UNIVERSITY OF GHANA, LEGON

MODELLING INSURANCE ATTRITION USING SURVIVAL ANALYSIS – A CASE STUDY OF GHANA

BY MERCY JESSIE ASARE (10805437)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF GHANA IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF PHILOSOPHY, ACTUARIAL SCIENCE

May, 2022

Declaration

I hereby declare that this submission is my own work towards the award of the Master of Philosophy degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

		5/30/2022
MERCY JESSIE ASARE	....	....
Student		
(10805437)	Signature	Date

Certified by:		5/30/2022
ISAAC BAIDOO	....	....
Principal Supervisor	Signature	Date

Certified by:		5/31/2022
GODWIN DEBRAH	....	....
Co-supervisor	Signature	Date

Dedication

I dedicate this work to my family, and friends.

Abstract

Life insurance operations immensely contributes to the economic growth and development of a nation while also serving as an alternative form of internal fund mobilization for developing economies. This notwithstanding, life insurance companies tend to face challenges. One of these challenges they face is insurance attrition. This condition arises when insurance policies are terminated by the insurer as a result of discontinuation of premium payment after a specified period of time called the grace period, and also by the policy holder.

Many factor(s) contribute to insurance attrition. The study focused on the length of survival time to attrition and the covariates that are likely to influence attrition. Randomly selected data was used in the study. Data was provided by an insurance company in Ghana for the period May 2018 to April 2021.

The study employed Kaplan-Meier estimators, log-rant test and Cox regression model for the analysis of data. The study revealed the survival time of a new client is 16 weeks after subscribing on to policy. It also revealed assuming a three year period, attrition will occur after 15 weeks of being in force. The study concludes that marital status, product type, base rate change, deduction source are the factors that influence insurance attrition in Ghana.

Acknowledgement

I express my gratitude and thanks to the Almighty God for his divine protection and direction throughout my study.

I am grateful to my supervisors, Dr. Isaac Baidoo and Dr. Godwin Debrah, for their time, consistent support, patience, and encouragement in helping me write this thesis.

I also express my heartfelt gratitude to my grandmother Mercy Jessie Martins, my father Rev Dr. Francis Yao Amaglo, my mother Josepha Atiamo, and siblings Jennifer, Erica, Grace and Emmanuel.

I am also grateful to Rev. Francis Saah-Ayisah for his prayers and support. And specially thank Joshua Hodinya for always being there. Thank you to my friends and course mates who have supported in their own special ways. God bless you all.

Contents

List of Abbreviation

PHM .... Proportional Hazard Model

NIC .... National Insurance Commission

K-M .... Kaplan- Meier

LCL .... Lower confidence bound

UCL .... Upper confidence bound

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction to Background of study

According to Alhassan & Biekpe (2015), there has been long-held beliefs that insurance market operations stimulate economic growth by serving as a conduit for long-term money to financial markets. The purpose of life insurance according to Goonetilleke & Caldera (2013) is to pay an agreed-upon amount called premium to the insurer on a specific date in exchange for a financial gain to the policyholder's dependents in the event of his or her death. This type of contract include uncertainties such as accidents, illness, or medical charges. Linder & Ronkainen, (2004) stressed that the insurance industry is important not just because of this important fact, but also because of its diverse business operations. The African Development Bank (2012) emphasized; the life insurance market is a better financial intermediary since it provides as both an income replacement and a savings mechanism for consumers in the case of loss of the breadwinner.

According to the African Development Bank (2012), another essential characteristic of life insurance is that the long-tail form of the contracts ensures that revenues generated by life insurance consumption are made available to financial agents through financial agents' and markets' intermediation operations. As a result of this feature, life insurance can be used as an alternative form of internal fund mobilization for developing economies such as Ghana, which have traditionally relied heavily on foreign financial allocations.

Linder & Ronkainen (2004), further stated that on the European Union (EU) stock market, insurance companies constitute the largest investors. Customers in the insurance sector across the world just like other sectors have the choice of selecting from a range of life insurance products from the provider who offers the most attractive benefits. According to Nazrina & Shahirah (2019), the insurance market is rapidly evolving, and hence, it is difficult for any one insurance firm to dominate the industry due to strong competition among the insurance companies.

Customers are critical to a company's long-term viability and survival, so acquiring new ones is its top priority. Normally, it is fairly easy to persuade customers to purchase insurance policies; but the challenge however becomes maintaining the customer's loyalty. Businesses cannot exist without customers. The Economist conducted a global survey in which 65 percent of respondents (senior executives of multinational corporations) revealed that customers were their top priority in achieving their business objectives (Gupta & Zeithaml, 2006). Thus, retention of customers is very pivotal to an insurance company's growth and profitability. Fu & Wang (2004) are of the view that, the retention/attrition ratio influences the two most important goals of the company, i.e. growth and profitability. It is also a key aspect in the insurance company's marketing, underwriting, pricing, and customer service initiatives, in addition to profit and growth.

Marin (2005) in a thesis, contributing to the understanding of elements in the assessment of the business risk of an insurance company notes that, loosing a customer is one of the operational risks an insurer assumes. In order for the company to measure its operational risk, it is required of the insurer to account for the possibility of losing a customer, i.e. policyholder staying with the company for at least a year. According to Fu and Wang (2019), in determining whether a policyholder will leave or not, the popularized binary logistic regression is the natural choice of methodology that is employed. In their study of attrition using survival analysis, Fu and Wang (2004) highlighted that though the binary retention model (such as logit and probit regression) have several advantages; which includes determining the event of interest exceeding the specified time, and also being simple to understand, there are many shortcomings of the model.

In a paper by Frederick Reichheld (2000), customers across a range of companies generate increasing profits each year they stay with the company. The author further revealed that in the financial institutions for example, a 5% increase in client retention improves profitability by 25% to 95%. This is so because the existing customer tends to buy more from the company over time, which in effect reduces the company's operating cost in serving the customer. Since attrition/retention is an important factor in the company's growth and profitability, understanding "when" or the timing difference of attrition in insurance is very important. According to the NIC, the insurance penetration rate as at 2019 is less than 2% and so increasing attrition rate or low retention rate could negatively impact the insurance industry greatly and the country's economy as a whole.

However, these firms tend to face many challenges which sometimes are company-specific as well as industry as a whole. Life insurance companies in Ghana like any other across the world face challenges as well. These challenges could be internal or external in nature. Many of such challenges are competition among insurers, insurance attrition, mismanagement, economic instability, inflation and many more. A major of these challenges in the life insurance sector is insurance attrition. Nazrina & Shahirah (2018) states attrition could be in different forms, that is, job or insurance attrition. The problem of job or insurance attrition arises when there is a decrease in the number of employees in a company or a decrease in the number of customers who purchased a product. There are two forms of insurance attrition. Goonetilleke & Caldera (2013) defined one of the forms of insurance attrition as a situation in which a policyholder stops paying insurance premiums, and after a period of continuous nonpayment, the policy is terminated by the insurer. The insurer is mandated under the insurance law to terminate the policy after a specified period of time called "grace period". Payment not made on term life insurance policy during a 30-90 days period, becomes lapsed and will no longer be in force. This notwithstanding, this time period (grace period) however, is pretty standard by jurisdiction, insurer and the type of life insurance product (Forbes Advisor, 2021).

According to the NIC, in Ghana, this provision allows for a grace period of 31 days following the due date of a premium. This notwithstanding, the grace periods in Ghana differ from one insurer to the other; some have grace period of three months, and others six months. Life insurance coverage remains in effect during the grace period, after which the policy lapses. If the insured passes away during the grace period, the insurance company is permitted to deduct from the death benefit payable the amount of any unpaid premium.

The second form of insurance attrition is where the policy holder voluntarily cancels the insurance policy after a period of time. Studies have shown insurance attrition is influenced by many factors. the contributing factors are customer address change, core customer status, premium increase, competitive market, marital status, age, and gender of holder (Marin (2005), Nazrina & Shahirah (2019), Mittal and Kamakura (2001), Tang et al. (2014)). In the life insurance business line of operations in Ghana, the products sold are universal life, funeral insurance, microinsurance, group life, whole life, term life, endowment insurance, and credit life insurance.

Unlike survival analysis, binary logistic models does not take into account all information on censored data, and can also not maximize information from time-varying variables. The use of survival analysis has a wide range of applications. It is employed mostly in biostatistics. This is obvious because of the estimation of time-to-death or the measurement of risk of death in mortality studies. It is also an unavoidable part of actuarial science, with life insurance accounting for the majority of its statistical applications, (Marin, 2004). The response (event of interest) variable in survival analysis is a continuous time and not a discreet yes or no. According to Fu & Wang (2019), the growing need to include covariates in the analysis of time-to-event data has brought forth the popular model; cox proportional hazard model (PH model).

In this study, the author proposes to use survival analysis to model customer attrition instead of the conventional binary logistic regression.

1.2 Problem Statement

Although customer retention is critical to the insurance company's long-term growth and profitability, attrition is however, an issue to consider. Han et al. (2010) identified in their research about the link between insurance development and national growth that, insurance plays a larger role in emerging countries as compared to developed countries. However, the NIC (2019) annual report, policy-holder inflows which measures the growth or contraction in total premium inflows from policyholders (including savings premium) shows the average policyholder inflow for 2019 was 21%; indicating a decrement from 24% in 2018 and 30% in 2017.

According to Reichheld (2000), customers that stay with a company year after year generate increased earnings for the company. The author went on to add that in a financial institution, a 5% increase in client retention for example, improves profitability by 25% to 95%. Nevertheless, Boadu et al., (2014) in assessing the life insurance in Ghana, concluded customers are increasingly becoming intolerant and hence as problems develop, they readily split off their relationship with the company.

However, a study done by the National Insurance Commission in 2019, 76.9% of insurance businesses in Ghana agree that market research in the industry is inadequate; indicating that the need for more market research cannot be overemphasized. According to the insurers, results from the market research assists insurers to better understand the market, identify challenges and how they could be addressed, and also make informed decisions. Therefore, the need to analyse customer attrition and the contributing factors.

1.3 Objectives of Study

For the above identified research problem, the study seeks to analyse the problem of customer attrition in the life insurance industry. And to attain this objective, the study must;

- Estimate the survival time of a new client.
- Estimate the survival time to attrition.
- Examine the covariate(s) that are most likely to contribute to insurance attrition (policy lapse).
- Make appropriate recommendations to the various agencies that will help inform policy decision in the life sector.

1.4 Research Questions

The following research questions will be considered while assessing insurance attrition in the life insurance industry in Ghana.

- When will attrition occur?
- Which covariate(s) are most likely to influence insurance attrition?

1.5 Significance of Study

This study is immensely significant in various ways to practitioners, policy makers and the researchers. The study adds to our understanding of the problem of attrition and its deterrent effect on the insurance industry by providing new insight and understanding, as well as some conclusions. To insurance practitioners, the discovery of the explanatory elements that contribute to insurance attrition in Ghana will aid policymakers in successfully managing and dealing with policyholder characteristics that influence the customer's decision to stop paying premiums and to also cancel their policies before end of policy tenure. It will reveal vital data for management to improve strategic decisions on the low insurance premium creation in the life insurance sector in a variety of ways, taking into consideration the country's peculiar geographical characteristics. This study would be among the few studies on insurance attrition and its influencing covariates in the life industry in Ghana. Again, this research provides the information that helps to fill the gap in the insurance attrition literature in Ghana. It aims to contribute to the existing body of knowledge and literature on life insurance in Ghana, as well as serving as a springboard for future research in this field. To researchers, this will serve as reference for further future study since there is limited work in this area of study. To policy makers, this will better place them in a position to better monitor attrition/retention rate of the life insurance companies in Ghana.

1.6 Methodology

A longitudinal data from May 2018 to April 2021 from a Ghanaian life insurance firm, will be used. The covariates in the dataset are age, gender, base rate change, marital status, policy tenure, employment sector, policy status, mode of premium payment, payment frequency, and product type. The data will first be cleaned. Respondents' profiles will be analyzed using descriptive statistics, whereas the Kaplan-Meier estimator, the log-rank test, and the Cox model will be employed for the further analysis. The Cox model will also be validated using the Score (logrank) test, Likelihood ratio test and the Wald test.

1.7 Expected contribution

The study's findings may contribute to a better understanding of attrition in the life insurance industry.

Chapter 2

Literature Review

2.1 The Global Life Insurance Mechanism

Insurance has a long history dating back to ancient times. It has over the years evolved into a contemporary industry that protects people from numerous hazards. For many years, the sector has been profitable, and it has played a significant role in both private and public long-term finance (Cleary Insurance, 2021). Insurance can be defined from different viewpoint of several disciplines, however, according to (World Bank, 2008), it can be defined as a risk transfer mechanism that is achieved through risk pooling.

According to Anzovin (2000) and Amicable Society (1854), the first life insurance policies were issued in the early eighteenth century. The first organization to sell life insurance was the Amicable Society for a Perpetual Assurance Office, which was created in London in 1706 by William Talbot and Sir Thomas Allen. Each member of the original life insurance plan was expected to make a predetermined annual payment per share on one to three shares, depending on their age, which varied from twelve to fifty-five years old. At the end of the year, wives and children of dead members received a portion of the "amicable contribution" in proportion to their shareholdings. Between 1787 and 1837, more than two dozen life insurance companies were established in the United States, but only about half of them survived.

According to Alhassan and Biekpe (2016), (Swiss Re, 2011), Between 2001 and

2010, overall insurance premiums in emerging markets increased by 11%, compared to 1.3 percent in the industrialised business. The life sector contributed for 13 percent of the growth in these emerging markets, compared to 0.6 percent for the industrialised sector. According to Pierre-Ignace et al. (2020), the life insurance sector over the last decade globally has seen substantial changes where developing economies have transformed into global growth engines. However, this growth in life insurance penetration in emerging nations is concentrated in Asia and Latin America, with China accounting for over a third of all premium growth in the emerging markets, (Alhassan and Biekpe (2016), (Swiss Re, 2011), Pierre-Ignace et al. (2020)).

Nonetheless, according to Pierre-Ignace et al. (2020), New challenges arose within the decade. Life insurers did not benefit from the bull market, with worldwide insurance penetration falling to 3% and premium growth slowing in most developed countries. Notwithstanding, premium growth in most developed countries, which hovered just below 2% per year, struggled to match up with GDP. Interest rates have also been lowered around the world, reducing investment portfolio returns. Most recently, the COVID-19 pandemic has pushed global interest rates even lower than they were during the global financial crisis of 2007–2008, causing life insurance stocks to underperform the market.

2.2 History of Insurance in Ghana

In the late 1800s, British businessmen introduced insurance to Ghana. The first insurance business in Ghana was created in 1924 by Royal Exchange Assurance, a British corporation, Enterprise Group (2018). At the time, there was no local regulation governing the insurance business in Ghana, thus insurance was offered through the country's foreign trading enterprises, who acted as agents for insurance companies in the United Kingdom and elsewhere. Marine (cargo) insurance covered items exported (mainly cocoa and gold) as well as goods brought into the country. As a result, there was no insurance market in Ghana at the time in the traditional sense; any appearance of an insurance market at the time was merely an adjunct to the British insurance market.

The Gold Coast Insurance Company, which began operations in 1955, was the first local insurance company to focus completely on life insurance, particularly African life insurance. It started providing large-scale life insurance to Africans. Until this company was founded, expatriate insurance companies primarily insured the lives of Europeans. Because there were no local life statistics or tables to base the grading of life insurance premiums on, and so the rates were based on those of the United Kingdom and were accordingly adjusted.

Regarding the history of Ghana's insurance industry, the year 1972 is pivotal. A number of insurance-related legislation were enacted, with the NRCD 95 being the most prominent. Industry unions such as the Insurance Institute of Ghana (IIG), the Ghana Insurance Brokers Association (GIBA), and the Ghana Insurance Association (GIA) were created in the post-independence era. In the same year, the then-ruling National Redemption Council imposed a number of rules on insurance companies, including the requirement that all insurers operating in Ghana establish their head offices in the country and that Ghanaians own 40% of the company's shares.

As a result of the adoption of that directive, the Ghana Reinsurance Organisation, was founded in a bid to prevent the outflow of premiums from the country. 1987 saw the establishment of the National Insurance Commission (NIC) in Ghana as the industry regulator responsible in Ghana for the establishment of capital requirements, dividend policies and investment criteria. The Association has developed a Bachelor's (insurance option) administration at the University of Ghana and also finances the National Commission for Road Safety to improve road safety in Ghana.

The Ghanaian insurance sector, because of its historical ties, is quite similar to the British insurance sector in terms of practices and policy writing. As at the end of 2019, the industry had 22 life, 27 non-life, 3 reinsurance and 91 brokerage firms (NIC, 2019). The Insurance industry in Ghana, and Africa in general (except South Africa), is a largely underdeveloped, or at best, developing market.

The elements that have contributed to this underdevelopment of the insurance industry in Ghana and Africa as whole can be attributed to following:

Low Disposable Income: There is a common adage that insurance is sold rather than bought. This phrase expresses the fact that most people do not purchase insurance as a matter of course. . This is one of the last items most individuals would purchase and hence an individual will not prioritize the purchase of insurance if he or she has a limited amount of disposable income. This is one of the primary reasons why the low-income segment of the population in most countries do not purchase insurance.

Low Insurance Sector Capacity: In general, the larger an insurance company's assets are, the better its risk-absorbing capacity. The total amount of assets held by the insurance sector in Ghana is quite low. This situation restricts the insurance industry's ability to underwrite huge risks.

Poor Public View: Delays in claim payment and insurance firms' denial of legitimate claims have caused the public to lose faith in the insurance industry, resulting in a poor view of the industry. While public perception of insurance seems to be improving, the insurance industry's image among ordinary Ghanaians is not very positive.

2.2.1 The new Insurance Law 2020

Mr. Ray Ankrah, former Chairman of the National Insurance Commission of Ghana, mentioned in a statement on 1st February 2021, that the new insurance Bill (2020) has received Presidential assent and that, the new law replaces the Insurance Act of 2006, and will ensure that the industry is well regulated in compliance with the international framework and accepted supervisory norms. Middle East Insurance Review (2021) a part of the Beacon International Group Ltd., mentioned the new law aims to boost Ghana's insurance penetration.

2.3 Insurance Penetration in Ghana

Insurance Penetration is defined as the contribution of total insurance premiums to Gross Domestic Product (GDP). In the NIC (2019) research conducted on "increasing insurance penetration in Ghana - the challenges and strategies", insurance penetration in developed countries like Luxembourg as of 2017 was 38.8% with GDP of over $70 billion. This is followed by Hong Kong/China and Ireland, both of which have insurance penetration rates of 17.9 percent and 13.6 percent, respectively. The largest coverage of insurance in Africa in 2018 was South Africa (16.99%), with Namibia (6.69%) and Lesotho (4.76%) followed by Ghana (1%). According to statistics from the Ghana Living Standards Survey (Round 6), comparing household insurance penetration to individual, household has not been as high as individual uptake. Short-term insurance plans are held by more than 60% of metropolitan families. In the survey, people with commercial or business insurance make up nearly 90% of urban families, 83.3% have property policy, and 77.4% with vehicle or motor policy. Medical policy (38.1%), funeral (25.7%), and automobile or motor vehicle (22.6%); these are the most common types of insurance plans carried by rural residents. Household members who have long-term insurance policies (more than 70%) is the same as those who have short-term insurance policies. 78.6% of households in urban areas have at least one of their members with a pension plan, 70.6% have an education policy and 74.4% have employer-provided insurance policy. Compared to 39,5% of rural areas, more households in urban areas had no policy at all at around 60,5%, (GLSS, 2014) (NIC, 2019).

2.3.1 Life Insurance and its Operations in Ghana

According to Musselman & Hughes (1950), "the growth of life insurance in a given country can be traced back to the demand created by that country's rate of social and economic progress". According to the writers, as each country's economic status improved, insurance became a more significant aspect of its culture and success. It was used during the development of Western countries in Europe, North America, and the majority of Australia. According to Pierre-Ignace et al. (2020), developing economies, particularly emerging markets in Asia, have transformed into global growth engines, accounting for more than half of global premium growth and 84 percent of individual annuity growth.

According to Poku (2009), in terms of business volume and number of companies, Ghana's insurance market was the largest in West African sub-region in the year 2005. Over the years, insurance business in Ghana operated as robust composite firms, with a few exceptions operating as separate life and non-life corporations until 2006. The Insurance Act, 2006 (274) Section 26 (2) mandated that all composite firms split up and run as distinct businesses, resulting in life and non-life units (NIC, 2006). All stakeholders in the business were prohibited from operating as composite firms. Prior to the enactment, the only life insurance firms operating separately were Gemini Life Insurance (GLICO) and Ghana Life Insurance Company. In NIC's 2005 annual report, there were eighteen (18) insurance companies two (2) reinsurance companies in the industry before the insurance

Act (2006).

Presently, there are 17 life insurance companies in good standing under the life unit (NIC, 2020). In the 2019 NIC annual end of year report, Enterprise Life Assurance was leading the sector by market premium of 26.4%, followed by SIC Life Insurance with 23.4%, Star Life (15.5%) and Glico Life (8.2%) at the end of 2019. In the NIC 2018 annual report, the top ten market leaders (Enterprise Life, Star Life, SIC Life, Glico Life, Old Mutual, Metropolitan Life, Prudential Life, Donewell Life, Mi Life, and Quality Life) in terms of total assets in 2018 contributed GHC 2.83 billion to the sector, accounting for 91 percent of the sector's total assets. In terms of premium mobilization, the top ten life market leaders mobilized GHC 1.25 billion in 2018. Enterprise Life, SIC Life, Star Life, Prudential Life, Old Mutual, Metropolitan Life, Mi Life, GN Life and Ghana Life recorded the highest premiums during the period.

2.3.2 Types of Modern Life Insurance Policies Sold in Ghana

Life insurance plays an important financial role in the life of many families in Ghana and across the world. Over time, life insurance policies have evolved from simple contracts with merely a death benefit to sophisticated contracts with a variety of features and benefits. Life insurance can be used to meet a variety of personal and family needs; the death of an insured creates an immediate estate for family members. Things have become significantly more complicated in modern times; in response to the changing economic conditions, new products are now being introduced for investment purposes rather than solely pure protection. The following are the modern life insurance products sold in Ghana as defined by Musselman & Hughes (1950), Turner & Turner (1994) and Diacon & Carter (1988) as;

- **Term life:** This is the most basic and oldest type of life insurance. If the policyholder dies within a predetermined period or term, the sum insured is paid to the policyholder's estate. If the life insured lived to the end of the policy term, the contract will terminate and no payment will be made.

- **Universal life and investment products:** The policyholder sets the premium and the amount of life insurance coverage in this policy. Premiums are flexible as long as the total amount of premiums paid is sufficient to pay the contract's specified sum insured.

- **Whole life insurance:** Whole life insurance benefits are paid out when the insured person dies, regardless of when that happens. As a result, insurance coverage is provided for the duration of the policyholder's life.

- **Group life insurance:** this form of coverage is sold to a group of people usually businesses as part of a fringe-benefit program for employees of a company, union, or association. Individual proof of insurability is normally not taking into account when underwriting this sort of policy; instead, the underwriter analyzes the group's size and turnover, as well as the company's financial soundness. Individually acquired life insurance is frequently less expensive than group life insurance.

- **Credit Life:** this is sold to cover the outstanding loan should the policyholder die prematurely

- **Funeral Insurance:** under this policy, benefits are paid to the beneficiaries of the insured when he or dies.

The figure below depicts the percentage of various life products sold in 2019. From the figure, it is seen that Group life insurance policy was the highest sold insurance policy at 81% and whole life being the least at 32.4%.

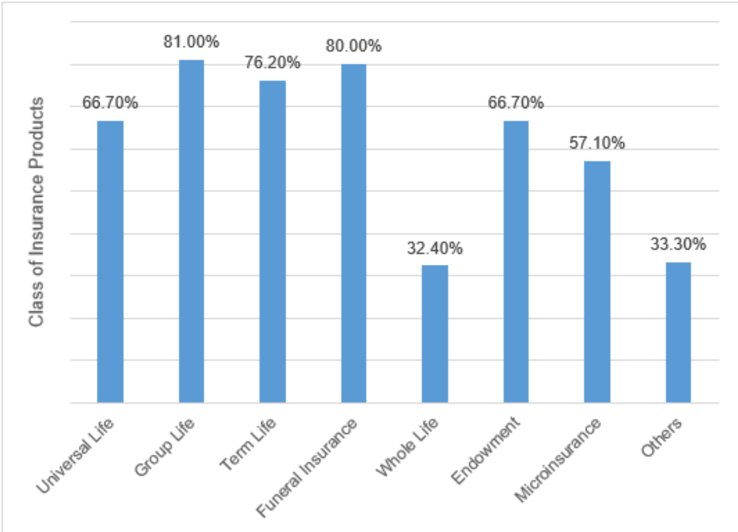

Figure 2.1: Histogram Showing Distribution of Types of Insurance
Source: NIC (2019)

2.3.3 Performance of the Life Sector in Ghana

According to Pierre-Ignace et al. (2020), globally, life insurers have struggled with growth and profitability over the past decade, and potential customers are beginning to doubt their value proposition. The performance of the sector depends heavily on the outcome of businesses insurers engage in and the effective collection of premiums. Premium generation is paramount to the success of the company and sector as a whole. The chart below shows annual premium collection of the sector over a five (5) year period. There is an increase in the year-on-year premium collection.

Table 2.1: Total Annual Life Premium over 5-year period.

Year	Life Insurance (GHS)
2019	1,651,886,133
2018	1,337,489,117
2017	1,082,083,312
2016	858,781,522
2015	705,853,360

Source: 2019 NIC Annual Report.

Group life insurance is not only the most sold life product as seen in Fig.1.1, it is also the most profitable of the products (Fig.1.2). Microinsurance on the other hand is the least profitable of all.

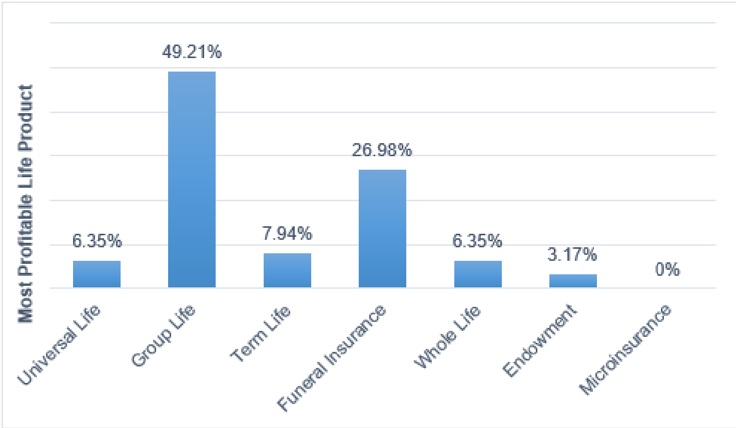

Figure 2.2: Histogram Showing the Profitability of Types of Insurance
Source: NIC (2019)

The chart below depicts the response of insurance providers in the research conducted by National Insurance Commission in 2019 to determine insurance penetration level, challenges and strategies in Ghana. It revealed 74.1% of life providers stating the insurance profitability currently is low compared to 4.6% who think otherwise. This report suggest this is a great challenge the these providers are faced with.

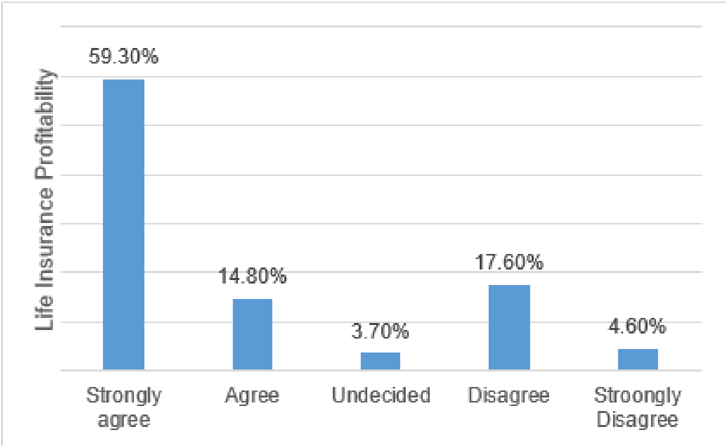

Figure 2.3: Histogram Showing the Opinions of Life Insurance Providers about the Profitability of Life Insurance
Source: NIC (2019)

2.3.4 Challenges of the Insurance Industry in Ghana – Life Unit

- According to the 2018 NIC annual report, the regulator expressed the problem of underwriting results. It expressed continued concern about the consistent failure of the insurer to remain profitable in its core business operations. According to the report, analysis of underwriting profits of life sector is not particularly helpful due to a significant majority of life insurance policies on the Ghanaian market include savings and investment components. The insurer relies heavily on the returns from the investment to compensate for the losses in their underwriting results. Moreover, because life insurers invest a large portion of the premiums they collect, fluctuation in interest rates affect returns from of the investment. As a result, over reliance on investment income to compensate for underwriting losses particularly threatens the growth and performance of the sector.

- Exchange rates: The Cedi depreciated by 12.90 percent, 15.65 percent, and 11.24 percent versus the UD Dollar, British Pound, and Euro, respectively, to conclude the year on a gloomy note. Following the policy rate drop, the Cedi had a period of weakness versus the greenback in the first quarter of 2019, resulting in large portfolio outflows. The Ghana cedi was trading at 5.5309/5.5365 per dollar by the end of 2019, compared to 4.8776/4.8224 at the end of 2018 (NIC, 2019)

- Development on the Ghana Stock Exchange: the GSE Composite Index, which measures the market's overall performance, fell by negative 12.25 percent to close at 2,257.15 points in 2019, down from 2,572.22 points in 2018. In 2019, the GSE Financial Stock Index fell by 6.23 percent to 2,019.65 points, down from 2,153.74 points in 2018. In 2019, market capitalization fell by 7.11 percent to GHS56,791.28 million, down from GHS61,136.53 million the previous year. As a result, domestic capitalization fell from GHS25,387.18 million in 2018 to GHS22,681.98 million in 2019. From GHS5,983.30 million in 2018, corporate bonds fell to GHS3,748.90 million in 2019. In comparison to the previous year, the performance of the Ghanaian stock exchange in 2019 was below expectations. The number of shares traded in the Insurance sector also decreased from 11,090,154 in 2018 to 8,788,271 in 2019.

- Another is inflation. According to Alhassan Biekpe (2015), most African countries experience higher inflation rates as a result of macroeconomic instability. Most empirical studies (Babbel (1981), Webb (2003), Hwang and Greenford (2005) have found that rising inflation reduces the value of life insurance products, which has a negative impact on insurance consumption.

- It's crucial and beneficial to have a better way to collect premiums. Payroll deduction has helped to alleviate the problem of non-payment of insurance premiums in the formal sector, although the problem remains significant with the informal sector (NIC, 2019)

2.4 Contributions of Insurance on the Economy

The insurance industry is a part of a larger economic and social context that influences market outcomes, demand-side realities, and a country's financial sector's operation. It is required to first understand the importance and evolution of the insurance market in order to appreciate the many functions of insurance in growth and development. The evolution of the insurance market is distinct in each country (World Bank, 2018). Weisbart (2018), in his paper on how insurance drives economic growth said, most individuals when asked what insurance does are likely to respond, it protects against the financial consequences of a premature death, accident, loss of property, loss of earning power, legal liability, or other unforeseen expenses. He says all of these are correct, but the industry's contribution to the economy, however, is significantly greater than what it portrays.

He went further to state the enormous contribution insurance makes to the economic growth of a country; claimants and beneficiaries reliably and rapidly reinstated-lessening the expenses of unanticipated losses and even benefiting people who may not be directly impacted by the loss, funding and promoting educational activities that save lives, protect and preserve property, contribute to the building of people's lives as well as the whole economy. Moreso, rather than having to set aside a large sum of money to cover unexpected losses, consumers and businesses can purchase insurance for a low premium, thereby investing more working capital in the economy, allowing them to produce and consume more goods and services, resulting in a higher standard of living, serve as credit facilitators for large purchases, among others.

What's more, millions of people work in insurance and related industries, who pay billions of cedis as income taxes and premium taxes, and tremendous philanthropic work being done. Some of the most important productive and growing industries including, mining, oil and gas, construction, and transportation and storage rely heavily on insurance. These industries depend on property, liability, disability, as well as engineering and other specialized insurance lines (NIC, 2016).

2.5 Empirical Frameworks

Insurance operation in Ghana has long been in existence. Yet, focus of research has not been on the demand-side. Little has been published on insurance attrition and policyholder characteristics, economic and financial factors that influence customer attrition.

An empirical study by Marin (2005), with emphasis on insurance client lifetime duration, employed survival techniques using insurance data from a Danish Insurance company consisting of 151,290 households between 1 January 1997 and 1 June 2001. The goal was to study the effects of other policy lapse rate owned by the same household for which at least one policy was lapsed. The demography, customer history and company experience of customer cancellations, and the dynamics of household portfolio are the way to better predict the probable time until the first client cancels taking into account the customer characteristics.

Using logistic regression to estimate the probability of a total cancellation with the company resulted in an overall significance of the model at a probability of 71%. Using the proportional hazard model, it was concluded that the amount of time it takes for a customer to remain with the company is totally dependent on the customer. It further demonstrated that one of the main determinants in determining the customer's lifespan is the first insurance type to be canceled. The survey therefore concluded that the factors accounting for the increased risk of all other plans are external firms, changes in address, claims and pruning.

Finally comparing the methodology employed, it was revealed that the proportional hazard regression gives a better lifetime duration, detecting customers with a high risk of cancellation within a short period of time whiles the Tobit model does within a longer period. Nevertheless, the overall performance of the proportional hazard model is preferable to the corresponding Tobit model specially taking into account potential under over-estimations of both models.

Poel and Lariviere (2003) conducted a research employing survival analysis to determine customer attrition in the European financial service companies which offerred both banking and insurance services to customers. The findings showed that attrition in general is high at the early years of becoming a customer, but stabilizes from 7 to 15 years, and then the probability of a customer to stay with the company reduces after 20 years of banking with the company.

Both Poel and Lariviere (2003), and Mittal and Kamakura (2001) agreed that age is a significant factor influencing attrition of a customer. It further stated because older people have stable preference, they are more inclined to stay with a company and less likely to end that relationship at a decreasing rate of 2.2% each year. Gender was significant as well, where men are at about 141% risk of leaving the company than their female counterparts. Poel and Lariviere (2003) supports the possible reason given by Mittal and Kamakura (2001) stating that women are more tolerant than men are. Another possible reason for this is that men are more involved in banking activities and products.

Though Mittal and Kamakura (2001) concluded that education plays an important role in customer attrition where better educated people have less rate of customer retention, both Colgate and Danaher (2000) and Dikempe and Degraeve (1997) on the other argued a no significant relationship between level of education of the holder and attrition. However, the findings by Poel and Lariviere (2003) that highly educated customers are less likely (8.2%) to end their relationship with the bank seems to be in contradiction to the findings of Mittal and Kamakura (2001). According to the authors of the paper, the possible reason accounting for this is that educated customers are able to make informed choices about the type of financial institution and product to invest in.

Further, Hasanthika and Jayesekara (2017) applied survival model to assess covariate impact over a period of time on the survival curve using life insurance data from a Sri Lanka insurance company between January 2013 and June 2016. The covariates considered in the study were age, gender, mode of payment, and policy type. Using the Cox proportional hazard model, age and gender were the covariates consistent with insurance attrition. The age coefficient of (-0.01) implied that older customer insurers were more likely than their younger holders to remain with the insurance provider. While overall gender is significant, there was nevertheless statistically no significant difference in survival at a 95% confidence interval between the two sex groups - men and women. The study concluded that in the expiration month, a major proportion of policies do not renew their policies and this scenario happens in future months.

According to Laryea (2015), motor insurance is a requirement in Ghana which offers protection for automobiles that operate on the roads. Using the Kaplan-Meier approach, a large sample of data from insurance data set with a considerable number of censored observations was used to assess the average time it takes for losses to occur and be paid by an insurance company in Ghana. The Cox proportional hazard model was used to analyze this portfolio to see if the kind of insurance impacts the time it takes to pay a claim and which variables contribute significantly to the time it takes to settle a claim. Using the log rank test and Cox proportional hazards regression model, the study revealed that age, gender, and marital status are significant risk factors that affect the occurrence of a loss but not the payment of claims. And the type of policy and the type of vehicle were important factors in determining the survival duration of settling claims. This clearly indicates that in Ghana, these risk factors are to be considered when quoting (calculating) premiums.

Tang et al. (2014) conducted a study to see how derived behavior information affects customer attrition typical of the banking sector in the United Kingdom. The purpose of the study was to examine whether derived information can be exploited by systematically analyzing customer financial behavior. The research switched its focus from directly observable data to derived data as a source of valuable estimation and prediction. While Chen et al. (2012) contends that customer behavior is dynamic and that relationships evolve over time, and as a result, an important component of the customer relation should be focused on the customer's lifetime value, Tang et al. (2014) on the other hand proposed customer attrition be identified and explained using directly observable information. There were 22 explanatory variables, 15 of which were directly observed and 7 of which were derived. The paper applied the joint probit and continuous hazard model. After the various relevant parameters had been estimated, the Nelson-Aalen estimation of the baseline cumulative hazard function was applied to predict the survival probability for each customer.

It was concluded that customers with lower survival possibilities were more likely to end their association with the firm. In all periods, the coefficients for the baseline log-hazard were statistically significant. Attrition risk begins to climb around the 12th month and continues to rise strongly until around the 18th month when it begins to decline until the 24th month. After two years, it dramatically rises again. The findings suggested that the likelihood of acquiring an investment strategy and the risk of attrition is influenced by social-demographic characteristics.

Gender was significant in the findings, with male customers experiencing a 15% higher rate of attrition than female customers. Also, because married couples are more risk averse than singles as a result of additional responsibilities, these people tend to have a longer relationship with their financial insurance company.

Customer attrition was also largely influenced by the consumer's age. When compared to the other age groups, customers aged 35 to 55 had the lowest attrition rate. Customers in the highest financial bracket (financial A) had a 17% lower attrition risk than those in the lowest financial bracket (financial C or beyond), while those in the lowest financial bracket (financial C or beyond) had a 19.2% higher attrition risk than those in the middle financial bracket (financial B). Another fascinating finding was that the rate of turnover seemed to be the same for all consumers, regardless of their geographical location in the United Kingdom. When all other factors are equal, the attrition rate for monthly payment clients was over 169% higher than for other frequency payment consumers.

Nazrina and Shahirah (2019) carried out a research in Malaysia with data from a life insurance agent from 2012 to 2015. The study used survival analysis to determine customer attrition cases and the policyholder characteristics that are likely to result in policy terminations. Kaplan Meier estimates, log rank test and Cox proportional model were used in analysing the data. The study looked at the following covariates in the dataset age, gender, marital status, policy tenure, hazard level, effective month, ineffective month, lapse status, and renewal month. According to the study, premium rates are charged according to the liability of the policy which are classified into four hazard classes. It ranged from class I (the lowest risk) to class IV (highest risk). Customers with low risk occupations fall under the hazard level I and follows suit. The findings from the research revealed that marital status of the customer is the only characteristic of the policyholder that influences the customer's attrition.

Alhassan & Biekpe (2015), using the low insurance penetration in Africa as motivation explored the factors that influence life insurance consumption in 31 African countries between 1996 and 2010, using both ordinary least squares and instrumental variables regressions. The variables of the dataset were demographic, economic and financial. It was discovered in the study that life insurance consumption in Africa is better explained by demographic factors than financial factors. While income, inflation, dependency ratio, and life expectancy were found to negatively impact life insurance consumption, financial development, health expenditure, and institutional quality were on the other hand found to positively impact life insurance consumption in Africa. According to the authors, a general rise in price levels contracts consumption life insurance and erodes the value of life insurance products. This outcome is in line with findings of Babbel (1981), Webb (2003), Hwang and Greenford (2005).

As expected by the authors, the results showed dependency ratio had a negative association with life insurance consumption, suggesting that as dependency ratio rises, life insurance consumption decreases. This result, however, contradicts Campbell (1980) and Lewis (1989) reasoning, as well as Li et al. (2007) findings on life insurance demand in OECD nations. Similarly, they concluded, a higher life expectancy ratio results in less insurance consumption. It was further revealed, while a positive relationship between education and life insurance consumption was expected, there turned out to be conflicting evidence in Africa. This result according to Alhassan & Biekpe (2015), can be attributed to Africans' low level of financial literacy, implies that improved levels of education in African countries have not been accompanied with a corresponding increase in awareness of the benefits of financial services. Their argument is based on the fact that the variable was captured by a percentage of the population enrolled in tertiary institutions, and this type of tertiary education may account for the negative effect of education on life insurance consumption in Africa, as higher education in Africa has not in essence been focused toward finance-related disciplines.

Health expenditure in juxtaposition to education, was found to be significant. Implying, a rise in health expenditure, will lead to an increase in demand for life insurance. This lends credence to the notion that social health insurance programs are not seen to be substitutes for private insurance consumption. This finding is consistent with those of Kjosevski (2012) and Li et al. (2007). Dube &Verma (2015), undertook a diagnostic and exploratory study on attrition trend and influencing factors in Telecomm and Its (BPO/KPO) industries in metros and semi metros in India, using both primary and secondary data from 50 professionals of telecomm and its related (BPO/KPO) companies in Mumbai. The study analysed the following variables, salary, family reason, working culture, community/culture bonding. It was concluded that for age group greater than 35 years, stability and social responsibility are the factors influencing customer attrition most whereas, salary, on-site opportunity and career growth are the contributing factors for age group less than 35. Das & Vijayalakshmi (2015), stated that because insurance is such a people-intensive sector, human resources play a critical role in the industry's success. Das & Vijayalakshmi (2015) mentioned, a steady increase in the rate of employee attrition is noted as a significant developing concern among Indian life insurance service providers in today's cutthroat market competition.

However, the research determined the causes of employee attrition and provided beneficial retention tactics. A market study of 85 workers of private life insurance companies was conducted in the city of Vijayawada in the state of Andhra Pradesh. The respondents were chosen using a convenience sampling method that included current employees and people who had already left the organization. Regression model was used as the statistical tool for the analysis. Its results revealed better job opportunity, target pressures, no time for family, job insecurity, relocation and bad work culture, are the reasons employees leave the insurance company.

According to Marina (2004), two approaches can be used to determine probability of insurance attrition i.e., survival analysis and binary logistic regression. According to the author, these two models and the appropriate approach for analyzing time dependent variable have been extensively discussed. One peculiar characteristic of survival analysis model is that it incorporates time-varying effects, which is great limitation of the regression model.

Adding on, Fu & Wang (2019) studied demographic, financial and macroeconomic factors affecting insurance attrition and when a policyholder will cancel (mid-term or cancellation end term non-renewal). They went on further to state the many time-varying macroeconomic variables, such as unemployment rate, GDP change, and stock market return, can have an impact on employee retention. In the case where time-varying variables are included in the model specification, Helsen and Schmittlein (1993) argue that, the time path of the explanatory variables will determine the suitable functional form. Fu & Wang (2019), outlined some of the advantages of survival analysis including the fact that it considers not only whether the policy will end but also when it will end, and also examines mid-term cancellation and end-term nonrenewal sequentially, providing a dynamic view of retention that improves on the static view derived from snapshot data, and can account for a time-varying economic factors.

Chapter 3

Methodology

3.1 Survival Analysis

In actuarial science, survival analysis approaches have wide range of applications. Any time-to-event variable would necessitate the application of such a procedure. The event is viewed in all situations as a transition from one state to another (either a good or negative occurrence) (Kleinbaum & Klein, 2012). The term ”survival time” is used in survival analysis to describe the time it takes for an event of interest to occur, while ”failure time” is used in survival analysis to indicate the occurrence of the event of interest (even if the event is a ”success” such as therapeutic recovery) (Stevenson, 2009).

The time to an occurrence of interest is the dependent (response) variable in survival models. One key element of survival data is that the response variable is a continuous non-negative random variable that represents the time between a well-defined origin and a well-defined event, (Moore, 2016). The time between the start of an individual's follow-up and the occurrence of the event of interest could be measured in days, weeks, months, or years. Time, according to Kleinbaum & Klein, (2012) can also refer to the age of the individual at the time of an event.

Randomized control trials, financial analyses, and industrial settings are all examples of situations where survival analysis has been employed. In epidemiology, it could be patient survival after surgery, or the time it takes for cows to conceive again after giving birth.

3.2 The Time to an Event

All of the functions and their numerous features serving as building blocks for modeling survival data are presented in this section. Let T be the amount of time until the event of interest occurs. This could be attrition, death, the return of a sickness, termination of an insurance coverage, or the breakdown of an equipment, among other things. T in this chapter is a continuous, nonnegative random variable from a homogeneous population. The T distribution is defined by the survival function, S(t), probability distribution function, f(t), the hazard function, h(t) and the cumulative hazard, H(t).

3.2.1 The Probability Density Function

The probability density function f(t), is the likelihood of an event occurring at time t. The cumulative density function F(t) is given as;

$$\text{F (t)} = \text{P(an individual fails before t)}$$

$$\text{F}\left(t\right) = P\left[T \leq t\right]; t \geq 0 \tag{3.1}$$

T has a density function $f(t)$ if it is a continuous random variable. The density function $f(t)$ is an unconditional probability written as;

$$f\left(t\right) = \lim_{dt \to 0} \frac{Pr\left[t \leq T < t + dt\right]}{dt} \tag{3.2}$$

If f is integrable, the distribution function or cumulative density feature is defined as the definite integral of 0 (that is, onset of risk) to time t, given as;

$$F\left(t\right) = Pr\left[T \leq t\right] = \int_0^t f\left(u\right) du. \tag{3.3}$$

The likelihood that the length of time T is less than or equal to any given value of t is expressed by the CDF, $F(t)$. The probability distribution function $f(t)$ is notable for having a nonnegative function with an area under the curve equal to one (Marin, 2005).

3.2.2 Survival Function

The survivor function determines the likelihood of living beyond a given period t. It is denoted by S(t). The survival function can be defined if T is a continuous random variable. This is given as;

S(t) = Pr(an individual survived beyond time t)

$$S\left(t\right)=\Pr\left(T>t\right) \;=\; \int_{t}^{\infty} f\left(x\right)ds = 1-F\left(t\right) \qquad (3.4)$$
$$\text{where } F\left(t\right)=Pr\left(T\leq t\right)$$

The probability density function f(t) can be expressed as a relationship with the survival function $S(t)$ as;

$$f\left(t\right)=-\frac{dS\left(t\right)}{dt} \qquad (3.5)$$

Similarly, the survival event density function can be given as

$$S\left(t\right)=S'=\frac{dS\left(t\right)}{dt}=\int_{t}^{\infty} f\left(t\right)ds=\frac{d}{dt}\left[1-F\left(t\right)\right]=-f\left(t\right) \qquad (3.6)$$

3.2.3 Hazard Function

The conditional failure rate, also referred to as the hazard function, $h(t)$, indicates the instantaneous potential for failing at time $t+1$, assuming survival up to time t. The $h(t)$, is the rate at which a randomly drawn individual known to be alive at time $(t-1)$ will die. The hazard function is concerned with an event happening, unlike the survival function which is concerned with not failing or an event happening. Mathematically, it is the probability that, assuming an individual has survived up to time t, he or she will fail in the next small-time interval, divided by the length of that interval.

The intensity function, force of mortality, death rate, or failure time are all terms used to describe the instantaneous hazard. The mortality rate at time t is a proportion of the population who fail or die between periods t and $t+1$ among the population surviving at time t, where t is commonly an integer in terms of time; e.g. day, months, years, etc. This is given as;

$$m\left(t\right) = P\left[t \leq T < t+1|T \geq t\right] \tag{3.7}$$

The hazard rate is the limit of the mortality rate within an extremely small-time interval (i.e. time interval approaches zero) divided by that time period. The hazard rate $h(t)$ is estimated as;

$$h\left(t\right) = \lim_{d(t)\to 0} \frac{P\left[t \leq T < t+dt|T \geq t\right]}{dt} \tag{3.8}$$

Therefore, if h(t) is small, we have

$$P\left[t \leq T < t+d\left(t\right)|T \geq t\right] \approx h\left(t\right)d\left(t\right) \tag{3.9}$$

From the definition of the hazard function implies

$$h\left(t\right)=\frac{\lim_{d(t)\to 0}\frac{P[t\leq T<t+dt|T\geq t]}{dt}}{P\left[T\geq t\right]}=\frac{f\left(t\right)}{S\left(t\right)}=-\frac{S'\left(t\right)}{S\left(t\right)}=-\frac{dlog\{S\left(t\right)\}}{dt} \tag{3.10}$$

From eqn. (3.10), we use the fact that S(0) = 1 integrating both sides, we obtain

$$H\left(t\right)=\int_{0}^{t}h\left(u\right)du=-\log S\left(t\right) \tag{3.11}$$

with $H(t)$ as the cumulative hazard function, we have

$$S\left(t\right)\ =\ e^{-H(t)}\ =e^{-\int_{0}^{t}h(t)dt} \tag{3.12}$$

According to Klein & Moeschberger, (2003), the hazard function is typically more informative about the underlying failure mechanism than the survival function.

3.3 Censoring

Censoring is a feature of survival analysis that occurs when the starting and ending events are not exactly observed (Moore, 2016). When just a section of the subject under research is known to have specific event times, and the remaining subject time is known to exceed the event time of interest, there is said to be censoring. Censorship happens when we have some knowledge of a subject's event time but not the exact moment. The fact that censored data is included in the analysis until it is eliminated from the risk set is one of the most appealing aspects of survival analysis. Right censoring, left censoring, and interval censoring are examples of censoring techniques.

Censorship may occur if:

- the subject does not encounter the event of interest before the study ends

- a person is lost to follow-up during the study period
- a subject withdraws from the research for reasons unrelated to the event of interest.

  Let T and C stand for the times of failure and censorship, respectively. The following are mathematical definitions and expressions for the censoring types:

Right censoring is said to occur when an individual experiences an event of interest after the specified time t. $T \in (C_r\ ,\infty)$ and the failure time T is larger than the observed censoring time Cr, but the exact value of the failure time is unknown.

Left censoring refers to a scenario in which a subject's beginning point is unknown. $T \in (0, Cl)$, and the failure time T is known to be shorter than the observed censoring time Cl, but its exact value is unknown.

Interval censoring: $T \in (Cl, Cr)$, the failure time T is known to be less than the observed right censoring time Cr and greater than the observed left censoring time Cl, but its exact value is unknown (Klein, 1997).

Random or non-informative censoring occurs when a subject's censoring time is statistically independent of their failure time. Subjects whose failure time is greater than their censoring time are right-censored, and the observed value is the minimum of the censoring and failure durations.

In addition to censoring, there is a characteristic known as truncation, which may be present in some time-to-event investigations. The causal impact and potential-outcome approach used in this thesis does not consider truncated data. The type of censoring type employed in this thesis is the independent or right censoring.

3.3.1 The Censoring Mechanisms

There are various distinct forms of censoring mechanisms, each of which results in a different likelihood function for inference, Cox and Oakes (1984). These are outlined below:

Type I Censoring: in Type I censoring, an event is only recorded if it occurs before a certain time. The censoring times may differ from one person to another. The investigators may end the trial or announce the results before all subjects experience the event due to cost or time constraints. If no subject withdrawals or accidental losses occur, censored observations have times equal to the length of the research period; the censored time for each individual is the same and can be treated as a fixed time for a particular trial.

Type II Censoring: In Type II censoring, the investigation continues until the pre-specified event of interest is experienced by a specified number of people. This form of censoring is frequently used in experiments to determine the time it takes for equipment to malfunction. In this instance, each individual's censored time may differ and can be regarded as a random variable. Type II censoring, however, is uncommon in human clinical studies.

Random censorship: random censoring involves a concept known as competing risks. Individuals may be withdrawn from the study as a result of other conflicting occurrences, than the primary event of interest.

3.4 Estimation of Survival Functions

3.4.1 Kaplan-Meier Estimate

The Kaplan-Meier (K-M) method is one of the best methods for determining the proportion of participants who live beyond a specific study period (censored data). It is a survival function estimator that is based on lifetime data. The Kaplan-Meier estimator, often known as the product-limit estimator, is the simplest method of determining survival over time (Kaplan & Meier, 1958). The K-M is a non-parametric statistic of S(t). It considers any point in time as a series of steps defined by observed survival and censored times, incorporating information from all accessible observations, both censored and uncensored.

The K-M is based on individual survival times. The K-M model makes the key assumption that censoring is independent of survival time (i.e., the reason an observation is censored is unrelated to the cause of failure). We also assume that the probability of survival are the same for participants who are recruited early and late. The K-M estimate is used in insurance to calculate the time it takes to file a claim or to attrition, and in medical research to calculate the proportion of patients who live after a treatment. In engineering, is the time at which a piece of equipment fails. In the job environment, it measures after a job loss the length of time for an individual to remain unemployed.

The Kaplan-Meier technique has the advantage of taking into account right censored data, which might occur when a patient withdraws from the trial, is lost to follow-up, or when the event of interest does not occur. Small vertical tick marks on the plot show losses, indicating that an insured survival time has been right-censored. The Kaplan-Meier curve complements the empirical distribution function when there is no censoring or truncation. Suppose $t_1 \leq \ t_2 \leq, \ldots, t_k$ are the ordered failure times.

For $t_k \leq t \leq t_{(k+1)}$, the probability of surviving beyond time t is;

$$S_{KM}(t) = P(T > t) = P\left(T \geq t_{(k+1)}\right) \tag{3.13}$$

$$S_{KM}(t) = P(T \geq t_1, T \geq t_2, \ldots, T \geq t_k) \tag{3.14}$$

$$S_{KM}(t) = P(T > t_1) \prod_{j=1}^{k} P(T \geq t_{j+1} \mid T > t_j) \tag{3.15}$$

but,

$$P(T > t) = S(0) = 1,$$

$$S_{KM}(t) = \prod_{j=1}^{k} P(T \geq t_{j+1} \mid T > t_j)$$

$$S_{KM}(t) = \prod_{j=1}^{k} [1 - P(T = t_j) | T > t_j] \tag{3.16}$$

Therefore, the Kaplan-Meier estimator of the survival function of S(t) is given as;

$$\hat{S}_{KM}(t) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{r_j}\right) \text{ , } for\ 0 \leq t \leq t^{+} \tag{3.17}$$

where,

$\hat{S}_{KM}(t) =$ Kaplan-Meier estimator of the survival at t

$d_j =$ Number of attrited policies at time t_j

$r_j =$ Number of individual policies still in force just before time t_j, including the risk set at t_j

t_j, j = 1,2, 3, ..., n, is the total set of attrited policies times recorded with t^{+} as the maximum attrition time.

3.4.2 Variance of the Kaplan-Meier estimator (Greenwood formula)

The Kaplan-Meier estimator is a statistic that employs a number of estimators to approximate its variance. The Greenwood formula is a popular use of such estimators. For $t_k \leq t \leq t_{(k+1)}$,

The risk set $r_j = r_i : r_i \geq r_j$, the number of failures makes it a binomial experiment (that is, either policy is terminated or not)

Thus $d_j \sim Binomial\,(r_j, \leftthreetimes_j)$, $\leftthreetimes_j$ = the hazard at t_j.

Let

$$q_j \;=\; 1 \;-\; \leftthreetimes_j \;\; for\; t_k \leq t \;\leq\; t_{(k+1)} \tag{3.18}$$

$$\hat{S}\,(t) = (1 - \leftthreetimes_1)\,(1 - \leftthreetimes_2)\ldots\left(1 - \leftthreetimes_{(j+1)}\right) \tag{3.19}$$

This implies;

$$\begin{aligned} \hat{S}\,(t) &= q_1 q_2 \ldots q_{(j+1)} \\ \log \hat{S}\,(t) &= \log\left(q_1 q_2 \ldots q_{(j+1)}\right) \\ Var(\log \hat{S}(t)) &= \; Var(\sum_{i=1}^{j} \log q_j) = \sum_{i=1}^{j} Var(\log q_j) \end{aligned} \tag{3.20}$$

By statistical theory, we can treat $\log \hat{q}_j$ as uncorrelated terms
Using Delta method; thus

$$Var\;(\log \hat{S}\,(t)) = \;[\frac{d}{dt}\log \hat{S}(t)]^2 Var\hat{S}(t) = \;[\frac{1}{\hat{S}(t)}]^2 Var\hat{S}(t) \tag{3.21}$$

$$Var\left(\log \hat{q}_t\right) = \left[\frac{1}{q_j}\right]^2 Var\left(\hat{S}\left(t\right)\right) \tag{3.22}$$

$$\begin{aligned} Var\left(\hat{S}\left(t\right)\right) &= \frac{\hat{S}\left(t\right)\left(1-\hat{S}\left(t\right)\right)}{r_j} = \frac{\hat{q}_j\left(1-\hat{q}_j\right)}{r_j} && (3.23)\\ Var(\hat{q}_j) &= [\frac{1}{q_j}]^2 q_j \lambda_j \\ &= \frac{\hat{\lambda}_j}{\hat{q} r_j} \\ \text{But } Var(\log \hat{S}(t)) &= \ Var(\sum_{i=1}^{j} \log q_j) && (3.24)\\ Var\left(\log \hat{S}\left(t\right)\right) &= \sum_{i=1}^{j} \frac{\hat{\lambda}_j}{\hat{q}_j r_j} && (3.25) \end{aligned}$$

Again, using the Delta method,

$$\begin{aligned} Var\left(\hat{S}\left(t\right)\right) &= Var(\exp(\log \hat{S}(t)) \\ &\approx [\frac{d}{\log \hat{S}(t)}]^2 Var(\log \hat{S}(t)) && (3.26)\\ Var\left(\hat{S}\left(t\right)\right) &= \left[\exp(\log \hat{S}(t))\right]^2 \ Var(\log \hat{S}(t)) \\ &= [\hat{S}(t)]^2 Var(\log \hat{S}(t)) && (3.27) \end{aligned}$$

From (3.26)

$$\begin{aligned} VarS(t) &= [\hat{S}(t)]^2 \sum_{i=j}^{j} \frac{\lambda_j}{q_j r_j} \\ VarS\left(t\right) &= [\hat{S}(t)]^2 \left[\sum_{i=j}^{j} \frac{d_j}{r_j}[\frac{1}{1-\frac{d_j}{r_j}}]\right] \end{aligned}$$

Hence, the Greenwood Formula is;

$$Var\left(\hat{S}\left(t\right)\right) = [\hat{S}(t)]^2 \left[\sum_{r_j \le t}^{j} [\frac{d_j}{r_j(r_j-d_j)}]\right] \tag{3.28}$$

$$Se\left(\hat{S}\left(t\right)\right) = \hat{S}\left(t\right)\sqrt{\sum_{r_j \leq\ t}^{j} \frac{d_j}{r_j\left(r_j - d_j\right)}} \tag{3.29}$$

A (1-α)% Confidence Interval for the Kaplan-Meier estimate is;

$$\hat{S}\left(t\right)\ \pm\ Z_{1-\frac{\alpha}{2}}\hat{S}\left(t\right)\sqrt{\sum_{r_j \leq\ t}^{j} \frac{d_j}{r_j\left(r_j - d_j\right)}} \tag{3.30}$$

3.5 Estimating the Median Survival Time to Attrition with Confidence Bounds

The average survival time is the expected value of the survival time given as;

$$\mu = E\left(T\right) = \int_0^\infty tf\left(t\right)dt \tag{3.31}$$

using integration by parts, and the fact that

$$f\left(t\right) = -\frac{d\left\{S\left(t\right)\right\}}{dt} \tag{3.32}$$

$$\mu\ =\ \int_0^\infty S(t)dt \tag{3.33}$$

The mean survival time can only be defined when

$$S(\infty)\ =\ 0 \tag{3.34}$$

This cannot hold in this case since it implies all policies are attrited at the end of the study period. The mean is not a suitable summary because survival times are not expected to be normally distributed. The appropriate summary to use is the median. The average survival time, which is quantified using the median, is another number that is frequently of importance in a survival analysis. The median survival duration for a period is determined by finding 0.5 on the Y-axis and reading down to the x-axis. This interval is terminated when an event occurs. The non-continuous nature of the K-M function emphasizes that they are not smooth but rather stepwise estimates. The cumulative probability of surviving up to a given time is read on the Y-axis. The sites at which this horizontal line crosses across the pointwise confidence intervals of $\hat{S}(t)$ define the confidence interval for $t_{0.5}$.

3.5.1 Comparing Survival Times of Two Different Variables

In statistics, often the interest is to test and compare if there are any significant differences in survival between and among different groups, individuals, or research participants. Typically, in an observational study, we are interested in comparing survival difference between the two groups e.g. the two genders, participants who are high risk and those who are low risk, or among the ages of participants. The interest is comparing the null hypothesis that the survival times of two covariates are equal to the alternative hypothesis that they are not (for two variables). The null hypothesis is rejected if the test statistic exceeds a certain constant. When the null hypothesis is true, the significance level of the test is the chance of rejecting it. This thesis will employ log rank test to compare survival between covariates.

3.6 The Log Rank Test

The log rank test, a statistical hypothesis, can be used to compare the survival times of different groups. It is used to test the null hypothesis that there is no difference in the survival curves of the populations (i.e. the probability of an event occurring at any time point is the same for each population). The test statistic is compared to the chi-square test statistic, which compares the number of observed events to the expected number at each time point during the follow-up period using the R statistical package. The test statistic is calculated as:

$$x^2{}_{log\ rank} = \frac{\left(\sum_{j=1}^{j} O_{jt} - \sum_{j=1}^{j} E_{jj}\right)^2}{\sum_{j=1}^{j} E_{jt}} \tag{3.35}$$

where;

O_{jt} is the total sum of observed events in the jth group over time, i.e., 1,2

E_{jt} is the total numbers of expected event for jth group

The sum of the expected number of events at the time of each event is the overall expected number of events for a group. The number of expected events at the time of an event can be determined by multiplying the probability of death at that moment by the number of people alive in the group. The risk of mortality (number of deaths/number of lives) can be computed from the combined data for both groups under the null hypothesis. The degrees of freedom for the log rank statistic is $k-1$, where k is the number of groups being compared. The log rank has the degree of freedom as 1.

3.6.1 Hypothesis

The hypothesis for comparing survival between variables is given as follows; H_0= there is no difference between survival groups.

H_1= there is difference between survival groups.

The level of significance, also known as the significance level, is a criterion by which a decision is made about the value provided in a null hypothesis. The criterion is based on the likelihood of obtaining a statistic measured if the null hypothesis value were true. The criteria or level of significance in behavioral research is usually set at 5%. We reject the null hypothesis and accept the alternative when the p-value is less than 5%.

3.7 Cox Regression Model

The Cox regression model is one of the models used in survival analysis to determine which combination of potential explanatory variables influence the shape of the hazard function. It computes hazard function estimates in a specific study. It is one of the most widely used regression techniques in predicting survival outcomes, Cox (1972). The Cox regression model is also known as the proportional hazard (PH) model. Explanatory covariates on survival times could be used in the model. The Cox regression (proportional hazard regression) method can be used to investigate the effects of many risk factors on survival. The hazard estimates the likelihood that the end point of the event of interest will occur. Though probability ranges from 0 to 1, the risk in a group can be greater than 1.

The Cox PH model, because it makes no assumptions regarding the nature or shape of the hazard function is referred to as a semi-parametric model. However, when using the Cox regression model there are a few key assumptions to consider:

- Survival times of subjects in the population are independent of one another.
- Regardless of time, changes in predictors result in proportionate changes in hazard.
- A hazard ratio remains constant over time. For analyzing censored data, the proportional hazards model is the most often used strategy., Van Den Poel et al. (2004).

The proportional hazard model is given as,

$$h\left(t|\ x_i\right) = h_0\left(t\right) e^{\beta'\ x_i} \tag{3.36}$$

dividing both sides of equations 3.33 by $h_0\left(t\right)$ and taking the natural logarithm of both sides, we obtain;

$$log\frac{h\left(t|x_i\right)}{h_0\left(t\right)} \ = \ \beta' x_i \tag{3.37}$$

where h_0 is the baseline hazard, also referred to as the hazard function where there are no covariates impacts.

$$x_i = \ \left(x_{1i}, x_{2i}, x_{3i,}, \ldots, x_{ki}\right) and \tag{3.38}$$

$$\beta = \left(\beta_1, \beta_2,\ \beta_3, \ldots, \beta_k\right) \tag{3.39}$$

k =number of covariates in the model.

The baseline hazard $h_0\left(t\right)$ is unspecified thus; non-parametric while β a parametric component hence, making the Cox proportional hazard a semi-parametric model. The corresponding estimates of the β's are called maximum likelihood (ML) estimates. In this study, the cox model will be employed to determine the hazard ratios without estimating the baseline hazard function. Estimating model coefficients is done with the partial likelihood function. Because the formula only considers probabilities for subjects who fail, the likelihood function is called "partial" likelihood function.

The partial likelihood estimate of β can be derived by maximizing this product;

$$\prod_{i=1}^{n}\left[\frac{e^{\beta'\chi_{i,t}}}{\sum_{j\in R_i} e^{\beta'\chi_{j,t}}}\right] \tag{3.40}$$

over observed n distinct ordered survival times. The dependent variable is the time it takes for a policy to attrite. Age, gender, marital status, product type, base rate change, pay frequency, mode of payment, policy tenure, premium, and employment sector are the independent variables (covariates).

3.7.1 Cox Proportionality Assumption

For the application of the Cox PH, the assumption of proportionality is particularly important. Although the proportional hazard assumption is unlikely to be fully satisfied in practice, significant violations of the PH assumption can lead to estimates that are erroneous and misleading (Von Dijk et. al., 2008).

The proportionality assumption is a crucial assumption for the proper application of the log rank test and the Cox proportional hazards regression model. It means that an event's hazards are proportionate, or the beta remains constant over time. We assume, in particular, that the hazards are proportionate over time, implying that the influence of a risk factor remains constant over time (the ratio of two hazards remains constant throughout time).

There are numerous approaches available for assessing the PH assumption. some of which rely on statistical testing and others on graphical(visual) evaluations. The assumptions of the Cox model and changes in predictors produce proportional changes in the hazard regardless of time. There is independence, and a linear association between the natural logarithm of the relative hazard and the predictors.

Using the statistical testing approach, predictor by time interaction effects are incorporated in the model and the statistical significance verified. The premise of proportionality is broken whenever one (or more) of the predictors by time interactions reaches statistical significance (e.g., $p<0.05$). Graphical analysis is an alternate method for determining proportionality. To determine whether the proportional hazards assumption is reasonable, there are numerous graphical displays that can be employed (Lemeshow and Hosmer, 1999).

In this thesis, we will employ statistical test and graphical diagnostic based on the scaled Schoenfeld residual to check for the proportional hazards (PH) assumption.

3.7.1.1 Statistical and Graphical Diagnostic Tests of Proportionality using Scaled Schoenfeld Residual

The Schoenfeld residuals are in principle, time-independent. A plot against time that exhibits a non-random pattern is evidence of PH violation. A non-significant link between residuals and time supports the proportional hazard assumption, whereas a substantial association between residuals and time refutes it.

3.8 The Stratified Cox Model

In some cases, the proportional risks assumption for a covariate(s) is violated. In such circumstances, there is the need for modification of the cox model. This can be done by stratification. The proportional model allows for stratifying on the variable(s) that voilates the assumption. This means taking the covariates voilating the assumption from the model and adding those meeting the assumption. This is possible with the proportional hazards model applied to the other covariates within each stratum. The effect of other explanatory factors on the hazard function in the jth stratum with an arbitrary baseline hazard function $h_{0j}(t\ )$ can be described by a proportional hazards model in that stratum as;

$$h_j[t|Z(t)] = h_{0j}(t)exp[\beta^t Z(t)], j = 1, 2, ..., s \qquad (3.41)$$

Although in this model the baseline hazard functions may be different and completely unrelated, the regression coefficients are considered to be the same in each strata. The partial likelihood function is given by

$$LL(\beta) = LL_1(\beta) + LL_2(\beta) + ... + LL_s(\beta) \tag{3.42}$$

where $LL_j(\beta)$ is the log partial likelihood using only the data for those individuals in the jth stratum. The derivatives for the log likelihood are derived by summing the derivatives across each stratum. The $LL_j(\beta))$ is, then, maximized with respect to β.

The modified cox proportional hazard model (stratified cox model) can be written as;

$$h_g(t, X) \;=\; h_{og}(t)exp[\beta_1 X_1 \;+\; \beta_2 X_2 \;+\; ... \;+\; \beta_k X_k] \tag{3.43}$$

where,

$$g \;=\; 1,\; 2, ..., k, \tag{3.44}$$

Chapter 4

Analysis, Results and Discussions

4.1 Overview of Data

This chapter presents the analysis of the various results obtained from the study. RStudio software was used to obtain all the estimated parameters and results reported in the study. The section also gives a descriptive overview of the entire data used for the analysis, this will include a summary and visualization of the data which will give more meaning and allow for simpler interpretation of the entire data. Histograms, and ratios were used for the descriptive analysis.

The analysis is categorized into preliminary and further analysis. the data used for the analysis was obtained from one of the life insurance companies in Ghana. Of the total dataset of 21007, 11634 (55.4%) of the data points were taken out because these data points are neither left censored or left truncated. These 11634 data points were not included in the further analysis because these policies have no payments made, are not cancelled and hence not in force. And therefore, 9204 data points will be used for the analysis. Under the policy status, inforce are active policies, not inforce are policies which are neither cancelled nor have payments made, cancelled policies are those that have been cancelled by the holder, while attrited(attrition) policies are those that are cancelled as a result of discontinuous payment of premiums.

The covariates used in this thesis are age, gender, marital status, product, employment sector, policy tenure, premium, base rate change, source type, and pay frequency with policy status as the dependent variable.

4.2 Preliminary Analysis

4.2.1 Characteristic Distribution of Continuous Variables

From Figure 4.1 below shows 18 and 62 are minimum and maximum ages respectively with 34.6 years as the average age. The data is moderately right skewed with a heavy right tail. Also, a kurtosis of 0.004 implying a broader peak and a thickened tail of the distribution as seen clearly in Figure 4.1 below; hence heavy tailed. The minimum and maximum premium is seen to be GHS4.20 and GHS5,590.00. A kurtosis of 8.8 indicates heavy tails.

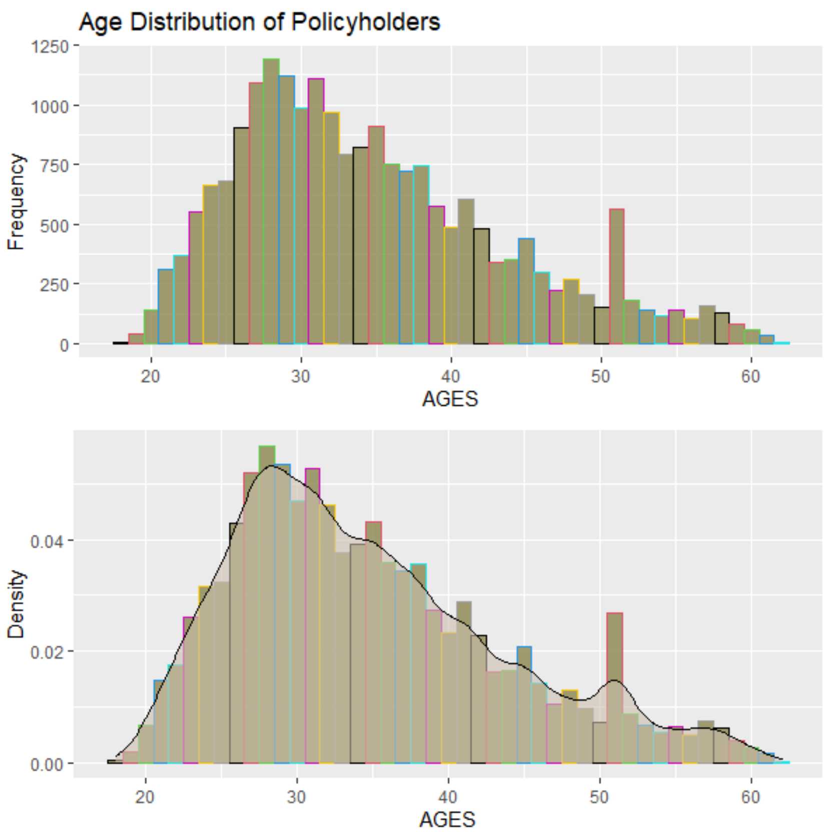

Figure 4.1: Histogram of Age Distribution of Policyholders

Table 4.1: Age and Premium Distribution of Policy Holders

Var.	Mean	Med.	Var.	Std.Dev	Skew.	Kurt.	Min.	Max.
Age	34.6	33.0	79.7	8.9	0.8	0.004	18.0	62.0
Premium	138.2	100.0	19489.2	139.6	8.8	185.5	4.2	5590.0

4.2.2 Characteristic of Categorical Variables

4.2.2.1 Status of Insurance Policies

From Figure 4.2, it can be seen that out of the total of 9204 policies for the period, 706 were cancelled by the holder representing 7.7%, 2487 attrited and therefore cancelled by the insurer representing 27%. There are a total of 6011 policies in force representing 65.3% of total policies.

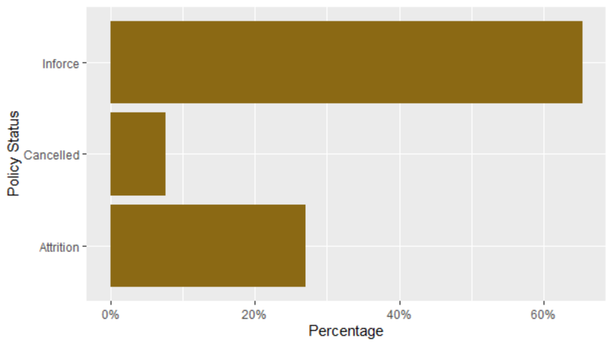

Figure 4.2: Histogram Showing Status of Policies

4.2.3 Estimating Correlation Between Premium and Policy Tenure

We see from Table 4.2 that there is a negative correlation between premium and tenure of policy. However, this does not imply causality.

Table 4.2: Correlation Between Premium and Policy Tenure

	PREMIUM	POLICY.TENURE
PREMIUM	-0.03	1.00
POLICY.TENURE	1.00	-0.03

4.3 Ratio Estimation for Covariates

4.3.1 Ratio Estimating for Product

Table 4.3: Estimated Ratios for Product Type

PRODUCT	POLICY STATUS	N	RATIO
EDUCATION	INFORCE	1919	0.208
EDUCATION	ATTRITION	1142	0.124
FUNERAL	INFORCE	1784	0.194
FUNERAL	ATTRITION	944	0.103
SAVING	INFORCE	1964	0.213
SAVING	ATTRITION	1052	0.114
TERM	INFORCE	345	0.037
TERM	ATTRITION	54	0.006

From Table 4.3, we see education having attrition ratio of 0.124 followed by savings 0.114 and term 0.006 as the lowest.

4.3.2 Estimated Ratios for Marital Status

From Table 4.4, the ratio (0.183) of attrition for singles is higher compared to the married of 0.159.

Table 4.4: Estimated Ratios for Marital Status

MARITAL STATUS	POLICY STATUS	N	RATIO
Divorced	INFORCE	37	0.004
Divorced	ATTRITION	16	0.002
Married	INFORCE	2720	0.296
Married	ATTRITION	1462	0.159
Separated	INFORCE	48	0.005
Separated	ATTRITION	11	0.001
Single	INFORCE	3160	0.343
Single	ATTRITION	1690	0.183
Widowed	INFORCE	47	0.005
Widowed	ATTRITION	13	0.001

4.3.3 Estimated Ratios for Source Type

From Table 4.5, mobile money mode of payment has a greater attrition ratio of 0.165 compared to direct debit of 0.132. Employer deduction is least compared to direct debit and mobile money.

Table 4.5: Ratio Estimates for Source Type

SOURCE TYPE	POLICY STATUS	N	Ratio
Cheque	INFORCE	72	0.008
Cheque	ATTRITION	56	0.006
Direct Debit	INFORCE	2193	0.238
Direct Debit	ATTRITION	1215	0.132
Employer Deduction	INFORCE	840	0.091
Employer Deduction	ATTRITION	406	0.044
Mobile Money	INFORCE	2907	0.316
Mobile Money	ATTRITION	1515	0.165

4.3.4 Estimated Ratios for Age

Table 4.6: Ratio Estimates for Age

AGE CATE.	POLICY STATUS	N	RATIO
18-29	INFORCE	2044	0.222
18-29	ATTRITION	1107	0.12
30-45	INFORCE	3172	0.345
30-45	ATTRITION	1674	0.182
46-59	INFORCE	770	0.084
46-59	ATTRITION	391	0.042
60+	INFORCE	26	0.003
60+	ATTRITION	20	0.002

From Table 4.6, the ratio of the middle-aged of 30-45 is seen to be higher than the younger of 18-29 and the older holders of 60+. From the computed ratios, among these three categories, the older holders have the lowest ratio of attrition and the middle-aged the highest.

4.4 Further Analysis

The section employed an inferential overview of the Kaplan-Meier, log rank test, and Cox proportional model on the data to estimate the survival time and key covariates that influence an insured's attrition in the life insurance sector.

4.4.1 Estimating the Survival Time Using Kaplan-Meier (Product-Limit) Approach on Life Insurance Data

4.4.1.1 Log-Rank Test of Significance Between Attrited and Inforce Policies

Table 4.7: Test of Equality of survival difference Between the Two Levels of Policy Status

Policies	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Inforce	6176	0	2437	2437	13409
Attrited	3028	3028	591	10044	13409

Chi-sq = 13409 on 1 degrees of freedom, p= <e-16

Table 4.7 shows the log-rank test of statistical difference between the survival time for a policy to stay in force or be attrited. We see from the table a p-value of <2e-16 and a χ^2 of 13409 on a degree of freedom of 1. The p-value obtained implies a high statistical significance at a 5% level. As a result, we reject H_0 and conclude that, there is a significant statistical difference between the survival time for a policy to stay in force or be attrited.

4.4.1.2 Log-Rank Test of Significance Difference of Customer Gender

Table 4.8: Log-rank Test of Statistical Difference for Customer Gender

Variable	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Female	3309	1065	1101	1.201	1.92
Attrited	5895	1963	1927	0.686	1.92

Chi-sq = 1.9 on 1 degrees of freedom, p= 0.2

From Table 4.8, with a p-value of 0.2, we fail to reject H_0 and conclude that there is no significant difference between the time it takes for a males or female policy holder to attrite a policy.

4.5 Estimating Median Survival Time to Policy Cancellation

Table 4.9: Estimated Time to Policy Cancellation

N	Events	Median	0.95LCL	0.95UCL
705	705	16	15	17

From Table 4.9 here, we see that the median time for a policy to lapse(attrition) is 16 weeks, and a 95% confidence interval of range 15 to 17 weeks. The median and its associated 95% confidence interval is represented in 4.5.1.

4.5.1 Kaplan-Meier Survival Curve Showing Time to Policy Cancellation

Figure Table 4.3 estimates the median survival time. The median survival time is traced on the red line. It is estimated to be 15 weeks. The 95% confidence intervals of 15 and 17 weeks was obtained and it is traced on the horizontal at 0.5.

4.6 Estimating Median Survival Time to Attrition

From Table 4.10 here, we see that the median time of attrition is 15 weeks, and a 95% confidence interval of range 14 and 15 weeks. The median and its associated 95% confidence interval is represented in Figure 4.3.

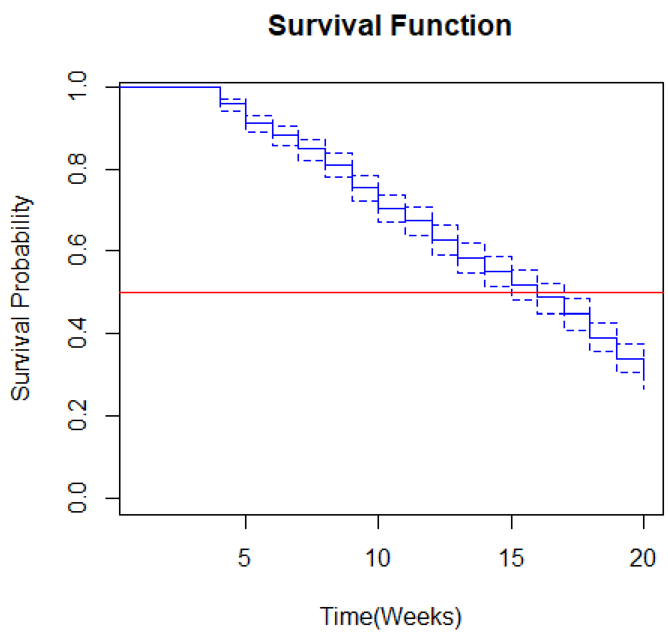

Figure 4.3: Kaplan-Meier Curve showing Time to Policy Cancellation

Table 4.10: Estimated Time to Attrition

N	Events	Median	0.95LCL	0.95UCL
2375	2375	15	14	15

4.6.1 Kaplan-Meier Curve Showing Median Survival Time

From Figure 4.4 estimates the median survival time. The median survival time to attrition traced on the red line is 15 weeks. The 95% confidence intervals can also be traced at 14 and 15 weeks intersecting at 0.5 on the horizontal.

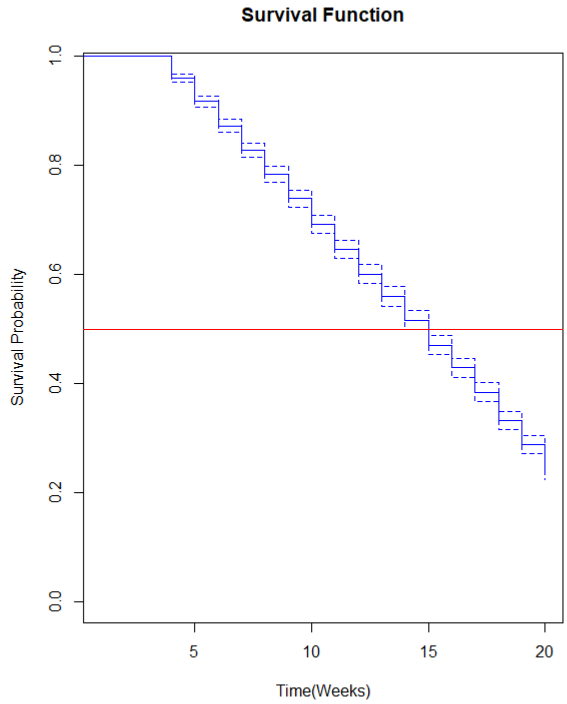

Figure 4.4: Kaplan-Meier Curve showing Time to Policy Lapse

4.7 Comparing Survival Times of Variables Using Kapaln-Meier Curves and Log Rank Test

4.7.1 Log Rank Test of Survival Difference between Gender and Age

From Table 4.11, using log rank test to compare survivorship of age against gender gives a p=0.05. And therefore, we fail to reject H_0 and conclude there is no statistical difference between survival curves.

Table 4.11: Test of Statistical Difference of Age Against Gender

Policies	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Female, 18-29	1309	437	429.87	0.1183	0.140
Female, 30-45	1561	494	523.01	1.6096	1.982
Female, 46-59	412	122	140.05	2.3259	2.484
Female, 60+	27	12	8.44	1.5062	1.538
Male, 18-29	1842	657	591.24	7.3143	9.258
Male, 30-45	3285	1057	1083.47	0.6469	1.026
Male, 46-59	749	244	245.56	0.0099	0.011
Male, 60+	19	5	6.36	0.2908	0.297

Chi-sq = 14.1 on 7 degrees of freedom, p=0.05

4.7.2 Kaplan-Meier Survival Curve of Age against Gender

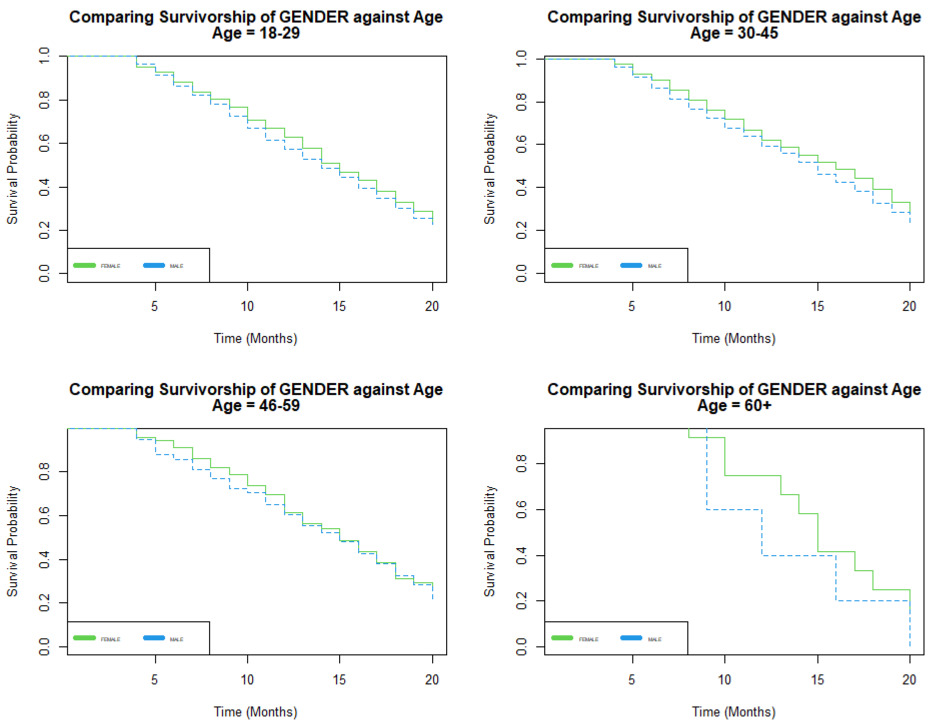

Figure 4.5: Survival Difference of Age against Gender

4.7.3 Kaplan-Meier Survival Curve of Gender against Source Type

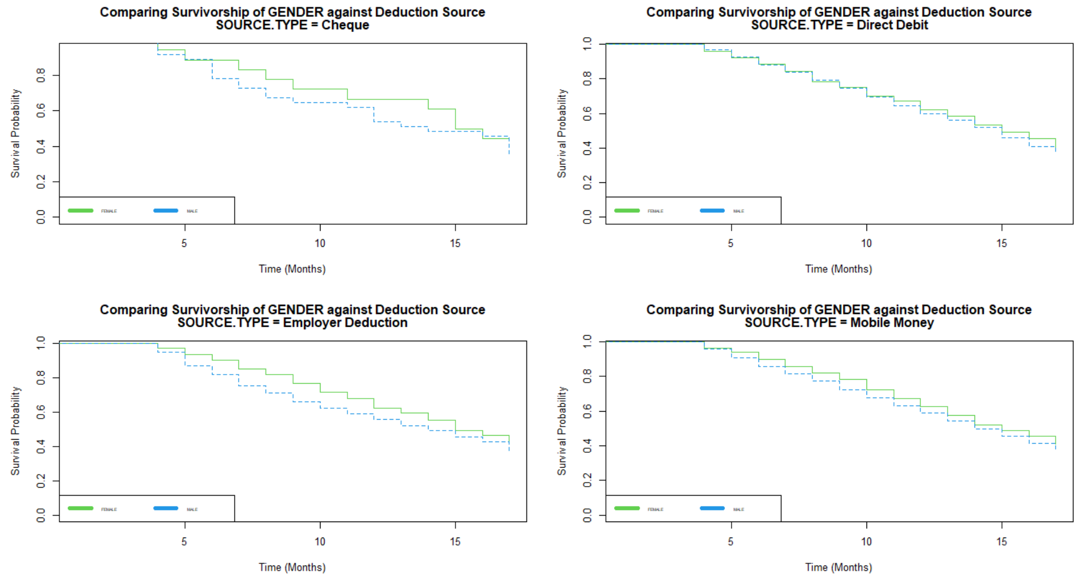

Figure 4.6: Survival Curve of Gender against Source of Deduction

4.7.3.1 Log Rank Test of Survival Difference between Gender and Source Type

Table 4.12: Test of Statistical Difference of Source Type Against Gender

Policies	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Female, Cheque	36	18	10.3	5.854	5.981
Female, Direct Debit	1118	391	365.4	1.792	2.076
Female, Employer Deduction	580	188	193.2	0.138	0.150
Female, Mobile Money	1575	468	532.5	7.824	9.671
Male, Cheque	92	37	28.5	2.555	2.627
Male, Direct Debit	2290	762	751.8	0.137	0.186
Male, Employer Deduction	666	198	219.4	2.091	2.297
Male, Mobile Money	2847	966	926.9	1.650	2.422
Chi-sq = 22.5 on 7 degrees of freedom, p=0.002					

From Table 4.12 using log rank test to compare survivorship of gender against source type gives a p=0.002 <0.05. Therefore, we reject H_0 and conclude there is statistical difference between survival curves.

4.8 Cumulative Hazard Estimation

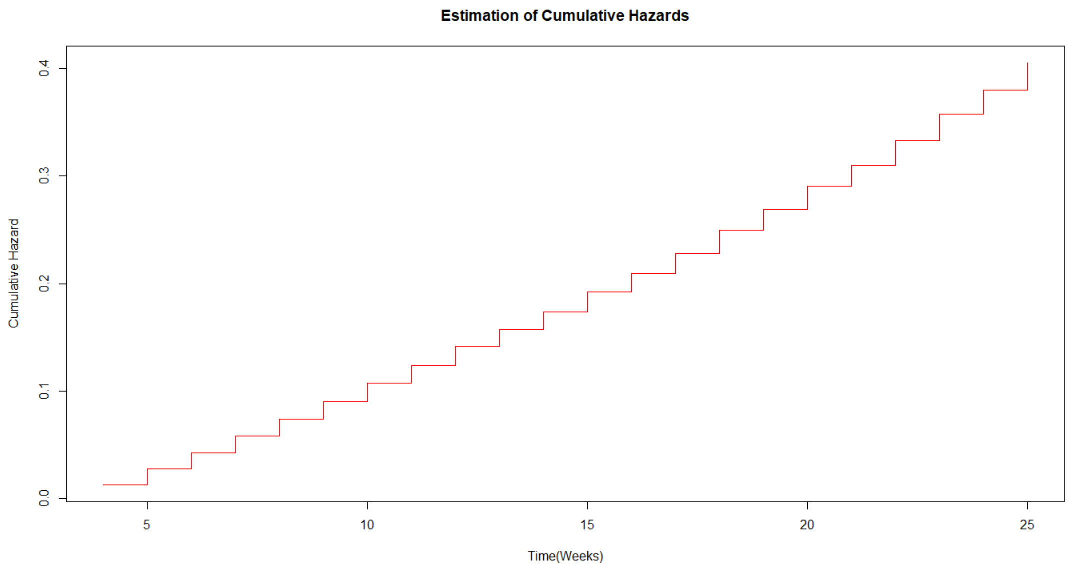

Figure 4.7: Cumulative Hazard

Figure 4.8 shows an increasing cumulative hazard function.

4.9 Cox Proportional Hazard Model - Full Model

Table 4.13: Analysis of Deviance

Variable	Log-lik	Chi-Sq	Df	Pr(>|Chi|)
NULL	-27061			
PRODUCT	-27019	84.32	3	<2.2e-16
GENDER	-27018	2.13	1	0.1441
MARITAL SCAN	-27008	18.36	4	0.0010
AGE CATEGORY	-27007	2.20	3	0.5316
SOURCE TYPE	-27002	9.86	3	0.0199
PREMIUM	-27002	0.04	1	0.8422
PAY FREQUENCY	-27001	2.98	3	0.3947
BASE RATE CHANGE	-26970	61.67	1	4.07e-15
EMPLOYMENT SECTOR	-26958	23.26	20	0.2762

Table 4.13 shows the output of the full cox model with their p-values.

4.10 Statistical and Diagnostic Tests for Proportionality

4.10.1 Schoenfeld Graphical Test for Proportionality

From 4.10.1, for the proportional hazards assumption to be correct or met, a plot of the residuals versus the covariate will yield a pattern of points that are centered at zero. From Figure 4.8, we see customer gender and pay frequency to voilate the principle of proportionality.

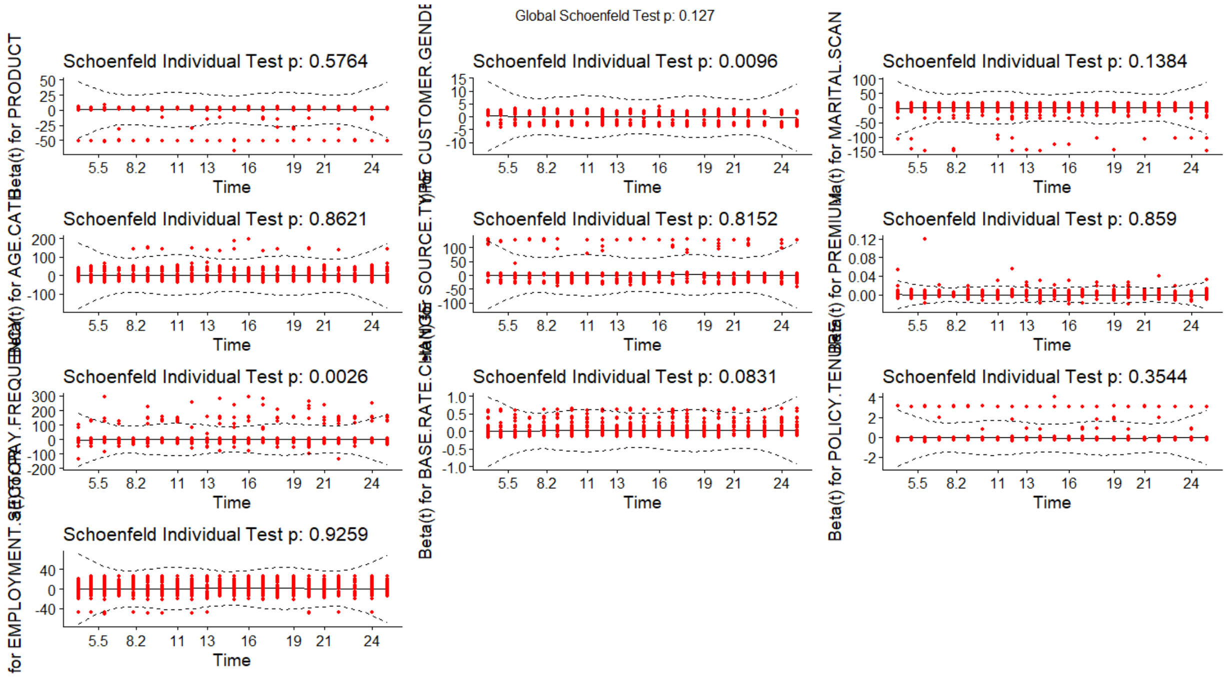

Figure 4.8: Schoenfeld Residuals Testing Cox Assumption

4.10.2 Statisitical Test the Scaled Schoenfeld Residuals

Table 4.14: Residual Estimates

Variable	Chi-Sq	Df	P-Value
PRODUCT	1.98	3	0.5764
CUSTOMER.GENDER	6.71	1	0.0096
MARITAL.SCAN	6.95	4	0.1384
AGE.CATH	0.75	3	0.8621
SOURCE.TYPE	0.94	3	0.8152
PREMIUM	0.03	1	0.8590
PAY.FREQUENCY	14.24	3	0.0026
BASE.RATE.CHANGE	3.00	1	0.0831
POLICY.TENURE	0.86	1	0.3544
EMPLOYMENT.SECTOR	11.70	20	0.9259
GLOBAL	50.32	40	0.1270

From the residual Table 4.14, it can be seen that pay frequency and customer gender have p-values of 0.00 and 0.0096 respectively. The criterion for violating the cox proportional hazard is a p-value <0.05. Pay frequency and customer gender are highly significant and hence violating the proportionality assumption. As a result, these covariates will be stratified(taken out) before fitting the model.

4.11 Anova Table of the Stratified Cox Model

Table 4.15: Analysis of Stratified Cox Proportional Model

Variable	Log-lik	Chi-Sq	Df	Pr(>|Chi|)
NULL	-24625			
PRODUCT	-24583	83.66	3	0.0000***
MARITAL SCAN	-24574	18.63	4	0.0009***
AGE CATH.	-24573	2.09	3	0.5535
SOURCE TYPE	-24569	7.81	3	0.0499*
PREMIUM	-24569	0.15	1	0.6963
BASE RATE CHANGE	-24538	61.79	1	0.0000***
POLICY TENURE	-24537	3.05	1	0.0805
EMPLOYMENT SECTOR	-24525	23.85	20	0.2492

From Table 4.15, the result of the Stratified Cox model for customers with a life insurance policies is shown above. Based on the findings, there is sufficient evidence to conclude that age, employment sector, policy tenure, and premium have no effect on policyholder insurance attrition at a 95% confidence interval. According to the results, marital status, product, base rate change, and source type are the covariates that significantlt influence insured attrition.

4.12 Analysis of Maximum Likelihood Estimation

From the results of the Table 4.16 below, we see some levels of covariates (predictors) statistically significant based on the Wald test on the sixth column at a 95% confidence interval. Term policy has the lowest hazard ratio (0.36), at a p-value of 0.00 as seen in the preliminary analysis. A hazard ratio of 1.04 of the funeral policy is seen to have the highest hazard, but contrast the preliminary.

Source type has all three levels significant with employer deduction having the lowest hazard of 0.65. mobile money and Direct debit have hazard ratios of 0.70 and 0.71 respectively.

Table 4.16: Analysis of Maximum Likelihood Estimates

Covariate	coef	exp(coef)	se(coef)	z	p	lower95%	upper95%
PRODUCT (Education)	-	-	-	-	-	-	-
FUNERAL	0.04	1.04	0.05	0.89	0.37	0.95	1.142
SAVING	-0.06	0.94	0.05	-1.29	0.20	0.8603	1.032
TERM	-1.01	0.36	0.14	-7.26	0.00	0.2757	0.4768
MARITAL SCAN (Divorced)	-	-	-	-	-	-	-
Married	0.16	1.17	0.26	0.60	0.55	0.7017	1.945
Separated	-0.45	0.64	0.40	-1.13	0.26	0.2929	1.391
Single	0.25	1.29	0.26	0.97	0.33	0.7728	2.150
Widowed	-0.33	0.72	0.38	-0.86	0.39	0.3428	1.517
AGE.CATH (18-29)	-	-	-	-	-	-	-
30-45	-0.07	0.93	0.04	-1.65	0.10	0.8509	1.014
46-59	-0.05	0.95	0.07	-0.74	0.46	0.8299	1.087
60+	0.17	1.18	0.25	0.67	0.50	0.7263	1.916
SOURCE.TYPE (Cheque)	-	-	-	-	-	-	-
Direct Debit	-0.34	0.71	0.14	-2.44	0.01	0.5392	0.9345
Employer Ded	-0.43	0.65	0.15	-2.93	0.00	0.4881	0.8673
Mobile Money	-0.36	0.70	0.14	-2.58	0.01	0.5291	0.9171
PREMIUM	-0.00	1.00	0.00	-0.07	0.94	0.9998	1.000
BASE RATE CH	0.02	1.02	0.00	8.27	0.00	1.0182	1.030
POLICY TEN.	-0.06	0.95	0.03	-1.61	0.11	0.8838	1.012

Likelihood ratio test= 180.6 on 16 df, p=<2e-16

Wald test = 159.9 on 16 df, p=<2e-16

Score (logrank) test = 165.6 on 16 df, p=<2e-16

Concordance= 0.569 (se = 0.005)

This shows there is not a huge significant difference in attrition though these two means of premium payment, as contrasting earlier preliminary analysis of mobile money mode having the highest attrition. The risk of attrition of direct debit is as much the same as those paying with cheque.

From the levels of the Marital Status, we see Singles with a log-hazard and hazard ratios of 0.25 and 1.29 respectively, indicating a 56.3% risk of insurance attrition among singles than the married, widowed and separated as seen earlier in the preliminary analysis.

Hazard ratio 1.18 of the age category 60+ revealed this category has an increased attrition than the younger categories. The preliminary analysis revealed holders of ages 60+ to have the lowest rate of attrition. The hazard ratio of premium shows no association between the amount of premium and attrition. Base rate change with a hazard ratio of 1.02 reveals subscribing on to a base rate change or not contributes to the risk of attrition

Chapter 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Insurance attrition affects the growth and sustainability of the insurance company. The conclusions in this section are made from both preliminary and further analysis. This chapter is based on the findings of the study and also, recommendations based on the conclusions drawn.

A significant finding of the study revealed that attrition is high at the early weeks of becoming a customer of the insurance company. Assuming a three year period, at a 95% confidence, survival time of a new client is 16 weeks after subscribing on to policy. Also, a policy will attrite after 15 weeks of being in force. This finding is consistent with Poel and Lariviere (2003), and Tang et al. (2014) that attrition in general is high at the early years of becoming a customer.

The findings from the study revealed 55.4% of total policies are inactive. These inactive policies which are to be terminated or removed by the insurer are still running. Of this, males are twice as much as females and married male thrice the number of single males.

Males are twice likely to have their policies lapsed compared to females, however, the statistical test does not reveal any significant difference between them.

Funeral policies have a higher tendency of attrition compared to term policy which has a very low tendency of attrition.

For marital status, singles are at a higher risk of attrition than married policy holders. This could be as a result of many factors. According to Tang et al. (2014), this could be attributed to the risk averse nature of the married, and also the burden of additional responsibilities on them.

Premium deduction source revealed direct debit has the highest risk of attrition followed by mobile money mode of payment, though there is not much significant difference between the likelihood of attrition (policy lapse) between these two. Employer deduction has the lowest hazard compared to both mobile money payment and direct debit.

Regarding age, Mittal and Kamakura (2001), Poel and Lariviere (2003), Tang et al. (2014), and Hasanthika and Jayesekara (2017) agreed that it is a significant factor influencing attrition of a customer. Their studies revealed that older insurance policy holders are more likely than their younger holders to remain with the insurance provider. It was further stated that this could be because older people have stable preference. However, the findings of this study is in contradiction and not consistent with the literature. The findings of this study revealed the older policy holders (60+) are at a greater risk of attrition than the younger policy holders. This could be because these older policy holders are not economically inactive.

A client subscribing to change in the base rate or not also influences attrition. Premium does not influence insurance attrition in any way. The study concludes that marital status, product type, base rate change, and deduction source type are the covariates that significantly influence attrition whiles age, employment sector, policy tenure, and premium have no significant effect on insurance attrition.

5.2 Recommendations

Attention must be given to policy holders who are 60 years and above since they are at a great risk of attrition than their younger holders.

The insurance regulatory authority should ensure compliance of the appropriate policy guidelines and regulations by insurers regarding grace and policy cancellation periods.

Further analysis can be conducted with data from other insurance companies to ascertain the findings of this study.

REFERENCES

Reference

Alhassan, A. L., & Biekpe, N. (2015). Efficiency, Productivity and Returns to Scale Economies in the Non-Life Insurance Market in South Africa. The Geneva Papers on Risk and Insurance - Issues and Practice, 40(3), 493–515. https://doi.org/10.1057/gpp.2014.37

Reichheld, F. (2001). Prescription for cutting costs : Loyal relationships. Bain & Company, Inc.; https://media.bain.com/Images/BB Prescription cutting costs.pdf.

Alhassan, A. L., & Biekpe, N. (2016). Determinants of life insurance consumption in Africa. Research in International Business and Finance, 37, 17–27. https://doi.org/10.1016/j.ribaf.2015.10.016

African Development Bank (2012). Briefing Notes for AfDB's Long-Term Strategy: Africa's Demographic Trends; http://www.afdb.org.

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J., & Perez-Marin, A. M. (2008). Survival Analysis of a Household Portfolio of Insurance Policies: How Much Time Do You Have to Stop Total Customer Defection? Journal of Risk & Insurance, 75(3), 713–737. https://doi.org/10.1111/j.1539-6975.2008.00281.x

Tang, L., Thomas, L., Fletcher, M., Pan, J., & Marshall, A. (2014). Assessing the impact of derived behavior information on customer attrition in the financial service industry. European Journal of Operational Research, 236(2), 624–633. https://doi.org/10.1016/j.ejor.2014.01.004

NIC. (2019). INCREASING INSURANCE PENETRATION IN GHANA: THE CHALLENGES AND STRATEGIES.

National Insurance Commission Report, (2019). Annual Report of the Insurance Industry, Ghana; www.nicgh.org

Nazrina, A., & Shahirah, A. R. (2019). Survival analysis in insurance attrition. AIP Publishing.

Goonetilleke, T. L. O., & Caldera, H. A. (2013). Mining Life Insurance Data for Customer Attrition Analysis. Journal of Industrial and Intelligent Information, 1(1), 52–58. https://doi.org/10.12720/jiii.1.1.52-58

Fu, L. & Wang, H. (2019). Estimating Insurance Attrition Using Survival Analysis. CASUALTY ACTUARIAL SOCIETY; VOLUME 8/ISSUE 1, Variance Advancing the Science of Risk.

Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 157(1), 196–217. https://doi.org/10.1016/s0377-2217(03)00069-9

Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behaviour: investigating the modearting effect of customers characteristics. Journal of Marketing Research 38, 131-142.

Goonetilleke, T. L. O., & Caldera, H. A. (2013). Mining Life Insurance Data for Customer Attrition Analysis. Journal of Industrial and Intelligent Information Vol. 1, No. 1. Engineering and Technology Publishing doi: 10.12720/jiii.1.1.52-58

Linder, U., & Ronkainen, V. (2004). "Towards a new insurance supervisory system in the EU". Scandinavian Actuarial Journal 6.

Marin, A.M.P. (2005). Survival methods for the Analysis of Customer Lifetime Duration in Insurance. Departmento de Econometria, Estadistica y Economia Espanola, Universidad de Barcelona.

Enterprise Group, (2018). Insuance Modules Index. https://www.enterprisegroup.net.gh. [Accessed 17 May 2018]

Van den Poel, D., & Larivière, B. (2004). Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 157(1), 196–217. https://doi.org/10.1016/s0377-2217(03)00069-9

Lublóy, Á., & Szenes, M. (2008). The network of corporate clients: customer attrition at commercial banks. Journal of Statistical Mechanics: Theory and Experiment, 2008(12), P12014. https://doi.org/10.1088/1742-5468/2008/12/p12014

N. H. E., H., & L. A. L. W., J. (2017). Analyzing the Customer Attrition using Survival Techniques. International Journal of Statistics and Probability, 6(6), 85. https://doi.org/10.5539/ijsp.v6n6p85

Oxford Business Group, (2021). Economic Research & Foreign Direct Investment Analysis. Oxford Business Group. https://oxfordbusinessgroup.com. [Accessed: 2 April 2021]

Cleary Insurance (2021). https://www.clearyinsurance.com/history-insurance-throughout-world [Accessed: April 19 2021]

Anzovin, S. (2000). Famous First Facts. H. W. Wilson Company. p. 121. ISBN 978-0-8242-0958-2.

Moore, D. F. (2016). Applied Survival Analysis Using R (Use R!) (1st ed. 2016 ed.). Springer.

Amicable Society, (1854). The charters, acts of Parliament, and by-laws of the corporation of the Amicable Society for a perpetual assurance office. p. 4

Swiss Re, (2011). World insurance in 2010-premiums back to growth, capital increases. In: No 2. Swiss Reinsurance Company, Zurich.

Pierre-Ignace, B., Ellingrud K., Godsall J., Kotanko, B., & Reich, A. (September, 2020). The future of life insurance: Reimagining the industry for the decade ahead. McKinsey & Company.

Van Dijk, P. C., Jager, K. J., Zwinderman, A. H., Zoccali, C., & Dekker, F. W. (2008). The analysis of survival data in nephrology: basic concepts and methods of Cox regression. Kidney International, 74(6), 705–709. https://doi.org/10.1038/ki.2008.294

Musselman, V. A., & Hughes, H. E. (1950). Introduction to modern business: analysis and Interpretation (6th ed). Prentice-Hall.

Diacon, S. R., & Carter, R. L. (1988). Success in Insurance. (2nd edition).

Babbel, D.F., (1981). "Inflation, indexation, and life insurance sales in Brazil". J. Risk Insur. 49, 111–135.

Hwang, T., & Greenford, B. (2005). A Cross-Section Analysis of the Determinants of Life Insurance Consumption in Mainland China, Hong Kong, and Taiwan. Risk Management Insurance Review, 8(1), 103–125.

Forbes Advisor, (2021). How to Reinstate a Life Insurance Policy that you Stopped Paying. https://www.forbes.com/advisor/life-insurance/reinstate-lapsed-policy

Guo, S., & Wells, K. (2003). Research on Timing of Foster Care Outcomes: One Methodological Problem and Approaches to Its Solution. Social Service Review, 77(1), 1–24.

Li, D., Moshirian, F., Nguyen, P., & Wee, T. (2007). The Demand for Life Insurance in OECD Countries. Journal of Risk & Insurance, 74(3), 637–652.

Campbell, R. A. (1980). The Demand for Life Insurance: An Application of the Economics of Uncertainty. The Journal of Finance, 35(5), 1155–1172.

Lewis, F.D. (1989). Dependents and the demand for life insurance. Am. Econ. Rev. 79, 452–467.

Kleinbaum, D. G., & Klein, M. (2011). Survival Analysis: A Self-Learning Text, Third Edition (Statistics for Biology and Health) (3rd ed. 2012 ed.). Springer.

Lemeshow, S., May, S., & Jr., D. H. W. (2008). Applied Survival Analysis: Regression Modeling of Time-to-Event Data (2nd ed.). Wiley-Interscience.

Klein, J. P., & Moeschberger, M. L. (2010). Survival Analysis: Techniques for Censored and Truncated Data (Statistics for Biology and Health). Springer.

Kaplan, E.L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of American Statistics Association, 53. 457–481.

Klein, J.P., & Moeschberger, M.L. (2003). Survival Analysis- Techniques for Censored and Truncated Data, 2nd Edition. Springer Publishers, New York.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. Biometrika, 69(1), 239–241.

Prentice, R. L., & Marek, P. (1979). A Qualitative Discrepancy between Censored Data Rank Tests. Biometrics, 35(4), 861.

Guillén, M., Nielsen, J. P., Scheike, T. H., & Pérez-Marín, A. M. (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. Expert Systems with Applications, 39(3),

The World Bank Annual report, (2018).

Ghana Living Standard Survey 6, (2014). Ghana Statistical Service.

Appendix A

Analysis of Maximum Likelihood Estimation

Table 5.1: Analysis of Maximum Likelihood Estimates for Employment Sector

	coef	exp(coef)	se(coef)	z	p
Agriculture	-0.19	0.82	0.18	-1.10	0.27
Artisan	-0.21	0.81	0.09	-2.31	0.02
Banks Ins. Fin.	-0.23	0.80	0.10	-2.22	0.03
Clergy	-0.15	0.86	0.26	-0.60	0.55
Construction	-0.23	0.79	0.16	-1.42	0.16
Education	-0.16	0.86	0.09	-1.73	0.08
Engineering	0.07	1.07	0.13	0.56	0.58
Health	-0.07	0.93	0.09	-0.83	0.41
Hospitality	0.00	1.00	0.15	0.03	0.98
Information Tech	0.04	1.04	0.21	0.21	0.84
Labour & Sanitation	-0.09	0.92	0.22	-0.39	0.70
Manufacturing & Industry	-0.04	0.96	0.14	-0.29	0.77
Media & Entertainment	0.06	1.06	0.23	0.25	0.80
Mining	-0.54	0.58	0.23	-2.38	0.02
Others	-0.29	0.75	0.22	-1.31	0.19
Sales & Marketing	-0.05	0.95	0.10	-0.54	0.59
Security Serv.	-0.09	0.92	0.10	-0.89	0.37
Services	-0.24	0.79	0.08	-2.77	0.01
Students	-0.14	0.87	0.14	-0.98	0.33
Trade	-0.17	0.84	0.08	-2.29	0.02

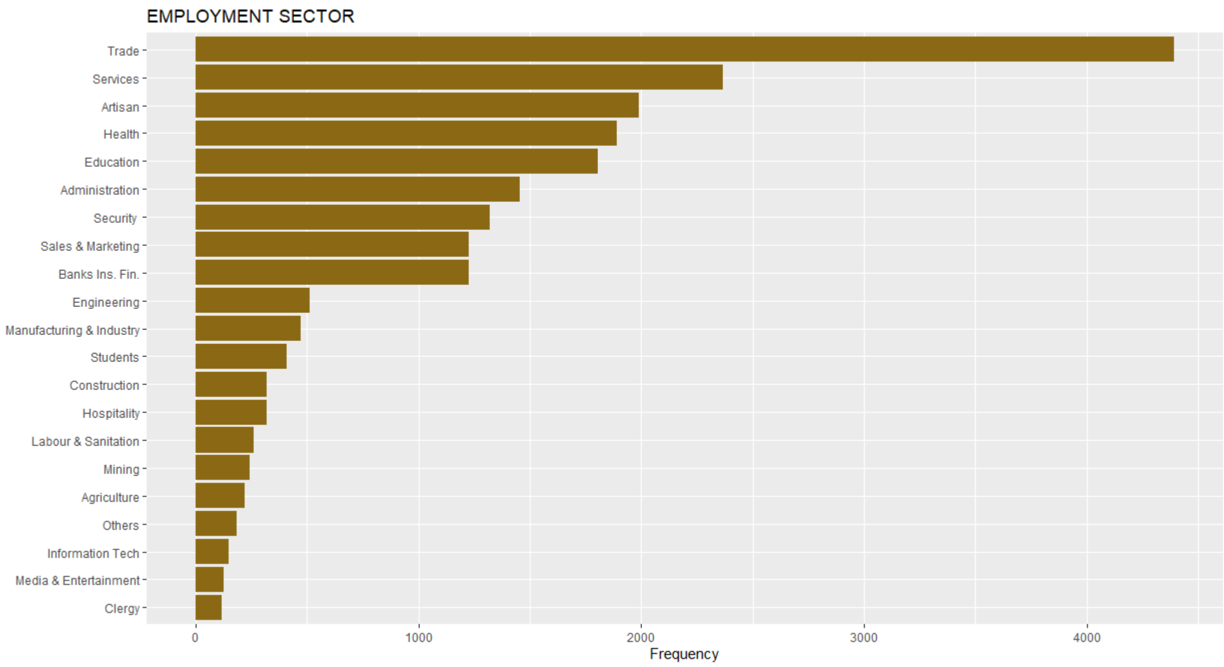

Figure 5.1: Bar Chart Showing Policyholders by Employment Sector