and survival experience, specifying the distribution of the baseline hazard/survival function according to some (defined) probability distribution,(Stevenson, 2009). With the parametric models, the outcome is assumed to follow a certain known distribution,(Cox, 1992; Buis, 2006); and can almost have the look and feel of a normal-errors linear regression analysis, (Kargarian-Marvasti, Rimaz, Abolghasemi, Heydari., 2017). It then follows that parametric models are used when the nature and form of the hazard functions are known.

Though Kaplan-Meier estimator is a very useful tool for estimating survival functions, sometimes, interest is to make more assumptions that allow for more detailed modeling. By specifying a parametric form for *S(t),* one can:

- easily compute selected quantiles of a distribution;
- estimate the expected failure time;
- derive a concise equation and smooth function for estimating *S(t), H(t) and h(t);*
- estimate *S(t)* more precisely than *KM* assuming the parametric form.

Parametric models can beexpressed in both proportional hazard form, andaccelerated failure time (AFT) form. Several parametric distributions are available but in

epidemiological and clinical studies, the most common used are theExponential and Weibull, (Cox, 1992)..

2.6.1 Exponential model

The exponential distribution probably is one of the most commonly usedparametric distributions for time-to-event data,(Kalbfleisch and Prentice 1980; Collet 2003, Montaseri, et al 2016). Statistical methods for the exponential distribution are fairly simple (Lawless, 2003) and the distribution has the memoryless property meaning that how long an individual has survived does not affect its future survival (Lee, 1992). It is used with ordered data, that is, the first individual to fail is the weakest, the second to fail is the second weakest, and so on (Epstein and Sobel, 1953).

Animportant distribution in survival studies and like a normal distributionin other statistical areas, exponential distributionhas played an important role in time to event analysis. Lawless (2003), Stevenson (2009), had shown thatdistribution is characterised by a constant function:

$$\lambda(t) = \lambda_0(t)\exp^{z\beta} \quad \ldots 2.29$$

Thus the hazard for a given *z*, is constant and this produces an exponential failure distribution but the failure rate depends on the *z,* the covariates. Exponential distribution is an accelerate failure time (AFT) model.

Where $\lambda > 0.$ the pdf and survivor function are:

$$f(t) = \lambda e^{-\lambda t} \quad \text{and} \quad S(t) = e^{-\lambda t}$$

$$\frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \quad \ldots 230$$ The

exponential density with mean parameter λ is

$$\lambda(t) = \lambda \quad \ldots 2.31$$

So mean survival time is:

$$\mu = E(T) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda}$$

Letting $S(t_{0.5}) = e^{-\lambda t_{0.5}} = 0.5,$ then the *Median Survival Time (MST)* is:

$$t_{0.5} = \frac{\log 2}{\lambda.} \qquad \ldots 2.32$$

The baseline hazard is assumed to beconstant within each time period, but can vary between time periods, (Stevenson, 2009), It involves one parameter λ (i.e. time-independent hazard rate),and otherimportant parameters (e.g., median survivaltime) can be computed based on the λ,and if the time between failures has the probability density function

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & for \quad t > 0, \lambda > 0 \\ 0 & otherwise \end{cases} \qquad \ldots 2.33$$

It also implies that the hazard function is constant over the time interval and the event rate is independent of t. The failure rate is:

$$z(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \qquad \ldots 2.34$$

why λ is called the rate parameter of the exponential distribution and more generally, the hazard need not be constant because it expresses the instantaneous risk of an event, the hazard rate is the natural response variable for regression models for survival data(Fox, 2014).

2.6.2 Weibull Model

Weibull distributionis known for its flexibility and has been used for many applications including product life and strength/reliability testing. It models the rate of failure as time increases (Nelson, 1982). Amodelthat can be used to describe varioustypes of observed failures of components and phenomena and it is the most widely used parametric survival model,(Lai, 2006; Kleinbaum and Klein, 2012). Described by a scale parameter λ and *p*shape parameter. If p < 1, the instantaneous hazard monotonically decreases with time, if $p = 1$, the instantaneous hazard is constant over time (equivalent to the exponential distribution) and if $p > 1$, the instantaneous hazard increases with time.The hazard at time t for an individual with covariates z is defined as:

$$\lambda(t,z) = \lambda p(\lambda t)^{p-1} e^{z\beta} \qquad (for \quad \lambda, \quad p > 0) \qquad \ldots 2.35$$

where $z = (z_1, z_2, \ldots z_s)$ is a vector of explanatory variables and is a vector of regression parameter; and the hazard is:

- monotone increasing if p > 1
- monotone decreasing if p < 1
- reduces to the constant exponential hazard if p = 1

According to Hallinan (1993) Chin-Diew (2006) the Weibull distribution has appeared in five different forms. The two common forms of the distribution function are:

$$F(t,\theta) = 1 - \exp\left[-\left(\frac{t-\tau}{\alpha}\right)^{\beta}\right], \quad t \geq \tau \qquad \ldots 2.36$$

and

$$F(t,\theta) = 1 - \exp\left[\lambda(t-\tau)^{\beta}\right], \quad t \geq \tau. \qquad \ldots 2.37$$

The parameters of the distribution are given by the set $\theta = \{\alpha, \beta, \tau\}$ with $\alpha > 0, \beta > 0$ and $\tau \geq 0$; where α is a scale parameter, β is the shape parameter that determines the appearance or shape of the distribution and τ is the location parameter. Frequently, the location parameter is not used, and the value for this parameter can be set to zero. When this is the case, the *pdf* equation reduces to that of the two-parameter Weibull distribution. When τ = 0, above equations become the two-parameter Weibull distribution with:

$$F(t,\theta) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \quad t \geq \tau \qquad \ldots 2.38$$

and

$$F(t,\theta) = 1 - \exp\left[\lambda(t)^{\beta}\right], \quad t \geq \tau. \qquad \ldots 2.39$$

There is also a form of the Weibull distribution known as the one-parameter Weibull distribution. This in fact takes the same form as the two-parameter Weibull *pdf*, the only difference being that the value of β is assumed to be known beforehand.Murthy et al (2003) refer to this as the standard Weibull model, but Johnson., et al.(1994) refer to a standard Weibull when α = 1 (or λ = 1).

The distribution is both a proportional hazards (PH) and accelerated failure time model, so both hazard ratios and time ratios can be estimated and if the AFT assumption

holds then the PH assumption also holds (and vice versa), which is unique to the Weibull distribution (Cox and Oakes, 1984;Kleinbaum and Klein, 2012) and holds if p the shape parameter, does not vary over different levels of covariates. Also for Weibull distribution, the $ln[-ln(S(t)]$ is a linear function of $ln(t)$ with slope p and intercept $p\ ln(\lambda),$ (Kleinbaum and Klein, 2012), and if the slope equals 1 then t follows an exponential distribution. This property allows a graphical evaluation of the appropriateness of a Weibull distribution for modelling time-to-event data. Uthman (2007) analysing 2003 Nigeria Demographic and Health Survey had shown relationship of low birth weight and other factors on infant mortality using multivariate survival regression procedure with Weibull hazard function. Chen Zhu (2012) on failure rate had also shown that the Weibull distribution is very flexible and powerful which could model different types of failure times.

The exponential distribution had been described as a special case of the Weibull distribution. The key property for the Exponential distribution is that the hazard is constant over time (not just the ratio). Both models can be run as a PH model or an AFT model, (Kleinbaum and Klein, 2012).

2.7Diagnostic Assessment of Survival Time and Distribution

2.7.1Goodness-of-fit testfor model selection criteria

Most times it is important to find out how much a model fits a data set,when used inappropriately, statistical models may give rise to misleading conclusions.Model validation is therefore important to assess the reliability and the ability of the models to predict future risks. Regardless of which type of model is fitted and how the variables are selected to be in the model, it is important to evaluate how well the model represents the data. A survival model is only adequate if it represents the survival patterns in the data to an acceptable degree. This aspect of a model is known as goodness of fit. In practice, the issues in choosing the most appropriate type of model and the most appropriate covariates are heavily related, and the adequacy of a model may be assessed in several ways, Bradburnet al, (2003).

Akaike's Information Criteria (AIC), log-log of survival against log of survival time, Schoenfeld's global tests, Baysian Information Criteria (BIC) and $R^2$are common test for a models best-of-fit. Stanley, Molyneux and Mukaka, (2016) compared the

performance of Cox, Weibull, and Exponential models in a randomized study, Akaike's Information Criteria (AIC), plots of log-log of survival against log of survival time, and the Schoenfeld's global tests were used to test suitability of the PH assumption. Results showed that Exponential model was the best fitting method,concluding that Exponential models can elicit more valid results than semi-parametric CoxPH model in a clinical trial with small sample size. Bradburn et al, (2003), however observed that AIC (Akaike, 1974), a statistic that trades off a model's likelihood against its complexity, may also be used when comparing the viability of different parametric models. A retrospective study on medical records of 178 patients by Saikia and Barman, (2016), AIC, Baysian Information Criteria (BIC) and R^2 were used to identify the best fitted model and it was found that Cox PH model was better than the other parametric counterparts for the esophagus cancer patients' data.

2.7.2Log-rank test for equality of survivor functions

Mantel's (1966) generalization of the Savage (1956) test, often referred to as the *log-rank test,* (a non-parametric test which makes no assumptions about the survival distributions), is the most widely used method of comparing two or more survival curves. It compares observed number of events, say O_i for treatment group i, to the expected number by calculating the test statistic

$$\chi^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i} \quad \ldots 2.40$$

This value is compared to a χ^2 distribution with (g-1) degrees of freedom, where g is the number of groups. In this manner, aP-value may be computed to calculate the statistical significance of the differences between the complete survival curves.

The null hypothesis is that there is no difference in survival between groups. The log rank statistic is approximately distributed as a chi-square test statistic.This test depends on a single assumption - that the hazards in one group are uniformly higher (or lower) than in the other group by some proportionality factor $\lambda \geq 0$, i.e.

$$h_{i,group\ 2} = \lambda \quad x \quad h_{i,group\ 1}$$

λ is regarded as the relative risk of medical records dormancy between the cohorts over a common dormancy time.

The Log–Rank Test can be extended to compare 2 population survival functions, Generally, to compare the distribution of survival times between 2 or moregroups, Kalbfleisch and Street (1990), suggested setting up a k 2×2 contingency tables:

	Failure	Survivals	At risk
Treatment	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Control	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
Total	d_i	$n_i - d_i$	n_i

A 2×2 table of Failures and Survivals at Failure Time 't'

We can then test whether or not the two or more survival functions differ by computing the following statistic and conducting the log–rank test, described below:

$$e_{1i} = \frac{n_{1i} d_i}{n_i} \qquad v_{1i} = \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n - 1)} \ldots 2.41$$

$$O_1 - E_1 = \sum_{i-1}^{k} (d_{1i} - e_{1i}) \qquad V_1 = \sum_{i=1}^{k} v_{1i} \qquad \ldots 2.42$$

Where the test hypothesis are

H_o : distribution are same

H_1 : that the distribution and different

A significant positive test statistics imply that distribution are different otherwise the distribution are same.

2.7.3Akaike's Information Criteria (AIC)

The AIC, AICc, mAIC and BIC had been used variously to assess the suitability of Cox regression, Weibull and Exponential models as best fit to time-to-event data with

good results (Kalbfleisch and Prentice, 1980., Efron, 1997; Oakes, 1997; Lawless, 1998; Saikia and Barman, 2016; and Stanley, Molyneux and Mukaka, 2016).

The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models.

Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model, then the AIC value of the model is (Akaike 1973, Burnham and Anderson 2003,Aho, Derryberry and Peterson, 2014),

$$AIC = 2k - 2\ln\left(\hat{L}\right) \quad \ldots 2.43$$

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages over fitting, because increasing the number of parameters in the model almost always improves the goodness of the fit.

2.7.4 Underestimation and overestimation in survival analysis

The under- or over-estimation techniques assess or measure the degree at which the reported time data or its analysis resulted into too low or too high estimate that quantify target population. This judgement of estimate that is unfavourable can be due to potential systematic bias that was not accounted for during estimation. Given that non-response rate has been accounted for in the data collection and gathering as well as outlying value and a well define censoring that accommodate all observe group that are loss to follow up, underestimation error may occur in the study if the estimate of survival-time difference between the observer point of analysis and the patient last time of contact is less than the survival-time difference between the patient'slast time of contact and the patient's penultimate time. If this is so then the assumption was that the study was carried out too early.

Unlike underestimation, Overestimation assess or measure the degree at which the reported/analysed data resulted into an estimate too high or too extreme than expected. This can be due to systematic error in data collection and gathering and sometimes the use of inappropriate statistical technique for data analysis. According to Mukangai, and

Odongo, (2016), the Kaplan Meier technique for estimating survival time sometimes overestimate in the presence of ties and may have severe implications particularly when using its estimate to inform healthcare planning and policy decision making. This may be due to non-interval measurement incorporated by the technique in estimating survival probabilities. This however can be minimized by incorporating a well define censoring indices in the estimation of survival probabilities in the presence of ties. This notwithstanding Kaplan-Meier method are frequently considered in survival analysis to estimate the survival parameters in the absence of any competing risk, (Beuscart, Pagniez, Boulanger, Lessore de Sainte Foy , Salleron, Frimat and Duhamel, 2012; Noordzij, Leffondré, van Stralen, Zoccali, Dekker, Jager, 2013; Mukangai, and Odongo, 2016). Thus, overestimation error may occur in a study if the estimate of the censored survival-time difference between the last time of patient contact and the patient penultimate time is higher than the estimated censored survival-time difference between the observer-study time and the patient last time of contact. Hence the need to check for overestimation.

2.8 Theoretical Framework for the Study

Records management cycle has been discussed severally, especially the aspect of retention, disposal and archiving of records. However not much have been done to estimate dormancy-time of medical records towards developing policies on retention, disposaland archiving. A well-known theory on records management is the records life cycle theory (Penn, Pennix and Coulson 1994) (figure 2.1), that records are born (created), lived anactive life through to semi-active life to an inactive life when the record is assumed dead (dormant).

The records lifecycle has been the subject of professional discourse particularly based on the historical experiences of the US National Archives in the 1930s and 1940s. During that time, Federal Agencies expanded exponentially leading to large volumes of records (Henry 1998). American archival scholar T. Schellenberg is credited with solidifying the concept in the 1950s with an emphasis on records professionals being involved in working with agencies at the earlier stages of the lifecycle (Bantin 1998; Borglund and Öberg 2006). At the core of the concept is that all records have a lifespan

beginning with record creation, use/maintenance and storage until final disposition or preservation. This concept has often been represented through linear illustrations (figure 2.2) and, on a few occasions, in circular illustrations (National Archives and Records Administration [United States], Office of Management and Budget [United States] et al. 2005). In practice there are aspects that are circular and others linear, however scholars had developed a model representing the records lifecycle in both linear and circular terms (Figure 2.3),as adapted from New Zealand's Digital Content Life Cycle (Digital NZ 2014).

An extension of this theory is the Recordsand Information Life Cycle Management Theory, (figure 2.4) which discusses the management of records at the various stages of the records life cycle theory. The records life cycle theory form the bases for the study.

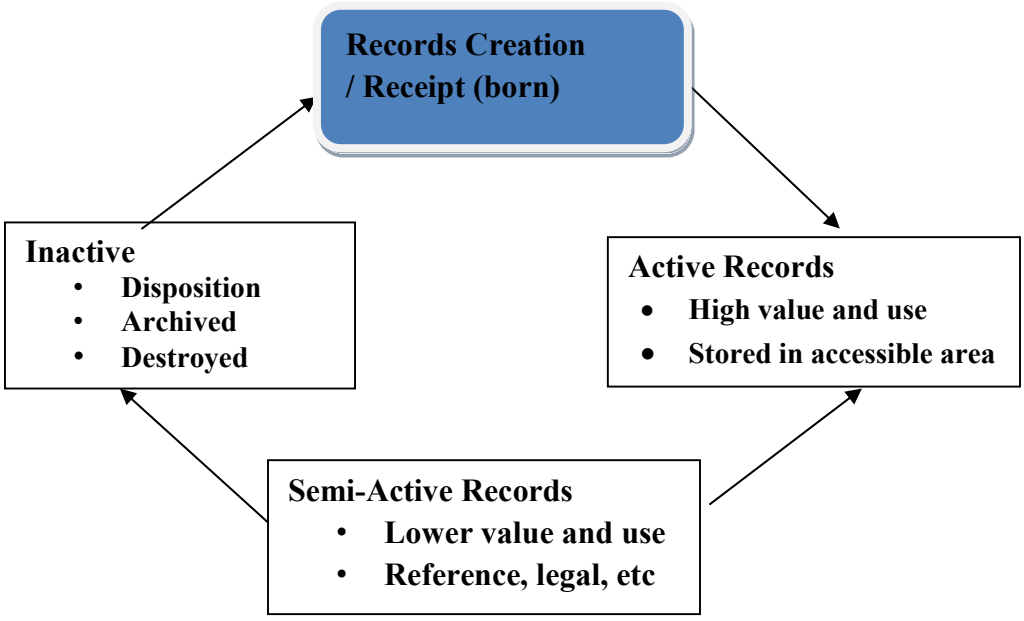


Figure 2.1: Records Life Cycle Model diagram (circular model)

Source: cms.montgomerycollege.edu

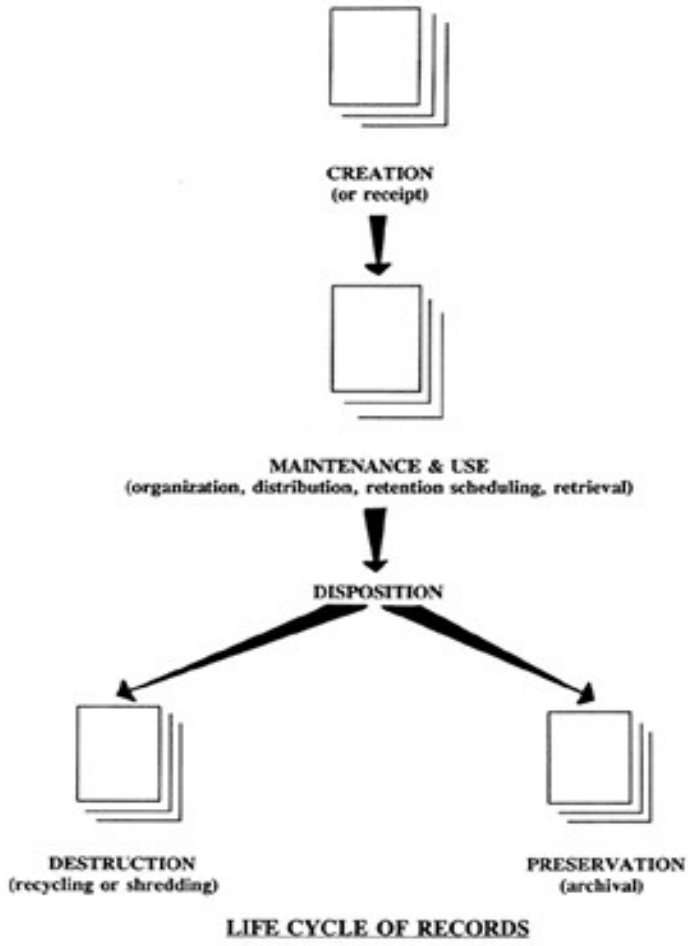


Figure 2.2 Records Life Cycle Model diagram (Linear)

Source: Caribbean centre for Development Administration (CARICAD)

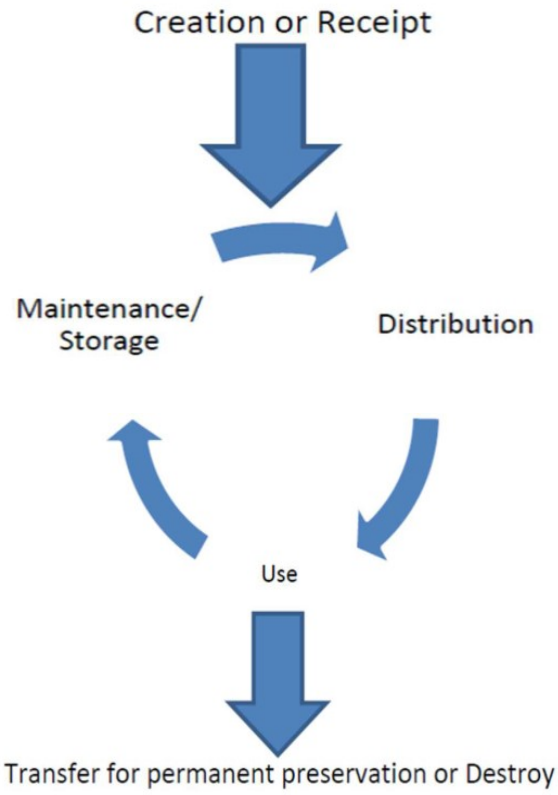


Fig 2.3 The lifecycle in both linear and circular terms

Source: adapted from New Zealand's Digital Content Life Cycle (Digital NZ 2014).

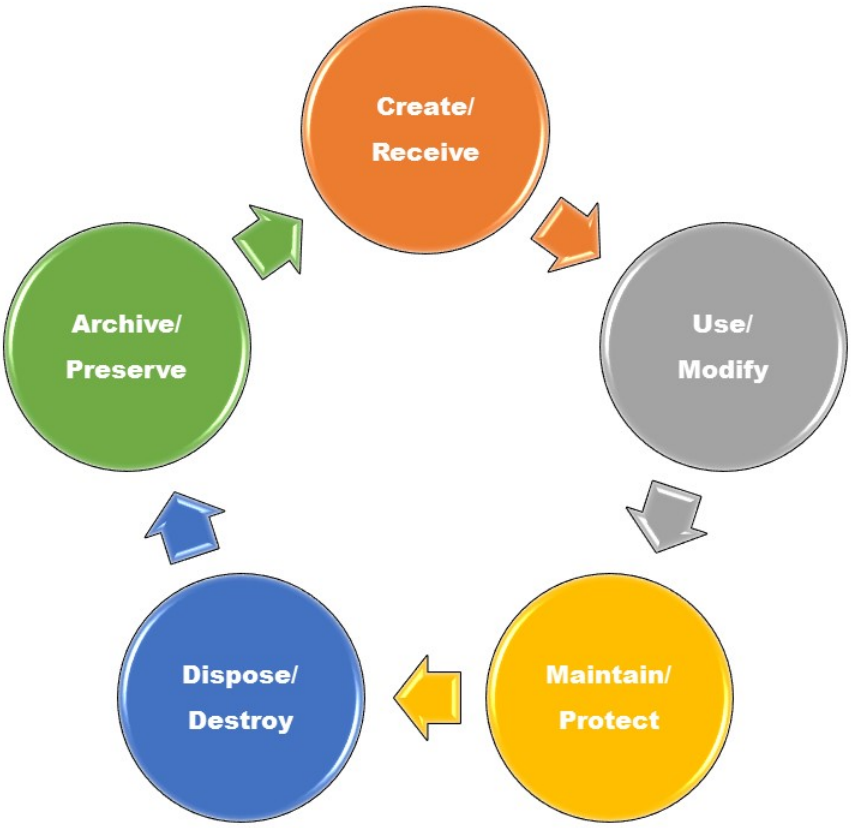


Figure 2.4 Records Life Cycle Management

Source: https://www.smartsheet.com/record-management

2.9Conceptual model

A conceptual model provides a working strategy, a scheme containing general, major and their interrelations. It orients towards specific sets of research questions and provides a guide for the researcher.

The conceptual model,figure 2.5a,assumesa specific time-to-dormancy between creation and dormancy for a patient record and also a relationshipbetweenindependent variables, patient characteristics(demographic, clinical and other factors) and medical record dormancy time. Intuitively, the value of any information (patient information inclusive) is a function of the frequency of use of such information over time, and can be describe by the ratio

$$\textit{The value of information} = f\left(\frac{u}{t}\right) = d$$

Where u is the frequency of use over time t, and t can be expressed in a unit of time say 6 months.

Where information is not used it is assumed to become inactive or dormant, and this dormancy time need to be quantified. Until now, serious deficiency in the records life cycle model is the failure to quantify in terms of survivorship time (time-to-dormancy) from the point of creation to death, that is, the life expectancy of a record, a limitation in records management.The study determined the statistical distribution, estimated the parameters of the dormancy time (time between creation and inactive), and the statistical model that best predicts factors associated with dormancy time for medical records of patients created in UCH, Ibadan. Findings is expected to guide the hospital management develop a retention policy for safe weeding ofdormant (inactive) records from the filing system. Figure 2.5b,show the linear concept of the study.

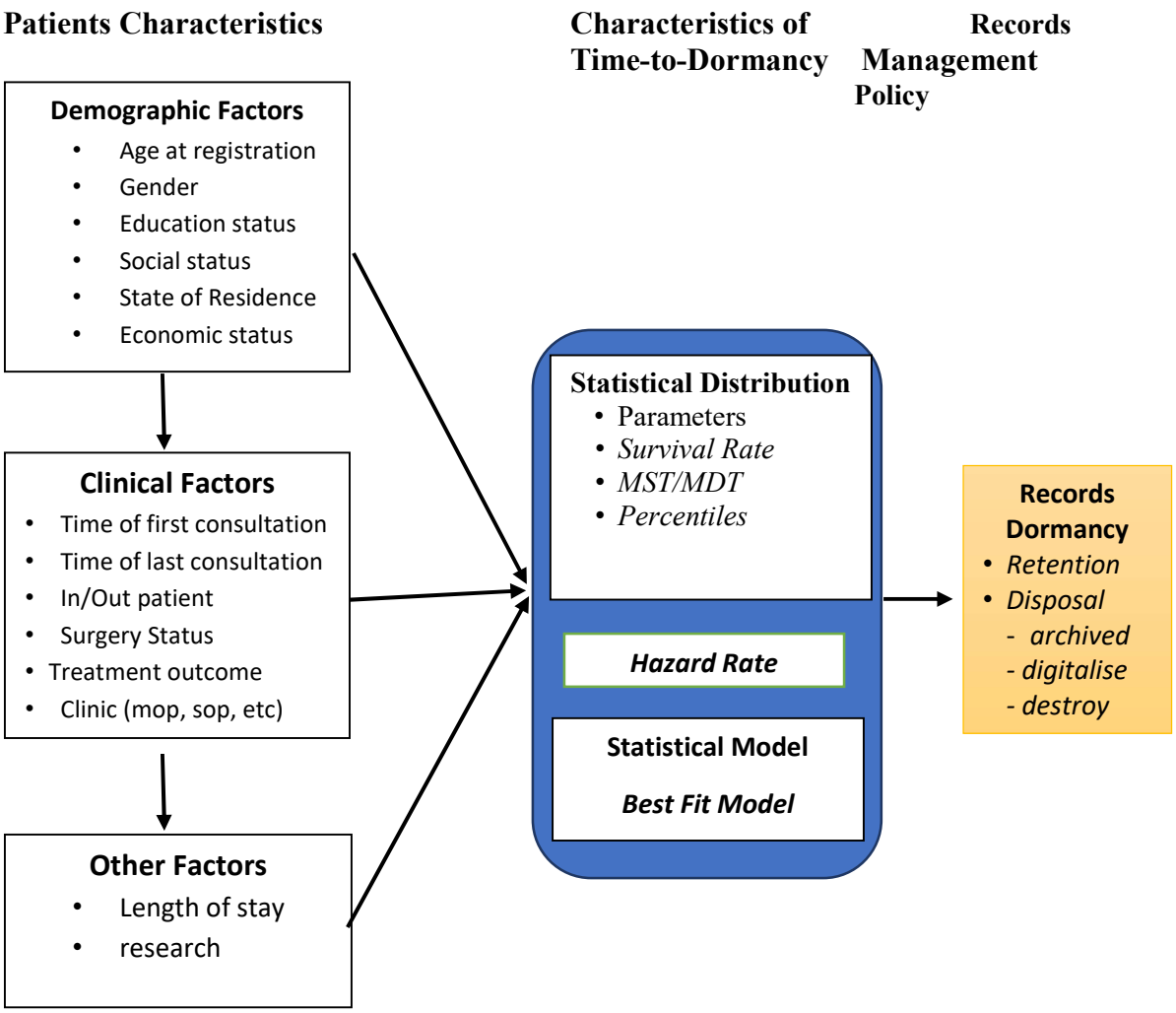


Figure 2.5a: Conceptual model for time-to-dormancy of patient's records
Source: Self developed conceptual model

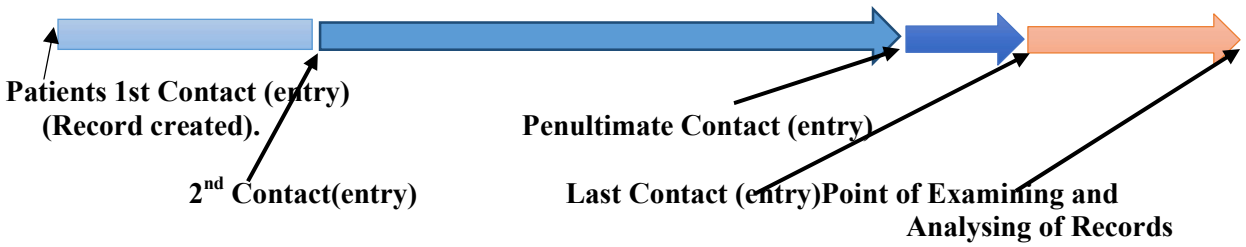


Figure 2.5b: Linear concept of conceptual model fortime-to-dormancy of patient records

Source: Self Developed Model

CHAPTER THREE

METHODOLOGY

3.1 Study Setting

The study site was the Health Records Department of the University College Hospital, Ibadan, the first teaching hospital in Nigeria, established in 1957 by the University College Hospital, London as a teaching hospital of the UniversityCollege, Ibadan, to be later known as the University of Ibadan. The Health Records Department of the hospital was established at the inception of the hospital to initiate, maintain, store and retrieve patients'medical records. This process of managing patient's records are undertaken by records management practitioners who are trained in the art of patient records management. Information on patients care are created and transmitted by longhand and processed manually to produce a paper-based patient medical/health records system over the years.

Nigeria operates a three tier healthcare system and creating medical record for a patient on first contact to a teaching (tertiary) hospital should be strictly onreferral from a secondary level health institution. However occasions may warrantregistering a patient from other sources and a record created without been referredbased on the assessment of a Consultant, especiallyif the case is of interest to research. The process of creating a record for a patient on first contact begins with allocation of a unique hospital number and a folder. This number serves as an identification toolthat appears on all documents relating to the particular patient; with this system the patient has only one folderthat holds all record of information relating to all activities during contacts. In UCH, Ibadan, records are filed centrally in the medical records library. In patient care management where every minute counts,proper records management is essential for prompt retrieval of patient information coupled withadequatemaintenance for patient's confidentiality.Therefore it is important that medical records practitionershave a goodknowledge of the principles ofrecords management practice.

A mandatory requirement in medical practice is that all documented entries in the medical records must be dated with time and signed, therefore every patient contact is indicated by an entry that is signed and dated with time. Conventionally, the date of the first entry therefore indicates when a record was created and the last entry suggest when the record was last used for the patient.

At the end of every use each record is put away on a filing shelfand arrangement follows a particular filing system determined by the adapted numbering system used by the medical records department.

In the University College Hospital, Ibadan, the unitary numbering system of medical record is operated.Patient'shospital numbers are generated through a patient number register. At first contact the patient is allocated the next unused number from the register and this number appears on all subsequent documentsrelating to the patient. Though patient medical records are created centrally, a patient's records is kept in a mini-library attached to each clinic. At the end of each use each patient medical record is filed away serially on a filing shelf located in the clinic. In addition to this mini-medical record libraries attached to each clinic, there is a Central Medical Records Library where records are filed serially on filing shelves. Observation however revealed that there are no rules or policy as to which records should be in the mini- or the central library. The result is medical records over-flowing the filing shelves in both the clinics and the central libraries, resulting to waste of resources in records retrieval time, a practice inimical to good patient care service.

Figure 3.1: The University College Hospital Ibadan, Nigeria

3.3 Research Design

This is a retrospective review of medical records of patients seen in UCH, Ibadan, Nigeria, from January 1990 to December 2014.

3.3 Study population and data source

The population for the study werepatient records created in the University College Hospital, Ibadan,Nigeria between 1st January 1990 and 31stDecember 2014. The Unit Patient Register maintained in the Medical RecordsDepartment, University College Hospital. Ibadan, revealed that **478,300** medical records were created between1st January 1990 and 31st December 2014.The study period 1990-2014 was divided into five consecutive intervals, 1990-1994, 1995-1999, 2000-2004, 2005-2009 and 2010-2014 to form five cohortsTable 3.1. Samples were selected from each of the cohorts for the study. The idea was to also compare dormancy pattern over time so as to find out if there would be changes in dormancy over time.

Table 3.1 Records of patient created in each of the five cohorts in UCH, Ibadan.

SN	Cohort Years	No. of Records created percohort
i.	1990 - 1994	84613
ii.	1995 - 2009	79417
iii.	2010 - 2004	87902
iv.	2005 - 2009	117384
v.	2010 - 2014	108984
Total		**478300**

Source: Medical Records Department, UCH Ibadan

3.4Sample Size determination and Sampling Methods

3.4.1 Estimation of Required Sample Size

The estimation of sample size for the study was approached as follows:

Ifit is assumed that the median duration of time-to-dormancy for a group of patients' record is τ_1 days, how many such patients' records must be studied to enable the estimation of τ_1 to within 5% of its true value with 95% confidence?

Thus the attention of the study is primarily on the distribution of time-to-dormancy of each patient's record and may be approached by classical survival analysis method. The focus then is the assessment of the hazard rate, λ, of the process for the estimation of the required sample size instead of the more complicated median survival time.

If the distribution of dormancy times in the population of the patient's records is approximately exponential, a sample size calculation by setting precision conditions for λ, the hazard rate, will also be adequate for the estimation of the median time-to-dormancy. This is because,for an exponential distribution, the distribution of survival times, and the median duration of survival time are directly obtainable from the hazard rate of the process.

The sample size estimation can now be determined as follows:

If the hazard rate of the process is λ, how many patients' records should be selected and followed-up for study to enable the estimation of λ, to 5% of its true value with 95% confidence.

As proposed by Lemeshow et al (1990), (see Appendix 2), let n be the required number of patients' records then,

$$n = \left[\frac{z_{1-\alpha/2}}{\varepsilon}\right]^2 \qquad \ldots 3.1$$

where: $Z_{1-\alpha/2}$= 1.96, the standardized z-value for α at 0.05; and

ε = 0.05 which is the error margin.

It follows that $n=\left[\frac{1.96}{0.05}\right]^2 = 1536.64 \cong 1537\ patients$ ...3.2

Thus, a sample size of 1,537 patients' records is statistically appropriate for follow-up in this study and for the five cohorts the sample size totalled to 7685.

3.4.2 Sampling Method

The systematic sampling techniquewas used to select1,537 patients' records from each of the five cohorts giving a total sample size of 7685 medical records for the study.Observation show that patients are randomly registered and allocated a unit number to their record with which medical records are filed in strict numerical order on the filing shelves. The number of records created varies with the years and hence within cohorts, as a result varying selection interval, *k,*was used in selecting medical records from each cohort,see Table 3.2. The study started on the 1st of July 2017 with the pulling of medical records and using selection interval, *k,* determined by the formula where *k* is:

$$Selection \quad Interval = \frac{\text{number of patient records created in cohort}}{\text{sample size}} = k$$

Table 3.2: Records created and sample sizeselected per Cohorts

SN	Cohort Years 1	No. of Records created percohort(y_i) 2	Sample Size*per cohort (n)* 3	Selection interval "*k*" (2/3) 4
i.	1990 - 1994	84613	1537	55
ii.	1995 - 2009	79417	1537	56
iii.	2010 - 2004	87902	1537	57
iv.	2005 - 2009	117384	1537	76
v.	2010 - 2014	108984	1537	71
Total		**478300**	**7685**	

3.5 Data Collection Instrument

Data was collected using a self-developed data extraction proforma (appendix 1). These patients' data are the demographic and clinical information recorded at time of first contact and subsequent visits of patient to the hospital. The following 13 variables were extracted from the patients' record:

- date first contact,
- date of last contact,
- Penultimate Contact Date,
- State of residence,
- gender,
- date of birth,
- age at first contact,
- procedure if any,
- whether were admitted or not,
- number of admission,
- length of stay,
- clinic attended,
- outcome of treatment (alive, Discharged Against Medical Advice (DAMA), died, referred).

3.6Inclusion and Exclusion Criteria of Patient Medical Records

Inclusion Criteria

The following medical records wereeligible for inclusion in the study:

i. records created between1st January 1990 and 31stDecember 2014;
ii. evidence of consultation with a doctor indicated by entry in the medical record;

Exclusion Criteria

The following medical records were excluded from the study:

i. temporary records, except where the original records are located and merged or date of initial creation established;
ii. records with one or more missing variable(s) indicated in (3.5) were excluded but replaced with a record with the next serial number.

3.7Data Collection Procedure

3.7.1 Training of data Extractors

Data extractorswere employed and trained to select patients' medical records from filing shelves using systematic sampling method. Patient records that do not meet the inclusion criteria were replaced with the next record so as to attain the required sample size. From each selected patient's record, relevant information were extracted using a specially designed data recording form, Appendix 1, administered by trained data extractors. This required each data extractor carefully going through each patient's record to extract the required information. The process was closely monitored and supervised by the researcher to ensure strict compliance.

3.7.2 Data Management

Data extracted from the patients' medical records were entered into the computer using the data entry software package of SPSS version 20.0. The data was verified and cleaned; using frequency counts and range checks to detect gross errors and outliers. Finally the data were analysed using STATA version 12

3.8List of variables and terms used in the study

Admission: a state of a patient having to occupy a bed for 24 hours and over within the hospital in an area provided for hospital care

Age at first contact: the age of patients at registration was derived from the *"date of registration – date of birth*". This was categorised into the following:

< 10	Children
10 - 20	Adolescents
21 – 30	Youths
31 – 60	Adults
61 +	Older adult

Clinic: a consultative outpatient units where patients receive care other than the GOPD

Clinical factors: admission status, number of admission, surgical if any, length of stay, clinic attended, treatment outcome.

Date of birth: date provided at registration as the date the patient was born.

Date of first contact: the date a patient was first registered for consultation in the hospital indicated by the creation of a medical record.

Date of last contact:date of last consultation as indicated in the medical record of the patient.

Date of penultimate contact: the date of the last but one visit to the hospital as indicated in themedical record;

Dead records: this are records that are due to their state of inactiveness and can be conveniently weeded off the shelves.

Demographic factors: gender, age at first contact, date of first contact, date of birth date, of last contact, date of penultimate contact, state of residence.

Dependent variables: the dependent variable for the multiple regression analysis was the hazard of medical records dormancy at UCH, Ibadan

Gender: this would be male or female

Independent variables: these are the selected demographic and clinicalfactors documented during first contact when a record is created.

Length of stay: the period over which a patient occupies a bed as an in-patient

Median Dormancy Time (MDT): is the time at which 50 percent of medical records become inactive and can be conveniently weeded off the shelves.

Number of admissions:the number of admission episodes for a patient in the hospital

Outcome of treatment: the condition under which a patient was at the time/point data was collected for this study. The patient is either alive, dead, referred to

another hospital or discharged against medical advice (DAMA) that is the patient decided on own volition to take leave from the hospital.

State of residence: the State of abode of a patient at the time of registration for consultation in the hospital

Surgery if any: any procedure that involved surgical operation or manipulation by Surgeons

SurvivalPercentile: The p^{th} survival percentile is the time t by which p% of patient medical records had experienced dormancy, while (100− P) % have not.

3.9 Statistical Analysis

Analyses were done in line with the study aims and objectives stated in sections 1.3 and 1.4 as follows:

Descriptive distribution of patient characteristics

Frequency distribution of patient demographic and clinical characteristics, (age, gender, clinic, state of residence, admission and surgery status and patient outcome) were presented for each cohorts and for all cohorts merged together. This was done to show the pattern of patients seen during the study period at UCH, Ibadan. Patient's age was categorised as:

< 10children
10 – 20 adolescents
21- 30 youths
31-60 adults
61 + older Adults.

State of residence was categorised into whether the patient resides in Oyo State or in other States, if patient was ever admitted or not, if patient had ever been operated on, and outcome of patient was categorised as alive at time of last contact, died, discharged against medical advice or referred to another hospital.

3.9.1 Determination of the general form and the distributionof dormancy time

For each record the dormancy time was calculated as:

Dormancy time = date of last entry – date of first entry (date of opening record).

The dormancy time of a record was censored if the time between the penultimate entry (visit) and the last entry (visit) was greater than the time between the last entry (visit) and the date of analysis.

i.e. *Censor record if: (date* of *last entry – date of penultimate entry) >(date of analysis – date of last entry).*

This is to prevent a premature assessment of the dormancy of the records.

A frequency distribution of dormancy time was done and presented in both tabular and graphical forms.

3.9.2 Estimation of Survival time, $\hat{S}(t)$

The survival function *S(t)* is the probability that an individual medical record remains active from the time of creation to sometime beyond time *t.*

From the frequency table, Survival functions were calculated fordormancy time using the K-M approach. The K-M estimator defined for all values of *t* in the rangeis definedas:

$$\hat{S}(t) = \begin{cases} 1 & if \quad t < t_1 \\ \prod_{t_i \le t} \left(1 - \frac{d_j}{n_j}\right) & if \quad t_1 < t \end{cases}$$

where t_i denotes the first observed dormancy time, d_i represents the number of record dormancy at time *t*, and n_i indicates the number of recordsthat had not experienced dormancy, and have also not been censored, by time *i.*

The Kaplan-Meier estimator of the survival function *S(t)* can then be defined as:

$$\hat{S}(t) = \prod_{t_s \le t}\left(1 - \frac{d_j}{n_j}\right)$$

$$= \prod_{t_j} \frac{n_j - dj}{n_j} \qquad \text{for} \quad 0 \le t \le \text{t} \qquad \ldots 3.3$$

where d_j is the number of recordsthat experience the event at time $t_{(j)}$, and $n_{(j)}$ is the number of recordsthat had not yet experienced the event at that time and are therefore still at risk for experiencing it, (Akram, et al, 2007; Zhao, 2008). The K-M survival curve was then plotted to examine the distribution of the data.

3.9.3 Estimation of percentiles of dormancy time

The p^{th} survival percentile is the time *t* by which p% of patient medical records had experienced dormancy, while (100− P) % have not.

The median dormancy time defined as the 50th percentile would be when 50% ofpatient records experienced dormancy.

The Kaplan-Meier Product Limit Method was employed to estimate the 25th, 50th, 75th and 95th survival percentiles along with their standard error and 95% Confidence Intervals for dormancy time of medical records created for patients at UCH, Ibadan.

3.9.4Estimation of the hazard rates, λ (t)

The hazard rate, **λ (t),** at time *t* for the study is the instantaneous dormancy rate among the records at that time.

The hazard rate has the following relationship with the *f(t)* and *S(t):*

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}. \qquad \ldots 3.4$$

or defined in terms of the cumulative distribution *F(t)*and probability density function as:

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t)) \qquad \ldots 3.5$$

3.9.5Hazard plotting

Hazard plotting (Nelson 1972, 1982) is analogous to probability plotting except that survival time, t, is plotted against the hazard function, $\hat{\lambda}(t)$, rather than the distribution function.To determine the form and shape of the hazard ratesthe hazard plot was constructed byplotting the estimated hazard, $\hat{\lambda}(t)$ against age of record *t.*The plot would suggest whether the hazard rate is constant, as with Exponential distribution,

increasing or decreasing overtime, has a bathtub shape as with Weibull distribution, or some other shape.

Secondly to determine if the dormancy time of medical records was from a particular theoretical distribution, the log of the Cumulative Hazard i.e. the log–log of the survival time, (*log(-log of S(t)*), was plotted against the log of time, *log(t),* for test of linearity and the slope of the plot was estimated to determine the parameter of the distribution. Thus:

$$\lambda(t) = \lambda p t^{p-1}$$

and p and $\lambda > 0$

This linearity of ln(t) of $S(t) = \exp(-N^p)$

$$\Rightarrow \ln[\ln S(t)] = \ln(\lambda) + p\ln(t)$$

Where the intercept = ln(**λ**), and slope = p

- If $p > 1$hazard increases over time
- $P = 1$ hazard is constant (the Weibull model reduces to exponential model **λ**(t) =1)
- $P < 1$= hazard decreases over time

This was done as a diagnostic test in assessing the validity for Weibull and Exponential distribution.

3.10 Comparing survival distributionsand their parameters for the five cohorts

To find out if the form and shape of the distribution of dormancy time of patient medical records created, 1990-2014 at UCH, Ibadan, are same, the survival distribution, their parameters and the hazard functions for the cohorts were compared using statistical diagnostic tests.

Both graphical and nonparametric test were used to compare time-to-dormancy of patients records created between 1st January, 1990 and31st December, 2014 in UCH, Ibadan. TheKaplan-Meier Product Limit Method (K-M) plot and hazard curve was plotted for each cohort to find out if the statistical distribution, the form and shape of the hazard function and the model that best fits time-to-dormancyof medical records were

same over the period of the study. The Kaplan-Meier log of the Cumulative Hazard*logH(t),* was plotted against the log of time, *log(t),* for the five cohorts and the merged data and these were compared for linearity, estimated values of shape parameter were interpolated from the intercept of the straight-line of the plots were compared.

The log-rank test of equality was used to assessthe differences in survivorship between the Kaplan-Meier survival curves for the five cohorts. Also, the Log-rank test of trend was used to assess the differences in survival between cohorts under the assumption that the record dormancy time data were in a naturally ordered sequence. The form of the test statistics used was;

$$\chi^2 = \sum \frac{\left(\sum O_{jt} - \sum E_{jt} \right)^2}{\sum E_{jt}} ,$$

Where $\sum O_{jt}$ represents the sum of the observed number of dormant records in the j^{th} dormancy timewith (g-1) degrees of freedom, where g equals 5, the number of cohorts in the study.

3.11 Modeling time-to-dormancy of patient medical records

Modelling the time-to-dormancy of medical records was done to determine which combination of documented demographic and clinical factors influenced dormancy time of patient records, and to obtain estimate of the level of hazard of records dormancy.

The analysis hereinvolved the use of semi-parametric (Cox proportional hazard model) and parametric models (Exponential and Weibull models) to explore how dormancy of patient records are influenced by demographic and clinical factors. Diagnostic test was further conducted to find out the model that best fit dormancy time data of patient records.

3.11.1Test of Cox Proportional Hazard Model Assumption

The use of the semi-parametric Cox proportional hazards regression model, was to avoid having to specify the hazard function completely. The utility of the proportional hazards model stems from the fact that a reduced set of assumptions is needed to provide hazard ratios that are easily interpreted and clinically meaningful.

Schoenfeld's global test was used to test the validity of the Cox's proportional hazard model assumption.This global goodness-of-fit test proposed by (Schoenfeld, 1980) was considered for testing the Cox PH assumption of the time-to-dormancy of medical records, because of its power to detect the insufficiency of covariates in describing the relative risks and the assumption of PH, when applied to the fitted model.With the global statistical significance of the model,output gives p-values for three alternative tests for overall significance of the model: The likelihood-ratio test, Wald test, and score log-rank statistics. These three methods are asymptotically equivalent. For large enough N, like in this study, they will give similar results. For small N, they may differ somewhat.

The log-log plots which are plots of $\log(-\log(S(t)))$ vs. $\log(t)$, where t = dormancy timewas used to evaluate the result of the test. This is a graphical test and an initial indication of failure of this assumption is when the survival curves under consideration cross and diverge.When these plots show a non-parallel pattern, the proportional hazards assumption is said to be violated, (Kleinbaum and Klein, 2012).

Cox regression, Exponential andWeibull regression models were fittedtotime-to-dormancyof medical records to identify independent factors associated withdormancyof medical records of patients created 1990-2014 at UCH, Ibadan.

3.11.2Cox Regression Modeling

Cox regression model, is a semi-parametric regression models that examines the relationship between independent variables with failure time (survival time) and estimated regression coefficients, as well as estimate hazard ratio (HR) of two individuals with different covariates. The major intend in fitting the Cox hazard model to time-to-dormancy of medical records was to determine the suitability of the model that best fit dormancy time for patient medical records.Because the model ability to evaluate simultaneously the effect of several factors on survival, it was used toinvestigate the effect of patient's demographics and clinical (explanatory) variables upon dormancy time of patient records.

The Cox model was expressed by the hazard function $\lambda(t)$, and interpreted as the risk of a patient record going into dormancy at time t. The cox model used in this study is of the form:

Let $X_i = \{X_{i1}...X_{ik}\}$ be the covariates for subject, the model is of the form:

$$h(t) = h_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k\}, \quad \text{... 3.6}$$

and can be expressed in the form

$$\log\left\{\frac{h_1(t)}{h_0(t)}\right\} = \{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k\} \quad \text{... 3.7}$$

Where t represent the dormancy time, *h(t)* is the expected hazard at dormancy time *t*, for a subject with explanatory values x_1,...., x_k and

x_1 = age of patient

x_2 = gender

x_3 = state of residence

x_4 = clinics

x_5 = patients status

x_6 = surgery status

x_7 = treatment outcome

and $h_0(t)$ is the baseline hazard that represents the hazard when all predictors

$x_1, x_2, ...x_7$ are equal to zero.

The assumption here is that the hazard is constant over time, or equivalently, and that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time. It therefore follows that the study assumed that:

i. the explanatory variable only changes the chance of failure and not the timing of periods of high hazard;
ii. the explanatory variable acts directly on the baseline hazard function and not on the failure time, and remains constant over time; or

iii. no assumption of any particular form of probability distribution for survival times. The Cox Model is different from ordinary regression in that the covariates are used to predict the hazard function, and not *Y* itself.

3.11.3 Parametric Survival Analysis Models

The study fitted the Exponential and Weibull models to dormancy time data of patient medical records to explore the influence of demographic and clinical factors would have on dormancy of patient medical records. With the parametric models, the outcome is assumed to follow a certain known distribution,(Cox, 1992; Buis, 2006); and can almost have the look and feel of a normal-errors linear regression analysis, (Kargarian-Marvasti, *et al*, 2017).

3.11.4Exponential Modeling

Given that the record dormancy time data is skewed distributed data, the exponential model was regressed on dormancy time on patients characteristics based on exponential model assumption of parameter λ=1. A one parameter λ time-independent hazard rate and because of its simplicity the exponential model is one of the most usedparametric distributions for time-to-event data,(Kalbfleisch and Prentice 1980; Collet 2003, Montaseri, et al 2016), The key property for the Exponential distribution is that the hazard is constant over time (not just the ratio) and can be run as a PH model or an AFT model, (Kleinbaum and Klein, 2012). The study used Exponential model of the form:

$$\log h_i(t) = \alpha(+\beta_1 x_{i1} + \beta_{i2} + ... + \beta_k x_{ik}) \quad \ldots 3.8$$

where the constant α represents the log-baseline hazard $h_0(t)$ when all the x's are zero, therefore equation 3.8 can be rewritten as

$$\log\left(\frac{h_i(t)}{h_0(t)}\right) = (+\beta_1 x_{i1} + \beta_{i2} + ... + \beta_k x_{ik}) \quad \ldots 3.9$$

and $x_1, x_2, x_3, ...x_k$ arethe explanatory variables where

x_1 = age

x_2 = gender

x_3= state of residence

x_4 = clinics

x_5= patients status

x_6 = surgery status

x_7 = treatment outcome

3.11.5 Weibull Modeling

Weibull distribution is unique for being a PH and AFT model.The two-parameter Weibull distribution has been described as one of the most widely applied probability distributions, particularly in modelings time-to-event data and correct estimation of the shape parameter of the Weibull distribution had placed a central role in the areas of statistical analysis and modeling, (Altin, 2013). An added advantage is that many different methods can be used to estimate this parameter, most of which utilise regression methods. Weibull distribution is very flexible and powerful and can model different types of failure times and the exponential distribution had been described as a special case of the Weibull distribution.The Weibull model was also fitted to the dormancy time data of patient records to explore the contributions of explanatory variables to dormancy time of patient records.

The Weibull model used in this study is of the form:

$$h_i(t) = \exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}\right) h_0(t), \qquad .. 3.10$$

for i = 1, 2, …, n. and this can be written in the form

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}, \qquad \ldots 3.11$$

Where $x_1, x_2, x_3, ...x_k$ are the explanatory variables and

x_1 = age

x_2 = gender

x_3 = state of residence

x_4 = clinics

x_5 = patients status

x_6 = surgery status

x_7 = treatment outcome

$h_0(t)$is the baseline hazard function, i.e., hazard function when all covariates equal zero. expis the regression coefficient for the i^{th}covariate and x_i the i^{th}covariate in the model, and $\beta_i$is the regression coefficient for the i^{th}covariate.

3.11.6Model selection criteria

Comparison was done among the three sets (Cox proportional hazard, Exponential and Weibull) of survival model that best fitted the records dormancy time data for each cohort.

Semi-parametric Cox regression and parametric Exponential and Weibull models have been used variously to analyze survival data; however, no study has focused on the comparison of survival models in dormancy association analysis of patient records.

The Log likelihood and Akaike Information Criterion were used to compare the Cox models and Exponential and Weibull models. The model with the minimum log likelihood and equivalently minimize the information lost (from the AIC value) was adjudged as the best model for each record dormancy time data among the five (5) cohort and all the five models combined.

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages over fitting, because increasing the number of parameters in the model almost always improves the goodness of the fit.

3.12Analysis of the One-day-Active records

A preliminary investigation was carried out by studying 1020 records selected at random from the filing cabinets of the Medical Record Department, UCH, Ibadan. The aim was to be acquainted with the type of information available in the files and its completeness.A brief analysis show that about 31.5% of the records had only one entry in them which was made on the day the record was created and no other entry, this showed that such records were active just for the day of their creation. This feature were to later found to manifest in all the five cohorts of the study necessitating attention.

Such records have been named "one-day-active records" in the study and have been excluded from the main survival analysis. They had a separate analysis aimed at their early identification in the process. The analysis included extraction of their International Classification of Diseases (ICD) codes; the use of key informants to investigate reasons for clinic attendance and the estimation of the timing of the second visit after the first visit froma subsample of patients continuing beyond the first visit to serve as indicator of due date for the second visit.

3.12.1 Estimate of time between first and second entry for records that survived beyond one day

The aim was to provide some information on how to recognise those who would probably not make a second visits after the first visit.This information was to serve as an indicator for weeding one-day-active records at appropriate time.To estimate the time between first entry (contact) and second entry (contact),a sub-sample using multiphase sampling technique, 150 records that survived beyond one day was selected from previously observed records for time-to-dormancy. The date of first and second entry were extracted and the 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles estimated for average time it takes a patients to return for a second visit. This was to provide useful information for recognising record of patients that are most likely to fail after the first contact.

International Classification of Diseases (ICD) codes

The International Classification of Diseasesproduced by the World Health Organisation (WHO) is the global choice classification of health conditions and used by member nations of the WHO for collecting and reporting statistics on hospitals morbidity and mortality at both the local and national level. Diagnoses extracted from the one-day-active records were coded with aid of the ICD-10 to find out the pattern of the disease condition.

3.12.2Key Informants interview

A Key Informants interview was conducted for Doctors, Nurses, Medical Records Officers and Patients. Five each of the key informants were selected at random and asked the following questions;

i. If they were aware that some patients may not return for a 2^{nd} visit after the 1^{st}?
ii. What could be responsible for a patient to decide not to return for a 2^{nd} visit after the first, despite being given appointment?

Result from the interview was analysed and findings would serve as indicator to causes of one-day-active records.

3.13 Validating dormancy time estimates of record dormancy time

To test the validity of estimated dormancy time, the study examined time difference between*penultimate entry-last entry time* and *last contact time - point of data analysis*of patient records seen between 2010 and 2014, Cohort 5, at the University College Hospital, Ibadan Southwest Nigeria. Intuitively the *"penultimate contact time-last contact time*" is the time it takes a patient to return for final check-up (last follow-up time) after the penultimate contact and the "*last contact time - point of data analysis"* is the period over which the medical records remain dormant after the last entry/contact.
Underestimation may result from lower survival-estimate if:

(Penultimate - Last-contact) > (Last contact - Point of data analysis),

the result would be that the record are censored and the study was carried out too early. However if:

(Penultimate - Last-contact) < (Last contact - Point of data analysis),

Then the estimation of time-to-dormancy, which is the time from record creation to point of dormancy is valid.

Survival estimate at 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles with their SE and CI wereestimated for underestimation. The distribution, survival and hazard plot of the two groups of survival time difference were plotted and survival curve compare using log rank test.

3.14Ethical approval

Ethical approval to conduct the study wasfirst obtained fromtheInstitutional EthicalReview Committee of University of Ibadan/University College Hospital, Ibadan, Nigeria, (approval protocol number NHREC/TR/02/06/2007a, dated Friday, May 12, 2017).A second approval was obtained from the management of the University College Hospital, Ibadan (approval letter dated June 14, 2018) to have access to patient medical records created 1990-2014, see appendix II and III.

CHAPTER FOUR
RESULTS

4.0 Introduction

The results of the analysis for this study are presented as follows:

i. records with 1 day dormancy time;
ii. separately for each of the cohorts, records surviving beyond first day of creation;
iii. all the five cohorts combined as a single sample;
iv. Diagnostic tests

The event of interest was on the time-to-dormancy of a medical records, that is, the active life, or survival time of the record. Record of some patients were however not used beyond the first day of creation or the patient stopped coming after the first contact, (one-day-active records).

4.1 Indication from preliminary pilot survey

Table 4.1 show the frequency distribution of the dormancy times of the 1020 records examined in the preliminary investigation pilot study. Close to one third (31.5%) of the records were active for one day, thatis,such records wereonly usedon one day after creation; this was established by a single entry in the record.

Excluding the one-day-active records, further analysis revealed that about 76% of the records were inactive (dormant) in 33.5 months of creation and close to 95% of the records became dormant in 147.5 months after creation. Other dormancy points can be seen on the table.

Table 4.1. Distribution of records by dormancy time from preliminary pilot survey

Time (t)		**Dormant record**	**Cum. Freq.**	**Cum.%**
1day active	***1day***	***321***	***321***	***31.5***
months		**Dormant record**	**Cum. Freq.**	**Cum.%**
< 6	3.5	302	302	43.20
7-12	8.5	118	420	60.09
13-18	15.5	41	461	65.95
19-24	21.5	25	486	69.53
25-30	27.5	24	510	72.92
31-36	33.5	21	531	75.97
37-42	39.5	9	540	77.25
43-48	45.5	9	549	78.54
49-54	51.5	9	558	79.03
55-60	57.5	15	573	81.94
61-66	63.5	7	580	82.98
67-72	69.5	14	594	84.98

73-78	75.5	2	596	85.26
79-84	81.5	6	602	86.12
85-90	87.5	6	608	86.98
91-96	93.5	5	613	87.70
97-102	99.5	6	619	88.56
103-108	105.5	5	624	89.27
109-114	111.5	7	631	90.27
115-120	117.5	9	640	91.56
121-126	123.5	5	645	92.27
127-132	129.5	6	651	93.13
133-138	135.5	4	655	93.71
139-144	141.5	7	662	94.71
145-150	147.5	3	665	95.14
151-156	153.5	2	667	95.42
157-162	159.5	5	672	96.14
163-168	165.5	8	680	97.28
169-174	171.5	3	683	97.71
175-180	177.5	7	690	98.71
181+	183.5	9	**699**	**100**
Total		**1020**		

4.1.1 Result of Analysisof One-Day-Active Records

Distribution of records with one-day-active period

The table 4.2 show the number of records created between1st January 1990 and 31st December 2014 was 478,300.Thenumber of records created was lowest with 79,417 in 1995 – 1999 and highest with 117384 records in 2005 – 2009. The Tablealso show the frequency distribution of the one-day-active records. The number of the one-day-active records ranged between 17.8% in the 2000 – 2004 to 30.6% in the 1990 – 1994. The overall one-day-active records was 24.6% for the five cohorts,(1990 – 2014) merged.

Table 4.2 One day active records.

Cohort	Period covered	Records created (N)	records selected *(n)*	One-day-active records	%
1	1990 - 1994	84613	1537	470	30.6
2	1995 - 1999	79417	1537	354	23.0
3	2000 - 2004	87902	1537	274	17.8
4	2005 - 2009	117384	1537	460	30.0
5	2010 - 2014	108984	1537	330	21.5
Merged	**1990 - 2014**	**478300**	**7685**	**1888**	**24.6**

4.1.2 Distribution of one-day-active records by some patients characteristics

Result from Table 4.3, show that records of male patients constitute about half of the whole records and this trend was observed in all the five cohorts. Records of patients residing in Oyo State were close to half compared to all other states put together except for cohort 3 and 4. However when the cohorts were merged records of patientsresiding in Oyo State was above 50%of all records put together. The result also show that records created for patients in targeted clinics was highest in MOP for all cohorts. Only 0.3% of the records indicated that patients had ever under gone surgery. None of the one-day-active records related to admitted case.

Table 4.3One-day active records by some patient characteristics

Variables n=1888	Level	COHORTS					
		1	2	3	4	5	Combined
Gender	Male	229	165	137	240	177	935
	Female	255	182	135	213	148	919
Clinic	MOP	85	45	136	79	207	546
	SOP	141	81	1	5	2	222
	CHOP	9	16	2	1	2	30
	GYNE	92	42	-	-	-	131
	Others	164	165	132	373	110	933
State of residence	Oyo	229	157	152	338	148	1008
	Others	242	178	122	122	180	880

Ever operated on	No	467	349	274	459	330	1880
	Yes	3	2	0	1	0	6
Admission status	Yes	-	-	-	-	-	nil

4.1.3 Results from ICD codes for diagnostic of One-Day-Active records

Result fromcoding statement of diagnosisextracted from the one-day-active records, Table 4.4,using the International Classification of Diseases and Health Related Conditions (ICD-10), revealed that 72% of the conditionsare not classifiable to any of the chapters (Chapters I to Chapter XXII) of the ICD-10.Whereas28% of the conditions classifiable to ICD-10, indicated that malariaconstituted 8%, conditions of the eyes14.1 %, road traffic accidents. 8%., diseases of the skin and subcutaneous tissue 8.1% among others.Other cases were 62.1%.

Table 4.4 ICD-10 Codes of Diagnosis for One-Day-Active records

Chapter No.	Chapter Title	No of Cases	REMARK
Chapter I	Certaininfectious and parasitic diseases (A00-B99)	31	Malaria 9
Chapter II	Neoplasms (C00-D48)	37	
Chapter III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)	47	
Chapter IV	Endocrine, nutritional and metabolic diseases (E00-E90)	11	
Chapter V	Mental and behavioral disorders (F00-F99)	5	
Chapter VI	Diseases of the nervous system (G00-G99)	14	
Chapter VII	Diseases of the eye and adnexa (H00-H59)	75	Conjunctivitis 18 cataract 13
Chapter VIII	Diseases of the ear and mastoid process (H60-H95)	15	
Chapter IX	Diseases of the circulatory system (I00-I99)	10	

Chapter X	Diseases of the respiratory system (J00-J99)	14	
Chapter XI	Diseases of the digestive system (K00-K93)	15	
Chapter XII	Diseases of the skin and subcutaneous tissue (L00-L99)	43	
Chapter XIII	Diseases of the musculoskeletal system and connective tissue (M00-M99)	16	
Chapter XIV	Diseases of the genitourinary system (N00-N99)	30	
Chapter XV	Pregnancy, childbirth and the puerperium (O00-O99)	5	
Chapter XVI	Certain conditions originating in the perinatal period (P00-P96)	8	
Chapter XVII	Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)	13	
Chapter XVIII	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	42	
Chapter XIX	Injury, poisoning and certain other consequences of external causes (S00-T98)	43*	Mostly RTA
Chapter XX	External causes of morbidity and mortality (V01-Y98)	14	
Chapter XXI	Factors influencing health status and contact with health services (Z00-Z99)	29	
Chapter XXII	Codes for special purposes (U00-U85)	0	
Total		**532**	

See appendix 3 for list of conditions

4.1.4 Key Informant Interview (KII) on patients fail 2nd visit

Table 4.5, revealed reasons advanced in the Key Informants Interview conducted to find out why there were no second entry in the one-day-active, or why would a patients decide not to come back after the first contact. All respondents were aware that many patients do not return for a second visit. Common reasons advancedwere that: mostof the patients with only one contact were referred for investigations from other hospitals; relatives of patients invited fordiagnostic screening e.g. Cataract and such person may not need to come back; patients treated for minor conditions that do not require follow-ups. Other reasons advanced were person that come for request for vision test to obtain driving license or eye glasses,cases as body pains or fever that do not require the patient being given appointments..

Table 4.5 Response from Key Informant Interview

SN	Group	Response
1	Doctors	i. patients referred for investigations from other institution ii. treatment of minor aliments for friends/relatives of staff among iii. stress from series of tests iv. high cost of medical care in UCH, Ibadan v. hospital policy not to turn patient back vi. preference for traditional medicine
2	Nurses	i. patients referred for investigations from other institution ii. treatment of minor aliments for friends/relatives of staff among iii. persons invited for diagnostic screening to trace diseases iv. hospital policy to create medical records for every patients attending

3	Medical/ Health Records Officers	i. treatment of minor aliments for friends/relatives of staff among ii. stress from series of tests iii. patients referred for investigations from other institution iv. high cost of medical care v. hospital policy that all patients should have a record
4	Patients	i. patients referred for investigations from other institution ii. stress through series of tests iii. high cost of medical care iv. long waiting time

4.2 Medical records of patients that survived beyond the first day of creation

Table 4.6 show the distribution of the medical records that survived beyond the first day of creation for each of the five cohorts and all the cohorts mergedtogether as a single sample.The study revealed that of the 7685 medical records sampled 75.6% survived beyond the first day of creation, indicated by two or more entries in the record. The highest records were observed in the 3rd and 5th cohorts while the least was observed in the 1st cohort.

Table 4.6.Frequency distribution of records that survived beyond the first day of creation

Cohort	Period covered	Records selected for study	Records that Survived beyond one day	%
1	1990 - 1994	1537	1067	69.4
2	1995 - 1999	1537	1183	77.0
3	2000 - 2004	1537	1263	82.2
4	2005 - 2009	1537	1077	70.0

5	2010 - 2014	1537	1207	78.5
Combined cohort	**1990 - 2014**	**7685**	**5797**	**75.4**

4.2.1 Estimates of time between patient 1st and 2nd contacts for patients who made 2nd and subsequent visits

Table 4.7 show results of the 25th, 50th, 75th and 95th percentiles for time between first and second contactsby patients (1st and 2ndentries in the records). The study revealed that 25%, 50%, 75% and 95% of records had a second entry/contact in 0.43, 0.72, 1.37 and 5.95 months respectively.

The study therefore show that 95% of the patients whose records did not fail on the first day of creation are most likely to return for a second visit/contact in about 5.95 months.

Table 4.7 Estimate of time between 1st and 2nd contacts by patients

Estimate	Percentiles			
	25th	50th	75th	95th

t = months	**0.43**	**0.72**	**1.73**	**5.75**

4.3. Cohort 1: Patient records created from January, 1990 - December1994

Between 1st January, 1990 and 31st December, 1994, 84,613 medical records were created in UCH, Ibadan, of which 1537 was selected for the study.Having excludedthe 470 (30.6%) one-day-active records,the result of analysis of the remaining 1067patient records that survived beyond the first day of creation are presented.

4.3.1 Frequency distribution of some demographic and clinical characteristics of the patients

Table 4.8 shows patient socio-demographic and other characteristics by dormancy time. The result reveal that 35.74%,patients were between 31-60 years, 148(13.88%) were aged 10-20, 210(19.70%) were below 10 years of age and 9.66% were above 60 years. Male patients constituted 51.11%,and 489(48.51%) were resident in Oyo

State.Records from Medical Outpatient Clinics (MOP) were 22.16%, Surgery Outpatient Clinics, (SOP), 2.89%, Children Outpatient Clinic (CHOP) records were35.84% and records from other clinics constitute 25.84%. Only 31.02% of thepatients were ever admitted, while 10.40%had at one time or the other undergone surgical operation. Almost all the patients,99.62%, were alive as at last entry/contact and 1 patient was discharge against medical advice while 3 patient died.

Table 4.8 Frequency distribution of patient's characteristics 1st cohort 1990-1994

Variables n=1067	**Level**	**Frequency**	**Percent**
Age at Registration	<10	210	19.70
	10-20	148	13.88
	21-30	224	21.01
	31-60	381	35.74
	61+	103	9.66
Gender	male	530	51.11
	female	507	49.89

State of residence	Oyo State	489	48.51
	Others	519	51.49
Clinic attended	MOP	230	22.16
	SOP	259	24.95
	CHOP	30	2.89
	GYNE	147	14.16
	Others	372	25.84
Ever admitted	No	736	68.98
	Yes	331	31.02
Ever operated on	No	956	89.60
	Yes	111	10.40
Treatment outcome	Alive	1055	99.62
	Died	1	0.09
	DAMA	2	0.28
	referred	-	-

4.3.2 Frequency distribution of records by dormancy timesfor cohort 1

Table 4.9 showed the frequncy distribution of the dormancy time for the 1067 records in the studythat survived beyond the first day of creation. The median dormancy time wasless than 3.5 months.About75.0 % was dormant at t = 15.5 months and close to 95.0% of records were dormant at about the age of 153.5months. The distribution is presented graphically in Figure 4.1. The graph show the distribution is skewed to the right.

Table 4.9 Distribution of records by dormancy times 1990-1994

month t*		Dormant records	Percent	Cum Percent
<1	0.5	405	37.96	37.96
1-6	3.5	270	25.30	63.26
7-12	9.5	81	7.59	70.85
13-18	15.5	55	5.15	76.01
19-24	21.5	32	3.00	79.01
25-30	27.5	18	1.69	80.69
31-36	33.5	18	1.69	82.38
37-42	39.5	16	1.50	83.88
43-48	45.5	13	1.22	85.10
49-54	51.5	9	0.84	85.94

55-60	57.5	8	0.75	86.69
61-66	63.5	7	0.66	87.35
67-72	69.5	9	0.84	88.19
73-78	75.5	4	0.37	88.57
79-84	81.5	9	0.84	89.41
85-90	87.5	6	0.56	89.97
91-96	93.5	7	0.66	90.63
97-102	99.5	9	0.84	91.47
103-108	105.5	5	0.47	91.94
109-114	111.5	4	0.37	92.31
115-120	117.5	3	0.28	92.60
121-126	123.5	5	0.47	93.06
127-132	129.5	2	0.19	93.25
133-138	135.5	7	0.66	93.91
139-144	141.5	3	0.28	94.19
145-150	147.5	3	0.28	94.47
151-156	153.5	3	0.28	94.75
157-162	159.5	5	0.47	95.22
163-168	165.5	2	0.19	95.41
169-174	171.5	2	0.19	95.60
175-180	177.5	4	0.37	95.97
181-186	183.5	7	0.66	96.63
187-192	189.5	2	0.19	96.81
193-198	195.5	6	0.56	97.38
199-204	201.5	3	0.28	97.66
205-210	207.5	3	0.28	97.94
211-216	213.5	5	0.47	98.41
month t*		**Dormant records**	**Percent**	**Cum Percent**
217-222	219.5	2	0.19	98.59
223-228	225.5	4	0.37	98.97
229-234	231.5	1	0.09	99.06
235-240	237.5	2	0.19	99.25
241-246	243.5	2	0.19	99.44
247-252	255.5	0	0.00	99.44
253-258	261.5	1	0.09	99.53
259-264	273.5	1	0.09	99.63
265-270	279.5	1	0.09	99.72
271-276	285.5	1	0.09	99.81
277 +	291.5	2	0.19	100
Total		1067	100.00	

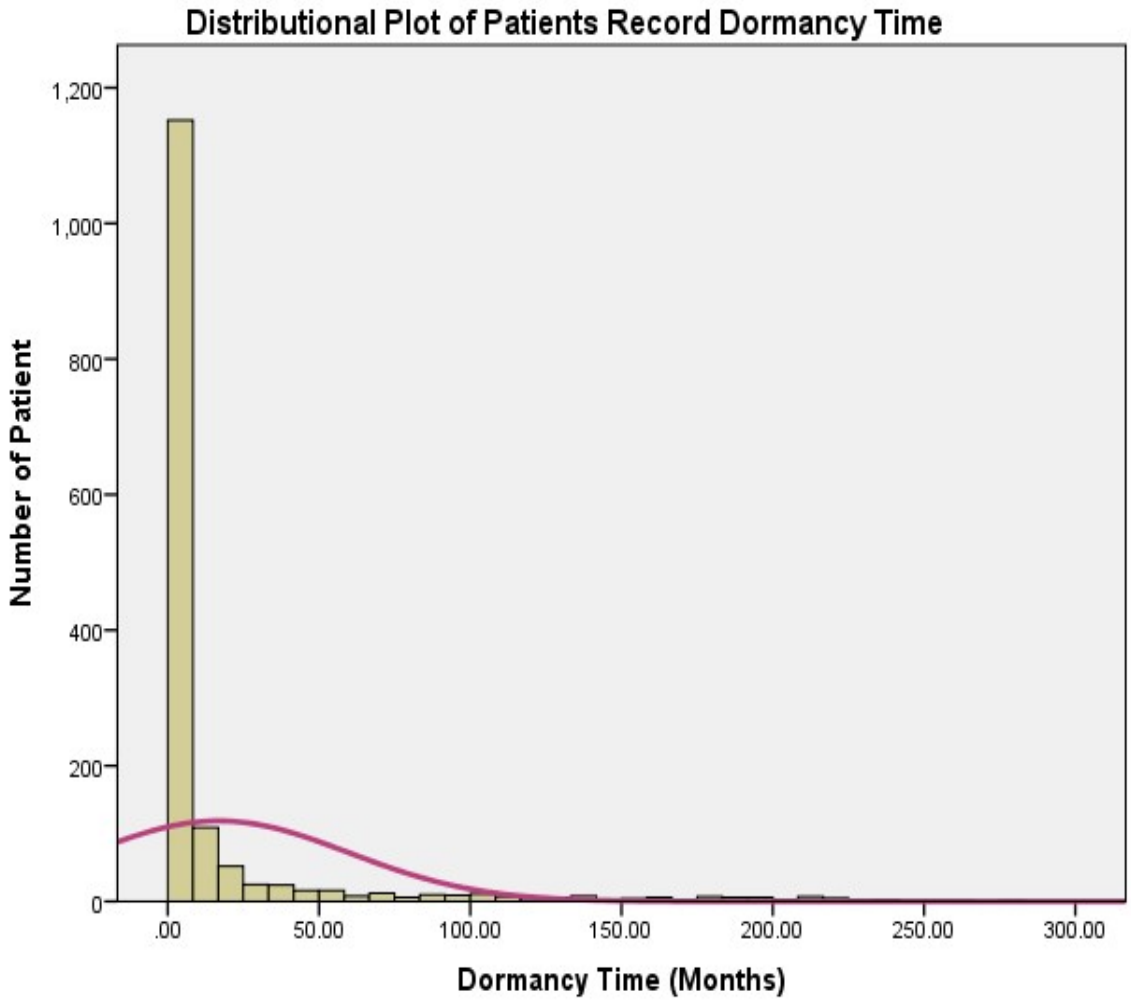


Figure 4.1 Disribution of records by dormancy time

4.3.3 Survival function of dormancy timescohort 1(1990-1994) data

Table 4.10 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals asobtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of

records or dormancy time increases and tends toward zero as time reaches end point. Result show that at about dormancy time, t = 291.5 months, dormancy of recordsapproaches 100%. The results are presnted graphically in Figure 4.2 for the survival curve.

Table 4.10 Distribution of Survival function of dormancy times (1990-1994)

Time (months)	Dormant records	Survival Function	Std. Error	95% CI	
0.5	405	0.98	0.02	0.86	1.00

3.5	270	0.96	0.03	0.84	0.99
9.5	81	0.94	0.03	0.82	0.98
15.5	55	0.92	0.04	0.79	0.97
21.5	32	0.90	0.04	0.77	0.96
27.5	18	0.88	0.05	0.74	0.94
33.5	18	0.85	0.05	0.72	0.93
39.5	16	0.83	0.05	0.69	0.91
45.5	13	0.81	0.06	0.67	0.90
51.5	9	0.79	0.06	0.65	0.88
57.5	8	0.77	0.06	0.62	0.87
63.5	7	0.75	0.06	0.60	0.85
69.5	9	0.73	0.06	0.58	0.83
75.5	4	0.71	0.07	0.56	0.82
81.5	9	0.69	0.07	0.54	0.80
87.5	6	0.67	0.07	0.51	0.78
93.5	7	0.65	0.07	0.49	0.76
99.5	9	0.63	0.07	0.47	0.74
105.5	5	0.60	0.07	0.45	0.73
111.5	4	0.58	0.07	0.43	0.71
117.5	3	0.56	0.07	0.41	0.69
123.5	5	0.54	0.07	0.39	0.67
129.5	2	0.52	0.07	0.37	0.65
135.5	7	0.50	0.07	0.35	0.63
141.5	3	0.48	0.07	0.33	0.61
147.5	3	0.46	0.07	0.31	0.59
153.5	3	0.44	0.07	0.30	0.57
159.5	5	0.42	0.07	0.28	0.55
165.5	2	0.40	0.07	0.26	0.53
171.5	2	0.38	0.07	0.24	0.51
177.5	4	0.35	0.07	0.22	0.49
183.5	7	0.33	0.07	0.21	0.47
189.5	2	0.31	0.07	0.19	0.44
195.5	6	0.29	0.07	0.17	0.42
201.5	3	0.27	0.06	0.16	0.40
207.5	3	0.25	0.06	0.14	0.38
213.5	5	0.23	0.06	0.12	0.35
219.5	2	0.21	0.06	0.11	0.33
Time (months)	**Dormant records**	**Survival Function**	**Std. Error**	**95% CI**	
225.5	4	0.19	0.06	0.09	0.31
231.5	1	0.17	0.05	0.08	0.28
237.5	2	0.15	0.05	0.06	0.26

243.5	2	0.13	0.05	0.05	0.23
255.5	0	0.10	0.04	0.04	0.21
261.5	1	0.08	0.04	0.03	0.18
273.5	1	0.06	0.03	0.02	0.15
279.5	1	0.04	0.03	0.01	0.13
285.5	1	0.02	0.02	0.00	0.10
291.5	2	0.00	.	.	.

Figure 4.2: Survival curve of time to dormancy of patient records created 1990-1994

Median dormancy time,*(MDT),* standard errors and confidence intervals by patient characteristics

Table 4.11show the 50th percentile of the dormancy time of records by categories of patient characteristics. The result show that the median dormancy time, (MDT) for records created between 1990 and 1994 wss 1.93, with a standard error of 0.16. This is equivalent to *S(t)=0.5.*

Records of patientswith aged less than 10 yearshad a MDT of 1.01 months, records of patients aged 10-20 was 1.47months and 31-60 years was 2.23 months, while the MDT for records of patients 60 years and above was 7.22 months. Records of male patients was dormant in 2.03 months compared to their female counterpart with 1.80 months, theMDT of records of patients resident in Oyo State was 1.97 months as against those from other states with 1.87 months. GYNE records *MDT*was 2.33 months, MOP records 1.47, SOP and CHOP records had *MDT* of 2.06 and 2.26 months respectively. Records of patients with history of admission was 1.14 months compared to patients not admitted with 2.52 months, and records of patients with history of surgery was 1.80 months compared to others with 1.97, while *MDT* of record of patients alive at time of last contact was of 1.93 months and those discharged against medical advice was found to be 0.09 months.

Table 4.11:Median-Dormancy-Time (MDT) by patient characteristics Cohort 1 (1990-1994)

Variables n = 1067	Level	n	t (months)	Std. Error	95% CI	
		1067	**1.93**	**0.16**	**1.70**	**2.33**
Age at Registration	<10	210	1.01	0.18	0.85	1.64
	10<20	147	1.47	0.22	1.01	2.46
	20-30	224	1.87	0.33	1.18	2.75
	31-60	380	2.22	0.29	1.83	2.98
	61+	103	7.22	3.77	2.75	12.15
Gender	male	529	2.03	0.22	1.60	2.46
	female	506	1.80	0.23	1.44	2.33
State of residence	Others States	518	1.87	0.21	1.60	2.36
	Oyo State	487	1.97	0.25	1.57	2.52
Clinic attended	MOP	230	1.47	0.30	1.05	2.52
	SOP	259	2.06	0.254	1.60	2.75
	CHOP	30	2.26	3.89	0.91	20.76
	GYNE	146	2.33	0.40	1.70	3.41
	Others	370	1.87	0.29	1.44	2.59
Ever admitted	No	734	2.52	0.26	2.06	3.03
	Yes	330	1.149	0.17	0.88	1.60
Ever operated on	No	954	1.97	0.19	1.64	2.36
	Yes	110	1.80	0.28	1.70	2.33
Treatment outcome	Alive	1052	1.93	0.17	1.70	2.33
	Died	1	-	-	-	-
	DAMA	3	0.09	0.05	0.03	-
	referred	-	-	-	-	-

Selected percentiles of the survival curve (1990-1994) data

Estimates of specific points of dormancy time for 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of observed survival distribution for patient records. The 25^{th} percentile survival estimate show that 25% records were dormant at 0.46 months, 50% (median dormancy time) records were dormant in 1.94 months. Also, the 75^{th} and 95^{th} percentiles showed that seventy five percent and ninety five percent of the records were dormant in 17.12 and 151.89 months respectively. Table 4.12 below shows the estimated record dormancy time, their standard error and confidence interval at each selected percentile point.

Table 4.12Selected percentiles of Dormancy Time 1990-1994

Percentiles	**t (months)**	**Std. Error**	**95% *CI***	
25^{th}	0.45	0.04	0.39	0.49
50^{th}	1.93	0.16	1.70	2.33
75^{th}	17.11	1.86	14.29	21.88
95^{th}	151.89	12.31	128.72	179.05
n = 1067				

4.3.4 Hazard function of dormancy timetime for records created 1990-2014

Table 4.13 shows the hazard function $\lambda(t)$ of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier method. The hazard plot that follows, Figure 4.3, show hazard curve of dormancy time, the hazard rate was high at the initial time, t, but decreased sharply as age of records (dormancy time) increases gradually until it reaches time point of dormancy time, t = 50 months, the plot then remain in a constant movement till time point of 150 months. from this point the plot increased with a sharp upward movement following constant and steady rise till it reaches end point making a bathtub shape.

Table 4. 13: Frequency distribution of hazard function (1990-1994) Cohort 1

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	665	405	0.38	0.01	0.35	0.41
5 -	408	253	0.62	0.01	0.59	0.65
10 -	368	39	0.66	0.01	0.63	0.68
15 -	291	77	0.73	0.01	0.70	0.75
20 -	249	43	0.77	0.01	0.74	0.79
25 -	222	26	0.79	0.01	0.77	0.82
30 -	205	17	0.81	0.01	0.78	0.83
35 -	191	14	0.82	0.01	0.80	0.84
40 -	176	15	0.84	0.01	0.81	0.86
45 -	169	7	0.84	0.01	0.82	0.86
50 -	157	12	0.85	0.01	0.83	0.87
55 -	148	9	0.86	0.01	0.84	0.88
60 -	141	7	0.87	0.01	0.85	0.89
65 -	136	5	0.87	0.01	0.85	0.89
70 -	131	5	0.88	0.01	0.86	0.89
75 -	121	10	0.89	0.01	0.87	0.91
80 -	119	2	0.89	0.01	0.87	0.91
85 -	111	8	0.90	0.01	0.88	0.91
90 -	109	2	0.90	0.01	0.88	0.92
95 -	102	7	0.91	0.01	0.89	0.92
100 -	96 -	6	0.91	0.01	0.89	0.93
105 -	88	8	0.92	0.01	0.90	0.93
110 -	84	4	0.92	0.01	0.90	0.94
115 -	80	4	0.93	0.01	0.91	0.94
120 -	76	4	0.93	0.01	0.91	0.94
125 -	72	3	0.93	0.01	0.92	0.95
130 -	68	5	0.94	0.01	0.92	0.95
135 -	67	1	0.94	0.01	0.92	0.95
140-	60	7	0.94	0.01	0.93	0.96
145 -	58	2	0.95	0.01	0.93	0.96
150 –	56	2	0.95	0.01	0.93	0.96
155 -	54	2	0.95	0.01	0.94	0.96
160 -	51	3	0.95	0.01	0.94	0.96
165 -	46	5	0.96	0.01	0.94	0.97
170 -	45	1	0.96	0.01	0.95	0.97
175 -	43	2	0.96	0.01	0.95	0.97
180 -	40	3	0.96	0.01	0.95	0.97
185 -	35	5	0.97	0.01	0.96	0.98
190 -	31	4	0.97	0.01	0.96	0.98
195 -	29	2	0.97	0.00	0.96	0.98

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
200 -	24	5	0.98	0.00	0.97	0.99
205 -	22	2	0.98	0.00	0.97	0.99
210 -	21	1	0.98	0.00	0.97	0.99
215 -	17	4	0.99	0.00	0.96	0.99
220 -	13	4	0.99	0.00	0.98	0.99
225 -	10	3	0.99	0.00	0.98	0.99
230 -	9	1	0.99	0.00	0.98	1.00
235 -	8	1	0.99	0.00	0.99	1.00
240 -	6	2	0.99	0.00	0.99	1.00
245 -	4	2	0.99	0.00	0.99	1.00
250 -	4	0	0.99	0.00	0.99	1.00
255 -	4	0	0.99	0.00	0.99	1.00
260 -	2	2	0.99	0.00	0.99	1.00
265 -	2	0	0.99	0.00	0.99	1.00
270 -	2	0	0.99	0.00	0.99	1.00
275 -	2	0	0.99	0.00	0.99	1.00
280 -	1	1	0.99	0.00	0.99	1.00

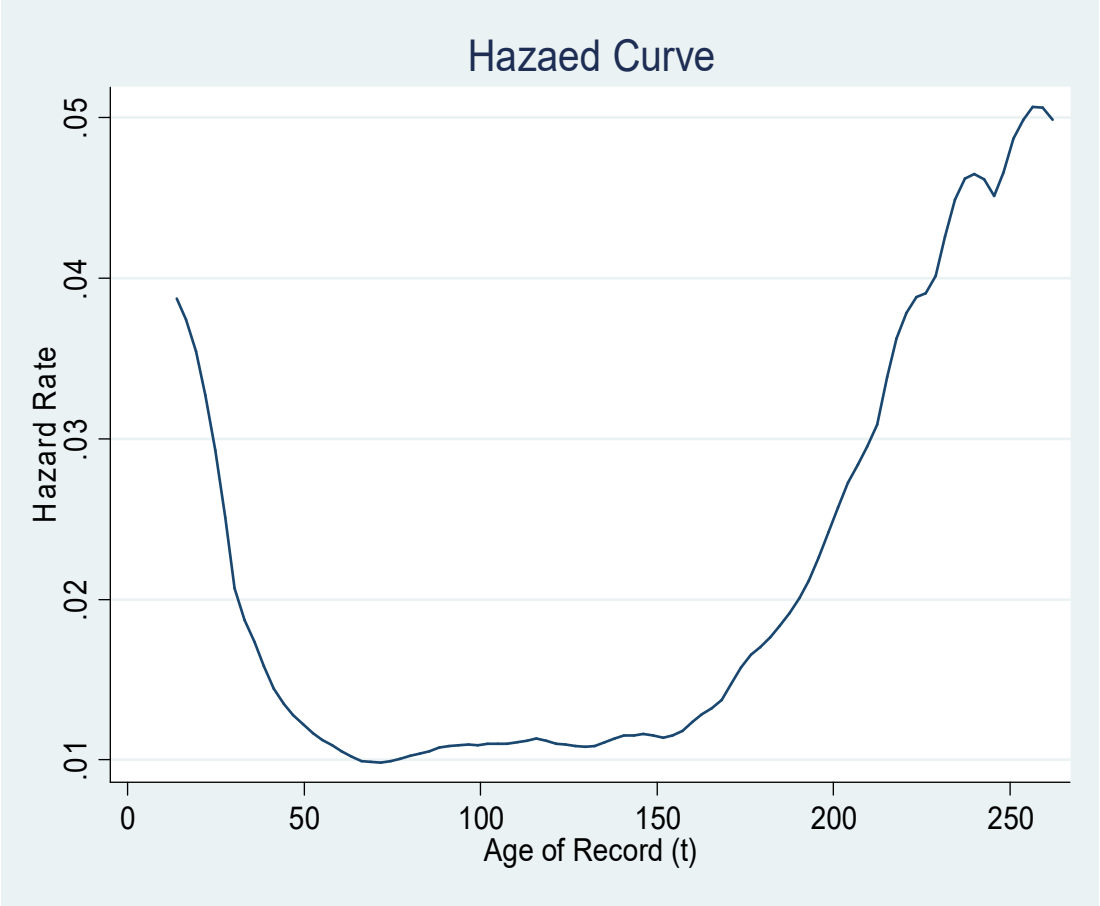


Figure 4.3: Hazard plot of time to dormancy of patient records created 1990-1994

Graphical evaluation of the form of the hazard rate of time-to-dormancy of patient records created 1990-1994

Considering the U-shape, (bathtub type), of the hazard plot, Figure 4.3., shows the result of the test for distribution assumption using Weibull probability plot of Kaplan-Meier log-log survival curves, *logH(t),* against log of survival time, *log(t).* The result show a straight line relationship between *logH(t)* against *log(t),* increasing monotonically suggesting a Weibull distribution. The intercept of the straight line is approximately - 0.5813 with a slope of 0.3581. From this results, the value of the shape parameter, γ, for two parameter Weibull distribution was estimated as:

$$\gamma^* = exp(-0.5813) = 0.5592 \text{ and}$$

the estimated hazard rate $\lambda^* = 0.3581$.

And since the estimated value of the shape parameter, γ, was less than unity, suggesting a decreasing hazard, λ, typical of Weibull distribution.

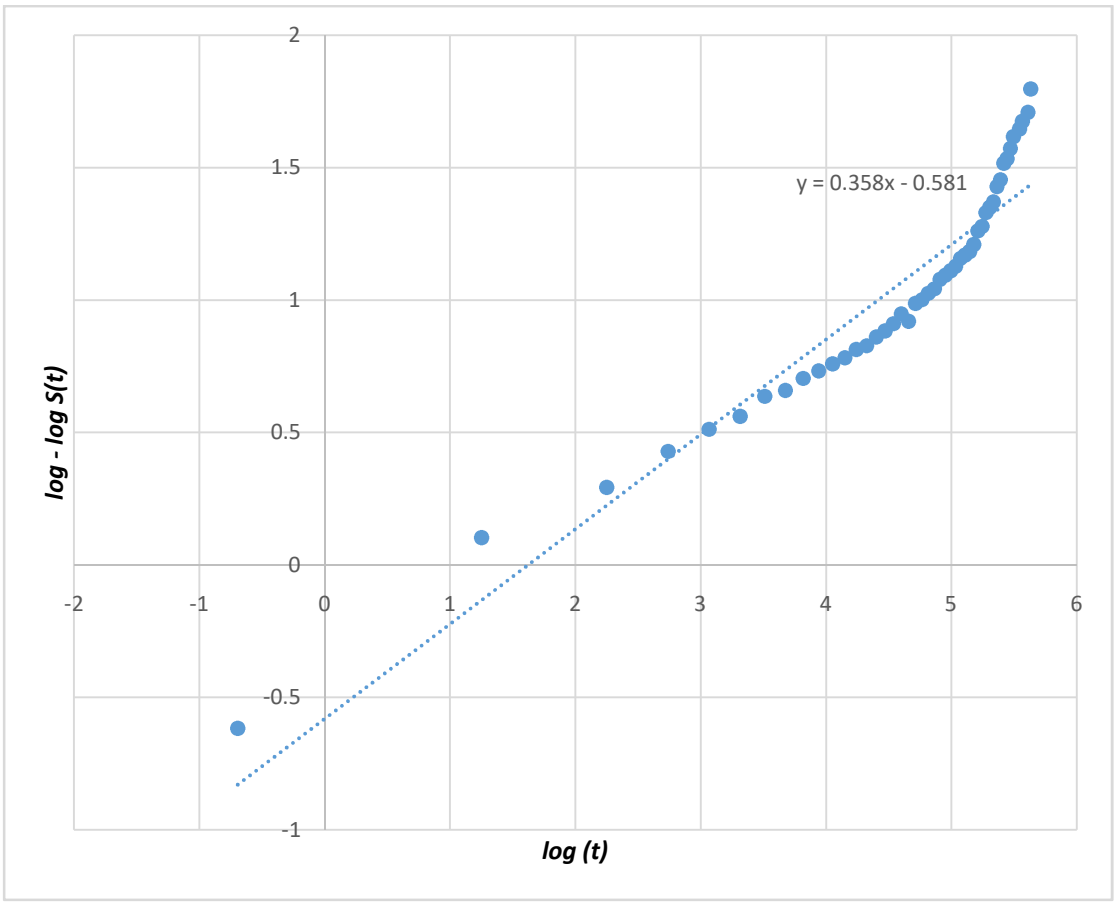


Figure 4.4.A Weibull plot of *log (t)* and *log–log S(t)* with line fitted

4.3.5 Influence of patient characteristics on hazard rate of patient records created 1990-1994

Results of semi-parametric (Cox Proportional Hazard) and Parametric (Exponential and Weibull) survival model used to measure the influence of patients demographic and clinical characteristics on dormancy time of records created between 1990 and 1994 (cohort 1) show as follow:

4.3.5.1 Non-parametric approach

Schoenfeld Test of Cox Proportional Hazard Model Assumption

Table 4.14 below show the global test for the proportional hazard assumption. The insignificant result of the test implies that the sample data did not violate the proportional hazard assumption, that the hazard of subject subgroup are proportional over follow-up period and therefore the global test indicated that for the data set used the assumption of PH is not violated.

.

Table 4.14: Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	Chi-square	df	p-value
Proportional Hazard Assumption	6.29	7	0.51

Graphical test for Proportional Hazard Assumption

The graph, figure 4.4, of the log-log Kaplan Meier estimate on dormancy time comparing patient's gender while adjusting for age, State of residence and clinics shows that the two line (male and female) are not parallel and indicating that the proportional assumption is invalid for TTD data patient records created in UCH, Ibadan:

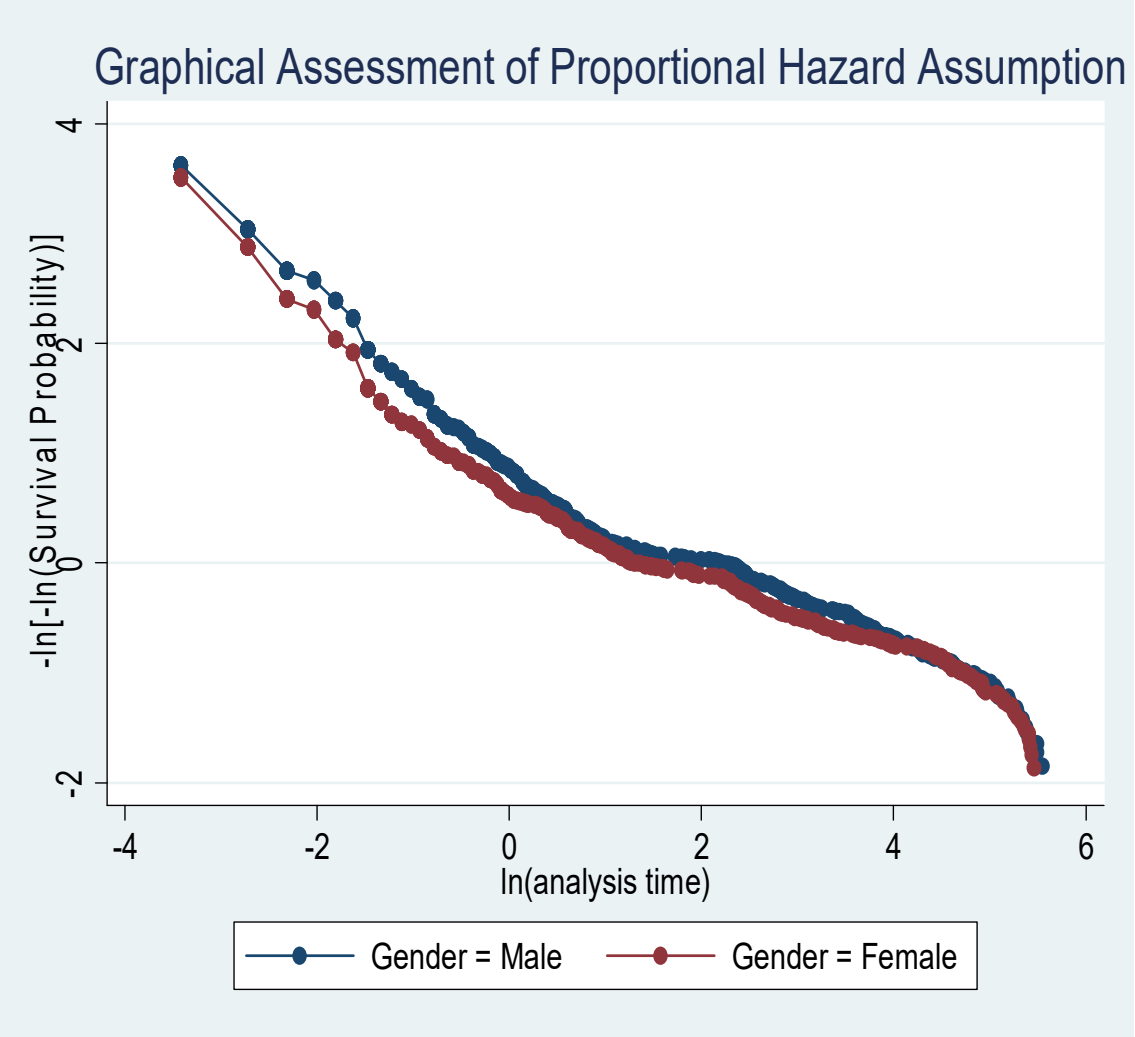


Figure 4.5 Graph testing for Cox Proportional Hazard Assumption.

Fitting Cox Proportional Hazard Model

Table 4.15 below shows result from the Cox regression analysis that succeed the global test above. Record dormancy time are affected by patient age (HR=0.92, p-value = 0.004), but failed to do so when categorised. State of residence (HR=0.89, p-value = 0.088), admission status (HR=1.19, p-value = 0.042) and treatment outcome (HR=4.01, p-value = 0.000) were significant at 1%, 10%, 5% and 1% respectively. Whereas, patients gender (HR=1.09, p-value = 0.153), clinics attended (HR=1.00,p-value = 0.887) and surgery status (HR=0.84, p-value = 0.192) will not influence record dormancy time as they are all insignificant and indicating failure to accept the research hypothesis.

Table 4.15 Cox Regression of Dormancy Time on Patients Characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.92**	**-2.90**	**0.00**	**0.87**	**0.97**
	60+	0.74	-2.21	-2.21	0.57	0.57
	31-60	0.76	-2.57	-2.57	0.62	0.93
	20-30	0.84	-1.50	0.13	0.67	1.05
	10>20	0.88	-1.08	0.27	0.69	1.10
	<10 years (rc)					
Gender		**1.09**	**1.43**	**0.15**	**0.96**	**1.25**
	female	1.12	1.58	0.11	0.97	1.29
	Male (rc)					
State of Residence		**0.89**	**-1.71**	**0.08**	**0.78**	**1.01**
	Oyo	0.89	-1.70	0.09	0.78	1.01
	Others (rc)					
clinics		**1.00**	**0.14**	**0.88**	**0.96**	**1.04**
	OTHERS	1.02	0.29	0.77	0.85	1.22
	GYNE	0.94	-0.21	0.832	0.76	1.23
	CHOP	0.89	-0.53	0.59	0.58	1.35
	SOP	1.04	0.44	0.66	0.85	1.26
	MOP (rc)					
Patient Admitted		**1.19**	**2.03**	**0.04**	**1.00**	**1.41**
	Yes	1.19	2.04	0.41	1.00	1.42
	No (rc)					
Surgery done		**0.84**	**-1.30**	**0.19**	**0.65**	**1.08**
	Yes	0.83	-1.38	0.16	0.64	1.079
	No (rc)					
Treatment Outcome		**4.01**	**3.99**	**0.00**	**2.02**	**7.94**
	DAMA	17.90	3.97	0.00	4.30	74.42
	Died	3.10	1.12	0.26	0.42	22.39
	Alive (rc)					

4.3.5.2 Parametric approach

Fitting Exponential Model on Patient characteristics

Giving that the dormancy time data is skewed distributed data, the result of the regression of dormancy time on patients characteristics based on exponential model assumption of parameter $\lambda=1$. Table 4.16, show that patient age at registration along with other characteristics like State of residence, gender and treatment outcome significantly (HR<1.00P<0.01, P<0.10) influence their dormancy time. The significant effect of gender HR= 1.11, clinic HR=1.05 imply that female patient's record have higher risk of being dormant compare to male patient's record.

Table 4.16: Exponential Model of Dormancy Time on Patient characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.87**	**-4.75**	**0.00**	**0.82**	**0.92**
	60+	0.63	-3.30	0.00	0.48	0.83
	31-60	.576	-5.10	0.0	0.46	0.71
	20-30	0.68	-3.41	0.00	0.54	0.84
	10>20	0.71	-2.76	0.00	0.56	0.90
	<10 years(rc)					
Gender		**1.11**	**1.61**	**0.10**	**0.97**	**1.27**
	female	1.16	2.07	0.00	0.56	0.90
	male(rc)					
State of Residence		**0.77**	**-3.91**	**0.00**	**0.67**	**0.87**
	Oyo	0.76	-4.02	0.00	0.67	0.8Z
	others(rc)					
clinics		**1.00**	**0.35**	**0.72**	**0.96**	**1.04**
	Others	1.06	0.74	0.46	0.89	1.27
	GYNE	1.1	0.92	0.35	0.88	1.42
	CHOP	1.1	0.57	0.57	0.74	1.72
	SOP	1.18	1.74	0.08	0.97	1.44
	MOP(rc)					
Patient Admitted		**1.05**	**0.56**	**0.57**	**0.88**	**1.24**
	Yes	0.99	-0.02	0.98	0.83	1.20
	No (rc)					
Surgery done		**0.89**	**-0.85**	**0.39**	**0.69**	**1.15**
	Yes	0.88	-0.94	0.34	0.67	1.14
	No (rc)					
Treatment Outcome		**19.85**	**9.12**	**0.00**	**0.00**	**37.74**
	DAMA	403.99	8.36	0.00	98.88	1650.48
	Died	58.06	4.03	0.00	8.05	418.47
	Alive(rc)					
_cons	**variables**	0.00	-15.95	0.00	0.00	0.00
	categories	0.06	-23.37	0.00	0.04	0.07

Fitting Weibull Model:

Result of Weibull model fitted to the skewed distributed dormancy time data under the assumption that the exponential model fail and the model fit Weibull model of parameter ϒ=λ=1. Similarly Patient Age (HR=0.93, P<0.01), state of residence (HR=0.87, P<0.05), admission status (HR=1.16, P<0.10) and treatment outcome (HR=2.97, P<0.01) significantly influence patient record dormancy time. However Patients age and type of Clinic attended will not determine patient record dormancy time. Table 4.17 below shows the Weibull regression model result.

Table 4.17: Weibull Regression Model of Dormancy Time on Patient characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.93**	**-2.68**	**0.00**	**0.88**	**0.98**
	60+	0.75	-2.11	0.03	0.578	0.97
	31-60	0.78	-2.37	0.01	0.63	0.95
	20-30	0.84	-1.46	0.14	0.68	1.05
	10>20	0.88	-1.04	0.29	0.70	1.11
	<10 years(rc)					
Gender		**1.10**	**1.43**	**0.15**	**0.96**	**1.25**
	female	1.12	1.60	0.11	0.97	1.29
	male(rc)					
State of Residence		**0.87**	**-2.02**	**0.04**	**0.76**	**0.99**
	Oyo	0.87	-2.03	0.04	0.76	0.99
	others(rc)					
clinics		**0.99**	**-0.13**	**0.90**	**0.95**	**1.03**
	OTHERS	1.01	0.14	0.88	0.84	1.21
	GYNE	0.97	-0.20	0.84	0.77	1.23
	CHOP	0.88	-0.57	0.56	0.58	1.34
	SOP	1.06	0.64	0.52	0.87	1.29
	MOP(rc)					
Patient Admitted		**1.169**	**1.82**	**0.06**	**0.98**	**138**
	Yes	1.17	1.81	0.07	0.98	1.39
	No (rc)					
Surgery done		**0.81**	**-1.53**	**0.12**	**0.63**	**1.05**
	Yes	0.80	-1.64	0.10	0.62	1.04
	No (rc)					
Treatment Outcome		**2.97**	**3.23**	**0.00**	**1.53**	**5.75**
	DAMA	8.79	3.02	0.0	2.14	36.09
	Died	3.59	1.27	0.20	0.49	25.91
	Alive(rc)					
_cons	**variable**	0.14	-5.18	0.00	0.06	0.29
	categories	0.42	-6.76	0.00	0.33	0.54
/ln_p	**variable**	-0.78	-32.28	0.00	-0.83	-0.73
	categories	-0.78	-32.24	0.00	-0.83	-0.73
P **1/p**	**variable**	0.45 2.19			0.43 2.09	0.47 2.30
	categories	0.45 2.19			0.43 2.09	0.47 2.30

4.4Cohort 2: Patient records created between 1st Jan. 1995 and 31st Dec, 1999

Between 1stJanuary, 1995 and 31stDecember, 1999, seventy nine thousand four hundred and seventeen (79,417) records were created in UCH, Ibadan, a sample of 1537 were selected for the study. Not less than 354 (23.00%) of the 1537 patients record were found to be inactive (dormant) afterthe first day of creation andthis was indicated by a single entry in the medical records.The 354 one-day-active records were excluded and results of the analysis of the remaining 1183 records are presented below.

4.4.1 Frequency distribution of some demographic and clinical characteristics of the patients

Result of analysis show that 33.75% of the patient who had two or more visits are between the ages of 31-60 years, 11.28% were within 10-20 years of age, 20.08% were less than 10 years and 13% were above 61 years of age. Male patients constitute 47.26%, whilepatient's residence in Oyo State were 49.47%.Records of patients from MOPclinic were 18.19%,SOP were 24.17%, CHOPhad 7.56%and 36.12% of the patients records were from other clinics.. Not less than 42% of the patients were admitted at one time or the other, while 15.10% of the patients went through surgical operation. Almost (98.97%) all the patients were alive at the end of their last contact, 1.03% were discharge against medical advice but no patient died during the period. Table 4.18 below shows the socio-demographic and clinical characteristics of the patient whose records were observed for dormancy time.

Table 4.18 Frequency distribution of patient's characteristics 2nd cohort 1995-1999

Variables n=1183	Level	Frequency	Percent	Cum
Age at Registration	<10	210	20.08	20.08
	10-20	118	11.28	31.36
	21-30	229	21.89	53.25
	31-60	353	33.75	87.00
	61+	136	13.00	100
Gender	male	551	47.26	47.26
	female	615	52.74	100
State of residence	Oyo State	564	49.47	49.47
	Others	576	50.53	100.00
Clinic attended	MOP	207	18.19	18.19
	SOP	275	24.17	42.36
	CHOP	86	7.56	49.91
	GYNE	159	13.97	63.88
	Others	411	36.12	100.00
Ever admitted	No	682	57.99	57.99
	Yes	494	42.01	100.00
Ever operated on	No	1001	84.90	84.90
	Yes	178	15.10	100.00
Treatment outcome	Alive	1149	98.97	98.97
	Died	-	-	-
	DAMA	12	1.03	100.00
	referred	-	-	-

4.4.2 Frequency distribution of dormancy times in cohort 2 (1995-1999)

The result on Table 4.19, show that of the 1537 records observed 354 or 23.0% became dormant on the first day of creation. As at half of the month, (t=0.5 months), 51.6 % of the records were already dorment, 80.3% of the records became dormant from day of creation up to when t = 9.5 months of creation. The result also showd that about 95.3% of records were dormant at the age of t=117.5 months, while at the age of 225.5 months almost all the records had become dormant.. The distribution is presented graphically in Figure 4.6. The distribution is skewed to the right.

Table 4.19 Frequency distribution of dormancy times1995-1999

Month (t*)		Frequency	Percent	Cum. Percent
0-<1	0.5	439	37.11	37.11
1-6	3.5	337	28.49	65.60
7-12	9.5	105	8.88	74.47
13-18	15.5	31	2.62	77.09
19-24	21.5	24	2.03	79.12
25-30	27.5	19	1.61	80.73
31-36	33.5	26	2.20	82.92
37-42	39.5	21	1.78	84.70
43-48	45.5	13	1.10	85.80
49-54	51.5	13	1.10	86.90
55-60	57.5	15	1.27	88.17
61-66	63.5	10	0.85	89.01
67-72	69.5	5	0.42	89.43
73-78	75.5	3	0.25	89.69
79-84	81.5	12	1.01	90.70
85-90	87.5	4	0.34	91.04
91-96	93.5	11	0.93	91.97
97-102	99.5	9	0.76	92.73
103-108	105.5	6	0.51	93.24
109-114	111.5	**4**	0.34	93.58
115-120	117.5	8	0.68	94.25
121-126	123.5	11	0.93	95.18
127-132	129.5	7	0.59	95.77
133-138	135.5	5	0.42	96.20
139-144	141.5	2	0.17	96.37
145-150	147.5	5	0.42	96.79
151-156	153.5	5	0.42	97.21
157-162	153.5	5	0.42	97.63
163-168	159.5	6	0.51	98.14
169-174	165.5	3	0.25	98.39
175-180	171.5	3	0.25	98.65
181-186	177.5	3	0.25	98.90
187-192	183.5	4	0.34	99.24
193-198	189.5	3	0.25	99.49
199-204	195.5	4	0.34	99.83
223-228	201.5	2	0.17	100.00
		1183	100.00	

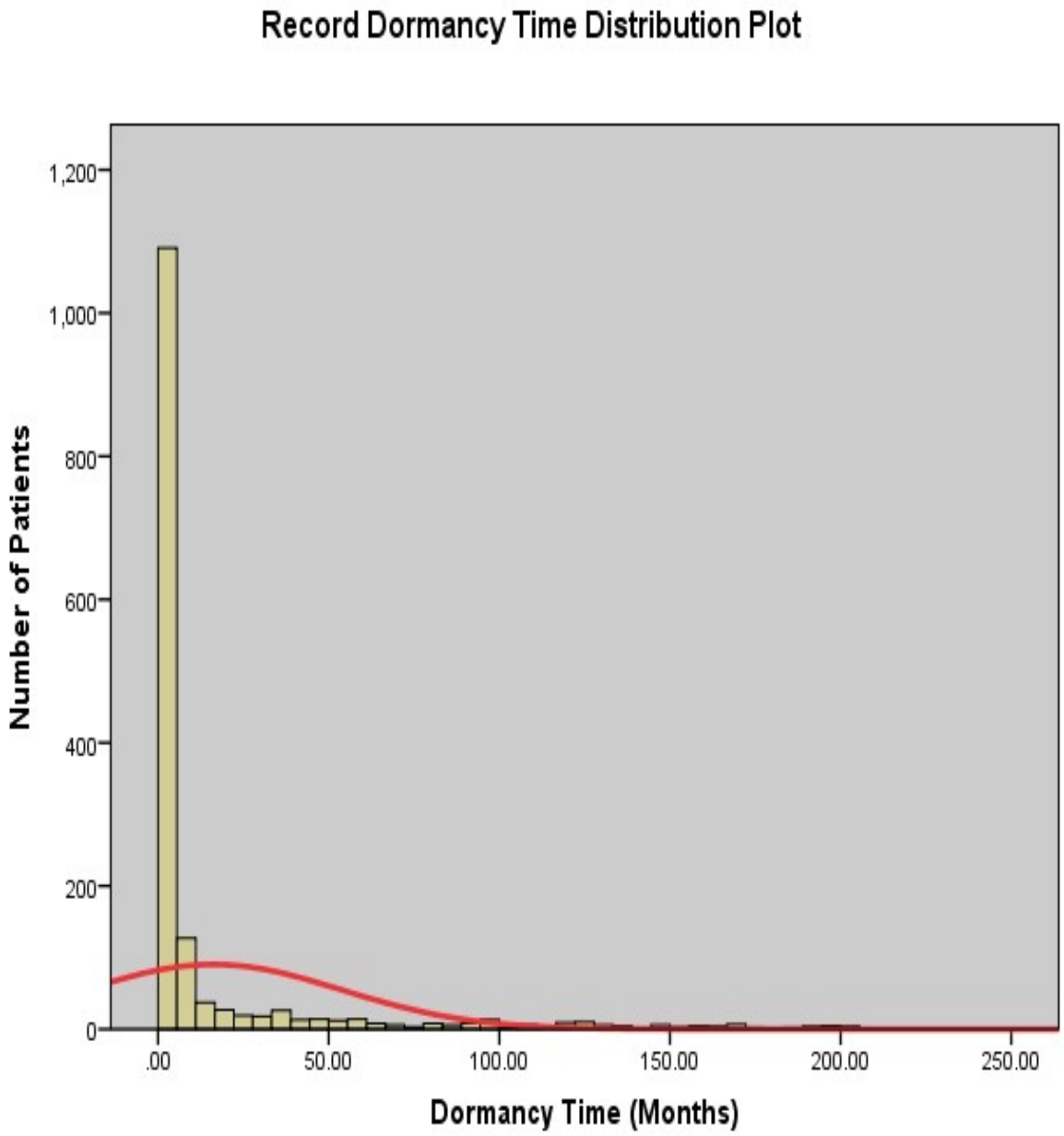


Figure 4.6 Distribution of dormancy times1995-1999

4.4.3 Survival function of dormancy times1995-1999

Table 4.20 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of records or dormancy time increases and tends toward zero as time reaches end point. Result show that at the dormancy timeof approximately 201.5 months, dormancy of records approaches 100%. The results are presnted graphically in Figure 4.7 for the survival curve.

Table 4.20Distribution of Survival function of dormancy times 1995-1999

Time (months)	Dormant records	Survival Function	Std. Error	95% CI	
0.5	439	0.97	0.03	0.82	1.00
3.5	337	0.94	0.04	0.80	0.99
9.5	105	0.92	0.05	0.76	0.97
15.5	31	0.89	0.05	0.73	0.96
21.5	24	0.86	0.06	0.70	0.94
27.5	19	0.83	0.06	0.67	0.92
33.5	26	0.81	0.07	0.64	0.90
39.5	21	0.78	0.07	0.60	0.88
45.5	13	0.75	0.07	0.57	0.86
51.5	13	0.72	0.07	0.55	0.84
57.5	15	0.69	0.08	0.52	0.82
63.5	10	0.67	0.08	0.49	0.80
69.5	5	0.64	0.08	0.46	0.77
75.5	3	0.61	0.08	0.43	0.75
81.5	12	0.58	0.08	0.41	0.72
87.5	4	0.56	0.08	0.38	0.70
93.5	11	0.53	0.08	0.35	0.67
99.5	9	0.50	0.08	0.33	0.65
105.5	6	0.47	0.08	0.30	0.62
111.5	**4**	0.44	0.08	0.28	0.60
117.5	8	0.42	0.08	0.26	0.57
123.5	11	0.39	0.08	0.23	0.54
129.5	7	0.36	0.08	0.21	0.51
135.5	5	0.33	0.08	0.19	0.49
141.5	2	0.31	0.08	0.17	0.46
147.5	5	0.28	0.07	0.14	0.43
153.5	5	0.22	0.07	0.10	0.37
159.5	6	0.19	0.07	0.09	0.34
165.5	3	0.17	0.06	0.07	0.30
171.5	3	0.14	0.06	0.05	0.27
177.5	3	0.11	0.05	0.04	0.24
183.5	4	0.08	0.05	0.02	0.20
189.5	3	0.06	0.04	0.01	0.16
195.5	4	0.03	0.03	0.00	0.12
201.5	2	0.00	.	.	.

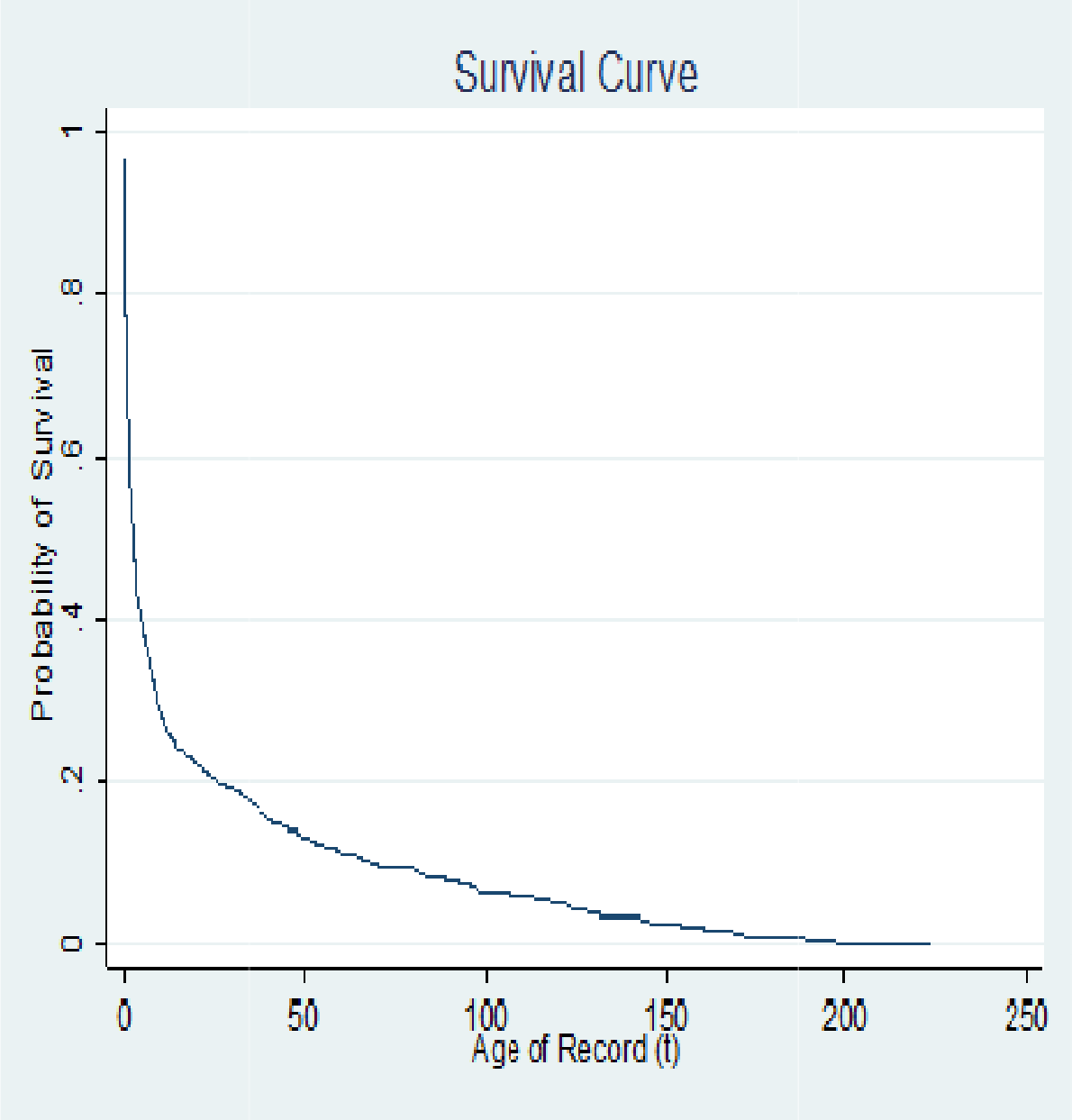


Figure 4.7: Survival Curve 2nd cohort 1995-1999

Median dormancy time, standard errors and confidence intervals by patient characteristics

Table 4.21showmedian dormancy time according to categories of patient characteristics.The median dormancy time is the point where survivorship *S(t)=0.5* and equivalently a point where 50% of dormancy time was 2.29 months with a SE of 0.19.

Records of patients aged below 10 years had a median dormancy time *(MDT)* of 1.41 months, for patients aged 10-20 years MDT was1.60 monthsand those aged 21-30 yearsin 2.69 months. Records of patients aged 31- 60 and those above 60 years were dormant in 3.1 months respectively. The *MDT* for records of male patients was 1.97 against those of females with 2,66 months,records of patients resident in Oyo State was dormant in 2.66 months compared to those from other state with 2.06. Records of patients attending GYNE clinic hadan *MDT* of 4.37months, MOP records 2.06 months, records in SOP and CHOP'sMDT was2.52 and 3.38 months respectively. Records of patients with history of admission was 2.79 months compared and non-admitted patients records of 2.10, records of patients with surgery was 5.19 months as against patients without surgery with MDT of 1,90 months. Record of patients alive at time of last entry/contact wasdormant in 2.33 months while the MDT of those with DAMA was 0.06 months.

Table 4.21 Median-Dormancy-Time by Patient Characteristics 2nd cohort 1995-1999

Variables n=1183	Level	n	months (t)	Std. Error	95% CI	
		1183	**2.29**	**0.19**	**1.90**	**2.75**
Age at Registration	<10	210	1.41	0.40	0.91	2.49
	10<20	118	1.60	0.40	1.11	2.82
	20-30	229	2.69	0.62	1.60	4.10
	31-60	353	3.12	0.56	2.29	4.53
	61+	136	3.12	1.38	2.06	6.86
Gender	male	551	1.97	0.25	1.60	2.75
	female	615	2.66	0.35	1.97	3.35
State of residence	Others States	576	2.06	0.27	1.57	2.75
	Oyo State	564	2.66	0.32	1.97	3.21
Clinic attended	MOP	207	2.06	0.39	1.60	3.35
	SOP	275	2.52	0.38	1.60	3.12
	CHOP	86	3.38	0.57	1.51	6.96
	GYNE	159	4.36	0.98	2.10	6.34
	Others	411	1.77	0.32	1.37	2.56
Ever admitted	No	682	2.10	0.21	1.64	2.66
	Yes	494	2.79	0.49	1.80	3.61
Ever operated on	No	1001	1.90	0.19	1.60	2.39
	Yes	178	5.19	0.19	1.87	2.75
Treatment outcome	Alive	1149	2.33	0.21	1.90	2.79
	Died	-	-	-	-	-
	DAMA	12	0.06	0.03	0.03	2.26
	Referred	-	-	-	-	-

Selected percentiles of the survival distribution

Table 4.22 shows the respective estimated *MDT*, their standard error and confidence interval at selected percentile point. Estimatesfor 25^{th}, 50th, 75^{th} and 95^{th} percentiles of dormancy time for records created between 1995 and 1999 show that twenty five percent of the records were dormant at 0.46 months, fifty percent of records were dormant in 2.30 months as shown from the 50^{th} percentiles. Also, the 75^{th} and 95^{th} percentiles show that not less than seventy five percent and ninety five percent of records were dormant in 13.93 months and 124.85 months respectively.

Table 4.22Selected percentiles of the survival curve 2nd cohort (1995-1999)

Percentiles	t (months)	Std. Error	95% CI	
25th	0.45	0.04	0.36	0.49
50th	2.29	0.19	1.90	2.75
75th	13.93	2,29	10.51	20.04
95th	124.84	8.99	117.35	143.17
n = 1183				

4.4.4 Hazard functions of dormancy times for records created 1995-1999

Table 4.23, show the the distributons hazard function, the standard error and 95% Confidence Interval for chohort 2. The hazard plot that follows, Figure 4.8, show the hazard curve was high at the initial time of records creation, then decreases as age of records (dormancy time) increases and only to remain constant with steady movement between about 30 months and about 120 months and increases with a sharp constant and steady rise till it reaches end point and therefore making a bathtub shape. A shape usually typical of Weibull distribution,

Table 4.23: Frequency distribution of hazard function (1995-1999) Cohort 2

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	747	439	0.37	0.01	0.34	0.40
5 -	468	279	0.61	0.01	0.58	0.63
10 -	337	129	0.72	0.01	0.69	0.74
15 -	288	50	0.76	0.01	0.73	0.78
20 -	269	18	0.77	0.01	0.75	0.80
25 -	248	21	0.79	0.01	0.77	0.81
30 -	232	16	0.80	0.01	0.78	0.83
35 -	214	18	0.82	0.01	0.80	0.84
40 -	189	25	0.84	0.01	0.82	0.86
45 -	180	9	0.85	0.01	0.83	0.87
50 -	166	14	0.86	0.01	0.84	0.88
55 -	156	10	0.87	0.01	0.85	0.89
60 -	144	12	0.88	0.01	0.86	0.90
65 -	138	6	0.88	0.01	0.87	0.90
70 -	128	10	0.89	0.01	0.87	0.91
75 -	124	4	0.90	0.01	0.88	0.91
80 -	122	2	0.90	0.01	0.88	0.91
85 -	111	11	0.91	0.01	0.89	0.92
90 -	109	2	0.91	0.01	0.89	0.92
95 -	100	9	0.92	0.01	0.90	0.93
100 -	88	12	0.93	0.01	0.91	0.94
105 -	86	2	0.93	0.01	0.91	0.94
110 -	80	6	0.93	0.01	0.92	0.95
115 -	77	3	0.94	0.01	0.93	0.95
120 -	71	6	0.94	0.01	0.93	0.95
125 -	60	11	0.95	0.01	0.94	0.96
130 -	54	6	0.96	0.01	0.94	0.97
135 -	51	3	0,96	0.01	0.94	0.97
140-	46	5	0.96	0.01	0.95	0.97
145 -	44	2	0.96	0.01	0.95	0.97
150 –	39	5	0.97	0.01	0.96	0.98
155 -	36	3	0.97	0.00	0.96	0.98
160 -	32	4	0.97	0.00	0.96	0.98
165 -	27	5	0.98	0.00	0.97	0.98
170 -	22	5	0.98	0.00	0.97	0.99
175 -	20	2	0.98	0.00	0.98	0.99
180 -	18	2	0.99	0.00	0.98	0.99
185 -	16	2	0.99	0.00	0.98	0.99
190 -	13	3	0.99	0.00	0.98	0.99
195 -	10	3	0.99	0.00	0.99	0.99

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
200 -	6	4	0.99	0.00	0.99	1.00
205 -	2	4	0.99	0.00	1.00	1.00
210 -	2	0	0.99	0.00	1.00	1.00
215 -	2	0	0.99	0.00	1.00	1.00
220 -	2	0	0.99	0.00	1.00	1.00
225 -	1	1				

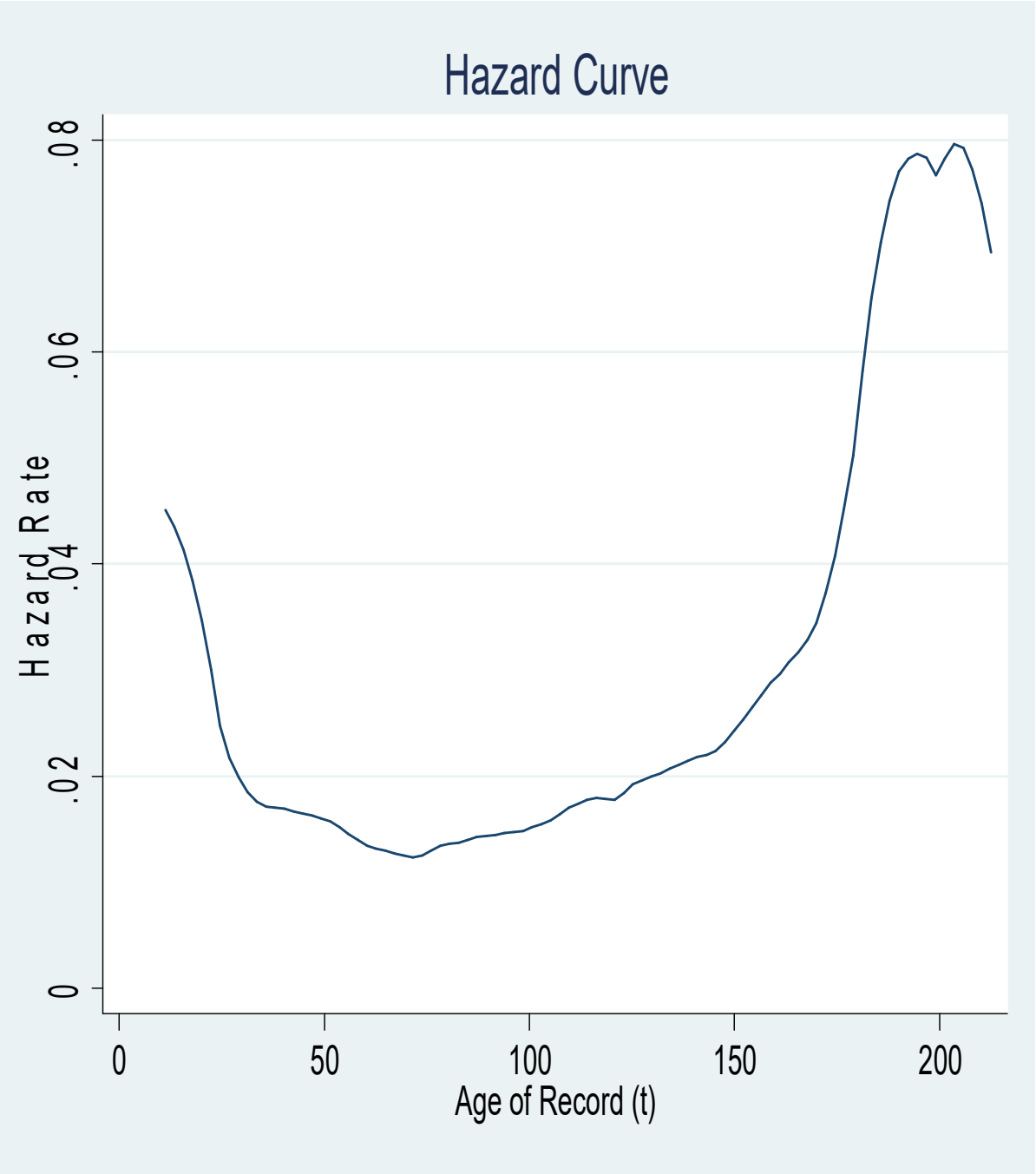


Figure 4.8: Hazard Curve 2nd cohort 1995-1999

Graphical evaluation of appropriateness of Weibull model

The hazard plot, figure 4.9, indicated a bathtub shape typical of Weibull distribution and to further test the validity of distribution a Kaplan-*Meier* log-log Survival curves, *log[H(t)],* against log survival time, *log(t),* was plotted, Figure 4.7. The plot indicated a straight line relationship between *logH(t)* against *log(t),* decreasing monotonically. The intercept was approximately - 0.3113 with a slope of 0.3260. From the value of γ, the shape parameter for two parameter Weibull distribution was estimated as:

$$\gamma^* = exp(-0.3113) = 0.8668 \text{ and}$$

the estimated hazard rate estimate as:

$$\lambda^* = 0.3260.$$

The estimated value of the shape parameter, γ, was less than unity, suggesting a decreasing hazard, λ, of the Weibull distribution.

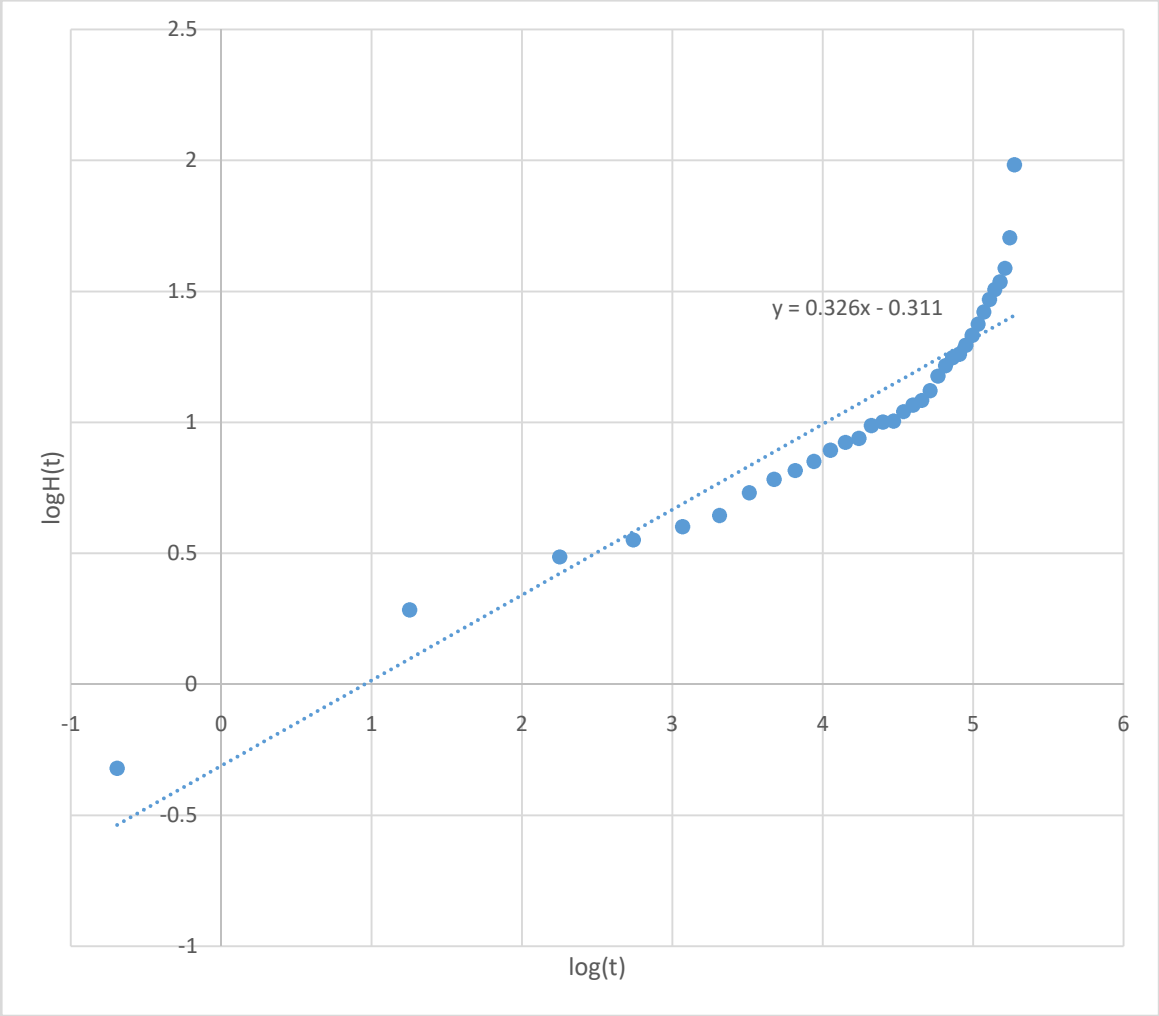


Figure 4.9:A Weibull plot of *log (t)* and *log–log S(t)* with line fitted 2nd cohort 1995-1999

4.4.5 Influence of patient characteristics on hazard rate

Results of semi-parametric (Cox Proportional Hazard) and Parametric (Exponential and Weibull) survival model used to measure effect of patients demographic and health characteristics on dormancy time of records created between 1995 and 1999 (2^{nd} cohort) show as follow:

4.4.5.1 Non Parametric approach

Schoenfeld Test of Cox Proportional Hazard Model Assumption

Table 4.24 show the global test for the proportional hazard assumption. The insignificant result of the test implies that the sample data is valid for the proportional hazard assumption that the hazard of subject subgroup are proportional over follow-up period and therefore the global test indicated that for the data set used the assumption of PH is not violated.

Table 4.24: Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	Chi-square	df	p-value
Proportional Hazard Assumption	2.55	7	0.92

Graphical test for Proportional Hazard Assumption

Result of the graph, Figure 4.10, comparing patient's gender while adjusting for age, zone, clinics, admission and surgery status, and treatment outcome shows that the two line (male and female) are parallel to each other and therefore substantiate the claim that the proportional assumption is valid for the data.

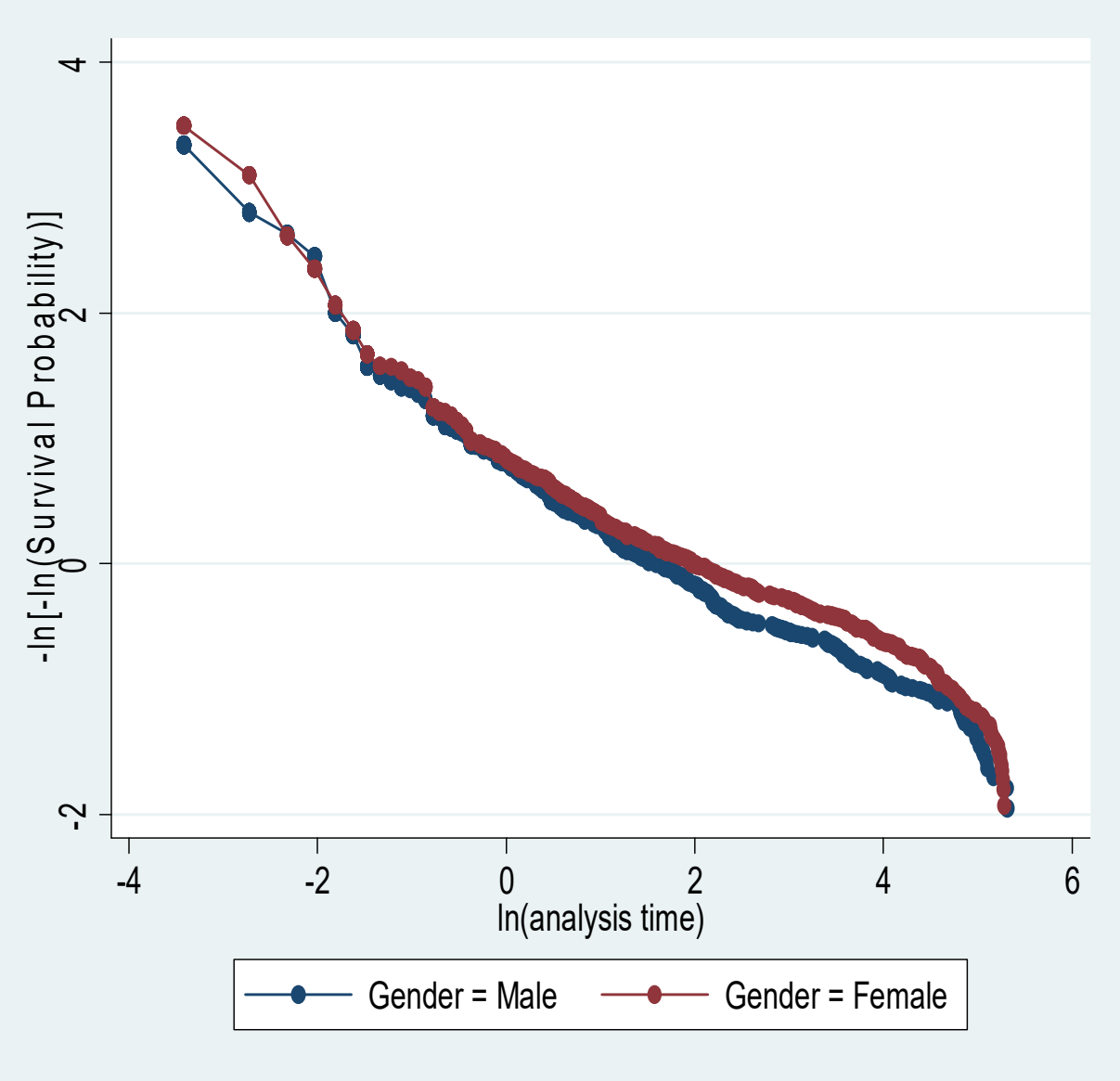


Figure 4.10 Graph Showing Violation of Proportional Hazard Assumption.

Fitting Cox Proportional Hazard Model

Variable	factor	H_Ratio	z	p>|z|	95% CI

Table 4.25 below shows the cox regression analysis that succeed the global test above. It can be inferred from the table that record dormancy time were affected by patients age (HR=0.90, p-value = 0.000), gender (HR=0.84, p-value = 0.014), State of residence (HR=0.87, p-value = 0.037), surgery status (HR=0.72, p-value = 0.002) and treatment outcome (HR=1.87, p-value = 0.000) as they are significant at both 1% and 5% level respectively. Whereas factors like patients admission status (HR=.0.90, p-value = 0.213) and type of clinic attended (HR=0.098, p-value = 0.392) did not influence record dormancy time as they are both insignificant.

Age group		**0.90**	**-4.06**	**0.00**	**0.85**	**0.94**
	60+	0.57	-4.20	0.00	0.44	0.74
	31-60	0.62	-4.25	0.00	0.50	0.77
	21-30	0.73	-2.66	0.00	0.57	0.92
	10-20	0.77	-2.01	0.04	0.59	0.99
	<10 years (rc)					
Gender		**0.84**	**-2.47**	**0.01**	**0.74**	**0.94**
	female	0.84	-2.40	0.01	0.72	0.96
	Male (rc)					
State of Residence		**0.87**	**-2.037**	**0.03**	**0.76**	**0.99**
	Oyo	0.85	-2.36	0.01	0.74	0.97
	Others (rc)					
clinics		**0.98**	**-0.86**	**0.39**	**0.94**	**1.02**
	OTHERS	0.92	0.84	0.39	0.76	1.11
	GYNE	0.98	-0.14	0.89	0.77	1.25
	CHOP	0.66	-2.45	0.01	0.48	0.92
	SOP	1.03	0.31	0.75	0.84	1.26
	MOP (rc)					
Patient Admitted		**0.90**	**-1.24**	**0.21**	**0.78**	**1.05**
	Yes	0.89	-1.34	0.75	0.77	1.04
	No (rc)					
Surgery		**0.72**	**-3.07**	**0.00**	**0.59**	**0.89**
	Yes	0.70	-1.34	0.17	0.77	1.04
	No (rc)					
Treatment Outcome		**1.87**	**3.87**	**0.00**	**1.36**	**2.57**
	Died	3.43	3.77	0.00	1.80	6.54
	Alive (rc)					

Table 4.25: Cox Regression of Dormancy Time on Patients Characteristics

4.4.5.2 Parametric approach

Fitting Exponential Model:

Variable	factor	H_Ratio	z	p>|z|	95% CI

Giving that the recorded dormancy time data is skewed distributed data, we regress dormancy time on patients characteristics based on exponential model assumption of parameter $\lambda=1$. Here only patient clinic type was marginally significant (HR=0.95, p-value = 0.05) while other patient characteristics like gender, HR=0.65, admission, HR0.79, and surgery status, 0.58 and treatment outcome were all significantly ($P<0.01$) influence their record dormancy time. This generally imply that record of older female patient admitted and discharge against medical advice after surgery will become dormant earlier than younger male patient that are alive as at time of last contact. Table 4.26 below shows the Exponential regression model:

Age group		**0.838**	**-7.14**	**0.00**	**0.79**	**0.87**
	60+	0.34	-8.09	0.00	0.26	0.44
	31-60	0.37	-8.76	0.00	0.30	0.47
	21-30	0.50	-5.68	0.00	0.39	0.63
	10-20	0.51	-5,18	0.00	0.39	0.66
	<10 years (rc)					
Gender		**0.65**	**-6.20**	**0.00**	**0.57**	**0.75**
	female	0.63	-6.11	0.00	0.55	0.73
	Male (rc)					
State of Residence		**0.76**	**-4.10**	**0.00**	**0.66**	**0.86**
	Oyo	0.6979849	-5.22	0.00	0.60	0.79
	Others (rc)					
clinics		**0.95**	**-1.96**	**0.50**	**0.91**	**0.99**
	OTHERS	0.80	-2.15	0.03	0.66	0.98
	GYNE	0.99	-0.07	0.944	0.77	1.26
	CHOP	0.44	-4.86	0.00	0.31	0.61
	SOP	1.02	0.25	0.80	0.83	1.26
	MOP (rc)					
Patient Admitted		**0.79**	**-3.05**	**0.00**	**0.68**	**0.92**
	Yes	0.74	-3.73	0.00	0.63	0.87
	No (rc)					
Surgery		**0.58**	**-5.27**	**0.00**	**0.47**	**0.71**
	Yes	0.55	-5.28	0.00	0.44	0.69
	No (rc)					
Trt_Outcome		**4.26**	**8.95**	**0.00**	**3.10**	**5.85**
	Died	17.54	8.74	0.00	9.22	33.34
	Alive (rc)					
_cons	**variable**	0.06	-1253	0.00	0.04	0.09
	categories	0.22	-10.51	0.00	0.16	0.29

Table 4.26 Exponential Model of Dormancy Time on Patient Characteristics

Fitting Weibull Model

We fit Weibull model to the skewed distribute dormancy time data under the assumption that the exponential model fail and the model fit Weibull model of parameter ϒ=λ=1. Similarly patient age (HR=0.90, p<0.001) gender (HR=.0.81, p<0.01), state (HR=0.87, p<0.10), surgery (HR=0.72, p<0.01) and treatment outcome (HR=1.94, p<0.01) significantly influence patient record dormancy time. However patient's admission status and type of clinic attended will not determine patient record dormancy time. Table 4.27 below shows the Weibull regression model result:

Table 4.27: Weibull Regression Model of Dormancy Time on Patient Characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.90**	**-4.05**	**0.00**	**0.85**	**0.94**
	60+	0.58	-4.15	0.0	0.45	0.75
	31-60	0.60	-4.57	0.00	0.48	0.75
	21-30	0.72	-2.74	0.00	0.57	0.91
	10-20	0.74	-2.25	0.02	0.58	0.96
	<10 years (rc)					
Gender		**0.81**	**-3.10**	**0.00**	**0.71**	**0.92**
	female	0.81	-2.83	0.00	070	0.93
	male (rc)					
State of Residence		**0.87**	**-2.01**	**0.04**	**0.76**	**0.99**
	Oyo	0.85	-2.36	0.01	0.75	0.97
	Others (rc)					
clinics		**0.98**	**-0.94**	**0.35**	**0.93**	**1.02**
	OTHERS	0.92	-0.78	0.43	0.76	1.12
	GYNE	0.96	-0.32	0.74	0.75	1.22
	CHOP	0.65	-2.62	0.00	0.47	0.89
	SOP	1.06	0.57	0.56	0.86	1.30
	MOP (rc)					
Patient Admitted		**0.90**	**-1.35**	**0.17**	**0.77**	**1.04**
	Yes	0.89	-1.42	0.15	0.76	1.04
	No (rc)					
Surgery done		**0.72**	**-3.12**	**0.00**	**1.41**	**2.67**
	Yes	0.69	-3.32	0.00	0.56	0.86
	No (rc)					
Treatment Outcome		**1.94**	**4.10**	**0.00**	**1.41**	**2.67**
	Died	3.72	4.02	0.00	1.96	7.08
	Alive (rc)					
_cons	**variable**	0.43	-3.81	0.00	0.27	-0.66
	categories	0.69	-2.58	0.10	0.53	0.91
/1n_p	**variable**	-0.72	-29.50	0.00	-0.77	-0.68
	categories	-0.72	-29.24	0.00	-0.77	-0.67

P	**variable**	0.48			0.45	0.50
1/p		2.07			1.97	2.17
	categories	0.48			0.46	0.50
		2.06			1.964	2.16

4.5Cohort 3: Patient records created between 2000 and 2004

Between 1st January 2000 and 31st December, 2004, the number of patient records that were created was 87,902 out of which 1537 was selected for the study out of which not less than 274 (17.8%) were found to be inactive (dormant) on the first day creation i.e. they never used beyond the day of creation as indicated by a single entry in the medical records.

4.5.1 Frequency distribution of some demographic and clinical characteristics of the patients

Table 4.28, shows some socio-demographic characteristics of the1263 records that survived beyond the first day of creation. Result shows that most, 42.23%, of the patient were between the ages of 31-60 years, while 10.46% were below the age of 10years, 10.54% were between 10-20 years of age and 18.86% were aged above 60 years. Male patients constitute 51.77%, and 54.17% of the patient wereresident in Oyo State, records from MOP clinicconstitute 52.73% from the total of 1246 patient records, 1.77% belongs toSOP clinic, 0.40%fromCHOP clinic while 44.54% of the recordswere from other clinics. Twenty-two percent(22%) of the patients were admitted and only 3.09% of patients were ever operatedon. Almost (99.68%) all the patients were alive at the time of last entry/contact, 0.32% were discharge against medical advice and no died recorded for the period.

Table 4.28 Frequency distribution of patient's characteristics 3rd cohort 2000-2004

Variables n=1263	Level	Freq.	Percent	Cum
Age at Registration	<10	132	10.46	10.46
	10-20	133	10.54	21.00
	21-30	226	17.91	38.91
	31-60	533	42.23	81.14
	61+	238	18.86	100.00
Gender	male	644	51.77	51.77
	female	600	48.23	100.00
State of residence	Oyo State	662	54.17	54.17
	Others	560	45.83	100.00
Clinic attended	MOP	657	52.73	52.73
	SOP	22	1.77	54.49
	CHOP	5	0.40	54.90
	GYNE	7	0.56	55.46
	Others	555	44.54	100.00
Ever admitted	No	984	77.91	77.91
	Yes	279	22.09	100
	Total	**1263**	**100.00**	
Ever operated on	No	1224	96.91	96.91
	Yes	39	3.09	100.00
Treatment outcome	Alive	1259	99.68	99.68
	Died	4	0.32	100.00
	DAMA	-	-	-

	referred	-	-	-

4.5.2 Frequency distribution of dormancy times for 3rd cohort 2000-2004

Table 4.29, showed that the frequency distribution of the 1263 records in the study that survived the first day of creation. Close to 46.6 % of the records were dormant in less than half of a month, in about 9.5 months, 72.1% of the records were dormantand about 95.5% were dormant at the age of 129.5 months. The distribution is presented graphically in Figure 4.11. The distribution is skewed to the right.

Table 4.29: Frequency distribution of records by dormancy times 3rd cohort 2000-2004

SN	month t*		Dormant records	Percent	Cum. Percent
1.	0-	0.5	443	35.08	3508
2.	1-	3.5	308	24.39	59.46
3.	7-	9.5	84	6.65	66.11
4.	13-	15.5	52	4.12	70.23
5.	19-	21.5	41	3.25	73.48
6.	25-	27.5	23	1.82	75.30
7.	31-	33.5	24	1.90	77.20
8.	37-	39.5	20	1.58	78.78
9.	43-	45.5	22	1.74	80.52
10.	49-	51.5	11	0.87	81.39
11.	55-	57.5	21	1.66	83.06
12.	61-	63.5	25	1.98	85.04
13.	67-	69.5	14	1.11	86.14
14.	73-	75.5	12	0.95	87.09
15.	79-	81.5	6	0.48	87.57
16.	85-	87.5	16	1.27	88.84
17.	91-	93.5	17	1.35	90.18
18.	97-	99.5	15	1.19	91.37
19.	103 -	105.5	9	0.71	92.08
20.	109-	111.5	6	0.48	92.56
21.	115-	117.5	7	0.55	93.11
22.	121-	123.5	6	0.48	93.59
23.	127-	129.5	10	0.79	94.38
24.	133-	135.5	12	0.95	95.33
25.	139-	141.5	17	1.35	96.67
26.	145-	147.5	20	1.58	98.26
27.	151-	153.5	16	1.27	99.52
28.	157-	159.5	3	0.24	99.76
29.	163-	165.6	3	0.24	100.00

Total	1263	100	

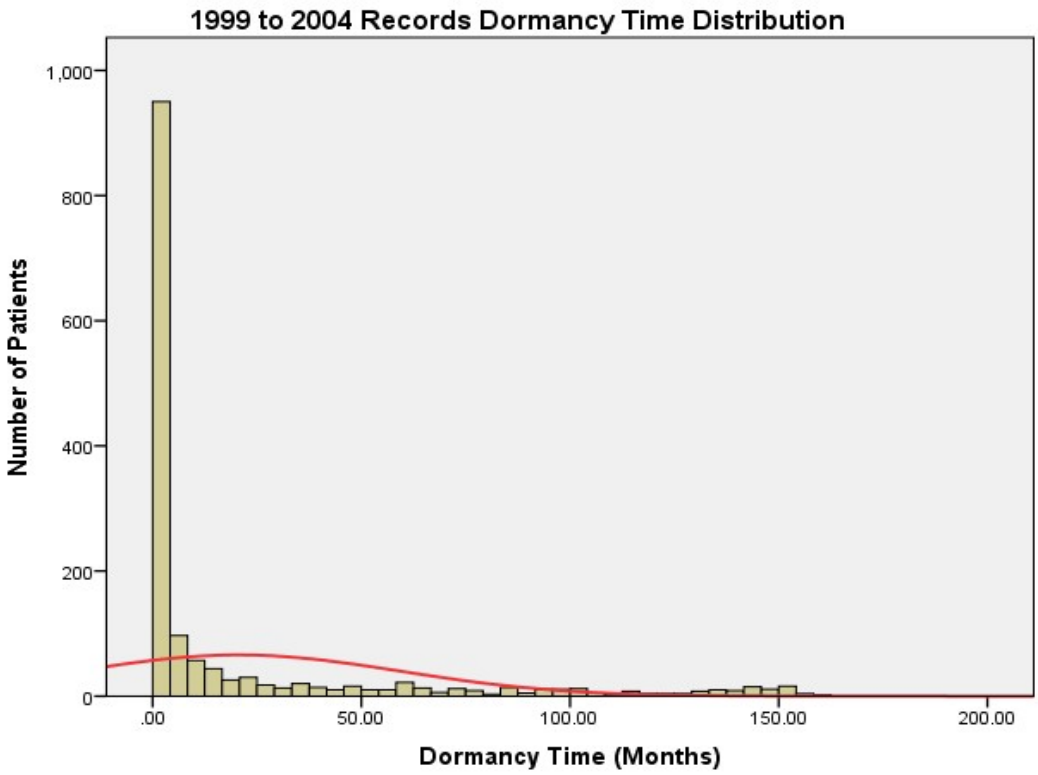


Figure 4.11 Distribution of dormancy times of records

4.5.3 Survival function of dormancy times of records created 2000-2004

Table 4.30 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of records or dormancy time increases and tends toward zero as time reaches end point. Result show that at the dormancy time of approximately 165.5 months, dormancy of records approaches 100%. The results are presnted graphically in Figure 4.12 for the survival curve.

Table 4.30Distribution ofSurvival function of dormancy times 2000-2004

Time (months)	Dormant records	Survival Function	Std. Error	95% CI	
0.5	443	0.97	0.03	0.78	1.00
3.5	308	0.93	0.05	0.75	0.98
9.5	84	0.90	0.06	0.71	0.97
15.5	52	0.86	0.06	0.67	0.95
21.5	41	0.83	0.07	0.63	0.92
27.5	23	0.79	0.08	0.60	0.90
33.5	24	0.76	0.08	0.56	0.88
39.5	20	0.72	0.08	0.52	0.85
45.5	22	0.69	0.09	0.49	0.82
51.5	11	0.66	0.09	0.45	0.80
57.5	21	0.62	0.09	0.42	0.77
63.5	25	0.59	0.09	0.39	0.74
69.5	14	0.55	0.09	0.36	0.71
75.5	12	0.52	0.09	0.33	0.68
81.5	6	0.48	0.09	0.29	0.65
87.5	16	0.45	0.09	0.27	0.62
93.5	17	0.41	0.09	0.24	0.58
99.5	15	0.38	0.09	0.21	0.55
105.5	9	0.34	0.09	0.18	0.51
111.5	6	0.31	0.09	0.16	0.48
117.5	7	0.28	0.08	0.13	0.44
123.5	6	0.24	0.08	0.11	0.41
129.5	10	0.21	0.08	0.08	0.37
135.5	12	0.17	0.07	0.06	0.33
141.5	17	0.14	0.06	0.04	0.29
147.5	20	0.10	0.06	0.03	0.24
153.5	16	0.07	0.05	0.01	0.20
159.5	3	0.03	0.03	0.00	0.15
165.5	3	0.00	.	.	.

Figure 4.12 Survival Curve of dormancy time for3rd cohort 2000-2004

Median dormancy time, standard errors and confidence intervals by patient characteristicsfor Cohort 3

Table 4.31, show median dormancy time (MDT) according to categories of patient characteristics. The resultsshow that MDT for recordsof patients below 10 years of age was2.82 months, records of patients 10-20 years was 4.20 months, 21-30 years was 1.83 months, 31-60 years was 3.21 months and records of thosewith age 60 years and above were dormant in 3.90 months. Records of male patients had a *MDT* of 2.29 months against female with 4.40, while those of patients resident in Oyo State was 4.07 months compared to all other states with 2.29. The *MDT*for record of patients attending MOP was 2.06, SOP 15.1. CHOP 102.76, GYNE 109.20 and 4.17 months in all other clinics combined. The *MDT* of record of patients ever admittedwas1.60 months, as against non-admitted with 3.48, while records of patients with history of surgery were dormant in 8.27 months compared to those without surgery with 3.02.The MDT of records of patients alive as at last contact was 3.05 months, whereas records of patients that died had a MDT of 0.19 months.

Table 4.31 Median-Dormancy-Time (MDT) by Patient Characteristics 3rd cohort (2000-2004) data

Variables n=1263	Level	n	t (months)	Std. Error	95% CI	
		1262	**3.05**	**0.37**	**2.36**	**3.74**
Age at Registration	<10	132	2.82	0.58	1.60	5.68
	10<20	133	4.20	1.18	2.06	7.52
	20-30	226	1.83	0.34	1.14	2.95
	31-60	532	3.21	0.68	2.06	7.52
	61+	238	3.90	0.89	2.33	5.42
Gender	male	644	2.29	0.36	1.54	3.21
	female	599	4.40	0.81	3.05	6.47
State of residence	Others States	560	2.29	0.41	1.47	3.21
	Oyo State	661	4.07	0.74	2.95	5.97
Clinic attended	MOP	656	2.06	0.32	1.54	2.69
	SOP	22	15.11	16.43	1.24	93.96
	CHOP	5	102.76	32.75	1.51	-
	GYNE	7	109.20	108.35	12.45	149.91
	Others	555	4.172	0.70	3.022	5.94
Ever admitted	No	983	3.48	0.52	2.85	4.89
	Yes	279	1.60	0.33	1.14	2.69
Ever operated on	No	1223	3.02	0.38	2.29	3.64
	Yes	39	8.27	4.26	2.6	16.09
Treatment outcome	Alive	1258	3.05	0.38	2.36	3.90
	Died	4	0.19	0.09		-
	DAMA	-	-	-	-	-
	referred	-	-	-	-	-

Selected percentiles of the survival curve (2000-2004)

Survival estimate of time-to-record dormancy time was measured from 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of observed record dormancy time. The 25^{th} percentile dormancy estimate showed that twenty five percent of the records were dormant at 0.49 months while the fifty percent (*MDT*) of records were dormant at 3.05 months as shown from the 50^{th} percentiles. Also, the 75^{th} and 95^{th} percentiles showed that not less than seventy five percent and ninety five percent of records were dormant (inactive) at the 28.45 and 134.34 months respectively. Table 4.32 below shows the respective record survival estimate, their standard error and confidence interval at each percentage point.

Table 4.32

Percentiles	t (months)	Std. Error	95% CI	
25th	0.49	0.03	0.45	0.59
50th	3.05	0.37	2.36	3.74
75th	28.45	3.66	23.75	36.69
95th	134.34	3.89	126.71	141.43
n = 1263				

4.5.4 The hazard curve of dormancy times for records created 2000-2004

Table 4.33 show the distribution of the hazard functions, the standard error and the 95% Confidence Interval. The hazard plotthat follows (Figure 4.13) showed a sharp decrease with age of records at the initial time, t, and continue to decreasese as age of records (dormancy time) increases before a constant and steady movement period between 30 and 130 months then increased with a sharp rise following constant and steady upward movement till it reaches end point and therefore making a bathtop shape.

Table 4.33: Frequency distribution of hazard function (2000-2004) Cohort 3

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	820	444	0.35	0.01	0.33	0.38
5 -	564	255	0.55	0.01	0.52	0.58
10 -	471	93	0.63	0.01	0.60	0.65
15 -	405	66	0.68	0.01	0.65	0.71
20 -	370	35	0.71	0.01	0.68	0.73
25 -	333	37	0.74	0.01	0.71	0.76
30 -	315	18	0.75	0.01	0.73	0.77
35 -	298	17	0.76	0.01	0.74	0.79
40 -	272	26	0.79	0.01	0.76	0.81
45 -	261	11	0.79	0.01	0.77	0.82
50 -	242	19	0.81	0.01	0.79	0.83
55 -	232	10	0.82	0.01	0.80	0.84
60 -	219	14	0.83	0.01	0.81	0.85
65 -	192	26	0.85	0.01	0.83	0.87
70 -	183	9	0.86	0.01	0.84	0.87
75 -	169	14	0.87	0.01	850	0.88
80 -	160	9	0.87	0.01	0.86	0.89
85 -	154	6	0.88	0.01	0.86	0.90
90 -	140	14	0.89	0.01	0.87	0.91
95 -	129	11	0.90	0.01	0.88	0.91
100 -	117	13	0.91	0.01	0.89	0;92
105 -	103	13	0.92	0.01	0.90	0.93
110 -	96	7	0.92	0.01	0.91	0.94
115 -	91	5	0.93	0.01	0.91	0.94
120 -	85	6	0.93	0.01	0.92	0.95
125 -	80	5	0.94	0.01	0.92	0.95
130 -	75	5	0.94	0.01	0.93	0.95
135 -	64	11	0.95	0.01	0.94	0.96
140-	54	10	0.96	0.01	0.95	0.97

145 -	39	15	0.97	0.00	0.96	0.98
150 –	23	16	0.98	0.00	0.97	0.99
155 -	6	17	1.00	0.00	0.99	0.1
160 -	1	5	-	-	-	.

.

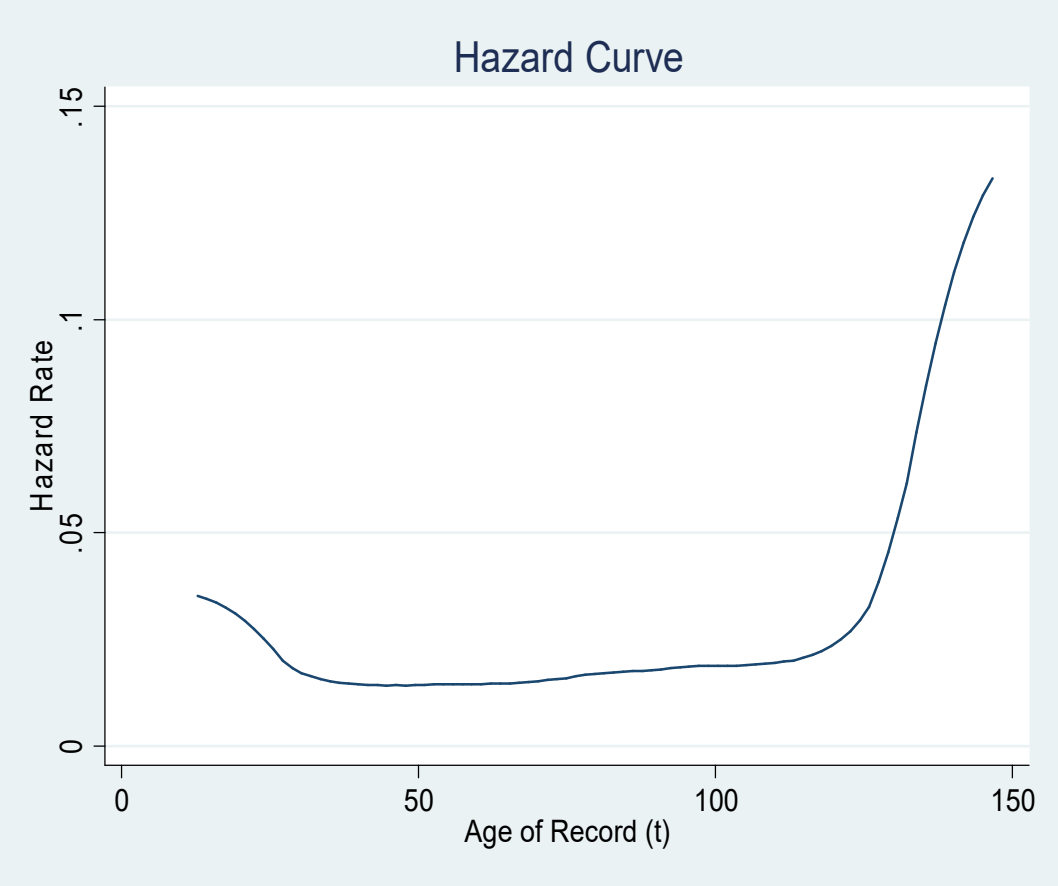


Figure 4.13: Hazard Curve of dormancy time for records created in 3rd cohort 2000-2004

Graphical evaluation of the form of the hazard function of time-to-dormancy

Considering the bathtub shape of the hazard plot, Figure 4.14, typical of a Weibull distribution we tested for validity of this assumption using Weibull probability plot test, a Kaplan-*Meier* log-log Survival curves, *log[H(t)],* against log survival time, *log(t),* Figure 4.16. The plot indicated a straight line relationship between *logH(t)* against *log(t),* increasing monotonically suggesting a Weibull distribution. The intercept was approximately - 0.4782 with a slope of 0.3571. From this the value the shape parameter of γ, for two parameter Weibull distribution was estimated as:

$$\gamma^* = exp(-0.4782) = 0.6198 \text{ and}$$

the estimated hazard rate estimate as:

$$\lambda^* = 0..3251.$$

Since the estimated value of the shape parameter, γ, is less than unity, suggesting a decreasing hazard, λ, of the Weibull distribution.

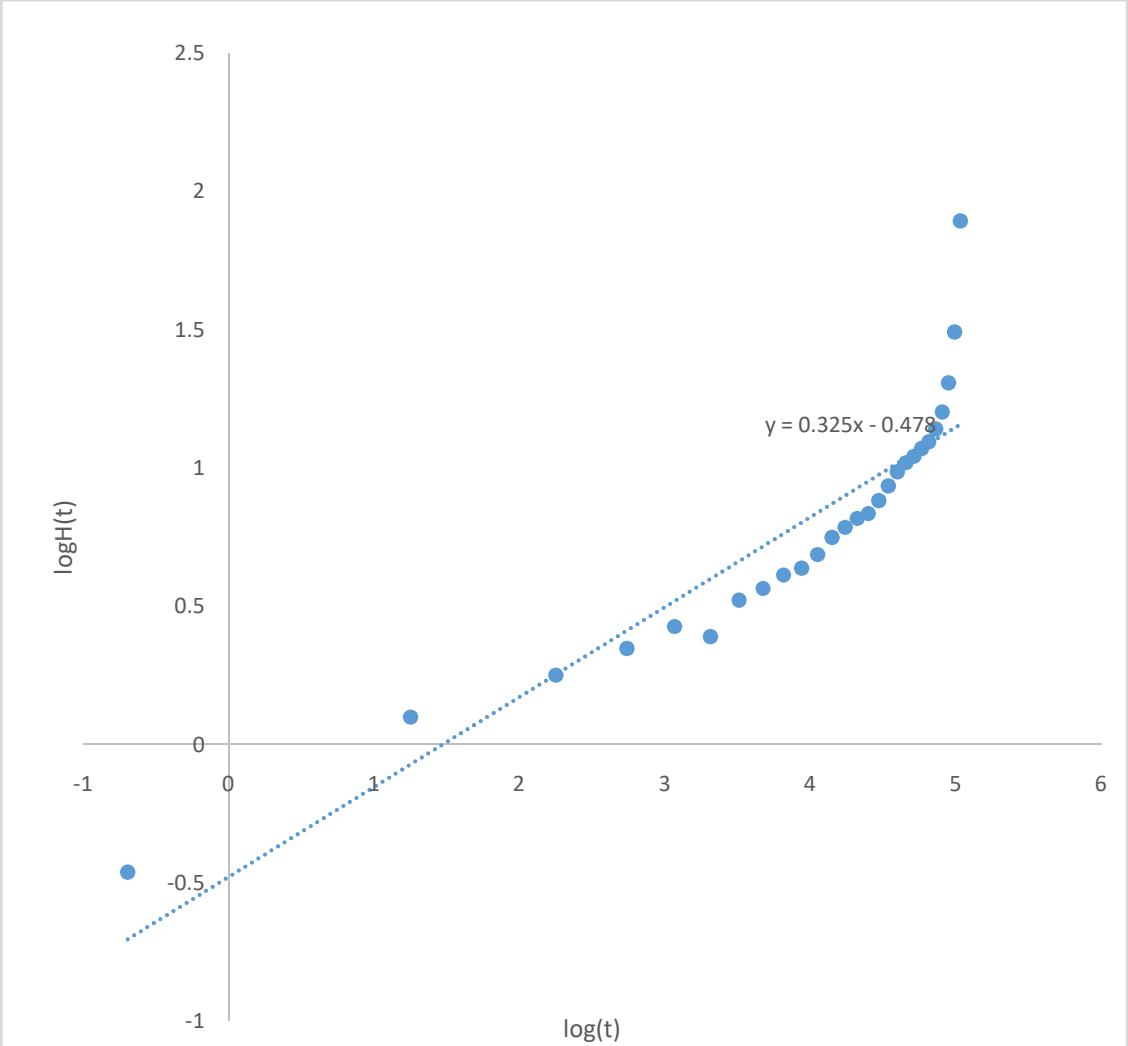


Figure 4.14. A Weibull plot of *log (t)* and *log–log S(t)* with line fitted 2000-2004 3rd cohort

4.5.5Influence of patient characteristics on hazard rate of patient records created between 2000 and 2004 (3rd cohort)

Results of semi-parametric (Cox Proportional Hazard) and Parametric (Exponential and Weibull) survival model used to measure effect of patients demographic and health characteristics on dormancy time of records created between 2000 and 2004 (3rd cohort) show as follow:

4.5.5.1 Non Parametric approach

Schoenfeld Test of Cox Proportional Hazard Model Assumption

Table 4.34 below shows the global test for the proportional hazard assumption. The insignificant (P>0.05) of the test implies that the sample data is valid for the proportional hazard assumption-that the hazard of subject subgroup are proportional over follow-up period and therefore the data can be conveniently analysed using Cox PH model.

Table 4.34: Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	**Chi-square**	**df**	**p-value**
Proportional Hazard Assumption	4.13	7	0.76

Graphical test for Proportional Hazard Assumption

Result, Figure 15, show the graphical test for the proportional hazard assumption by comparing patient's gender while adjusting for age, zone, clinics, admission and surgery status, and treatment outcome shows that the two line (male and female) are parallel to each other and therefore substantiate the claim that the proportional assumption is valid for the data.

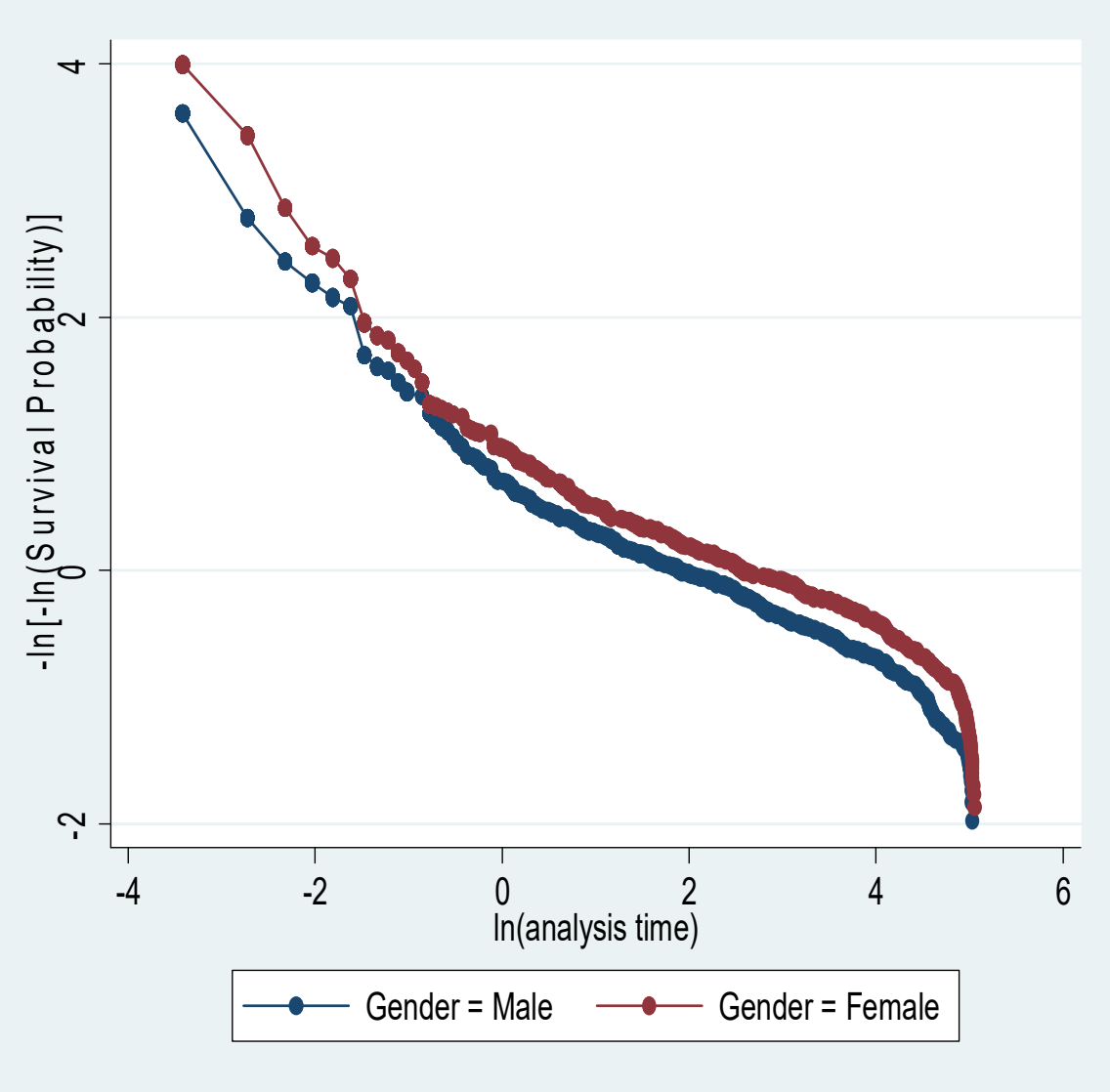


Figure 4.15 Graph Showing violation of Proportional Hazard Assumption.

Fitting Cox Proportional Hazard Model

Table 4.35 below shows the Cox regression analysis that succeed the global test above. It can be inferred from the table that record dormancy time were affected by all patient characteristics, which are; age (HR=0.93, p-value = 0.004), gender (HR=0.74, p-value = 0.000), state of residence (HR=0.83, p-value = 0.002), type of clinic attended (HR=0.95, p-value = 0.005), patient admission status (p-value=0.010), surgery status (p-value = 0.000) and treatment outcome (p-value = 0.003) as they are all significant at 1%, level respectively. Also, hazard ratio that assess the risk magnitude and likelihood are presented in the table for each group of patients characteristic.

Variable	factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.93**	**-2.85**	**0.00**	**0.88**	**0.97**
	60+	0.73	0.010	0.01	0.58	0.93
	31-60	0.80	-2.09	0.03	0.65	0.98
	21-30	1.00	0.02	0.98	0.79	1.26
	10-20	0.86	-1.14	0.25	0.67	1.11
	<10 years (rc)					
Gender		**0.76**	**-4.45**	**0.00**	**0.68**	**0.85**
	female	0.77	-4.33	0.00	0.68	0.86
	male(rc)					
State of Residence		**0.83**	**-3.04**	**0.00**	**0.74**	**0.93**
	Oyo	0.82	-3.27	0.00	0.73	0.92
	Others (rc)					
clinics		**0.95**	**-2.78**	**0.00**	**0.92**	**0.98**
	OTHERS	0.85	-2.55	0.01	0.75	0.96
	GYNE	0.34	-2,76	0.00	0.16	0.73
	CHOP	0.32	-2.20	0.02	0.12	0.88
	SOP	0.64	11.84	0.06	0.40	1.02
	MOP (rc)					
Patient Admitted		**1.21**	**2.57**	**0.01**	**1.04**	**1.41**
	Yes	1.28	3.20	0.00	1.10	1.49
	No (rc)					
Surgery		**0.51**	**-3.49**	**0.00**	**0.35**	**0.74**
	Yes	0.50	3.20	0.00	1.10	1.49
	No (rc)					
Trt_Outcome		**2.13**	**2.97**	**0.00**	**1.29**	**3.50**
	Died	4.42	2.92	0.00	1.63	12.02
	Alive (rc)					

Table 4.35: Cox regression of dormancy time on patient's characteristics

4.5.5.2 Parametric approach

Fitting Exponential Model

Given that the record dormancy time data is skewed distributed data, we regress dormancy time on patients characteristics based on exponential model assumption of parameter $\lambda=1$. Here all patient characteristics like age, HR=0.89; gender, HR=0.63; state of residence, HR=0.76; clinic attended, HR=0.93; admission status, HR=1.3 surgery and treatment outcome all significantly (HR=7.3, $P<0.01$) influence their record dormancy time. This generally implied that record of older female patient admitted, discharge against medical advice or under gone surgery will become dormant earlier than younger male patient that are alive at time of last contact. Table 4.36 below shows the Exponential regression model for the explanatory variable as factors and as sub-categorical factor.

Table 4.36: Exponential regression of dormancy time on patient characteristics 2000-2004 3rd cohort

Variable	factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.89**	**-4.37**	**0.00**	**0.85**	**0.94**
	60+	0.34	-8.09	0.00	0.26	0.44
	31-60	0.37	-8.76	0.00	0.30	0.47
	21-30	0.50	-5.68	0.00	0.39	0.63
	10-20	0.51	-5.18	0.00	0.39	0.66
	<10 years (rc)					
Gender		**0.63**	**-7.68**	**0.00**	**0.56**	**0.7**
	female	0.63	-6.11	0.00	0.55	0.73
	Male (rc)					
State of Residence		**0.76**	**-4.44**	**0.00**	**0.68**	**0.86**
	Oyo	0.69	-5.22	0.00	0.60	0.79
	Others (rc)					
clinics		**0.93**	**-4.32**	**0.00**	**0.90**	**0.96**
	OTHERS	0.80	-2.15	0.03	0.66	0.98
	GYNE	0.99	-0.07	0.94	0.77	1.2622
	CHOP	0.44	-4.86	0.00	0.31	0.61
	SOP	1.02	0.25	0.80	0.83	1.26
	MOP (rc)					
Patient Admitted		**1.31**	**3.56**	**0.00**	**1.12**	**1.52**
	Yes	0.74	-3.73	0.00	0.63	0.87
	No (rc)					
Surgery		**0.44**	**-4.32**	**0.00**	**0.31**	**0.64**
	Yes	0.55	-5.28	0.00	0.44	33.34
	No (rc)					
Trt_Outcome		**7.38**	**7.90**	**0.00**	**4.49**	**12.13**
	Died	17.54	8.74	0.00	9.22	33.34
	Alive (rc)					
_cons	**variable**	**0.02**	**-12.77**	**0.00**	**0.012**	**0.03**
	categories	0.22	-10.51	0.00	0.16	0.29

Fitting Weibull Model to dormancy time data for 2000 and 2004 (3rd cohort)

We further fitted Weibull model to the skewed distributed of time-to-dormancy of medical records for the cohort under the assumption that the exponential model failed and the model fit Weibull model of parameter ϒ=λ=1. Similar to exponential model above, all patient characteristics; age, HR=0.93; gender, HR= 0.77; state of residence, HR, 0.83; clinics, HR=0.95; admission status, HR=1.22; surgery 0.58 and treatment outcome, HR2.15 are significant at P<0.01. Table 4.37 below shows the Weibull regression model result with two significant (P<0.01) extended parameter for the categorical and sub-categorical characteristics:

Table 4.37: Weibull Regression Model of Dormancy Time on Patient Characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.93**	**-2.84**	**0.00**	**0.88**	**0.97**
	60+	0.72	-2.71	0.00		0.91
	31-60	0.78	-2.29	0.02	0'636	0,96
	21-30	1.02	0.17	0.86	0.81	1.28
	10-20	0.82	-1.54	0.12	0.63	1.05
	<10 years (rc)					
Gender		**0.77**	**-4.28**	**0.00**		**0.87**
	female	0.78	-4.08	0.00	0.69	0.89
	male (rc)					
State of Residence		**0.83**	**-3.06**	**0.00**	**0.92**	**0.98**
	Oyo	0.81	-3.37	0.00	0.72	0.91
	Others (rc)					
clinics		**0.95**	**-2.76**	**0.00**	**0.92**	**0.98**
	OTHERS	0.84	-.2.61	0.00	0.7521275	0.96
	GYNE	0.36	-2.61	0.00	0.17	0.77
	CHOP	0.29	-2.42	0.01	0.10	0.79
	SOP	0.58	-2.27	0.02	0.36	0.93
	MOP (rc)					
Patient Admitted		**1.22**	**2.62**	**0.00**	**1 .051**	**1.41**
	Yes	1.28	3;23	0.00	1.10	1.49
	No (rc)					
Surgery done		**0.59**	**-2.77**	**0.00**	**0.41**	**0.86**
	Yes	0.58	-2.87	0.00	0.40	0.84
	No (rc)					
Treatment Outcome		**2.15**	**3.02**	**0.00**	**1.30**	**3.54**
	Died	4.59	3.00	0.00	1.69	12.45
	Alive (rc)					
_cons	**variable**	0.32	-3.73	0.00	0.18	0.58
	categories	0.49	-5.94	0.00	0.38	0.62
/1n_p	**variable**	-0.72	-32.44	0.00	-0.77	-0.68
	categories	-0.71	-31.99	0.00	-0.76	-0.67
P	**variable**	0.48			0.46	0.50
1/p		2.07			1.98	2.16
	categories	0.48			0.46	0.51

		2.04			1.96	2.14

4.6Cohort 4: Records created between 1st January 2005 and 31st December2009

The analysis for the 1077 records that survived beyond the first day o contact, indicated by second entry in the medical record was carried out in line with the previous cohorts.

4.6.1 Frequency distribution of some demographic and clinical characteristics of the patients

Table 4.38show that revealed that 41.04% of the records were those of patients 31-60 years old, 9.72%belong to patients aged 10-20,patients whose age were below 10 years constitute15.75%, whilepatients above 61 years of age made up 19.81%. Male patients constitute 55.52%, and 76% of the patients were residence in Oyo State. Records from Medical Outpatient Clinics were 29.31%, Surgery Outpatient Clinics 1.89%, Children Outpatient clinic 0.85% while 66.73were from other clinics. About 25.96% of patients were ever admitted and 11.72% of the whole patients ever went through surgical operation, Almost all the patients observed for time-to-dormancy were (98.21%) were alive as at time of last entry/contact, 1.69%) were discharge against medical advice and only 1 patient died during the period. Table 4.32 below shows details of the patients' socio-demographic and health characteristics based on their dormancy time.

Table 4.38 Distribution of patient's characteristics 4th cohort 2005-2009

Variables n=1077	Level	Freq.	Percent	Cum
Age at Registration	<10	167	15.75	15.75
	10-20	103	9.72	25.47
	20-30	145	13.68	39.15
	31-60	435	41.04	80.19
	61+	210	19.81	100.00
Gender	male	588	55.52	55.52
	female	471	44.48	100.00
State of residence	Oyo State	788	75.84	75.84
	Others	251	24.16	100.00
Clinic attended	MOP	311	29.31	29.31
	SOP	20	1.89	31.20
	CHOP	9	0.85	32-05
	GYNE	13	1.23	33.27
	Others	708	66.73	100.00
Ever admitted	No	793	74.04	74.04
	Yes	278	25.96	100.00
Ever operated on	No	949	88.28	88.28
	Yes	126	11.72	100.00
Treatment outcome	Alive	1044	98.21	98.21
	Died	1	0.09	98.31
	DAMA	18	1.69	100.00
	referred	-	-	-

4.6.2 Distribution of dormancy times for 4th cohort(2005-2009) data

Table 4.39 showed the frequncy distribution for the 1077 records in the studythat survived beyond the first day of creation.Above 50% of the records were already dormant at 3.5 months of creation, 75% were dormant at the record age of 27.5 months and 95% at the end of 84 months. The distribution is presented graphically in Figure 4.16. The distribution is skewed to the right.

Table 4.39 Frequency distribution of dormancy times 4th cohort - 2005-2009

SN	Month t*		Dormant records	Percent	Cum. percent
1.	<1	0.5	345	32.03	32.03
2.	1-6	3.5	281	26.09	58.12
3.	7-12	9.5	92	8.54	66.67
4.	13-18	15.5	57	5.29	71.96
5.	19-24	21.5	38	3.53	75.49
6.	25-30	27.5	41	3.81	79.29
7.	31-36	33.5	37	3.44	82.73
8.	37-42	39.5	11	1.02	83.75
9.	43-48	45.5	23	2.14	85.89
10.	49-54	51.5	20	1.86	87.74
11.	55-60	57.5	24	2.23	89.97
12.	61-66	63.5	14	1.30	91.27
13.	67-72	69.5	10	0.93	92.20
14.	73-78	75.5	17	1.58	93.78
15.	79-84	81.5	13	1.21	94.99
16.	85-90	87.5	21	1.95	96.94
17.	91-96	93.5	17	1.58	98.51
18.	97-102	99.5	7	0.65	99.16
19.	103 -108	105.5	4	0.37	99.54
20.	109-114	111.5	2	0.19	99.72
21.	115120	117.5	2	0.19	99.91
22.	121-126	123.5	1	0.09	100.
Total			1077	100	

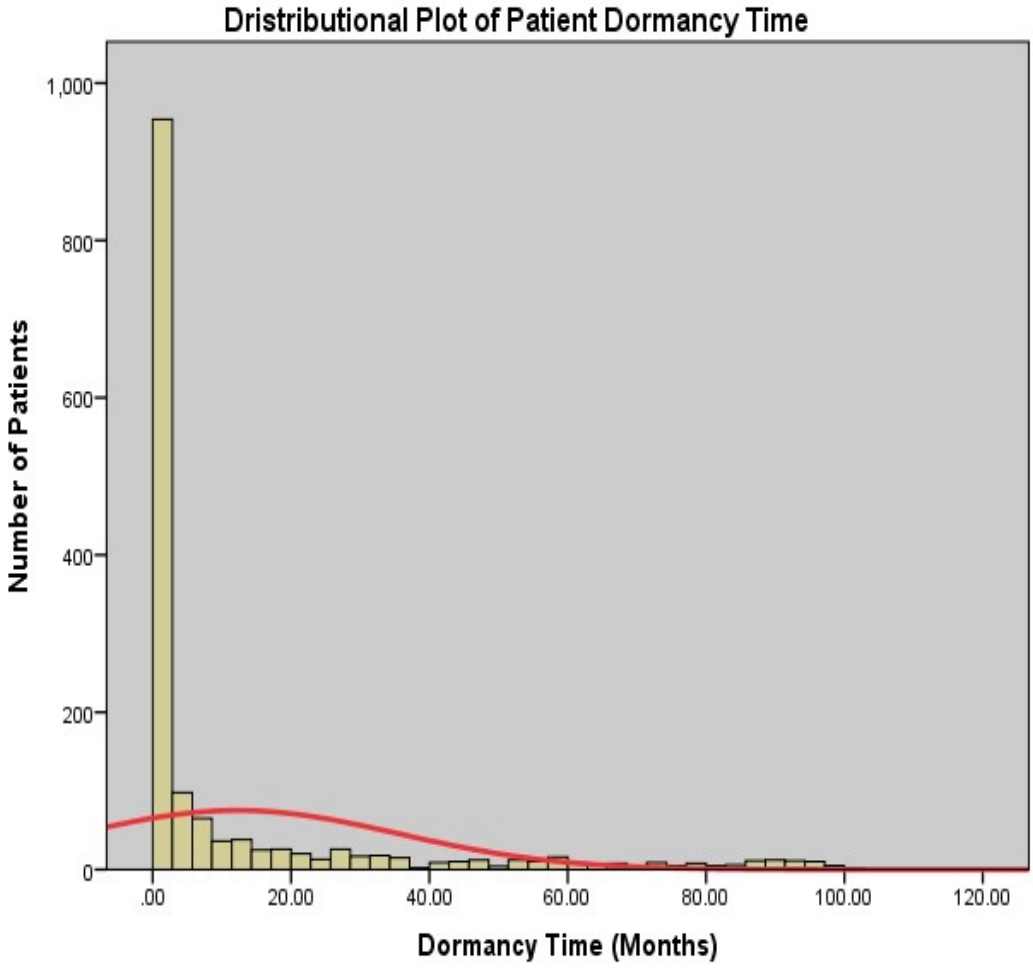


Figure 4. 16Frequency distribution of dormancy times 4th cohort - 2005-2009

4.6.3 Survival function of dormancy times for 2005-2009 data

Table 4.40 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of records or dormancy time increases and tends toward zero as time reaches end point. Result show that at the dormancy time of approximately 123.5 months, dormancy of records approaches 100%. The results are presnted graphically in Figure 4.17 for the survival curve.

Table 4.40Distribution ofSurvival function of dormancy times 2005-2009

Time	Dormant records	Survival Function	Std. Error	95% CI	
0.5	345	0.95	0.04	0.72	0.99
3.5	281	0.91	0.06	0.68	0.98
9.5	92	0.86	0.07	0.63	0.95
15.5	57	0.82	0.08	0.59	0.93
21.5	38	0.77	0.09	0.54	0.90
27.5	41	0.73	0.10	0.49	0.87
33.5	37	0.68	0.10	0.45	0.83
39.5	11	0.64	0.10	0.40	0.80
45.5	23	0.59	0.10	0.36	0.76
51.5	20	0.55	0.11	0.32	0.72
57.5	24	0.50	0.11	0.28	0.68
63.5	14	0.45	0.11	0.24	0.64
69.5	10	0.41	0.10	0.21	0.60
75.5	17	0.36	0.10	0.17	0.56
81.5	13	0.32	0.10	0.14	0.51
87.5	21	0.27	0.10	0.11	0.46
93.5	17	0.23	0.09	0.08	0.41
99.5	7	0.18	0.08	0.06	0.36
105.5	4	0.14	0.07	0.03	0.31
111.5	2	0.09	0.06	0.02	0.25
117.5	2	0.05	0.04	0.00	0.19
123.5	1	0.00	.	.	.

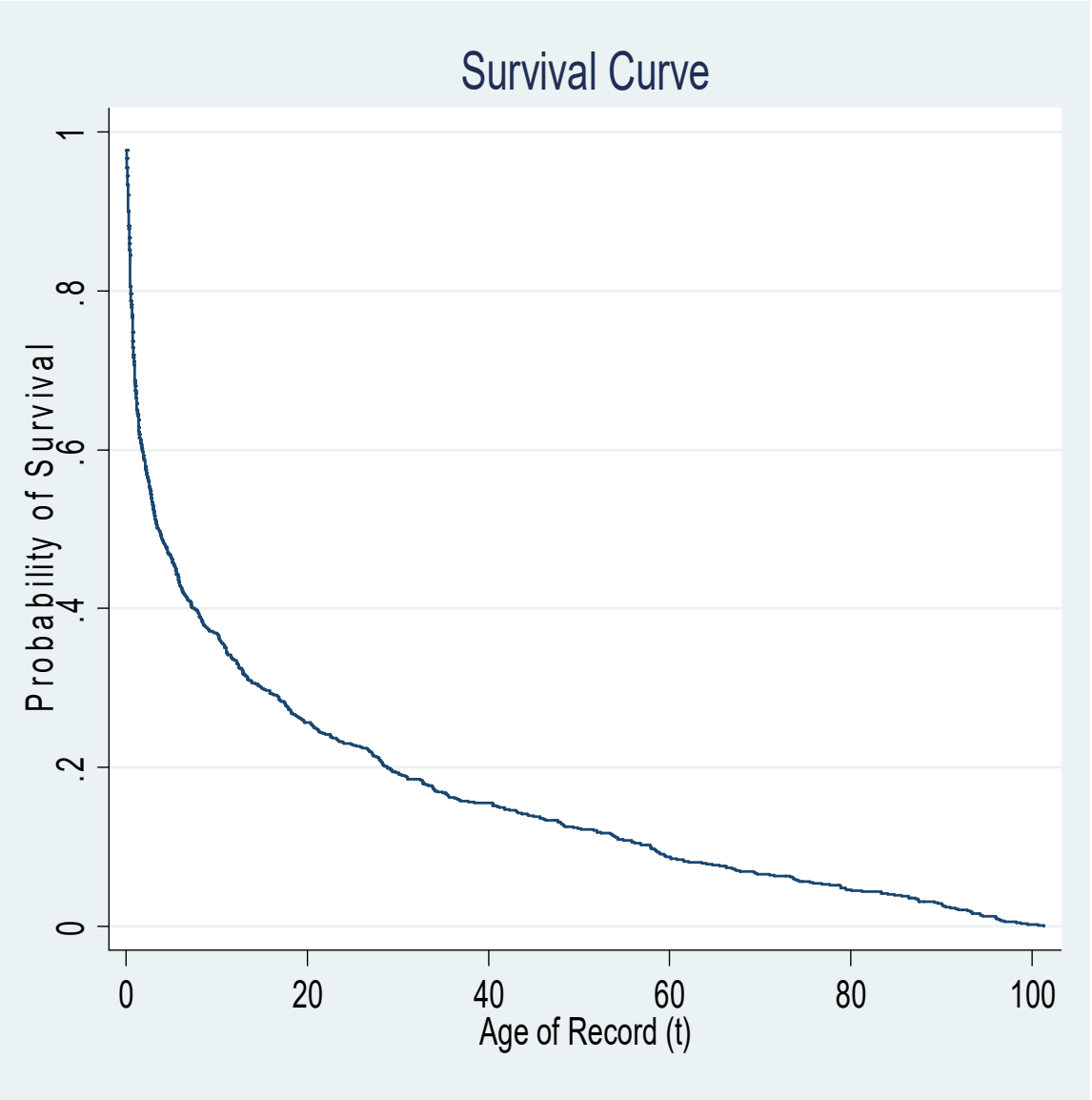


Figure 4.17: Survival Curvedormancy time for 4th cohort 2005-2009

Median dormancy times, standard error and confidence intervals by patient chracteristics

Table 4.41, show the median dormancy time (MDT)according to categories of patients characteristics

The median dormancy time for records of patient less than 10 years of agewas 2.98 months,records of patients that arebetween 10-20 years was observed to be 2.95 months. Patients 31-60 years and those above 60 years of age category had *MDT* of 5.28 and 5.74 months respectively. Male patient records *MDT* was observed to be 2.75 lower than female with 5.51 months,records of patients from other States *MDT*was 2.95 months as against those of patient's residence in Oyo State with 5.57 months. With respect to clinics, CHOP records had the highest*MDT*of 61.20 months and MOP records lowest with *MDT* of 2.23 months The *MDT* of records from SOP and GYNE were however 8.70 and 48.06 months respectively. MDT forrecords of admitted patients was 3.05 months, which was slightly lower than that never admitted patientswith 4.00 months. Records of patients with history of surgery had a *MDT of* 8.34 months as against 3,21 months for records of patients without history of surgery that was 3.21 months. Record of patients alive as at last contact had a longer *MDT* of 4.00 months against DAMA with a *MDT* of 0.29 months.

Table 4.41 Median-Dormancy-Time by Patient Characteristics 4th cohort 2005-2009

Variables n=1077	Level	n	t (months)	S. Error	95% CI	
		1075	**3.84**	**0.44**	**2.98**	**4.96**
Age at Registration	<10	167	2.98	1.19	1.47	6.24
	10-20	103	2.95	0.77	1.37	4.96
	21-30	145	1.18	0.34	0.78	2.23
	31-60	434	5.28	0.75	3.31	6.34
	61+	210	5.28	0.75	3.31	9.56
Gender	male	586	2.75	0.39	2.16	3.74
	female	471	5.51	0,88	4.17	7.65
State of residence	Others States	250	2.95	0.57	2.29	4.96
	Oyo State	787	4.17	0.53	3.12	5.45
Clinic attended	MOP	311	2.23	0.50	1.37	3.38
	SOP	20	8.70	6.94	0.91	44.32
	CHOP	9	61.20	4.99	13.40	75.82
	GYNE	13	48.06	14.31	5.45	73.56
	Others	706	4.23	0.51	3.05	5.42
Ever admitted	No	792	4.00	0.51	3.02	5.28
	Yes	277	3.05	0.87	2.06	5.45
Ever operated on	No	948	3.21	0.43	2.52	4.17
	Yes	125	8.34	2.13	5.74	11.79
Treatment outcome	Alive	1042	4.00	0.44	3.12	5.19
	Died	1	-	-	-	-
	DAMA	18	0.29	0.24	0.06	1.37
	referred	-	-	-	-	-

Selected percentiles of the survival distribution

Estimate of specific points of time-to- dormancy for patient records was measured for 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of observed Tables 4.42 below shows the respective patients record dormancy time estimate, their standard error and confidence interval at each percentile point.

. The 25^{th} percentile survival estimate showed that that twenty five percent of the records were dormant in 0.69 months while the fifty percent (Median survival time) of records were 3.84 months as shown from the 50^{th} percentiles. Also, the 75^{th} and 95^{th} percentiles showed that not less than seventy five percent and ninety five percent of records were dormant (inactive) in 23.65 and 84.07 months respectively.

Table 4.42Selected percentiles of survival curve 4th cohort 2005-2009

Percentiles	t (months)	Std. Error	95% CI	
25th	0.68	0.05	0.59	0.78
50th	3.84	0.44	2.98	4.96
75th	23.65	1.97	19.25	28.12
95th	84.07	2.71	78.58	88.04
n = 1077				

4.6.4 Hazardplot of time to dormancy for patient records created 2005-2009

Table 4.43 show the distribution of the hazard functions, standard errors and the 95% Confidence Intervals. The hazard plot that follows, Figure 4.18, showed that the hazard rate was high at the initial time point (at creation of patient records) and continue to decreasese gradually as age of (time-to-dormancy) increases until it reaches time point of 20 months, then becomes constant and steady movement until at 80 months then increase with a sharp upward rise till it reaches end point and therefore making a bathtop shape.

Table 4.43 : Frequency distribution of hazard function, SE and CI(2005-2009) Cohort 4

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	733	345	0.32	0.01	0.29	0.35
5 -	507	225	0.53	0.02	0.50	0.56
10 -	409	97	0.62	0.01	0.59	0.65
15 -	339	70	0.69	0.01	0.66	0.71
20 -	295	45	0.72	0.01	0.70	0.75
25 -	263	31	0.76	0.01	0.73	0.78
30 -	226	37	0.79	0.01	0.77	0.81
35 -	198	28	0.82	0.01	0.79	0.84
40 -	183	15	0.83	0.01	0.81	0.85
45 -	166	17	0.85	0.01	0.82	0.87
50 -	149	17	0.86	0.01	0.84	0.88
55 -	131	18	0.88	0.01	0.86	0.90
60 -	110	21	0.90	0.01	0.88	0.92
65 -	98	12	0.91	0.01	0.89	0.93
70 -	86	12	0.92	0.01	0.90	0.94
75 -	74	12	0.93	0.01	0.92	0.95
80 -	63	11	0.94	0.01	0.93	0.96
85 -	53	10	0.95	0.01	0.94	0.96
90 -	38	15	0.97	0.00	0.95	0.98
95 -	15	23	0.99	0.00	0.980	0.99
100 -	4	12	1.0	0.00	0.99	0.1
105 -	1	2	-	-	.-	-

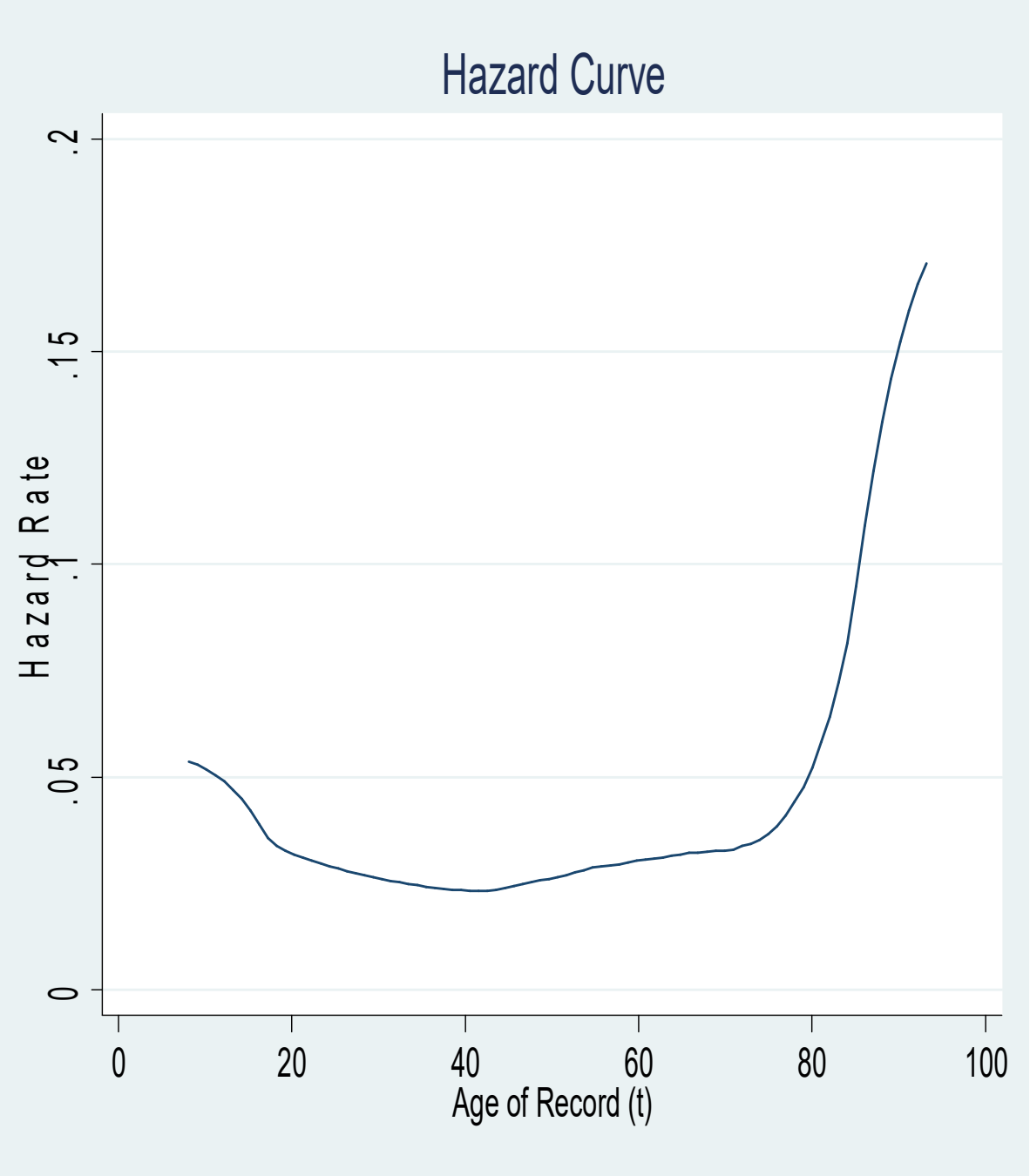


Figure 4.18 Hazard Curveof dormancy time for 4th cohort 2005-2009

Graphical evaluation of the form of the hazard function

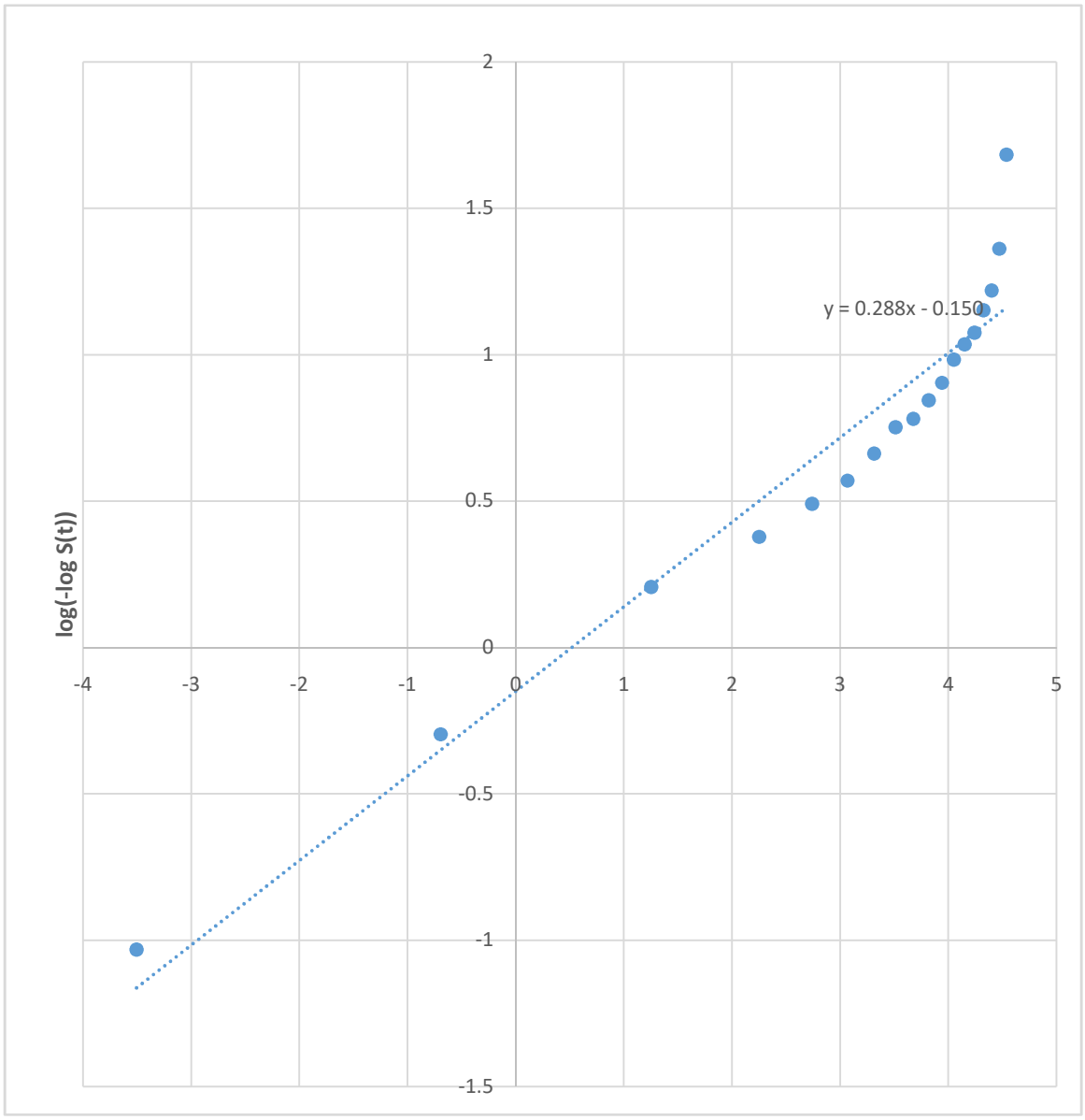
Considering the bathtub shape of the hazard plot, typical of a Weibull distribution we further tested for validity of Weibull distribution assumption by using the Weibull probability plot of Kaplan-*Meier* log-log Survival curves, *log[H(t)],* against log survival time, *log(t),* Figure 4.19. The plot indicated a straight line relationship between *logH(t)* against *log(t),* increasing monotonically suggesting a Weibull distribution. The intercept the fitted line was approximately - 0.1501 with a slope of 0.2888. From this value the shape parameter, γ, and hazard rate for two parameter Weibull distribution was estimated as:

$$\gamma^{*} = exp(-0.1501) = 0.8606$$

and

$$\lambda^{*} = 0.2888.$$

Since the estimated value of γ, the shape parameter of the Weibull distribution is less than unity, suggesting a decreasing hazard, *λ,* the time-to-dormancy of medical records we assumed that the *TTD* data of patient records created from 2005-2009 follows Weibull distribution.



log(t)

Figure 4.19.A Weibull plot of *log (t)* and *log–log S(t)* with line fitted

4.6.5 Influence of patient characteristics on hazard rate,of patient records created between 2000 and 2004 (cohort 4)

4.6.5.1 Non Parametric approach

Results of semi-parametric (Cox Proportional Hazard) model used to measure effect of patient's demographic and clinical characteristics on dormancy time of records created between 2005 and 2009 (4^{th} cohort) show as follow:

Schoenfeld Test of Cox Proportional Hazard Model Assumption:

Table 4.44 below shows the global test for the proportional hazard assumption. The significant result of the test implies that the sample data violate the proportional hazard assumption-that the hazard of subject subgroup are proportional over follow up period and therefore the null hypothesis was rejected.

Table 4.44 Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	**Chi-square**	**df**	**p-value**
Proportional Hazard Assumption	19.39	7	0.00

Graphical test for Proportional Hazard Assumption

Figure 20 show the result of the graph comparing patient's gender while adjusting for age, state of residence and clinics shows that the two line (Male and Female) intersect and also indicated that the proportional assumption is not valid for records dormancy time data created in 2005-2009:

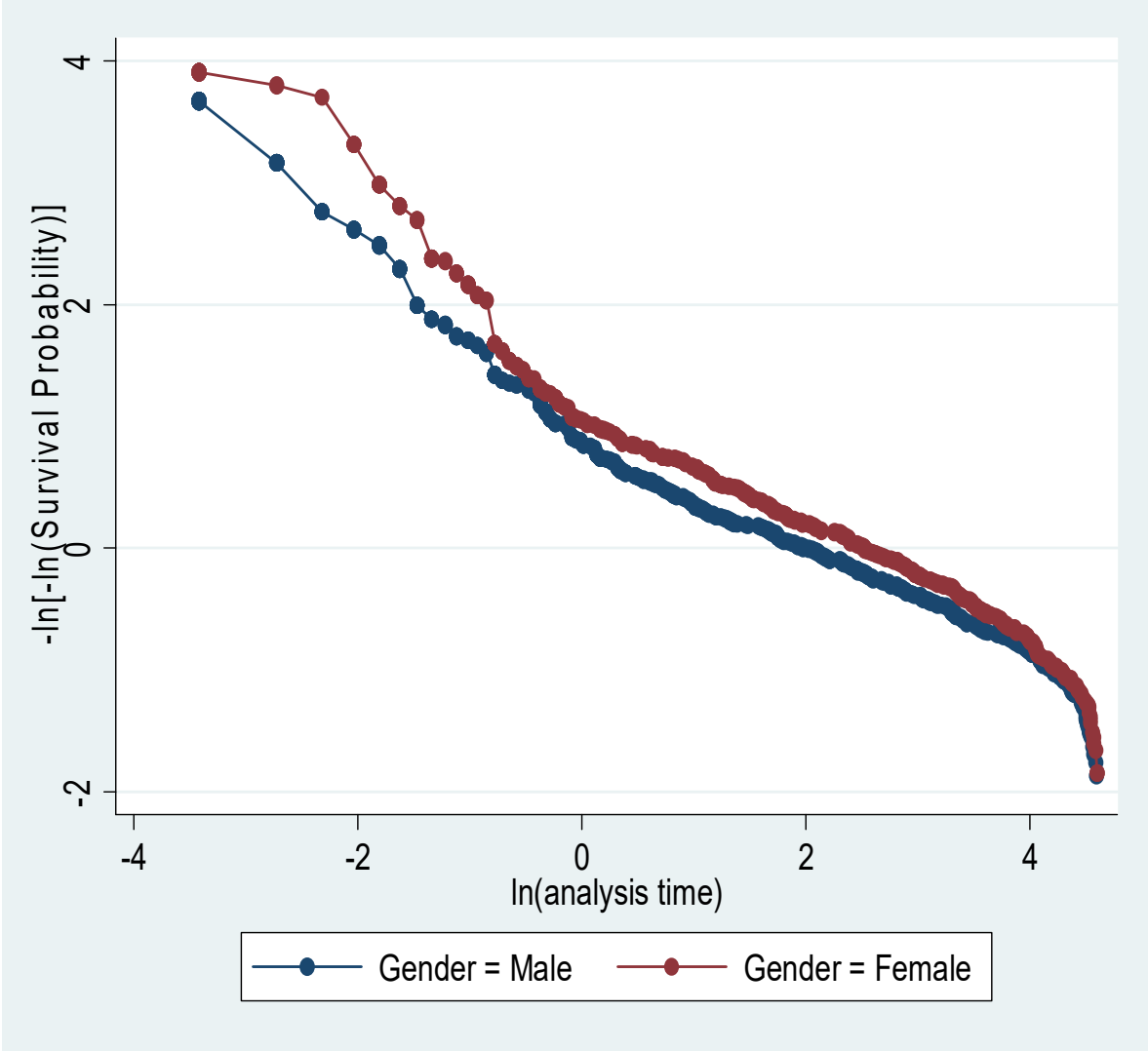


Figure 4.20: Graph Showing Violation of Proportional Hazard Assumption 4th cohort 2005-2009

Fitting Cox Proportional Hazard Model

Table 4.45 below shows the cox regression analysis that succeed the global test above. Results from the table that record dormancy time were affected by gender (HR=0.84. p-value = 0.013), patient admission status (HR=1.36, p-value = 0.002), surgery status (HR=0.60, p-value = 0.000) and treatment outcome (HR=1.76, p-value = 0.000) as they were significant at 5%, 1%, 1% and 1% respectively. Whereas, patients Age (HR=0.96, p-Value = 0.206), State of residency (0.90, p-value = 0.174) and Clinic attended (HR=1.00, p-Value = 0.911) will not influence their record dormancy time as they were all insignificant and indicating failure to accept the research hypothesis.

Table 4.45: Cox regression of dormancy time on patient's characteristics 2005-2009 4th cohort

Variable	factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.96**	**-1.26**	**0.20**	**0.92**	**1.01**
	60+	0.91	-0.82	0.41	0.73	1.13
	31-60	0.87	-1.32	0.18	0.71	1.49
	21-30	1.16	1.20	0.23	0.90	1.49
	10-20	1.14	0.98	0.32	0.87	1.48
	<10 years (rc)					
Gender		**0.84**	**-2.49**	**0.01**	**0.74**	**0.96**
	female	0.88	-1.80	0.07	0.77	1.01
	No (rc)					
State of Residence		**0.90**	**-1.36**	**0.17**	**0.77**	**1.04**
	Oyo	0.90	-1.36	0.17	0.77	1.04
	Others (rc)					
clinics		**1.00**	**0.11**	**0.91**	**0.96**	**1.04**
	OTHERS	0.99	-0.00	0.99	0.85	1.16
	GYNE	0.32	-3.45	0.00	0.17	0.61
	CHOP	0.42	-2.31	0.02	0.20	0.87
	SOP	0.83	-0.73	0.46	0.52	1.34
	MOP(rc)					
Patient Admitted		**1.36**	**3.06**	**0.00**	**1.11**	**1.04**
	Yes	1.43	3.45	0.00	1.16	1.76
	No (rc)					
Surgery		**0.60**	**-3.69**	**0.00**	**0.46**	**0.79**
	Yes	0.61	-3.51	0.00	.0.46	0.80
	No					
Trt_Outcome		**1.76**	**4.64**	**0.000**	**1.39**	**2.25**
	DAMA	3..06	4.48	0.00	1.87	5.01
	Died	18.22	2.84	0.00	2.45	135.17
	Alive (rc)					

4.6.5.2 Parametric approach

Results of parametric (Exponential and Weibull) survival models used to measure effect of patients demographic and health characteristics on dormancy time of records created between 2005 and 2009 (4^{th} cohort) show as follow:

Fitting Exponential Model to dormancy time data for 4^{th} cohort 2005-2009

Given that time-to-dormancy (age of records) data is skewed distributed data, we regress dormancy time on patients characteristics based on exponential model assumption of parameter y=1. Here patient State of residence, HR=0.83, along with other characteristics like gender, HR=0.80; patient admission status, HR=1.53; surgery HR=0.51 and treatment outcome, HR=2.74 significantly ($p<0.05$, $p<0.01$) influence their dormancy time. Implying that Female patient have lower hazard of having dormant record compare to Male patient. Table 4.46 below shows the Exponential regression model:

Table 4.46: Exponential Regression of Dormancy Time on Patients Characteristics,

Variable	factor	H_Ratio	z	p>|z|	95% CI

2005-2009 4th cohort

Age group		**0.99**	**-0.41**	**0.68**	**0.94**	**1.03**
	60+	1.00	0.01	0.99	0.80	1.24
	31-60	0.88	-1.19	0.23	0.72	1.08
	21-30	1.21	1.53	1.12	0.94	1.55
	10-20	1.29	1.91	0.05	0.99	1.68
	<10 years (rc)					
Gender		**0.80**	**-3.25**	**0.68**	**0.94**	**1.03**
	female	0.86	-2.17	0.03	0.75	0.98
	Male (rc)					
State of Residence		**0.83**	**-2.45**	**0.01**	**0.71**	**0.96**
	Oyo	0.82	-2.53	0.01	0.70	0.98
	Others (rc)					
clinics		**0.99**	**-0.20**	**0.84**	**0.95**	**1.03**
	OTHERS	0.94	-0.68	0.49	0.81	1.10
	GYNE	0.23	-4.55	0.00	0.12	0.43
	CHOP	0.29	-3.27	0.00	0.14	0.61
	SOP	0.69	-1.51	0.13	0.43	1.11
	MOP (rc)					
Patient Admitted		**1.53**	**4.32**	**0.00**	**1.26**	**1.86**
	Yes	1.66	4.95	0.00	1.35	2.03
	No (rc)					
Surgery		**0.51**	**-4.91**	**0.00**	**0.39**	**0.67**
	Yes	0.52	-4.75	0.00	0.39	0.68
	No (rc)					
Trt_Outcome		**2.74**	**8.28**	**0.00**	**2.16**	**3.48**
	DAMA	7.60	8.15	0.00	4.66	12.38
	Died	118.67	4.74	0.00	16.45	855.97
	Alive (rc)					
_cons	**variable**	**0.03**	**-15.51**	**0.00**	**0.02**	**0.05**
	categories	0.07	-19.70	0.00	0.05	0.09

Fitting Weibull Model to Dormancy Time data on Patients Characteristics, 2005-2009 4th cohort

The result of Weibull model fitted to the dormancy time data (Table 4.47) under the assumption that the exponential model fail and the model fit Weibull model of parameter y=k=1, showed patient gender (HR=0.84, P<0.01), State (HR=0.88, P<0.10), admission status (HR=1.42, P<0.01), surgery (HR=0.56, P<0.01) and treatment outcome (HR=1.79, P<0.01) significantly influence patient record dormancy time. However Patients Age, HR=0.97, and type of Clinics attended, HR=0.99, will not determine patient record dormancy time.

Table 4.47: Weibull regression model of dormancy time on patient characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.97**	**-1.03**	**0.30**	**0.92**	**1.02**
	60+	0.89	-0.95	0.34	0.72	1.12
	31-60	0.90	-0.94	0.34	0.74	1.10
	21-30	1.16	1.22	0.22	0.91	1.49
	10-20	1.12	0.91	0.36	0.86	1.46
	<10 years (rc)					
Gender		**0.84**	**-2.60**	**0.00**	**0.74**	**0.95**
	female	0.87	-2.00	0.04	0.76	0.99
	male(rc)					
State of Residence		**0.88**	**-1.65**	**0.09**	**0.76**	**1.02**
	Oyo	0.88	-1.69	0.09	0.75	1.02
	Others (rc)					
clinics		**0.99**	**-0.03**	**0.97**	**0.96**	**1.03**
	OTHERS	0.98	-0.23	0.81	0.84	1.14
	GYNE	0.34	-3.31	0.00	0.18	0.64
	CHOP	0.38	-2.56	0.01	0.18	0.79
	SOP	0.77	-1.05	0.29	0.48	1.24
	MOP (rc)					

Patient Admitted		**1.42**	**3.48**	**0.00**	**1.16**	**1.74**
	Yes	1.48	3.77	0.00	1.20	1.81
	No (rc)					
Surgery done		**0.56**	**-4.20**	**0.00**	**0.43**	**0.73**
	Yes	0.58	-3.85	0.00	0.44	0.76
	No (rc)					
Treatment Outcome		**1.79**	**4.74**	**0.00**	**1.40**	**2.28**
	DAMA	3.12	4.55	0.00	1.91	5.10
	Died	9.46	2.22	0.02	1.30	68.61
	Alive (rc)					
_cons	**variable**	**0.22**	**-6.74**	**0.00**	**0.14**	**0.35**
	categories	0.32	-8.11	0.00	0.24	0.42
/1n_p	**variable**	**-0.57**	**-23.33**	**0.00**	**-0.62**	**-0.53**
	categories	-0.56	-22.82	0.00	-0.61	-0.51
P **1/p**	**variable**	**0.56** **1.78**			**0.53** **1.70**	**0.58** **1.87**
	categories	0.56 1.76			0.54 1.67	0.59 1.84

4.7Cohort 5: Medical records created between 1st January 2010and 31st December2014

In cohort 5, one thousand two hundred and seven (1207) records was analysed for dormancy time and the results are presented below.

4.7.1 Frequency distribution of some demographic and clinical characteristics of the patients

Table 4.48 shows the socio-demographic and clinical characteristics of the patients.

Most, 47.68%, of the records between the age 31-60 years, 8.13% were aged 10-20 years, 8.71%) were under 10 years old, and 19.49%) were above 60 years of age. Male patients were 56.44%, while patients residing in Oyo Stateconstituted 50.13%. Out of the records observed, records from MOP constituted 69.93%, Surgical Outpatient clinic were 1.46%),Children Outpatient clinic 0.60% while 26.98% of the records were fromother clinics. About 31% of patients were admitted and only 2.49% of the patients ever went through surgical operations,almost all (97.34%) of the patients were alive at the time of

last contact. 2.57% were discharge against medical advice but only 0.08% of the patients died.

Table 4.48: Frequency distribution of patient's characteristics in cohort 5(2010-2014)

Variables n=1207	Level	Frequency	Percent	Cumulative percent
Age at Registration	<10	105	8.71	8.71
	10-20	98	8.13	16.83
	21-30	193	16.00	32.84
	Adult 31-60	575	47.68	80.51
	61+	235	19.49	100.00
Gender	male	657	56.44	56.44
	female	507	43.56	100.00
State of	Oyo State	592	50.13	50.13

residence	Others	589	49.87	100.00
Clinic attended	MOP	814	69.93	69.93
	SOP	17	1.46	71.39
	CHOP	7	0.60	71.99
	GYNE	12	1.03	73.02
	Others	314	26.98	100.00
Ever admitted	No	834	69.10	69.10
	Yes	373	30.90	100.00
Ever operated on	No	1177	97.51	97.51
	Yes	30	2.49	100.00
Treatment outcome	Alive	1172	97.34	97.34
	Died	-	2.57	99.92
	DAMA	31	0.08	100.00
	referred	-	-	

4.7.2Frequency distribution of records by dormancy times 5thcohort 2010-2014

Table 4.49 showed the frequncy distribution for the 1537 records in the studythat survived beyond the first day of creation. Close to 50.0% of the records were already dormant as at the end of the first month of creation and about 95.0% in 33.5 months. The distribution is presented graphically in Figure 4.21. The distribution is skewed to the right.

Table 4.49 Distribution of dormancy times 5th cohort 2010-2014

month t*		Dormant records	percent	Cum. percent
0-<1	0.5	531	43.99	43.99
1-6	3.5	341	28.25	72.25
7-12	9.5	74	6.13	78.38
13-18	15.5	55	4.56	82.93
19-24	21.5	47	3.89	86.83
25-30	27.5	63	5.22	92.05

31-36	33.5	52	4.31	96.35
37-42	39.5	43	3.56	99.92
115-120	117.5	1	0.08	100.00
Total		1207	100.00	

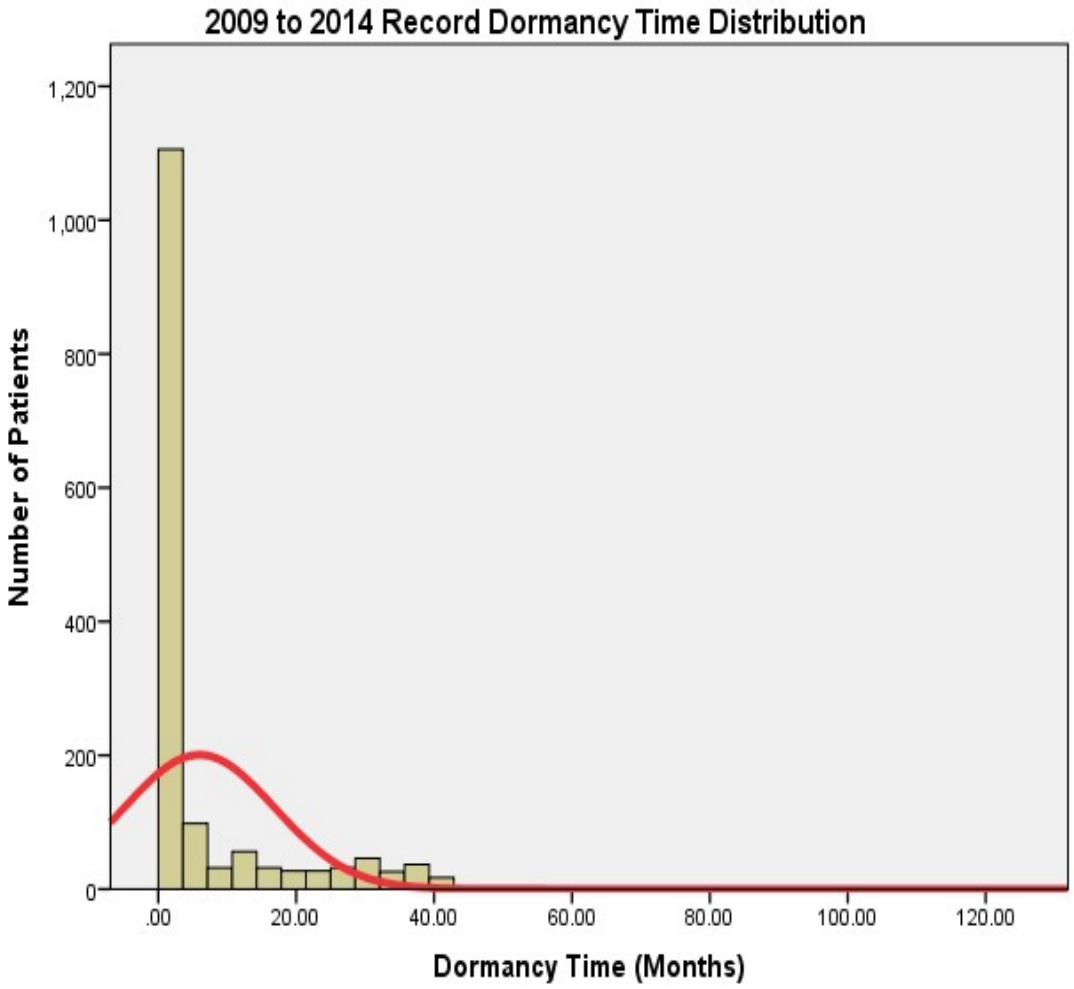


Figure 4.21: Graph showing distribution dormancy time

4.7.3Survival function of dormancy times cohort 5 (2010-2014) data

Table 4.50 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of records or dormancy time increases and tends toward zero as time reaches end point. Result show that at the dormancy time of approximately 117.5 months, dormancy of records approaches 100%. The results are presnted graphically in Figure 4.22 for the survival curve.

Table 4.50 Distribution of Survival function of dormancy times cohort 5 (2010-2014

Time	Dormant records	Survival Function	Std. Error	95% CI	
0.5	531	0.89	0.10	0.43	0.98
3.5	341	0.78	0.14	0.36	0.93
9.5	74	0.67	0.16	0.28	0.88
15.5	55	0.56	0.17	0.20	0.80
21.5	47	0.44	0.17	0.13	0.72
27.5	63	0.33	0.16	0.08	0.62
33.5	52	0.22	0.14	0.03	0.51
39.5	43	0.11	0.10	0.01	0.39
117.5	1	0	.	.	.

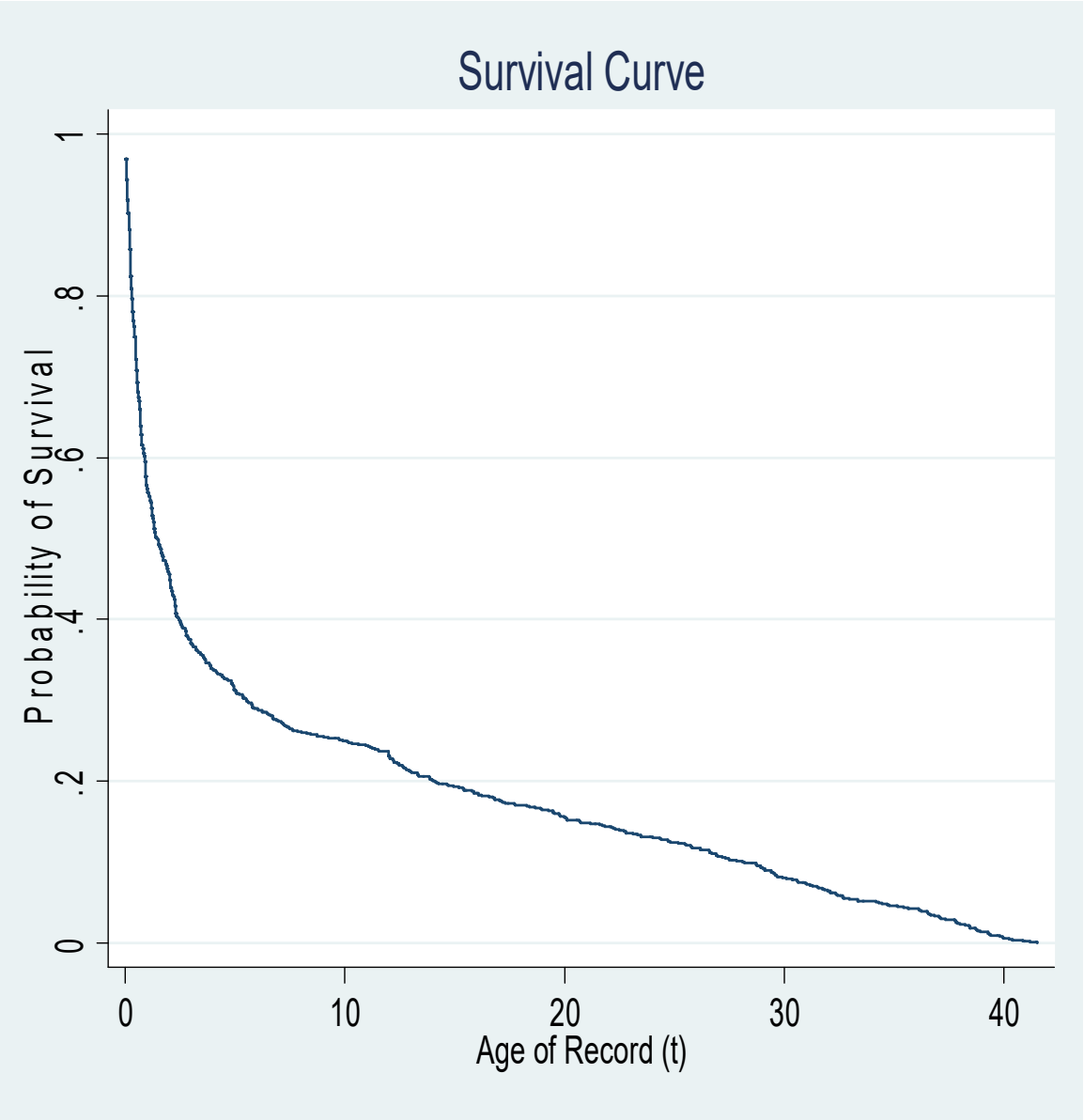


Figure 4.22 Survival Curve of time to dormancy cohort5 (2010-2014) data

Median dormancy times, standard errors and confidence intervals by patient characteristics

Table 4.51 show the median dormancy time according to categories of patient's characteristics.The Median Dormancy Time (*MDT)* of record of patients aged <10 years was 2.66 months, that of patients aged 10-20 was1.18 months and patients between 30-60 years of agehad MDT of 1.31 months for their records, while those 60 years and above hada *MDT* of 1.21 months for their records. Male patient records had a MDT of 1.31 compared to female patient with 1.74 months, records of patients' residence in Oyo State had*MDT*of 1.37 months as against records of patients from other States with 1.51 months. SOP records with *MDT* of 9.75 months was highest compared to MOP records with lowest *MDT* of 1.28 months; records from GYNE and CHOP showed a MDT of 6.60 and 2.03 months respectively. Records of admitted patients had MDT of 0.72 months lower than records of those that had never been admitted with MDT of 2.13 months. MDT of records of patients that never had surgery with 1.41 months, compared to 2.89months for records of patients that had under gone surgery. The MDT of record of patients that were alive at time of last contact was 1.54 compared to the median dormancy time for DAMA which was 0,19 months.

Table 4.51 Median-Dormancy-Time by Patient Characteristics 5th cohort 2010-2014

Variables n=1207	Level	n	t (months)	Std. Error	95% CI	
		1207	**1.51**	**0.12**	**1.24**	**1.80**
Age at Registration	<10	105	2.66	0.67	1.77	4.63
	10-20	98	1.18	0.21	0.72	2.00
	21-30	193	1.64	0.31	0.95	2.10
	31-60	575	1.31	0.20	1.11	1.87
	61+	235	1.21	0.30	0.72	1.90
Gender	male	657	1.31	0.17	1.05	1.74
	female	507	1.74	0.22	1.28	2.23
State of residence	Others States	589	1.51	0.17	1.21	1.93
	Oyo State	592	1.37	0.21	1.11	2.00
Clinic attended	MOP	814	1.28	0.16	1.05	1.74
	SOP	17	9.75	9.19	0.32	1.74
	CHOP	7	2.03	0.68	0.16	10.15
	GYNE	12	6.60	9.10	0.16	31.27
	Others	314	1.64	0.20	1.31	2.10
Ever admitted	No	834	2.13	0.16	1.87	2.43
	Yes	373	0.72	0.07	.0.59	0.85
Ever operated on	No	1177	1.41	0.13	1.21	1.74
	Yes	30	2.89	1.52	1.11	7.58
Treatment outcome	Alive	1172	1.54	0.13	1.31	1.90
	Died	-	-	-	-	-
	DAMA	31	0.19	0.02	0.13	1.90
	referred	1	-	-	-	-

Selected percentiles of the survival distribution

Table 4.52 shows the respective record survival estimate, their standard error and confidence intervals at each percentage point.

Estimates of dormancy time of record measured for 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of observed record revealed thattwenty five percent of the records were dormant in 0.42 months while the fifty percent (MDT) of records were dormant in 1.51 months as shown from the 50^{th} percentiles. Also, the 75^{th} and 95^{th} percentiles shows that not less than seventy five percent and ninety five percent of records were dormant in 10.61 and 34.76 months respectively.

Table 4.52Selected percentiles of the survival distribution 5th cohort 2010-2014

Percentiles	t (months)	Std. Error	95% CI	
25th	0.42	0.02	0.36	0.45
5th	1.51	0.12	1.24	1.80
7th	10.61	1.10	7.19	12.45
95th	34.75	5.65	32.65	36.63
n = 1207				

4.7.4 Hazard curve of dormancy times

Table 4.53 show the distribution of the hazard functions, standard errors and 95% Confidence Interval. The hazard plot that follows, Figure 4.23, show a cvruve with sharp decrease with increase in age of records observed at the initial time, t, and continue to decrease as dormancy times increase. This was followed by a constant and steady movement between 10 and 20 months and then an increase with a sharp rise following constant and steady upward movement till it reaches end point and thereby making a bathtub shape.

Table 4.53: Frequency distribution of hazard function of dormancy times (2010-2014) Cohort 5

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	682	531	0.44	0.01	0.41	0.47
5 -	383	295	0.68	0.01	0.66	0.71
10 -	308	74	0.76	0.01	0.72	0.77
15 -	241	67	0.80	0.01	0.78	0.82
20 -	197	44	0.84	0.01	0.81	0.85
25 -	160	37	0.87	0.01	0.85	0.87
30 -	105	55	0.91	0.01	0.9	0.93
35 -	61	44	0.95	0.01	0.94	0.96
40 -	10	51	0.99	0.00	0.99	1.00
45 -	2	8	1.00	0.00	1.00	1.00
50 -	2	0	1.00	0.00	1.00	1.00
55 -	2	0	1.00	0.00	1.00	1.00
60 -	2	0	1.00	0.00	1.00	1.00
65 -	2	0	1.00	0.00	1.00	1.00
70 -	2	0	1.00	0.00	1.00	1.00
75 -	2	0	1.00	0.00	1.00	1.00
80 -	2	0	1.00	0.00	1.00	1.00
85 -	2	0	1.00	0.00	1.00	1.00
90 -	2	0	1.00	0.00	1.00	1.00
95 -	2	0	1.00	0.00	1.00	1.00
100 -	2	0	1.00	0.00	1.00	1.00
105 -	2	0	1.00	0.00	1.00	1.00
110 -	2	0	1.00	0.00	1.00	1.00
115 -	2	0	1.00	0.00	1.00	1.00
120 -	2	0	1.00	0.00	1.00	1.00
125 -	1	1	-	-	-	-

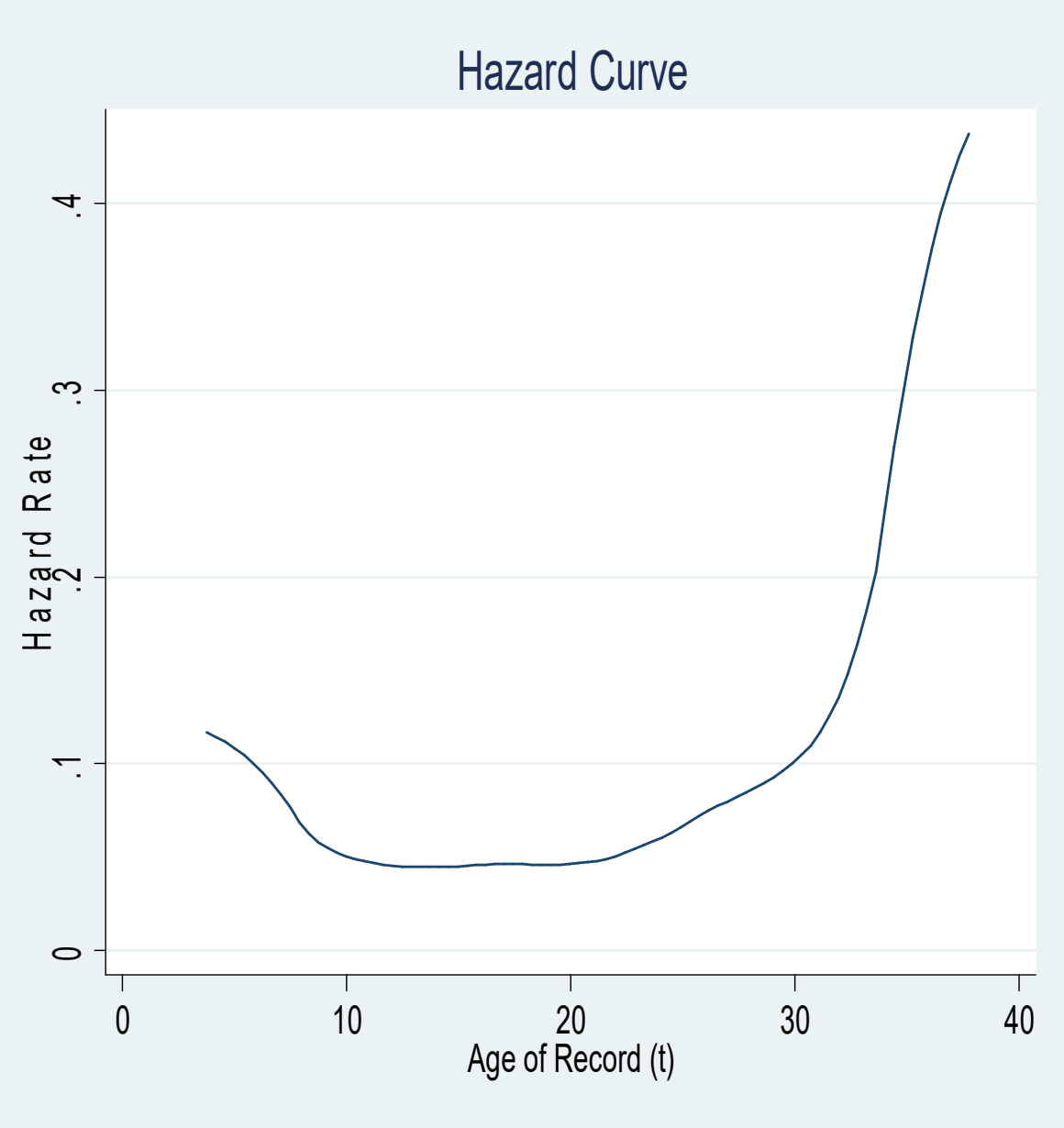


Figure 4.23 Hazard Curve of dormancy timefor cohort 5 (2010-2014)

Graphical evaluation of the form of the hazard function

Considering the bathtub shape of the hazard plot, typical of a Weibull distribution we tested for validity of Weibull distribution assumption using a Kaplan-*Meier* log-log Survival curves, *log[H(t)],* against log of survival time, *log(t),* Figure 4.24. The plot indicated a straight line relationship between *logH(t)* against *log(t),* increasing monotonically. The intercept of the fitted line was approximately -0.0847 with a slope of 0.3819. From this the value of the shape parameter, γ, and the hazard rate for two parameter Weibull distribution was estimated as:

$$\gamma^* = exp(-0.0847) = 0.9187 \text{ and}$$

$$\lambda^* = 0.3819$$

respectively. Since the estimated value of the shape parameter, γ, was less than unity, suggesting an decreasing hazard, λ, the TTD data of medical records created between 2010 and 2014 can be said to follow Weibull distribution.

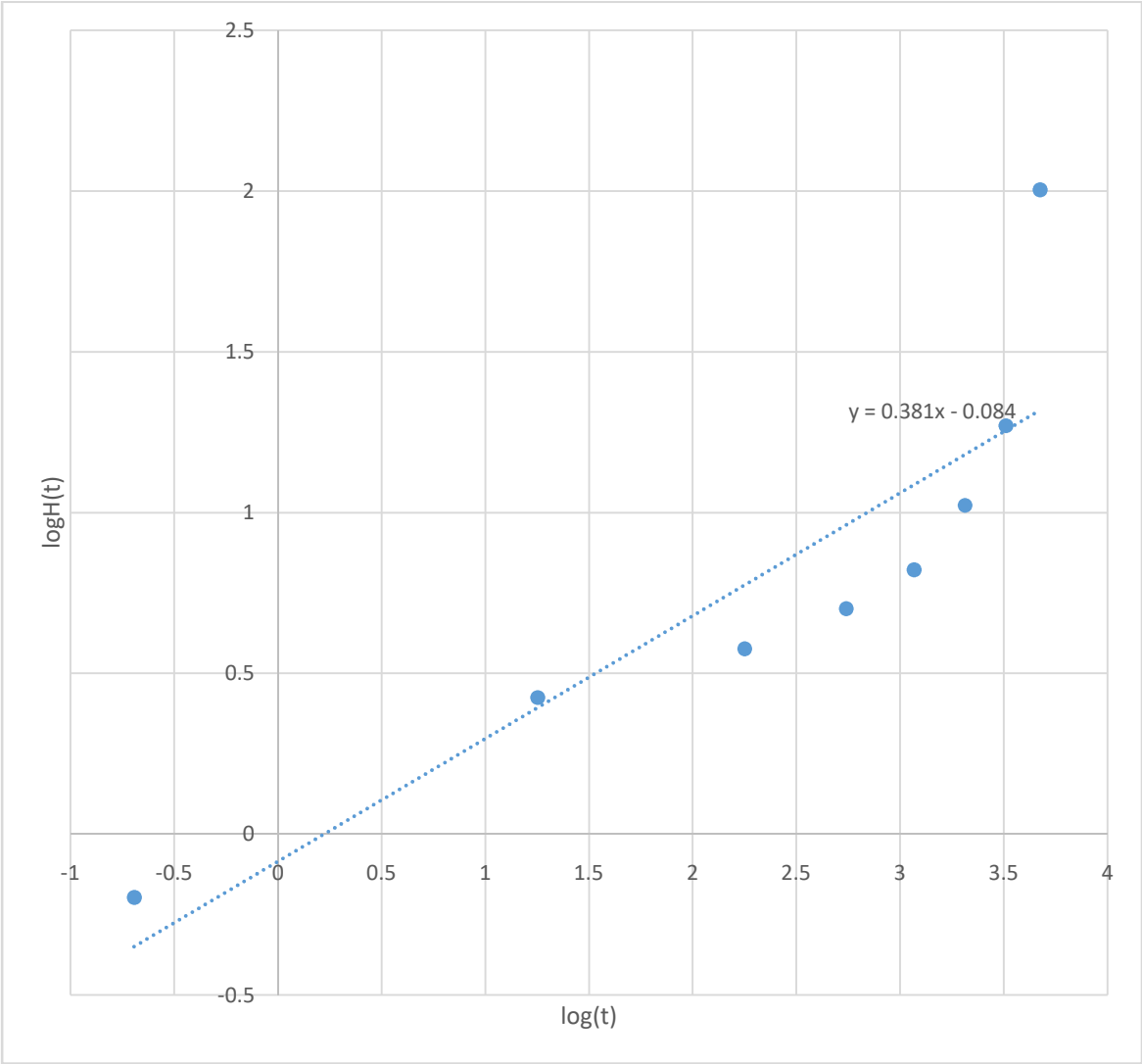


Figure 4.24. Plot of log Cum. Hazard on Log of time

4.7.5 Influence of patient characteristics on hazard rate

4.7.5.1 Non Parametric approach

Results of semi-parametric (Cox Proportional Hazard) survival model used to measure effect of patients demographic and health characteristics on dormancy time of records created between 2010 and 2014 (5^{th} cohort) show as follow:

Schoenfeld test of Cox Proportional Hazard model assumption

Table 4.54 below shows the results of the global test for the proportional hazard assumption. The significant ($p<0.05$) of the test implies that the sample data is invalid for the proportional hazard assumption-that the hazard of subject subgroup are proportional over follow up period and thereforethe global test indicated that for the data set used the assumption of PH isviolated.

.

Table 4.54: Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	Chi-square	df	p-value
Proportional Hazard Assumption	22.61	7	0.00

Graphical test for Proportional Hazard Assumption

Thus the respective graph comparing patients' gender while adjusting for age, zone, clinics, admission and surgery status, and treatment outcome shows that the two line (male and female) intersect (non-parallel) each other and therefore substantiate the claim that the proportional assumption is valid for the data, Figure 4.25.

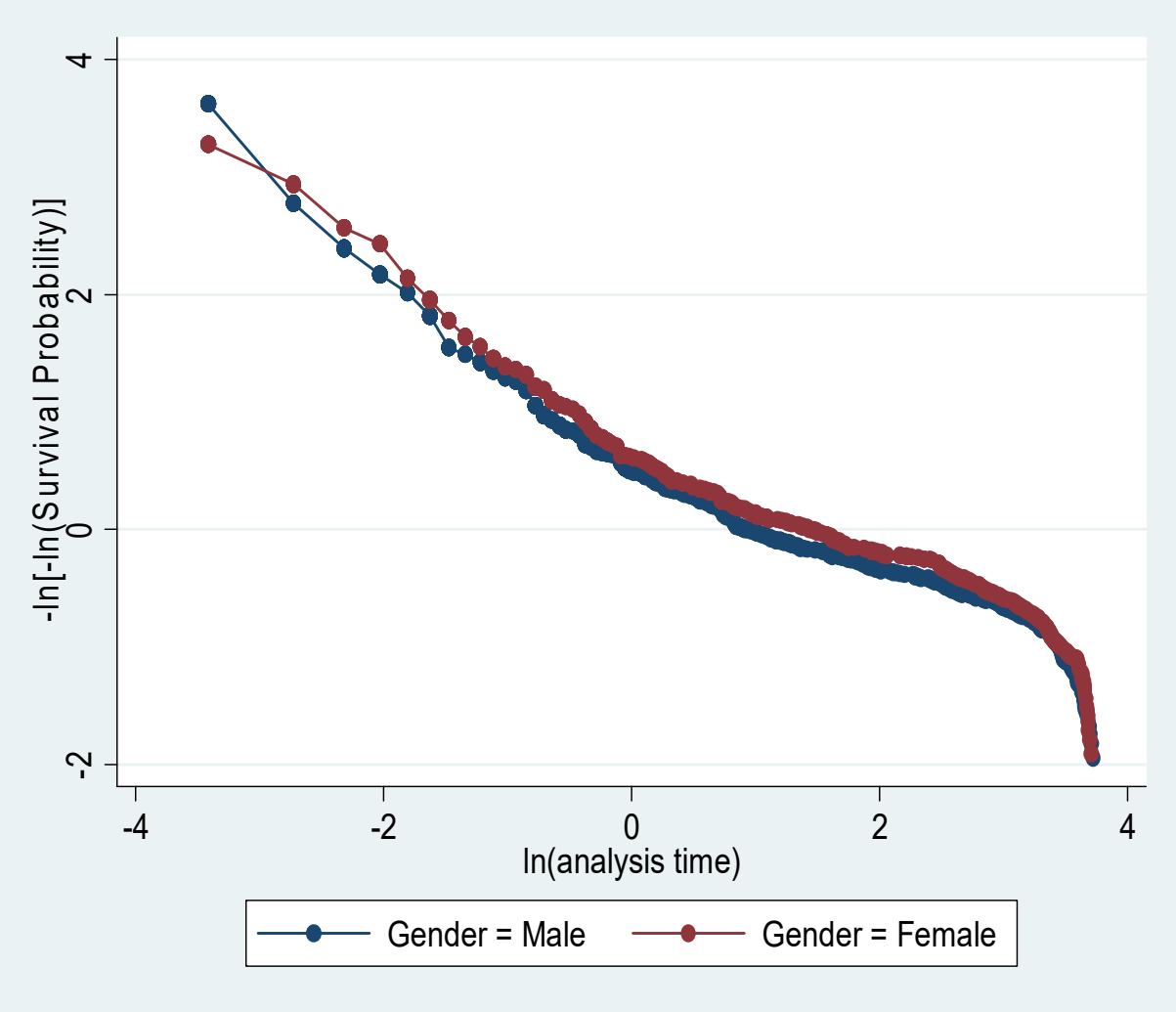


Figure 4.25 Graph Showing violation of Proportional Hazard Assumption 5th cohort 2010-2014

Fitting Cox Proportional Hazard Model

Table 4.55 below shows the Cox regression analysis that succeed the global test above. It can be inferred from the table that time-to-dormancy of record were affected by patient admission status (HR=1.50, p-value=0.000), surgery status (HR=0.51, p-value = 0.003) and treatment outcome (HR=1.27, p-value = 0.012) as they are significant at 1% and 5% α level respectively. Whereas, patients age, HR=1.01; gender, HR=0.90; state of residence, HR=0.98 and type of clinic attended, HR=1.00, do not predict patient record dormancy time. Also, hazard ratio that assess the risk magnitude and likelihood were presented in the table for each group of patients characteristic.

Table 4.55 Cox regression of dormancy time on patient characteristics 2010-2014

Variable	factor	H_Ratio	z	p>|z|	95% CI	
Age group		**1.01**	**0.65**	**0.51**	**0.96**	**1.07**
	60+	1.19	1.28	0.20	0.91	1.56
	31-60	1.22	1.60	0.10	0.95	1.56
	21-30	1.26	1.69	0.09	0.96	1.65
	10-20	1.34	1.88	0.06	0.98	1.84
	<10 years (rc)					
Gender		**0.90**	**-1.58**	**0.11**	**0.80**	**1.02**
	female	0.89	-1.82	0.06	0.79	1.00
	Male (rc)					
State of Residence		**0.98**	**-0.17**	**0.86**	**0.87**	**1.11**
	Oyo	0.98	-0.25	0.80	0.87	1.10
	Others (rc)					
clinics		**1.00**	**0.46**	**0.64**	**0.97**	**1.04**
	OTHERS	1.06	0.87	0.38	0.91	1.24
	GYNE	0.82	-0.61	0.54	0.43	1.54
	CHOP	0.86	-0.36	0.71	0.39	1.90
	SOP	0.72	-1.31	0.18	0.44	1.17
	MOP (rc)					
Patient Admitted		**1.50**	**5.71**	**0.00**	**1.30**	**1.73**
	Yes	1.89	-1.82	0.18	1.44	1.00
	No (rc)					
Surgery		**0.51**	**-2.97**	**0.00**	**0.32**	**0.79**
	Yes	0.98	-0.25	0.80	0.87	1.10
	No (rc)					
Trt_Outcome		**1.27**	**2.51**	**0.01**	**1.05**	**1.53**
	DAMA	3.75	1.32	0.18	0.52	26.89
	Died	1.59	2.35	0.01	1.08	2.34
	Alive (rc)					

4.7.5.2 Parametric approach

Fitting Exponential Model

Given that the time-to-dormancy of record is skewed distributed data, results of the regressed dormancy time on patients categorical and sub-categorical characteristics, based on exponential model assumption of parameter $\lambda=1$, show patient characteristics gender, HR=0.86; clinic attended, HR=1.04, admission status, HR=1.90, surgery, HR=0.44, and treatment outcome, HR=1.20, significantly ($p<0.01$ and $p<0.05$) influence their record dormancy time. Although, age, HR=1.00 and state of residence, HR=0.96 were not significant. This generally imply that record of female patient admitted and discharge against medical advice after surgery will become dormant earlier than younger male patient that are alive after treatment. Table 4.56 below shows the Exponential regression model for the explanatory variables.

Table 4.56 Exponential regression of dormancy time on patient characteristics 2010-2014

Variable	factor	H_Ratio	z	p>|z|	95% CI	
		1.00	**0.05**	**0.95**	**0.94**	**1.05**
Age group	**60+**	1.18	1.18	0.23	0.89	1.55
	31-60	1.32	2.23	0.02	1.03	1.69
	21-30	1.34	2.15	0.03	1.02	1.77
	10-20	1.53	2.65	0.00	1.11	2.09
	<10 years (rc)					
Gender		**0.86**	**-2.44**	**0.01**	**0.76**	**9.97**
	female	0.84	-2.68	0.00	0.74	.95
	Male (rc)					
State of Residence		**0.96**	**-0.53**	**0.59**	**0.86**	**1.09**
	Oyo	0.96	-0.63	0.52	0.85	1.08
	Others (rc)					
clinics		**1.04**	**2.11**	**0.03**	**1.00**	**1.08**
	OTHERS	1.24	2.86	0.00	1.07	1.45
	GYNE	0.63	-1.39	0.16	0.33	1.20
	CHOP	1.26	0.59	0.55	057	2.77
	SOP	0.59	-2.06	0.03	0.36	0.97
	MOP (rc)					
Patient Admitted		**1.90**	**8.95**	**0.00**	**1.65**	**2.19**
	Yes	1.90	8.88	0.00	1.65	2.19
	No (rc)					
Surgery		**0.44**	**-3.54**	**0.00**	**0.28**	**0.69**
	Yes	0.53	-2.74	0.00	0.33	0.83
	No (rc)					
Trt_Outcome		**1.20**	**1.97**	**0.04**	**1.00**	**1.46**
	DAMA	18.48	2.91	0.00	2.58	132.07
	Died	1.39	1.67	0.09	0.94	2.05
	Alive (rc)					
_cons	**variable**	**0.10**	**-12.03**	**0.00**	**0.07**	**0.15**

Fitting Weibull Model

Results of the fit Weibull model to the skewed distributed time-to-dormancy under the assumption that the exponential model fail and the model fit Weibull model of parameter $\Upsilon=\lambda=1$. Similar to Cox model above, only patient characteristics like; admission status, HR=1.61, surgery, HR=0.54 and treatment outcome, HR=1.20 are significant (at $p<0.01$ and $p<0.05$) predictors of record dormancy time. Table 4.57 below shows the Weibull regression model result with two significant ($p<0.01$) extended parameter for the categorical and sub-categorical characteristics:

Table 4.57: Weibull regression model of dormancy time on patient characteristics 2010-2014

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**1.00**	**0.22**	**0.82**	**0.95**	**1.06**
	60+	1.16	1.09	0.27	0.88	1.52
	31-60	1.24	1.78	0.07	0.97	1..59
	21-30	1.25	1.63	0.10	0.95	1.64
	10-20	1.41	2.16	0.03	1.03	1.92
	<10 years (rc)					
Gender		**0.89**	**-1.82**	**0.06**	**0.79**	**1.00**
	female	0.88	-1.95	0.05	0.78	1.00
	Male (rc)					
State of Residence		**0.97**	**-0.44**	**0.65**	**0.86**	**1.09**
	Oyo	0.96	-0.51	0.61	0.86	1.09
	Others (rc)					
clinics		**1.02**	**1.51**	**0.13**	**0.99**	**1.06**
	OTHERS	1.16	2.06	0.04	1.00	1.35
	GYNE	0.72	-1.02	0.31	0.38	1.35
	CHOP	1.22	0.50	0.61	0.55	2.67
	SOP	0.67	-1.58	0.11	0.41	1.09
	MOP (rc)					
Patient Admitted		**1.61**	**6.67**	**0.00**	**1.40**	**1.86**
	Yes	1.62	6.70	0.00	1.40	1.87
	No (rc)					
Surgery done		**0.54**	**-2.70**	**0.00**	**0.34**	**0.84**
	Yes	0.62	-2.07	0.03	0.39	0.97
	No (rc)					
Treatment Outcome		**1.20**	**1.96**	**0.05**	**1.00**	**1.45**
	DAMA	3.72	1.31	0.19	0.52	26.68
	Died	1.41	1.75	0.07	0.96	2.08
	Alive (rc)					
_cons	**variable**	**0.33**	**-5.93**	**0.00**	**0.23**	**0.47**
	categories	0.30	-8.51	0.00	0.22	0.40
/1n_p	**variable**	**-0.55**	**-24.00**	**0.00**	**-0.59**	**-0.50**
	categories	-0.54	-23.83	0.0	-0.59	-0.50
P	**variable**	**0.57**			**0.55**	**0.60**
1/p		**1.73**			**1.66**	**1.81**
	categories	0.57			0.55	0.60
		1.72			1.65	1.80

4.8. Analysis of dormancy time for records created between 1st January 1990 and 31st December 2014

Result of analysis for the combining data for cohort 1, 2, 3, 4, and 5, are presentedbelow.

4.8.1 Frequency distribution of some demographic and clinical characteristics of patients

Table 4.58 shows socio-demographic and clinical characteristics of the 5797 patients whose records were observed for dormancy time. Result revealed that 40.4% were between the aged of 31-60 years,10.6% between 10-20 years of age, 14.6% were <10 years and 16.4% were above 60 years of age. Male patients constitute 52.4% and patients resident in Oyo State were 55.4%. Medical Outpatient Clinic records (MOP) constituted 39.3%, 10.5%of the records were from Surgical Outpatient clinic, and the least, 2.4%from Children Outpatient clinic, records from other clinics put together were41.8%. About 70% of patients were never admitted while only 8.36% went through surgical operations. Almost all the patients (98.77%) were alive at time of last contact, 1.2%) were discharge against medical advice and 0.03% died at the end of last contact.

Table 4.58 Frequency distribution of some patient's characteristics for the combined data (1990 – 2014)

Variables n=5797	Level	Frequency	%
Age at Registration	<10	824	14.61
	10 - 20	600	10.64
	21-30	1017	18.03
	31-60	2277	40.37
	above 61	922	16.35
Gender	Male	2970	52.38
	Female	2700	47.62
State of residence	Oyo State	3095	55.37
	Others	2495	44.63
Clinic attended	MOP	2219	39.30
	SOP	593	10.50
	CHOP	137	2.43
	GYNE	338	5.99
	Others	2360	41.79
Ever admitted	No	4029	69.66
	Yes	1755	30.34
Ever operated on	No	5307	91.64
	Yes	484	8.36
Treatment outcome	Alive	5679	98.77
	Died	2	0.03
	DAMA*	68	1.18
	Referred	1	0.02

*** Discharge Against Medical Advice**

4.8.2 Frequency distribution of records by dormancy time 1990-2014.

Table 4.59: shows the frequency distribution for the 5797 records (the five cohorts merged into a single sample) in the study. These are the records that survived beyond the first day of creation.In less than a months of creation, close to 37.31% of the records were already dormant, above 50% in 3.5 months, about 75% in 15.5 months dormancy time,and over 95% in a dormancy time of 105 months.

The distribution is presented graphically in Figure 4.26. The distribution is skewed to the right.

Table 4.59Frequency distribution of records dormancy time1990 to2014.

month (t)		Dormant records	%	Cum Percent
<1	0.5	2163	37.31	37.31
1-6	3.5	1538	26.53	63.04
7-12	9.5	436	7.52	71.36
13-18	15.5	250	4.31	75.67
19-24	21.5	181	3.12	78.79
25-30	27.5	164	2.83	81.61
31-36	33.5	157	2.71	84.31
37-42	39.5	111	1.91	86.22
43-48	45.5	71	1.22	87.44
49-54	51.5	53	0.91	88.35
55-60	57.5	68	1.17	89.52
61-66	63.5	56	0.97	90.48
67-72	69.5	38	0.66	91.13
73-78	75.5	36	0.62	91.75
79-84	81.5	40	0.69	92.44
85-90	87.5	47	0.81	93,25
91-96	93.5	59	1.02	94.26
97-102	99.5	40	0.69	94.95
103-108	105.5	20	0.35	95.29
109-114	111.5	14	0.24	95.53
115-120	117.5	20	0.35	95.87
121-126	123.5	22	0.38	96.24
127-132	129.5	20	0.35	96.58
133-138	135.5	24	0.41	96.99
139-144	141.5	22	0.38	97.36
145-150	147.5	28	0.48	97.84
151-156	153.5	24	0.41	98.25
157-162	159.5	12	0.21	98.45
163-168	165.5	8	0.14	98.58
169-174	171.5	5	0.09	98.66
175-180	177.5	7	0.12	98.78
181-186	183.5	9	0.16	98.93
187-192	189.5	5	0.09	99.01
193-198	195.5	11	0.19	99.19
199-204	201.5	7	0.12	99.31
205-210	207.5	2	0.03	99.34

month (t)		Dormant records	%	Cum Percent
211-216	213.5	5	0.09	99.42
217-222	219.5	2	0.03	99.45
223-228	225.5	5	0.09	99.53
229-234	231.5	1	0.02	99.54
235-240	237.5	2	0.03	99.57
241-246	243.5	2	0.03	99.63
253-258	255.5	1	0.02	99.60
259-264	261.5	1	0.02	99.61
271-276	273.5	1	0.02	99.62
277-282	279.5	5	0.09	99.70
278-283	285.5	4	0.07	100
Total		5797	100	

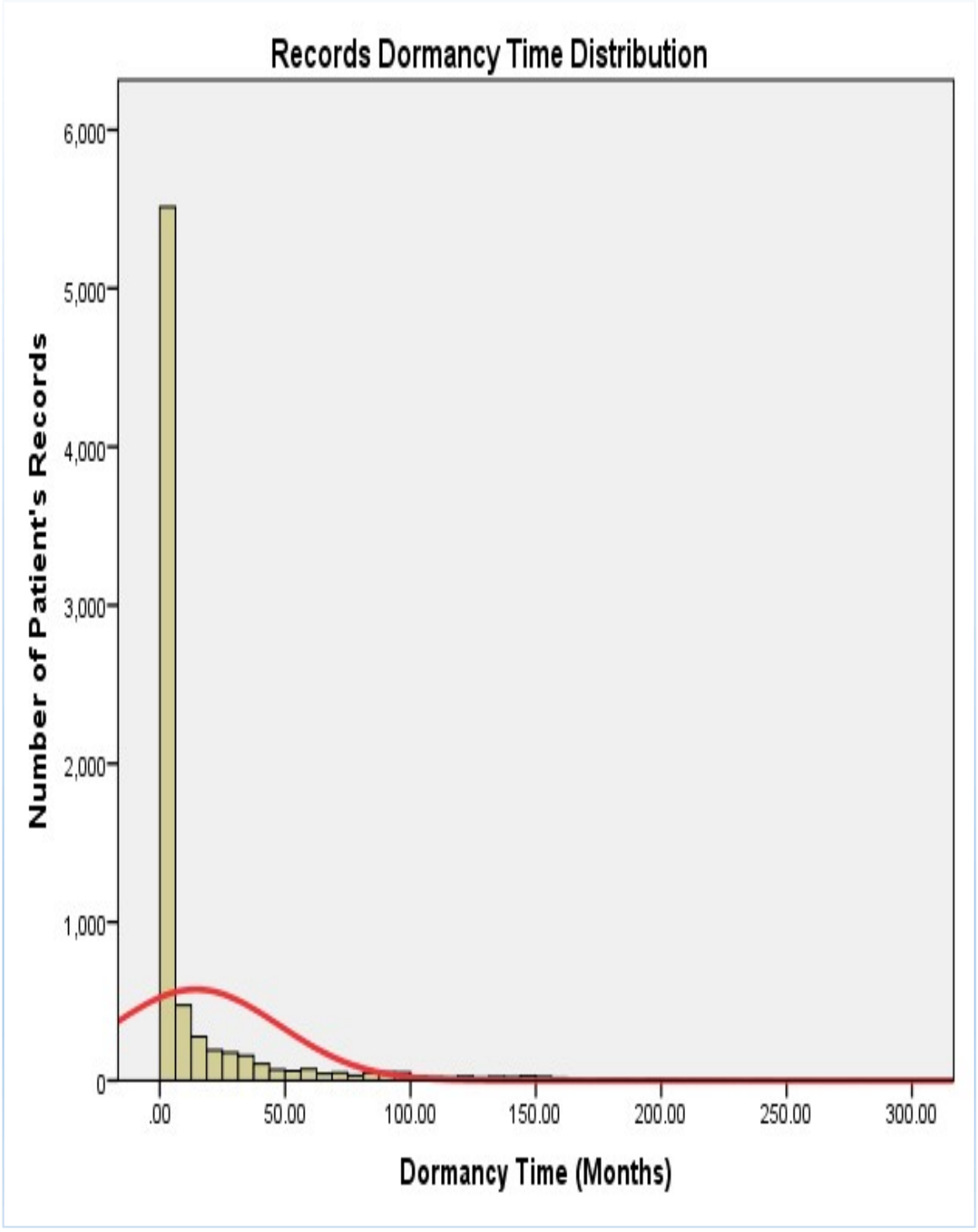

Records Dormancy Time Distribution
Number of Patient's Records
6,000
5,000
4,000
3,000
2,000
1,000
0
.00
50.00
100.00
150.00
200.00
250.00
300.00
Dormancy Time (Months)

Figure 4.26Distribution of patient record dormancy 1990-2014

4.8.3 Survival function of dormancy timesCohort 1-5 1990-2014 data

Table 4.60 shows the survival function *S(t)* of the procss, the standard errors and confidence intervals as obtained from the Kaplan-Meier mthod. The survival functions ranged between 0.0 and 1.0. The survival time of the records decreases as the age of records or dormancy time increases and tends toward zero as time reaches end point. Result show that at the dormancy time of approximately 285.5 months, dormancy of records approaches 100%. The results are presnted graphically in Figure 4.27 for the survival curve.

Table 4.60Distribution ofSurvival function of dormancy times, 1990-2014 merged

Time	Dormant records	Survival Function	Std. Error	95% CI	
0.5	2163	0.98	0.02	0.86	1.00
3.5	1538	0.96	0.03	0.84	0.99
9.5	436	0.94	0.04	0.82	0.98
15.5	250	0.91	0.04	0.79	0.97
21.5	181	0.89	0.05	0.76	0.95
27.5	164	0.87	0.05	0.74	0.94
33.5	157	0.85	0.05	0.71	0.93
39.5	111	0.83	0.05	0.69	0.91
45.5	71	0.81	0.06	0.66	0.90
51.5	53	0.79	0.06	0.64	0.88
57.5	68	0.77	0.06	0.62	0.86
63.5	56	0.74	0.06	0.59	0.85
69.5	38	0.72	0.07	0.57	0.83
75.5	36	0.70	0.07	0.55	0.81
81.5	40	0.68	0.07	0.53	0.79
87.5	47	0.66	0.07	0.51	0.78
93.5	59	0.64	0.07	0.48	0.76
99.5	40	0.62	0.07	0.46	0.74
105.5	20	0.60	0.07	0.44	0.72
111.5	14	0.57	0.07	0.42	0.70
117.5	20	0.55	0.07	0.40	0.68
123.5	22	0.53	0.07	0.38	0.66
129.5	20	0.51	0.07	0.36	0.64
135.5	24	0.49	0.07	0.34	0.62
141.5	22	0.47	0.07	0.32	0.60
147.5	28	0.45	0.07	0.30	0.58
153.5	24	0.43	0.07	0.28	0.56
159.5	12	0.40	0.07	0.26	0.54
165.5	8	0.38	0.07	0.25	0.52
171.5	5	0.36	0.07	0.23	0.50
177.5	7	0.34	0.07	0.21	0.47
183.5	9	0.32	0.07	0.19	0.45
189.5	5	0.30	0.07	0.18	0.43

Time	Dormant records	Survival Function	Std. Error	95% CI	
195.5	11	0.28	0.07	0.16	0.41
201.5	7	0.26	0.06	0.14	0.38
207.5	2	0.23	0.06	0.13	0.36
213.5	5	0.21	0.06	0.11	0.34
219.5	2	0.19	0.06	0.09	0.31
225.5	5	0.17	0.05	0.08	0.29
231.5	1	0.15	0.05	0.07	0.26
237.5	2	0.13	0.05	0.05	0.24
243.5	2	0.11	0.05	0.04	0.21
255.5	1	0.09	0.04	0.03	0.19
261.5	1	0.06	0.04	0.02	0.16
273.5	1	0.04	0.03	0.01	0.13
279.5	5	0.02	0.02	0.00	0.10
285.5	4	0	.	.	.

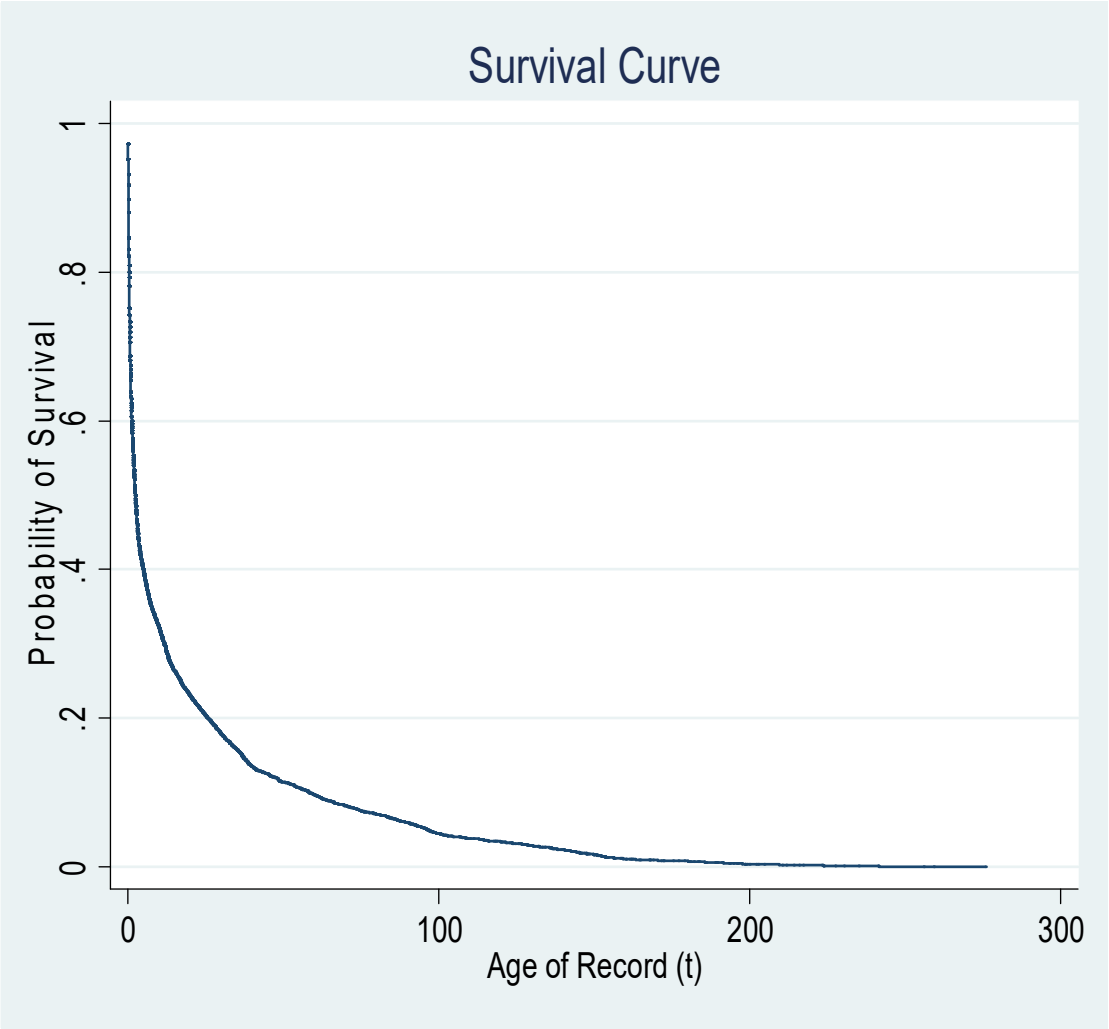


Figure 4.27: Survival Curve of dormancy time for combined cohort 1990-2014

Median survuval time, standard errors and confidence intervals by patient chracteristics

Table 4.61shows Median Dormancy Time *(MDT)* according to categories of patient characteristics. The *MDT* was measured as equivalent of the Median Survival Time.Fifty percent (50%) of records of patients aged less than 10 yearswere dormant in 1.93 months, patients aged 10-20 years had MDT of1.80 months, those aged 31-60 years had MDT of 2.43 andpatients 60years and above had MDT of 3.25 months respectively. Records of male patients*was*dormant in2.03 months compared to their female counterpart with 2.75 months, record of patientresidence in Oyo State had MDT of 2.69 as against 2.03 for patients from other states. The MDT of records of patients attending MOP clinic were dormant in 1.08 months, SOP in 2.52, CHOP in 4.10 and GYNE in 3.48 months, while other clinics were in 2.72 months. MDT for records of admitted patients were 2.75 months whereas non-admitted patients was 2.10,those that underwent surgeryhad MDT of 5.42 months, while the MDT for those patients alive as at last entry/contact was dormant in 2.29 months, DAMA in 0.19months andthose that diedas at last entry 0.06 months.

Table 4.61 Median-Dormancy-Time (MDT) by Patient Characteristics 1990-2014

Variables n=5797	Level	n	t (months)	Std. Error	95% CI	
n=5797			**2.29**	**0.09**	**2.10**	**2.49**
Age at Registration	<10	824	1.93	0.23	1.51	2.43
	10-20	599	1.80	0.26	1.44	2.49
	21-30	1017	1.83	0.16	1.47	2.23
	31-60	2274	2.43	0.18	2.10	2.89
	61+	922	3.25	0.42	2.75	4.20
Gender	male	2967	2.03	0.11	1.80	2.23
	female	2698	2.75	0.20	2.39	3.12
State of residence	Others States	2493	2.03	0.12	1.80	2.16
	Oyo State	3091	2.69	0.16	2.33	2.98
Clinic attended	MOP	2218	1.80	0.11	1.51	2.00
	SOP	593	2.52	0.26	2.06	2.98
	CHOP	137	4.10	1.61	2.29	8.57
	GYNE	337	3.48	0.71	2.52	5.45
	Others	2356	2.72	0.20	2.33	2.98
Ever admitted	No	4025	2.75	0.13	2.46	2.98
	Yes	1753	1.44	0.11	1.28	1.70
Ever operated on	No	5303	2.10	0.10	1.93	2.29
	Yes	482	5.42	0.66	3.97	6.43
Treatment outcome	Alive	5673	2.29	0.10	2.13	2.52
	Died	2	0.06	-	0.06	-
	DAMA	68	0.19	0.02	0.13	0.29
	Referred	1	-	-	-	-

Estimates of selected percentiles of the survival distribution

Estimates of specific points of dormancy time for patient records was measured for 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of observed survival distribution.Table 4.62 below shows the respective record survival estimate, their standard error and confidence interval at each percentage percentiles.

The 25^{th} percentile survival estimate shows that 25% of the records were dormant at t = 0.45 months, 50% at t = 2.30 months as shown from the 50^{th} percentiles. The 75^{th} and 95^{th} percentiles show that seventy five percent and ninety five percent of records were dormant by the 17.58 and 101.52 months respectively.

Table 4.62 Selected percentiles of the survival curve (1990-2014 data)

Percentiles	t (months)	Std. Error	95% CI	
25%	0.45	0.01	0.45	0.49
50%	2.29	0.09	2.10	2.49
75%	17.57	0.84	16.13	19.64
95%	101.51	6.09	96.82	110.16
n = 5797				

4.8.4 Hazard plot of dormancy time for records created 1990-2014

Table 4.63 show the distribution of the hazard functions, the standard errors and 95% Confidence Intervals. The hazard plot that follows, Figure 4.28, shows the resultof the hazard plot with sharp decrease with age of records until time, t, reaches around 60 months and then assumed an irregular trend following a zigzag pattern of an uneven movement until about $t = 125$ months. From this point the hazard increased sharply but still with irregular trends as age (dormancy time) of patient records increases and thereby forming a bathtub shape.

Table 4.63: Frequency distribution of hazard function (1990-1014) Cohort 1-5

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
< 1	0	0	0.00	-	-	-
1 -	3647	2164	0.37	0.01	0.36	0.39
5 -	2324	1307	0.60	0.01	0.59	0.61
10 -	1891	432	0.67	0.01	0.67	0.69
15 -	1557	330	0.73	0.01	0.72	0.74
20 -	1375	185	0.76	0.01	0.75	0.77
25 -	1223	252	0.79	0.01	0.78	0.80
30 -	1079	143	0.81	0.01	0.80	0.82
35 -	958	121	0.83	0.00	0.83	0.84
40 -	826	132	0.86	0.00	0.85	0.87
45 -	774	52	0.87	0.00	0.86	0.88
50 -	712	62	0.88	0.00	0.87	0.89
55 -	665	47	0.89	0.00	0.88	0.89
60 -	612	54	0.89	0.00	0.89	0.90
65 -	562	49	0.90	0.00	0.90	0.91
70 -	526	36	0.91	0.00	0.90	0.92
75 -	486	40	0.92	0.00	0.91	0.92
80 -	462	24	0.92	0.00	0.91	0.93
85 -	427	35	0.93	0.00	0.92	0.93
90 -	394	33	0.93	0.00	0.93	0.94
95 -	345	50	0.94	0.00	0.93	0.95
100 -	301	43	0.95	0.00	0.94	0.95
105 -	276	25	0.95	0.00	0.95	o.95
110 -	259	17	0.96	0.00	0.95	0.96
115 -	247	12	0.96	0.00	0.95	0.96
120 -	231	16	0.96	0.00	0.96	0.97
125 -	211	20	0.96	0.00	0.96	0.97
130 -	195	16	0.97	0.00	0.96	0.97
135 -	180	15	0.97	0.00	0.96	0.97
140-	158	2	0.97	0.00	0.97	0.98
145 -	139	19	0.98	0.00	0.97	0.98
150 –	116	23	0.98	0.00	0.98	0.98
155 -	94	22	0.98	0.00	0.98	0.99
160 -	82	12	0.99	0.00	0.98	0.99
165 -	72	10	0.99	0.00	0.98	0.99
170 -	66	6	0.99	0.00	0.99	0.99
175 -	62	4	0.99	0.00	0.99	0.99
180 -	57	5	0.99	0.00	0.99	0.99
185 -	58	7	0.99	0.00	0.99	0.99
190 -	43	7	0.99	0.00	0.99	0.99

Time (months)	n	Records failing	Hazard function	Std. Error	95% CI	
195 -	38	5	0.99	0.00	0.99	1.00
200 -	29	9	0.99	0.00	0.99	1.00
205 -	23	6	1.00	0.00	0.99	1.00
210 -	22	1	1.00	0.00	0.99	1.00
215 -	18	4	1.00	0.00	0.99	1.00
220 -	14	4	1.00	0.00	1.00	1.00
225 -	10	4	1.00	0.00	1.00	1.00
230 -	9	1	1.00	0.00	1.00	1.00
235 -	8	1	1.00	0.00	1.00	1.00
240 -	6	2	1.00	0.00	1.00	1.00
245 -	4	2	1.00	0.00	1.00	1.00
250 -	4	0	1.00	0.00	1.00	1.00
255 -	4	0	1.00	0.00	1.00	1.00
260 -	2	2	1.00	0.00	1.00	1.00
265 -	2	0	1.00	0.00	1.00	1.00
270 -	2	0	1.00	0.00	1.00	1.00
275 -	2	0	1.00	0.00	1.00	1.00
280 -	1	1	1.00	0.00	1.00	1.00

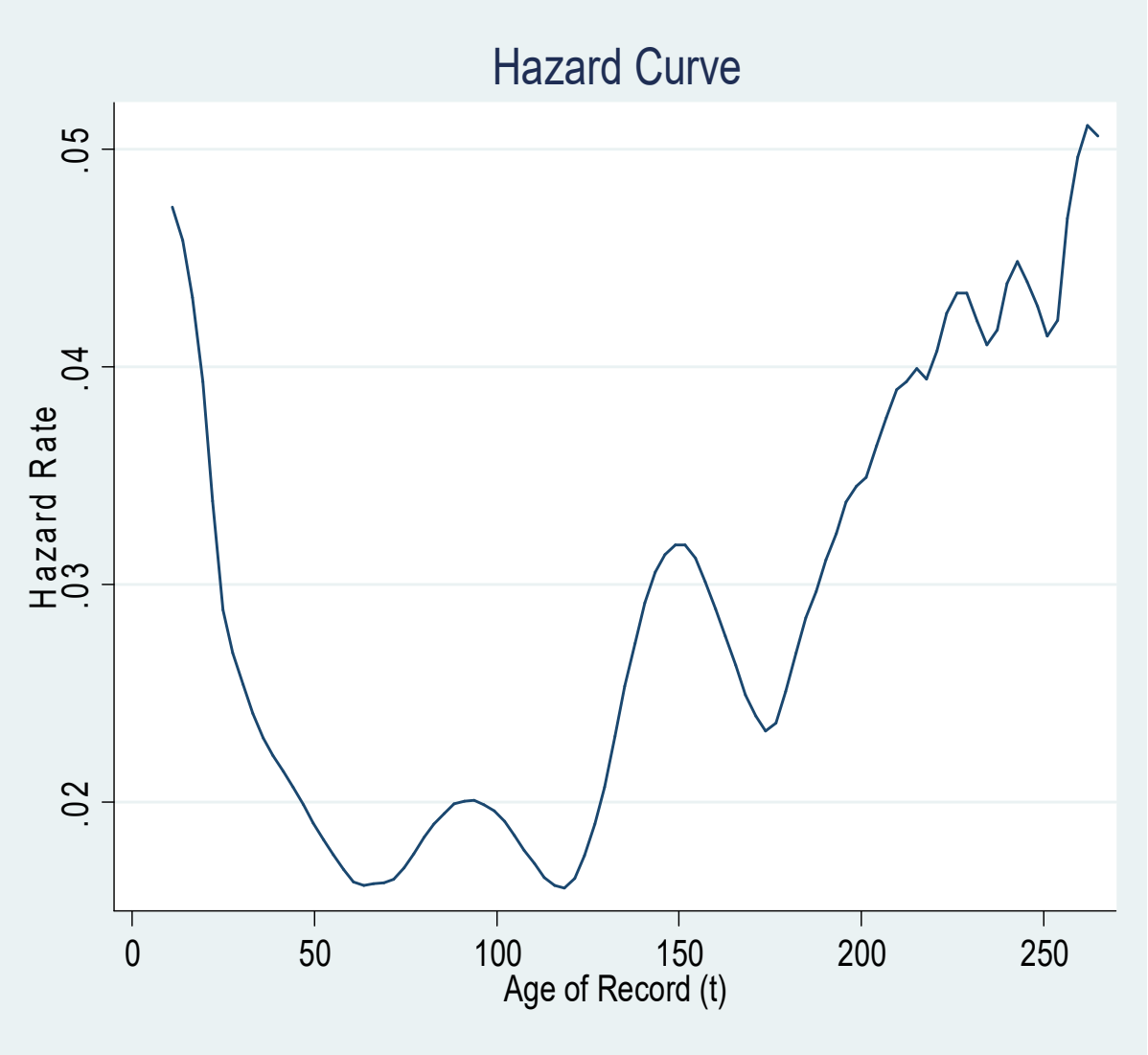


Figure 4.28: Hazard Curve of dormancy time 1990 – 2014 data

Evaluation of the form of the hazard rate (modelling the hazard form)

The result of the test for validity of Weibull distribution assumption using Weibull probability plot of Kaplan-Meier log-log Survival plot, *logH(t),* against log survival time, *log(t),* is shown on Figure 4.29. The plot indicated a straight line relationship between *logH(t)* and *log(t),* increasing monotonically. The intercept of the straight line was approximately (– 0.3980) with a slope of approximately 0.3968. From this results, the value of the shape parameter, γ, for two parameter Weibull distribution was estimated as:

$$\gamma^* = exp(-\ 0.3980\ ) = 0.6717 \text{ and}$$

the estimated hazard rate, $\lambda^* = 0.3968.$

$$\text{and } \lambda(t) = \lambda p t^{p-1}$$

where p and $\lambda, > 0$

The linearity of ln(t) of $S(t) = \exp(-N^p)$

$$\Rightarrow \ln[\ln S(t)] = \ln(\lambda) + p\ln(t)$$

Where the intercept ln(λ,), and the slope = *p.*

The estimated value of the shape parameter, γ, is less than 1, suggesting a decreasing hazard, λ, for the dormancy time of patient medical records created from 1990-2014.

Given the *pdf* of two parameter Weibull distribution as:

$$f(t) = \frac{\lambda}{\gamma}\left(\frac{t}{\gamma}\right)^{\lambda-1} e^{-\left(\frac{t}{\gamma}\right)^{\lambda}}$$

And substituting the estimated values, we can then write:

$$f(t) = \frac{0.3968}{0.6717}\left(\frac{t}{0.6717}\right)^{0.3968\ -1} e^{-\left(\frac{t}{0.6717}\right)^{0.3968}}$$

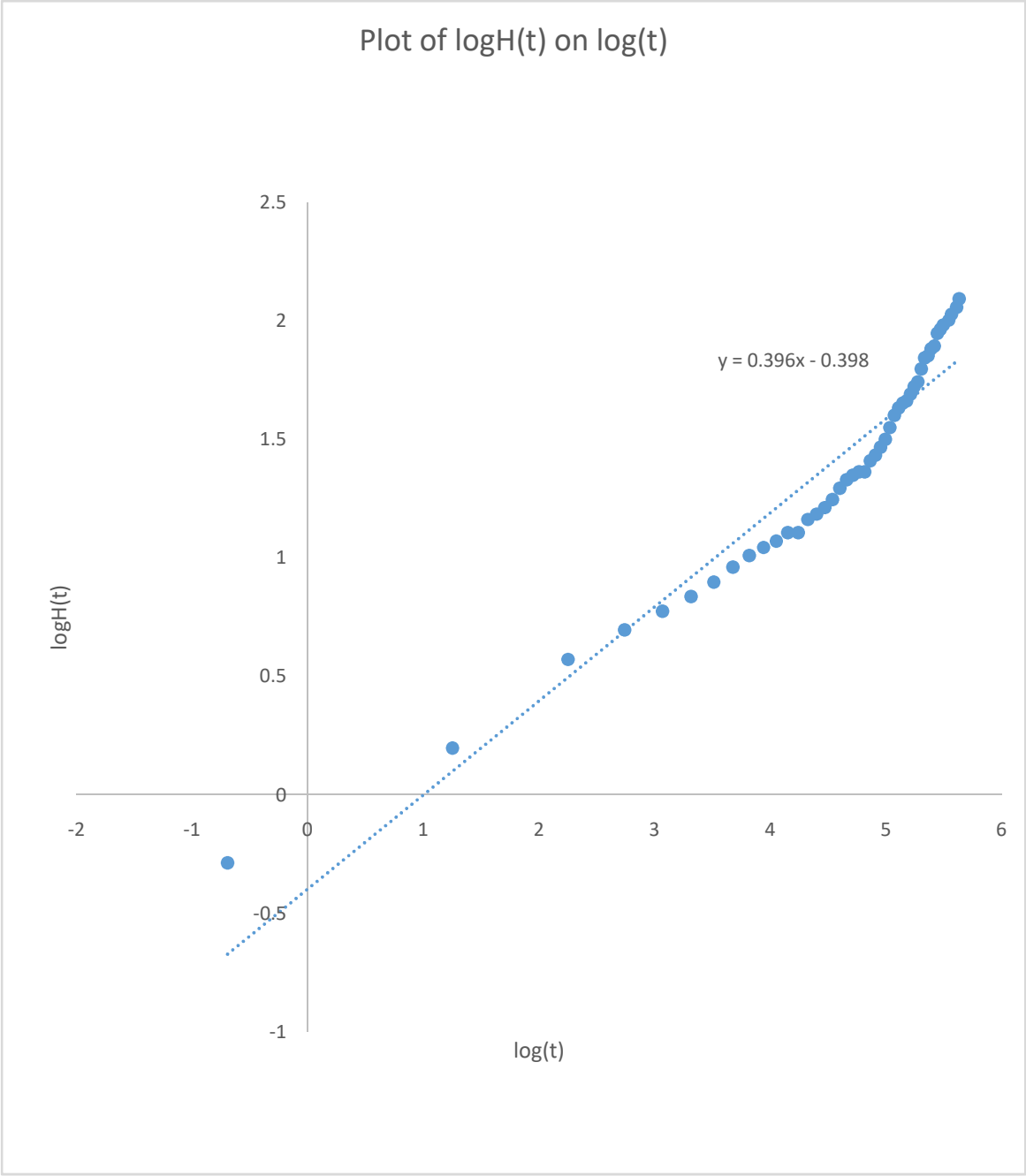


Figure 4.29A Weibull plot of *log–log S(t)* on *log (t)* with line fitted

4.8.5 Influence of patient characteristics on hazard ratefor records created 1990-2014

A semi-parametric model (Cox Proportional Hazard) and parametric models (Exponential and Weibull) were fitted to dormancy time data of patient records to test for model with the best fit and to identify effects of some selected patients characteristics on records dormancy.

4.8.5.1 Non-parametric approach

Schoenfeld's global test of Cox proportional hazard model assumption

Table 4.64 show the result of the global test for the proportional hazard assumption. The significant ($P<0.05$) of the global test implies that the sample data is invalid for the proportional hazard assumption and that the hazard of subject subgroup are proportional over follow up period and thereforethe global test implies indicated that for the data set used the assumption of PH is violated.

Table 4.64: Global Test for Proportional Hazard Assumption

Dormancy time Assumption test	Chi-square	df	p-value
Proportional Hazard Assumption	30.94	7	0.00

Graphical test for Proportional Hazard Assumption

The graph of the log-log Kaplan Meier estimates by dormancy time comparing patient's gender while adjusting for age, zone, clinics, admission status, and surgery and treatment outcome shows that the two line (male and female) are not parallel to each other and therefore substantiate the claim that the proportional assumption is not valid for the dormancy time data, figure 4.30.

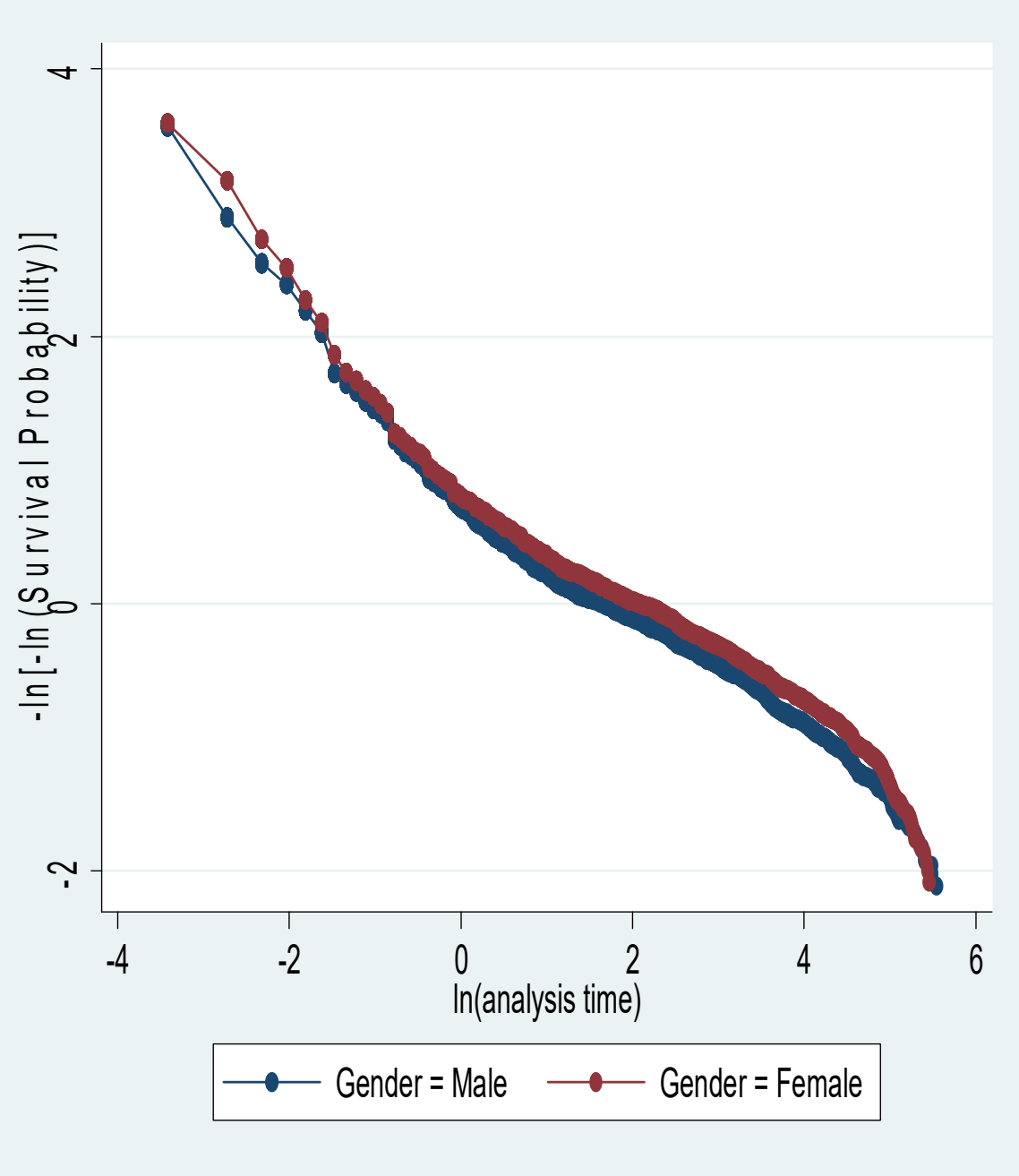


Figure 4.30: Graph showing violation of Proportional Hazard Assumption 1990 - 2014 merged.

Fitting Cox Proportional Hazard Model:

Table 4.65 below shows the Cox regression analysis result that succeed the global test above. From the result time-to-dormancy of patients record are affected by patient age (HR=0.95, p-value=0.000), gender (HR=0.87, p-value=0.000), State of residence (HR=0.89, p-value=0.000), clinic attended (HR=0.97, p-value=0.002), admission status (p-value=0.000), surgery status (HR=0.64, p-value = 0.000) and treatment outcome (HR=1.62, p-value = 0.000) as they are all significant at 0.05% α level respectively. Also, hazard ratio that assess the risk magnitude and likelihood are presented in the table for each group of patients characteristic. Hazard ratio with interval that pass through 1 are insignificant as seen for the adolescent (10-20) group under Age and the referred group under treatment outcome:

Table 4.65: Cox regression of medical record dormancy time on patient characteristics

Variable	factors	Haz. Ratio	z	p>|z|	95% CI	
Age group		**0.95**	**-4.26**	**0.00**	**0.93**	**0.97**
	60+	0.79	-4.29	0.00	0.71	0.90
	31-60	0.82	-4.06	0.00	0.71	0.90
	21-30	0.94	-1.14	0.25	0.84	1.04
	10-20	0.91	-1.61	0.10	0.81	1.02
	<10 years (rc)					
Gender		**0.87**	**-4.94**	**0.00**	**0.08**	**0.91**
	female	0.85	-4.54	0.00	0.82	0.92
	male (rc)					
Residence in Oyo State		**0.89**	**-4.02**	**0.00**	**0.84**	**0.94**
	Oyo	0.88	-4.27	0.00	0.83	0.93
	others (rc)					
clinics		**0.97**	**-3.15**	**0.00**	**0.96**	**0.99**
	Others	0.89	-3.31	0.00	0.84	0.95
	GYNE	0.78	-3.77	0.00	0.68	0.88
	CHOP	0.67	-3.85	0.00	0.0.5	0.82
	SOP	0.85	-3.08	0.00	0.77	0.94
	MOP(rc)					
Patient Admission status		**1.22**	**5.64**	**0.00**	**1.13**	**1.30**
	Yes	1.24	6.01	0.00	1.15	1.33
	No (rc)					
Surgery done		**0.64**	**-7.58**	**0.00**	**0.57**	**0.71**
	Yes	0.89	-7.00	0.00	0.58	0.73
	No (rc)					
Treatment Outcome		**1.62**	**7.55**	**0.00**	**1.43**	**1.84**
	Transfer	4.35	1.47	0.14	0.61	31.03
	DAMA	2.59	7.24	0.00	2.00	3.35
	Died	4.15	2.01	0.04	1.03	16.67
	Alive (rc)					

4.8.5.2 Parametric approach

Fitting Exponential Model:

Given that the record time-to-dormancy is skewed distributed data, the result of regressing dormancy time on patients characteristics based on exponential model assumption of parameter $\lambda=1$ show that like the Cox model, all patient characteristics; age, gender, state, clinic attended, admission status, surgery and treatment outcome significantly ($HR<1.00$, $p<0.01$) influence their record dormancy time. The significant categorical result generally imply that record of admitted older female patient from Oyo State that are dead, referred or discharge against medical advice after surgery will become dormant earlier than non-admitted younger male patient from other state that are alive and didn't pass through surgery treatment. Table 4.66 below shows the Exponential regression model for the explanatory variables.

Table 4.66: Exponential regression model of record dormancy patient characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.93**	**-5.88**	**0.000**	**0.91**	**0.95**
	60+	0.72	-5.91	0.00	0.65	0.81
	31-60	0.73	-6.60	0.00	0.66	0.80
	21-30	0.91	-1.75	0.08	0.82	1.01
	10-20	0.82	-3.23	0.00	0.74	0.92
	<10 years (rc)					
Gender		**0.74**	**-10.26**	**0.00**	**0.70**	**0.79**
	female	0.75	-9.46	0.00	0.71	0.80
	male(rc)					
State of Residence		**0.81**	**-7.32**	**0.00**	**0.77**	**0.85**
	Oyo	0.80	-7.60	0.00	0.76	0.85
	others(rc)					
clinics		**0.95**	**-6.05**	**0.00**	**0.93**	**0.96**
	Others	0.80	-6.61	0.00	0.75	0.85
	GYNE	0.63	-6.88	0.00	0.56	0.72
	CHOP	0.53	-6.24	0.00	0.43	0.64
	SOP	0.71	-6.57	0.00	0.64	0.79
	MOP(rc)					
Patient Admitted		**1.22**	**5.69**	**0.00**	**1.14**	**1.30**
	Yes	1.26	6.44	0.00	1.17	1.35
	No (rc)					
Surgery done		**0.53**	**-10.72**	**0.00**	**0.47**	**0.59**
	Yes	0.56	-967	0.00	0.49	0.63
	No (rc)					
Treatment Outcome		**2.53**	**14.48**	**0.00**	**2.23**	**2.87**
	Transfer	46.17	3.83	0.00	6.49	328.44
	DAMA	6.01	13.64	0.0	4.65	7.79
	Died	37.27	5.11	0.00	9.29	149.40
	Alive (rc)					
_cons	**variable**	**0.05**	**-13.91**	**0.00**	**0.04**	**0.06**
	categories	**0.10**	**-43.31**	**0.00**	**0.09**	**0.11**

Fitting Weibull model

The result of fitting Weibull model to the skewed distributed records time-to-dormancy under the assumption that the Exponential model fail and the model fit Weibull model of parameter ϒ=λ=1. Similar to Cox and Exponential model above, all patient characteristics like; Age, Gender, State, Clinics Admission Status, Surgery and Treatment Outcome are significant at HR< 1 and P<0.01. Therefore all the characteristics are important predictors of record dormancy time. Table 4.67 below shows the Weibull regression model result with two significant (P<0.01) extended parameter for the categorical and sub-categorical characteristics.

Table 4.67: Weibull Regression Model of medical record dormancy time on patient characteristics

Variable	Factor	H_Ratio	z	p>|z|	95% CI	
Age group		**0.95**	**-4.29**	**0.00**	**0.93**	**0.97**
	60+	0.79	-4.35	0.00	0.71	0.87
	31-60	0.82	-4.02	0.00	0.75	0.90
	21-30	0.95	-0.86	0.38	0.86	1.05
	10-20	0.91	-1.54	0.12	0.81	1.02
	<10 years (rc)					
Gender		**0.85**	**-5.48**	**0.00**	**0.81**	**0.90**
	female	0.86	-5.07	0.00	0.81	0.91
	male (rc)					
State of Residence		**0.88**	**-4.53**	**0.00**	**0.83**	**0.93**
	Oyo	0.87	-4.75	0.00	0.82	0.92
	others					
clinics		**0.97**	**-3.51**	**0.00**	**0.95**	**0.98**
	Others	0.88	-3.56	0.00	0.83	0.94
	GYNE	0.77	-3.83	0.00	0.68	0.88
	CHOP	0.65	-4.22	0.00	0.53	0.79
	SOP	0.87	-2.69	0.00	0.78	0.96
	MOP (rc)					
Patient Admitted		**1.22**	**5.85**	**0.00**	**1.14**	**1.31**
	Yes	1.24	6.22	0.00	1.16	1.34
	No (rc)					
Surgery done		**0.63**	**-7.76**	**0.00**	**0.56**	**0.71**
	Yes	0.64	-7.25	0.00	0.57	0.72
	No (rc)					
Treatment Outcome		**1.16**	**7.42**	**0.00**	**1.41**	**1.82**
	Transfer	0.09	1.41	0.15	0.57	29.11
	DAMA	2.55	7.12	0.00	1.97	3.30
	Died	4.14	2.01	0.04	1.03	16.61
	Alive (rc)					
_cons	**variable**	0.36	-10.77	0.00	0.30	0.43
	categories	0.48	-12.08	0.00	0.42	0.54
/ln_p	**variable**	-0.70	-67.09	0.00	-0.72	-0.68
	categories	-0.70	-66.77	0.00	-0.72	-0.68
P 1/p	**variable**	0.49 2.02			0.48 1.98	0.50 2.07
	categories	0.49 2.02			0.48 1.98	0.50 2.06

Regression models on patient characteristics Cohort 1-5 (1990-2014) merged

Regression on patient characteristics from the three models of dormancy time for Cohort 1-5 merged (1990-2014) show, Table 4.68, the hazard ratios and p-values for records of admitted patients and treatment outcome with high risk compared to other patient characteristics. The hazard ratios for surgery status, gender and state of residence to be less than one while those for age and clinic attended are close to one for all the three models.

Table 4.68: Regression models on patient characteristics Cohort 1-5 (1990-2014) merged

Explanatory Variables	Cox	Exponential	Weibull
Age of Patients (rc = <10)	HR=0.95 (P=0.000)	HR=0.94 (P=0.000)	HR=0.93 P=0.000)
Gender (rc=male)	HR=0.87 (P=0.000)	HR=0.75 (P=0.000)	HR=0.86 (P=0.000)
Residence (rc=other States	HR=0.89 (P =0.000)	HR=0.81 (P=0.000)	HR=0.88 (P=0.000)
Clinics (rc=MOP)	HR=0.97 (P=0.000)	HR=0.95 (P=0.000)	HR=0.97 (P=0.000)
Patient Admitted*	HR=1.22 (P=0.000)	HR=1.22 (P=0.000)	HR=1.23 (P=0.000)
Surgery status	HR=0.64 (P=0.000)	HR=0.53 (P= 0.000)	HR=0.64 (P=0.000)
treatnt outcome* (rc=Alive	HR=1.62 (P=0.000)	HR=2.53 (P=0.000)	HR=1.16 (P=0.000)

4.9 Diagnostic assessment of distribution, survival time and model of time-to-dormancy of medical records (cohorts 1 to 5)

4.9.1 Comparing some patient characteristics of the cohorts and the combined data

Table 4.69 shows the distribution of some patient characteristics for cohorts 1 – 5. The number of records that became dormant on the day of creation (one-day-active) were 30.6%, 23%, 17.8%, 30.0% and 21.5%, in the 1st, 2nd, 3rd, 4th and 5th cohorts respectively, the highest was observed in cohort 1, (1990-1994),while the lowest was observedin the cohort 3, (2000-2004). However on the whole 24.5% of the records were never used beyond the day of creation and consequently became dormant. The percentage of male patients were slightly higher than that of their female counterpart for all cohorts and the combined data except for cohort 2. Most of the patients registered were adults between the ages of 31-60 years, while the least were adolescents aged 10 - 20 years. Patients whose records were observed,were mostly from Oyo Statecompared to all other states, even when all other states were put together, except for the 1st and 2nd cohorts. The 1st and 2nd cohorts (1990-1994 and 1995-1999) recorded the highest attendance rate for Surgical Outpatient Clinic, while in the 3rd, 4th, 5th cohorts, and the combined data, the attendance rate was highest at the Medical Outpatient Clinic. On the whole, most of the records observed were from MOP (39.30%), which is about the number of records for other clinics put together. Between 22% and 45% of the medical records observed for dormancy time were records of admitted patients, however on the whole(1990-2014), 30.34% of the records belong to patients that were admitted at one time or the other. The 2nd cohort had the highest number of surgery cases(15.5%) and lowest (3.09%) for the 3rd cohort. Result of treatment outcome show that deaths were not recorded except in cohorts 1 and 2 which this was as low as 1.3%. The number of patients Discharge Against Medical Advice (DAMA) ranged between 0.0% in cohort 2 to 2.57% in cohort 5.

Table 4.69 Comparing distribution ofsome patient characteristics over the cohorts and the combined data

Parameters	Factors	COHORT					Merged
		1	2	3	4	5	1-5
		1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	1990-2014
	Sample size	1537	1537	1537	1537	1537	7685
t = 1 day excluded for failing at point of creation		470 (30.6%)	354 (23%)	274 (17.8%)	460 (30.0%)	330 (21.5%)	1888 (24.5%)
t > day		1067	1183	1263	1077	1207	5797
Gender	male	51.11%	47.26%	51.77%	55.52%	56.44%	52.38%
	female	48.89%	52.74%	48.23%	44.45%	43.56%	47.62%
Age range (years)	< 10	19.70%	20.08%	10.46%	15.75%	8.71%	14.61%
	10 - 20	13.88%	11.28%	10.54%	9.72%	8.13%	25.25%
	21 -30	21.01%	21.89%	17.51%	13.68%	16.00%	18.02%
	31- 60	35.75%	33.75%	42.23%	41.04%	47.68%	40.37%
	> 60	9.66%	13.00%	18.86%	19.81%	19.49%	16.35%
Clinic attended	MOP	22.16%	18.19%	52.73%	29.31%	69.93%	39.30%
	SOP	24.95%	24.17%	1.77%	1.89%	1.46%	10.50%
	CHOP	2.89%	7.56%	0.40%	0.85%	0.60%	2.42%
	GYNAE	14.16%	13.97%	0.56%	1.23%	1.03%	5.99%
	Others	35.84%	36.12%	44.54%	66.73%	26.98%	41.79%
State of residence	Oyo	48.51%	49.47%	54.17%	75.84%	50.13%	55.37%
	others	51.49%	50.53%	45.82%	24.16%	49.87%	44.63%
Admitted	Yes	31.02%	42.01%	22.09%	25.96%	30.90%	30.34%
Surgery	Yes	10.4%	15.10%	3.09%	11.72%	2.49%	8.36%
Treatment outcome	Alive	99.62%	98.97%	99.68	98.21%	97.34%	98.77%
	Died	0.09%	1.03%	0.0%	0.09%	0.0%	0.03%
	DAMA*	0.28%	0.0%	0.32%	1.69%	2.57%	1.18%

4.9.2Comparing Kaplan-Meier Survival curves of the cohorts

Graphical approach

Figure 4.31 show the graphic assessment of survival curves for the five cohorts and the combined data. Survival functions ranged from 0.0 to 1.0 with curves decreasing sharply with increasing age of records/dormancy time for all cohorts The curve for all five cohorts and combined data followed samilar shape and pattern,tending toward zero as time reaches end point for all five cohorts and the combined data. End points of each curve decreased as the period covered by each cohorts get closer to point of data analysis. The curve for the first cohort (1990-1994) and that of the combimed data (1990-2014) have the similar pattern,shape and lenght.

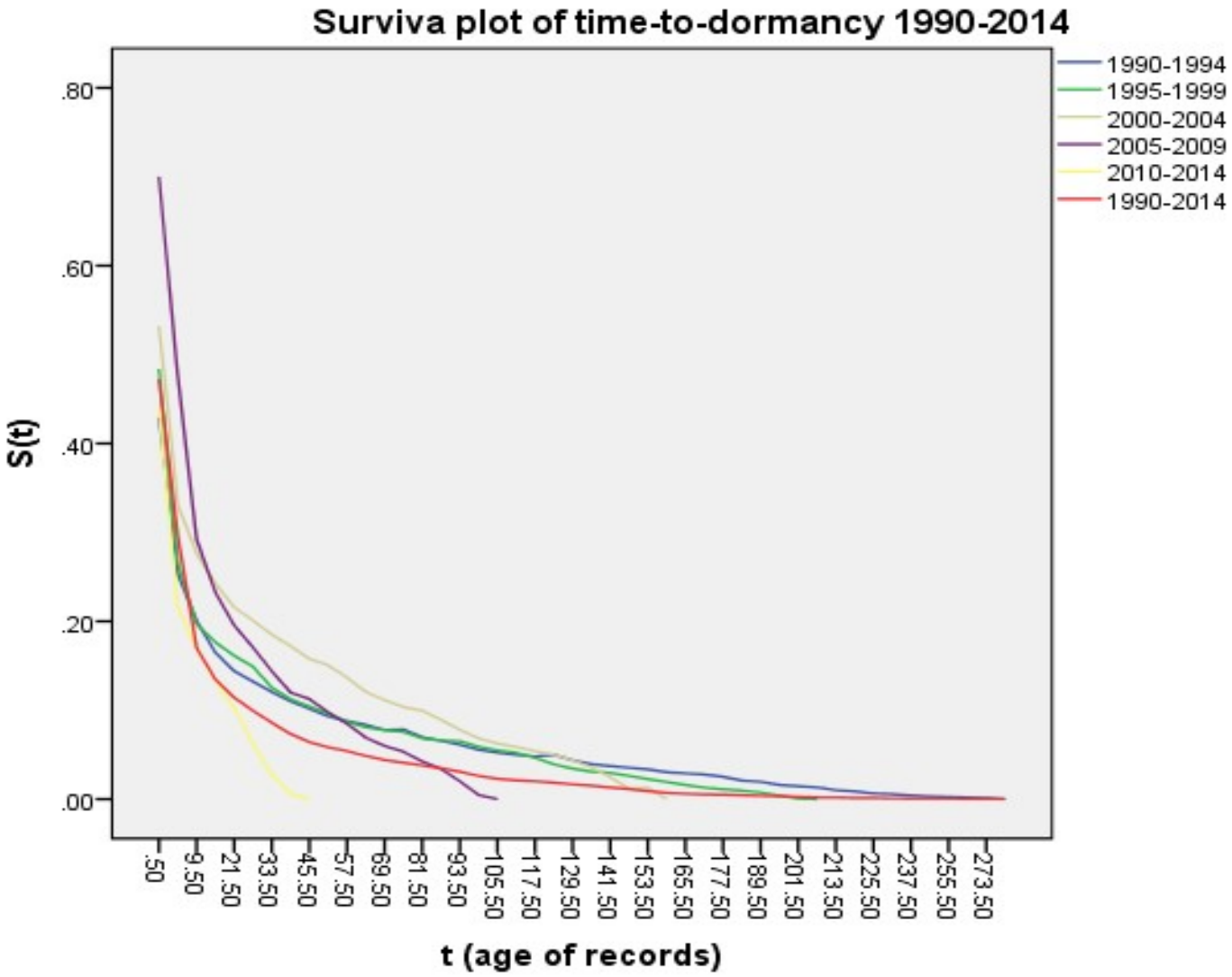


Figure 4.31 Survival plots of dormancy time for the five cohorts and merged data

Log-rank test to compare survival curve of cohorts 1 to 5

Table 4.70 show results of the log-rank test for equality of the survival curve for the five cohorts and the log-rank test for trend that assessed the differences in survival between cohorts under the assumption that the time to event data are in a naturally ordered sequence (such that patients' records in early years are either censored or observed earlier) and the grouping of dormancy time data was based on ordinal or ranking scale.

Both test were however significant ($P<0.001$), which implied that the survival curve for the five cohorts were not equal and also there was no significant trend among the survival curves of the cohorts

Table 4.70Log-rank test for equality of survivor curves of cohort 1-5

Cohort	Events observed	Events expected
1	1064	1179.71
2	1183	1209.51
3	1262	1431.45
4	1075	1099
5	1207	870.48
Total	**5791**	**5791**

Test of equality

Chi 2 (4) = 172 04
Pr. > chi.2 =0.000

Test of trend

Chi 2 (4) = 172 04
Pr. > chi.2 =0.000

4.9.3 Comparing estimates of percentiles of dormancy time in cohorts 1-5 andcombined data

Table 4.71 show assessmentof selected percentiles for 25th, 50th, 75th and 95th in respect ofthe fivecohorts and combined data.

Estimates of the dormancy time of observed medical records for 25thpercentiles show that 25% of the observed records were dormant in about 0.46 month in all five cohorts and when the data was combined into a single sample,except forthe 4th cohort(0.69 months) that was slightly higher.

The 50thpercentiles (the median dormancy time)which is the point at which 50% of the records became dormant was observedto increasein dormancy time with subsequent cohorts as they get closer to the point of analysis. The median dormancy timewas 1.93, 2.30,3.05 and 3.84 month for the1st,2nd,3rd, and 4thcohorts respectively. The median dormancy time however dropped to 1.51 months in the 5th cohort (last cohort),while the observed median dormancy timefor the combined data was 2.29 months.Estimatesfor the 95th percentiles showed a reversed pattern to the 50th percentile estimate by decreasing with subsequent cohorts. Estimateshow that for the 1st cohort 95% of the records observed had the highest dormancy time of 151.89months (12.66 years)and the lowest of 34.75 months (2.90 years) in the 5th cohort (last cohort). A dormancy time of 101.51months (8.46 years) was observed for the combined data. The results showed that the5thcohort recorded the lowest dormancy time for all percentiles points.

Table 4.71Estimates of selected percentiles of the survival curve for cohorts 1-5andcombined data

Cohort		Percentiles							
		25th		50th		75th		95th	
		t	SE	t	SE	t	SE	t	SE
1990 -	**1st**	0.46	0.04	1.93	0.16	17.12	1.86	**151.89**	12.332
1995 -	**2nd**	0.46	0.04	2.30	0.19	13.93	2.30	129.85	8.99
2000 -	**3rd**	0.49	0.03	3.05	0.37	28.45	3.66	134.34	3.66
2004 -	**4th**	**0.69**	0.05	**3.84**	0.44	23.65	1.90	84.07	2.70
2010 -	**5th**	0.42	0.02	1.51	0.12	10.61	1.10	34.75	5.65
1990-2014	**Merged**	**0.45**	**0.01**	**2.29**	**0.09**	**17.57**	**0.84**	**101.51**	**6.09**

***estimates in months**

4.9.4Comparingforms and shapes ofhazard curve of cohorts and the combined data

Graphical approach

Figure 4.32 show the hazard curves for the five cohorts and the combined data. The curves show a similar form and shape, exhibiting a sharp decrease by all the curves at the initial time, t=3.5 months and continue to decrease as dormancy time (age of records) inceases for all curves. This was followded by a period of constant steady irregular movement as dormancy time(age of records)increased and then the curves exhibiting a sharp upward rise with a steady upward increase to end points, all the curves making a bathtub shape. The dormancy time (age of records) at which the hazard curve begin to decrease and attain constant movement appeared same for all cohorts and combined data, however the period over which the curves remain contant to when each curve starts the upward rise varies with dormancy time, t, (age) of records. The hazard curves for the first cohort (1990-1994) and that of the combined data, 1990-2014, have similarshape and pattern. The closer a cohort is to the point of analysis (end of the study) the shorter the period over which it remained constant. Cohort 5 (2000-2014) had the shortest constant period while the 1st cohort (1990-1994) and the combined(1990-2014) had the longest period of constant hazard rate.

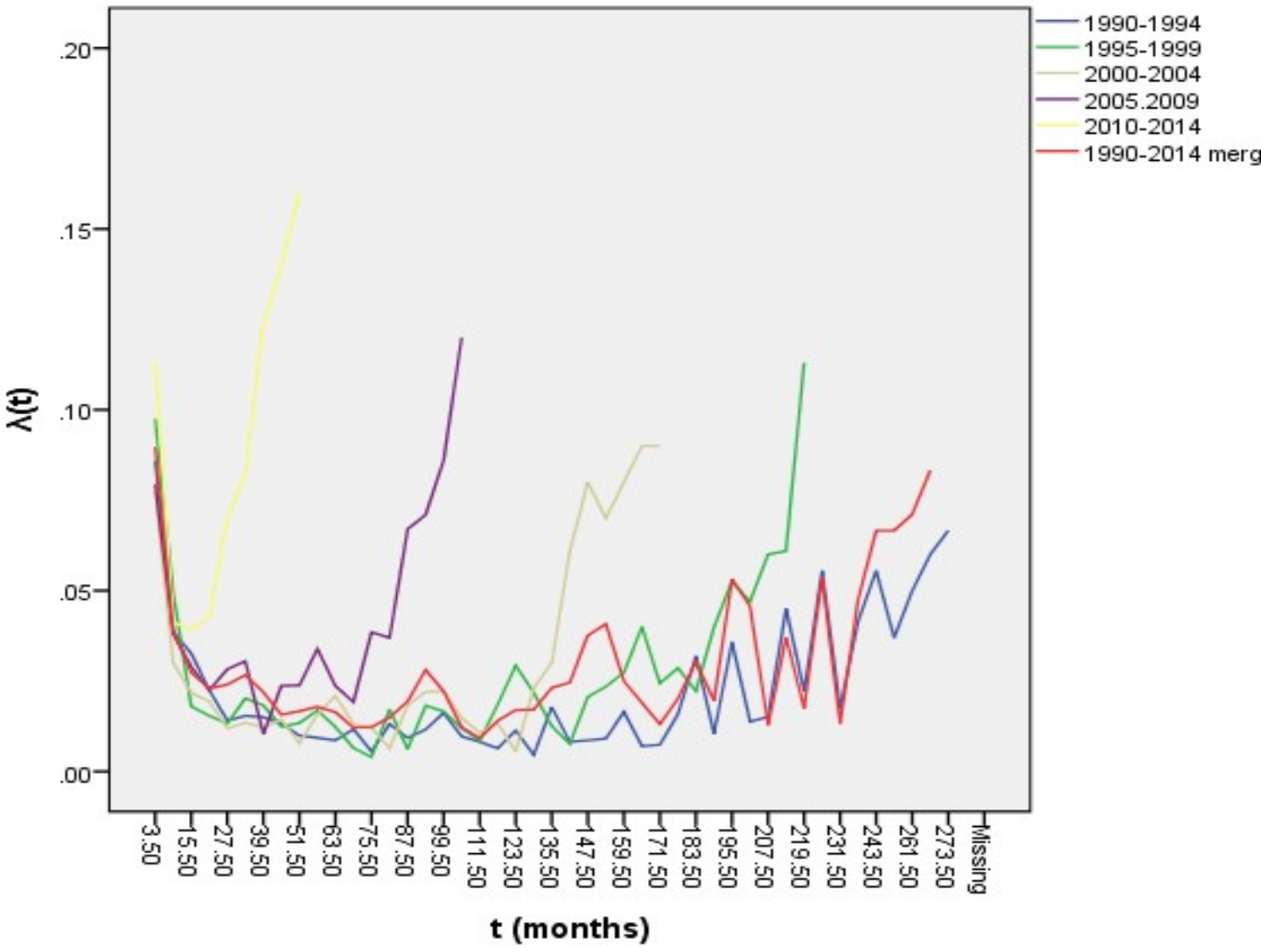


Figure 4.32 Hazard Curves of the five cohorts and merged sample

Comparing estimates of shape parametersforcohorts 1-5 and combined data

Table 4.72, show estimatedshape parameter for test of Weibull distribution from the intercepts of the fitted lines to the plots of log-log of S(t) *(log H(t))* on log of time, t, for the cohorts and the combined data. The results show estimated value of the shape parameters, γ, for the first four cohorts and the combined data were less than unity indicating a decreasing hazard rate while estimate for the 5^{th} cohort, 2010-2014 indicated a shape parameter greater than unity indicating increasing hazard rates. The non-unity of any of the shape parameter is an indication that the distribution of time-to-dormancy data of medical records cannot be exponential.

Table 4.72. Estimates of shape parametersof cohorts 1-5 and combined data

Parameter	Cohorts					
Estimate values*	1	2	3	4	5	1 – 5
Intercept	-0.55	-0.14	-0.14	-0.15	0.10	-0.39
shape parameter (γ*)	0.57	0.86	0.88	0.86	1.82	0.67
hazard rate (λ*)	0.35	0.35	0.35	0.28	0.60	0.39

4.9.5. Comparing model of the cohorts and the combineddata

Global Test for Proportional Hazard Assumption

Table 4.73 show results of the global tests for the proportional hazard assumption for combined data and the five (5) cohorts. The result of the test for cohort 1, 2, and 3 were insignificant an indication that the sample data did not violate the proportional hazard assumption, that the hazard of subject subgroup are proportional over dormancy time and therefore the null hypothesis was not rejected. However the results of the test for cohorts 4, 5 and combined data were significant, indicating that the sample data did violate the proportional hazard assumption.

Table 4.73: Global test for PH of cohorts of 1-5 and combined data

Cohort	Period	Dormancy time Assumption test	Chi-square	p-value
1	1990-1994	Proportional Hazard Assumption	6.29	0.50
2	1995-1999		2.55	0.92
3	2000-2004		4.13	0.76
4	2005-2009		19.39	0.00
5	2010-2014		22.61	0.00
Merged data	1990-2014		30.94	0.00

4.9.6 Comparing hazard ratios of record dormancy time on some patient characteristics for Cox, Exponential and Weibull models

Cox model

Table 4.75 show the results of the hazard ratios of fitting regression of dormancy time on some patient characteristics with Coxregressions models. The risk of dormancy among records of admitted patients and those with treatment outcomes are higher than for other explanatory variables, whereas age and clinic attended by patients was close to 1 an indication that clinic attended do not have much influence on record dormancy.

Table 4.74:Hazard ratios of dormancy time on patient characteristicsmodelled with Cox

Predictors	COHORTS					
	1	2	3	4	5	1-5
Age (rc <10)	0.93	0.90	0.93	0.97	1.102	0.95
Gender (rc male)	1.10	0.85	0.76	0.85	0.91	0.87
Oyo State (rc others)	0.90	0.87	0.84	0.90	0.99	0.89
Clinics (rc MOP)	1.00	0.98	0.96	1.00	1.01	0.97
Admission status*	1.19	0.91	1.22	1.37	1.51	1.22
Surgery status	0.84	0.73	0.52	0.61	0.51	0.64
Trt_Outcome* (rc alive)	4.01	1.87	2.13	1.77	1.27	1.62

Exponential model

Table 4.76 show the results of fitting regression of dormancy time on some patient characteristics with Exponentialregressions models. The risk of dormancy among records of admitted patients and those with treatment outcomes are higher than for other patients' characteristics, whereas age and clinic attended was close to one an indication that MOP as reference category do not have influence records dormancy.

Table 4.75:Hazard ratios of dormancy time on patient characteristicsmodelled with Exponential

Predictors	COHORTS					
	1	2	3	4	5	1-5
Age (rc <10)	0.88	0.82	0.90	0.99	1.00	0.93
Gender (rc male)	1.11	0.66	0.64	0.81	0.86	0.75
Oyo State (rc others)	0.77	0.76	0.77	0.83	0.97	0.81
Clinics (rc MOP)	1.01	0.96	0.93	1.00	1.04	0.95
Admission status*	1.05	0.79	1.31	1’53	1.91	1.22
Surgery status	0.89	0.58	0.45	0.52	0.45	0.53
Trt_Outcome* (rc alive)	19.86	4.26	7.39	2.75	1.21	2.53

Weibull model

Table 4.77: show the results of fitting regression of dormancy time on some patient characteristics withWeibull regressions models. The risk of dormancy among records of admitted patients and those with treatment outcomes are higher than for other explanatory variables, as indicated by hazard ratio of above one.However age and clinic attended by patient was close to one which is an indication that they do not influence dormancy time.

Table 4.76:Hazard ratios of dormancy time on patient characteristicsmodelled with Weibull

Predictors	COHORTS					
	1	2	3	4	5	1-5
Age (rc <10)	0.93	0.90	0.93	0.98	1.01	0.95
Gender (rc male)	1.10	0.81	0.78	0.84	0.89	0.86
Oyo State (rc others)	0.88	0.88	0.84	0.88	0.97	0.88
Clinics (rc MOP)	1.0	0.98	0.96	1.00	1.03	0.97
Admission status*	1.17	0.90	1.22	1.43	1.62	1.23
Surgery status	0.82	0.72	0.60	0.57	0.54	0.63
Trt_Outcome* (rc alive)	2.97	1.9	2.2	1.80	1.21	1.16

4.9.7 Test of Models best fit for time-to-dormancy of patient records

Table 4.74 below show the results of comparing the three survival models, Cox proportional hazard, Exponential and Weibull models that best fit the records dormancy time for each cohort data. The model with the minimum log likelihood and equivalently minimise the information lost (from the AIC value) was adjudged as the best model for each record dormancy time data cohort. This result shows that for all the cohorts Weibull parametric model with most number of model parameter (9) and minimum log likelihood value minimises the information lost (with the least AIC value) in estimating the true model than the Cox proportional hazard and the exponential model counterpart. Weibull model in cohort 4 with -2logL of 4170.97 and AIC value of 4188.97 best provide a model close to the true model of record dormancy time among all the five cohorts.

Table 4.77: Test of Models for best fit for dormancy time data of patient records

Cohorts	Period	Model	K-Parameter	-2LogL	AIC-value
1	1990-1994	Cox	7	11061.35	11075.35
		Exponential	8	5904.87	5920.87
		Weibull	9	**4371.85***	**4389.85***
2	1995-1999	Cox	7	10883.71	10897.71
		Exponential	8	5480.28	5496.28
		Weibull	9	**4234.19***	**4252.19***
3	2000-2004	Cox	7	14429.75	14443.75
		Exponential	8	6878.98	6894.98
		Weibull	9	**5422.50***	**5440.50***
4	2005-2009	Cox	7	11488.14	11502.14
		Exponential	8	4878.68	4894.68
		Weibull	9	**4170.97***	**4188.97***
5	2010-2014	Cox	7	132207.74	13234.74
		Exponential	8	5364.29	5380.29
		Weibull	9	**4619.30***	**4637.30***

4.10Verifying estimated dormancy time

Estimate of selected percentiles toverify dormancy time

Estimates of survival timeTable 4.78, for 25^{th}, 50^{th}, 75^{th} *and* 95^{th} percentiles of the result from time difference between *penultimate entry - last entry* and *last entry-point of data analysis*for observed recordsshow that the later (*last-entry to point of data analysis*) is higher than the former (*penultimate entry-last entry)* for all selected percentile points.That is for estimated P_i

$$LC - PC < PS - LC,$$

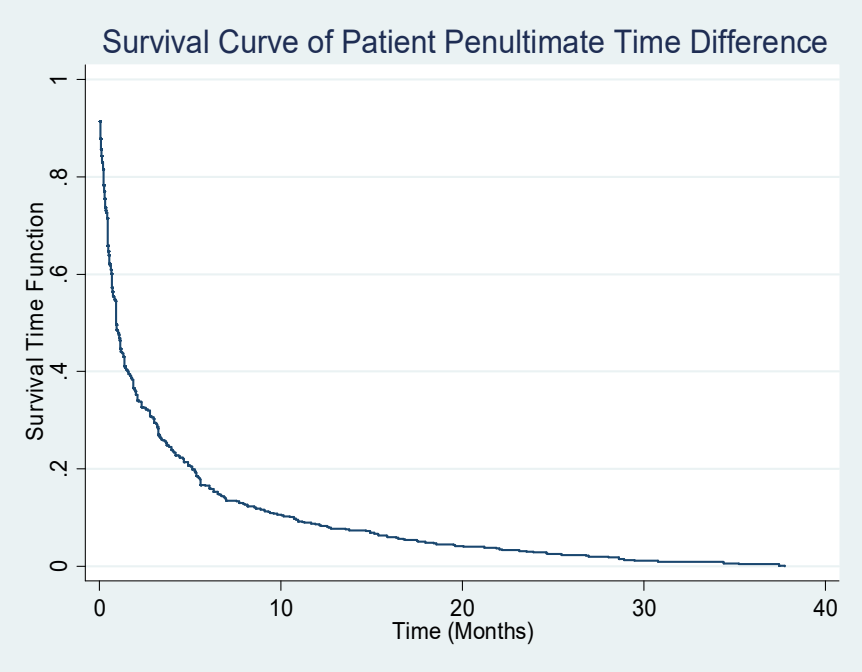
indicating that enough time was allowed for, before the start of data analysis.

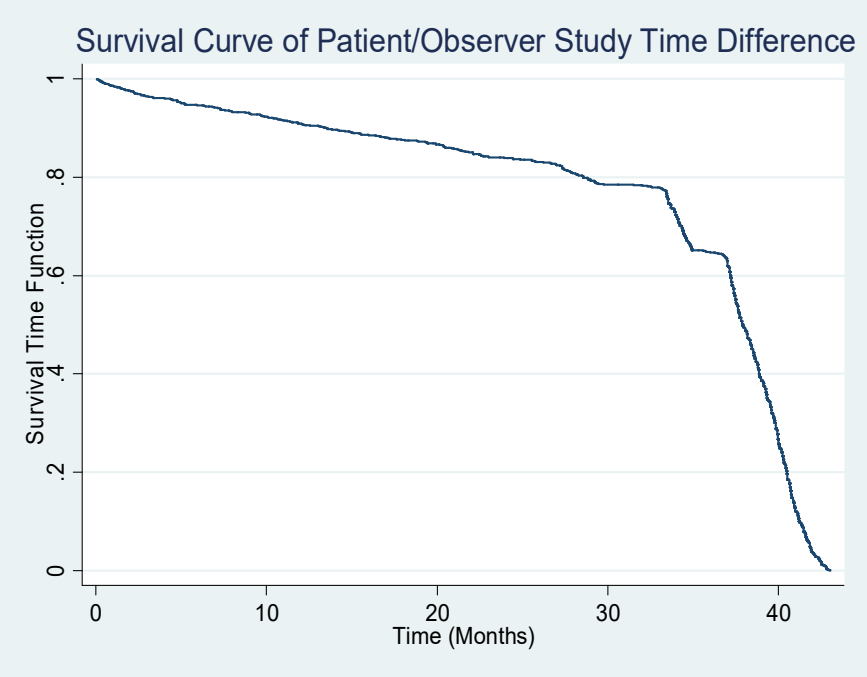
Table 4.78:Estimates of selected percentiles to verify dormancy time

P^{th}	25^{th}	SE	95%CI		50^{th}	SE	95%CI		75^{th}	SE	95%CI		95^{th}	SE	95%CI	
Lc-Pc	0.33	.033	.23	.43	0.93	.06	.93	1.1	3.60	.03	3.6	4.6	17.96	.2.3	15.1	.23.1
Ps-Lc	33.5	0.6	33.3	33.9	37.96	0.1	37.7	38.3	40.2	.07	40.0	40.4	41.93	0.08	41.8	42.1

Survival curve for verification of estimated dormancy time

The survival function ranged between 0.0 and 1.0, decreasing sharply and later gradually as record dormancy time from penultimate to last contact increases and while thesurvival function of the last contact to point of data analysis decreases gradually and later sharply as records dormancy time increases, both tending toward zero as time reaches end point.





. **Fig4.33a Survival curve penultimate/lastcontact time**

Fig4.33b: Survival curve lastcontact/observer-study time

CHAPTER FIVE
DISCUSSION

5.1 Introduction

This study has four goals. First was to find out the statistical distribution of dormancy time of medical records of patients seen between 1990 and 2014 at University College Hospital, Ibadan. Secondly was to estimate the dormancy time for the medical records, thirdly to find out if the distribution and parameters are the same for medical records of patients created over time and fourthly to determine the demographic (age, gender…) and clinical factors (clinic, treatment outcome, …) associated with the time-to-dormancy of medical records of patients seen at the University College Hospital, Ibadan. Based on findings from the study, guidelines would be highlighted to enable the Authority of UCH, Ibadan draft policies to guide the management of medical records that will include schedules for retention, disposal and archiving of medical records.

These issues are important because it is good practice for every healthcare organisation to have medical records management policy in place. Whereas the large volumes of medical records created in the treatment of patient daily require effective and efficient management and control.There are no evidence of documented empirical policy guidelines on retention, disposal and archiving of patient records in Nigerian hospitals particularly in UCH, Ibadan. While no known previous studies on medical records management had attempted to determine the statistical distribution or estimate the dormancy time nor determine factors associated with dormancy of medical record of patients, this study had been able to determine the characteristics and form of the statistical distribution of dormancy time data, estimate dormancy time and determined factors associated with dormancy time of medical records of patients seen at the University College Hospital, Ibadan. Key findings from this study discussed in this chapter includes:

i. Statistical distribution of time-to-dormancy of medical record of patients seen at UCH, Ibadan followed a Weibull distribution
ii. The estimated 95^{th} percentile revealed that 95% of the records had a dormancy time of 151.89 months

iii. The Weibull model provides the best fit compared to Cox PH and Exponential models
iv. Records of patients that died, admitted, attending SOP clinic and those discharged against medical advice had higher hazard ratio,indicating higher risk of dormancy.

5.2 Statistical distribution of time-to-dormancy of medical records

The time-to-dormancy of medical record of patients seen at UCH, Ibadan was found to follow a typical *time-to-event* data that can be addressed by survival analysis techniques. This was evident by the form and shape of the characteristics of the K-M survival curve that exhibited a distribution with a decreasing survival function that range between 0.0 and 1.0. The curve decreases as dormancy time (age) of patient records increases and tends towards zero as time reaches end point.

The hazard plot indicated a rise at the initial time point and decreasesedsharply as dormancy time increased and then assume a steady constant movement for some time before increasing with a sharp rise followingsteady constant upward movement till it reaches end point and therefore making a bathtop shape. This combination of decreasing, constant and increasing hazard rates, often referred to as early life, useful life and old age of the hazard curve, surpports the Records Life Cycle theory of record management which likened records to an organisim that have a creation phase (being born), active phase and lastly inactive phase when the record become dormant,(Penn, Pennix and Coulson 1994, Shepherd and Yeo 2003).

These three phases of decreasing, constant and increasing hazard functions often combine to produce the well-known bathtub curve typical of Weibull distribution, (Wang, 2000; Xie, 2002; Collet, 2003; Lee and Wang, 2003; Kleinbaum and Klein, 2012). It then follows that hazardof dormancy of medical records created at UCH, Ibadan, was relatively high at the early life ofcreation and this can be explained by high activity period from patient influx of both short (acute) and long term (chronic) conditions, coupled with the one-day-active records. This phase was then followed by a relativelyconstant hazard rate of dormancy as a result of none return of patients with one-day-active records and the dischargeof acute cases leaving morerecords of patients with

long-term (chronic) conditions.After this period of relatively constant hazard rate, the rate of dormancy from improved health status, discharges (including deaths), and other factors ledtoa rise in hazard rate and consequently to the bathtubshape. The relationship between bathtub-shaped hazard plot and Weibull distribution had been established in various studies of time-to-event data, (Wang, 2000; Xie, Lai, and Murthy, 2002; Zhang, 2004; Wajid and Khan, 2006; Mustafa, El-Desouky and AL-Garash, 2016)

5.3Dormancy time of medical records of patients seen at UCH, Ibadan

One-day-active records

Findings in this study show that large number of patient records were only active for one day, evident by a single entry.Hence they became dormant on first day of creation. This was as a result of the patients who do not return for a second visit/contact hence their records were never used after the first contact. The implication is retaining large number of dormant recordsamong active records resulting to about a quarter of the filing shelves (primary storage) are occupied by inactive/dormant records. Keeping these one-day-active records in the primary storage area are not cost effective and inconsistent with good records management practice. The practice was viewed as uneconomical, inefficient and ineffective approach to records management and it is not in line with global practice. These one-day-active records need to be weeded from the filing shelves and moved to a warehouse (secondary storage) to create space for new records. Since more than 95% of patients return for a second contact within 5.95 months of the first visit, a patient that failed to return for a second contact after the first contact within this periodcan have the record removed from the filing shelves. This is consistent with Aduku and Abdul (2012) and Records Management Bulletin (2012) that records become less valued with time and that 90% of the use of a record takes place during the first 3 months after it was created. The habit of retaining the one-day-active records establish the fact that large number of dormant records are kept along with active records in UCH, Ibadan. This is in agreement with Ngulube (2011) that it is common to see records that do not support current operations clog the records department or unused records occupying expensive space and equipment due to lack of policies regarding records retention, disposal and archiving. Retaining such records is rather uneconomical to the hospital

management considering the cost of maintenance and accessingneeded records buried among dormant/inactive records.

Dormancy time of medical records that survived beyond the first day of creation

Estimate of survival (dormancy) time of medical records that survived beyond the first day of creation, having excluded the one-day-active records, showedthat the active life (dormancy time) of medical records created at UCH, Ibadan was151.89 months (12.66 years) with a 95% CI 72-179.06. This was seen as the most convenient time to safely weed dormant medical records. This findings is consistent with the study onretention of medical records of 30 hospitals in Isfahan, Iran, by Tavakoli and Jahanbakhsh, (2013), where 44%of the hospitals retained inpatient records for 15 years, 20% for less than 15 years, and 36% for more than 15 years, while 26.1% of hospitals retained outpatient records for 5 years, 56.2% for less than 5 years, and 15.4% for more than 5 years. On the whole, the reported retention period vary between 3 to 25 years. From the foregoing a retention of 13 years would be ideal formedical records of patients seen in UCH, Ibadan,at this point records can be moved to secondary storage. The retention, disposal and archiving policy ofmedical records can therefore be anchored on 13 yearsestimate as a convenient weeding point.

Findings from the study however revealed a pattern in dormancy time in the 25th, 50th and 95th percentiles. The estimated dormancy time (age of records) for P_{25} (the point at which 25% of the records became dormant) was found to be constant for all cohorts and merged data. Estimates at the P_{50}(the median dormancy time)increased with subsequent cohorts however thepattern was reversed in the estimated dormancy time at P_{95} percentileswhich decreased with subsequent cohorts. Estimates for the 5th cohort was generally found to have the lowest dormancy time for all estimated percentiles pointsin all the cohorts. This trend in dormancy time observed in the 5thcohorts could have been as a result of being closer to the end of thestudy,and most likely the medical records are still active.

5.4 Factors associated with dormancy time

The Weibull model with minimum -2logL and AIC valuesproved the best fit model for determinant of dormancy time for medical record having minimised the information lost compared to the Cox proportional hazard and the Exponential models.

Though the three models identified all patient characteristics, (age, gender, state, clinic attended, admission status, surgery and treatment outcome), as important predictors of record dormancy time as they were all significant at $p<0.05$, Weibull model was the preferred predictor of time-to-dormancy. The risk of medical record dormancy was found to be high with records of females, admitted patients, and patients seen at Surgical Out-patient Clinic. Similarly records of patients'with treatment outcome of death and DAMA were also higherthan other patient characteristics. Intuitively, a patient on admission will definitely receive more attention and continuous medical care with a faster recovery period than an outpatient. Expectedly their records are most likely to have higher risk of dormancy when compared to records of an outpatient.Records of dead patients and DAMA are most likely to become inactive instantly.

5.5 Limitations and strengths of the study

Some limitations need to be borne in mind when interpreting the findings of this study. First is that the observed absence or poor documentation of clinical and socio-demographic information of patients in the medical records. Socio-demographic information had been found to be a strong factor influencing health status of a patient. Some vital information about the patient were either totally omitted or poorly documented. For example, economic and educational status of patients were not recorded, making it impossible to determine the association of these important variables withtime-to-dormancy of patient's record. The age of patients and date of births recorded were in many instances not correlated with each other or were totally omitted from patient's record resulting to such records being excluded from the study. Clinical diagnosis were either absent, or non-diagnostic terms documented as diagnosis and in few instances statements such as abdominal pains, patient can't lift right hand, poor vision etc. were recorded as diagnosis. Many of the medical records were created for none medical issues such as request for eye glasses, medical test for driving license, etc. Secondly, a number of the medical records created were used only once, indicated by a

single entry. Considering the large number of records involved they were excluded and analysed separately as "one-day-active records". This exclusion was done to avoid possible bias in estimation. Thirdly, the way and manner both socio-demographic and clinical data were collected do not suggest need for future statistical analysis, and finally, there was paucity of literature in the area of empirical estimation of records dormancy time and this made the study quite difficult in providing a base for comparison. These limitations however notwithstanding the strength of the study includes the fact that it is a longitudinal study spanning over a period of twenty-five (25) years which makes it valid for generalisation. Secondly the study is the first of its kind that attempts to estimate the duration of time-to-dormancy of records, its distribution and the associated factors,empirically particularly in Nigeria. The study can therefore be classified as a novel research establishing a base for developing benchmarks for policy for the retention, disposal and archiving of patients medical records.

CHAPTER SIX

SUMMARY CONCLUSION AND RECOMMENDATION

6.1 Summary

Records management usually are by-products of business processes and it is a fundamental activity of every organisation for economic, efficiency and effective management. The importance of patient information management have long been recognised as an integrated part of medical practice. A good record management system must incorporate a retention, disposal and archival policy. The creation of records is easy to establish, however most organisations do not have concerns when creating or using information that retention, disposal and archiving issues may arise. These issues necessitate the development of a robust record management policy that must incorporate how long records are retained, mode of disposal and archiving guidelines.

Most health organisations in Nigeria, including the University College Hospital, Ibadan, do not have medical records management policy, let alone a retention, disposal and archiving policy guidelines. The result is keeping both active and inactive (dormant) patient records together on the filing shelves at the expense of efficient and cost effective records management that could support quality healthcare delivery. In the records management context, the records life cycle looks at the creation, active and disposal of records and these three life cycle stages, require different management strategies.

No known study had specifically been done to determine the statistical distribution, estimate the dormancy time and determine factors that could influence dormancy time of medical records. To fill the gap this study was conducted to address these issues and the information from the result will be useful for the development of a medical records management retention, disposal and archiving policies in UCH, Ibadan.

The results of the study was analysed, interpreted and discussed in light of the available evidence, the limitation and strength of the study was briefly discuss. The main conclusion, some recommendation and implications are surmised in the next sections.

.

6.2 Conclusion

The study had indicated that time-to-dormancy of medical records of patients seenat UCH, Ibadan, was of the time-to-event type, and can best be analysed with survival techniques. The performance of the distribution that best fit the dormancy time data was tested variously and found to be the Weibulldistribution. This was evident by the:

- hazard plots of the dormancy time exhibitinga bathtub shape;
- the plot of *log-log (S(t)) on log(t)* test of linearity which exhibited a straight line;
- the estimated shape parameter of the fitted line from the plot that is not unity;and
- *-2logL* and *AIC* value test that indicated Weibull model adjudged as the best predictor

The Kaplan-Meier survival curve tested for the five cohorts and then merged as a single sample indicated that bothsurvival and the hazard curvefor all five cohorts had similar characteristics, forms andshapes, suggesting that the distribution and its parameters are same for dormancy time of medical records of patients seen at different periods at UCH, Ibadan.

A considerably large number, close to 25%, of medical records of patients seen at UCH, Ibadan had a single entry, suggesting that the records were only used once andthus became dormant (inactive) immediatelyafter the first contact. Estimates however show that among those returning for a second visit, 95% had doneso at 5.95 months after the first visit. The implication of this is that patients that failed to return for a 2ndcontact after 5.95 months maynever returned again and can have their records declared dormant and safely weeded.

However, for those records that survived beyond one day after creation,(indicated by two or more entries),the estimatedP_{95}showthat 95% of the medical records have an active life of151.89 months. This findings has the implication that95% of the medical records became dormant in 151.89 months (12.66 years), that is a 0.95 probability of a record becoming dormant in 151.89 months, and such a record can be safely weeded and disposed of to create space for new records with the confidence that at most less than5% are likely to return. Ninety-five percent of the penultimate appointments was in 17.92 months.

The hazard proportionality assumption test was found to be significant in some of the cohorts and with the data combined into a single sample,therefore the dormancy time data was adjudged invalid for the proportional hazard assumption which was corroborated by the adjusted gender stratification plot of –ln-ln(survival time) against dormancy time, t,withthe lines intersecting each other. Though Cox Proportional Hazard, Exponential and the Weibull models all identified patients age, gender, state of residence, clinics attended, admission status, surgery status and treatment outcome as predictors of record dormancy time at UCH, Ibadan, as they were all significant at $P<0.01$,however with a hazard ratio of above 1 for admission status and treatment outcome these two factors are better predictors of dormancy time. However the Weibull distribution was observed as the best model for predicting association between patient characteristics and dormancy time of records of patient seen at UCH, Ibadan.

6.3 Recommendation/ policy implications

The findings in this study have important policy implications on management of patients'medical records. Records are vital tools in medical practiceand good quality medical record is essential and beneficial to the patient, clinicians and the hospital. Healthcare is information driven hence the effectiveness and efficiency of patient care are dependent on the availability of patient information held in the medical record. It then follows that medical records contain information that support clinical decisions and actions and badly managed medical records will adversely affect patient care. It is important that hospital authorities pay attention to the development of proper management of patientmedical records,supported with good policyguidelines both at the

institutional and national levels, especially when the institution is a tertiary health institution of the type of the University College Hospital, Ibadan.

Studies had shown that not all records created deserve to be kept permanently or for longer period than required since significant costs associated with creation, maintenance and storage of records can be economical with defined records management policy, retaining what is needed and doing away with what is not.

Findings of this study will provide a frame work for the development of a management policy guidelines on retention, disposal and archiving of medical recordsin UCH, Ibadan. The frame work will also serve as a guide (benchmark) towards the development of a national policy for the retention, disposal and archiving of patient records in Nigeria.

The study in addition revealed the urgent need for improved documentation of both the socio-demographic and clinical patient information. There should be a policy guideline on clinical documentation to ensure standard and uniformity in patient documentation, and to improve the quality of patient information. Education and economic status are important socio-demography variables and should be documented for every patient. Attention should be given to clear indication of diagnosis that enable efficient and good quality data.

The management of the UCH, Ibadan need toappreciate the importance of patient information and the management beyond the used for individual patient care, healthcare data serve a useful decision making tool for healthcare management at community, population and national levels.

6.4 Proposed guidelines formedical records policy development

From the findings of this study the following points are highlighted to guide the Management of the University College Hospital, Ibadan in developing"***a medical records management policy"***, that will incorporate particularly schedules for the retention, disposal and archiving of medical records.

i. Medical records of patients that failed to return for a 2nd contact after 6 months should have their records declared dormant and safely weeded from the active files in the respective clinics;

ii. Medical records that attained the age of 13 years should be weeded from the active files and moved to a secondary storage to create space for new once;

iii. Medical records of dead patients, discharged against medical advice and referred cases should beweeded for disposalimmediately;

iv. Clinical documentation guidelinesfor doctors should be developed to improve the quality of patients' information;

v. Economic and educational status of patients were not captured in the medical records, hence these factors wereexpunged from the study. This made it impossible to determine theirrisk factorsontime-to-dormancy of patient's medical record.

6.5 Contribution to knowledge

i. Until now, a serious deficiency in the records life cycle model is its failure to quantify in terms of survivorship,the time from the point of creation to death, (time-to-dormancy) of a records, a limitation in records management. This study providesa technique, for estimating the time-to-dormancy of records, thereby filling the gap created in the records life cycle model;

ii. Application of Survival Analysis techniques had been popular in clinical trials and cancer studies until now; findings of the study now advances the use of survival analysis methods to records management. Using Survival Analysis techniques to study the longevity and obsolescence of patient's medical records for the first time in Nigeria.

iii. No known empirical studies had been done to determine the time–to-dormancy (longevity and obsolescence) or factors associated with patient medical record dormancy, the study provides estimate for time-to-dormancy for medical records, its distribution and the associated factors. The findings established a base for developing benchmarks for policy for the retention, disposal and archiving of patients medical records;

iv. The absence of a documented policy for the retention, disposal and archiving inNigeria health institutions had created a gap in management of patients' medical

records. The findings established a base for developing benchmarks for policy development on retention, disposal and archiving of patients medical records.

References

Abankwah, R., 2008. *Management of Audiovisual Materials in the member states of the Eastern and Southern Africa Regional Branch of the International Council on Archives (ESARBICA).* Pietermaritzburg, University of KwaZulu-Natal. Phd thesis.

Abdulrahaman, A. B. 2015. Management of university records for effective administration of universities in North Central Nigeria. International journal of Library and Information Science *academicjournals.org/journal/IJLIS/article-full-text-pdf/A4DD33451096*

Acna Professional Counsel 2007. *Guide for Lawyers and Law Firms, Creating A Record Retention & Destruction Policy*. Retrieved December 12, 2016 from https://www.gilsbarpro.com/gilsbarpro/media/GilsbarPro/LPL%20Docs%20 (Generic)/NewsUpdates/Creating-A-Record-Retention.pdf

Aduge-Ani, D. 2013. Nigeria: What is happening to Medical Records at Wuse General Hospital? *Leadership Weekend online Publication*. Retrieved 7/23/2017 from https://www.google.com.ng/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=what%20is%20happening%20to%20medical%20records%20at%20wuse%20general%20hospital

Aduku, B. S. and Abdul, O. A. 2012. Management of Records of the Judicial Service Committee of the Federal Capital Territory Abuja, *The Information Manager*: Vol. ***12***, No ***1-2*** (2012)

Agbaje, A. A. 1991. *Medical Records in University College Hospital, Ibadan.Immanuel College of Theology, Ibadan, Nigeria: A thesis*. Retrieved 8/12/2017 from dv.sagepub.com/content/7/1/19.short

Agere, S. L. V. and Mazikana. P. 1999. *Better Information Practices: Improving Records and Information Management in the Public Sector, Managing the Public Service Strategies for Improvement Series*: No. 6 London: Commonwealth Secretariat.

Akor, P. U. and Udensi, J. 2013. An Assessment of Record Management System in Establishment Division of Two Universities In Nigeria. *Mediterranean Journal of Social Sciences* 4. 12. Retrieved October 2 2017 from 10.5901/mjss.2013.v4n12p87

Akussah, H., 1996. Records management: an overview. *African Journal of Library, Archives and Information Science* 6. 2: 101-106.

Akram, M., AmanUllah, M., and Taj, R. 2007. Survival Analysis of Cancer Patients Using Parametric and Non-Parametric Approaches. *Pakistan Vet* 27. 4: 194-198.

Aljumah, A.A.,Ahamad, M.G. and Siddiqui, M.K. 2013. Application of data mining: Diabetes health care in young and old patients. *Journal of King Saud University - Computer and Information Sciences* 25. 2: 127–136

Allison, P. D., 1995. *Survival Analysis Using the SAS System. A Practical Guide*. SAS Institute, Cary, NC.

Altin, A. 2013. Estimation of the Shape Parameter of the Weibull Distribution Using Linear Regression Methods: Non-Censored Samples. *Quality and Reliability Engineering* 29. 8. Retrieved 7/17/2017 from http://www. 10.1002/qre.1472

Altman, D. G.; 1992. Analysis of Survival times In :*Practical statistics for Medical research*. London (UK): Chapman and Hall

American Health Information Management Association. 2008. *Medical Records and the Law*, 4th Edition. Jones and Bartlett Publishers, Sudbury, Massachusetts.

American Institute of Certified Public Accountants (undated) Records Management: Integrating Privacy using generally Accepted Privacy Principles

American Psychological Association (*APA*), *2007. Revised Record Keeping Guidelines. Incomplete*

Andersen, P.K. 1992. Repeated assessment of risk factors in survival analysis. *Statistical Methods in Medical Research* 1: 297-315.

Anderson, A. and Semmelroth, D. 2013. Quantile-Quantile (Q-Q) Plots: Graphical Technique For Statistical Data. Retrieved 7/17/2017 from //www.dummies.com/programming/big-data/data-science/quantile-quantile-qq-plots-graphical-technique-for-statistical-data/

Arruda, M. E., Prinzing M. and Rana Shruti, 2013. Documents, What Documents?" Business Law Today January/February 2003 issue of at page 23

ARMA International. 2016. *Principle of Retention*. Retrieved December 12, 2016 from ww.arma.org/r2/generally-accepted-br-recordkeeping-principles/retention

Ata, N. and Tekin, S. M. 2007, Cox Regression Models with Nonproportional Hazards Applied to Lung Cancer Survival Data. Hacettepe Journal of Mathematics and Statistics Volume 36 (2) (2007), 157 – 167

Atherton, J., 1985. "From Lifecycle to Continuum: Some Thoughts on the Records Management–Archives Relationship." Archivaria 1(21): 43-51.

Ayoade, B. A., Thanni, L. O. and Shonoiki-Oladipupo, O. 2012. *Mortality Pattern in Surgical Wards of a University Teaching Hospital in Southwest Nigeria: A Review*. Retrieved December 17, 2018 from https://www.researchgate.net/publication/233892401_Mortality_Pattern_in_Surgical_Wards_of_a_University_Teaching_Hospital_in_Southwest_Nigeria_A_Review

Axel, G. 2005. *Directed Model Checks for Regression Models from Survival Analysis, A PhD thesis*. Retrieved Month Day, Year from https://wwwf.imperial.ac.uk/~agandy/papers/06/diss.pdf

Bantin, P. C. 1998. Strategies for Managing Electronic Eecords: A New Archival Paradigm? An Affirmation of our Archival Traditions? *Archival Issues* 23. 1: 17-34.

Beuscart, P., Boulanger, L., Salleron, F. and Duhamel, 2012. Overestimation of the probability of death on peritoneal dialysis by the Kaplan-Meier method: advantages of a competing risks approach. *BMC Nephrology**13***. 31. Retrieved Month Day, Year from https://doi.org/10.1186/1471-2369-13-31

Billinton, R. and Allan, R.N. 1983. *Reliability Evaluation of Engineering Systems: Concepts and Techniques*. Pitman Books Limited, Boston.

Bellavia, A. 2015. *A Percentile approach to time-to-event outcomes*. Thesis for Doctoral Degree (Ph.D.)

Bester Sandra, 2013. Protection of Personal Information Act 3 of 2013, Medical Protection Society, South Africa. *http://www.justice.gov.za/inforeg/docs/InfoRegSA-POPIA-act2013-004.pdf*

_____2014. *Retention of medical records, Medical Protection Society, South Africa*. Retrieved Month Day, Year from www.medicalprotection.org/southafrica

Borglund, E. and Öberg, L. M. 2006. *Operational Use of Records. Paradigms Politics.* 8/23/2017 https://www.academia.edu/26870376/Operational_use_of_records

Bou-Hamad, I., Larocque, D. and Ben-Ameur, H. 2011. A review of survival trees. *Statistics Surveys* 5. 44–71.

Bradburn, M. J., Clark, T. G, Love, S. B., and Altman, D. G. 2003. Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. *British Journal of Cancer* 89. 605 – 611.

Bruno, R. M. 1994. Statistical Analysis of Survival Data. *UNF Theses and Dissertations. Paper 150*. Retrieved Month Day, Year from http://digitalcommons.unf.edu/etd/150

Buis, M.L., 2006. An introduction to Survival Analysis, Department of Social research Methodology, VrijeUniversiteit Amsterdam

Chachage, B and Ngulube, P, 2008. *Management of business records in Tanzania: an exploratory case study of selected companies*. Published by InterWord Communications for Department of Information and Knowledge Management, University of Johannesburg

Charman, D. 1984. *Records surveys and schedules: a RAMP study with guidelines*. Paris: UNESCO.

Chattopadhyay, G. and Kumar, S. 2008. Parameter estimation for rail degradation model. *International Journal of Performability Engineering*.

Chen Zhu, B. E. 2012. *Equipment data analysis study - Failure time data modeling and analysis*. The Degree of Master of Science in Engineering thesis, University of Texas at Austin.

Chibambo, M.L.N., 2003. Records Management: The Key to Good Governance and Sustainable Development. *XVII BIENNIAL Eastern and Southern Africa Regional of Branch of the International Council on Archives (ESARBICA) General Conference on Archives, Society and Good Governance. Mozambique, Maputo, JULY 22-26, 2003.*

Chin-Diew, L. 2006. *Weibull Distributions and their applications*. Retrieved Month Day, Year from https://www.researchgate.net/publication/37628953

Clark, T. G., Bradburn, M. J., Love, S. B. and Altman, D. G. 2003. Survival analysis. Part I: basic concepts and first analyses. *Br J Cancer* 89. 232–238.

Cochran, W. G., 1963. *Sampling Techniques*. 2nd Ed., New York: John Wiley and Sons, Inc.

Coetzer, X. P., 2012. *The status of Records Management at the University of Zululand Ulundi, University of Zululand*. Master's thesis.

Collett, D. 1994. *Modelling Survival Data in Medical Research*. Chapman & Hall, London, UK).

_____ 2003.*Modeling survival data in medical research* (2nd Ed.). London: Chapman and Hall, CRC.

Columba University Mailman School of Public Health, (n.d.).*Time-To-Event Data Analysis*. Retrieved Month Day, Year fromhttps://www.mailman.columbia.edu/research/population-health-methods/time-event-data-analysis

Cook, T. 1984. From Information to Knowledge: An Intellectual Paradigm for Archives. *Archivaria* 1. 19: 28-49.

Cox, D. R. and Snell, E. J., 1968. A general definition of residuals. *Journal of the Royal Statistical Society, Series B*, 30. 248-275.

_____ 1968. A general definition of residuals (with discussion). *Journal of the Royal Statistical Society*, Series B 30:

Cox D. R. 1972. Regression models and life-tables. *J R Stat Soc Series B StatMethodol* 34. 2:187–220.

_____1992. Regression models and life-tables. *Journal of the Royal Statistical Society Series* B 34. 187: 220

Cox, D.R and Oakes, D. 1984. *Analysis of Survival Data*. Chapman and Hall, London

Crowley, J. and Hu, M. 1977. Covariate analysis of heart transplant survival data. *Journal of American Statistical Association* 78. 27-36.

Crumer, A.M. 2011. Comparison Between Weibull And Cox Proportional Hazards Models

Cunningham, P. and Wiedemann, L. A. 2011. **AHIMA's Recommended Retention Standards** [appendix D from the 2011 update

CMartinez, C. and Maria, R. L. 2007. Diagnostics for Choosing between Log-Rank andWilcoxon Tests. *Dissertations.* Paper 895. Retrieved Month Day, Year from http://scholarworks.wmich.edu/dissertations

Dana, C. and McWay, 2010. *Legal and Ethical Aspects of Health Information Management*, 3rd edition.

Donald S. Skupsky, D.S.*Law and Records Management Legal Requirements for Records Retention*. Retrieved Month Day, Year from Hppts://.rch.com/legal-requirements-for-records-retention-the-three-year-presumption/

Danso, J. 2015. A Study of Records Management Practice at Health Facilities in Upper Denkyira West District of Ghana. *Advances in Life Science and Technology* 31. Retrieved Month Day, Year fromwww.iiste.org

Dearstyne, B.W. 1985. *The Management of Local Government Records: A Guide for Local Officials*. New York: American Association for State and Local History.

Department of Health, 2006. *Records Management: NHS Code of Practice Part 1* www.dh.gov.uk/publications

Department of Health South Africa, 2012. *Medical Records Management Policy*. Limpopo provisional Government, Republic of South Africa.

Department of Health Social Services and Public Safety (DHSSPS), 2004. *Good Management, Good Records*, retrieved www.dhsspsni.uk 15/09/2014.

Department of Health, Social Services and Public Safety, 2004.An introduction to Good Management Good Records. https://www.health-ni.gov.uk/articles/introduction-good-management-good-records

DezanShira and Associates. 2016.*Accounting Records in China: Keeping up to Date with the New Regulations*. China Market Watch: Relaxed Foreign Investment Control and Liberalization of Stock Market

Dollar, C. 2002. *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods*. Macerata: University of Macerata Press.

Digital, N. Z. 2014. "Make it Digital." Retrieved February 16, 2017 from http://www.digitalnz.org/about.

Kathy, D and Marcy, P. 2013. *Retention and Destruction of Health Information*. AHIMA.

Ebadifar F, Hajavi A, Maidani Z. 2004. A Comparative Study of Medical Record Standards in Selected Countries. Management and Medical Information Journal. 2004; 7(17): 37-41.

Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* 557–565.

Epstein, B. and Sobel, M. 1953. Life Testing. Journal of the American Statistical Association, 48, 486-502. http://dx.doi.org/10.1080/01621459.1953.10483488

European Document Retention guide, 2014. *A quick guide to records management and retention*. Retrieved Month Day, Year from http://www.legalsupportnetwork.co.uk/sites/default/files/IM%20EU%20Full%20Retention%20Guide_UK_Oct14_ONLINE_web.pdf

FEA, 2005. *Federal Enterprise Architecture Records Management Profile*, Version 1.0. December 15, 2005.

Federal Government of Nigeria. 1958. *Federal Government Financial Regulation, Control and Management Act (1958)*.9/23/2017. https://www.lawyard.ng/wp-content/uploads/2016/01/FINANCE-CONTROL-AND-MANAGEMENT-ACT.pdf

_____ 2000. *Revised Financial Regulations, Control and Management Act (1958)*International Journal of Finance and Accounting. P-ISSN: 2168-4812 e-ISSN: 2168-4820
2013; 2(8): 446-451. doi:10.5923/j.ijfa.20130208.07

Federal Taxation Committee. 2004. *The Record Retention Guide. The Massachusetts Society of Certified Public Accountants*, Inc.

Feinstein, A.R. 1985. The architecture of clinical research. *Clinical Epidemilogy* 226-227

Ford, C 2015. *Understanding Q-Q Plots. Research services and sciences*, University of Virginia library

Fox, J. 2006. "Introduction to Survival Analysis. (2006) Lecture Notes for 'Statistical Applications in Social Research', <http://socserv.mcmaster.ca/jfox/Courses/soc761/survival-analysis.pdf>.

_____ 2014. *Introduction to Survival Analysis. Survival Analysis Using the SAS System*, Second Editionsocserv.mcmaster.ca/jfox/Courses/soc761/survival-analysis.pdf

Fraser, N. E. 2010. Medical Records Management. Retrieved Month Day, Yearfrom http://EzineArticles.com/expert/Nancy_E_Fraser/615155

Galani, O. and Nikiforou, A. 2006. Electronic Health Records. *Handbook of Research onInformatics in Healthcare and Biomedicine*. A. L. Athina. Hershey, Idea Group Reference.

Griffith University 2018. *Records Management Policy*. Retrieved from http://policies.griffith.edu.au/pdf/Records Management Policy.pdf

Garaba, F. 2011. *An Investigation into the Management of the Records and Archives of Former Liberation Movements in East and Southern Africa held by National and Private Archival Institutions*. Pietermaritzburg, University of KwaZulu-Natal Phd thesis.

Garrib, A., Stoops N., Mckenzie A., Diamini L., Govender T., Rohde D. and Herbst K.., 2008. An Evaluation of the District Health Information System in Rural South Africa. *South African Medical Journal* 98.7: 549-552

Goel,Manish Kumar, Khanna,Pardeep, and Kishore,Jugal. Understanding survival analysis: Kaplan-Meier estimate

Goodman, E.C. 1994. Records management as an information management discipline: a case study from SmithKline Beecham Pharmaceuticals. *International Journal of Information Management* 14. 2:134-143

Government of Canada, 2011. *Retention Guidelines for Common Administrative Records of the Government of Canada Section 1, Part 1 General Administration*

Function. https://www.collectionscanada.gc.ca/obj/007002/f2/007002-3000.1-e.

Hallinan Jr., A. J., 1993. A review of the Weibull distribution. *J. Qual. Technology*. 25. 85–93

Hardcastle, S. 1989. Providing storage facilities. In: Peter, E. (ed.) *How to manage your records: a guide to effective practice*. Cambridge: ICSA Publishing.

Hare, C.E. and McLeod, J, 1997. *Developing a records management programme*. London: Aslib.

Hardcastle, S., 1989. Providing storage facilities. In: Peter, E. (ed.) *How to manage yourrecords: a guide to effective practice.* Cambridge: ICSA Publishing.

Harrell, F. E. 2001. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. British Journal of Cancer 89. 781–786. Retrieved Month Day, Year from http://www. doi:10.1038/sj.bjc.6601117

Health Professionals Council of South Africa, 2008. *HPCSA guidelines for good practice in the health care professions: Guidelines on the keeping of patient records.* Booklet 14. Pretoria. Retrieved May 15, 2012 from: http://www.hpcsa.co.za/downloads/conduct_ ethics/rules/generic_ethical_rules/booklet_14_keeping_of_patience_records

Health Service Executive, 2013. *Record Retention Periods*. Retrieved from https://www.hse.ie/eng/services/yourhealthservice/info/dp/recordretpolicy.pdf

Henry, L. J. 1998. Schellenberg in Cyberspace. *The American Archivist* 61. 2: 309-327.

Hess, D. R. 2004. The ABCs of Research, 49th International Respiratory Congress, held December 8–11, 2003, inLas Vegas, Nevada. *Respiratory Care Journal* 49. 10: 1171–1174.

Hoke, G.E.J. 2011. Records Life Cycle: A Cradle-to-Grave Metaphor, RIM Fundamentals, Information management. *ARMA International* . Retrieved Month Day, Year from www.arma.org

Hosmer D.W. and Lemeshow, S. 1980. *Applied Logistic Regression*. Wiley, New York.

_____ 1999. *Applied Survival Analysis; Regression Modeling of Time to Event Data*. New York: John Wiley & Sons.

Hosseinia, R. and Takemurab, A. 2014. *An objective look at obtaining the plotting positions for QQ-plots*. Retrieved Month Day, Year from https://arxiv.org/pdf/1409.6885.

Howell Jr. R. Thomas and Cogar, R. N. 2003. Records Retention – An Essential Part of Corporate Compliance retrieved 8/23/2017 https://apps.americanbar.org/buslaw/newsletter/0019/materials/recordretention.pdf

_____ 2003. Record Retention and Destruction Current Best Practices. Retrieved 3/21/2017 from https://apps.americanbar.org/buslaw/newsletter/0019/materials/recordretention.pdf

Hribar, L. and Duka, D. 2010. *Weibull Distribution in Modeling Component Faults 82210*.ELMAR_2010_symposium international conference. Retrieved Month Day, Year from https://ieeexplore.ieee.org/document/5606115

Huffman, E. K. 2014. *Health information management, 10th Edition, Physicians' Record* Company, Berwyn, Illinois.

Hurvich, C.M. & Tsai, C.-L., 1989. Regression and Time Series Model Selection in Small Samples. *Biometrika* 76. 297-307.

Ifedili, C. J. and Agbaire, J. J. 2011. Management of Educational Records in Nigerian Universities for Better Results. Review of European Studies. Retrieved from http://dx.doi.org/10.5539/res.v3n2p52

Igbokwe-Ibeto, C. J. 2013. Record Management in The Nigerian Public Sector and Freedom Of Information Act: The Horn of Dilemma. *International Journal of Development and Management Review (INJODEMAR)* 8. 1, June 2013.

International Records Management Trust. 1999. *Principles of records and archives management*. London: International Records Management Trust.

_____ 1999.*Managing Public Sector Records: A Study Programme*. London WC1N 2EB

_____, 2000. *Managing Records as the Basis for Effective Service Delivery and Public Accountability in Development: An Introduction to Core Principles for Staff of the World Bank and Its Partners*

_____ 2002. Evidence-Based Governance in the Electronic Age. Case Study Financial Records Systems in Nigeria. A World Bank/International Records Management Trust Partnership Project

Iron Mountain, 2005. Records Management Best Practices Guide. A Practical Approach to Building a Comprehensive and Compliant Records Management Program. https://www.agnesscott.edu/facultyservices/files/documents/bestpracticesguide.pdf

Iron Mountain, 2005. *Records Management Best Practices*. Retrieved 23/7/2017 http.www.Guideww.ironmountain.com

Iron Mountain, 2005. *Records Management Best Practices*. Retrieved 23/7/2017 http.www.Guideww.ironmountain.com

Ishwaran, H., Kogalur, Udaya, B. B., Eugene, H. and Lauer, M. S. 2008.. Random survival forests. *Annals of Applied Statistics* 2, 841–860.

ISO 15489, 2001. *Definitions of records management*, 4/6/2017https://irp-cdn.multiscreensite.com/b4ebdc6d/files/uploaded/ISO_15489

Iwhiwhu, E. B. 2005. Management of records in Nigerian universities: Problems and prospects. *The Electronic Library* 23. 3:345-355. 10/5/2016 https://doi.org/10.1108/02640470510603741

Jamoom Eric W, Partel Vaishali; FurukawaMichael F., and King Jennifer, 2014. Clinical Benefits of Electronic Health Record Use: National Findings. 10.1111/1475-6773.12135

Jenkins, S. P. 1997. Discrete time proportional hazard regression. *Stata Technical Bulletin* 39. 17-32.

Johnson N. L., Kotz, S. and, Balakrishnan, N. 1994. *Continuous Univariate Distributions* 1.2nd edn. Wiley: New York.

Kalbfleisch, J.D and Prentice, R.L 1980. *The statistical analysis of failure time data*. John Wiley and Sons, London.

Kalbfleisch, J. D. and Prentice, R. L. 2011. *The Statistical Analysis of Failure Time Data, Second Edition*. Wiley Online Library

Johnson N. L., Kotz, S. and, Balakrishnan, N. 1994. *Continuous Univariate Distributions* 1. 2nd edn. Wiley: New York

Kalbfleisch, J.D and Prentice, R.L 1980. *The statistical analysis of failure time data*. John Wiley and Sons, London.

Kaplan, E. L. Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53. 457–81.

Katuu, S. A., 2015. *Managing Records In South African Public Health Care Institutions – A Critical Analysis*, University of South Africa, Pretoria, A PhD Thesis. http://hdl.handle.net/10500/19058

Kanzi, N. 2010. *An Investigation of the Role of Records Management with Specific Reference to Amathole District Municipality*. Port Elizabeth, Nelson Mandela Metropolitan University. A Master thesis.

Kemoni, H, Ngulube, P and Stilwell, C. 2007. Public records and archives as tools for good governance: reflections within the recordkeeping scholarly and practitioner communities. *ESARBICA Journal* 26. 3-23.

Kenosi, L. 2008. *Records, the Truth Commission and National Reconciliation: Accountability in Post-Apartheid South Africa*. Pittsburgh, PA, University of Pittsburgh. Phd thesis.

Kim, H. and Lee, M. 2015. *Factors associated with health services utilisation between 2010 and 2012 I Korea: using Andersen's Behavioral model. Retrieved 8/12/2017* http://creativecommons.org/licenses/by-nc-nd/4.0/

Kleinbaum, D.G. 2005. *Survival analysis. A self-learning text*. USA: Springer;

Kleinbaum, D. G. and Klein, M. 2012. *Survival Analysis, A Self-Learning Text*. Springer New York Dordrecht Heidelberg London Third Edition

Kooperberg, C., Stone, C. J. and Truong, Y. K. 1995. Hazard regression. *Journal of the American Statistical Association* 90. 78–94.

Krit, M. 2014. Goodness-of-fit tests in reliability: Weibull distribution and imperfect maintenance models. General Mathematics [math.GM]. *Université de Grenoble.*

Kumar, S. 2006. *A Study of the Rail Degradation Process to Predict Rail Breaks. Licentiate Thesis. Division of Operation and Maintenance Engineering*. Luleĺ University of Technology, Luleĺ, Sweden.

Kumar, S., Espling, U. and Kumar, U. 2007. A holistic procedure for rail maintenance in Sweden. Accepted for publication in Journal of Rail and Rapid Transit: *Proceedings of the Institution of Mechanical Engineers*, Part F.

Kumar, S., Gupta, S. and Ghodrati, B. 2007. *Rail defect prioritization and risk assessment using a hybrid approach.* Retrieved 8/23/2017 *ltu.diva-portal.org/smash/get/diva2:989929/FULLTEXT01.pdf*

Lawless J. F. 1998. *Parametric models in survival analysis. In Encyclopaedia of Biostatistics*, Armitage P, Colton T(eds). Wiley: New York, 1998; 3254–3264.

_____ 2003. *Statistical Models and Methods for Lifetime Data 2nd Edition*. A John Wiley & Sons, INC., Publication.

Lee E. 1992. *Statistical Method for survival data analysis*. New York: Wiley; pp. 8–17.

Lee E. T and Go O. T. 1997. Survival analysis in public health research. *AnnuRevPublic Health.*18. 105–34. doi: 10.1146/annurev.publhealth.18.1.105. [PubMed: 9143714].

Lee, E. T. and Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis Third Edition*. A John Wiley & Sons, Inc., Publication.

Lemeshow, S., HosmerJr, D. W., Klar, J. and Lwanga, S. K., 1990. *Adequacy of Sample Size in Health Studies*. John Wiley and Sons. Published on behalf of World Health Organization.

Levant, S. and DeFrances, C. J. 2013. *Trends in Inpatient Hospital Deaths: National Hospital Discharge Survey, 2000–2010 Margaret Jean Hall*. NCHS Data Brief. No. 118. March 2013. U.S. Department of Health And Human Services Centers for Disease Control and Prevention National Center for Health Statistics

Limpert E. W., Stahel, A. and Abbt, M. 2001. Log-normal distributions across the sciences:
Keys and clues. *BioScience.*51**.** 5:341–52. doi: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

Lin, D. Y. and Wei, L. J. 1991. *Goodness-of-Fit Tests for the General Cox Regression Model* .Institute of Statistical Science, Academia Sinica, Vol. 1, No. 1

(January 1991), pp. 1-17. Retrieved 7/10/2017 from https://www.jstor.org/stable/24303990

Madu, E. C. 2004. *Technology for Information Management and Service*. Ibadan: Coleman. p 201 -210.International Standards Organisation.

Mampe, G., and Kalusopa., 2012, Records management and service delivery: the case of Department of Corporate Services in the Ministry of Health in Botswana

Mccracken, B. and Hyun, S. 2010. *On The Comparison of Two Survival Functions*. U S C U p s t a t e U n d e r g r a d u a t e R e s e a r c h J o u r n a l, V o l. III, F a l l 2 0 1 0

McKemmish, S. 1997. Yesterday, today and tomorrow: a continuum of responsibility. *Proceedings of the Records Management Association of Australia 14th National Convention, 15–17 Sept 1997*. Perth, Australia: RMAA.

McKemmish, S., Piggott, M., Reed, B. and Upward, F,. 2005. *The Records Continuum. Archives: Recordkeeping in Society*. WaggaWagga, NSW: Centre for Information Studies.

McWay, D. C. 2002. Legal Aspects of Health Information Management. *Delmar Learning* 78. 75-76.

_____ 2008. *Today's Health Information Management: An Integrated Approach*. CliftonPark, NY, Thomson Delmar Learning.

_____ 2013. *Today's Health Information Management: An Integrated Approach*. CliftonPark, NY, Delmar Cengage Learning.

Medical Council of New South Wales. 2010. *Medical Records – Regulation*. Retrieved 4/23/2017 from http://www.mcnsw.org.au/page/65/resources/legislation/medical-records-act---regulations/

Medical and Dental Defence Union of Scotland. 2013. *Essential guide to medical and dental records.*Retrieved from: https://www.mddus.com/training-and-cpd/training-for-members/dental-risk-toolbox/record-keeping

Medical Protection Society, SA, 2012. Health care in South Africa. https://www.brandsouthafrica.com › SA Fast Facts › Health

Medical Protection Society. 2012. *South African Medical Association T 0800 225 677. R*etrieved 4/23/2017. Retieved fromwww.mps-group.org/za-mla

Medical Protection Society. 2013. Retention of medical records. *SA Orthopaedic Journal Autumn* 12.1 *R*etrieved 4/23/2017 fromfrom.http://www.scielo.org.za/pdf/saoj/v12n1/14.

Medical Records. 2015. *Professional support and expert advice*. Medical Protection Society, England

Marubini, E., Valsecchi, M.G. andd Chichester. 1995. *Estimation of Survival Probabilities. Analysing survival data from clinical trials and observational studies*. John Wiley and Sons; Chichester (UK) pp. 41–8.

Mayosi, B. M. and Lawn, J. E. 2012. Health in South Africa: Changes and Challenges since 2009. *The Lancet* 380. (9858): 2029-2043.

Moore, D.S. and Spruill, M.C., 1975. Unified large sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics* 3. 3.

Moreau T., O., Quigley, J. and Lellouch, J. 1986. On D. Schoenfeld's approach for testing the proportional hazards assumption. *Biometrica* 73. 2.

Mphatswe, W., Mate K. S., Bennett. B., Ngidi H. W., and Reddy J., Barker P. M., Rollins N., 2012. Improving Public Health Information: A Data Quality Intervention in KwaZulu-Natal, South Africa. *Bulletin of the World Health Organization* 90. 3: 176-182.

Miller, J. E. 2008. *Writing about Hazards Models: Practical Guidelines for Effective Presentation*. Institute for Health, Health Care Policy and Aging Research Rutgers University

Miller, R.G. 1981. *Survival Analysis*. John Wiley and Sons, New York

Millar, L. 2001. *Additional Resources for Electronic Records Management*. ISO 15489-1, International Organization for Standardization

Ministry of Health, NSW North Sydney Australia. 2012. *Health Care Records – Documentation and Management.R*etrieved 4/23/2017 from http://www.health.nsw.gov.au/policies/15/09/2014

Montaseri M., Yazdani, C. and Fateme, E. 2016. *Application of Parametric Models to a Survival Analysis of Hemodialysis Patients*. Nephrourol Mon. 2016 Sep 13;8(6):e28738. eCollection 2016 Nov.

Moreau, T., O'Quigley, J., and Mesbah, M., 1985. A global goodness-of-fit statistic for the proportional hazards model. *Applied Statistics* 34. 212-218.

Mrwebi, S. E. 2000. *Records Management in a Management Consulting Firm. Johannesburg, University of Johannesburg*. Masters thesis.

Mukangai, J. I. and Odongo, L. O. 2016. Survival Analysis: An Overestimation of Kaplan-Meier Method in the Presence of Ties. *American Journal of Theoretical and Applied StatisticRetrieved 4/23/2017 from*http://article.sciencepublishinggroup.com/html/10.11648.j.ajtas.20160505.17.html

Murthy D. N. P., Xie M. and Jiang R. 2003. *Weibull Models*. Wiley: New York 2003

Nardi, A. and Schemper, M. 2003. Comparing Cox and parametric models in clinical studies. *Statist. Med*. DOI: 10.1002/sim.1592)

National Archives and Records Administration [United States], Office of Management and Budget [United States], et al., 2005. *Federal Enterprise Architecture Records ManagementProfile*. Retrieved January 17, 2014, from http://www.archives.gov/records-mgmt/pdf/rmprofile.pdf.

. National Health Act, 2014. Explanatory Memorandum. *Retrieved 4/23/2017 from*http://www.nassnig.org/document/download/7990

NationalHospitalsOffice. 2007.*NHOCodeofPracticeforHealthcareRecordsManagement.Version 2.0 (illustratedRetrieved 4/23/2017from* https://www.hse.ie/eng/services/publications/hospitals/nho-code-of-practice-for-healthcare-records-management-version-2-0.pdf

National Hospital Services, Portsmouth Hospitals, 2011.Clinical Records management Policy. Retrieved 23rd May 2013 from http://porthosp.nhs.uk/clinicalrecords

Noordzij M., Leffondré, K., Stralen, K.J., Zoccali, C., Dekker, F.W. and Jager, K. J. 2013. When do we need competing risks methods for survival analysis in nephrology? *Nephrology Dialysis Transplantation*, 28. 1: 2670–2677 *Retrieved 4/23/2017 from*https://doi.org/10.1093/ndt/gft355

Nwogu, B. G. 2005. *Educational research basic issues and methodology*. University. University trust publishers.

Naidoo, S. 2013. Patient's Access to Records. *South African Dental Journal* 68. 1: 36-37.

NECCC , 2004. *Challenges in Managing Records in the 21st Century. National Electronic Commerce Coordinating Council*. *Retrieved 4/23/2017 from* https://library.osu.edu/assets/Uploads/RecordsManagement/Challenges-in-21st-e-recs-neccc.pdf

Nell, J. L. 2006. *Aspects of Confidentiality in Medical Llaw*. Pretoria, University of Pretoria. Masters thesis.

Newman University Birmingham. 2005. Policy and Process for record management: a discussion and consultation Paper.*Retrieved 4/23/2017 from*https://www.newman.ac.uk

Ngoepe, M. S. 2008. *An Exploration of Records Management Trends In The South African Public Sector: A Case Study of The Department of Provincial And Local Government*. a master's degree thesis

Ngoepe, M. 2013. *Fostering a Framework to embed the Records Management Function into the Auditing Process in the South African Public Sector. Pretoria, University of South Africa*. Phd thesis.

Ngulube, P. 2000. Professionalism and ethics in records management in the public sector in Zimbabwe. *Records Management Journal* 10. 3:161-173.

_____ 2003. *Preservation and Access to Public Records and Archives in South Africa. PhD Thesis. Pietermaritzburg: University of Natal*. *Retrieved 4/23/2017 from* http://www.hs.unp.ac.za/infs/thesispn.pdf#search='patrick%20ngulube'

_____. Cost analysis and the effective management of records throughout their life cycle. *Journal of the South African Society of Archivists* 44.

Ngulube, P. and Tafor, V.F. 2006. An overview of the management of public records and archives in the member countries of the East and Southern Africa Regional Branch of the International Council on Archives (ESARBICA). *Journal of the Society of Archivists* 27. 1: 58-83.

Nihal, A. and TekinSÄozer, M. 2007. Cox Regression Models With Nonproportional Hazards Applied To Lung Cancer Survival Data. *Hacettepe Journal of Mathematics and Statistics* 36. 2: 157 – 167

NHS Foundation Trust. 2011. *NHS policy and the record management code of practice*. England

NHS Foundation Trust. 2011. *Retention, Disposal and Destruction of Clinical Records Procedure, NHS policy and the record management code of practice*. England

_____ 2012, 2015. Health Records Management Policy & Strategy, Northern Lincolnshire and Goole Hospitals NHS Foundation Trust Retrieved 4/23/2017 from https://www.nlg.nhs.uk/content/uploads/.../Health-records-management-strategy.pdf

Office of the Federal Register. 1994. *Guide to Record Retention Requirements in the Code of Federal Regulations.* Washington, DC: National Archives and Records Administration.

Oakes, D. 1997. The asymptotic information in censored survival data. *Biometrika* 64. 441–448.

Osundina, K. S., Kolawole, J. A. and Ogunrewo, J. O. 2015. The Role Of Records Management In Secondary Health Care Delivery System In Selected State Hospitals In Osun State, Nigeria. Retrieved 4/23/2017 from https://www.iiste.org/Journals/index.php/IKM/article/viewFile/27823/28550

Oweghoro, B. M. 2015. Health Records Retention and Disposal in Nigerian Hospitals: Survey of Policies, Practices and Procedures,*Afr. J. Lib. Arch. & Inf. Sc*. 25. 1: 69-76.

Paradoxes (n.d.) *29th Information Systems Research Seminar in Scandinavia, Copenhagen, the IRIS Association*. Reytieved 4/23/2017 from https://books.google.com.ng/books?isbn=1317238923

Parzen, M. and Lipsitz, S. R. 1999. A Global Goodness-of-Fit Statistic for Cox Regression Models. *Biometrica* 55.2 *Retrieved 4/23/2017 from* file:///C:/Users/USER/Desktop/MY%20WORK%202017A/RELATED%20MAT

Penn, I.A., Pennix, G. B. and Coulson, J. 1994. *Records management handbook*. 2nd ed. Vermont: Gower.

Ponnuraja, C and Venkatesan P. 2010. Survival models for exploring Tuberculosis clinical trial data-an empirical comparison. *TuberculosisResearchCentre(ICMR)Chennai*. 3. 600. |

Pickett, F. 2012. *Health Information Management and the Health Care Institution. HealthInformation Management: Principles and Organization for Health Information Services*. M. A.Skurka. San Francisco, CA, John Wiley & Sons.

Pearce-Moses, R. 2005. *A Glossary of Archival and Records Terminology.* Chicago, IL: The Society of American Archivists, Information and Knowledge Management

Pelz, C. J., and Klein, J. P. 1996. *Analysis of survival data: A comparison of three major statistical packages (SAS, SPSS, BMDP*). Tech. Rep. 17, Division of Biostatistics, Medical College of Wisconsin, Milwaukee.

Ramadurai. M., and Ponnuraja, C. 2011. Non-Parametric Estimation of The Survival Probability of Children Affected by TB Meningitis. -*Journal of Arts, Science & Commerce*2231-4172.

Ramakrishnan, M and Ravana, R. 2013. Non-parametric methods for comparing two survival distributions. *Journal of Arts, Science & Commerce* . Retrieved 9/12/2017 from www.researchersworld.com

Records Management Bulletin, 2002. *Records management advice prepared for GNWT employees by the Records Management Unit of Public Works and Services* No. 6-July 2002 Retrieved 9/12/2017 from https://www.inf.gov.nt.ca/.../files/gnwt

_____ 2012. *Records Management Unit of Public Works and Services, Northwest Territories*. . Retrieved 9/12/2017 from https://www.inf.gov.nt.ca/.../files/bulletin_29

Records Management: *NHS Code of Practice* (2nd Edition). Business and Corporate (Non-Health) Records Retention Schedule, Annex D2

Records Management University of Washington. 2014. *Maintaining a Filing System, Identifying Inactive Files.* . Retrieved 9/12/2017 from https://finance.uw.edu/recmgt/files-mgt

Reitz, J.M. 2004. *Dictionary for Library and Information Science*. Westport, Conn: Libraries Unlimited.

Rinehart-Thompson, L. A. 2008. Storage Media Profiles and Health Record Retention Practice Patterns in Acute Care Hospitals. *Perspectives in Health Information Management* 5. 9.

Rich, J. T.J., Gail, N., Paniello, R.C., Voelker, C. C. J., Nussenbaum, B. and Wang, E. W. 2010. A Practical Guide to Understanding Kaplan-Meier Curves.*Otolaryngol Head Neck Surg* 143. 3: 331–336. doi:10.1016/j.otohns.2010.05.007.

Rinehart-Thompson, L. A. 2008. Record retention practices among the Nation's Most Wired hospitals. *Perspectives in Health Information Management* 5. 8. [PMCID: PMC2430773][PubMed: 18563205]

_____ 2008. Storage media profiles and health record retention practice patterns in acute care hospital. *Perspectives in Health Information Management*. 5. 9. [PMCID: PMC2435263] [PubMed: 18574517]

Rinku, S. and ManashPratim B. 2016. Comparing Cox Proportional Hazard Model and Parametric Counterpart in the Analysis of Esophagus Cancer Patient Data. *IOSR Journal of Mathematics* 12. 5: Ver. V (Sep. - Oct.2016), PP 16-21 Retrieved 9/12/2017 from www.iosrjournals.org

Ritesh, S. and Keshab, M. 2011.Survival analysis in clinical trials: Basics and must know areas *Perspect Clin Res* 2. 4: 145–148. doi: 10.4103/2229-3485.86872

Roach, W. H., Hoban,R. G., Brocolo, B. M., Roth, A. B., and Blanchard, T. S. ., 2006. *Medical Records and the Law.* Jones and Bartlett Publishers.Sudbury, Mass,

Rockefeller Archive Center. 2008. Records Retention and Disposition Guidelines. Collaborative Electronic Records Project. 21 Retrieved 9/12/2017 from http://siarchives.si.edu/cerp/cerpindex.htm

Rodrigues, A. D. S. 2013. *An Archival Collecting Framework for the records generated by South Africa's Portuguese Community-Based Organisation in Gauteng*. Pretoria, University of South Africa. Phd thesis

Rodríguez, G. 2007. Survival Models. 21 Retrieved 9/12/2017 from http://data.princeton.edu/wws509/notes/c7.pdf

Royal Cornwall hospital NHS trust. 1999. Health Records Retention/Destruction Policy. Retrieved 9/12/2017 from https://www.igt.hscic.gov.uk/.../RCHT_Health%20records%20destruction-retention%2

Saurabh, K. 2008. *Reliability Analysis and Cost Modeling of Degrading Systems*. Doctoral Thesis, Luleå University of Technology, Division of Operation and Maintenance Engineering 2008.

Scherm, H., and Ojiambo, P. S. 2004.. Applications of survival analysis in botanical epidemiology. *Phytopathology* 94. 1022-1026.

Schneider, H., P. Barron, et al. 2007. *The Promise and the Practice of Transformation in South Africa's Health System*. State of The Nation: South Africa 2007. S. Buhlungu, J. Daniel, R. Southall and J. Lutchman. Cape Town, HSRC Press**:** 289-311.

Schoenfeld, D. 1980. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* 67. 1: 145-153.

Schoenfeld, D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika* 69. 1.

Shortliffe Edward H. 1999. The Evolution of Electronic Medical Records Academic Medicine 1999;74(4):414-419

Scottish Government Records Management: NHS Code of Practice (Scotland) Version (2012) ANNEX B - 'The Management, Retention and Disposal of Personal Health Records

Shepherd, E. and Yeo, G. 2003. *Managing records: a handbook of principles and practices.* London: Facet Publishing.

Singapore Medical Association Council. 2000. Medical Records: Making and Retaining Them. Retrieved 9/12/2017 from.
https://www.sma.org.sg/UploadedImg/files/.../medical_records.pdf

Sibanda, R. 2011. *Developing a Service Quality Measurement Instrument for Archival Institutions*. Pretoria, University of South Africa. Phd thesis.

Singh, H. 2011. Formulating A Document Retention Policy. Retrieved 9/12/2017 from http://www.nixonpeabody.com/linked_media/publications/TIPA_04282004

Singapore Medical Association Council. 2000. Medical Records: Making and Retaining Them
Retrieved 9/12/2017 from
https://www.sma.org.sg/UploadedImg/files/.../medical_records.pdf

Skurka, M. A. 1998. *Health Information Management: Principles and Organization for Health Record Services*. Chicago, AHA Press.

SmithTyler, S. and Ryan, M. A. K. (nd). *Survival Analysis Using Cox Proportional Hazards Modeling For Single And Multiple Event Time Data*. Department of Defense Center for Deployment Health Research,Naval Health Research Center, San Diego, CA. Paper 254-28

Smith, T. and Smith, B. 2000. *Survival Analysis and The Application Of Cox's Proportional Hazards Modeling Using SAS Incomplete*

Society of American Archivists, 2016. Records Continuum". *A Glossary of Archival and Records Terminology*. *Society of American Archivists*. *Retrieved Month Day, Year from the website*

Stanger, K. C. and Olson, J. K. 2007. *Record Retention and Destruction, solution for Healthcare Professionals.* Hawley TrozellEnnia & Hawley.

Stanley, C., Molyneux, E. and Mukaka, M. 2016. Comparison of performance of exponential, Cox proportional hazards, weibull and frailty survival models for analysis of small sample size data. *Journal of Medical Statistics and Informatics*. Retrieved 9/12/2017 from http://www.hoajonline.com/journals/pdf/2053-7662-4-.pdf

Stanley, C., Molyneux, E. and Mukaka, M. 2016. *Comparison of Cox's Regression Model and Parametric Models*, Retrieved 9/12/2017 from

Stevenson, M. 2009. *An Introduction to Survival Analysis R*etrieved 4/23/2017 from http://www.massey.ac.nz/massey/fms/Colleges/College%20of%20Sciences/Epicenter/docs/ASVCS/Stevenson_survival_analysis_195_721

Stine, R. A. 2016. *Explaining Normal Quantile-Quantile Plots through Animation: The Water-Filling Analogy Incomplete*

Sturm, C. 2010. Record keeping for practitioners; *"CE Corner"* APA's guidelines help psychologists steer through the sometimes murky waters of how best to document and protect patient information. February 2012, Vol 43, No. 2

Sullivan, F. and Wyatt, J. C. 2006. *ABC of Health Informatics*. Blackwell Publishing

_____ 2009. *ABC of Health Informatics*. Retrieved 9/12/2017 from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1444312804.html

Sujatha, V. and Kalpanapriya, D. *2016.* A study on survival analysis International *Journal of Development Research 06. 01: 6439-6443.*

Tang Y, Xie M, and Goh T. N., (2003). Statistical analysis of a Weibull extension model. Communication in Statistic. Theory and Methods. 32: 913- 928.

Tanbakuchi, A. 2009. *Introductory Statistics Lectures Assessing Normality*. Retrieved 9/12/2017 from http://www.u.arizona.edu/~kuchi/Courses/MAT167/Files/LH_LEC.0450.RandVars.AssesNorm

Tavakoli N. 2007. *Investigation of retention and destruction process of medical record and provide guidelines in hospitals in Isfahan* [In Persian]. Isfahan University of Medical Science, Faculty of Medical Informatics & Management,

Tavakoli, N., Saghaianneja, S., Isfahan, M and. Reza, H. 2012. A Comparative Study of Laws and Procedures Pertaining to the Medical Records Retention in Selected Countries University of Medical Sciences, Isfahan, Iran

The American Health Information Management Association (AHIMA), 2011,.in a guide for retention and disposal of health information (patient medical records)

The American Health Information Management Association (AHIMA), 2013.in a guide for retention and disposal of health information (patient medical records)

The National Hospitals Office. 2007. *NHO Code of Practice for Healthcare Records Management. Version 2.0 (illustrated Retrieved 4/23/2017 from* https://www.hse.ie/eng/services/publications/hospitals/nho-code-of-practice-for-healthcare-records-management-version-2-0.pdf

The National Electronic Commerce Coordinating Council. 2004. *Challenges in Managing Records in the 21st Century*. United States of America

The South African Institute of Chartered Accountants, 2013.Guide on the Retention of Records

The Health Service Executive. 2013. *Standards and Recommended Practices for Healthcare Records Management*Retrieved 9/12/2017 from http://www.hse.ie/eng/services/list/3/acutehospitals/hospitals/ulh/staff/resources/ppgs/rm/recret2013.pdf

The Medical and Dental Defence Union of Scotland. 2013. *Essential guide to medical and dental records*. Retrieved 9/14/2017 from https://www.medicalprotection.org/uk/.

The University of Massachusetts, 2009. *Records Retention/Disposition Matrix.* Retrieved 9/14/2017. *iuc-ohio.org/wp-content/uploads/2018/02/IUC-Model-Schedule1.pdf*

The University of Montana. 2002. A Hybrid Entity as Defined by HIPAA Destruction/Disposal of Patient Health Information UK

Therneau, T. M. 2000. *Modeling s urvival data: extending the Cox model*. USA: Springer;

Therneau, T. M. and Grambsch, P. M., 2000. *Modelling Survival Data: Extending the Cox Model*, (Springer, New York, 2000).

Thomas, H., Jr. R. and Rae N. Cogar1**.** 2003. *Records Retention: An Essential Part of Corporate Compliance*Retrieved 9/12/2017 from pps.americanbar.org/buslaw/newsletter/0019/materials/recordretention.pdf

Viscomi, S., Pastore, G., Dama, E., Zuccolo, L., Pearce, N. and Merletti F, 2006. Life expectancy as an indicator of outcome in follow-up of population-based cancer registries: the example of childhood leukemia. *Ann Oncol* 17. 1:167–71. doi: 10.1093/annonc/mdj050.[PubMed: 16249212].

UK department of Health, 2006. *Records Management: NHS Code of Practice, Part 1*. (2nd Edition)

_____ 2006. *Records Management: NHS Code of Practice*. (2nd Edition)

_____ 2006. *Records Management: NHS Code of Practice*. Part 2 (2nd Edition)

University of Strathclyde. 2012. *Best Practice Guidance on Information and Records Management*. v1.0, Glasgow

University of Waterloo. 2016. Introduction to the University of Waterloo Records Classification and Retention Schedules (WatClass) policies procedure guidelines Policy 46 - Information Management.

Upward, F. 2000. Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond – a personal reflection. *Records Management Journal* 10. 3: 15-139.

_____ 2003. *Structuring the records continuum – part one: post-custodial principles and properties*. Retrieved 9/12/2017 from http://.sims.monash.edu.au/research/rcrg/publications/recordscontinuum/fupp1.h

Uthman O. 2007. Effect of low birth weight on infant mortality: Analysis Using Weibull Hazard Model. *The Internet Journal of Epidemiology* 6. 1.

Wang and Chow. 2007.Sample Size Calculation For Comparing Time-to-Event Data**.** *Wiley Encyclopedia of Clinical Trials*, Copyright © 2007, John Wiley & Sons, Inc.

Wang, F. K.(Incomplete) 2000. A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliability Engineering and System Safety* 70 (2000) 305±31

Were, S. M. 2015. *Management of Records in Health Institutions International Journal of Science and Research*. 2319 Retrieved 9/12/2017 from 7064https://www.ijsr.net/archive/v4i2/OCT141017.pdf

Weibull W. 1951. A statistical distribution function of wide applicability. *JApplMechanics* **18**. 3: 293–7.

Wessels,W. R. 2007. Use of the weibull versus exponential to model part reliability. *Annual Reliability and Maintainability Symposium*, pp. 131-135, 2007.

Wey, A., Connett, J. and Rudser, K. 2014. *Combining parametric, semi-parametric, andnon-parametric survival models with stacked survival models*. Retrieved 9/12/2017 from ArXiv:1309.7936v6 [stat.ME] 20 Dec 2014

Wilk, M.B.; Gnanadesikan, R. 1968. "Probability plotting methods for the analysis of data", *Biometrika*, Biometrika Trust, **55**(1): 1–17. doi:10.1093/biomet/55.1.1, JSTOR 2334448, PMID 5661047

World Bank Group. 2000. *Managing Records as the basis for Effective Service Delivery and Public Accountability in Development: An Introduction to Core Principles for Staff of the World Bank and its Partners*. Retrieved May 29, 2012 from http://web.worldbank.org/.

Wissmann, S. 2015. Addressing Challenges to the Health Information Management Profession: An Australian Perspective. *Perspectives in Health Information Management*. Retrieved 12/9/2017 from https://perspectives.ahima.org/addressing-challenges-to-the-health-information-manag

World Bank and International Records Management Trust, 2000. Managing Records as the Basis for Effective Service Delivery and Public Accountability in Development: An Introduction to Core Principles for Staff of the World Bank and Its Partners

World Health Organization. 2006. *Medical records management manual. A guide for developing countries*. Western pacific region.

Yamane, Taro. 1967. *Statistics: An Introductory Analysis*. 2nd Ed., New York: Harper and Row.

Yaya, J. A., Asunmo, A. A., Abolarinwa, S. T. and Onyenekwe, N. L. 2015. Challenges of Record Management in two Health Institutions in Lagos State, Nigeria International *Journal of Research in Humanities and Social Studies* 2. 12: 2394-6288.

Zegers M., de Bruijne, M.C., Spreeuwenberg, P., Wagner, C., Groenewegen, P.P. and, van der Wal, G. 2009. Quality of patient record keeping: an indicator of the quality of care? *BMJ QualSaf*. Retrieved May 12 2012 from: http://qualitysafety.bmj.com/ content/early/2011/02/08/bmjqs.2009.038976.

Zhao, G. 2008. *Nonparametric and Parametric Survival Analysis of CensoredData with Possible Violation of Method Assumptions.* A Thesis Submitted tothe Faculty of The Graduate School atThe University of North Carolina at Greensboro

Zhu, Ruoqing and Kosorok, Michael R. 2012. Recursively imputed survival trees. *Journal of the American Statistical Association* 107, 331–340.

APPENDICES

APPENDIX 1

DATA EXTRACTION SHEET TO PATIENTS RECORDS TIME-TO-DORMANCY

	Form No.
***P*atients Regt No.**	
Date of first contact	
Penultimate Contact	
Date of last Contact	
State of Residence	
Date of Birth	
Age at registration	

Gender	***M***	***F***					
Level of education	***None***	***Pry***	***Sec***	***Tertiary***			
Clinic(s) attended	**MOP**	**SOP**	**CHOP**	**GYNE**	**OTHERS**		
Principal Diagnosis							
Was Surgery done	Yes	No					
Was Patient Admitted	Yes	No					
No. of Admission.							
Length of Stay (Days)							
Treatment Outcome	Alive	Died	DAMA	Referred			

PLEASE USE PENCIL ONLY

APPENDIX 2

Estimation of required sample size explained

(derived from: 'Adequacy of Sample Size in Health Studies authored by Stanley Lemeshow, David W. Hosmer Jr, Janelle Klar and Stephen K. Lwanga. 1990)

To develop the expressions needed for ascertainment of sample size for this study, a model with the statistical distribution of survival times in the population following an exponential distribution will be assumed.

Under this model, the probability that an individual will survive for more than t time units is

$$S(t) = e^{\lambda t}$$

and that the survival time is less than or equal to t is

$$P(t) = 1 - e^{\lambda t}$$

$$=1 - S(t)$$

The conditional probability that a patient's medical record will become dormant given that the record is still active prior to that interval, is constant and equal to λ. In

other words the hazard function for this model is constant and equal to λ. Under the assumption of exponential survival the maximum likelihood estimate of the hazard function is

$$\hat{\lambda} = \frac{d}{F}$$

where d is the number of events and F is the total follow-up time. Thus the incidence density is the estimate of the hazard under exponential survival. This result will provide the basis for development of formulae to determine necessary sample size for this study.

Let $t_{1...}$ t_n the observed dormancy times for a group of patient medical records.

In this case $\hat{\lambda} = \frac{1}{t}$, where

$$t = \left(\frac{1}{n}\right)\sum_{i-1}^{n} t_i .$$

It follows from the theory of maximum likelihood estimation that where n is sufficiently large, $\hat{\lambda}$ is normally distributed with mean λ. and variance $\frac{\lambda^2}{n}$. This information may be used in the same way that similar information was used to develop sample size formulae for estimation and tests about proportions. The sample size which is necessary to estimate λ to within ε of its true value with probability (1- α) is given by the formula

$$\text{n} = \left[\frac{z_{1-\alpha/2}}{\varepsilon}\right]^2$$

from $|\hat{\lambda} - \lambda| = z_{1-\alpha/2}\left[\lambda/\sqrt{n}\right]$ with $\varepsilon = \frac{|\hat{\lambda} - \lambda|}{\lambda}$.

The value of n may be looked up directly from the table 'Sample Size to Estimate the Incidence Rate to within ε percent, with 99%, 95% or 90% Confidence Level'

APPENDIX 3

List of diagnosis of 1 day active records

1. Presbyopia
2. Jaundice
3. Fever
4. Cough
5. URTI
6. Anatomical mid pontial lesion
7. RIH
8. Endemic goitre
9. Fever
10. Abdominal discomfort
11. 1st deg perineal tear
12. Incomplete abortion
13. Itching and watering of both eyes
14. Vaginal bleeding (abnomal)
15. Catarat
16. Scanty menses
17. Poor Vision
18. Primigravida
19. Poor Vision
20. Fever
21. Laryngo malacia
22. Chronic PID
23. Hydrocephalis
24. Nephritis
25. Epilepsy
26. Haemorrhoid
27. Viral conjuctivitis
28. Ophthalmitis Neonatorum
29. Cornual ulcer
30. CSOM
31. Assault with multiple injury
32. Cystocele
33. Congenital hydrocoele with PPV
34. Amenorrhoea 2nd deg
35. infertility
36. Acute urine retention
37. Infertility 2nd
38. Nasopharyngeal cancer
39. Infertility 1st deg
40. Portum hypertrophic scar of R breast
41. Functional Ovarian cyst

42. Tongue tie
43. Endophytic cancer with pyometra stage I
44. Medical Examination
45. Fever
46. Bilateral cataract
47. Primary infertility
48. severe nystagmus
49. Vertigo
50. pityriasis vensicilor
51. Ca cervix
52. FILARIASIS
53. Tibia_Oblique
54. Ca cervix
55. Hortous disease
56. Paraphagia
57. ANDI
58. Ca thyroid
59. Poor Vision
60. Left orbital swelling
61. Gastric fuling defects
62. post burn contracture fingers
63. Scrotal swelling Rt
64. seizure disorder
65. Acute subdural heamatoma
66. Hypertension
67. Idiopathic with nerve palsy
68. Painful erection
69. Severe Hyperpigmented
70. Aptic astrocytoma
71. Family planning
72. Impaired memory
73. complex partial seizure
74. Ca Breast
75. Filariasis
76. Mal-united distal ends tibia fibula
77. Breast lump
78. Hypertension
79. Diarrhea
80. Fracture of the distal fibula
81. Left shoulder injury
82. Wound stiches
83. Severe Abdominal discomfort
84. Diabetes
85. Papular rash
86. Peripheral Neuropathy
87. Fever
88. Psychodomeitic
89. Infertily
90. Missed abortion
91. Neonatal jaundice
92. Neonatal septicacemia
93. Acute tonsillitis
94. Allergic Rhinutis
95. Abdominal pain
96. Neonatal septiceamia
97. Dysfunctional uterine bleeding
98. PID
99. Short femur
100. hearing lost
101. Hernia (inguinoscrotal)
102. Hernia (ingunial)
103. infertility
104. Dislocation (Knee joints)
105. Anxiety Neurosis
106. Foreign Body in R Ear
107. Myelomeningocele
108. Tumour of the Brain
109. Bilateral hand deformation
110. Burns on the left hand
111. hermia (Bilateral inguinal)
112. Ante natal booking
113. Ante natal booking
114. RISH
115. Bilateral cataract
116. Epigastic pain
117. Soprintemor
118. Cervical lesion
119. Delayed Developmental milestone
120. Ricketts
121. birth Asphyxia associated Cerebral palsy
122. Febrile convulsion
123. Maxillary sinusitis
124. Hypertension
125. Chronic Myeloid Leukaemia
126. conjuctivitis

127. Bilateral watery discharge
128. Malaria
129. Severe birth asphyxia
130. Lower limb deformity
131. Febrile Convulsions
132. URTI
133. Foreign body in the nose
134. Uterine Bleeding (Abnormal)
135. Malaria
136. Preterm baby
137. PID
138. Ante Natal booking
139. Bilateral Hydroceles
140. Poor Vision
141. Neonatal sepsis
142. conjuctivitis trauma
143. Fibrocystic dx of the breast
144. Vomitting / impaired sensorium
145. Lid laceration
146. Head injury
147. Obstructed labour
148. RTA.
149. Diverticaticlar of Rectum
150. Hepatosplenopathy
151. Urinary difficulties
152. Itching inside the R ear
153. Progressive weight loss of obscure etiology
154. Convulsion (seizure)
155. Foreign body in the left eye
156. infertility
157. Measles
158. Chronic open angle glaucoma
159. High blood Pressure
160. Cerebral palsy
161. Left visual impairment
162. Post infective flexion deformity
163. infertility
164. RTA Recurrent Headaches
165. Thyroglosal cyst
166. Redness of the eye
167. infertility
168. infertility
169. PTB
170. Lymphoma
171. Mental Retardation
172. Corneal keratopathy/ staphyloma
173. Congenital toxoplasmosis
174. RTA (L Tibia fibula fracture)
175. Fracture of the mid femur
176. Depressed skull fracture
177. TB spine
178. Distal 1/3 R Femur
179. RTA
180. Carcinoma of the cervix
181. infertility
182. Fibriroadenoma Rt breast
183. Carcinoma of Rt breast
184. appendicitis
185. Foreign body
186. Simple Goitre
187. Congenital Hydrocele of the left scrotum
188. Ante natal booking
189. Pocidentia
190. Ligament lesion (Rt) knee
191. Lumbosacral pain
192. Fugal infection of the nails
193. Wax in both ears
194. prolapse
195. Seizure
196. TB
197. conjuctivitis
198. Vesico_Vaginal fistula
199. Orchdiitis
200. Obesity
201. Ruptured Appendicitis
202. Inability to talk well
203. perforated typhoid enteritis
204. Cleft lip palate
205. Conjuctivitis
206. mass in the (R) Ear
207. Diabetics milletus
208. Epilepsy
209. Cystocele
210. Painless swelling of scrotal sac
211. NIDDM
212. Amemorrhae

213. Eczema
214. Ca Cervix
215. Infection of Pinnia
216. Cervical spondylosis with radiculopathy
217. Oblique Termonalphalanix with separation of epiphysis ring f
218. Amenorrhoea
219. Complex fistula in-anohypertension
220. Haeomorrhis
221. hydrocelle
222. conjunctivitis
223. Fractured hand
224. Ca cervix
225. Eczema
226. hyperhydrosis (Indiopathic)
227. Amenorrhoea & consequenl primary infertility
228. Frequent priapism
229. Medical Examination
230. Pyogenic granuloma
231. Keloid & poste
232. Diabetics Milletus
233. Cervical spondylosis
234. Uterine fibroid
235. Tinea crisis + ulcer
236. Ca of the rectum
237. Medical Examination
238. Cranosynostosis
239. BPH
240. Early Osteomyelitis
241. Post abortal infection
242. Chronic osteomyelitis
243. Recurrent dislocation L shoulder
244. Recurrent pyomyositis
245. fall Lt leg
246. Ante natal booking
247. Severe Oligospermia
248. infertility
249. R inguino scrotal hernia
250. Partial Fistula
251. Cervical lymphadenopathy
252. Malaria
253. Acute intestinal obstruction
254. Foreign body
255. Assault
256. conjuctivitis
257. Asthma
258. Viraemia
259. Mental Retardation , cerebral palsy
260. Icemyoxis
261. Hearing impairment R ear
262. Abortion
263. lumbosacral pain
264. Lower resp tract infection
265. Malaria
266. Spinal cord compression syndrone
267. Ganglion
268. Refractive error
269. Ca prostate
270. Herpes zooter of labial
271. High Blood pressure
272. Polio paralysis
273. Malaria
274. Infretility
275. L CSOM with Rhinitis
276. RTA
277. Sepsis
278. Fever
279. Acute Epididynminitis
280. Burns
281. Hearing impairment
282. Antenatal booking
283. Fibroid in pregnancy
284. Otitis Media Chronic
285. Malaria
286. Chest injury
287. Body itching
288. Rt supra orbit scar
289. Typhoid enteritis
290. Malaria
291. Malaria
292. Trichomoniasis
293. Vaginal bleeding (abnomal)
294. Ante natal visit
295. infertility
296. Ca rectum
297. bilateral eye ache

298. Injury to Arm
299. Chronic allergic conjuctivitis
300. Fever
301. Loss of Right side of Nose
302. Brain damage
303. Pterygium and immature cataract
304. Pterygium
305. Post burns contracture of Lt hand
306. Keloids
307. Perforation of R tympanic membrane
308. Ambigious genitalia
309. Establish labour
310. conjuctivitis
311. examination for obtain a driving license
312. Uterine Fibroid
313. Flexion constrictive of 5th R finger
314. infertility
315. Abdominopelvic Malignancy
316. Herpes Facial
317. Impated wax
318. Cataract
319. conjuctivitis
320. GTD
321. vaginal Bleeding (abnormal)
322. Cataract
323. Gynaecomastia
324. Gynaecomastia
325. cataract
326. Removal of IUCD
327. Left interior retina detachment
328. cataract complicated Cornea irregular
329. Ovarian Cyst
330. Vascular injury Bronchial artery
331. Cataract
332. URTI
333. CVA
334. Ante Natal booking
335. Pneumonia
336. Sickles in Haemolytic crisis
337. rhinosinusitis Chronic
338. Haemorrhoids
339. Ceacal tumour
340. Ante-natal booking
341. Missed Abortion
342. Ca breast
343. Cataract
344. Corneal opening
345. Occupational Accident
346. RTA (finger injury)
347. Head injury
348. Crush injury R little finger
349. Ptergyium
350. spleenomegaly Massive
351. Congenital Anormaly
352. deformity
353. Forehand injury
354. Haemorrhoids
355. Swelling on the right side of the forehead
356. Diplopia & Blurring vision
357. Eye pain & itching
358. Psychiatric
359. Non - ulcer dyspepsia
360. infertility
361. Urethral stricture
362. Congenital deafness & dumbness
363. Partial deafness
364. Left breast lump
365. Cervical Lymphademopathy
366. Bil congenital Hydrocoele
367. Refractive Error
368. Tumour
369. Bilateral red eye
370. Surgical contraception
371. Bilateral cataract
372. Amenorrhea
373. Amenorrhea
374. infertility
375. Pain in right ear
376. conjunctivitis
377. PID
378. conjuctivitis
379. Blurring of vision
380. conjuctivitis
381. Otitis media
382. Fracture
383. Ulcer (PUD)

384. infertility
385. Pulmonary TB
386. Amblyopia
387. Refractive error
388. Foreign body in the L Nasal cavity
389. Testicular tumour
390. Injury to the L 2nd toe . Recurrent ulcer885064
391. infertility
392. URTI
393. Urethral injury
394. Poor vision
395. Refractive error Visual Acuity
396. Vision blurring
397. RTA
398. Keratoderma
399. External Ampular demand
400. Injury to the left eye
401. Fever
402. Cough
403. Vernal conjuctivitis
404. Vernal conjuctivitis
405. Abdominal pain
406. Diabetes Mellitus
407. Right sided Hydrocode
408. Congenital cyst Rt eye
409. Foreign body in the right ear
410. Acoustic neuromal
411. Haemorrhoids
412. Abdominal cramp
413. Collagen vascular disease
414. Fibrocystic disease of the breast
415. Bilateral geno vara
416. RTA
417. cancer L Breast
418. Short sighteness
419. Antenatal booking
420. Antenatal booking
421. conjuctivitis
422. Cerebral palsy
423. Cataract
424. Beningn Protate Hypertrophy
425. Loss of function of finger
426. Intraabdominal mass
427. Fall
428. Bilateral (Persistent watery eye discharge)
429. Pitutary tumour
430. Diabetics
431. Abdominal pains and menstural irregularities
432. Atopic Eczema
433. Wax in the right ear
434. Skin disease
435. facial mass
436. conjuctivitis
437. pleural effusion
438. Astigmatism
439. Poor vision
440. Fungal (Trauma in Rt eye)
441. Bilateral cataract
442. Otomycosis
443. Discharge in the ear
444. cataract
445. infertility
446. infertility
447. Poor vision Rt eye
448. Nasophanyngeal cancer
449. infertility
450. CSOM
451. Poor vision
452. Malaria
453. In growing toe nail L hallux
454. Lipoma of the Burn's spale
455. Antenatal care booking
456. Nephrotic syndrome
457. intestinal obstruction
458. fracture Mid shaft Tibia & fibula
459. RTA
460. facial fistulla
461. infertility
462. Hypertenstimilus
463. Lymphocytic leukaemia
464. Leukoplakie
465. Microcephaly
466. Duct cell carcinoma of L breast
467. HX of swelling and pain R shoulder region
468. Fibrous ankyloglosis
469. Rt. check sebaceous cyst

470. Lump at Rt breast
471. carcinoma of the right breast
472. Acute head injury with compound skull bone #
473. RTA
474. Gun shot injury
475. Chronic osteeomyelitis
476. Bilatteral putting pedal oedema
477.
478. Anorexia
479. URTI
480. Simple nodular Gritre
481. Lumbago
482. Occupational Asthma
483. Dysfunctional uterine bleeding
484. Ca breast
485. Bi-ventricular cardiac failure
486. Urinary tract infection
487. Lichenoid dermatitis
488. mass
489. Locally Advanced breast Ca
490. 2nd degree Menogenic bladder
491. Motor Neuro Disease
492. Alcoholic liver cirrhosis
493. breast CA
494. Rhabdomyosacroma
495.
496. RTA
497. Cyesis
498. Abdominal Hysterectomy

Appendices 4: Approval by UI/UCU Ethics Committee

Appendices 5:

Approval by UCH, Ibadan management to collect data