



Strathmore
UNIVERSITY

A Generalized Linear Model for Rating the Premiums of a Micro Health Insurance Policy

Felista Muganda: 072060

**Submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Business Science Actuarial at Strathmore University**

**School of Finance and Applied Economics
Strathmore University
Nairobi, Kenya**

November, 2015

**This Research Project is available for Library use on the understanding that it is
copyright material and that no quotation from the Research Project may be published
without proper acknowledgement.**

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Project contains no material previously published or written by another person except where due reference is made in the Research Project itself.

© No part of this Research Project may be reproduced without the permission of the author and Strathmore University

Felista Muganda



.....
30/11/2015

This Research Project has been submitted for examination with my approval as the Supervisor.

Dr. Carolyn Njenga



.....
30/11/2015

School of Finance and Applied Economics
Strathmore University

ABSTRACT

This study focuses on the modeling of a risk premium for a micro health insurance policy. It was carried out as a response to the gap highlighted by scholars and industry players which relates to the difficulty in developing insurance products which are both affordable and sustainable to cater for the needs of the low income market, which is a large segment in developing countries (KPMG, 2013). It adopts the generalized linear model in the rating of premium factors to parameterize their significance in predicting the claim amount. The product being analyzed is a group micro health insurance policy in the Kenyan market, and claims data has been obtained from three schemes whose policies have been in force for at least three years since 2011. A policy covers not only the scheme's members but also their nuclear family. Factors rated include the claimant's age, gender, and their relation to the insured life. The impact of the interaction between these factors is also analyzed. The results show that for a group policy offering family cover, all the three factors are significant in the prediction of claim amounts and thus the risk premium. They also show that interactions between the claimant's age and gender and between their age and relation to the policy holder significantly improve the pricing model.

Contents

1. Introduction	1
1.1 Background	1
1.1.1 Need for Micro-health Insurance	1
1.1.2 The Micro Insurance Market in Kenya.....	3
1.1.3 The Target Market for Micro Insurance in Kenya	3
1.1.4 The Risk Profile of the Low Income Market.....	4
1.1.5 Benefits of Micro Health Insurance to the Low Income Earners	5
1.1.6 Difficulty in the Pricing of a MHI product	5
1.2 Motivation for Study	6
1.2 Problem Statement	7
1.4 Research Objective.....	9
1.5 Research Questions	9
1.6 Scope of Research	9
1.7 Summary	10
2. Literature Review	11
2.1 Introduction	11
2.2 The Basics of Premium Calculation for General Insurance	11
2.3 Application of Generalized Linear Models in Premium Rating	11
2.3.1 Appropriateness of a GLM for premium rating.....	11
2.3.2 Further Description of a GLM.....	14
2.4 Need for Partnerships and Complementary Relationships between MHIs and other Organizations	16
2.4.1 The Kenyan Scenario: What is the place for micro health insurance products with the existence of NHIF?	19
3. Methodology.....	22
3.1. Introduction	22
3.2. Research Design	22
3.3. Sampling Design	22
3.4. Empirical Approach.....	23
4. Results	26
4.1 Introduction	26

4.2	Data	26
4.2.1	Data Refining	26
4.3	Analysis	27
4.3.1	Summary Statistics of Annual Costs	27
	Figures showing Split among Claimants	27
4.3.2	Fitting the Data to a Statistical Distribution	29
4.3.3	Relevance of Risk Factors	30
5.	Discussion and Recommendations	38
5.1	Introduction	38
5.2	Discussion in the context of the study:.....	38
5.3	Recommendations	39
5.3.1	Regarding Data Recording	39
5.3.2	Regarding Pricing	39
5.3.3	Scope for Further Research	39
6.	Conclusion	40
7.	Bibliography	41

List of Figures

Figure 1 : Split by Relation..... 27
Figure 2: Split by Gender am Equation (4.6)..... 34
ong Child Claimants 28
Figure 3: Split by Gender among Spouse of Policy Holder Claimants 28
Figure 4: Box Plot 1: Natural Log of Claim vs. Age of Claimant 31
Figure 5: Box Plot 2: Natural Log of Claim vs. Gender of Claimant..... 32
Figure 6: Box Plot 3: Natural log of Claim vs. Relation of Claimant to Policy Holder 33

List of Tables

Table 1: Summary Statistics of annual settled claims 29
Table 2: Evaluation of the log-likelihood function and standard deviation output to find distribution
of best fit 29

List of Equations

Equation (2.1) 11
Equation (2.2) 14
Equation (2.3) 15
Equation (2.4) 15
Equation (3.1) 23
Equation (3.2) 24
Equation (3.3) 24
Equation (3.4) 25
Equation (3.5) 25
Equation (3.6) 25
Equation (3.7) 25
Equation (4.1) 34
Equation (4.2) 34
Equation (4.3) 34
Equation (4.4) 34
Equation (4.5) 34
Equation (4.6) 34
Equation (4.7) 38

List of Abbreviations Used

AKI	-	Association of Kenya Insurers
GLM	-	Generalized Linear Model
IFC	-	International Finance Corporation: World Bank Group
ILO	-	International Labor Organization
KNBS	-	Kenya National Bureau of Statistics
Ksh	-	Kenya Shillings
MHI	-	Micro Health Insurance
NHIF	-	National Hospital Insurance Fund
UHC	-	Universal Health Coverage
WHO	-	World Health Organization

1. Introduction

1.1 Background

1.1.1 Need for Micro-health Insurance

Universal health coverage (UHC) is the focus of many countries and organizations in the world. An example of efforts made to achieve this objective of UHC is the formation of the global coalition of more than 500 health and development organizations worldwide in December 2014, urging governments to accelerate reforms that ensure everyone, everywhere, can access quality health services without being forced into poverty (WHO, 2015).

Access to quality health care is also enshrined in the Kenyan constitution as a fundamental right, and is one of the major objectives of Vision 2030 (Government of the Republic of Kenya, 2007). Currently however, many Kenyans are unable to enjoy this right due to the unaffordable cost of health services in the country (The World Bank, 2014). There are two main modes of payment for health services in Kenya, namely; through health insurance or through out-of-pocket payments.

The majority of Kenyans are not insured. A study done by Deloitte (2011) indicated that the National Hospital Insurance Fund, which is the state run health scheme in Kenya, was covering only 18% of the Kenyan population by the year 2010. Out of this 18%, only 19% of members were from the informal sector. Private health insurance schemes were covering another 2% of the population. In total only 20% of Kenyans were under any kind of health insurance plan (Deloitte, 2011).

The World Bank published a similar report in 2014 regarding the Kenyan health sector. The report stated that NHIF enrollment rates for the informal sector had grown at an aggregate of around 38% since 2010. However, the population coverage in terms of overall percentage of insured individuals still remained relatively the same.

By October 2014, still only 20% of Kenyans were under any sort of medical cover; either offered by private or public health schemes. This stagnation was explained by factors such as population growth and significant dropout rates of members from the scheme (The World Bank Group, 2014). By October 2014; according to the same report, 80% of Kenyans (about 35 million people) still

had no health insurance cover whatsoever. This means that the uninsured 80% of Kenyans settle medical costs through the out-of-pocket means of payment.

The economic status of the majority of Kenyans does not allow them to access the public or private health care system without insurance (World Bank, 2014). According to a report done by Smith et al. (2010), a relatively large percentage of the Kenyan population is considered poor; about 40% earn less than US\$ 1.25 per day, and most of this 40% is made up of informal workers. The formal health care system is not within their economic capability; especially for illnesses which are expensive to treat and manage.

Conventional insurance products are not affordable for the low income earners. As stated by a KPMG report (2013), insurance companies in Africa traditionally target only the richest 5% of the adult population, with most poor people having no insurance. This is also true in Kenya where the majority of the population cannot afford traditional health insurance policies in the market.

The subsidized rates in public hospitals and the National Hospital Insurance Fund are the only options available for the working poor. While NHIF offers affordable contribution rates, the scheme mainly covers public hospitals which in the past have had issues to do with drug shortages, lack of adequate personnel and inadequate facilities for treatment (Transparency International, 2011). The case for the feasible co-existence between NHIF and MHI schemes is however further elaborated under section 2.4.1 of the literature review.

Out-of-pocket health expenses drive many Kenyans into poverty every year. An article featured by the World Bank group (2014) indicated that nearly one million Kenyans fall below the poverty line every year, due to health related expenditures. He suggests that expanding health care access will reduce this burden. Micro health insurance products are a positive step towards increasing health care access to the poor.

In general, micro insurance products offer protection to the lower income earners in society against specific perils in exchange for regular premium payments that are proportionate to the likelihood and cost of risk involved (Churchill, 2006). Micro-health insurance (MHI) therefore covers the health needs of the poor in exchange for premiums that reflect the risk of illness and the cost of care involved.

1.1.2 The Micro Insurance Market in Kenya

The need for enhanced access to insurance at affordable rates to the lower end of the market is well stipulated in the Kenya Vision 2030 and other government policy documents (IRA, 2014). Compared to some other countries in Africa, such as South Africa, Namibia and Mauritius the insurance and consequently micro insurance market in Kenya is still under developed, with respective insurance penetration rates at 14.28% for South Africa, 7.5% for Namibia and 5.78% for Mauritius being relatively higher than Kenya's 3.2 % (KPMG, 2013). Income has been identified as the greatest determinant for insurance penetration growth and the high levels of poverty in Kenya correspond to the low insurance penetration rate (KPMG, 2013). However, insurers in Kenya are increasingly becoming active in the micro insurance market and informal "insurers" are expanding into the formal insurance system thus increasing the penetration rate. (IRA, 2014).

To counter lack of affordability as a factor hindering wider insurance coverage, Kenya like many other developing countries has identified micro insurance as a key enabler for enhancing financial inclusion, mitigating risks and increasing insurance penetration rates.

1.1.3 The Target Market for Micro Insurance in Kenya

The target market for micro insurance adopted for Kenya is defined as "insurance that is accessed by or accessible to the low income population, potentially provided by a variety of different providers and managed in accordance with generally accepted insurance practices". This definition is the one provided by the International Association of Insurance Supervisors (IAIS) and is supported by Makove (2011).

Makove (2011) however illustrates that the target market for micro insurance in Kenya is not those individuals below the poverty line because these individuals cannot afford to mitigate risk through insurance. This argument is supported by Yao (2013). According to the United Nations Development Program (UNDP) any person who earns less than US\$2 per day is considered low-income. These individuals may not be in a position to mitigate risk through insurance.

Makove (2011) describes the target market for micro-insurance in Kenya as those that have a higher level of income than the poverty line and are able to: use micro-insurance as a risk pooling instrument for the protection of their households, benefit sufficiently from the small benefits

provided by micro-insurance and are able to meet their most basic needs as well as pay the low levels of premiums. This description of the target market will be adopted for this study and will be implied whenever the term 'low income earners' or 'informal sector' is referred to.

1.1.4 The Risk Profile of the Low Income Market

The existence of a need for more MHI products in the Kenyan insurance space creates a substantial potential market for insurance companies. However selling insurance to the working poor comes with its own challenges. The working-poor seemingly carry more risks associated with their health, risks which are likely to lead to large losses yet their economic capability does not allow the insurer to charge them a particularly high premium.

It is therefore difficult to design a value adding product for the insured that also ensures profitability for the insurer. Traditional health insurance products are not suitable for the working poor because of one main factor: price (KPMG, 2013). Yet as much as they cannot afford it, this group of the population particularly need health insurance because they have comparatively more risk both of ill health and financial loss to mitigate than the well to do in society.

The greater risk of illness that the working poor face can be broken down as follows: for one, the lower income earners in society are more prone to certain kinds of illnesses than those who are more economically able. This is because the poor are more likely to be exposed to unhealthy living conditions like congestion and poor sanitation among other factors, which may contribute to illness.

The working poor also lack enough resources to invest in preventive healthcare in the form of better nutrition and regular hospital checkups. The main disadvantage of this is that eventually individuals end up contracting illnesses that would have been prevented and also diagnosis of disease is made at an advanced stage. All this boils down to wastage of resources because money is spent treating diseases that could have been avoided.

Even for illnesses that are not particularly preventable; the ultimate cost of treatment ends up being higher than could have been if the individual had access to regular health checkups and thus earlier diagnosis. For the working poor, because they have few resources to spend on healthcare, they only visit the hospital when the trip cannot be avoided. Therefore by the time of diagnosis, illness

is usually at an advanced stage and thus the treatment needed is likely to be more complex and thus more expensive (Leatherman, Christensen, & Holtz, 2010).

As illustrated, both the exposure that the working poor face and their ultimate health costs are likely to be higher than for say an individual with access to more income. This is a general conclusion which assumes that the wealthier persons in society generally live a healthy lifestyle i.e. they eat better quality food, have access to preventive healthcare and they generally live healthy.

However, despite having greater exposure to illness, the working poor have little choice when it comes to the insurance products currently offered in the market. Most products are designed for the rich or middle class and are not pocket friendly to the less fortunate. The cost of health care in the country has been classified by several researchers including the World Bank (2014) as expensive, thus out-of-pocket medical treatment as highlighted by Leatherman et al. (2010), is not affordable to many in the informal sector.

1.1.5 Benefits of Micro Health Insurance to the Low Income Earners

Accessibility to affordable health insurance gives significant health and financial relief to poor households but the problem is this: insurers can only accept to take on risk at a price above the expected loss associated with the risk being covered.

Cost notwithstanding, the benefits of insurance to the poor are many. According to Hamid, Roberts & Mosley (2011) and Yao (2013), micro-health insurance products which offer both preventive and curative care may improve the health status of the working poor through providing greater access to formal health care, promoting increased health awareness and improved health practices. Another positive implication mentioned by Yao (2013) is the economic safety net provided by micro-health insurance (MHI). According to this author, by reducing the effect of out-of-pocket healthcare costs, MHI improves the living standards of the working poor and concurrently avails resources for future investment in their human capital or economic status.

1.1.6 Difficulty in the Pricing of a MHI product

There is a need for value adding micro health insurance products, but researchers such as Leatherman et al. (2010), Hamid, Roberts & Mosley (2011) and Yao (2013) have found that a MHI product is a difficult product to run profitably and sustainably. Lack of profitability arises

mainly due to two factors: one is inaccurate pricing that arises due to differences in risk between the working poor and the conventional market, and the other is difficulty in ensuring and sustaining a sufficiently large membership once members are initially enlisted.

Hamid et al. (2011) explain the problem of pricing and sustainability as follows: they explain that because MHI products cannot charge high premiums; they have little capacity for large losses. They therefore may minimize on product quality to meet premium requirements or increase co-insurance levels with the insured to minimize the amount of claims that the insurer has to pay out. Measures that reduce the provisions or services offered under the MHI product are a common way of reducing the risk covered by the insurer. They usually involve setting more service exclusions like taking out the option of dental care among others, or totally excluding out-patient care from the health options available under the product. Leatherman, Christensen & Holtz (2010) explain that such measures reduce the value of the MHI product to the insured. Exclusions of needed services and getting rid of out-patient care may become counter-productive as policy holders have no incentive to stay in the policy. Insurers end up experiencing high dropout rates.

Aside from high dropout rates, many MHI products are loss making for the first number of years; this even before charging of administrative expenses on the premium (Angove & Tande, 2011). In the study by Leatherman et al. (2010), CIC Insurance's *Bima ya Jamii* product experienced losses until it came into a partnership with NHIF; where NHIF agreed to carry the risk of outpatient care for the product.

1.2 Motivation for Study

The low income market is substantial; especially in developing countries (McCord, Steinmann, Tatin-Jaleran, Ingram, & Mateo, 2012). Insurers could therefore benefit from a wider market if they are able to run profitable and sustainable MHI products. The large informal sector in Kenya for example which stands at 82.7% of the total workforce is a substantial market for micro health insurance (KNBS, 2015).

Any insurance product has to be profitable to the insurer and also add value to the insured for it to be sustainable. Proper premium calculation is therefore fundamental for the success of micro-health insurance products as it is to any other insurance product. In the past premiums charged on micro-health insurance products have often not succeeded in meeting the tradeoff between

affordability and being able to cover claim amounts, and as such many products have led to losses to the insurer (Angove & Tande, 2011).

If micro-health products can be made profitable, they will also enhance the quality of life of the low income earners, who will have access to better health care and protection from financial ruin which may arise due to unaffordable out-of-pocket health care expenses.

1.2 Problem Statement

Universal health coverage (UHC) is a current discussion both in Africa and in the world. It has been identified as one of the millennium development goals (MDGs). It is also one of the objectives the Kenyan government is working to achieve in Vision 2030. Access to quality health care is also a constitutional right stipulated in the Kenyan constitution.

Several authors identify that in many countries in Africa, the low income earners; who form the majority of the population; are unable to access quality healthcare. This lack of accessibility is mainly attributed to the high poverty levels in developing countries, which makes the cost of health care or conventional health insurance products unaffordable to the greater part of the population (KPMG, 2013). Commercial insurers for example cover less than 10% of all those insured for health in Africa. This is because most health products in the market target the middle and upper classes which cover only a minority of the population, with the informal sector being particularly neglected (McCord, Steinmann, Tatin-Jaleran, Ingram, & Mateo, 2012).

Studies done by Deloitte (2011) and The World Bank Group (2014) focus on the Kenyan health sector; and in general both studies conclude that the informal sector in Kenya, like in most African countries, has little access to health insurance. In the Economic Survey published in 2015, KNBS states that even in NHIF, the informal sector only accounts for 33.7% of the total membership of the scheme (KNBS, 2015).

The International Labor Organization (ILO) and other organizations have identified micro health insurance schemes as one among the most viable methods economies can use to increase accessibility of health care to the informal sector. Micro health insurance is however relatively new to most developing countries, and research done on existing MHI products indicates that many of them are loss making at least in the first number of years (Yao, 2013). This conclusion is

consistent with studies done by Leatherman et al. (2010), Angove & Tande (2011) and Holtz et al. (2014).

There are two basic problems which have been identified to contribute to the high loss ratio associated with micro insurance products, MHI products included. These problems and their possible solutions are discussed below.

The first as identified by Angove & Tande (2011) is that claims tend to exceed premiums charged. This means insurers are charging an inadequate net premium for MHI products therefore a problem of pricing exists. Inadequate premiums could either be due to excessive undercutting, or due to inaccurate quantifying of risk sources. Risk factors that drive claim amounts in the MHI market such as gender, residence area, age, occupation and the like may not have been properly quantified. However there is also an element of difficulty in pricing that is attributable to the lack of sufficient data for insurance and micro health insurance in general (KPMG, 2013).

Scholars acknowledge that exposure to health risks is not as easy to price as say mortality risk (Leatherman, Christensen, & Holtz, 2010) (Hamid, Roberts, & Mosley, 2011) (Yao, 2013). They also state that the low end market bears more morbidity risk than the middle class and upper class markets. However not much research has gone into quantifying the characteristics of the low end market which make it more prone to illness; causing higher claim amounts than those of the conventional market.

This study seeks to quantify the morbidity risk of the low income market, by rating policy holder characteristics and also the characteristics of the policy holder's dependents, since most medical policies in the market cover the family of the main insured life as well. Pricing, at least in the Kenyan market, as a general practice, does not consider the dependents' characteristics but only the main policy holder's characteristics such as age and gender. This is as obtained from interviews with different insurers on the nature of pricing for micro health insurance products.

The second problem identified is that of high distribution expenses which in turn lead to losses to the insurer (Angove & Tande, 2011) (Holtz, Hoffarth, & Phily, 2014).

Literature review suggests direct partnerships and complementary relationships between micro health insurers and other organizations; such as the state, as a possible remedy which increases sales volumes and reduces distribution costs (Holtz, Hoffarth, & Phily, 2014). The higher the sales

volumes, the higher the number of policy holders over whom overhead costs can be spread and thus the lower the distribution expenses.

This study therefore seeks to evaluate the viability and effectiveness of such direct partnerships in the Kenyan market and to evaluate the possibility of complementary relationships between the MHI provider and the state so as to establish whether there is a need for the existence of micro health insurance products despite the existence of the state run scheme; NHIF. This will mainly be expounded on in the literature review.

1.4 Research Objectives

1. This study seeks to analyze the risk an insurer faces due to the uncertainty of the size of medical claims made on a micro health insurance product by the main policy holder. This uncertainty of claim sizes and frequency per insured life arises as a result of certain unique characteristics of each policy holder such as age and gender.

2. This study also seeks to evaluate the significance of the same characteristics for the dependents of the policy holder i.e. age and gender, on the overall claim made on a medical policy in the case of a family cover, so as to be able to form a basis for premium calculation for a self-sustainable micro health insurance product.

1.5 Research Questions

1. What is the significance of the policy holder's characteristics such as age and gender to the claim amounts of a micro health insurance policy and the volatility of the same?
2. What is the significance of the policy holder's dependents' characteristics such as age and gender to the overall health claims settled on a family cover micro health insurance policy?

1.6 Scope of Research

This study seeks to effectively measure the risk characteristics of the low income market using a Generalized Linear Model (GLM). The covariates that will be considered will be age of policy holder, gender and the same covariates will also be evaluated for the dependents of the policy holder for the case of a group cover that covers both the scheme/Sacco member and their direct family members. The GLM will be parameterized with respect to data from a group micro health

insurance product sold in the Kenyan market by a recognized insurer. The period of study is three years when the three schemes/groups of interest had a cover in force i.e. 2011, 2012 and 2013.

The study also seeks to evaluate the effect of partnerships between insurers and other organizations especially the state, in the provision of MHI products, and thus determine whether such relationships are likely to enhance the profitability and sustainability of MHI products to the insurer. A comparative analysis of products whose performance was linked to strategic partnerships will be done; to compare performance of products where no partnership with other organizations was existent and performance where such a partnership was established. Other MHI products of countries with similar economic and demographic characteristics as Kenya will also be evaluated subject to availability of information. This however will mainly be analyzed in the literature review due to data constraints.

1.7 Summary

With increasing interest in achieving Universal Health Care by governments worldwide there has been a growing interest in offering micro-health insurance products. Kenya is also working towards making health care more affordable and accessible to all its citizens. The government is seeking to make NHIF more effective with the capitation of a section of the scheme being implemented in 2012 and the new premium rates and comprehensive in patient cover introduced in April 2015 (NHIF, 2015). The literature review section of this study has further details on the changes in NHIF and the impact these have on the viability of micro health insurance schemes.

Apart from recent NHIF reforms, private insurers are increasingly bringing MHI products to the market in order to help the working poor mitigate risks associated with ill health. However for many of these products; coming up with a premium that is adequate and affordable, and getting the numbers of policy holders needed to sufficiently pool risk and sustain the product has been a challenge. Through research, several papers have been written in recent years on how to run profitable MHI products. This study seeks to add to this research; mainly by attempting to find a more suitable model for the pricing of MHI premiums and evaluating a possible structure for distribution.

2. Literature Review

2.1 Introduction

This chapter highlights the theoretical and conceptual framework which forms the basis of this study. Section 2.2 highlights the basics of premium calculation for any insurance product. Section 2.3 is an explanation regarding Generalized Linear Models and the properties of a GLM that make it suitable to model premium factors. Finally section 2.4 explores the different arguments for partnerships between micro – health insurance providers and other organizations.

2.2 The Basics of Premium Calculation for General Insurance

1. MHI falls under health insurance which is a general insurance product. According to Finan (2012) , basic premiums for general insurance are modeled as per the loss distribution or the distribution of past claims. This statement is based on the equivalence principle which is represented as

$$EPV \text{ Premiums} = EPV \text{ of outgo}$$

(2.1)

Outgo for an insurance contract according to the same source; refers to claim payments and expenses associated with the policy and may include a profit loading. This study will focus on the basic premium and thus will ignore expenses and profits loading since these are particular to the company.

2.3 Application of Generalized Linear Models in Premium Rating

2.3.1 Appropriateness of a GLM for premium rating

A Generalized Linear Model is one of the most widely used ways to extend or generalize a linear model in actuarial practice (Frees, 2009). It is defined by Hilbe (1994), as a method of extending standard linear regression to incorporate a variety of responses including count, binary, proportions and positive-valued continuous distributions.

A GLM is an extension of a simple linear regression model, thus several properties of simple linear regression make it suitable for this study.

A GLM like simple linear regression; forms a relationship between the mean of the response variable Y (what we want to predict) and the factors which affect the response variable. A GLM can therefore be used to rate premium factors by determining their significance to claim severity.

A GLM like simple linear regression models allows for non - additivity. Where non – additivity exists; the effect of a variable X on Y depends on the value of a third variable. The property is demonstrated in GLMs in the form of interactions and is important in the modeling of claim severity in several ways. For example we are able to study the combined effect of age and gender on the health claim of the insured rather than analyzing the independent impact of the age of the policy holder on claim severity and that of their gender separately. This is important because the relationship between drivers of risk is often multiplicative rather than additive. This is according to Jong & Heller (2010).

The usefulness of the non-additive property is seen in the case of analyzing the expected claims of a medical insurance policy for policy holders with the same age of entry but who are of different genders. Take an example of a 25 year old male and a female entrant of the same age; who both take out a one year medical policy. When modeling future claim severity on this policy, it makes sense to consider the policy holder's age and gender as a combined factor. Simply analyzing the effect of the policy holder's age and gender independently may be misleading because the model will effectively be saying that age as a measure of risk is gender neutral, and gender as a measure of risk does not change with age. But the truth is that the probability of incurring health costs for a policy holder of age X also depends on the gender of the policy holder.

For the policy described above, the factor age as a gender neutral risk measure concludes that the 25 year old has an advantage of youth and thus is not prone to as high health care costs as an older person; holding all other factors constant; factors such as lifestyle choices and the like. Age as a gender sensitive factor concludes the same, but also says that a 25 year old male is more prone to high risk activities such as dangerous sports, reckless driving etc. compared to a female policy holder of the same age. On the other hand, a 25 year old female is likely to incur more costs related to reproductive health compared to their male counterpart.

On the other hand, looking at gender as a factor independent of age only captures general health characteristics of either gender. For example it is an accepted fact in medicine and social science that despite women having greater longevity compared to men, they have higher rates of morbidity

than their male counter parts¹ (Birda & Rieker, 1999) (Sen, 2008). However despite differences in health risks inherent to the different genders, the age of the individual determines the level of risk that is relevant at that particular time of the policy holder's life. Therefore to analyze the overall risk of morbidity there is need to assess the interaction between factors such as age and gender.

The multiplicative property of GLMs allows us to evaluate such interactions and thus is crucial in modeling medical claim severity. The property enables more accurate evaluation of morbidity risk by making it possible to assess combined risk arising from interactions between related covariates such as age and gender.

Perhaps the most important advantage of GLMs over simple linear regression is that GLMs allow for non – normal data. Actuarial data is seldom normally distributed and thus Generalized Linear Models which allow for data from other distributions are important in analysis. Several authors support this such as Jong & Heller (2010), Rodríguez (2015) and McCullagh & Nelder (1989) . Dobson (2001) and McCullagh & Nelder (1989) also state that apart from the fact that they are more flexible in the specification of Y 's distribution, another advantage of using GLMs over a simple linear regression is that GLMs extend linear models by allowing for non-linearity. This implies that GLMs allow for interactions between a covariate and itself (self-interaction). Non-linearity occurs where the effect of a variable X on another variable Y depends on the specific value of X . Applied to this study for example; the effect of the term of a policy on morbidity is not the same regardless of age; i.e. For a one year policy, the risk of morbidity attached to a 50 year old is different from that of a 40 year old for the same period. Non-linearity allows for self-interaction between covariates. For example effect of age on Y could be defined in terms of the log of age X , the square and the like.

In conclusion of this section; properties of linear models like allowing for non-additivity and those specific to GLMs like allowing for non-normal data and non-linearity make GLMs appropriate for rating premium factors for a MHI policy.

¹ A further discussion of these differences in morbidity and how they arise can be found in the cited articles under the journals: Social Science and Medicine, and Global Public Health.

2.3.2 Further Description of a GLM

A GLM is characterized by a response or outcome variable Y , which is described by Dobson (2001) as a dependent random variable whose value varies according to other predictor or explanatory values. An example of a simple GLM for a normally distributed data set (a simple linear regression) obtained from Faraway, Peterson & Baker (2010), can be written as:

$$E(y_i) = \mu_i = X_i\beta, \quad \text{where } y_i \sim N(\mu_i, \delta^2) \quad (2.2)$$

In the above equation y_i is the dependent variable which we are interested in predicting while $X_i\beta$ is the function of an independent factor X_i that determines the dependent variable.

In a GLM the independent factors may be referred to as covariates, predictor or explanatory variables. This is as stated by Hilbe, Francis, Green and Payne in 1993, Dobson (2001), and Faraway et al. (2010).

According to Hilbe (1994), the explanatory variables are mostly systematic in nature. Annette Dobson explains that the explanatory random variables could be nominal, ordinal or continuous. Nominal variables take binary numbers i.e. fall into either of two categories for instance male or female.

If categories are more than two, then the predictor variable is called a multinomial. Ordinal variables are progressive and fall between categories but are still discrete; factors like age, degree of illness and the like. Then finally continuous variables fall anywhere on a continuum. The term covariates as generally will be adopted to refer to both binary and ordinal values for the purpose of this study.

As further explained by McCullagh & Nelder (1989), a GLM is characterized by three main components. These are expounded upon in the coming paragraphs.

The Components of a GLM

The first component is **the distribution of the response variable**. The second is a **linear predictor** ' η '; where η is a function that describes which covariates are likely to affect the response variable Y and how they are going to affect this response variable.

For our study an example of a linear predictor could be

$$\eta = \alpha_i + \gamma_j + \rho_k + \beta_0 x + \beta_1 z_l \quad (2.3)$$

Where:

α_i : is a factor for gender which takes different values for male or female

γ_j : is a factor for pre- existing condition such that if the policy holder had a pre-existing condition at policy inception, it is indicates as a different value from if none was present.

ρ_k : represents a factor for the relation of the claimant to the primary policy holder

$\beta_0 x$: is a function of age and

$\beta_1 z_l$: is a function of the insured's duration of hospitalization.

Finally the third characteristic of a GLM is a **link function**. This is described by McCullagh & Nelder (1989) as a function which connects the mean response $E(Y)$ to the linear predictor denoted as η .

The link function gives a straight forward equality of the mean response and the linear predictor. It is denoted as in the following equation:

$$g(\mu) = \eta \quad (2.4)$$

A GLM uses a link function to establish the relationship between the mean of the response variable Y , which we want to predict, and the predictor variables about which we have information.

A GLM is suitable for modeling claim severity since claim sizes are affected by a number of factors such as age, gender, area of residence, pre-existing conditions, disease classification and the interaction between these factors. It therefore can be able to provide a relationship between expected claims and these covariates. The objective of our study is to rate premium factors and a GLM does this through the linear predictor.

In this way a GLM is suitable for rating premium factors for a MHI product, and will be determined by establishing the relationship between the covariates outlined.

2.4 Need for Partnerships and Complementary Relationships between MHIs and other Organizations

Micro Health Insurance providers could benefit greatly from fostering relationships with other organizations. Beneficial relationships can be in terms of direct partnerships between the MHI provider and another organization, or they could also involve the MHI playing complementary roles to those played by the relevant organization(s). Direct relationships may be easier to foster within the private sector, while complementary roles may be more relevant when relating to state run health schemes; which may be more difficult to partner with directly.

Several research papers such as Leatherman, Christensen, & Holtz (2010), Kimbal, Phily, Folsom, Lagomarsino & Holtz (2013) and Holtz, Hoffarth, & Phily (2014) have highlighted the importance of partnerships between MHI providers and other organizations.

Although direct partnerships are more difficult to achieve, where MHI providers have managed to secure such a relationship, the marketing costs of the MHI provider have reduced significantly and depending on the strategic soundness of the partnerships, the dropout rate of the already enlisted policy holders of the micro health insurer has been seen to drastically reduce.

A study done by Angove & Tande (2011) evaluates the profitability of different micro insurance products mostly in developing countries, where the provider had some kind of direct partnership with either the state scheme or some other private or non – government organization. The products assessed were not necessarily health products but the study is still relevant to the current study in that it portrays the impact of partnerships between a micro insurer and other organizations. The providers evaluated were CIC, Kenya with the product *Bima ya Jamii*; ICICI Lombard, India with

the product MAS health insurance; Old Mutual South Africa with the micro product Group Funeral; ASR, Guatemala with the products Life cover and Student cover and finally Malayan, Philippines with the micro insurance business cover.

CIC Insurance Kenya as per the paper by Angove & Tande (2011) managed to secure a partnership with the National Hospital Insurance Fund (NHIF) which is the state health insurance scheme in Kenya. According to the data collected the micro insurance product; *Bima ya Jamii* was only able to start making profit, after CIC outsourced the services of NHIF to cover the out-patient care claims of the product. NHIF is a state run health scheme and has more experience in administering both out-patient care and generally micro health insurance than any private insurer in Kenya. The experience of NHIF, their considerably larger pool of funds and having a greater number of scheme members might be the reason for their success in administering out-patient care where CIC had previously encountered challenges.

The results of the study by Angove & Tande (2011) agree with Kimbal et al. (2013) and Holtz et al. (2014) who explain through other similar studies that complementary partnerships especially with the state; are key for the success of micro-health insurance. Some complementary roles that the state can play are identified to be those such as providing access to a wider market, offering experience in administering out-patient care through state run health schemes, carrying part of the risk insured and possible subsidies of premiums. This is as illustrated by Kimbal et al. (2013) and Holtz et al. (2014). In the paper Leveraging Health Micro insurance to Promote Universal Health Coverage; the role of the insurer in bridging weaknesses of state run health schemes includes mainly expertise in raising awareness and enrolment.

The example of CIC insurance, in addition to studies that also proved success of partnerships in the Philippines according to Kimbal, Phily, Folsom, Lagomarsino, & Holtz, (2013) shows that a complementary relationship between MHI private providers and the state increases profitability of MHI products.

Based on the above analysis; the effect of direct partnerships of MHI providers with the state and other large organizations in the sales and sustainability of MHI products seems viable and may be applicable for micro health insurers in the Kenyan market.

Besides direct collaboration; the private sector MHI providers can complement state health services by offering products not covered by the state. This is a simpler relationship to achieve since direct collaboration or partnership is not required. Kimbal et al. (2013) give several viable examples of complementary products that micro health insurers could provide to meet the needs not met by the state insurer.

According to this authors, micro health insurers could offer products such as those which cover transport costs to the health facility, risk of lost income during illness and the like where the state already provides medical care. This complementary relationship where the MHI provider is not exactly competing with the state; reduces negative energy between the private and public sector and saves on marketing resources since the target market is essentially the same. It also adds value to the insured that now has greater access to a wider variety of products which meet his or her needs more adequately.

In the Kenyan scenario, NHIF fully covers all costs of medical treatment, maternity and surgery so long as services are accessed in a public hospital. For non-government hospitals however, the cover is partial and co-payment is required. Micro health insurance policies could be designed to cover some portion of whatever claim is unpaid by the state scheme. This is however further elaborated in the coming sections. Transport costs to the health facility are also not covered by NHIF. This could be a viable product addition that a micro health insurance product could offer.

Where the state does not insure the entire population, micro health insurance products could also be designed to cater for the segment of the population not catered for by the state. This is particularly applicable in the Kenyan market as is further described in the coming section.

2.4.1 The Kenyan Scenario: What is the place for micro health insurance products with the existence of NHIF?

2.4.1.1 About the capitation of NHIF: Impact on benefits and premium rates

The National Hospital Insurance Fund (NHIF) which is the state run health scheme in Kenya was recently restructured with the controversial capitation plan being fully implemented on 1st April 2015. This restructuring has seen a change in both premium contribution rates and benefits payable as well.

The last revision of the plan's premium rates in 1989 saw most employed Kenyans pay Ksh 320 per month. This was the highest contribution rate, and was paid by all formal employees earning a monthly salary of Ksh 15 000 and above. Now the highest contributor to NHIF pays Ksh 1,700 monthly, and the rate is applicable to those who earn a salary of Ksh 100 000 or above. The lowest contribution has now changed from Ksh 30 monthly to Ksh 150, while the contribution for voluntary members or all informal workers formerly at Ksh 160 per month is now at Ksh 500. This is according to NHIF (2015) and Business Daily (2015).

The proposed benefits being offered by NHIF have also changed. Under the previous arrangement, NHIF mainly covered a portion of the bill upon admission to accredited hospitals i.e. covered rebates on daily bed ranging between 400 Kenyan shillings to Ksh 2400 depending on the category of the hospital.

Under the new arrangement NHIF offers a full and comprehensive in-patient cover for the member and acknowledged dependents whenever they are treated in any NHIF accredited public hospitals. This means that a NHIF member or their listed dependent will not pay anything upon admission to a public hospital. The cover includes all medical services, maternity and surgery. Public hospitals are under category A of service providers. If the patient is admitted under category B of providers which includes mission hospitals; NHIF still covers full hospital fees except where surgery is involved, in which case a co-payment may be necessary. The terms of co-payment have not been elaborated by the scheme. For those visiting category C providers or private hospitals, NHIF will continue to pay specified daily benefits under the current arrangements. This information is as per the NHIF website (2015).

The new benefit package also includes an outpatient cover for all categories of providers, with different arrangements for the different categories and different particular group C providers.

2.4.1.2 The place for micro health insurance with the restructuring of NHIF

It is undeniable that better health care is accessible to the employed low income earners under the revised benefits and premium rates of NHIF. The scheme now offers a fully comprehensive medical cover for public hospitals, as opposed to the previous regime where NHIF only offered rebates on daily bed charges regardless of the health facility visited (NHIF, 2015).

Since NHIF is compulsory for all employed Kenyans; those earning a low income have no option but to contribute to the state scheme (Carrin, 2007). After making the compulsory NHIF contribution, workers are entitled to the described benefit package.

There will therefore be little incentive for the formally employed low income earners to spend another amount in premiums to pay for a privately provided micro health insurance product. All they stand to gain from an additional coverage under a micro health insurer is access to private hospitals. The additional premium required however, given that they are already covered under NHIF may not seem a worthwhile expense. It is therefore unlikely that low income earners in formal employment will buy micro health insurance. The formal sector is thus not the main target market for MHI providers.

However formally employed low income earners may buy micro health insurance to cater for the residual claim left in case of visits to non – public hospitals. Surgery in a mission hospital; which is a category B facility, as well as any medical service undergone in a private hospital; which falls in category C, requires co-payment. Patients are likely to visit category B and C hospitals over public hospitals due to perceived better service delivery or out of closer proximity to the facility. Visiting a mission or private hospital means that NHIF does not fully cover health claims incurred. Lower income earners in formal employment may therefore buy MHI products to cater for the residual claim left after NHIF rebates are settled.

The CIC *Medisure* - Family product for example covers the cost of hospital bed accommodation minus the NHIF rebate (CIC Insurance, 2015). The *essential* package which offers a benefit of Ksh 300,000 may especially be attractive to the low income earners, because it ensures the outstanding claim above what NHIF covers is catered for by the MHI product.

The major market for micro health insurers however exists in the informal sector. NHIF is only compulsory for formal workers, thus informal workers (mainly the self-employed) are not obliged to enlist. The current statistics for NHIF membership, according to the Kenya National Bureau of Statistics (KNBS, 2015), indicate that the formal sector made up a majority of the membership, at 66.3% of the 18% of Kenyans under NHIF in the year 2013/14. Informal sector workers are therefore not a big market for NHIF; only having had a residual of 33.7%. According to the same report by KNBS, the informal sector employment as at 2014 made up 82.7% of total employment in the country (KNBS, 2015). This is an overwhelming majority.

The increase in premiums for self-employed members of NHIF from Ksh 160 to Ksh 500 has also effectively made a case for micro health insurance providers who will now seem more attractive to the low income earners, given that they give access to private hospitals at more favorable rates and thus even if premiums may be comparatively higher, value is expected to be substantially higher as well.

3. Methodology

3.1. Introduction

To achieve the objective of this study, a GLM model will be used as suggested by several authors as a recommended model for rating premiums such as Jong & Heller (2010) and McCullagh & Nelder (1989). The model will be based on GLMs as illustrated by Dobson (2001) and other writers on the same. In an attempt to rate a premium for MHI products, a micro health insurance product offered by a recognized and sufficiently large insurer in the Kenyan market will be used. A linear relationship will be established between claim severity and covariates such as age, gender, pre-existing condition, disease classification, duration of hospitalization and whether the policy holder took up the option of health check-ups or not.

Analysis of variance between different models including different combinations and interactions of the covariates will be used to determine the best model to be used. Box plots and Q-plots will also be generated using R software to determine significance of covariates to the claim size.

To evaluate the impact of partnerships between MHI providers and other organizations on financial performance, a comparative analysis will be carried out between the financial performance of the MHI products before the partnership with the other organization(s), and the performance after the partnership. The trend of premium income and profits will be compared to determine stability measured by variance, and the mean of the actual amounts received.

3.2. Research Design

This study is exploratory in nature as it seeks to assess the different characteristics of the policy holder in the micro population in Kenya, and the effect these characteristics have on their claim severity in relation to the peril of sickness.

3.3. Sampling Design

The study will be based on MHI products from an insurer in Kenya because data will be more accessible from local companies than otherwise, and preferably from a company that engages in partnerships with other organizations because this is a contributing factor to the success of any micro insurance business; based on studies by different authors like Angove & Tande (2011) and Kimbal, Phily, Folsom, Lagomarsino & Holtz (2013). The most recent MHI products offered by

the selected insurer will be studied because they are likely to reflect current products and trends in the MHI market.

The structure of the insurance company in terms of its partnerships with NHIF, SACCOS, Co-operative societies and any other organization, will also be evaluated. This is because the partnerships have been highlighted by several authors as one of the drivers of any micro insurer's success.

3.4. Empirical Approach

3.4.1 Rating premium factors based on claims data

A GLM as described in the literature review will be used to test several models of obtaining the expected value of response variable 'y'. Our response variable which we wish to predict is the claim severity on the *Medisure – Family* product. The general relationship to be derived is

$$E(y_i) = \mu_i = \eta \tag{3.1}$$

Where $E(y_i)$ is the expected claim size for a policy holder 'i', and η is the linear predictor.

The first step in generating a GLM is to identify the distribution of response variable 'y_i'.

The most suitable and widely used models for claim sizes are distributions of continuous random variables that assume positive values only and have "fat tails" (or "heavy tails"), that is distributions which allow for occasional occurrences of very large values. Other distributions however can also be considered. A possible distribution of claim sizes arising from MHI claims that will be tested is the log-normal distribution which unlike the normal distribution only allows for positive values. Another possible distribution that will be tested for claim severity is the gamma distribution. This is based on a study done by Kastelijn.W.M & Wit (1977).

R software will be used to run the data based on the log-normal distribution and several other exponential families as well so as to find the distribution which best describe the data.

The linear predictor or model which will be found of best fit: based on analysis of variance (ANOVA), box plots and Q-plots will be the one which will be adopted in calculating a premium.

For the variance, we will be looking for the simplest model that gives us the required accuracy levels. The model that gives the lowest residual variance with an adequate but not excess number of parameters is the model of best fit. I.e. no significant reduction in residual variance is observed with increase in parameters. This analysis is as described by Rodríguez (2015) in the website *Introducing R* of Princeton University.

Q-plots will also be used to support the choice of the model of best fit where the model with the least correlation in plots is desired. Box plots will be used to visually explain the impact of different covariates on the response variable.

The second component of a GLM is the **linear predictor**. A possible linear predictor is equation (2.3) as described in the literature review:

$$\eta = \alpha_i + \gamma_j + \rho_k + \beta_0 x + \beta_1 z_l \quad (3.2)$$

Where α_i is a factor for gender which takes different values for male or female, γ_j is a factor for pre-existing condition: Yes or No, ρ_k one for option of health checkups: Yes if insured took up option, No if they did not. ' $\beta_0 x$ ' is a function of age and ' $\beta_1 z_l$ ' a function of the insured's duration of hospitalization.

Another possible model allowing for all possible interaction between age and gender i.e. age*gender that will be analyzed is as follows:

$$\eta = \alpha_i + \gamma_j + \rho_k + \beta_0 x + \vartheta_k + \beta_1 z_l \quad (3.3)$$

Where the term ϑ_k represents the single term interaction between age and gender: '*age.gender*'. This linear predictor will be tested in order to analyze the effect of age and gender of the policy holder combined. This is because health needs at the same age are different for the respective genders.

The symbol ' η '; represents the linear predictor. After obtaining the most suitable distribution for the data, the best model will then be tested using different linear predictors (different combinations and interactions of covariates).

Finally the link function is in the form of equation three restated as below:

$$g(\mu) = \eta \tag{3.4}$$

It will be determined based on the distribution that is found to be of best fit. If the proposed log-normal distribution is found to be appropriate, then the natural logs of the data will be obtained and the identity link function for the normal distribution will be used to establish the relationship between the linear predictor η and the mean response of claim amounts.

The identity link function is as shown below:

$$g(\mu) = \mu \tag{3.5}$$

The identity link function states a direct relationship between the expected claim amounts and the linear predictor or model that will be used to describe the covariates.

If the gamma distribution is found to be of best fit, then the inverse link function as stated below will be used:

$$g(\mu) = \frac{1}{\mu} \tag{3.6}$$

If the Inverse Gaussian distribution is found to be of best fit, then the inverse link function as stated below will be used:

$$g(\mu) = \frac{1}{\mu^2} \tag{3.7}$$

Different linear predictors developing progressively from the simplest model (nested models) will then be tested based on the distribution of best fit, to find the factors that are most significant in determining expected claims and thus the possible premium.

4. Results

4.1 Introduction

This section gives a practical understanding of the methodology outlined in the Methodology of this report. The results as obtained using R-statistical software and the analysis will be presented and interpretations of the same explained.

4.2 Data

The data used for this project was obtained from a Kenyan Insurer. It was data for a micro health insurance group policy which offers both in-patient and out-patient care services to policy holders. Data covered three membership schemes for three years during which they were in force i.e. 2011, 2012 and 2013.

Data from all schemes was analyzed, but the analysis described on this report was for scheme I which had the most number of members: 2667, compared to the other two schemes which had 199 and 120 members respectively.

4.2.1 Data Refining

Data refining process included elimination of incomplete data points such as policy holders whose claim information was present but whose other details were not among the three schemes provided i.e. details such as age, gender and dependents' demographic characteristics.

Negative claim amounts; which indicate a refund to the insurer from the policy holder were also not considered. This is because the aim of the study is on modeling the medical costs incurred by persons under micro health insurance, and not necessarily on the payment procedures; which might include the insurer being able to recover part of claims earlier paid; for instance if claims are discovered to be fraudulent.

The data was granulated into the demographic characteristics of the claimant under each policy; which their age, gender and relationship to the policy holder as either self, spouse or relation. These three characteristics of the policy holder: age, gender and relation formed the covariates of the tested linear predictors under the generalized linear model. Under this particular policy, relation only referred to the claimant being the child of the policy holder.

The age of claimants was classified into four groups according to the similarity of interaction between age and claim amounts. This interaction was determined through plotting of box plots of claim amounts (y axis) against age of claimant (x-axis) under R software. This will be further explained in the interpretation of results section of this report. The covariate gender was assigned a dummy variable of "0" for the female claimants and "1" for the males.

4.3 Analysis

The main aim of analysis as per the objective of this report was to measure the significance of as many risk factors describing micro health insurance policy holders as could be obtained; and as such find a regression model which could be used to determine the premiums of a micro health insurance policy in the Kenyan market.

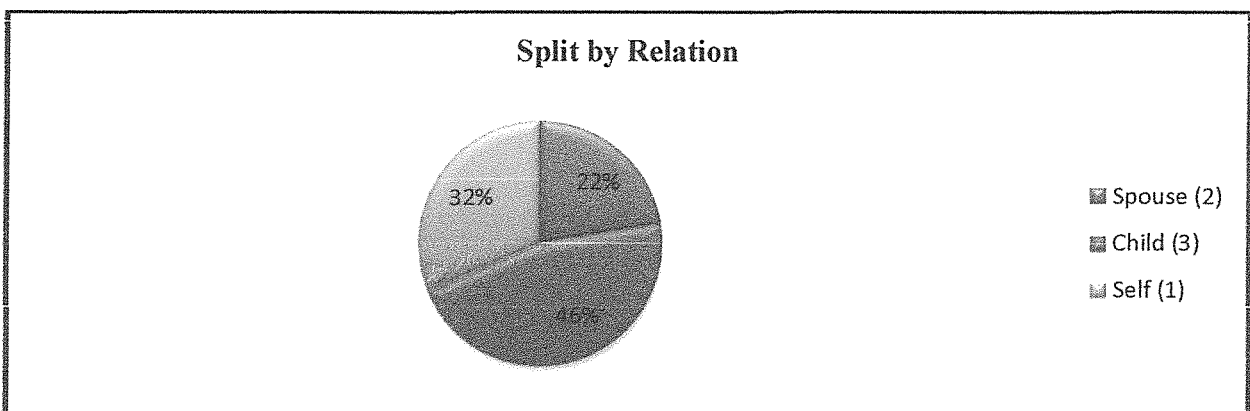
The covariates or risk factors obtained were age, gender and relation of the claimant to the policy holder as earlier explained under the data refining section above.

4.3.1 Summary Statistics of Annual Costs

The data analysis began with an evaluation of key statistics of the average annual medical costs for each individual of the medical scheme and is summarized below:

Figures showing Split among Claimants

Figure 1 : Split by Relation. The pie chart is obtained from the analysis of combined scheme data, for the three years 2011, 2012 and 2013. Claim frequency from a child of the insured life are the most significant, followed by claims from the insured life itself, and lastly the spouse of the insured has the lowest claim frequency at 22% but not much lower than the insured's claims which are 32% of total number of claims made.



Of the 1534 claimants of total, majority were children (704 claims); accounting for 46% of total claims settled, then followed by claims from the insured life itself (484 claims) standing at 32%, then claims from the category of Spouse (345 claims) being the least at 22% of total claims paid out to the members of the scheme.

Figure 2: Split by Gender among Child Claimants. This is as per combined scheme data analysis for the years 2011, 2012 and 2013. There is no significant difference in terms of claims frequency between male and female child claimants. Gender is therefore not a significant factor in determining possible claims from the insured's children. A child in this case is of the age range 0 to 20 years.

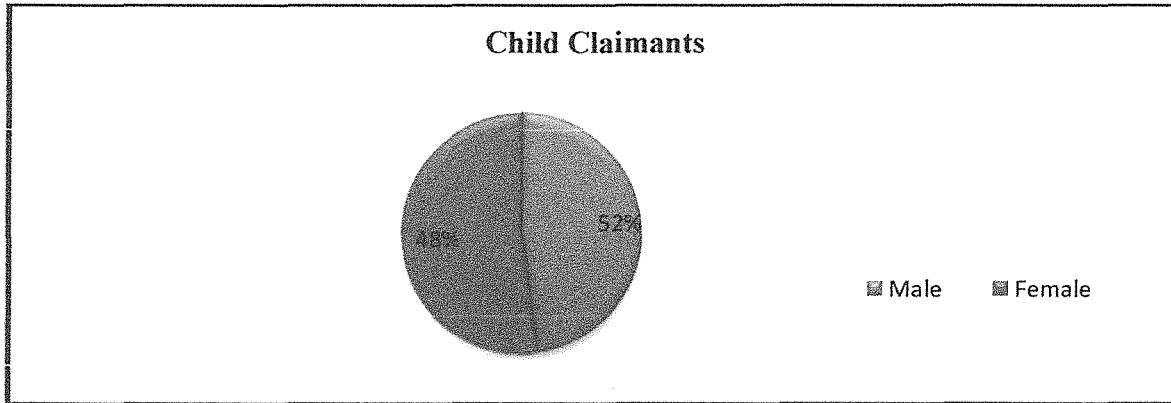
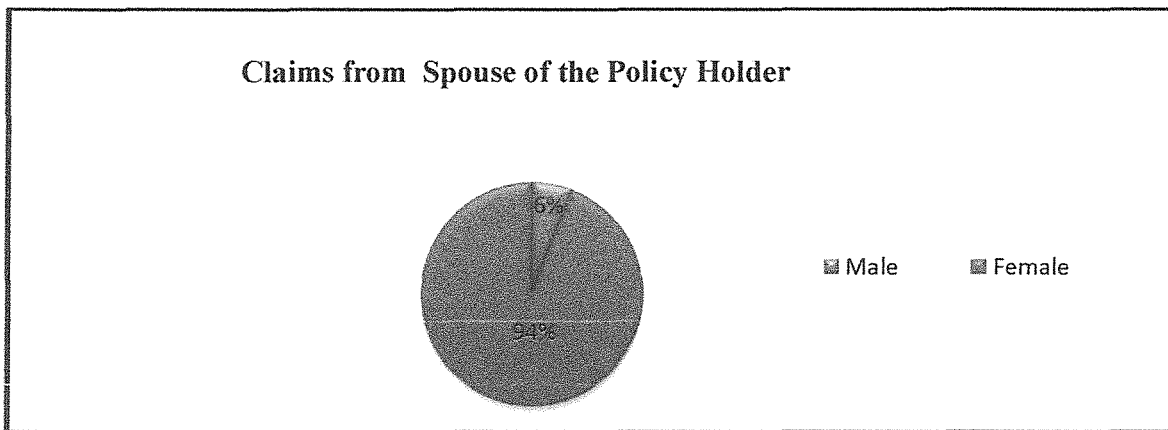


Figure 3: Split by Gender among 'Spouse' category of Claimants. This is as per the combined scheme data for the years 2011, 2012 and 2013. Majority of claimants related to the insured by marriage are female. This could indicate inversely that majority of policy holders are male. Gender will therefore be a significant bias in analyzing the claim patterns from the 'self' and 'spouse' categories.



There were a total of 1534 claimants of which 48% were male (765 of them), and 52% female (769 of claimants). Among child claimants; there was no gender bias with an almost equal proportion of 48% male claimants and 52% female claimants. However among spouse claimants;

there was a significant gender bias with 94% of spouse claimants being females, and only 6% being male.

Table 1: Summary Statistics of Annual Settled Claims. This is as obtained from the combined scheme data. There is a great variation of claims, ranging from Ksh. 33 to Ksh. 275,000 as explained in the paragraph below.

Min	Q1	Median	Q3	Mean	Max
33	200	400	801.5	1665.09304	275000

As shown in the table above, the claim amounts were varied, ranging from a minimum of Ksh. 33 to a maximum of Ksh. 275,000. This variation was based on the fact that the policy being studied has both an inpatient and outpatient component under one package, each with a different cover limit; and when these benefits are combined; claims paid out can take on a wide range thus the volatility in the claims observed. Every person (insured or dependent) has an outpatient limit of Ksh. 50,000, but the inpatient limit of Ksh.250, 000 is per family. The claims per person could therefore range between Ksh. 0 to Ksh. 300,000 thus the variation observed.

4.3.2 Fitting the Data to a Statistical Distribution

From literature review; recommended distributions for medical claims include the Gamma, Gaussian, Inverse - Gaussian and Log Normal models. The log likelihood function was used to determine the statistical distribution that best fit the data. The Inverse Gaussian distribution had the most positive log likelihood and was thus considered the distribution of best fit as shown below:

Table 2: Evaluation of the log-likelihood function and standard deviation output to find distribution of best fit. The Inverse Gaussian distribution had the most positive log likelihood thus is the distribution of best fit as further explained below.

Distribution	Log likelihood	Standard deviation
Normal/Gaussian	-13701.37	3216.73296
Lognormal	-2155.59	1.07776
Inverse Gaussian	6176.249	0.00335
Poisson	-1956111	1136.53015
Gamma	-2470.22	1.03590

The log likelihood of a distribution applied on a set of data is directly proportional to how well the distribution fits the data. The log likelihood is the natural log of the likelihood function. The

likelihood function is proportional to the probability of observing the data given the parameter estimates and the model being tested. A larger likelihood function means a larger probability of observing the data. The likelihood in itself is not a probability, however it is proportional to the probability, thus it can be greater than one. The log likelihood therefore can also take values greater than 1, and is used in the place of the likelihood because it gives a smaller value, but is a direct reflection of the likelihood and thus a measure of how well the data fits into the distribution whose likelihood function is applied.

The third column in the table indicates the results of a secondary test of standard deviations which was also done, where the distribution with the least standard deviation was deemed to be the distribution of best fit. In both tests, the Inverse Gaussian distribution seems to be the distribution best suited for the data.

Following this conclusion, in the analysis that followed, the inverse of the claims was obtained and then the Gaussian distribution applied to the data.

4.3.3 Relevance of Risk Factors

The risk factors affecting the claim amount were considered to be gender, age and relation of claimant to the policy holder. The correlation of these factors to the claim amount was determined by use of box plots and also by the evaluation of residual variance as elaborated later in this section.

4.3.3.1 Box Plots

Box plots of the response variable – Claims against the risk factors were obtained and interpreted as illustrated in this section.

General Interpretation of box plots

The width of the box plots indicates the variation of the average claim amounts against the characteristic of the claimant described in the x - axis; and is thus a measure of risk. The black bar across the box indicates the average claim per class of characteristic as described by the box.

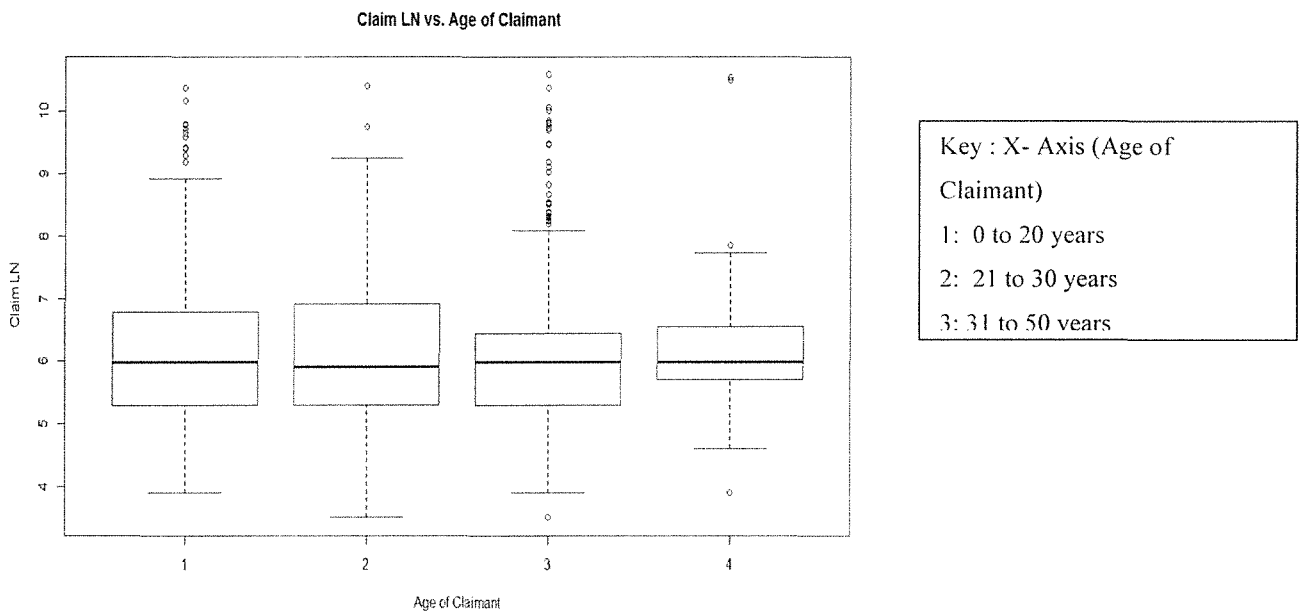
Since the claims were very varied i.e. from a minimum of Ksh.33 to Ksh.275. 000; the natural logs of the claims were used instead of the absolute claim amounts, in order to reduce the variation and thus improve visibility. This is as demonstrated by (Sarkar, 2012).

The Box Plots

There were three box plots generated showing:

1. The relationship between claim amounts and the age of the claimant
2. The relationship between claim amounts and the gender of the claimant
3. The relationship between claim amounts and the relation of the claimant to the policy holder

Figure 4: Box Plot 1: Natural Log of Claim vs. Age of Claimant. This box plot was obtained from the combined scheme data for the years 2011, 2012 and 2013. The natural log of claims is a representation of absolute claim amounts. From the plot it is evident that claim amounts and variation of claim amounts vary considerably with age.

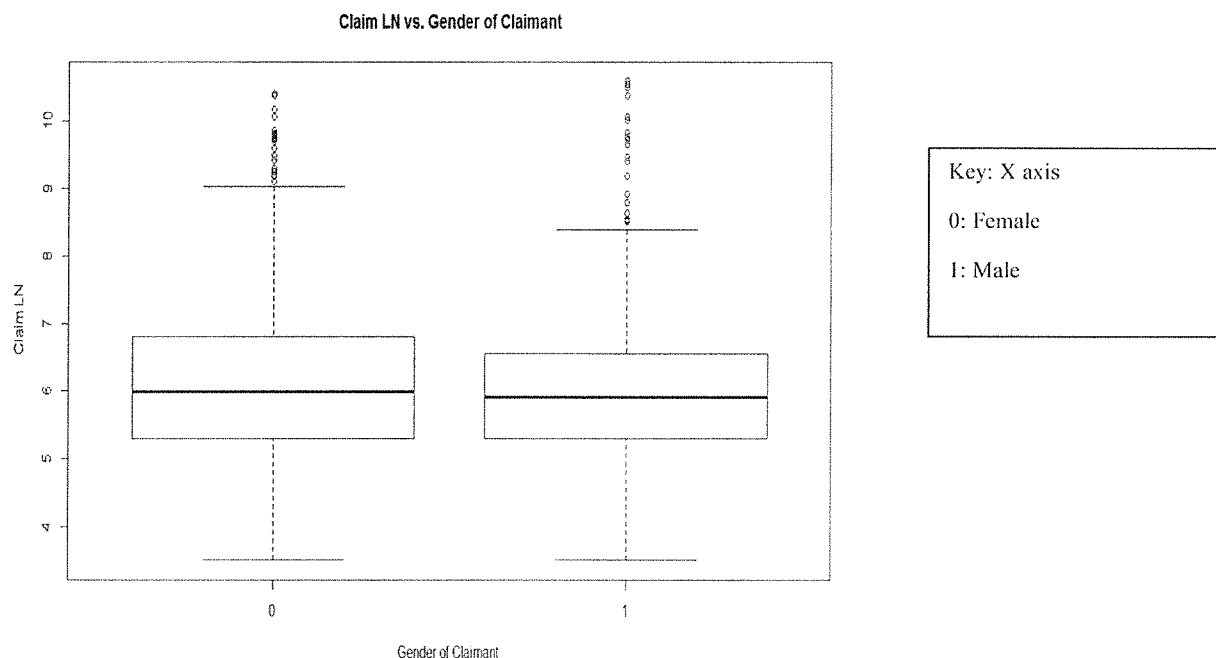


In the above box plot; the variability of claims is highest in the age group between 21 and 30 years. It is then followed by claimants between 0 to 20 years who are mostly child claimants. The claimants between ages 50 to 60 have the least variability of claims, but with a higher level of average claim. Since variance is a measure of risk, the insurer faces the biggest risk of a claim from insured lives and dependents who are between the ages of 21 and 30 years, followed

dependents of the insured; the majority of whom are between 0 and 20 year. There seems to be a decreasing risk for older claimants, although an increasing mean claim is observed.

The average claim paid is highest for age groups 3 and 4 who are the older claimants. This is expected as it indicates that medical costs are likely to rise with age, since the older population is likely to be more prone to illness. However the rise in claims for age group 3 which has a significant middle – aged group is also likely to be attributable to increasing reproductive costs such as maternity fees and health complications i.e. since it includes the child bearing age.

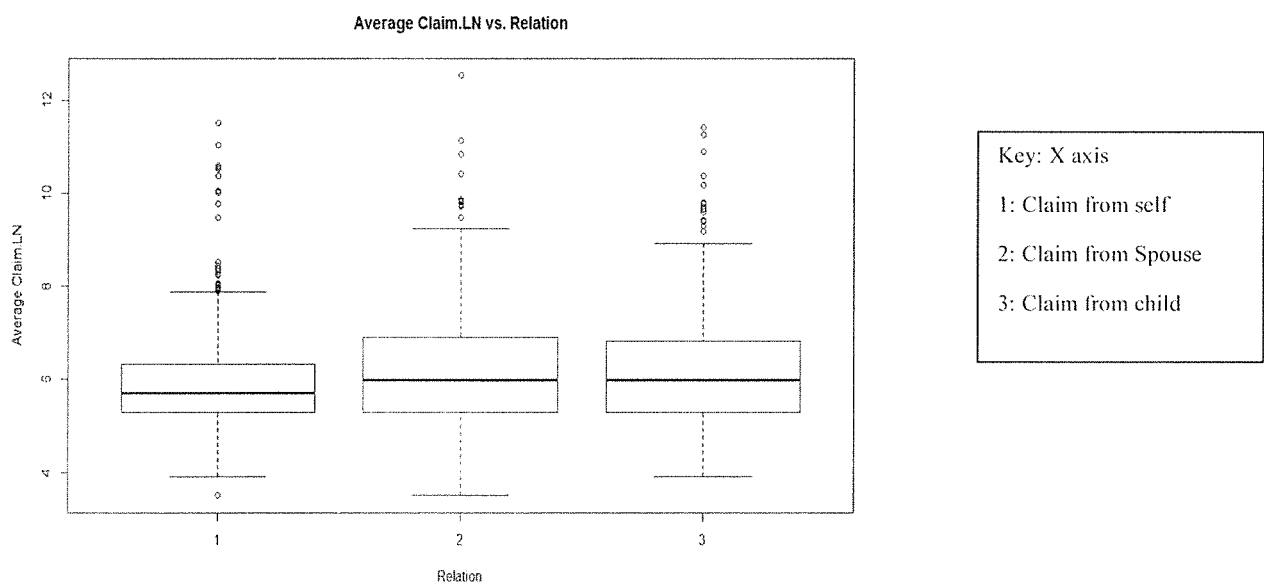
Figure 5: Box Plot 2: Natural Log of Claim vs. Gender of Claimant. This box plot was obtained from the combined scheme data for the years 2011, 2012 and 2013. The natural log of claims is a representation of absolute claim amounts. Claim amounts and claim amount variation are higher for females than for males.



The variability of claims from females is significantly higher than that of claims from males. This is expected since women are likely to have more health issues thus costs than men; particularly reproductive costs. The average claim for women is also slightly higher; although not too clearly visible from the box plot above; with the mean claim from females at Ksh. 1184, while that from males standing at Ksh. 1059. These figures are as calculated from the data. Overall, the insurer

therefore faces a greater risk of a claim arising, and also a higher amount of claim arising from a female covered life.

Figure 6: Box Plot 3: Natural log of Claim vs. Relation of Claimant to Policy Holder. This box plot was obtained from the combined scheme data for the years 2011, 2012 and 2013. The natural log of claims is a representation of absolute claim amounts. The average claim amounts for spouse and child claimants are relatively the same, but significantly higher than those from the policy holder (self). The variation of claims is highest in the claims by the spouse, followed by variation in child claims and is least in the claims by self.



There is a higher variation in claim amounts; thus risk to the insurer, for the Spouse of the policy holder than for the policy holder himself or herself. The Spouse (category 2 in the x axis) also seems to experience a higher average claim amount indicated by the higher log.

Seemingly this might not necessarily mean that spouses tend to claim way more than the policy holder. Of the spouses who claimed, most were females i.e. 303 out of 323 of policy holder's spouses who made a claim are females. The variation under the class of spouses is really a reflection of the impact of gender earlier explained in the second box plot.

The variability of claims by a child of the policy holder (category 3) is also considerably higher compared to the variability of claims made by the insured life (1). The average claim made by children of the policy holder is also significantly higher than that made by the policy holder himself

or herself. The gender split among the child claimants is almost proportional¹ at 48% male and 52% female. Gender is therefore not a bias in the claims by a child of the insured. However the interaction between age and average claims of most children (falling between ages 0 to 20) is very similar. The difference in variability between claims by self and those by the child of the policy holder is likely to be attributable to age.

4.3.3.2 The Linear Predictor

As described in the methodology (chapter 3), a generalized linear model can be used to find the most significant covariates that influence the response variable; through the assessment of the residual variance (Analysis of Variance – ANOVA), of different possible models. In this case a model refers to a linear predictor. The linear predictors which were tested are as summarized below:

$$1. \textit{Claim} \sim \textit{Age} \tag{4.1}$$

$$2. \textit{Claim} \sim \textit{Age} + \textit{Gender} \tag{4.2}$$

$$3. \textit{Claim} \sim \textit{Age} + \textit{Gender} + \textit{Relation} \tag{4.3}$$

$$4. \textit{Claim} \sim \textit{Age} + \textit{Gender} + \textit{Relation} + \textit{Age} * \textit{Gender} \tag{4.4}$$

$$5. \textit{Claim} \sim \textit{Age} + \textit{Gender} + \textit{Relation} + \textit{Age} * \textit{Gender} + \textit{Age} * \textit{Relation} \tag{4.5}$$

$$6. \textit{Claim} \sim \textit{Age} + \textit{Gender} + \textit{Relation} + \textit{Age} * \textit{Gender} + \textit{Age} * \textit{Relation} + \textit{Gender} * \textit{Relation} \tag{4.6}$$

While assessing the predicting power of a model, parsimony is fundamental. The model should not be overly complex. Additional variables tend to increase the accuracy of the model, but a balance is required between the impact of an additional variable on accuracy as well as on the complexity of the model. The smaller the scaled deviance of the model the more accurate it is. However simplicity in terms of number of parameters employed to achieve accuracy is also a

factor to consider. The model has to be parsimonious. The results of the evaluation of the above models are illustrated in the following table:

Table 3 : Model evaluation results under the Inverse Gaussian distribution. This table demonstrates the results of the evaluation of progressive nested models through the analysis of variance.

Models Tested	Formula	Scaled Deviance	Degrees of Freedom (DF)	Scaled Deviance	Difference in DF	2(p-q)
Models	S	2120.10	1442			
				2.00	1	2
model1	Age	2119.10	1441			
				2.00	1	2
model2	Age + Gen	2117.10	1440			
				35.40	1	2
model3	Age + Gen + Rel	2103.20	1439			
				24	1	2
model4	Age + Gen + Rel + Age.Gen	2078.80	1438			
				142.90	1	2
model5	Age + Gen + Rel + Age.Gen + Age.Rel	1935.90	1437			
				1.10	1	2
model6	Age + Gen + Rel + Age.Gen + Age.Rel + Gen.Rel	1934.80	1436			

Key

Gen	Claimant's gender
Rel	Claimant's relationship with policy holder
	Preferred model - because it has the greatest difference between A and B
A	S1 - S 2
B	2(p-q)

Based on the comparison between scaled deviance and the chi-square distribution i.e. column 2(p-q) in the table, Model5 was found to be the best model for predicting claim amounts from the covariates age, gender and relation of claimant to the directly insured life.

This result was established from the analysis of the scaled deviance and the difference in degrees of freedom I.e. we compare the difference in deviance between models and the chi square

distribution. The deviance of model 1 less the deviance of model 2 gives us 2, we compare this value to twice the difference in parameters $2(p-q)$, where p is the number of parameters in model 1 and q those in model 2. The number of parameters is reflected in the degrees of freedom. This comparison holds because model 1 is a subset of model 2; i.e. model 2 includes parameter age which is already in model 1. All the models being compared are nested models.

Where S_i represents the scaled deviance of model i , if $S_1 - S_2 > 2(p - q)$, then the subsequent model is a better fit than the previous because it has a higher level of accuracy with less parameters used.

In the above table, model 1 and 2 are not better than the saturated model, but all the subsequent models up to model 5 are an improvement of the previous since the inequality given above holds. Model 6 is not an improvement of model 5 because this inequality does not hold.

The significant interactions are the interaction between *age.gender* and that between *gender.relation*.

As much as model 5 was found to be the most parsimonious model, upon further analysis of the p-values of all possible covariates, the following results were obtained:

Table 4: this table outlines the p-values of all the covariates in model 6 which is the most comprehensive, and their interactions, as obtained from analysis of the model.

Covariate/Interaction	p-value
Age	< 2e-16
Gen	0.2070
Rel	< 2e-16
Age*Gen	0.0137
Age*Rel	< 2e-16
Gen*Rel	0.3669

The lower the p-value, the more significant the covariate is to the response variable. With the commonly applied alpha of 0.05, i.e. a confidence interval of 95%, then if the p-value is less than 0.05, we reject the null hypothesis that the covariate is not significant in the model. From this interpretation of p-values, the factors *Age*, *Relation* and the interaction between *Age.Gender*

and *Age.Relation* are the most significant to the pricing model. The most parsimonious model thus becomes modified as:

$$\textit{Claim} \sim \textit{Age} + \textit{Relation} + \textit{Age} * \textit{Gender} + \textit{Age} * \textit{Relation}.$$

(4.7)

5. Discussion and Recommendations

5.1 Introduction

This section evaluates the results and analysis in the context of the study. In section 5.1, the overall significance of different risk factors and their interactions is discussed by evaluating both the analysis from box plots and the tested models. In section 5.2 recommendations regarding data recording and pricing models are stipulated.

5.2 Discussion in the context of the study:

Model 5 which was chosen as the best pricing model indicates that all the factors given of age of claimant, their gender and relation are significant in forecasting the claim amount that the insured will make. This is as indicated both in the box plots and also in the summary of the possible models that were assessed i.e. model 1, 2 and 3. Model 3 particularly, which has an addition of relation to the model, indicates that knowing the relation of dependents to the insured is helpful in determining future claims from a group micro health insurance policy.

Extending this interpretation from the box plots; the underlying factors causing the difference in claims between the directly insured, their spouse and child is mainly age and gender. i.e. the details of the relations of the insured who are covered by the policy; in terms of age and gender are also fundamental in the prediction of claims for a micro health insurance policy's claims thus premiums. Also the number of children directly influences the claim amount expected from the policy.

The interactions that are important however are only *age.gender* and *age.relation*. The interaction between *gender.relation* which is the additional component in model 6 is insignificant. I.e. it is a more accurate model than model 5 judging by the drop in scaled deviance, but the tradeoff between a desired drop in scaled deviance compared to the consequent increase in parameters is not met thus this addition is not necessary.

This could indicate that age is a more important factor in predicting health costs and thus claim amounts than gender. This may be due to the fact that most claimants of the policy tend to be the child dependents, as depicted by *Figure 7* in the previous chapter, whose claims are not biased by gender but are greatly influenced by age, as illustrated in *Figure 8* and also in box plot 1 which is under *Figure 9*.

5.3 Recommendations

5.3.1 Regarding Data Recording

More elaborate models with more covariates would be examinable if information such as pre-existing condition of the claimants, their disease classification and period of hospitalization would be available. An important recommendation to this effect would be that insurers could record more detailed and further granulated claims data, so as to enable proper assessment of factors which influence micro health insurance health costs.

5.3.2 Regarding Pricing

Rather than charging a flat rate premium for all policy holders, who may have different characteristics in terms of age, gender and a different risk profile of dependents; micro health insurance providers should consider taking into account differences in these characteristics of those directly insured and the dependents under the cover in the case of a group policy; since from the results and discussion above; the characteristics of the policy holder's relation have an impact on the claim amounts expected from a policy. Also the claim patterns of the policy holder's dependents, in this study referred to as 'relations', are likely to be significantly different from the claims of the insured; and this is because of the difference in the age and gender of dependents from the insured life.

5.3.3 Scope for Further Research

Evaluation of other possible risk factors such as information regarding the pre-existing condition of the claimants, their disease classification and period of hospitalization may be carried out, subject to availability of data, to determine their significance in predicting claim amounts and thus premiums of micro health insurance products. This could lead to the development of a better pricing model.

A possible topic for further research as highlighted in the literature review but not extensively covered in this study, would be the evaluation of the possible effect of partnerships between insurers and other organizations such as NHIF, or other institutions such as banks, as well as the effectiveness of different channels of distribution for micro health insurance policies in terms of cost effectiveness and sales volumes achieved, with the aim of increasing the profitability of micro health insurance products especially in developing markets.

6. Conclusion

The main objective of the study is to contribute towards the pricing of MHI products through a generalized linear model; in an effort towards ensuring better profitability of MHI products to the insurer. From previous research, it had already been established that the age and gender of the insured are important risk characteristics that should be considered in the pricing of health products, and also that the interaction between factors like *age.gender* may be more informative than an independent observation,of the two factors.

This study has established that in group health insurance offering a family cover to members; not only the gender and age of the insured needs to be considered in pricing, but also the number of dependents who are primarily children, and the ages and gender of all dependents. As demonstrated in the results and discussion sections of this study, the inclusion of an interaction between the claimant's age and their relation to the policy holder (*age.relation*), could lead to a better pricing model for micro health insurance products which offer a family cover.

However as highlighted in the recommendations, further research is needed in both pricing of micro health insurance products i.e. incorporating other characteristics of the policy holder as are not as easily accessible currently due to data constraints, as well as on the impact of partnerships on sales volumes and overall profitability of micro health insurance products.

7. Bibliography

AKI. (2013). *Insurance Industry Annual Report 2013*.

Angove, J., & Tande, N. (2011). Is microinsurance a profitable business. *Protecting the Poor* , 378.

Birda, C., & Rieker, P. (1999). Gender matters: an integrated model for understanding men's and women's health. *Social Science and Medicine* , 745 - 755.

Business Daily. (2015, July 22). *New NHIF rates to take effect in April*. Retrieved July 2015, 2015, from Business Daily: <http://www.businessdailyafrica.com/New-NHIF-rates-to-take-effect-in-April/-/539546/2622888/-/xhwqlgz/-/index.html>

Carrin, G. e. (2007). Health financing reform in Kenya – assessing the social health insurance proposal. *South African Medical Journal* , 130-135.

Churchill, C. e. (2006). *Protecting the Poor: A Microinsurance Compendium* . Geneva ILO , 12.

CIC Insurance. (2015). *CIC Medisure – Family*. Retrieved July 27, 2015, from CIC Insurance: <http://cic.co.ke/portfolio/cic-medisure-family/>

CIO East Africa. (2014). Safaricom, Britam, Changamka Micro health, PSI, launch affordable healthcare option for Kenyans. *CIO East Africa* .

Deloitte. (2011). *NHIF Strategic Review & Market*. Nairobi.

Dobson, J. a. (2001). *An Introduction to Generalized Linear Models*. New York: CHAPMAN & HALL/CRC.

Faraway, J. J., Peterson, P., & Baker, E. M. (2010). In *Generalized linear models* (pp. 178-183). Elsevier: International Encyclopedia of Education.

Finan, M. (2012). *A Reading of the Theory of Life Contingency Models: A Preparation for Exam MLC/3L*. Arkansas Tech University.

Frees, E. (2009). Regression Modelling with Actuarial and Financial Implications. *Annals of Actuarial Science* , 362.

Government of the Republic of Kenya. (2007). *Kenya Vision 2030: The Popular Version*.

- Hamid, S. A., Roberts, J., & Mosley, P. (2011). Can Micro Health Insurance Reduce Poverty? Evidence from Bangladesh. *The Journal of Risk and Insurance* , 57-82.
- Hilbe, J. (1994). Generalized Linear Models. *The American Statistician* , 255 - 265.
- Holtz, J., Hoffarth, T., & Phily, c. (2014). Making Health Microinsurance Work: Ten Recommendations for Practitioners. *International Labour Organization* .
- IRA. (2014). *The Kenya Microinsurance Policy Paper*. Nairobi.
- Jong, P. d., & Heller, G. (2010). Generalized Linear Models for Insurance Data. *Cambridge Books Online* , 64.
- Kastelijn.W.M, & Wit, G. W. (1977). An Analysis of Claim Experience in Private Health Insurance to Establish a Relation between Deductibles and Premium Rebates. *Cambridge Journals* , 16.
- Kimbal, M., Phily, C., Folsom, A., Lagomarsino, G., & Holtz, J. (2013). Leveraging Health Microinsurance to Promote Universal Health Coverage. *International Labour Organization* .
- KNBS. (2015). *Economic Survey 2015*. Nairobi: KNBS.
- KPMG. (2013). *Insurance in Africa*.
- Leatherman, S., Christensen, L. J., & Holtz, J. (2010). Innovations and barriers in health microinsurance. *ILO's Microinsurance Innovation Facility* , 112.
- Makove M., S. (2011). African Policy Approaches: Microinsurance in Kenya . *AIO – A2ii Regulator' s Workshop*. Victoria Falls, Zimbabwe.
- McCord, M., Steinmann, R., Tatin-Jaleran, C., Ingram, M., & Mateo. (2012). The landscape of microinsurance in Africa 2012. *Munich Re Foundation* .
- McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models*. Boca Raton: Chapman & Hall.
- NHIF. (2015). *Accredited Hospitals*. Retrieved July 22, 2015, from NHIF: <http://www.nhif.or.ke/healthinsurance/>
- NHIF. (2015). *Premium Rates*. Retrieved July 22, 2015, from NHIF: <http://www.nhif.or.ke/healthinsurance/>

Rodríguez, G. (2015). *Generalized Linear Models*. Retrieved June 22, 2015, from Introducing R : <http://data.princeton.edu/R/glms.html>

Sarkar, R. (2012). Analysis of human immune responses in quasi experimental settings: tutorial in biostatistics. *Biomedical Central Journals* , 1471 - 2288.

Sen, G. &. (2008). Gender inequity in health: why it exists and how we can change it. *Global Public Health* , 1 - 12.

Smith, A., Chamberlain, D., Smit, H., & Grieve, S. N. (2010). *Kenya microinsurance landscape: market and regulatory analysis*.

The World Bank Group. (2014, October 28). *Improving Health Care for Kenya's Poor*. Retrieved July 3, 2015, from The World Bank: <http://www.worldbank.org/en/news/feature/2014/10/28/improving-healthcare-for-kenyas-poor>

The World Bank. (2014, October 8). *Improving Health Care for Kenya's Poor*. Retrieved July 2015, from The World Bank: <http://www.worldbank.org/en/news/feature/2014/10/28/improving-healthcare-for-kenyas-poor>

Transparency International. (2011). *The Kenya Health Sector Integrity Study Report*. Nairobi: Transparency International - Kenya.

WHO. (2015, July). *Universal health coverage*. Retrieved July 30, 2015, from World Health Organization: http://www.who.int/universal_health_coverage/en/

World Bank. (2014). *Data*. Retrieved 2015, from The World Bank: <http://data.worldbank.org/country/kenya>

Yao, Y. (. (2013). Development and Sustainability of Emerging Health Insurance Markets: Evidence from Microinsurance in Pakistan. *The Geneva Papers* , 160 - 180.