INSTITUTE OF DISTANCE LEARNING

KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

MODELING OCCURRENCE OF DEATH IN GHANAIAN FEMALE PENSIONERS

by

GLORIA ALA ADJETEY-DIN (BSC. MATHEMATICS)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS,

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR

THE DEGREE OF MASTER OF SCIENCE, ACTUARIAL SCIENCE

October 2016

Declaration

I hereby declare that this submission is my own work towards the award of the MSc. 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.

Gloria Ala Adjetey-Din (PG) ……………………. …………………
Student Signature Date

Certified by:

Nana Kena Frempong ……………………. …………………
Supervisor Signature Date

Certified by:

Prof. S. K. Amponsah ……………………. ……………………..
Head of Department Signature Date

Dedication

I dedicate this work to my lovely family

Abstract

The study considered count data with excess zeros and over-dispersion. The models used for this situation of the study data are the zero-inflated negative binomial (ZINB) and the negative binomial (NB) regression model. This was to establish whether the parameter age is a function of mortality for the various cohort groups. Also the Akaike Information Criteria (AIC) goodness of fit testing was used to determine a best fit for the cohort groups with the various model. The ZINB regression model best fitted the female pensioners" data for the 2005 cohort group while the NB regression model best fitted the 1990 cohort group. Other parameters like causes of death and standard of living after retirement must be considered in future model.

Acknowledgement

Special thanks to my lovely and caring husband Joe Ala Adjetey-Din who motivated, took good care of our children any time am away and also supported me financially throughout the programme. And to my wonderful and lovely daughters Elizabeth Naa Astwei Adjetey-Din, Dorothy Naa Ayorkor Adjetey-Din and Rebecca Kai Adjetey-Din for their understanding and co-operation. To my sister Dorothy Teye-Doe for her timely financial assistance, I say thank you and God bless you richly. Also to Mr. Moses Amoah for your encouragements and financial assistance, I say thank you and God bless you. To my entire family and friends who supported me through prayers and motivations, thank you very much.

Many thanks go to Nana Kena Frempong for supervising my thesis and also helping me out with my data analysis, God richly bless you. Thank you to Yaw Ayekple for the inputs and assistance through the data collection to analysis. To all lecturers at the Statistics and Mathematics Department, KNUST I say thank you so much for your inputs and corrections.

To Dr. Mettle, Dr. Samuel Idi, Calous and Enoch Nii Boi all of University of Legon, I say a very big thank you for your assistance and input during my data analysis, God bless you richly. Also to Robert Anum Modin, thank you very much for always holding the forte when am busy with my studies. And to members of Presec-Osu Mathematics Department, I thank you for always holding the forte in my absence throughout my programme. Finally, to Josephine Aba Quainoo for your support both in cash and in kind, thank you so much.

List of Table

List of Figures

Table of Contents

CHAPTER ONE

INTRODUCTION

The background of the study, a statement of the problem, objectives of the study, the significance of the study and the organization of the study were discussed in this chapter.

1.1 Background to the study

Social roles of women in Ghana varied throughout history. The overall impact of women in Ghanaian society has also been significant as their way of life by culture, location and generation. ("Women in Ghana", 2013)

During the pre-modern era women in Ghanaian society were seen as bearers of children, retailers of fish and farmers. Also in the traditional setup, the ability of women to bear children was explained as the means by which ancestor lineages were allowed to be reborn. ("Pre-modern era women", 1994)

Rates of female-headed households are on the increase in Ghana. The number of female-headed households who are widowed, divorced or unemployed has also increased. Contrary to worldwide findings that female poverty is correlated with higher rates of female-headed households, findings from the Ghana Living Standard Survey indicate that female-headed households may not actually experience higher poverty than male-headed households. ("Women in Ghana", 2013)

The proportion of the female elderly population is 56% as compared with 44% of the male elderly population, an indication of higher life expectancy of the female population. Almost three quarters of the female elderly compared to less than half of the males are illiterate. Also about three quarters of the female elderly compared to less than half the males have no formal education. There are disparities in educational attainment which also exist between the elderly in the rural and urban areas. The resistance to female education stemmed from the conviction that women would be supported by their husbands. The proportion of the elderly males who are economically active is slightly higher than that of the females. (Population Census, 2010)

During pre-modern Ghanaian society, in rural areas where noncommercial, agricultural production was the main economic activity, women worked on the farm. The coastal women also sold fish caught by men. The financial benefits that these women accrued went into the upkeep of the household. ("Pre-modern era women", 1994) Starting the mid-1970s to the early 1980s, however, urban market women, especially those who specialized in trading manufactured goods, gained reputations for manipulating market conditions and were accused of exacerbating the country‟s already difficult economic situation. Today, women make up 43.1% of economically active population in Ghana, the majority working in the informal sector and in the food crop farming. About 91% of women in the informal sector experience gender segregation and typically work for low wages. Within the informal sector, women usually work in personal services (2010 PHC).

Most working Ghanaian women are self-employed due to the low level of education. ("Women in Ghana", 2013) With only a hand full of females with well paid jobs occupying managerial positions. This has an effect on the number of females who go on retirement every particular year. A growing challenge for many nations is population ageing. As birth rates drop and life expectancy increases and ever-larger portion of the population is elderly, this leaves fewer workers for each retired person. A retirement plan is an arrangement to provide workers with an income during pension when they no longer earn a steady income from employment.

The pension is an arrangement by which an employer and an employee pay money into a fund that is invested to provide the employee with a benefit retirement plan that provides monthly income by companies or governments. According to SSNIT (2008), to qualify for full pension one must be at least 60 years and must have contributed for a minimum of 180 months in aggregate. The basis of calculation of pension depends on age, average best three year salary and earned pension right (i.e. Number of months one has contributed to the scheme). The pension is paid to the employee from the beginning of the retirement until death.

Looking at the number of years one works before retiring, one is expected to be fully established to be able to rest after hard working years and enjoy the benefits. However, due to socioeconomic challenges and poor working conditions pensioners rather become stressed out due to high demands since the pension cannot cater for their basic amenities. At pension people still have lots of responsibilities to take care of. All these expose pensioners to some form of mortality risk. How long these pensioners live is clearly a crucial factor in the ultimate cost of providing pensions. There are two elements to assessing life expectancy; first, and most important the current level of mortality and secondly the possible future path of mortality rates.

According to World Bank (2013) the mortality rate for female adult (per 1,000 female adults) in Ghana was measured at 233.49. The adult mortality rate is the probability of dying between the ages of 15 and 60; that is the probability of a 15 year old dying before reaching age 60 is subject to current age-specific mortality between those ages.

Over more human existence, mortality has shown a decreasing pattern (Berin, Stolnitz and Tenebein 1999; Friendland 1998, page 49). Sharp improvements have been experienced in particular over this century (Charlton, 1997; Goss, Wade and Bell 1998). An improvement scale assumes that agespecific death rates fall by a fixed percentage per year and this percentage is very often estimated by trending past observations, coupled with some "professional judgment". These improvements pose a challenge for pricing and reserving in life insurance and the management of public pension (James C. Hickman**,** 2007). For public policy, as well as the management of financial institutions, it is important to forecast future mortality rates. It is important to be able to accurately measure changes in mortality over time since the policyholders‟ benefits depend on survival.

When mortality rate used to be very high the cause was not due to genetics but rather epidemics, malnutrition, wars etc. Now, due to improved health facilities and medical procedures, innovations and technological improvement, life expectancy is increasing and research in mortality improvement scale indicates that it will increase further in the near future. A continuous mortality investigation will give estimates and projections of mortality rates and forecast using actuarial models. To make progress in predicting the frequency of death occurrences, life expectancy has been increasing over the years in several countries. The implication on mortality reduction have been widely recognized and affirmed. However, the predictions of mortality improvements made since the late 1960s have generally substantially underestimated actual mortality improvements (Continuous Mortality Investigation Mortality Sub-committee- CMI, 2004).

Of late the approach to model mortality improvements using CMI methodology, are based on projections by age, period and cohort effects. It is very important to have information on the estimation of life chances and distribution of death statistic nationwide. However, there are various uncertified or undocumented death cases which make forecasting very tedious and complex.

The occurrence of death in Ghanaian female pensioner has excess zero count outcome. The two popular models for count data are the Poisson when counts are unbounded and the binomial model for bounded counts.

1.2: Problem of the study

SSNIT continue to pay monthly benefits to pensioners from the date of retirement until death. Deaths occurrences recorded were not always reported, but at times assumed when pensioners‟ failed to renew their pension certificate. Also at age 73 and above most death occurrences are not reported.

Occurrence of death data has a count outcome variable which can be zero or non-zero. The zeros count can be due to two distinct processes; the first by sampling zeros that occurred by chance that can be assumed as a result of a dichotomous process and the second due to true zeros which are inevitable which are part of the counting process. The main problem is the choice of model to use to best fit between the observed data and the predicted values.

Researchers often used Poisson regression to model count data. There is a challenge to use Poisson Regression Model for count data with excess zeros which leads to over dispersion. So Negative Binomial (NB) regression model and Zero Inflated Negative Binomial (ZINB) regression model was considered.

1.3: Objectives of the study

The researcher seeks to answer the problem using the following objectives:

1. to model the occurrence of death as a function of age in each cohort using Negative Binomial (NB) regression model and the Zero Inflated Negative Binomial (ZINB) regression model
2. to determine which of the modeled cohort group best fit the data of the study modeled with zero inflated negative binomial (NB) and Zero Inflated Negative Binomial (ZINB) regression models using AIC

goodness-of-fit test

1.4: Significant of the study

The mortality rate is needed by Actuaries to perform actuarial calculations. Predicting future mortality rates are a problem of fundamental importance in the insurance and pension industry. This is why the Institute of Actuaries, UK has a committee (Continuous Mortality Investigation – CMI) in place to continuously investigate mortality and make projections. It will inform pension manager and insurance companies in premium pricing and pension plans.

The SSNIT pension fund calculates pension benefits on the basis of age, best three year salary and number of years worked. The scheme pays benefits to pensioner until death. This research will inform fund managers to be able to calculate their reserves to take care of the policy holders. Policy makers will be informed about the pensioner welfare through the observed mortality experience. It will also aid the government in implementing appropriate welfare programmes for the aged.

There have been several researches like mortality forecasting and Mortality Trends of Males and Females over the Ages on mortality where the method of modeling has been either Gumpertz or Lee Carter. However, the method of modeling using zero inflated negative regression model and negative binomial regression model have not been used in any literature to model pensioner mortality. Therefore, this research will also add to literature as there is little relevant literature in this area.

1.5: Limitation of the Study

1. Accuracy of data cannot be guaranteed as the data used are solely secondary data

2. Reported death at older ages is scanty

3.` Some individuals present unrealistic date of birth

1.6: Organization of The Study

This section of the study is partition into five chapters. Chapter one is the introduction discussed the background of the study, a statement of the problem, objective of the study and organization of the study.

The chapter two of the study reviews related works to the study. It also reviews some theories and actuarial models.

Chapter three of this study discussed the scope the data, data collection and methodology of models used to achieve the objectives and analysis of the findings. The methods used in this study are; exposed-to-risk, calculation of crude mortality rates, description of female pension data, Poisson regression model, zero inflated Poisson (ZIP) regression model, negative binomial (NB) regression model, zero inflated negative binomial (ZINB) regression model, the likelihood function model selection and AIC goodness-of-fit test.

The fourth chapter outlined the data analysis which discussed the mortality experience and description of the data. The chapter also looked at the data output findings of results and discussions. .

Finally the chapter five elaborate on the summary of the findings. It has also drawn conclusion from the discussions and suggested recommendations.

CHAPTER 2

LITERATURE REVIEW

2.0: Introduction

This chapter reviewed some related work that had been done. It also reviewed some of the methods and models that were used for the thesis.

2.1: The Elderly in Ghana

Category of adults who have attained advanced ages is the concept elderly. The demographic understanding of the concept refers to persons aged 60 or 65 years and above. In the developed countries where life expectancy is high (more than 70 years) and the age of retirement from active public economic activity is 65 years, the elderly are defined as persons aged 65 years and above (Population Reference Bureau, 2012). Developing countries on the other hand, since life expectancy at birth is lower, around 60 years (though it is gradually increasing), the age of retirement is 60 years and the elderly are considered as persons aged 60 years and above. At the international level, age 60 is now being used as the cutoff age for defining the elderly (United Nations Population Fund (UNPF) and Help Age, 2012).

The concept of ageing is an associated word that refers to the process by which persons or adults are attaining chronological ages that are classified as old ages. Ageing also refers to a process by which the elderly or older persons constitute a higher proportion of the total national population and initial period. This process has been a characteristic of the populations of the developed countries Weeks 2012for a long time, but is now being experienced by developing countries. (Ham-Chandeet al. 2009). When the proportion of the aged in a total national population is more than 10% or 15%, the population is referred to as an older population.

Population ageing is occurring as a result of declining fertility rates and increasing life expectancy. The efforts made by the UN to address the challenges, began three decades ago when in 1983 it held the First Assembly on Ageing in

Vienna and later designated 1999 as "The Year of the Older Person" and October 1st as the day of the aged, which was spearheaded by the UNFPA in 2012. The adoption of the Madrid Plan Action marked the beginning of taking concrete steps towards addressing the needs of the elderly with a globally concrete effort. (2012)

Ghana"s response to the UNFPA recommendation led to the adoption of the National Ageing Policy for the country in July 2010. The policy has the elements of the Madrid Plan of Action that qualifies its sub-caption to be rendered as „Ageing in Security Dignity. But for Ghana and other sub-Saharan African countries, their present socioeconomic conditions can militate against the political will that will be required to implement the policy and thereby contribute to the well being of the elderly. The process of ageing is taking place in an era in which the traditional systems that support elderly care have been transformed by the processes of modernization and globalization and in the absence of public welfare systems. Furthermore, knowing the situation of the elderly can contribute to the adoption of programs and other forms of intervention that can ensure that the aged in Ghanaian society enjoy a life of security and dignity.

Demographic and Socioeconomic Characteristics of the Elderly

According to the 2010 Population and Housing Census report the population of the elderly has increased by more than sevenfold since the 1960 census, rising from 213,477 in 1960 to 1,643,381 in 2010. The proportion of the female elderly population is 56% as compared with 44% of the male elderly population, an indication of higher life expectancy of the female population. A higher proportion of the elderly population (54%) resides in the rural areas whilst 47% of the females and 44% of the males are residing in the urban areas. Almost half (49.1%) of the females as compared with the 8.8% of the males were widowed, an indication of the highest female life expectancy. However, a higher proportion of the men tend to remarry, even at advanced ages, as compared with their female counterparts. ("The Elderly in Ghana", 2013)

Literacy and Educational Attainment

The 2010 Population and Housing Census report also reveal that literacy rate and educational status is generally low among the elderly and their various subgroups. Sex differentials in educational status are very wide among the elderly. Almost three quarters of the female elderly compared to less than half of the males are illiterate. Also about three quarters of the female elderly compared to less than half the males have no formal education. There are disparities in educational attainment which also exist between the elderly in the rural and urban areas. About 13% the females attained the highest education up to the Middle/JSS/JHS level whilst a negligible proportion has higher education (tertiary). ("The Elderly in Ghana", 2013)

Economic Activity Status of the Elderly

Relatively about 58.8% of the elderly are economically active. Out of this 74.4% of the elderly within 60 and 64 years old are economically active compared with 73.1% of those in the working age group 15 – 59 years. The proportion of economically active, however, declines to as low as 38.5% among the elderly at age 80 years and above. The proportion of the elderly males who are economically active is slightly more than that of the females. Most of the elderly who are economically active are engaged in the private, informal sector while 5.8% are engaged in the formal sector. Reintegrating the elderly into the economic productive sector should be expected to be successful. ("The Elderly in Ghana", 2013)

Living conditions of the Elderly

Most of the elderly (71.1%) live in a dwelling owned by a household member and almost 23% resides in houses that are owned by a relative. Ownership of the house is quite low among the elderly, considering that the elderly are expected to have their own houses and command some respect accordingly. The young, old, especially, have the lowest proportion that owns a house. A high proportion of the elderly reside in dwellings, particularly compound houses, with limited access to sanitation facilities and amenities. As high as 22% of the elderly have no access to a toilet facility and 34% use public toilet facility. ("The Elderly in Ghana", 2013)

2.2: Mortality

Survival date methods have been developed to analyse the problems encountered when studying the occurrence of events in time. Several aspects of time that may be important are: i. age ii. calendar time iii. time since first exposure to some influence or treatments. Problems are specified in terms of one time scale, the values of which will be denoted by *t*. A random variable denoting the time of occurrence of an event of interest (death) will be denoted by *dx*. The factors that affect mortality includes age, sex, standard of living occupation, nutrition, housing, climate, education and genetics.

Pension data consist of count variable outcome interest which might contain too many zeros. And with this count data the expected number of occurrence of death is the dependent variable and the age is the predictor variable. Different models were proposed to fit count data with too many zeros than expected: Lambert (1992) described the zero-inflated Poisson regression models with an application to defects in manufacturing; Hall (2000) also described the zero-inflated binomial regression model and incorporated random effects into ZIP and ZIB models.

Many count datasets have the joint presence of excess zero observations and long right tails features that may be accounted for by over-dispersion in the data, which are both relative to the Poisson assumption, Gurmu and Trivedi (1996). The proportion of the zeros increases whenever there zeros are too many relative to the Poisson assumption, so the Negative Binomial Regression And Zero-Inflated Negative Binomial Regression Model tend to improve the fit of the data. The model selection was done using AIC and BIC test to choose the one that best fit the study data.

Mortality has shown a decreasing pattern over the year (Berin, Strolnitz and Tenebein 1999; Friedland 1998, p. 49). Also over the century, sharp improvements have been experienced (Charlton, 1997; Goss, Wade and Bell 1998). The subject of mortality improvements has been the focus of much work in recent years.

2.3: Pensioners' Mortality

In 2008, the Research Committee of the Canadian Institution of Actuaries (Institute) formed the Pension Experience Subcommittee to: Review pensioner mortality experience in Canada and develop and maintain Canadian pension mortality tales and improvement scales. The primary objective of these studies was to build base mortality tables and mortality improvement scales and may be used for actuarial valuations for funding and/or financial reporting purposes for a broad range of Canadian pension plans.

Furthermore, it was expected that such tables and scales may be considered for use under actuarial standards of practice for the determination of pension commuted values and the division of pension benefits on marriage breakdown. This report presents a set of mortality tables based primarily on the experience observed from the RPP Study and mortality improvement scales based on the experience observed from the C/QPP Study and assumptions used in the 26th CPP Actuarial Report.

Mortality Tables And Mortality Improvement Scales

In the RPP Study, the mortality experience for calendar years 1999 to 2008 of a subset of Canadian public sector and private sector registered pension plans was reviewed. Based on the results of the RPP Study, the following base male and female mortality tables for the year 2014 are provided:

- 2014 Mortality Table (CPM2014) --- based on the separate experience exhibited under the public sector plans included in the Study
- 2014 Public Sector Mortality Table (CPM20) --- based on the separate experience exhibited under the public sector plans included in the RPP Study
- 2014 Private Sector Mortality Table (CPM2014) --- based on the separate experience exhibited under the private sector plans included in the RPP Study

As part of the RPP Study, the subcommittee reviewed the mortality experience by industry. Generally, there are insufficient data to develop mortality tables by industry. However, data on actual to expected (A/E) ratios, relative to the CPM tables, by industry may be useful to actuaries where specific plan experience or similar plan experience is not available. The RPP Study, and the C/QPP Study, identified significant experience variation by the size of pension. Accordingly the subcommittee developed size adjustment factors that can be used with the table mortality.

It is expected that practitioners will adopt a table that is most reasonable and appropriate in the circumstances of the particular plan under review. The subcommittee believes that it is best practice to consider whether modifications to the based tables are appropriate to reflect actual, credible experience of the pension plan under review. If lacking fully-credible experience, the subcommittee suggest that the actuary might consider using experience from other similar plans, the RPP Study size adjustment factor and/or industry data for adjusting the base table.

Mortality Improvement Scales

The C/QPP Study reviewed the trends of mortality experience since 1967, the first year that pensions became payable under those programs. Based on the results of the C/QPP Study, the following male and female improvement scales are provided:

- CPM Improvement Scale B (CPM-B) – improvement rates by age that decrease in a linear fashion for years 2012-2030 and ultimate rates applicable for all years after 2030

- CPM Improvement Scale B1 – 2014 (CPM – B1D2014) – improvement rates by age only designed to approximate the CPM Improvement scale B for pension valuations in 2014 and 2015.

The subcommittee proposes that practitioners consider adopting the twodimensional mortality improvement scale, CPM-B. However, the subcommittee recognizes that some pension valuation and administration system may not currently accommodate a two-dimensional scale. Based on these considerations, the subcommittee also developed the transitional, one-dimensional (age only), gender-specific mortality improvement scale, CPM-B1D 2014, that approximates in the near term the financial effect of the two-dimensional scale, assuming both sets of rates are applied on a generational basis.

For each age, the mortality improvement rates developed for the onedimensional scale take into account the evolution of improvement rates anticipated over the next several decades. The two-dimensional scale assumes a slowdown in mortality improvement after 2014 compared to earlier years. As such, it may be inappropriate to apply the one-dimensional scale for the purpose of actuarial valuations after 2016 since it may result in an overstatement of
actuarial liabilities.

Notation for mortality rates and improvement rates by year does not appear to be standardized within the profession. The subcommittee uses the following definitions, which incidentally were also used by the Society of Actuaries in connection with the two-dimensional Scale BB.

q_x^y --- means the improvement rate in mortality for persons aged *x* nearest birthday at the beginning of the calendar year *y*, will die before reaching the end of calendar year. Note that both *x* and *y* are defined at the beginning of the one-year period. I_x^y --- means the improvement rate in mortality for persons aged *x* nearest birthday at the start of the calendar year *y* – 1 to those aged *x* at the start of the calendar year *y*. In this case *x* is constant through the one-year period and *y* defined at the end of the period. $q_x^y = q_x^{y-1} (1 - I_x^y)$

Development of Mortality tables and size Adjustment factors

The Institution commissioned MIB Solutions to gather data from Canadian pension plan contributors on lives covered by their pension plans. Nineteen contributors submitted data for calendar years 1999 to 2008, from both public and private sector, for active lives for pensioners and for beneficiaries after the death of pensioners. Not all contributors provided data for all years and one contributor subsequently withdrew from the study.

Not all data submitted by contributors were of uniformly high quality. Individual records were excluded if they had been flagged by MIB Solutions as excluded. If a record was marked as unresolved, all records for that life were excluded. Also not all contributors provided sign-off to MIB Solutions indicating their agreement that he date were sufficiently accurate. All records for a contributor were rejected for a particular year if any of the following criteria were met;

- Unresolved records exceeded 10% of the number of deaths in the year ;

- The A/E ratio based on annualized pension was an outlier by more than three standard deviations;
- The number of deaths in the year was less than 20.
- The A/E ratio based on annualized pension was an outlier by more than three standard deviations; or
- The number of deaths in the year was less than 20.

Mortality experience varies significantly by industry. However, the data submitted to the RPP Study is not distributed by industry in the same proportions found in the full population of Canadian pension plans. For example, education is over-represented while construction and finance are under-represented in the data. The subcommittee decided to adjust the data by industry so that it would be more representative of Canadian pension plans membership.

Size Adjustment Factors ---- RPP Study

It is evident from both the C/QPP Study and RPP Study that mortality rates vary significantly with size of pension (other factors being equal). Size adjustment factors were derived that reflect the difference in that RPP Study experience by income band, for the males and females separately. When considering the possible application of the size adjustment factors, actuaries should be aware that:

- Pension amounts on which the mortality tables and size adjustment factors are based include bridge benefits and indexed pensions where applicable;
- There is no indicator in the data to whether included plans are closed to future accruals.

All calculations in this report employing size adjustment factors use the factors as stated to age 85, grading linearly to 1.0 for over 100 and higher.

The subcommittee believes that it is generally a satisfactory approximation to use the factors for all ages rather than using the more complicated grading at high ages.

Development of Mortality Improvement Scales

Assumptions in respect of future mortality improvement rates are subject to a high level of uncertainty. These rates are subjective as they vary by income, level of education and place of residence. The RPP Study did not have enough data to produce a mortality improvement scale but the C/QPP Study did in its Phase III report. The male, female Mortality Improvement Rate Charts which illustrate that the Scale AA Improvement Scale is too low and that actual mortality improvement has been much higher in Canada since 1967. The genderspecific improvement scales were developed as follows:

- Short-term rates applicable to years 2000-2011 are set equal to the smoothed 10 years experience based on the C/QPP income class 4 (35 percent of maximum pension and above) from the C/QPP for ages 65 and higher.
- Short-term rates for years 2000-2011 for ages 50 are set equal to the CPP assumption for 2010 as reported in 26th CPP Actuarial Report. Note there are no mortality rates available at these younger ages;
- Short-term rates for years 2000-2011 for ages 51-64 are linear interpolations between the above rates for ages 50 and 65;

- Ultimate rates (applicable for years 2030 and beyond) for ages 0-114 are set equal to the CPP year 2030 actuarial assumptions for those ages, as disclosed in the 26th CPP Actuarial Report;
- Rates for ages 115 and higher are zero;
- Rates for years 2012-2019 are derived by linear interpolation between the short-term rates and the ultimate rates.

The subcommittee also encourages the use of the two-dimensional improvement scale versus the one-dimensional table provided for the year 2014 and 2015

2.4: Review of related literature on ZINBRM and NBRM

This section of the study review related works in different areas. The research was on Analyzing Death Rate of Age Model with Excess Zeros using Zero Inflated Negative Binomial and Negative Binomial Death Rate: Mortality AIDS CoInfection Patients, Kelantan Malaysia. In this study the researchers used a secondary data from Kota Bharu, Kelantan Malaysia on death of AIDS which consist of 945 measurement of gender, national, race, marital status, occupation and mode transmission. (QQE 2012)

The data analysis was done by accessing the high zeros using the Negative Binomial and Zero-Inflated Binomial regression models with death rate dependent variable employed. The goodness of fit statistic did not indicate over dispersion exist in the study using the deviance and Pearson Chi-Square test. And the best model that fit was selected using the Akaike Information Criteria (AIC) and the

Bayesian Information Criteria (BIC). At the end the model which has the small value of AIC and BIC was accepted as a good model and that was the ZINBDR.

Though the model ZINBDR was chosen, there was a possibility that not all the assumptions for this model were met. More especially in regard to the underlying dual-state distribution of this model.

Famoye1 and Singh (2006), in their journal used Zero-Inflated Generalized Poisson Regression Model with an Application to model Domestic Violence Data. In their research work they observed that the zero-inflated negative binomial regression models couldn"t fit the data sets even when the zeroinflated models were inadequate. This major problem in these cases was the fact that the iterative technique to estimate the zero-inflated negative binomial regression model parameters failed to converge. However, they realized that this observation was similar to the one Lambert (1992) made and they quote : "of course, inflating a negative binomial model with „perfect zeros" might provide an even better model for the printed –wiring-board data than ZIP regression doe.

Such a model was not successfully fit to these data, however." Hence they were motivated to develop a zero-inflated generalized Poisson regression model for modeling over-dispersed count data with too many zeros.

The model parameters were estimated using the maximum likelihood method. Then the number of zeros was tested to whether it is too large for the generalized Poisson model to adequately fit the domestic violence data using a score test. They observed that though the ZIGP regression model was a good competitor of ZINB regression model, they don"t know the condition under which one will be better.

According to Fang (2013), he used Zero-Inflated Negative Binomial (ZINB) Regression Model for Over-Dispersed Count Data with Excess Zeros and Repeated Measures, An Application To Human Microbiota Sequence Data. His focus was on modeling the over-dispersed count data with excess zeros and repeated measure in human Microbiota sequence data. The data analysis employed ZINB regression model for over-dispersed count with excess zeros in the study data. However, he observed that the mixture model contains component to model the probability of excess zero values and the other negative binomial parameters which allow for repeated measures using independent random effects between these two components.

Also, Ismail *et al* in his paper focus on Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and ZeroInflated Generalized Poisson Regression Models. In this paper the study through the mean-variance relationship relates the negative binomial and generalized Poisson regression models through and suggests the application of these models for over dispersed or under dispersed count data.

Also, they suggests the application of these zero-inflated models for zeroinflated and over-dispersed count data by relating zero-inflated negative binomial and zero-inflated generalized Poisson regression models using the mean-variance relationship. The researchers used data from Malaysian OD claim count data and the German healthcare count data. They model the Malaysian OD claim count data using the negative binomial and generalized Poisson regression models. And also fitted the zero-inflated negative binomial and zero-inflated generalized Poisson regression models to the German healthcare count data.

Greene (March, 1994) paper is on Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. In the research he observed that there are excess zeros in the data. So he modified the Poisson and the negative binomial models for count data. This is to enable the models accommodate cases in which the number of zeros in the data exceed what would by predicted by any of the model.

To distinguish between the zero inflation and over-dispersion of the data he presented a new test procedure. Also, he developed a model for sample selection which is analogous to the Heckman style specification for continuous choice models. Finally, he presented an application to a consumer loan behaviour data in which both of the phenomena are clearly presented.

However, if the occurrences of count events depend on specific conditions and/or time, such as the case of deductible or no claim discount in insurance data, the zero-inflated models are more appropriate.

CHAPTER 3

METHODOLOGY

3.0: Introduction

This chapter discusses the scope of data, data processing and actuarial modeling process. This chapter also discusses the various methods adopted for this study. The study focused on pensioners and the occurrence of deaths.

3.1: Scope of Data

The study focused on the occurrence of death in Ghanaian female pensioners who retired from1990 to 2005 at SSNIT. These pensioners included those who retired voluntarily between 54 to 60 years and those who retired from 60 to 65years. All the pensioners were exposed to investigation from the day of retirement to 2010. Each one of the pensioners was observed from the age of retirement to 80 years and occurrence of death recorded over a year period. The study ended investigation pensioners at 80 years because deaths recorded after 80 years were insignificant which will affect the output of the analysis.

Secondary data from Social Security and Nation Insurance Trust (SSNIT) which consisted of 2,178 female pensioners was sampled from a five year period pension interval; 1990, 1995, 2000 and 2005. The total occurrence of deaths recorded within the five years interval period was 424.The table below describes the selection of the cohort groups:

Table 3.1: sample data from 1990 – 2005 with five year interval cohort group

Cohort group	Pensioners	Occurrence of death
1990	61	28
1995	358	162
2000	670	168
2005	1,089	66
Total	2,178	424

3.2: Data collection

The study uses a quantitative research to model the occurrence of death in Ghanaian female pensioners. Secondary data was obtained from SSNIT which consist of female pensioners‟ for the periods 1990 to 2010 age 55 to 80 years. The data contain information on the date of birth, death if any, and year of retirement of pensioners. The life certificates were updated pension year 02/06/2014 as at the time data was retrieved for the study purpose. However, any pensioner whose life certificate had not been updated as at that date was assumed dead until otherwise proved.

The general pensioners‟ population includes invalidity pensioners, hazardous workers pensioners, old age pensioners, and early retirees. But for this study purpose, the target population comprises of both old age and early valid retirees. The old age retirees were individuals who go on retirement at the normal retirement age of 60 years while the early retirees include individuals who voluntarily go on retirement from the ages 55 to 59 years.

In other to obtain a homogenous group of early retirees and old age pensioners for a period of five intervals pension years 1990,1995, 2000 and 2005, purposive sampling was employed to select individuals from the general pensioners population to form the cohort group.

The total sample obtained from the sampling consists of 2,178 female pensioners. These cohort groups of pensioners were grouped according to the year of retirement from age 55 to 80 years. For the pension year 1990 we have 61 pensioners,1995 pension year we have 358 pensioners, pension year 2000 we have 670 pensioners and 1,089 pensioners for 2005 pension year.

For the purposes of this study some of the data that had major inconsistencies were discarded. These inconsistencies included inaccurate or blank date of birth, retirement and death, very late entry into the pensioner category. On the average about 10% of the total population was excluded before arriving at the sample size stated. The total general population was about 120,000 pensioners for both male and female. As at the time of the study, pensioners who have not renewed their life certificate and have had their pension payments seized were assumed dead at the date of last update. Out of the 2,178 female pensioners selected from the general population 424 deaths were recorded.

For confidentiality purpose member identification numbers were removed and data were regrouped to have three essential details; date of retirement, date of death or last update and current age if still alive. Data was further sorted and regrouped to obtain in each target year, age at pension, number of deaths at each age, and the exposed to risk at each age. Pensioners were exposed to investigation from the pension year to June 2014 and were observed from ages 55 to 80 years. The investigation was done only up to age 80 years because after 80 years reported deaths were very scanty and to avoid distorted or misleading results.

3.3: Methodology

Secondary data was used for the research which gives the number of workers who retired at a certain age x to $x+1$ as the exposed (Ex) within the year. It also counts the number of pensioners who died in a particular year (dx). The crude mortality rate (qx) produced at a particular year is discrete and not smooth. Graduation is done to change the discrete to continuous and for smoothness using Poisson model. But the data on the female mortality has excess zeros which the Poisson model did not fit. A zero inflated Poisson (ZIP) logit model was proposed.

Exposure-to-risk (Ex):

The Ex denotes the number of person years lived during year by people aged x at the start of the year. Assuming that people who die during a year have on average been alive during half of the year, the exposed-to-risk can be approximated by the number of survivors plus half the number of deaths in this group. (Pitacco et al, 2009). The differences in observation periods are accounted for by the count model by including the log of the exposure variable in the model with coefficient constrained to be one. The exposure makes use of the correct probability distributions that is why it is superior in many to analyse rates as response variables. Also the exposure is used to adjust counts on the response variable and it is possible to various kinds of rates, indexes or per capita measures as predictors.

Production of Crude Mortality Rates for 1990, 1995, 2000 and 2005

The crude mortality rate for a given age for any given year is the probability that a person at age *x* dies that year. Crude mortality rates are usually calculated by simply dividing the relevant number of deaths by the number of life-years that were exposed to the risk of death over that period. The crude mortality rates for each plan year 1990, 1995, 2000 and 2005 were developed accordingly.

Description of Female Pension Data

Pension data is considered to be of the form of number of deaths and number of living pensioners who are exposed to death which are in cells by year of death and age at death. The study focus on the occurrence of death for a year which gives a count (discrete) variable outcome. A total of 424 deaths occurred within the five year interval period from age 55 to 80 years. The data was cleaned by discarding all pensioners who are over 80 years since much record was not recorded. The R software was then used to analyse the data by finding the descriptive statistics for each cohort group. The result from the output which shows there were excess of zeros with large variation was used to propose the model to be used for the data. The following models were proposed to model the data; zero inflated negative Poisson and negative binomial. Before discussing them let"s consider Poisson regression model and the zero inflated model. The response variable is the number of death that occurred in the year and is represented by ***y*** and the predictor variable is the age at which death occurred and is represented by ***x*.**

Models

Pension data consist of count variable outcome interest which might contain too many zeros. And with this count data the expected number of occurrence of death is the dependent variable and the age is the predictor variable. Different models were proposed to fit count data with too many zeros than expected: Lambert (1992) described the zero-inflated Poisson regression models with an application to defects in manufacturing; Hall (2000) also described the zero-inflated binomial regression model and incorporated random effects into ZIP and ZIB models.

Many count datasets has the joint presence of excess zero observations and long right tails features that may be accounted for by over-dispersion in the data, which are both relative to the Poisson assumption, Gurmu and Trivedi (1996). The proportion of the zeros increase whenever there zeros are too many relative to the Poisson assumption, so the negative binomial regression and zeroinflation negative binomial regression model tend to improve the fit of the data.
The model selection is done using the likelihood ratio test.

Poisson regression model

Poisson regression model is used to model count data. It is a discrete probability distribution that is used to model the number of events occurring within a given time interval. The Poisson distribution models the log-odds as a linear function of the observed covariates. This gives the generalized linear model with Poisson response and ling log.

If the number of occurrence has a variable *Y* which has a Poisson distribution with parameter μ and it takes integer values of $y = 0, 1, 3, \ldots$ then the probability distribution is given by

$$P(Y = y) = \frac{\mu^{y} e^{-\lambda}}{y!} \quad ; \quad \lambda > 0 \qquad 3.1$$

where λ is the shape parameter which indicates the average number of events in the given time interval.

The Poisson distribution has mean and the variance that can be shown as

$$E(Y) = var(Y) = \mu$$

If it is true that the mean is equal to the variance, then any factor that affects one will also affect the other. The Poisson distribution can only be applied under the following assumptions;

1. the event is something that can be counted in whole numbers;

2. occurrences are independent, so that one occurrence neither diminishes nor increases the chance of another

$$\log(\mu) = \beta_0 + \beta_1 x \qquad 3.3$$

where; ***x*** denotes the vector of explanatory variables and **β** the vector of regression parameters.

However, this model was not did not fit the data for the study since the mean is not equal to the variance even though it is a count data. This was due to the excess zeros in the data which were not sampling error but outcome. A ZeroInflated-Poisson was proposed.

Zero-Inflated-Poisson (zip)

The data that has excess of zero counts is model by zip regression model. Theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros are modeled independently. The zip model has two parts, the first part use Poisson to mode the count model and the second use logit model to predict excess zeros. Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros.

$$Pr(y_i = 0) = \pi + (1 - \pi)e^{-\mu} \qquad 3.4$$

$$Pr(Y_i = y_i) = (1 - \pi) \frac{\mu^{y_i} e^{-\mu}}{y_i}, \quad y > 0 \qquad 3.5$$

Where y_i is the outcome variable with any non-zero value, μ is the expected Poisson count for the ith individual and π is the probability of the extra zeros. The zip regression model has mean to be $(1 - \pi)\mu$ and the variance is $\mu(1 - \pi)(1 + \mu\pi)$. This model fit best if the data is not over dispersed with the mean larger than the variance.

Negative Binomial Regression Model (NB)

The negative binomial regression model is a parametric model that is more dispersed than the Poisson, which can handle the over dispersed situation in the data. Given y to be the respondent variable of the number of death occurrence in a year and that $y \sim$ Poisson (μ), whereas μ is a random variable with a gamma distribution. Now if

$$\frac{y}{\mu} \sim \ Poisson\ (\mu) \ \ and \ \ \mu \sim Gamma(\alpha, \beta),$$

Where the gamma distribution has mean $\alpha\beta$ and variance $\alpha\beta^2$, with probability density

$$P(\mu) = \frac{1}{\beta^{\alpha}\Gamma(\mu)} \mu^{\alpha-1}\exp(-\mu/\beta); \ \mu > 0 \qquad 3.8$$

Then the negative binomial with unconditional distribution of y is

$$P(y) = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y\,!}\left(\frac{\beta}{1+\beta}\right)^{y}\left(\frac{1}{1+\beta}\right)^{\alpha}, \ y \ = \ 0, 1, 2, \ldots \qquad 3.9$$

This distribution has mean

$$E(y) = E[E(y/\mu)] = E(\mu) = \alpha\beta$$

and variance $Var(y) = E[Var(y/\mu)] + Var[E(y/\mu)]$

$$= Var(\mu) + E(\mu) \quad = \quad \alpha\beta + \alpha\beta^2 \qquad 3.10$$

Expressing the negative binomial distribution in terms of the parameters

$\mu = \alpha\beta$ and $k = \frac{1}{\alpha}$, that the $E(y) = \mu$ and $Var(y) = \mu + k\mu^2$ (function is quadratic)

Therefore the distribution of y is given by

$$P(y) = \frac{\Gamma(k^{-1}+y)}{\Gamma(k^{-1})\,y\,!}\left(\frac{k\,\mu}{1+\beta}\right)^{y}\left(\frac{1}{1+k\,\mu}\right)^{\frac{1}{k}}, \qquad 3.11$$

Note that the negative binomial distribution approaches Poisson (μ) as $k \rightarrow 0$.

To model the negative binomial, let $y_i \sim$ Negative (μ_i, k) with the log link, so that

$$\text{Log}\ \mu_i = \beta_0 + \beta_1 x_1 + \ldots \quad \text{(for offset)} \qquad 3.12$$

Zero-Inflated Negative Binomial Regression

Data with excess zeros that uses the zero-inflated model assumes the outcome of the zeros is due to two different processes. The study data considered occurrence of death in Ghanaian female pensioners. the occurrence have two process; first that a pensioner death occurred which give a count outcome (nonzero death) and the second no death occurred which give a possible outcome of zero. The first part of the process which is the zeros is modeled by the logit whereas the negative binomial model is used to model the second part of the process which is the count.

Zero inflated negative binomial distribution is a mixture of distribution which assigns a mass of p to extra zeros and mass of $(1 - p)$ to a negative binomial distribution $0 \leq p \leq 1$. It is a continuous mixture of Poisson distribution with mean μ o be gamma distributed and modeled the over dispersion. For better understanding of the zero-inflated negative binomial regression, review the negative binomial model;

$$P(Y = y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\, y\,!}\left(\frac{\mu}{1+\beta}\right)^{y}\left(\frac{1}{1+k\,\mu}\right)^{\alpha},\ \ y = 0, 1, 2, \ldots; \mu, \alpha > 0 \qquad 3.13$$

Where μ = *E(Y)*, α is the shape parameter which quantifies the amount of over dispersion and the response variable of interest is *Y* and the variance of *Y* is $\alpha + \frac{\mu^2}{\alpha}$. The ZINB distribution is given by

$$P(Y = y) = \begin{cases} p + (1-p)\left(1 + \frac{\mu}{\alpha}\right)^{-\alpha}, & y = 0 \\ (1-p)\frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)}\left(1 + \frac{\alpha}{\mu}\right)^{-y}, & y = 1, 2, \ldots \end{cases} \quad 3.14$$

The zero inflated negative binomial distribution has mean E(Y) = (1 – p) μ and variance to be Var (Y) = (1 – p) $\mu\left(1 + p\mu + \frac{\mu}{\alpha}\right)$, respectively. Note that the zero inflated negative binomial distribution reduces to Poisson distribution if both $\frac{1}{\alpha}$ and $\theta \approx 0$.

Model selection

Comparing the two models to select the one that best fit the study data, the Akaike Information Criteria and the Bayesian Information Criteria were used. The model that has the lowest AIC and the BIC is selected to be the best fit.

Likelihood function

Suppose a set of parameter value θ, with given x outcomes, then the likelihood function is the probability of those observed outcomes;

$$\mathcal{L}(\theta|x) = P(x|\theta)$$

Suppose a given parameterized family of probability functions in the discrete distribution case;

$$x \mapsto f(x \mid \theta),$$

where θ is the parameter, the **likelihood function** is

$$\theta \mapsto f(x \mid \theta),$$

written

$$\mathcal{L}(\theta \mid x) = f(x \mid \theta),$$

with x being the observed outcome of the data. Alternatively, when $f(x \mid \theta)$ is viewed as a function of x with fixed θ, it is a probability density function, and when viewed as a function of θ with x fixed, it is a likelihood function.

From a geometric standpoint, if we consider $f(x, \theta)$ as a function of two variables then the family of probability distributions can be viewed as a family of curves parallel to the x-axis, while the family of likelihood functions are the orthogonal curves parallel to the θ-axis.

Akaike Information Criteria (AIC)

One way of selecting a model from a set of models is by using the Akaike Information Criterion (AIC). The model that minimizes the Kullback-Leibler distance between the model is chosen. The criteria seek a model that has a good fit to the true but few parameters. It is defined as: AIC = -2LL + 2 K

where likelihood is the probability of the data given a model and K is the number of free parameters in the model. AIC scores are often shown as ΔAIC scores, or difference between the best model (smallest AIC) and each model (so the best model has a ΔAIC of zero).

CHAPTER 4

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.0: Introduction

The data from SSNIT on female pensioners from the cohort groups 1990, 1995, 2000 and 2005 was analyzed using R software. The data sets give information on number of females who retired voluntarily and involuntarily on pension within a cohort group and the number of female pensioners who died within a one year period. This chapter focuses on the data analysis using the methods described in chapter 3 and discusses the results under the following headings: descriptive analysis of data, model1990, 1995, 2000 and 2005 cohort group using zero inflated negative binomial regression model, model 1990,1995, 2000 and 2005 cohort group using negative binomial regression model and compare the two models.

Researches in mortality often consist of count variables as an outcome. With the count data the variable death which represents the response or dependent variable is non-negative and the data has only one predicted variable age. The proposed model for count data is Poisson which holds when the conditional mean equals the conditional variance. But in real life situation this does not really happen. That is why the Poisson cannot be used which indicate that the data consist of excess zeros.

4.1: Descriptive Analysis Mortality Experience

The research investigates estimates of the level of mortality for members aged 55 to 80 years with retired status over a five (5) year interval period; 1990, 1995, 2000 and 2005. The resulting mortality rates represent, on the average, the mortality experience over the retired year.

Female Pensioners

Table 4.1: Female Pensioners data from 1990 – 2005, five year interval cohort group

Cohort group	Pensioners	Percentage (%)
1990	61	2.8
1995	358	16.4
2000	670	30.8
2005	1,089	50.0
Total	2,178	100.0

The number of female pensioners who retired at the end of each year from 1990 – 2005 five years intervals cohort group is shown in table 4.1 above. Out of 2,178 female pensioners‟ who were observed under the study, 61 pensioners representing 2.8% retired in 1990, 358 pensioners representing 16.4% retired in 1995, 670 pensioners representing 30.8% retired in 2000 and 1,089 pensioners representing 50 % retired in 2005. There is increase in the number of females who retired over the five year intervals.

Deaths

Table 4.2: Deaths Aged 55 – 80 years of female pensioners for the cohort group 1990, 1995, 2000 and 2005

55	0	0		0
56	0	0	1	0
57	0	0	0	0
58	0	1	1	0
59	0	0	1	2

Age	1990 (*dx*)	1995 (*dx*)	2000 (*dx*)	2005 (*dx*)
60	0	0	0	3
61	0	1	6	10
62	0	4	3	8
63	0	2	3	24
64	1	7	11	12
65	1	7	7	1
66	1	9	9	2
67	0	4	6	2
68	0	9	2	1
69	1	9	5	0
70	1	5	4	0
71	0	4	9	0
72	5	18	25	0
73	6	33	47	1
74	1	14	13	0
75	0	6	4	0
76	2	6	7	0
77	1	8	0	0
78	2	8	2	0
79	3	2	1	0
80	3	5	1	0
	28	162	168	66
percentage (%)	7	38	40	16

To estimate mortality rates, first we count the number of deaths by year over the period. The age at death determined at t age before the next birthday. Table 4.2 below represents the number and proportion of deaths for each year by age. Out of the 424 observed deaths over the period, 28 occurred in 1990 representing 7%, 162 occurred in 1995 representing 38%, 168 occurred in 2000 representing 49% and 66 occurred in 2005 representing 16%. From the table the age at which maximum number of deaths occurred is age 72 and 73 for the period 1990, 1995 and 2000. For the period 2005 the maximum number of deaths occurred at age 63 and 64 years.

Exposures

Exposure is the number of female pensioners who were exposed to the risk of death during the period of retirement at a particular age. The exposures were determined by the exact age method period to the earlier time of death or the end of the observed period. In this study the exposed to risk was calculated by adding the number at risk for the current year and the cumulative survivors from the previous year. The exposed to risk for the 1990-2000 cohort group increased from 55 years to 68 years since the number of deaths are less than the new female pensioners into the exposed. However, the exposed to risk decrease from 69 to 80 years as the number of deaths are more than the new female pensioner entrant. The 2005 exposed to risk increased from 55 to 73 years as the entrants of new female pensioners are less than the number of deaths.

Table 4.3below shows that the number of female pensioners over the five year period varies by age who are exposed to risk. Female pensioners" exposure increased rapidly over the five year interval period. The exposed to risk in 2005 calendar year is very high followed by the 2000 calendar and 1990 has the lowest.

Table 4.3: Exposed Aged 55 – 80 years of female pensioners for the cohort 1990 – 2005 for five years interval

Age	1990	1995	2000	2005

55	11	41	75	45
56	18	68	131	104
57	25	107	175	157
58	33	137	228	253
59	38	165	270	400
60	52	237	446	1922
61	57	293	555	2241
62	59	315	580	2388
63	59	323	602	2442
64	59	328	612	2463
65	59	326	615	2472
66	58	322	617	2478
67	57	318	613	2486
68	57	315	613	2487
69	58	308	611	2487
70	57	302	607	2492
71	56	298	603	2493
72	54	287	587	2495
73	48	262	551	2495
74	45	238	523	2494
75	44	228	514	2495
76	43	222	510	2495
77	42	215	506	2495
78	40	207	505	2495
79	38	202	504	2495
80	35	199	503	2495

Description of the data

A total of 2,178 female pensioners were observed over the period 1990, 1995, 2000 and 2005 cohort groups. Each group recorded the number of deaths; in the 1990 group there were 7 death, in the 1995 group there were 38 deaths, in the 2000 group there were 40 deaths and the 2005 group there 16 deaths. The highest death occurred in 2000 and 1995. It is obvious that if a person went on pension, the count could be zero (no death) and non-zero (death) at a particular age. The summary of the output data is discussed in tabular form and representation on a histogram.

Tabular Representation

Table 4.4 shows the average number of deaths by year. The minimum death for all the cohort group is zero. The 2000cohort group recorded the highest range (47), the 1995 cohort group recorded the second highest (33), the 2005 group recorded the third highest (24) and the 1990 group recorded the lowest range (6). The 2000cohort group recorded the highest mean, approximately 7 deaths over ages 55 - 80, followed by 1995 cohort group mean, approximately 6 deaths over ages 55-80 and the 2005 cohort group recorded mean, approximately 3 deaths over ages 55 – 80 and the 1990 cohort group recorded the lowest.

Considering the range and mean of the various cohort group, it is observed that the variation of each death from the mean is wide.

Table 4.4: summary of death count for the period 1990, 1995, 2000 and 2005

	1990	**1995**	**2000**	**2005**
Minimum	0.000	0.000	0.000	0.000
1st Quartile	0.000	1.250	1.000	0.000
Median	0.500	5.000	3.500	0.000
Mean	1.077	6.231	6.462	2.538
3rd Quartile	1.000	8.000	7.000	2.000
Maximum	6.000	33.000	47.000	24.000

Graphical Representation

A total of 424 deaths were recorded in between the period 1990, 1995 2000 and 2005 cohort group. Within each cohort group it was not all the ages under consideration that recorded death, some recorded zero deaths. As the number of occurrence of death is predicted there is the need to also predict the occurrence of excess zeros. The outcome is represented on a histogram with a plot of count against death.

Histogram representation of 1990 cohort group

Figure 4.1.a below display the plot of count against death for the 1990 cohort group. The count of zero death was 16 and the count of 1 death was 7 within the 1990 calendar year. Also there were 2 counts for 2 and 3 deaths. Likewise 1 count recorded for 5 and 8 deaths. On the whole the number of zeros outnumbered the non-zeros with minimal counts for higher deaths.

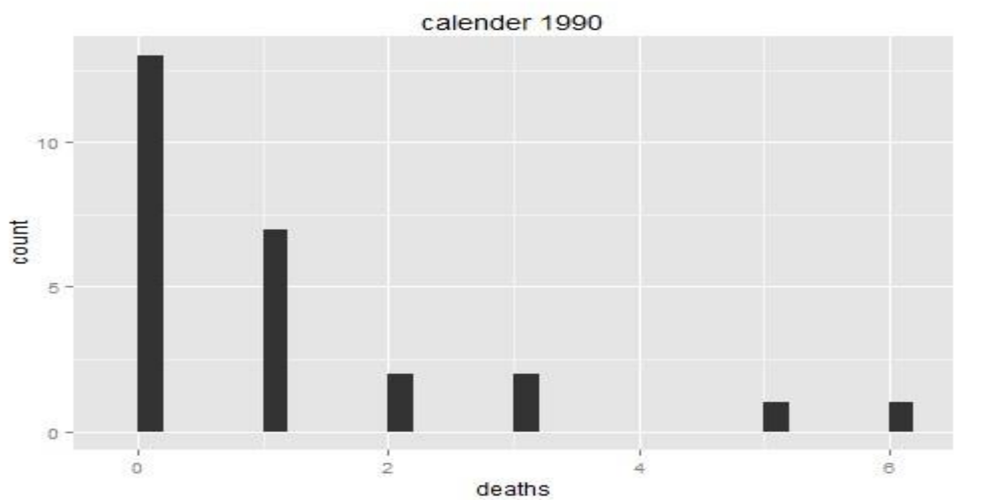

Figure 4.1a: count the number of death with 1990 calendar year

Histogram representation of 1995 cohort group

Figure 4.1.b above shows the plot of count against death for the 1995 cohort group. The count of zero death was 7 followed by the count of 4 and 10 deaths which was 3. Also the count of 5, 6, 7, 8, 9 deaths was 2 and the count of 13, 18, 34 deaths was only 1. On the whole the count of zeros outnumbered the non-zeros.

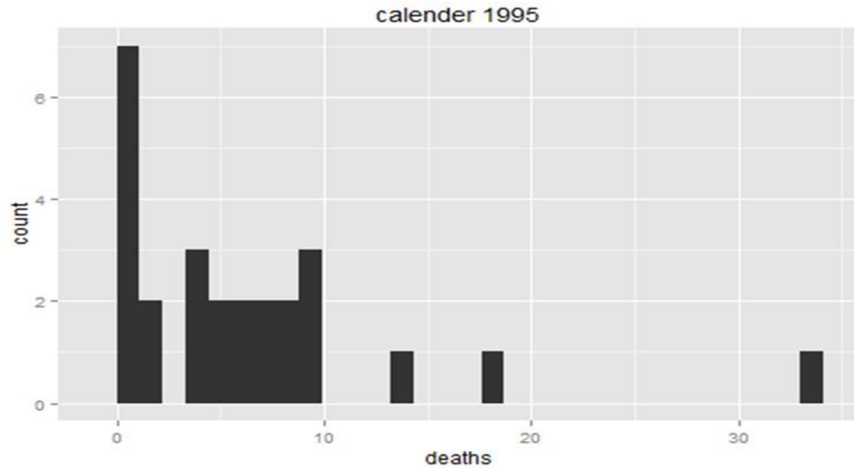

Figure 4.1b: count the number of death with 1995 calendar year

Figure 4.1.b above shows the plot of count against death for the 1995 cohort group. The count of zero death was 7 followed by the count of 4 and 10 deaths which was 3. Also the count of 5, 6, 7, 8, 9 deaths was 2 and the count of 13, 18, 34 deaths was only 1. On the whole the count of zeros outnumbered the nonzeros.

Histogram representation of 2000 cohort group

Figure 4.1.c below display the plot of count against death for the 2000 cohort group. The count of zero death was less than 10 and the count of 1 death was less than 5 within the 2000 calendar year. Also there were less than 2 counts for the other non-zero deaths. On the whole the number of zeros outnumbered the non-zeros.

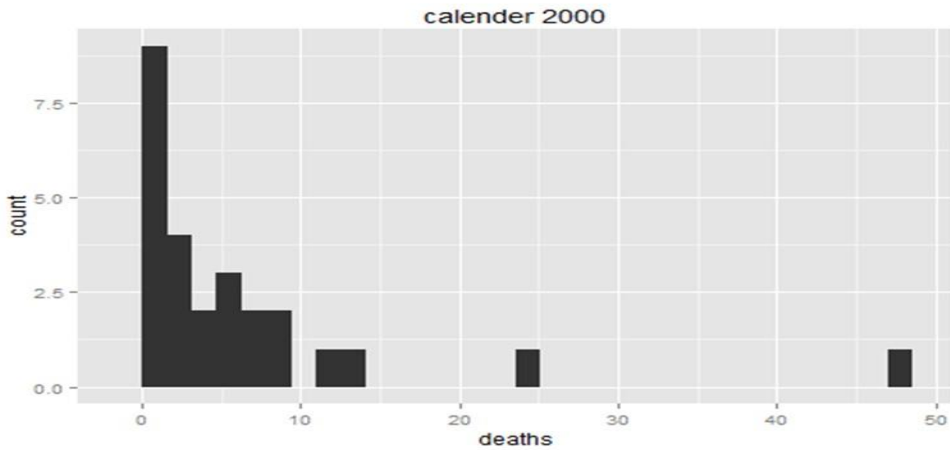

Figure 4.1.c: count the number of death with 2000 calendar year

Histogram representation of 2005 cohort group

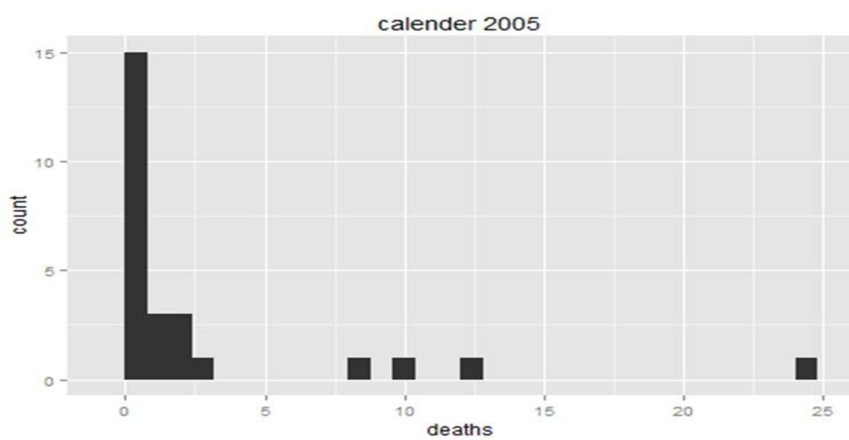

Figure 4.1d: count the number of death with 2005 calendar year

Figure 4.1.d above display the plot of count against death for the 2005 cohort group. The count of zero death was 15 and the count of 1 death was3 within the 2005 calendar year. We observed only 1 count for the other non-zero deaths. On the whole the number of zeros outnumbered the non-zeros. Also the deaths are highly sparse.

Summary

From figure 4.1.(a), (b), (c) and (d)the histograms above it could be observed that the count of zeros which represent no death in each calendar year far outnumbered the count of death. This could be a result of over dispersion and also note that data was highly non-normal. With the excess zeros one could consider the zero inflated Poisson regression model for this data. But with the zeros outnumbering the non-zeros shows the excess zeros are over dispersed. Therefore the zero inflated Poisson regression model could not be appropriate since it does better when data is not over dispersed. The models to consider for these data sets are the:

1. zero inflated negative binomial regression

2. negative binomial regression

4.2: Modeling the data

The summary from the descriptive data analysis shows that there were excess zeros in the data which are over dispersed. The models to be considered to for this data are the zero inflated negative binomial and negative binomial regression. Age was the variable to be used to model the count and the zero inflation. This is done by first estimating the model with the variables of interest.

Results and Discussions

Table 4.5: Estimates from zero inflated negative binomial and negative regression model

		Zero Inflated Negative Binomial			Negative Binomial		
Cohort	Age	Coefficient Estimate	Standard Error	P - Value	Coefficient Estimate	Standard Error	P - Value
1990	55 - 80	0.08540	0.04993	0.0872	0.148591	0.006257	< 0.0001
1995	55 - 80	0.06213	0.03258	0.0565	0.097710	0.001618	< 0.0001
2000	55 - 80	0-08256.	0.04908	0.0926	0.058027	0.001491	< 0.0001
2005	55 - 80	-0.25436	0.09395	0.00678	-0.343554	0.003149	< 0.0001

Table 4.5 shows the output of zero inflated negative binomial and negative binomial regression models regression models used for the study data. The output consists of coefficient estimate for each cohort group along with the standard error and the p-value.

Considering the 1990 cohort year group, the estimated coefficient for the zero inflated negative binomial regression models is 0.08540 and that of the negative binomial is 0.148591. The predictor variable for the negative binomial with offsets is statistically significant (p-value = 2×10^{-16}). This means that the occurrence of death is a function of age. However, age is not significant in the zero inflated negative binomial regression models.

The 1995 cohort group has coefficient estimate to be 0.06213, standard error to be 0.03258 and p-value of 0.0565 for the zero inflated binomial regression model. The significance of age in the model is on the border line. Age in the negative binomial regression model is statistically significant with coefficient estimate of 0.097710 along with standard error of0.001618 and pvalue 2×10^{-16}.

Also the cohort group of 2000 modeled by zero inflated negative binomial regression model has 0.0825 as coefficient estimate, standard error 0.0908 and pvalue of 0.09395 for the zero inflated negative binomial regression model which is not significant. The negative binomial regression model used to model the age for this cohort was statistically significant with p-value 2×10^{-16}.

The zero inflated negative binomial regression model used to model 2005 cohort group has a negative estimate (-0.25436) and that of negative binomial regression model also has negative estimate (-0.343554).both the zero inflated negative binomial and the negative binomial regression model used to model the age for the 2005 cohort group are statistically significant with p-value 0.00678 and p-value of 2×10^{-16} respectively. This indicated that occurrences of death is a function of age in the two models

Observations from the four cohort groups reveal that age all the cohort groups modeled by negative regression model are statistically significant but with the zero inflated negative binomial regression model, the 2005 cohort group which is statistically significant, the 1995 cohort group though significant is on the border line and the 1990/2000 are all not statistically significant.

The direction of coefficient estimate for each of the four cohort group for both the zero inflated negative binomial regression model and negative regression model are positive for cohort 1990, 1995 and 2000. However, the 2005 cohort shows a negative coefficient estimate. The effect of age as a significant predictor for the

2005 cohort group means for any one unit increase in age, there is a decrease in expected log count of the number of death.

Goodness-of-fit test

The female pension data which has age as the predictor variable and the number of occurrence of death as the response variable was put into four different cohort groups. Each cohort group was modeled by the ZINB and NB regression model. Both models have only one parameter being the age. The Aikaike Information Criteria (AIC) statistic was used to measure the goodness-of-fit of each of the cohort group model by ZINB and NB regression model. A good cohort model has the least value of the AIC criterion that is determined within each fitted model.

Table 4.6: cohort group selection criteria for ZINB and NB regression model

Cohort Group	**AIC (ZINB)**	**AIC (NB)**
1990	171.12	2856.80
1995	135.70	34640.00
2000	151.28	76218.00
2005	101.30	40714.00

The model selection to determine the goodness-of-fit is basically based on theoretical procedure of some criteria information. To judge the quality of a model we use the AIC criteria to select the best fit model. Table 4.6 shows AIC of the fitted ZINB and NB regression model for each of the cohort group.

Observing the ZINB regression model fitted for the various cohort group, it was found that the selection criteria for the 2005 cohort group model was low. That is the model with the smallest AIC of 101.30. The occurrence of death is linearly dependent on the age of the female pensioner. The lower the age there is a possibility of fewer or no death and higher age will result several count of death as the life expectancy of female is higher. Moreover, the ZINB regression model best fit 2005 cohort group data because $\alpha = 1.36$ which indicate over dispersion with excess zeros and is an assumption for ZINB.

The NB regression model fitted for the various cohort group shows that the selection criteria for the 1990 cohort model has the smallest value AIC of 2856.8. Amongst the four statistically significant cohort groups, it is the 1990 cohort group model that best fit the female pension data. The response variable (number of occurrence of death) is linearly dependent on the predictor variable (age).The 1990 cohort has $\alpha = 0.42$ which shows over dispersion but no excess zeros that is why the NB regression model best fit the 1990 cohort model.

With the fitted ZINB regression model for the female data, the 2005 cohort group is highly statistically significant but the 1995 cohort model though significant is on the border line.

CHAPTER 5

SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATION

5.0: Introduction

This chapter outlines the summary of the findings of the study, draw conclusions and make recommendations based on the findings.

5.1: Summary of the study

The study considered Ghanaian female pensioners who retired at age 5580 years within the calendar year from 1990-2005 five years interval period. The number of occurrence of death observed within the period was model using ZINB and NB regression model. Goodness-of-fit test for each cohort group of the various models was done using AIC.

From the analysis, it was observed that the number of females who retired over the five year intervals increased steadily. That is for the 1990 cohort group only 61 retired and by the end of the five years 1995 cohort group the number increased to 358, 2000 cohort group 670 and 2005 cohort group 1080 giving a total of 2,178 female pensioners. Out of the total female pensioners under the study, a total of 424 deaths occurred within the period. Maximum number of death occurred at age 72-73 for the period 1990, 1995 and 2000 whereas the 2005 cohort group recorded the maximum death at age 63 and 65.Female pensioners‟ exposure increased rapidly over the five year interval period.

Also the descriptive statistics revealed that each of the cohort group has a higher number of zero count and minimal count for higher deaths (non-zeros) of the study. Especially the 2005 cohort year group has excess zeros with highly sparse deaths.

Furthermore, the two models ZINB and NB regression model fitted to the study data revealed from 1990-2000 five years interval cohort group have positive estimates. It is only the 2005 cohort year group that has negative estimates. The fitted NB regression models for the entire four cohort group model are statistically significant with p-value (2×10^{-16}). However, the fitted ZINB regression model used to model the female pension data shows that only 2005 cohort year group was statistically significant and the 1995 cohort group significance is on the border line.

Finally, AIC used to test the goodness-of-fit for the various cohort group modeled shows that the ZINB regression model best fit the 2005 cohort group.
That of NB regression model best fit the 1990 cohort year.

5.2: Conclusions of the study

The ageing of Ghana‟s population has been rapid over the past two decades and will continue into the future with increasing number of Ghanaians surviving to 60 years and beyond.

This study used ZINB and NB regression model to model the Occurrence of death in Ghanaian Female Pensioners data from 1990 to 2005 with five years interval. The data which were obtained from SSNIT was put into four cohort groups of 1990, 1995, 2000 and 2005 then modeled using the two models. The main parameter used in this model is the age which is to be a function of mortality.

The NB regression model was significant to all the four cohort groups while the ZINB was significant to only the 2005 cohort group. The coefficient estimates for

1990, 1995 and 2000 cohort groups obtained from the two models are positive signifying an increase in the expected count of death with a unit increase in age.

Conversely, the coefficient estimate for the 2005 cohort group obtained from the two models was negative. This signified that for a unit increase in age there is a decrease in the expected count of death.

Finally, the application of the ZINB and NB regression models to the Ghanaian female pensioners‟ data shows the usefulness of the model. The ZINB regression model best fitted the 2005 cohort group while the NB regression model best fitted the 1990 cohort group.

5.3: Recommendations of the study

Because SSNIT base their calculations on age and pay the pension benefit until death, I recommend that that modeling the occurrence of death with age be extended and used to forecast future occurrence to enable SSNIT manage their reserves.

Due to time constraint only one parameter was considered; is recommend that more parameters (i.e. causes of death and standard of living) be considered to model the study data since a lot of factors affect mortality.

Recommend that the 2005 cohort group should be investigated to know why the expected count of death of female pensioners tend to decrease per unit increases age.

REFERENCE

Berin, B., G. Stolnitz, And A. Tenenbein. 1989. "Mortality Trends of Males and Females over the Ages." Transactions of the Society of Actuaries 41: 927.

Benefit; Old age or Retirement Pension, (© 2015 SSNIT), Retrieved from www.**ssnit**.org.gh/ataglance.php

Continuous Mortality Investigation CMI. Projecting future mortality: a discussion paper. Working Paper No. 3, 2004.

Charlton, J. 1997. Trends in All-Cause Mortality 1841-1994. The Health of Adult Britain, 1841-1994. Vol. 1, Decennial Supplement No. 12, pp. 17-29. London: Office for National Statistics.

Famoye, F., Wulu, J.T., Singh, K.P. 2004. On the generalized Poisson regression model with an application to accident data. *Journal of Data Science*. 2: 287-295.

Famoye, F., Singh, K.P. 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*. 4: 117130.

Greene WH (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. working paper, Department of Economics, Stern School of business, New York University, New York.

Hall DB (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics 56: 1030–1039.

Goss, S. A. Wade, And F. Bell. 1998. "Historical and Projected Mortality for Mexico, Canada, and the United States." North American Actuarial JownaZ 2(4): 108-26.

Hall. D and Berenhaut. K. S, *Score test for heterogeneity and over dispersion in Zero-inflated Poisson and Binomial Regression Models,* The Canadian Journal of Statistics,1-15,30,(3),(2002)

James C. Hickman, (2007). “An Actuary Who Made a Difference”. North American Actuarial Journal Volume 11, Issue 1,

Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*1992;**34**: 1–14.

APPENDIX A

The 2 x 26 Matrix for the Analysis

Age	**1990**		**1995**		**2000**		**2005**	
	dx	*dx*	*lx*	*lx*	*dx*	*Lx*	*dx*	*lx*
55	0	11	0	41	0	75	0	45
56	0	18	0	68	1	130	0	104
57	0	25	0	107	0	175	0	157
58	0	33	1	136	1	227	0	253
59	0	38	0	165	1	269	2	399
60	0	52	0	237	0	446	3	1920
61	0	57	1	292	6	552	10	2236

62	0	59	4	313	3	578	8	2384
63	0	59	2	322	3	600	24	2430
64	1	58	7	324	11	606	12	2457
65	1	58	7	322	7	611	1	2471
66	1	57	9	317	9	612	2	2477
67	0	57	4	316	6	610	2	2485
68	0	57	9	310	2	612	1	2486
69	1	57	9	303	5	608	0	2487
70	1	56	5	299	4	605	0	2492
71	0	56	4	296	9	598	0	2493
72	5	51	18	278	25	574	0	2495
73	6	45	33	245	47	527	1	2494
74	1	44	14	231	13	516	0	2494
75	0	44	6	225	4	512	0	2495
76	2	42	6	219	7	506	0	2495
77	1	41	8	211	0	506	0	2495
78	2	39	8	203	2	504	0	2495
79	3	36	2	201	1	503	0	2495
80	3	33	5	196	1	502	0	2495

APPENDIX B

The output of the analysis

1990 Cohort Group

```
zeroinfl(formula = morts$dx ~ morts$age, data = morts, dist =
"negbin", EM = FALSE)

Pearson residuals:
      Min        1Q    Median        3Q       Max
-1.138778 -0.341494 -0.017415  0.004738  2.850545

Count model coefficients (negbin with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.73789    3.67116  -1.563   0.1181   morts$age
```

```
0.08540    0.04993   1.710   0.0872 . Log(theta)   1.35572
1.16813   1.161   0.2458
```

```
Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  169.384    198.641   0.853    0.394 morts$age
-2.677      3.123  -0.857    0.391
---
glm.nb(formula = morts$dx ~ morts$age, data = morts, weights =
offset(morts$lx),
init.theta = 2.479668836, link = log)
```

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-11.4436   -5.6189   -2.8414   -0.1433   12.9920
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.431766   0.452419  -23.06   <2e-16 *** morts$age
0.148591   0.006257   23.75   <2e-16 ***
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(2.4797) family taken
to be 1)
```

```
    Null deviance: 1624.5  on 25  degrees of freedom
Residual deviance:  992.2  on 24  degrees of freedom
AIC: 2856.8
```

```
Number of Fisher Scoring iterations: 1
```

```
              Theta:  2.480
          Std. Err.:  0.327
2 x log-likelihood:  -2850.806
```

1995 Cohort Group

```
zeroinfl(formula = morts$dx.1 ~ morts$age, data = morts, dist =
"negbin", EM = FALSE)
```

```
Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.2122 -0.6142 -0.3453  0.4086  3.7918
```

```
Count model coefficients (negbin with log link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.35656    2.30004  -1.025   0.3056   morts$age
0.06213    0.03258   1.907   0.0565 . Log(theta)   0.89914
0.40191   2.237   0.0253 *
```

```
Zero-inflation model coefficients (binomial with logit link):
```

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  53.5056    31.1604   1.717   0.0860 . morts$age
-0.9035     0.5252  -1.721   0.0853 .
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 2.4575
Number of iterations in BFGS optimization: 12
Log-likelihood: -64.85 on 5 Df


glm.nb(formula = morts$dx.1 ~ morts$age, data = morts, weights = offset(morts$
init.theta = 2.17551166, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-29.294  -13.599  -10.311    6.989   32.771

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.856791   0.112497  -43.17   <2e-16 *** morts$age
0.097710   0.001618   60.40   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(2.1755) family taken to be 1)
    Null deviance: 9277.6  on 25  degrees of freedom
Residual deviance: 6724.4  on 24  degrees of freedom
AIC: 34640

Number of Fisher Scoring iterations: 1


              Theta:  2.1755
          Std. Err.:  0.0524

 2 x log-likelihood:  -34634.4470
```

2000 Cohort Group

```
zeroinfl(formula = morts$dx.2 ~ morts$age, data = morts, dist =
"negbin", EM = FALSE)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-0.9175 -0.5917 -0.2982  0.2343  3.7001

Count model coefficients (negbin with log link):
```

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.80281    3.39849  -1.119   0.2632   morts$age
0.08256    0.04908   1.682   0.0926 . Log(theta)  -0.10445
0.33577  -0.311   0.7557

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   71.852    231.564    0.31    0.756 morts$age
-1.295      4.178   -0.31    0.757
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 0.9008
Number of iterations in BFGS optimization: 144
Log-likelihood: -72.67 on 5 Df


glm.nb(formula = morts$dx.2 ~ morts$age, data = morts, weights =
offset(morts$lx.2),
init.theta = 0.9864830251, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-50.258  -21.068   -9.998    4.348   49.748

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.034556   0.103601  -19.64   <2e-16 *** morts$age
0.058027   0.001491   38.93   <2e-16 ***
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(0.9865) family taken
to be 1)

    Null deviance: 14807  on 25  degrees of freedom
Residual deviance: 13956  on 24  degrees of freedom
AIC: 76218

Number of Fisher Scoring iterations: 1


              Theta:  0.9865
          Std. Err.:  0.0139

2 x log-likelihood:  -76212.3460
```

2005 Cohort Group

```
zeroinfl(formula = morts$dx.3 ~ morts$age, data = morts, dist =
"negbin", EM = FALSE)

Pearson residuals:
     Min       1Q   Median       3Q      Max
-0.56390 -0.45063 -0.30682  0.03751  2.86223
```

```
Count model coefficients (negbin with log link):
             Estimate  Std. Error  z value  Pr(>|z|)
(Intercept) 17.36905    6.14417   2.827  0.00470 **
morts$age   -0.25436    0.09395  -2.707  0.00678 **
Log(theta)  -1.13479    0.40614  -2.794  0.00520 **

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|) (Intercept)
-11.60828          NA       NANA
morts$age     0.03572    1.55485   0.023    0.982


glm.nb(formula = morts$dx.3 ~ morts$age, data = morts, weights =
offset(morts$lx.2),
init.theta = 0.6516333755, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-34.354  -22.121   -9.026   -2.231   33.248

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 23.468479   0.204390   114.8   <2e-16 *** morts$age
-0.343554   0.003149  -109.1   <2e-16 ***
---
Signif.codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for Negative Binomial(0.6516) family taken
to be 1)

    Null deviance: 21196  on 25  degrees of freedom
Residual deviance:  9124  on 24  degrees of freedom
AIC: 40714

Number of Fisher Scoring iterations: 1


              Theta:  0.6516
          Std. Err.:  0.0121

2 x log-likelihood:  -40708.1480
```