Namibian Journal for Research, Science and Technology

Volume 4. Issue 1. May 2022

Special Issue

Guest Editor: Opeoluwa Oyedele, PhD

NCRST
National Commission on Research Science & Technology

a Journal of the National Commission of Research Science and Technology

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Editorial Commentary

Science for All: Boosting knowledge through communication

ARTICLE INFO

Received: March 2022
Accepted: April 2022

"Excellence is an art won by training and habituation"
— Aristotle

The need for science, technology, engineering and mathematics (STEM) to spearhead development has increased steadily, while communication and dissemination fall behind. In a world continuously influenced by scientific development, STEM communication grows ever more important to enable active engagement and participation of citizens in society. The intention is to engage society to increasingly recognize and illustrate the important of STEM for daily life, while boosting science motivation among students to learn STEM subjects, thus creating a critical mass of scientists, engineers and technologists for the world's tomorrow. This need is even more of a challenge in developing countries, like Namibia, where there is limited funding towards STEM.

Since 2015, the National Commission for Research, Science and Technology, in Namibia, has engaged in science fairs, public talks, or participatory projects, including publications as channels on reaching and engaging underserved audiences of STEM. One such publication is this journal. Now in its fourth volume, a special collection of articles on applications of statistics is showcased in this journal issue. The articles, edited by Dr Opeoluwa Oyedele, a senior lecturer in the Department of Computing, Mathematical and Statistical Sciences – formerly Department of Statistics and Population Studies, in the School of Science, are drawn from a selection of research done by postgraduate students and staff in that department. The collection tries to showcase the breath to which statistical techniques and methods are applied in a real world. All case studies are drawn from the Namibian context. The list of the articles is:

- Statistics: an intrinsic part of everyday life - commentary
- Modelling state preferences among airline travellers in Namibia: a case study at Eros airport and Hosea Kutako international airport
- Zero-augmented models for exploring the factors affecting the pass rate of 2016 grade 10 learners in Khomas region, Namibia
- Assessing the impact of proximate and non-proximate determinants of fertility in Namibia: a structural equation modelling approach
- Household poverty levels in Namibia and their associated sociodemographic factors: An empirical investigation of the 2015/16 Namibia household income and expenditure survey
- An application of survival analysis on the determinants of employment longevity in Namibia: evidence from 2018 Labour Force Survey
- Customer relationship management (CRM) and passenger loyalty in delivering high quality service at Air Namibia: A structural equations approach
- Application of longitudinal analysis to crime Data: Windhoek case study (2011-2016)
- Socio-demographic variations on age-sex mortality in Namibia: an analysis of the 2016 civil registration and vital statistics data
- A logistic regression model to assess factors influencing schizophrenia symptoms in Namibia
- An ecological adjusted random effect model for property crime in Windhoek, Namibia (2011-2016)

This collection is the first of its kind, and we look forward to many more, in line with the four main goals of NJRST, which is to inform, innovate, educate and debate, hence boosting scientific communication. We invite our readers to appreciate the beauty of statistics as presented here.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Guest Commentary

Statistics: An Intrinsic Part of Everyday Life

Opeoluwa Oyedele[1*]

[1]Department of Computing, Mathematical and Statistical Sciences, University of Namibia, Windhoek, Namibia

ARTICLE INFO

Received: March 2022
Accepted: April 2022

***Guest Editor: Special issue**

--
opeoluwaoyedele@gmail.com

"Learning is not child's play; we cannot learn without pain."
— Aristotle

Fast becoming an ever-present reality in our day-to-day lives, statistics play an intrinsic and contemporary role in our everyday activities, especially, in today's data-driven world. Owning to its definition as the science of collecting, summarizing, presenting and interpreting information, statistics helps us understand the world a little bit better through numbers as well as other quantitative and qualitative source of information.

Although statistics originated many centuries ago, its impacts and applications have evolved in recent years as modern statisticians have advanced applications of statistics through innovative, problem-solving approaches, far beyond its historic use by governmental offices to capture censuses in order to track population sizes and growths (Michigan Technological University, 2021). One of these modern statisticians was Sir Ronald Aylmer Fisher (1890 - 1962) who was active as a mathematician, statistician and geneticist. Although trained as an (evolutionary) biologist, Fisher was referred to as the "father of the modern science of statistics" as he single-handedly created the foundations for modern statistical science (Anders, 1998). He further pioneered the design of experiments principles, small samples statistics and the analysis of real data. He went on to published a book titled "*Statistical Methods for Research Workers*" in 1925, which later became one of the 20th century's most influential books on statistical methods and even to date. His notable works includes the popularly used F-test, F-distribution, Fisher's exact test, maximum likelihood estimation, random effects models and analysis of variance, to mention a few.

The importance of statistics being relevant to nearly every area of our lives cannot be overemphasized enough, to the extent that all countries over the world have at least one national statistical agency operating within their respective countries that manages critical information related to labour trends, health, education, political campaigns and many more. Statistics has influenced and is still influencing the operations of industries such as the sales and financial markets, profit and non-profit organizations, meteorological services, medical care services, manufacturing, urban planning, education, law, and even social media analytics (to mention a few). In this present world, we as individuals and organizations regularly use statistics to make daily financial and non-financial planning and budgeting decisions that affects our lives. For example, consider the daily forecasted weathers, lending risks at banks, impacts of economy crises, healthcare financial policies, traffic flow operations, investment payoffs, urban planning with respect to population growths and declines, predicting diseases, stock markets, human psychology behaviours, insurance pay-outs, and political election results. All these, and more, are statistics and/or derived from statistics. Another simple life application of statistics is the global daily recorded number of COVID-19 cases, deaths, recoveries and vaccination numbers. As of 23 August 2021, a total of 212,679,403 COVID-19 cases were recorded, with 4,446,610 deaths and 190,301,359 recoveries, with majority of these cases reported in the United States of America, India, Brazil, United Kingdom, France, Russia, Turkey and Italy (Worldometer, 2021). Again, all these basic figures were all compiled through the use of statistics in various regions, states, countries, provinces and continents all over the world. Thus, statistics is heavily used in many different fields for a variety of applications as showcased in this journal issue.

In this journal issue, a total of ten original research articles have been assembled, with different fields of statistics applications and techniques. These statistical applications were in aviation, education studies, health sciences, economics and management, population studies, customer relation management, crime analysis and curriculum studies, using several statistical techniques such as generalized linear models, zero-inflated generalized linear models, structural equation modelling, Kaplan-Meier and Cox proportional hazard, and longitudinal data analysis with generalized estimating equations. To be precise, Amwaama, Oyedele and Kazembe used a binary logit, probit and latent class modelling technique on information collected from departing air passengers at Eros and Hosea Kutako International airports in Windhoek, Namibia, to model passengers' stated preferences in choosing between Low-Cost Carriers (LCC) and Full-Service Carriers (FSC), in addition to examining the determinants of carrier choice between LCC and FSC in Namibia. They found that factors such as passenger's airfare, age, income and purpose of travel were significantly important with respect to the passenger choice and that passengers had different preferences for different destination be it domestic, regional (short haul) and international (long haul). For domestic and regional flights, the passengers preferred LCC, while the FSC was preferred for international flights. Their study further recommended that the best, viable and appropriate carrier in and within Namibia should be an LCC, which can ensure sustainability. Similarly, Nakunipa, Pazvakawambwa and Iiyambo study used a binary logistic regression technique to establish the prevalence and factors influencing schizophrenia symptoms using information obtained from the Namibia demographic and health survey. It was revealed in their study that the prevalence of schizophrenia symptoms in Namibia was 12.4%. Also, females were more likely to have schizophrenia symptoms, while people who resided in urban areas and did not consume alcoholic drinks were less likely to have schizophrenia symptoms Their study further recommended that there is a need to step up gender-specific mental health programs especially in rural areas.

Mumbuu, Pazvakawambwa and Oyedele used the zero-inflated generalized linear modelling technique that caters for excess zeros within datasets to explore the influencing factors that affects grade 10 learners' pass rate in the Khomas region of Namibia, based on information extracted from the directorate of national examination and assessment. In their study, it was revealed that the age of the learners, school location and the type of school had significantly influenced the pass rate of grade 10 learners. Their study further revealed that the zero-inflated negative binomial technique was the best-fit model to use in the modelling of the factors which influenced Khomas region grade 10 learners' pass rate. For these reasons, it was recommended that more schools be built in densely populated areas so that classrooms are not overcrowded per subject, in addition to overaged learners being given extra teaching assistance and attention.

Likewise, Shipanga, Oyedele and Matengu used a negative binomial regression modelling technique to perform an inference mortality analysis across all ages and both sexes in Namibia as well as across regions and marital status using information collected from the civil registration vital systems. Their study revealed that there was a significant relationship between mortality and the individuals' age, sex, marital status and region. In addition, Oshana, Kavango East, Khomas, Hardap and Omaheke regions had high mortality rates, while infants and elderly individuals had a high probability of dying. For these reasons, their study recommended that interventions such as affordable and proper health care and well-being services targeted at the (most) vulnerable age groups, marital group and regions be immediately made available, in order to meet goal 3 of the health-related sustainable development goals of the United Nations.

Moreover, Shinyemba and co-authors used the structural equation modelling technique to model the direct and indirect effects of socio-economic, socio-demographic and health attributes on fertility, as well as the proximate and non-proximate determinants of fertility in Namibia. Their study showed that the proximate determinants had a direct negative impact on the number of children ever born, while a positive effect existed between the non-proximate determinants and the number of children ever born. In addition, women who had their first birth at the beginning of their reproductive period as well as those who had their first marriage at younger ages were more likely to have more children. For these reasons, their study recommended that there is a need to promote contraceptive use among Namibian women to further reduce fertility rates among women from poor households as the cost of raising children has become high as the year progresses. Similarly, the study by Isaacs, Lwendo and Kazembe used the structural equation modelling technique to establish the impact of passenger loyalty on Customer Relationship Management (CRM) in delivering high quality service to passengers and value creation using survey data collected from international, regional and domestic passengers using Air Namibia for passengers travelling through the Hosea Kutako International and Eros Airports. Their study revealed that factors such as customer orientation, operational specialties, domain expertise, and service recovery and information technology contributed to passenger's satisfaction with Air Namibia's value chain activities, while factors that contributed to passenger's retention and loyalty towards Air Namibia's products and services included marketing and promotional activities, loyalty aspects, value for money and comfort issues. Their study also found that interpersonal relationships between staff and the customers were crucial to CRM initiatives as they can result in a better understanding of customer needs, which in turn leads to passenger loyalty.

Oyedele, Angula and Mutorwa used the Kaplan-Meier and Cox proportional hazard techniques to estimate the survival of employment longevity for employed adults in Namibia using information extracted from the Namibia labour force survey. Their study revealed that employed adults' characteristics such as age group, type of employer, highest education attained, marital status, region, current schooling status and sex had a significant association with their survival of employment longevity. Additionally, employees aged 30-39 and 40-49 years, employed in non-profit institutions, parastatals and government institutions, and from the Oshikoto, Omaheke, Oshana, Khomas, Erongo and Otjozondjupa regions had a high survival of employment longevity, while employees employed in privately-owned informal enterprises and had already attained a technical or vocational certificates/diplomas, junior and senior secondary education had a low survival. Their study further recommended that all relevant organizations and governmental ministries that deals with employment and labour matters should frequently engage all employers through their respective human resources departments, to further assist in the creation and implementation of favourable employment contracts that best suits their respective employees. Unandapo and Ntjamba analyzed reported crime data in the Windhoek municipal area using longitudinal data analysis with generalised estimating equations to compare snapshots of crime over time while considering the correlations within the data. Their study revealed that the best correlation structure was the exchangeable correlation structure, which assume constant correlation over time.

Conversely on the same Windhoek crime data, Amunyela, Neema and Pazvakawambwa used a zero-inflated negative binomial – generalized linear mixed model to model factors related to property crime (theft and burglary) - where they assumed the data is aggregated at zone level. Random effects were then assigned to each of the 59 policing zones in City, with the goal of highlighting areas likely to experience property crime. Crime was high during spring and winter time during the study period. The study further discovered that areas with high population densities have a high crime intensity.

In a nutshell, supporting the words of the famous English writer Herbert George "H. G." Wells (1866 - 1946) who once said:

> *"Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write",*

the applications of statistics and its methods are fast becoming an unavoidable intrinsic need in every area of our day-to-day lives. It might be useful to mention at this junction that the Department of Computing, Mathematical and Statistical Sciences at the University of Namibia, Windhoek, currently offers a range of statistics-affiliated qualifications:

- *BSc in Statistics (NQF level 8)*
- *BSc in Population Studies (NQF level 8)*
- *MSc in Applied Statistics & Demography (NQF level 9)*
- *MSc in Biostatistics (NQF level 9)*
- *PhD in Applied Statistics (NQF level 10)*
- *PhD in Population Studies (NQF level 10)*

as well as new data science affiliated programs as from 2023 onwards:

- *Diploma in Applied Statistics (NQF level 6)*

- *Diploma in Computing (NQF level 6)*
- *BSc on Computing (NQF level 7)*
- *BSc in Data Science (NQF level 7)*
- *Postgraduate Diploma in Applied Research Methods (NQF level 8)*

For more information about these qualifications and programs, interested persons are encouraged to contact Dr. Opeoluwa Oyedele at ooyedele@unam.na or Dr. Samuel Nuugulu at snuugulu@unam.na.

References

Anders, H. (1998). A History of Mathematical Statistics. Wiley, New York.

Fisher, R.A. (1925). Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Michigan Technological University. (2021). Every Number Counts: The Importance of Applied Statistics in Our Daily Lives. Available at https://onlinedegrees.mtu.edu/news/every-number-counts-importance-applied-statistics-our-daily-lives-infographic. Accessed date: 22 July 2021.

Worldometer. (2021). COVID-19 Coronavirus Pandemic. Available at https://www.worldometers.info/coronavirus/. Accessed date: 23 August 2021.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Modelling state preferences among airline travellers in Namibia: a case study at Eros airport and Hosea Kutako international airport

Alisa M. Amwaama[1,2], Opeoluwa Oyedele[1], Lawrence N. Kazembe[1]*

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
[2]Air Namibia, Windhoek, Namibia
*Corresponding author: lkazembe@yahoo.com

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
state preferences
airline travellers
probit
logit
Namibia

ABSTRACT

Stated preference experiments are become an increasingly popular survey methodology for investigating air travelers' choices. Analysis of this behaviour, which is an element of the demand prediction, helps for a better future planning and development of competing airlines. In this paper, emphasis is stressed on the stated preferences of passengers in choosing between low cost carriers (LCC) and full service carriers (FSC). A binary logit, probit and latent class models were employed on the primary data collected from departing air passengers at Eros and Hosea Kotako International airport in Windhoek, Namibia, to model passengers' stated preferences and examine the determinants of carrier choice between LCC and FSC in Namibia. Major findings show that airfare, age, income, and purpose of travel are significantly important with respect to passenger choice. Furthermore, it was observed that passengers have different preferences for different destination region be it domestic, regional and international. For domestic and regional flights (short haul) they prefer LCC, while for international flights (long haul) they opted for FSC. In addition, majority of the passengers were travelling for business purposes, hence their tickets were bought by their respective employers. Most passengers indicated that they were willing to fly LCC if it was available in Namibia because of it's low fares. There was an indication that air tickets were not affordable and these are a big concern to passengers. Presumably, if ticket prices can come down or introduce a LCC in Namibia then many will consider flying. This study concluded that, based on the interviewed passengers' profiles, the best and appropriate carrier in Namibia is a low-cost carrier. Introducing a LCC in Namibia might be a viable alternative which may ensure sustainability.

1. Introduction

Discrete choice models, such as the binary logit and multinomial logit, are used to predict the probability a decision-maker (often an individual, group of individuals or corporates) will choose one alternative among a finite set of mutually exclusive and collectively exhaustive alternatives. Currently, there is a growing interest in applying discrete choice models in the airline industry. This interest is driven by the desire to more accurately represent why an individual makes a particular choice and how the individual makes trade-offs among the characteristics of the alternatives.

Integrating discrete choice and other models grounded in behavioural theories with traditional revenue management, scheduling, and other applications is also being driven by several factors, including the increased market penetration of low-cost carriers, wide-spread use of the internet, elimination and/or substantial reduction in travel agency commissions, and introduction of simplified fare structures by network carriers (Garrow, 2010). The presence of low-cost carriers has reduced average market fares and increased the availability of low fares. Moreover, Garrow (2010) indicated that the internet has reduced individuals searching costs and made it easier for individuals to both find these fares and compare fares across multiple carriers without the assistance of a travel agent. In addition, the elimination of commissions has removed the incentive of travel agencies to concentrate sales only on those carriers offering the highest commissions.

According to Garrow (2010), the introduction of simplified fare structures by network carriers was motivated by the need to offer products competitive with those sold by low-cost carriers. Often, low-cost carrier products do not require Saturday night stays and have few fare-based restrictions. However, these simplified fares have been less effective in segmenting price-sensitive leisure passengers willing to purchase weeks in advance of flight departure from time-sensitive business passengers willing to pay higher prices and needing to make changes to tickets close to flight departure. All of these factors have resulted in the need to better model how passengers make purchasing decisions, and to determine their willingness to pay for different service attributes. Moreover, Garrow (2010) further detailed that unlike traditional models based solely on an airlines internal data, there is now a perceived need to incorporate existing and/or future market conditions of competitors when making pricing, revenue management, and other business decisions.

A wide range of studies have investigated air travel choice behaviour. Mamdoohi et al. (2013) used binary logit to model the origin airport choice of resident and non-resident travellers' from the city of Tehran. Results show that the difference in the two groups is affected by "age", "income", "travel Destination", "trip Purpose" and "marital Status". Further model results show that variables "public access", "flight frequency" and "airport tax" are more important for non-resident air travellers' in choosing their origin airport. Ashford and Bencheman (1987) developed a multinomial logit model to analyze air passengers' choice in central London. This study showed that for business and inclusive tour travel, the most important variables of choice were access time to the airport and frequency to the chosen destination. For domestic and leisure trips, there were three factors: airfare, access time, and frequency of available flights, in that order of importance. Davidson and Ryley (2010) performed a binary logit modelling in airport choice in which the air fare was the most meaningful variable whereas the travel time was the second one. Hess and Polak (2005) extended a mixed multinomial logit model to analysis of the choice of airport, airline and access-mode for travellers' living in the San Francisco Bay area. Results indicated that the most important variables affecting travellers' choices were in-vehicle access time, access-cost and flight frequency.

In a related study, a binary logit was used for airport selection in which the most meaningful variables were airfare, access time and frequent flyer benefits (Hess et al. 2007). Another study by Pels et al. (2001) developed a nested logit model to investigate low-cost airline and airport competition in greater London. They analyzed most important factors affecting air travellers' choices such as airfare, surface-access costs and frequency. In a related study, Stefano (2012) used discrete choice random utility models (multinomial logit, mixed multinomial logit and cross-nested logit models) to investigate and model airport choice behaviour in a multi-airport region in Campania, southern Italy. He found that access time, airfare, age, experience and income were the most significant variables.

When passengers choose a carrier, they may base their decision on a combination of factors, including the airline's market presence, schedule convenience, low fares, on time performance, re- liability and the availability of frequent flyer programmes (Proussaloglou and Koppelman, 1999). Hess et al. (2007) studied the airport and airline choice behaviour with the use of stated preference survey data. This paper analyzed factors affecting passenger choice behaviour, including air fare, access time, flight time and airline and airport allegiance using multinomial logit model. Similarly, Pels et al. (2003) used nested logit model and found that passengers are sensitive to fare, frequency, airport access time and airport access cost. Later, same authors studied the competition between full service and low-cost airlines by analyzing the demand structure. They estimated not only the competition for passengers occurring between airports and airlines, but also the own-and cross-price elasticities based on a nested logit model (Pels et al. 2009). There are significant differences in choice behaviour between business travellers' and non-business travellers (Chang and Sun, 2012). Most business travellers have strict requirements regarding travel time and will seldom strive for lower prices because they are restricted by time inflexibility. On the contrary, leisure travellers' will choose the lower price among two acceptable flight choices (Xiao et al 2008). Overall, discrete choice models provide one framework for accomplishing these objectives. In this study we model stated preferences among airline travellers in Namibia.

2. Modelling Travellers' Choices and Preferences

A dichotomous-choice response question is examined, "Why does a traveller choose a particular airline (Low-Cost Carrier [LCC] = 1) over its alternative (Full Service Carrier [FSC] = 0) in his/her travel decision making?" A log-odds model is adopted and estimated using logit analysis of the form (Greene and Hensher, 2010):

$$log[P/(1-P)] = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon \qquad (1)$$

where P is the probability of the respondent to travel by a particular carrier (i.e. LCC); X_i is explanatory variable hypothesized to influence this probability; while β is coefficients for the explanatory variables; ε is stochastic disturbance term; and $P/(1-P)$ is the ratio of the probability that the respondent travels by LCC to the probability that he/she travels by FSC. It can also be considered as the odds of the respondent to travel by LCC over FSC.

The set of socio-demographic explanatory variables employed are: age groups, ethnic categories, gender, sector of employment, monthly income levels, and educational level. In addition, several behavioural variables are included: concerns for airfare, method of booking, purpose of travel, and destinations of travel Table 1. The predictor variables were identified in line with the objectives of this study. We seek answers to the objectives of the study. These variables will assist us identify what determinants inform the stated preferences based on passengers profiles and when this is assessed, the Namibian airline industry can be informed accordingly given the SP knowledge of their passengers.

In a binary response model, two approaches are available; logit and probit. A logit is obtained if cumulative logistic model is used, whereas a probit applies when ε is assumed to follow a cumulative standard normal distribution.

2.1. Estimation: Maximum Likelihood Estimation of Binary Response Models

Estimation and inference for probit and logit models for binary choices are usually based on maximum likelihood estimation. Because the dependent variable is discrete, the likelihood function cannot be defined as a joint density function as with a continuously-distributed dependent variables. Each observation is a draw from a Bernoulli distribution (binomial with one trial). The model with success probability $F\ (\gamma' x_i)$ and independent observations leads to the joint probability, or likelihood function (Greene and Hensher, 2010),

$$Prob(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n | x_1, x_2, \dots, x_n) = \prod_{y_i=0}[-F\ (\gamma' x_i)] \prod_{y_i=1} F\ (\gamma' x_i).$$

Let X denote the sample of n observations, where the ith row of X is the ith observation on x_i (transposed, since x_i is a column) and let y denote the column vector that is the n observations on y_i. Then, the likelihood function for the parameters may be written

$$L(\gamma|X, y) = \prod_{i=1}^{n}[1 - F\ (\gamma' x_i)]^{1-y_i}[F\ (\gamma' x_i)]^{y_i}$$

Taking logs, we obtain the log likelihood function,

$$lnL(\gamma|X, y) = \sum_{i=1}^{n}(1 - y_i)ln\,[1 - F\ (\gamma' x_i)] + y_i lnF\ (\gamma' x_i)$$

By limiting our attention to the normal and logistic, as symmetric distributions, this permits a useful simplification if we let $q_i = 2y_i - 1$. Thus, q_i equals −1 when y_i equals zero and +1 when y_i equals one. Because symmetric distributions have the property that $F(t) = 1 - F(-t)$, we can combine the preceding into:

$$lnL(\gamma|X, y) = \sum_{i=1}^{n} lnF[q_i\ (\gamma' x_i)]$$

T

he maximum likelihood estimator (MLE) of γ is the vector of values that maximizes this function. The MLE is the solution to the likelihood equations, The elasticities are simple to obtain from the estimated partial effects. However, since it is a ratio of percentage changes, the elasticity is not likely to be useful for dummy variables such as marital status, or for discrete variables such as age and education level. Like a partial effect, an elasticity for a dummy variable or an integer valued variable will not necessarily produce a reasonable result. The computation for a dummy variable or an integer variable would be a semi-elasticity, [%ΔProb]/Δx, where Δx would equal one. Whether a percentage change in an integer valued x would make sense would depend on the context.

3. Application: Analysis and Results

Primary data from 285 departing passengers at the two Windhoek airports were analyzed to model four binary logit models. The data set include aspects that affect choice of carrier; behavioural aspects and socio-demographic factors. The dependent variable is defined as passenger's stated preference for k flights = 1 if LCC otherwise 0 if FSC, where k is either domestic, regional, international or general flights. The following variable were used in the regression part of the model, x_i = *(constant, gender, income, education level, maritalstatus, age, nationality).*

The predictor variables were identified in line with the objectives of this study. We seek answers to the objectives of the study. These variables will assist us identify what determinants inform the stated preferences based on passengers' profiles and when this is assessed, the Namibian airline industry can be informed accordingly given the SP knowledge of their passengers.

In the original data set, income is divided into three parts as INCOM1, INCOM2, INCOM3 representing low, mid and high income respectively. Education level is measured by TERTIARY which is a binary variable, indicating whether or not the respondent has attended tertiary level. Descriptive statistics for the data used in the analysis are shown in Table 1 and 2. Estimates of the parameters of the logit models are shown in Tables 3 and 4.

The assumptions of binary response model are that the outcome must be discrete, otherwise explained as, the dependent variable should be dichotomous in nature (e.g., LCC vs. FSC); There should be no outliers in the data, which can be assessed by converting the continuous predictors to standardized, or z scores, and remove values below -3.29 or greater than 3.29; There should be no high intercorrelations (multicollinearity) among the predictors. This can be assessed by a correlation matrix among the predictors. Tabachnick and Fidell (2012) suggest that as long correlation coefficients among independent variables are less than 0.90 the assumption is met. Hence the binary logit assumptions are met and analysis proceeds.

The analysis are presented according to regions of destinations, which are domestic, regional and international plus the passengers general flying preference. All analysis were carried out in R (3.1.0) statistical package.

Table 1: Description and summary statistic of variables in the statistical model

VARIABLE	DESCRIPTION	MEAN	STD.DEV
WHYTHEAIRLINEGEN	1= THEY CHOOSE AN AIRLINE DEPENDING ON THE FARE, 0=OTHERWISE	1.51	1.143
FLYMOSTLCCFSC	1= LCC, 0=FSC	0.91	0.288
WHYAIRLINESPC	1= FARE, 0=OTHERWISE	2.57	1.484
TICKETPAYER	1= MYSELF, 0=OTHERWISE	0.41	0.584
RESERVATIONSPOINT	1= RESERVATION MADE ONLINE, 0=OTHERWISE	2.18	1.138
ONLINESVCS	1= ONLINE SERVICES IS CONVINIENT, 0 OTHERWISE	0.32	0.727
LONGHAULS	1= PREFER LCC ON LONGHAULS, 0= PREFER FSC ON LONGHAULS	0.31	0.463
DOMESTIC	1= PREFER LCC ON DOMESTIC, 0= PREFER FSC ON DOMESTIC	0.81	0.395
REGIONAL	1= PREFER LCC ON REGIONAL, 0= PREFER FSC ON REGIONAL	0.66	0.475
GENERALFLIGHT	1= PREFER LCC IN GENERAL, 0= PREFER FSC IN GENERAL	0.6	0.491
NATIONALITY	1= NAMIBIAN, 0= NON-NAMIBIAN	0.61	0.488
GOVERNMENT	1= GOVERNMENT EMPLOYEE, 0= OTHERWISE	0.13	0.333
TERTIARYEDU	1= TERTIARY EDUCATED, 0= OTHERWISE	0.6	0.491
INCOM1	1= 0 -9999 (LOW INCOME), 0= OTHERWISE	0.16	0.369
INCOM2	1= 10 000 -29 999 (MID INCOME), 0= OTHERWISE	0.51	0.501
INCOM3	1= 30 000 - 40 000+ (HIGH INCOME), 0= OTHERWISE	0.31	0.461
MARITAL STATUS	1= SINGLE, 0= EVERMARRIED	0.41	0.492
GENDER	1= MALE, 0= FEMALE	0.59	0.493
YOUTH	1= 15-34 YOUTH, 0= OTHERWISE	0.38	0.485
ADULT	1= 35-54 ADULT, 0= OTHERWISE	0.51	0.501
SENIORCITIZV	1= SENIOR CITIZEN, 0= OTHERWISE	0.1	0.303
FAREMATTERS	1= CHOICE BASED ON FARE, 0= OTHERWISE	0.36	0.48
ONLINECONVINIENT	1= CONVINIENT, 0= OTHERWISE	0.18	0.384
FLYREASON	1= BUSINESS, 0= OTHERWISE	0.89	0.316

3.1 Passenger Stated Preferences for Domestic flights

Table 2 shows the frequencies of passenger stated preferences for all destination regions. On average, about 81% of respondents stated that they prefer LCC on domestic routes. Table 3 shows the results of the logit analysis for Domestic preferences. Only one socio demographic variables (TERTIARYEDU) and two behavioral factors (FARE MATTERS AND FLY- REASON) were statistically significant in affecting the choice of carriers (Table 3). The odds of flying domestic with LCC is 0.391 times less for passengers with tertiary education as opposed to passengers with other level of education other than tertiary level. Further, tertiary educated respondents were less likely to fly LCC on domestic routes as compared to other passengers with non-tertiary educated respondent. This implies that higher educated individuals had a higher tendency of traveling by FSC on domestic routes. Being concerned over fares is also a statistically significant factor on the probability of carrier choice as those who value airfares have a 2.493 more in the log-odds of flying LCC, holding all other independent variables constant. This result is consistent with the findings of O'Connell and Williams (2005) and Ong and Tan (2010), whereby fare is the principle reason for carrier selection among low-cost airline passengers.

Table 2: Frequency table for passengers SP for all regions of destinations and in general.

		DOMESTIC		REGIONAL		INTERNATIONAL		GENERAL	
VARIABLE	CATEGORY	FSC % (n)	LCC % (n)	FSC % (n)	LCC % (n)	FSC % (n)	LCC % (n)	FSC % (n)	LCC % (n)
GENDER	FEMALE	15% (18)	85% *100)	26% (31)	74% (87)	61% (72)	39% (46)	35% (41)	65% (77)
	MALE	22% (37)	79% (130)	40% (66)	61% (101)	75% (125)	25% (42)	44% (73)	56% (94)
NATIONALITY	NAMIBIAN	20% (34)	80% (140)	29% (50)	31% (124)	66% (114)	34% (60)	35% (61)	65% (113)
	NON-NAMIBIAN	19% (21)	81% (90)	42% (47)	58% (64)	75% (83)	25% (28)	48% (53)	52% (58)
MARITAL STATUS	EVER MARRIED	18% (30)	82% (139)	34% (58)	66% (111)	68% (115)	32% (54)	36% (60)	64% (109)
	SINGLE	22% (25)	78% (91)	34% (39)	66% (77)	71% (82)	29% (29)	47% (54)	53% (62)
EDULEVEL	LOWER	17% (1)	83% (5)	17% (1)	83% (5)	50% (3)	50% (3)	17% (1)	83% (5)
	HIGHER	25% (27)	75% (82)	29% (32)	71% (77)	66% (72)	34% (37)	33% (36)	67% (73)
	TERTIARY	16% (27)	84% (143)	38% (64)	62% (106)	72% (122)	28% (48)	45% (77)	54% (93)
AGE	0-34 (YOUTHS)	16% (23)	84% (83)	25% (32)	75% (74)	86% (80)	14% (26)	39% (51)	61% (55)
	35-54 (ADULTS)	28% (27)	72% (123)	35% (52)	65% (98)	65% (97)	35% (53)	32% (49)	68% (101)
	55+	17% (5)	83% (24)	45% (13)	55% (16)	75% (52)	25% (17)	48% (14)	52% (15)

The reason why respondents fly is significantly related to carrier choice as those that fly for business are less likely to use LCC as opposed to respondents that fly for non-business purposes. The odds of business travelers to fly domestic with LCC is 0.233 times less than for respondents traveling for other reasons other than business.

Table 3: Domestic and regional flights results of logit analysis

	DOMESTIC			REGIONAL		
VARIABLE	COEF.	SIG	OR	COEF.	SIG	OR
CONSTANT	4.215	0.141	67.667	1.695	0.539	5.447
NATIONALITY	0.195	0.615	1.215	-0.212	0.499	0.809
GOVERNMENT	-0.117	0.841	0.890	-1.622	0.014 (*)	0.197
TERTIARYEDU	-0.939	0.009 (*)	0.391	0.023	0.939	1.023
INCOM1	1.331	0.188	3.785	0.701	0.446	2.016
INCOM2	-0.583	0.533	0.558	-0.064	0.939	0.938
INCOM3	0.674	0.475	1.962	0.985	0.240	2.678
MARITAL STATUS	0.145	0.714	1.156	-0.098	0.760	0.907
GENDER	0.179	0.635	1.196	0.173	0.574	1.189
YOUTH	-1.140	0.278	0.320	-0.413	0.698	0.662
ADULT	-1.428	0.158	0.240	-0.414	0.690	0.661
SENIORCITIZV	-1.768	0.119	0.171	-0.14	0.899	0.87
FARESMATTERS	0.914	0.01 (*)	2.493	0.345	0.229	1.412
ONLINECONVINIENT	0.499	0.402	1.648	1.034	0.043 (*)	2.812
FLYREASON	-1.459	0.01 (*)	0.233	-0.492	0.243	0.611

3.2. Passenger Stated Preferences for Regional flights

Passengers stated different preferences for different fleets. In this paragraph we are examining the stated preferences for passengers on regional fleet. The model explains 19.9% of the variability of the response data around its mean. Among the interviewed passengers, 66% indicated that their stated preference on regional fleet is LCC and only 34% stated to prefer FSC (see Table 2). Table 3 shows the results of the logit analysis for regional flights. Results in Table 3 indicate that government employee's respondents were less likely to fly LCC on Regional routes as compared to respondents from any other sector. Therefore the odds of flying Regional with LCC was 0.197 less times more for respondents who were worked for the Government opposed to respondents who are non-government employees. This is supported by the fact that the most tickets are company/government paid and companies/government usually just pay for full paid tickets for their employees' business trips. This is tied to business travelers being less likely to fly LCC on domestic routes.

The use of online services in making airline reservations is significantly related to carrier choice as those who fly LCC on Regional are more likely to use Information Communication Technology (ICT) booking channels. Moreover, for every one-unit increase in online services, we expect a 1.034 increase in the log-odds of flying LCC on Regional routes, holding all other independent variables constant. Research show that many that fly LCC use ICT booking channels (Hess et al. 2007).

3.3 Passenger Stated Preferences for International flights

On international fleet, which are usually long hauls, passengers stated preferences are quite different from those of domestic and regional fleet. Table 2 displays that about 69% of interviewed passenger stated that they will prefer FSC on International routes because they are quite comfortable than LCC and on a long haul one needs to travel in comfort.

Factors related to being a government employee and online services convenience were statistically significant in affecting the choice of airlines (Table 4). Specifically, government employees respondents were less likely to fly LCC on regional routes as compared to respondents from any other sector. Therefore, the odds of flying regional with LCC was 0.359 less times more for respondents who were worked for the government opposed to respondents who are non-government employees. This is supported by the fact that the most tickets are company/government paid and companies/government usually just pay for full paid tickets for their employees' business trips. Similar to Regional the use of online services in making airline reservations is significantly related to carrier choice as those who fly LCC on International are more likely to use information communication technology (ICT) booking channels. This implies that for International or long haul flights respondent prefer FSC over LCC due to the comfort found in FSC.

Table 4: International and general flights results of logit analysis

	INTERNATIONAL			GENERAL		
VARIABLE	COEF.	SIG	OR	COEF.	SIG	OR
CONSTANT	-1.739	0.506	0.176	3.753	0.179	42.648
NATIONALITY	-0.222	0.519	0.801	-0.800	0.014 (*)	0.449
GOVERNMENT	-1.024	0.02 (*)	0.359	-0.963	0.088	0.382
TERTIARYEDU	-0.061	0.845	0.941	0.076	0.801	1.078
INCOM1	0.584	0.501	1.792	1.027	0.278	2.792
INCOM2	0.803	0.301	2.232	-0.203	0.812	0.816
INCOM3	1.364	0.089	3.913	0.688	0.425	1.989
MARITAL STATUS	-0.127	0.709	0.881	0.201	0.531	1.222
GENDER	0.419	0.180	1.521	-0.316	0.311	0.729
YOUTH	0.555	0.593	1.742	-0.806	0.441	0.446
ADULT	-0.065	0.949	0.937	-1.702	0.097	0.182
SENIORCITIZV	-0.100	0.927	0.905	-1.305	0.234	0.271
FARESMATTERS	0.580	0.069	1.785	0.936	0.002 (*)	2.549
ONLINECONVINIENT	-2.099	0.006 (*)	0.123	0.587	0.209	1.798
FLYREASON	-0.653	0.109	0.520	-1.793	0.0 *)	0.166

3.4. Passenger Stated Preferences for General flights

Now turning to preferences flights in general, results show that more than half (60%) (see Table 2) of the interviewed group stated that they prefer LCC. Table 4 shows the results of the logit analysis for general lights. Both NATIONALITY and two behavioural factors (FAREMATTERS AND FLYREASON) were statistically significant in affecting the choice of airlines.

Namibians were less likely to prefer LCC more than non-Namibians. This is because the non-Namibians were possibly exposed to LCC in their respective countries, unlike in Namibia where there is not a single LCC. The odds of flying domestic with LCC is 0.449 times less for Namibian respondents as opposed to non-Namibian respondents. Further, even though over all majority preferred LCC, for a Namibia it was less likely compared to non-Namibians. Overall, both Namibians and Non- Namibians respondents had a higher tendency of traveling by LCC than FSC in general- regardless of the route. Being concerned over fares is also a statistically significant factor on the probability of carrier choice as those who value airfares have a 2.551 increase in the log-odds of flying LCC, holding all other independent variables constant. The reason why respondents fly is significantly related to carrier choice as those that fly for business are less likely to use LCC as opposed to respondents that fly for non-business purposes. The odds of business travellers to fly domestic with LCC is 0.166 times less than for respondents traveling for other reasons other than business.

4. Conclusion

This study aims to inspect the likelihood of passengers to choose between two air carriers with dissimilar operating structures: low cost and full-service carrier. The findings provide additional support to the concept that passengers' socio-demographics (occupation, education level) and behavioural choices (concerns about ticket prices, fares, online services, and purpose journey) are main determinants of air- line choice. The model results show that the difference in the four groups is affected by age, income, purpose of travel, fares, and occupation. Furthermore, passengers indicated that for domestic, regional and in general flights they prefer LCC while for international flights they prefer FSC.

References

Ashford N, Bencheman M (1987). Passengers' choice of airport: an application of the multinomial logit model. Transport Research Record, 1147.

Chang LY, Sun PY (2012). Stated-choice analysis of willing-to-pay for low cost carrier services. Journal of Air Transport Management, 20: 15e17

Davidson L, Ryley T (2010). The development of airport choice models using stated preference data. Transport Studies Group. Loughborough University, Leicestershire, UK.

Garrow LA (2010). Discrete choice modelling and air travel demand: Theory and applications

Greene WH, Hensher DA (2010). Modelling ordered choices. Cambridge University Press.

Hess S, Polak JW. 2005. Accounting for random taste heterogeneity in airport-choice modelling. Transport Research Record 1915: 36-43.

Hess S, Adler T, Polak JW (2007). Modelling airport and arline choice behaviour with the use of stated preference survey data. Transportation Research. Part E 43(3): 221-233.

Mamdoohi AR, Yazdanpanah M, Teherpour A, Saffarzadeh M (2013). An analysis of resident and non-resident air passenger behaviour of origin airport choice. International Journal of Transport Engineering, 2:1

O'Connell JF, Williams G (2005). Passengers' perceptions of low cost airlines and full service carriers: a case study involving Ryanair, Aer Lingus, Air Asia and Malaysian Airlines. Journal of Air Transport Management, 11:259-272.

Pels E, Nijkamp P, Rietveld P (2001). Airport and airline choice in multi-part region: an empirical analysis for the San Francisco Bay area. Regional Studies, 35: 1-9.

Pels E, Nijkamp P, Rietveld P (2003). Access to and competition between airport: a case study for the San Francisco Bay Area. Transport Research. Part A 37(1): 71-83.

Proussaloglou K, Koppelman FS (1999). The choice of air carrier, flight and fare class. Journal of Air Transport Management 5(4): 193-201

Stefano DL (2012). Modelling airport choice behaviour for direct flights, connecting flights and different travel plans. Transport Geography 22: 148-163.

Tabachnick BG, Fidell LS (2012). Using multivariate statistics (6th Edition). Boston, MA: Person.

Xiao YB, Chen J, Liu XL (2008). Joint dynamic pricing for two parallel flights based on passenger choice behaviour. System Engineering: Theory and Practice, 28(1): 46-55.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Zero-augmented models for exploring the factors affecting the pass rate of 2016 grade 10 learners in Khomas region, Namibia

Rapikama Mumbuu[1*], Lilian Pazvakawambwa[2] and Opeoluwa Oyedele[2]

[1]Department of Mathematics & Statistics, Namibia University of Science and Technology, Windhoek
[2]Department of Statistics and Population Studies, University of Namibia, Windhoek
*Corresponding author: rmumbuu@nust.na; Tel: 061-2072543

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Hurdle model, zero augmented models, Generalized Linear Models, Zero-Inflated models

ABSTRACT

The poor performance of grade 10 learners has been a big concern over the last few years and in the effort to understand this phenomenon there has been efforts to present models that explain it. This study aimed at exploring the factors which influence Khomas Region grade 10 learners' pass rate using Generalized Linear Models (GLMs). The data used for this study was obtained from the Directorate of National Examination and Assessment for the year 2016, with permission from the Permanent Secretary of the Ministry of Education (DNEA). With the presence of excess zeros in the study data, six GLMs were explored (Poisson, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial) to assess their goodness of fit on modelling the zero-inflated DNEA count data. Afterwards, the better performing GLM was used in achieving the study aim. The Zero- Inflated Negative Binomial performed better based on its lowest Akaike Information Criterion (AIC) values among the six fitted GLMs. Results from the fitted Zero- Inflated Negative Binomial model revealed that the age of the learner, school location and the type of school (private/state) had significant differential in the pass rate of grade 10 learners, with p-values < 0.05 in the Zero- Inflated Negative Binomial model. Thus, it is recommended that for densely populated areas, emphasis should be put on building more schools in these areas so that classrooms are not overcrowded per subject. In addition, overaged learners should also be given extra assistance such as extra classes and extra motivation.

1. Introduction

In the era of globalization and technological revolution, education is considered as the first step for every human activity. Education plays a vital role in the development of human capital and is linked to an individual's well-being and opportunities for a better living (Saxton, 2008). It is one of the most powerful instruments known for reducing poverty and for laying the basis for a sustainable economic growth. It ensures the acquisition of knowledge and skills that enable individuals to increase their productivity and improve their quality of life (Battle & Lewis, 2002). This increase in productivity also results in more employment opportunities which enhance the economic growth of a country. In addition, education can be viewed as a process through which the intellectual, moral capacities, proper conduct, and technical competency of individuals are developed to make them cultural members of their respective societies (Tuan, 2009).

Studies by Miller-Grandvaux & Yoder(2012) on secondary schools education revealed that the main challenges in secondary school education seem to be the academic performance of learners. Generally, the academic performance of learners varies from learner to learner, school to school, location to location and country to country. Anecdotal evidence indicates that the school location, environment, inadequate facilities and infrastructure are some of the factors that account for the differences in academic performance of learners across different subjects. Although this study focused on the grade 10 learners' performance in the Khomas region, the problem of poor academic performance is a national debacle.

In Namibia, the education system is divided into three stages, namely primary level, secondary level and tertiary level. The primary level, Grade 1 to 7, prepares children for secondary education. In other words, primary education is the basic education provided at primary school level. On the other hand, the secondary level stretches over a period of 5 years from Grade 8 to Grade 12 (Namibia Government, 2001). Learners are presented with a Junior Secondary School Certificate (JSC) after successfully completing Grade 10, and they get a Senior Secondary Certificate at the end of Grade 12 (NSSC).

Since 1993, grade 10 learners in Namibia, regardless of the type of school attended have written the National Junior Secondary Examinations (NJSE) administered by the Directorate of National Examinations and Assessment (DNEA) in Namibia. The NJSE is compulsory for all registered grade 10 learners in Namibia and is used to assess the achievement of learners in a curriculum in order to provide an estimate of the learners' achievement level in the education system. In addition, it is used to make performance comparisons among the 14 regions of Namibia, to further identify schools/regions in need of interventions.

Factors leading to poor performance in secondary schools

Poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents' educational level, occupational status and income level. It is believed that a low socio-economic status negatively affects the academic achievement of learners in secondary schools (David, 2014). David (2014) further elaborated that learner performance is dependent on the socio-economic background (SEB). They observe that, "High school learners' level of performance is with statistically significant differences, linked to their sex, grade level, school location, school type, learner type and socio-economic background (SEB)." Considering the physical geographical location of most secondary schools in the Sumbawanga District is rural, and the physical infrastructure is poor and limited, the communities might be affected by low socio-economic which influence academic performance (David, 2014).
Several studies have been carried out to identify and analyze the numerous factors that affect academic performance in various centres of learning. Their findings identified factors such as the learners' efforts, literacy level of parents' education, parental involvement (Jeyness, 2012); self-motivation, the age of learners, learning preferences (Obiero, Mwebi, & Nyang'ara, 2017); class attendance and entry qualifications as factors that have a significant effect on the learners' academic performance in various settings
The influence of age and sex on academic performance has been investigated in a number of studies with widely differing conclusions. Research has also shown that men perform better than women in certain settings while women outperform men in other settings (Sommerville & Singaram, 2018). Scholarly observations show that recent changes in educational policies around the world have led to an increase in the number of mature-age admissions in educational institutions (Sommerville & Singaram, 2018).
The relationship between sex and the academic achievement of learners has been contested. However, a gap between the achievement of boys and girls has been found, with girls showing better performance than boys in certain instances (David, 2014). According to Considine and Zappala (2002), the educational performance in school has also been influenced by the learner's sex. Boys suffer an educational disadvantage relative to girls, especially in terms of performance in literacy. Several explanations for this increasing sex gap which include: biological differences; sex biases (such as reading the fact is seen as not being masculine); teaching, curricula and assessment (for instance less structured approaches to teaching grammar) may have weakened boys (Considine & Zappala, 2002).
According to Jeyness (2012) poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents' educational level, occupational status and income level. In addition, the low Socio-Economic Status (SES) negatively affects academic achievement of learners in secondary schools (Hansen & Mastekassa, 2013). While a positive relationship between self-motivation and academic performance has been established, the effect of the family's income and parents' level of education on academic performance is far from being unraveled without equivocation. The Socioeconomic status of learners and their families show moderate to strong relationships with academic performance (Jeyness, 2012). However, these relationships are contingent upon a number of factors such that it is nearly impossible to predict academic performance using SES.
A study conducted by Orlu (2013) among six hundred teachers and learners aimed at establishing the environmental influence on the academic performance of secondary school learners. It was found that the school environment has a significant

influence on academic performance, and that its location can affect learners' performance. For example, when a school is situated in a noisy area like an airport or in the heart of a city where activities disrupt the teaching and learning of the learner, one would not expect the learners to do well academically. In fact, noise in any learning environment interferes with the teaching and learning process.

Overcrowding is another factor that affects the teaching and learning enviroment. Chuma (2012) observes that overcrowding in classrooms makes it difficult for pupils to write. The teacher is also unable to move around the class freely to assist needy pupils and this affects the teaching-learning process. This means that crowded classroom conditions not only make it difficult for learners to concentrate but inevitably limit the amount of time teachers can spend on innovative teaching methods such as cooperative learning and group work (Chuma, 2012).

Parents' occupations also influence the learners' achievement in academic work. This is a result of the parents' levels of investment in their children's education which determine their level of purchasing capacity. Learners' academic achievement is negatively correlated with the low level of the parent's income which hinders the individual from gaining access to sources and resources of learning (Jeyness, 2012).

Those studies carried out did not identify the determinants of poor academic performance of the learners, such as, age, sex, location of the school and types of school. Most of the studies done focused on factors influencing the performance of learners in secondary school not specifically in grade 10. Nevertheless, the literature review was used to provide general information on factors influencing learners' performance in general.

Statistical models for count data

The models that are developed to handle count data are normally the Poisson regression and the negative binomial. However due to excess zeros, the hurdle and zero inflated models become very important models in studies on count data. General linear models although very useful, have limitations, such as when the response is restricted to binary and count and when the variance of the response depends on the mean. However, the Generalized linear models (GLM's) extend the general linear framework to address both of the above issues (Zeileis, Kleiber, & Jackman, 2008). GLM involves probability that can be expressed in exponential form. Such distributions are members of the exponential family of distributions written as:

$$f(y;\theta,\phi) = \exp\left[\frac{(y\theta - b(\theta))}{a(\phi)} + c(y,\phi)\right] \quad (1)$$

where a(.), b(.) and c(.) are some functions, with θ being a function of the location parameter of the distribution (e.g. the mean). This exponential family of distributions include well-known distributions such as the normal distribution, the Poisson distribution and binomial distribution (Zeileis, Kleiber, & Jackman, 2008).

Zero-inflated count models

An alternative approach for modelling zero-inflated data is the zero-inflated count model proposed by Lambert (1992). This model assumes that data are from a mixture of a regular count distribution, such as the Poisson distribution, and a degenerate distribution at zero. The EM algorithm or the Newton–Raphson method can be used to obtain the maximum likelihood estimates. Compared to the hurdle model, this model is more complex to fit, as the model components must be fitted simultaneously.

A zero-inflated negative binomial regression model (With hidden Markov chain)

Wang & Alba (2006) consider a random variable Y of event counts with a piece of data set of k subjects, $\{(y_{ij}, x_{ij}); i = 1, \ldots, k, j = 1, \ldots, n_i\}$, where y_{ij} is observed event counts for subject i during the jth period, associated with a vector of covariates x_{ij}, and the total sample size $n = \sum_{i=1}^{k} n_i$. The proposed model assumes that:

(1) for observed event counts y_{ij} for subject i during period j, there corresponds a partially observed binary random variable, S_{ij}, representing the state of a two-state discrete time Markov chain with $S_{ij} = 1$ when $y_{ij} > 0$ and $S_{ij} = 0$ or 1 when $y_{ij} = 0$;

(2) The partially observed binary random vector $(S_{i1}, S_{i2}, \ldots, S_{in})$ for subject i follows the two-state discrete time Markov chain with transition probabilities defined by

$$\Pr(S_{ij} = 0|S_{i(j-1)} = 0) = p_{00}, \Pr(S_{ij} = 1|S_{i(j-1)} = 0) = p_{01} = 1 - p_{00} \quad (2)$$

$$\Pr(S_{ij} = 1|S_{i(j-1)} = 1) = p_{11}, \Pr(S_{ij} = 1|S_{i(j-1)} = 1) = p_{10} = 1 - p_{11} \quad (3)$$

where p_{00}, p_{01}, p_{10} *and* p_{11} are unknown parameters. For observed count y_{ij}, (3) conditional on $S_{ij} = 1$ follows a Negative Binomial distribution with

$$f_1(y_{ij}|x_{ij}, \alpha, \beta, S_{ij} = 1) =$$

$$\frac{\Gamma(y_{ij}+\alpha^{-1})}{y_{ij}!\Gamma(\alpha^{-1})}\left(\frac{\alpha\lambda_{ij}}{1+\alpha\lambda_{ij}}\right)^{y_{ij}}\left(\frac{1}{1+\alpha\lambda_{ij}}\right)^{\alpha^{-1}} \quad (4)$$

where $\lambda_{ij} = \exp\,(\beta' x_{ij}, \beta = (\beta_1, \ldots, \beta_d)$ is an unknown parameter vector and $\alpha \geq 0$ is the dispersion parameter; conditional on $S_{ij} = 0, y_{ij} = 0$, i.e.

$$f_0 = (y_{ij}|S_{ij} = 0) = \begin{cases} 1, & if\ y_{ij} = 0 \\ 0, & if\ y_{ij} > 0 \end{cases} \quad (5)$$

The Zero-inflated models are useful when the data contains excess zeros that are both structural and non-structural (sampling zeros) (Hall, 2000). These models have been extensively used in fields such as econometrics and medical fields (Hall, 2000).

This paper adopts the conceptual framework of David (2014) who stated that the academic performance of learners in secondary schools over a given period of time may be influenced by socio-economic factors originating from their families, school environment and the learners themselves. In addition, the article notes that socio-economic, socio cultural and socio-political variables influence learners' performance directly or indirectly, either by the increase or decrease of the number of learners' average scores of grades A, B and C for Pass and grade D and F for Fail. The main goal of the study was to investigate factors affecting the pass rate of grade 10 learners across schools in the Khomas region, using the Junior Secondary Certificate (JSC) examination results for the year 2016 obtained from the DNEA. The specific objectives of the study were to explore the various models that can be potentially applied to analyse the relationships between the pass rate and the demographic and socio-economic variables; apply the best model to analyse the relationship between the pass rate and demographic and socio-economic variables; estimate the effects of demographic and socio-economic variables on the pass rate based on the best model and to suggest measures and strategies that can be used to improve the pass rate of grade 10 learners in the Khomas region.

2. Methods

The quantitative cross-sectional study was based on secondary data obtained from the Directorate of National Examinations and Assessment (DNEA) in the Ministry of Education's database for the year 2016. The population of this study were the 45 schools in the Khomas region that offer grade 10 education whereby all the 45 schools were used for analysis thus, there was no sampling performed. The dependent variable for this study was the number of subjects passed and the independent variables used were learner's sex, age, school location (urban, semi-urban, or rural), and type of school (government or private). Descriptive statistics was done to graphically explore the data and provide some basic summary statistics. Six models were explored, namely the Poisson regression, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial models. The hurdle Poisson and the hurdle Negative binomial and Zero-Inflated models are used to account for variables with many zeros, particularly in our case to analyze the variable "number of subjects passed. The Poisson regression is susceptible to over dispersion and the Quasi Poisson as well as the Negative Binomial are useful when there is over dispersion, which means that the variance is higher than the mean. The analysis of this study did not explore the Quasi Poisson due to some limitations experienced during the R programming. Models with the lowest AIC's were more preferable. The data was analysed using both the R programming software version 3.3.1 and the Statistical Packages for Social Sciences (SPSS) software. Several in-built R packages (such as *MASS, pscl* and *AER* packages) were used to handle cases involving the hurdle models and the Zero-Inflated models. Data for this research was obtained from the DNEA broad sheets provided to each region that contains distribution of subject's rankings, results per school, school name etc.

3. Results

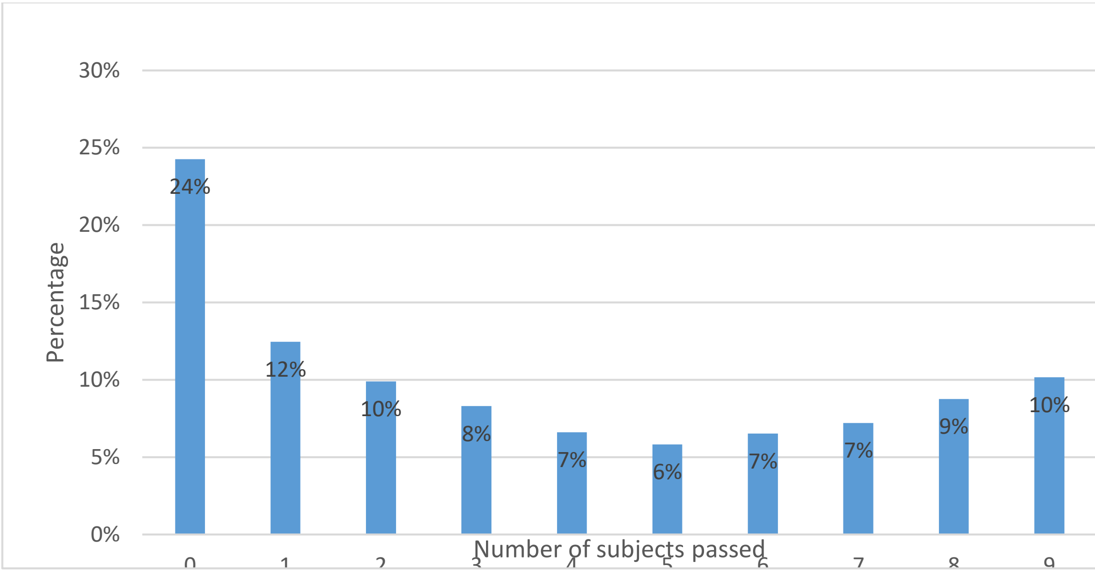
Descriptive Profile of the Pass Rate across Sex

Table 1 shows that the failure rate in 2016 from the Khomas Grade 10 results was higher (55%) than the pass rate (45%). Hence, the study's intention is to explore factors that affect the pass rate in the Khomas region.

Table 1: Pass rate vs Sex of grade 10 learners

	Sex					
	Female		Male		Total	
	Count	%	Count	%	Count	%
Fail	1527	53.4%	1305	57.0%	2832	55.0%
Pass	1333	46.6%	985	43.0%	2318	45.0%
Total	2860	100.0%	2290	100.0%	5150	100.0%

Figure 1: Bar chart showing the distribution of the number of subjects passed.



The bar chart, Figure 1, reveals that zeros are in excess, compared to the remaining groups. Thus, a hurdle model or zero inflated model was more appropriate to use in the inferential data analysis section to achieve the objectives of this study.

Table 2: Distribution of the types of school across school location

		Location							
		Low Density		High Density		Rural		Total	
		Count	%	Count	%	Count	%	Count	%
Type	Private	51	2.2%	314	11.8%	69	45.4%	434	8.4%
	State	2296	97.8%	2337	88.2%	83	54.6%	4716	91.6%
	Total	2347	100.0%	2651	100.0%	152	100.0%	5150	100.0%

Table 2 shows that of the total number of schools in the high density location, 434 (8.4%) were privately owned while 4716 (91.6%) were state owned.

Models Comparison

Six different GLMs (Poisson regression, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial) were fitted to analyse the effect of the independent variables on the pass rate of grade 10 learners in the Khomas Region. Table 3 shows the obtained AIC and log-likelihood values for these fitted models. The purpose here was to make a comparison that would yield the best model for the analysis of the variability in the pass rate of grade 10 learners in the Khomas Region. The Poisson regression and Negative Binomial models in Table 4 were fitted using the *glm* package in R, while Hurdle Poisson model, Hurdle Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models were fitted using the *pscl* and *AER* packages in R.

Table 3: The AIC and log-likelihood values of the 6 fitted GLMs for the DNEA data

	Model	AIC	**Log-likelihood**
1	Poisson regression	26496	-13242.2
2	Negative Binomial	23771	-11878.35
3	Hurdle Poisson model	23448.58	$-\mathbf{1.71} \times \boldsymbol{e^4}$
4	Hurdle Negative Binomial	22995.1	$-\mathbf{1.148} \times \boldsymbol{e^4}$
5	Zero Inflated Poisson	23448.55	$-\mathbf{1.71} \times \boldsymbol{e^4}$
6	Zero- Inflated Negative Binomial*	22988.84	$-\mathbf{1.148} \times \boldsymbol{e^4}$

*best model

From Table 3 the lowest AIC value (22988.84) and the highest log likelihood occurred when using the Zero-Inflated Negative Binomial (model 5), which means that the Zero- Inflated Negative Binomial model was the best option for explaining the variability in the grade 10 results of the Khomas region of 2016. Results from the Zero-Inflated Negative Binomial model are presented in Table 4.

Table 4: Zero – Inflated Negative Binomial model coefficients (Binomial with logit link)

Variable	Estimate	Std. Error	P-value	OR(Odd Ratio)
Intercept	-15.122	0.879	<0.001***	0.000
Sex				
Male	-0.015	0.087	0.861	0.985
Female (ref)				
Age	0.675	0.039	<0.001***	1.964
Location				
High density	0.131	0.255	0.608	1.140
Low density	-0.788	0.261	0.0033***	0.455
Rural (ref)				
School type				
State	1.831	0.337	<0.0001***	6.240
Private (ref)				

***significant at 5 % level of significance.

Table 4 above reveals that the odds of having no subject passed among male learners is 0.985 times lower than the odds of having no subject passed among females. However, the p-value of 0.861 (in Table 4) is larger than 0.05, hence one can conclude that the male children do not necessarily fail more subjects than female learners. A one unit increase in the number of years (being one year older), increases the odds of not passing a subject by 1.964 times. A learner in a low-density area has a reduced chance of 0.455 (45.5%) of failing a subject compared to a learner in a rural school. Being in a state school, increases the chance of not passing a subject by 6.240 compared to a learner in a private school.

Table 5: Zero-Inflated Negative Binomial model coefficients (with log link):

Variable	Estimate	Standard error	P-value	Expected Rate (ER)
Intercept	5.425	0.219	0.000***	227.011
Sex				
Male	0.022	0.021	0.295	1.022
Female (ref)				
Age	-0.220	0.012	0.000***	0.803
Location				
High density	-0.454	0.065	0.000***	0.635
Low density	-0.208	0.065	0.001***	0.812
Rural (ref)				
School type				
State	0.300	0.051	0.000***	1.350
Private (ref)				

***significant at 5 % level of significance.

Table 5 reveals that a male learner has a higher chance of 1.022 times of passing than a female learner. However, among learners with a positive number of subjects passed, the p-value of 0.295 is larger than 0.05, hence male children do not necessarily pass more subjects than female learners. The value of 0.803 indicates that if a learner's age increases by one year among learners with a positive number of subjects passed, it will lead to the reduction of the number of subjects passed by a learner by 0.803 times. A learner who is attending school in a highly populated area has a reduced chance of 0.635 of passing subjects compared to a learner in a rural school, given that the learner has a positive number of subjects passed. Moreover, a learner attending school in a low-density area has a low chance of 0.812 times of passing subjects than a learner in a rural area, given that the learner has a positive number of subjects passed. The value of 1.350 indicates that the learner at the state school has a higher chance of 1.350 times of passing more subjects than a learner at the private school, given that the learner has a positive number of subjects passed.

4. Discussion

This study concluded that the poor performance of the grade 10 learners is a challenge in the Khomas region, Windhoek. It was found that the age of the learners, location and school had an effect on the performance of learners. The study also established that the sex of the learners had no impact on their overall performance. As such, the study concurred with Considine & Zappala's (2012) research on factors influencing the educational performance of learners from disadvantaged backgrounds. Although they fitted a Binomial Logistic regression to estimate the extent to which individual, family, behavioural and socio-economic factors contribute to learners' achievement, the results from the Wald test revealed that the coefficients were statistically significant for sex, ethnicity, and parental education. However, in terms of the location, the study revealed that schools in low-density areas performed better than rural schools. They concluded that the geographical location did not significantly predict school performance outcomes. This study however yielded the same results in the case of learners' performance in highly populated areas, where the variable was insignificant in the Zero Inflated Poisson model.

5. Conclusions

The Zero-Inflated Negative Binomial performed better based on its lowest AIC values among the six fitted GLMs. The results revealed that the age of the learner, school location and the type of school (private/state) had significant differential in pass rate with p-values less than 0.05 in the Zero- Inflated Negative Binomial model.

Based on the findings of this study, the following recommendations are made that could be implemented to alleviate the poor performance of the Grade 10 learners on the national examinations, especially in Khomas region.

1. State owned schools should strive to have the same privileges, infrastructures and teaching methodologies as private schools.
2. Emphasis should be put on building more schools in the area so that classrooms are not overcrowded.
3. Rural schools should be given the same attention as urban schools. The current bush allowance should be improved to attract qualified teachers to rural schools.
4. The state schools should bring the teacher-learner ratio of 1:40 in secondary schools on par with the private schools' ratio of 1:25 to give more attention to the slow learners and ease the marking load which consume teachers' time of preparation for lessons that will promote effective teaching and learning.
5. The Ministry should make sure that the teachers employed in state schools are qualified to teach at the relevant levels unlike the current prevailing situation whereby principals appoint and keep unqualified teachers at the expense of qualified ones.
6. Students should be motivated to focus on studies when they are still young.

References

Battle, J., & Lewis, M. (2002). The increasing significance of class: The relative effects of race and socioeconomic status on academic achievement. *Journal of Poverty*, 21-35.

Chuma, P. C. (2012). *Challenges Affecting Teaching-Learning in Primary Schools in Kenya. A case study of Central Division Mandera East District.* Nairobi: Moi University.

Considine, G., & Zappala, G. (2002). Factors influencing the educational performance of students from disadvantage background. *Competing visions*, 91 107.

David, N. M. (2014). *Determinants of poor academic performance of secondary school in Sumbwanga district, Tanzania.* Sokoine: Sokoine University of Agriculture.

Hall, D. (2000). Zero-inflated Poisson and binomial regressions with random effects: a case study. *Biometrics*, 1030-1039.

Hansen, N. M., & Mastekassa, A. (2013). *Social origins and academic performance at university.* Oxford: Oxford University Press.

Jeyness, W. (2012). A Meta-Analysis of the Efficacy of Different Types of Parental Involvement Programs for Urban Students. *Urban Education*, 706-742.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* .

Miller-Grandvaux, y., & Yoder, K. (2012). *A literature review of community schools in Africa.* Washington DC: Bureau for Africa.

Namibia Government. (2001). *Edcuation Act No. 16.* Windheok: Ministry of Education.

Obiero, K., Mwebi, B., & Nyang'ara, N. (2017). Teacher Absenteeism in Public Secondary Schools in Borabu Sub-County. *International Journal of Education and Research*, 130-133.

Orlu, C. (2013). Environmental Influence on the Academic Performance of Secondary School Students in Port Harcourt Local Government Area of River State. *Journal of Economic and Sustainable Development*, 23.

Saxton, J. (2008). Investment in Education. *Private and Public Returns*.

Tuan, L. (2009). Teaching English Discrete Sounds through Minimal Pairs. *Journal of Langauage Teaching and Research*, 540-561.

Wang, P., & Alba, J. (2006). *A zero-inflated negative binomial regression model with hidden Markov chain.* Singapore: Nayang Technological University .

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Jouranal for Statistical Software*.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Assessing the Impact of Proximate and Non-Proximate Determinants of Fertility in Namibia: A Structural Equation Modelling Approach

Tobias Shinyemba[1*], Opeoluwa Oyedele[1], Lawrence N. Kazembe[1], Martin Shaalukeni[1]

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: tshinyemba@unam.na

ARTICLE INFO

Received: March 2021
Accepted: October 2021

Keywords:
Fertility, Namibia, Structural Equation Model

ABSTRACT

Fertility rate has been declining over the years in Namibia, and a number of studies have been conducted to investigate how socio-economic and physiological factors influenced fertility decline. This study was aimed at modelling the direct and indirect effects of socio-economic, socio-demographic and health attributes on fertility, as well as the proximate and non-proximate determinants of fertility using the Structural Equation Modelling (SEM) technique and the data from the 2013 Namibia Demographic Health Survey. To be precise, the confirmatory factor analysis part of the SEM technique was used to test the theorized model of the proximate and non-proximate determinants of fertility, while the factor modelling part was used to measure the effects that these two constructs of fertility determinants had on fertility. Results from this study showed that the proximate determinants had a direct negative impact $\beta =$ -0.023) on the number of children ever born, while there was a (significant) positive effect $\beta =$ 0.053) between the non-proximate determinants and the number of children ever born. In addition, age at first birth had a (significant) positive effect on the number of children ever born by Namibian women while the effect of contraceptive use was found to have a minor effect. Moreover, women who had their first birth at the beginning of their reproductive period were more likely to have more children born to them, while women who had their first marriage at younger ages were more likely to have more children. It is therefore recommended that there is a need to promote contraceptive use among Namibian women to further reduce fertility, especially among women from poor households as the cost of rising children has become high as the year progresses. Additionally, there is also a need to promote and strengthen the education of young females in order to increase their age at first birth and at first marriage in Namibia.

1. Introduction

Globally, the Total Fertility Rate (TFR) has declined from 2.58 children per woman to 2.47 children per woman between 2010 and 2019 respectively (UN-DESA, 2019). A similar trend was also observed in Sub-Saharan Africa with the TFR dropping from 5.40 children per woman to 4.72 children per woman during the same period (UN-DESA, 2019). In Namibia, the TFR dropped from 5.4 children per woman in 1992 to 3.6 children per woman in 2013, and it was projected in 2011 to still have a linear decline to 2.4 children per woman by 2041 (NSA, 2014b, p. 2). Additionally, the TFR varied between the urban and rural areas and across regions in Namibia. To this effect, it was found that fertility was lower in the urban areas with 3.2 children per woman compared to the rural areas with 4.9 children per woman (NSA, 2014a). With respect to regions, a low TFR was recorded for the Khomas (3.0), Oshana (3.2) and Erongo (3.2) regions, whereas the Kunene (5.3), Ohangwena (4.9) and Omaheke (4.7) regions recorded a high TFR (NSA, 2014a). Moreover, fertility (levels) differs among women due to socio-economic factors such as educational attainment, wealth status and occupation.

NSA (2014a) reported that education attainment, occupation, and wealth status were closely related to fertility, in that as the years of education rises, the TFR varied from 5.7 children per woman among women with no education to 3 children among women with secondary and tertiary education. In addition, women who were widowed recorded a lower TFR of 4.2 children per woman, followed by women who were divorced/separated and never married with a TFR of 3.9 and 3.1 children respectively, while women who were in consensual unions and married women with certificate/traditionally recorded a TFR of 6.3 and 6.0 children respectively (NSA, 2014a). By economic activities, women who were homemakers had a high TFR (5.9) compared to unemployed women (5.13), employed women (4.01) and those who were students (1.98) (NSA, 2014a).

Over the year, most studies done on fertility in Namibia and other parts of the world found that recent fertility decline was attributed to factors such as increment in the costs of rising children, mother's educational attainment, household's wealth, years at first marriage and first birth, contraception use, abortion and induced abortion, postponement of first birth, employment status, postponement of marriages, changing marriage patterns, and postpartum infecundability (Chola & Michelo, 2016; Indongo & Pazvakawambwa, 2012; Islam et al., 2016; Johnson, Abderrahim, & Rutstein, 2011; Majumder & Ram, 2015; Palamuleni, 2017; Shinyemba, 2014). In addition to the foresaid factors, the Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) was found to have a moderate influence on fertility decline in some countries in recent years (Fortson, 2009; Johnson et al., 2011; Milly Marston et al., 2017; Marston, Zaba, & Eaton, 2017). However, the implication of HIV/AIDS on fertility in Namibia has not yet been sufficiently quantified. A study done in Namibia by Palamuleni (2017) concluded that the presence of HIV/AIDS might have changed the attitudes and behaviour of individuals regarding pre-marital sexual intercourse, having multiple sexual partners and postponement of marriages. His argument concurs with that of Johnson et al. (2011) who claimed that women fertility preferences were changing due to the Human Immunodeficiency Virus (HIV) pandemic as women who are HIV positive were more likely to use contraceptives to avoid infecting and/or re-infection from their partners, thereby reducing their risk of falling pregnant.

In Namibia, fertility decline can be attributed to several factors that had both direct and indirect effects, with each factors affecting fertility differently. Although there are number of studies done in Namibia documenting fertility decline and factors contributing to the trend (Indongo & Pazvakawambwa, 2012; Palamuleni, 2017; Shemeikka, Notkola, & Siiskonen, 2005; Shinyemba, 2014), no studies have been done (in Namibia) to assess the impact that socio-economic, socio-demographic and health attributes might have on fertility. Hence, the effect that these attributes have on fertility remains unknown, and presently lawmakers and researchers cannot quantify the enormity of these attributes on fertility. Therefore, the aim of the study was to examine and model the direct, indirect and joint effects of socio-economic, socio-demographic and health attributes on fertility for Namibian women, using the Structural Equation Modelling (SEM) approach.

2. Methods

2.1. Data and Sample

The data used in this study was obtained from the 2013 Namibia Demographic Health Survey (NDHS) carried out by the Ministry of Health and Social Services (MoHSS). This survey was designed to provide demographic, socio-economic and health information necessary for policy making, planning, monitoring and evaluation of national health and population programme in Namibia. For the NDHS, 9,176 women in the reproductive age group of 15 to 49 years were interviewed. For this study, the number of Children Ever Born (CEB) per woman was used as a proxy of fertility (for the women). The independent variables of this study (which were the socio-economic, socio-demographic and health attributes of the women) were classified into the indirect cause (age, place of residence, level of education, working status, exposure to mass media, fertility preferences, health care during pregnancy and wealth index), and the direct cause (contraceptive use, age at first marriage, age at first birth, duration of breastfeeding and age at first sexual intercourse). The choice of attributes used was derived from reviewed literature. These variables were further classified as proximate and non-proximate determinants of fertility. Table 1 shows the list of variables used in this study as well as their codenames.

2.2 Statistical Analysis

Structural Equation Modelling (SEM) is a complex statistical modelling technique that attempts to describe the structural relationships between observed and unobserved variables, with a basic goal of providing a quantitative test of a theoretical model hypothesized by a researcher (Sánchez, Budtz-Jørgensen, Ryan, & Hu, 2005). To be precise, SEM is a combination of the Confirmatory Factor Analysis (CFA) and multiple regression analysis (often termed the factor modelling) approaches.

Table 1 List of variables used and their codenames

Variable codename	Variable
V012	Respodents current age
V025	Type of place of residence
V130	Religion
V106	Educational Attainment
V157	Fequency of reading Rewspapers
V158	Fequency of listening to the Radio
V159	Frequency of watching Television
V190	Wealth Index
V201	Total Children Ever Born
V212	Age at first birth
V312	Current contarceptive method
V404	Breastfeeding
V501	Marital status
V511	Age at first cohabitation
V525	Age at first sex
V602	Fertility preference
V714	Working status

It works by combining the CFA and factor modelling approaches to analyze the structural relationship between the observed (manifest) variables and unobserved (latent) variables. CFA enables a researcher to confirm or reject the defined theorized model, which state that there is a relationship between the observed (manifest) and their underlying latent (unobserved) variables (Hair, Black, Babin, & Anderson, 2009). CFA forms part of SEMs, as the techinique uses standardized regression scores to relate unobserved variables to observed variables using using path diagrams. There are three steps involved during the CFA model building. The first step deal with the estimation of the (theorized) model parameters, while the second step deals with the identification of parameters in the model. The third stage assesses how well the estimated model predicts the covariance matrix of the manifest variables (Schumacker & Lomax, 2010; Shinyemba, Nickanor, & Kazembe, 2019). The CFA model can be summarized as follows. First, the observed variables $x's$ are collected into a vector x for each individual subject, with x_i denoting the observable variable i, for $i = 1,2, \dots, m$. That is,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

Secondly, the vector of variables x is assumed to be a random vector sampled from a population with a mean vector μ, where μ is defined as

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}$$

Here, the unobserved common factors f_i are collected in a vector f as follows:

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

On the other hand, the factor modelling part of the SEM is a series of multiple regressions, which are predicting each of the observable variables x_i from the values of the unobserved common factors f_i. The regression coefficients are obtained as follows:

$$\begin{gathered} X_1 = \mu_1 + l_{11}f_1 + l_{12}f_2 + \cdots + l_{1m}f_m + \varepsilon_1 \\ X_2 = \mu_2 + l_{21}f_1 + l_{22}f_2 + \cdots + l_{2m}f_m + \varepsilon_2 \\ \vdots \\ X_p = \mu_p + l_{p1}f_1 + l_{p2}f_2 + \cdots + l_{pm}f_m + \varepsilon_p \end{gathered}$$

where μ_i to μ_p are the intercept terms of the regression equation and ε_1 to ε_p are the error terms. The regression coefficients $l_{ij,}$ for $i = 1,2, \dots, p$ and $j = 1,2, \dots, m$ are referred to as the factor loadings, which can be written in a matrix form as:

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdot & \cdot & \cdot & l_{1m} \\ l_{21} & l_{22} & \cdot & \cdot & \cdot & l_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ l_{p1} & l_{p2} & \cdot & \cdot & \cdot & l_{pm} \end{pmatrix}$$

The error terms can also be collected into a vector ε as follows:

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_p \end{pmatrix}$$

Here, the error terms (ε_i) are often referred to as the specific factors of variable *i*. Thus, the regression coefficients in the factor modelling can be written in matrix notation as

$$X = \mu + \mathrm{Lf} + \varepsilon$$

In this study, the SEM was fitted to identify the underlying structural relationship between the indirect and direct determinants of fertility as well as their joint effect. To be precise, the CFA part of the SEM was used to test the theorized model of the proximate and non-proximate determinants of fertility as per Bongaarts (1978) classification, while the factor modelling part was used to assess the relationship between these two constructs of fertility determinants and afterwards model their effects on fertility.

The R software was used to perform the data analyses of this part, with the R packages 'lavaan', 'semTools' and 'semPlots' were used to fit the SEM.

3. Results

Table 2 shows the output obtained from the basic descriptive analysis of all the women considered in the 2013 NDHS. It can be observed that the sexual debut for the women was at the early age of 10 years. The youngest woman gave birth to her first child at the age of 12 years, while the eldest woman gave birth to her first child at the age of 42. The highest numbers of CEB by a woman was 13 children as shown in Table 2. Furthermore, the average age of women in the NDHS was 29.11 years, whereas the average age at first birth was 20.37 years, which meant that most women had their first child birth before they reached the age of 21 years. Moreover, from Table 2, it can be observed that most women started cohabiting at the age of 23.12 years on average, although the average age at first sex was found to be 18.19 years. On average, a Namibian woman had 2.80 children, which implies that each woman had at least 2 children born to her as shown in Table 2.

Table 2 Descriptive analysis of respondents' background

Variables	Minimum	Maximum	Mean
Age of respondents	15	49	29.11
Age at first birth	12	42	20.37
Age at first Cohabitation	10	48	23.12
Age at first Sex	10	46	18.19
Total Child Ever Born	0	13	2.80

3.1 Model Evaluation

Several fit indices can be used to evaluate the appropriateness of the CFA models. The most common fit index used to assess the goodness of fit is the Chi-Square (χ^2) test, where χ^2 value > 200 indicates a good fit for the model. However, Schumacker & Lomax (2010), Brown (2006) and Jöreskog (1969) found the Chi-Square (χ^2) test to be sensitive to sample size effect. Alternatively, the Root Mean Square Error Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) indices can be used to assess the validity of a fitted CFA model. The values of the RMSEA, SRMR, CFI and TLI indices ranges for 0 to 1, with 1 indicating an unacceptable fit (Hu & Bentler, 1999). In this study, a model was regarded as good fit if the RMSEA value ≤ 0.09, the SRMR value ≤0.08, the CFI value ≥ 0.9 and the TLI value ≥ 0.9.

3.2 Model fit: Validating constructs

Two models were fitted to validate the two constructs of fertility determinants and both models were found to be good fit model for the data using the model evaluation criterion discussed above. Even though all the fitted models were found to be good fit, upon inspecting variables contributing to the non-proximate determinants, it was found that the religion (V130) variable was not a significant ($p > 0.05$) predictor of this construct. This model was re-fitted, after the removal of the religion (V130) variable from the model. Comparing the Akaike Information Criterion (AIC) and the Expected Cross-Validated Index (ECVI) values for both the initial model (with religion included) and the re-fitted model (without religion included), the re-fitted model had a lower AIC and ECVI values compared to the initial model. According to Schreiber, Nora, Stage, Barlow, and King (2006), the model with the lowest AIC and ECVI values can be concluded to be a good model fit. Thus, the re-fitted model was selected as the approriate model to use for the non-proximate determinants.

3.3 Full model fitting: linking CEB to proximate and non-proximate determinants of fertility

Here, the full model fitting was done in two stages. In the first stage, the fitted model was assessed to see if the model fits the data well (Tables 3 & 4) and validated to see if all the manifest variables in the model were significant predictors of the study constructs (Table 5). In the second stage, the two constructs of fertility (proximate and non-proximate) were linked to the number of CEB, to capture their direct and indirect effects (Table 6).

Table 3 shows the overall goodness of fit model for the proximate and non-proximate determinants was satisfactory with χ^2_{88}=2,033.328 (>200), with a significant p-value (p<0.001) at a 5% significance level. The result also showed a RMSEA value of 0.082 (90% CI: 0.079, 0.085), and a SRMR value of 0.068, which are below the suggested cut-off value, indicating a good model fit. Additionaly, the GFI value of 0.917 (from Table 4) suggests that the model has a good fit. However, from Table 3, the CFI value of 0.830 and TLI value of 0.797 were below the cut-off value of at least 0.90 or 0.95, implying that the model has an unacceptable fit. Thus, using the model fit indices of Table 4, it can be concluded that the (re-fitted) model has a good fit.

Table 1 Baseline model and fit indices for proximate and non-proximate determinants of fertility

Number of observations	n= 9,176
Estimator	ML
Minimum fit test statistic	2033.328
Degree of freedom	88
P-value(Chi-square) <=0.05	<0.001***
User model versus baseline model	
Comperative Fit Index (CFI)	0.830
Tucker-Lewis index (TLI)	0.797
Root Mean Square Error of Approximation	
RMSEA	0.082
90% confidence interval	(0.079, 0.085)
P-value rmsea <=0.05	<0.001***
Standardized Root Mean Square Residual	
SRMR	0.068

**** Significant at a 5% significance level*

Table 4 Model fit indices of proximate and non-proximate (re-fitted model)

Index	Model Magnitude	Threshold of acceptance
Chi-Square	2,033.328	>200
RMSEA	0.082	<=0.06
SRMR	0.068	<=0.08
GFI	0.917	>0.90

Table 5 shows the result of the estimated model parameters. Since all the manifest variables had sigificant p-values at a 5% level of significance, as shown in Table 5, it can be concluded that all the manifest variables in the model were significant predictors of both non-proximate and proximate determinants. Furthermore, looking at the regression coefficient value for the non-proximate determinants in Table 5, it can be concluded that a significant direct (positive) effect exists between the non-proximate determinants and the number of CEB. To be precise, for every positive change in the non-proximate determinants the number of CEB would increase by 0.502 children. On the other hand, looking at the regression coefficient value for the proximate determinants, it can be concluded that the proximate determinants had a direct (negative) effect on the number of CEB to a Namibian woman. That is for every positive change in the proximate determinants, the number of CEB would reduce by 0.020 children as shown in Table 5.

Table 5 Parameter estimates for proximate and non-proximate determinants of fertility

Latent Variables	Estimates	Std.Err	P-value
Non-proximate =~			
V025	1.000		
V106	-1.387	0.049	<0.001***
V190	-3.986	0.106	<0.001***
V157	-1.824	0.057	<0.001***
V158	-0.848	0.049	<0.001***
V159	-2.273	0.066	<0.001***
V714	-0.699	0.032	<0.001***
V602	-0.152	0.065	0.021***
Proximate =~			
V212	1.000		
V312	17.988	4.902	<0.001***
V404	-0.236	0.076	0.002***
V501	-0.710	0.230	0.002***
V511	15.066	4.134	<0.001***
V525	12.208	3.323	<0.001***
Regression:			
Non-proximate ~ V201	0.052	0.003	<0.001***
Proximate~V201	-0.020	0.006	<0.001***

**** Significant at 5% significance level*

Table 6 displays the results of the standardized factor loadings between the manifest variables and latent variables, with a graphical display shown in Figure 1. From Table 6, it can be observed that wealth index (V190) contributed the most towards the non-proximate determinants construct, with a factor loading of -0.880 and approximately 77.2% of explained variation in non-proximate determinants. Likewise, the fertility preference (V602) contributed the least towards the non-proximate determinants, with a factor loading of 0.040 and approximately 0.2% of explained variation in non-proximate determinants. On the other hand, it can be concluded that age at first birth (V212) and age at first sex (V525) contributed more towards the proximate determinants construct, with a factor loading of 0.820 and 0.720 respectively, and approximately 66.8% and 52.0% of explained variations in proximate determinants respectively as shown in Table 6. Moreover, it can be concluded that marital status (V501) and breastfeeding (V404) had the least contribution towards the proximate determinants construct, with both having a factor loading of -0.11, and 1.2% of explained variation in proximate determinants.

Table 6 Factor loadings for proximate and non-proximate determinants

	Standardized factor loadings	R-squared
	Non-proximate determinants (*n*=9,176)	
V025	0.620	0.385
V106	-0.580	0.336
V157	-0.670	0.452
V158	-0.330	0.109
V159	-0.760	0.573
V190	-0.880	0.772
V714	-0.430	0.187
V602	0.040	0.002
	Proximate determinants (*n*=9,176)	
V212	0.820	0.668
V312	0.070	0.005
V404	-0.110	0.012
V501	-0.110	0.012
V511	0.430	0.187
V525	0.720	0.520

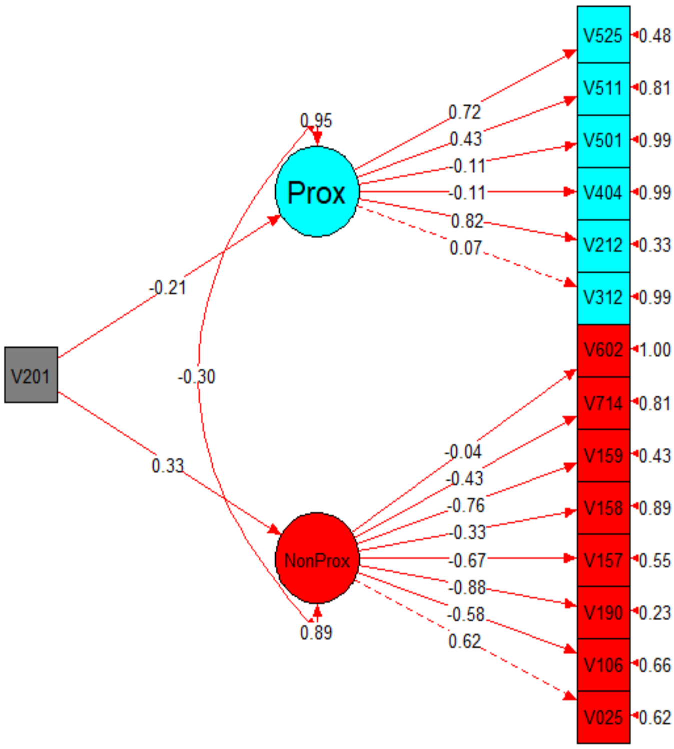


Figure 1 Proximate and non-proximate determinants of fertility factor model structure, where V201=Children Ever Born, Prox=Proximate, NonProx=non-proximate, with factor loadings (middle) and errors (end)

4. Discussions

The study objective was to model the relationship between proximate and non-proximate determinants of fertility using SEM approach. The identified latent constructs were validated using CFA and the analysis was concluded by linking the two constructs (proximate and non-proximate determinants) to the number of CEB, with the relationship between the variables depicted in a path diagram (Figure 1), a process that complete SEM.

The study findings revealed that most women in Namibia started to engage in sexual intercourse at an early age of 10 years old, and this might have increased their likelihood of bearing more children by the time they reach the end of the reproductive period. The study also found that, half of the women had their first union when they were age below 20 years, and this had a significant contribution to high fertility rate. The study further found that educational attainment, wealth index, exposure to media (newspaper, television and radio), employment status were significant predictors of fertility in Namibia.

It was also noted that education had a negative effect on the number of CEB. This maybe due to Namibian women with high level of educational attainment being more likely to have access to better family planning services, and more likely to (choose to) have few(er) children as a result of their contraceptive use and late marriage (postponement of marriage owing to the numbers of years spent in school). This finding is similar to those of Islam et al. (2016), who pointed out that educated women who lived in mid-town areas had fewer children than their uneducated counterparts.

In this study it was found that religion was not a significant predictor of fertility in Namibia. This finding is similar to those of Shinyemba (2014) who concluded that religion was not a significant predictor of fertility among adolescents in Namibia. The data sets used in these two studies were 6 years apart, thus, suggesting that religion still does not play a significant role in reducing fertility.

Exposure to mass media (newspaper, radio and television) was found to have a strong negative effect on the number of CEB to a Namibia woman. This means that women who were exposed to mass media were more likely to have few children compared to those who do not have access to mass media. This is because mass media communication provides a wide variety of information on sexual and reproductive health rights of both women and men, that covers contraceptive use, danger of sexually transmitted disease, violence against woman and girls amongst other. Exposure to mass media also teach women about child spacing and the cost of raising more children. However, Shinyemba (2014) found that exposure to mass media among adolescents had a positive impact on the number of children born by adolescents. This could be attributed to that fact that adolescent girls who watch TV and listen to the radio were more likely get exposed to sexual content broadcasted in the mass media which then triggers their sexual activities, even though contraceptive use among adolescents is known to be very low, consequently resulting in the girls engaging in unprotected sexual acts which most often leads to unplanned pregnancies.

Furthermore, the study showed that among the variables on the proximate determinants construct, age at first birth and age at first sex had the greatest positive effect on fertility. This means that, women who have had first sex at younger ages or a first birth at younger ages were more probable to have more children born, compared to women who had sexual debut at older ages or those that started childbearing at older ages. These findings were similar to those of Motsima and Malela-Majika (2016) who found age at first sexual intercourse to be a significant predictor of fertility outcomes in Lesotho. Similarly, Heywood, Patrick, Smith, and Pitts (2015) also found age at sexual debut to be related to the number of children born by women in the US.

It was also found that contraceptive use had minor negative effect on fertility. Palamuleni (2017) explained that it could be due to stagnant use of contraceptives from 2006. In Zambia, Chola and Michelo (2016) also found contraceptive use to have had a slight influence on fertility decline. However, issues contributing to unprogressively contraceptive use were beyond the scope of this study.

The study also found breastfeeding to have had a minor significant negative impact on the number of children. Suggesting that women who breastfeed were more likely to have fewer children compared to women who did not breastfeed, because breastfeeding increases postpartum infecundability. This finding concurs with those of Alene and Worku (2009) who also found prolonged breastfeeding to be the most important proximate determinants of fertility reduction in Ethiopia.

Moreover, this study showed that the proximate (direct) determinants had a negative significant impact on the number of children, suggesting that for every positive change in proximate determinants the number of children born were more likely to decline. Conversely, Alene and Worku (2009) and Hinde (2014) reported that proximate determinants played a major role in reducing fertility. On the other hand, the non-proximate (indirect) determinants were found to have a significant positive effect on the number of children born, implying that for every positive change in non-proximate determinants, the number of children were more likely to increase.

5. Conclusion

In conclusion the groupings of the socio-economic, socio-demographic and health attributes of fertility differed from the Bongaarts classification of proximate and no-proximate determinants of fertility. The study found that, proximate determinants of fertility had played an important role in the reduction of fertility in Namibia over the years. On the other hand, non-proximate determinants had played as significant role on the increase of fertility.

However, the findings of this study are based on the 2013 NDHS data that was collected about 9 years ago, the results might not reflect the current fertility performance of the population. Also, this study only focused on women fertility, which could have been of a great importance if men fertility were studies together with the women fertility and observe how fertility pattern has changed among men over time. In addition, studying male fertility can help policy makers to formulate policies geared to control male fertility.

The study recomends that there is a need to educate and promote the use contraceptives and prolong breastfeeding among the poor to further reduce fertility, especially among women from poor families as the cost of rising children has become high as the year progresses. Promotion of girl child education also need to strengthened especially in rural areas in order to increase the age at first marriage and first birth. Additionally, it is recommend that further studies be done to investigate factors that caused stagnation in contraceptive use between 2006 and 2013 in Namibia. Further investigation should also focus on evaluating the effect of HIV/AIDS pandemic on fertility in Namibia.

Acknowledgment

The DHS program team is thanked for their approval to use the NDHS data for this study.

References

Alene, G. D., & Worku, A. (2009). Estimation of the total fertility rates and proximate determinants of fertility in North and South Gondar zones, Northwest Ethiopia: An application of the Bongaarts' model. *Ethiopian Journal of Health Development, 23*(1).

Bongaarts, J. (1978). A Framework for Analyzing the Proximate Determinants of Fertility. *Population and Development Review, 4*(1), 105-132. doi:10.2307/1972149

Chola, M., & Michelo, C. (2016). Proximate Determinants of Fertility in Zambia: Analysis of the 2007 Zambia Demographic and Health Survey. *International Journal of Population Research, 2016*, 5236351. doi:10.1155/2016/5236351

Fortson, J. G. (2009). HIV/AIDS and Fertility. *American Economic Journal: Applied Economics, 1*(3), 170-194. doi:10.2307/25760176

Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate Data Analysis* (7 ed.). New Jersey: Pearson Prentice Hall.

Heywood, W., Patrick, K., Smith, A. M. A., & Pitts, M. K. (2015). Associations Between Early First Sexual Intercourse and Later Sexual and Reproductive Outcomes: A Systematic Review of Population-Based Data. *Archives of Sexual Behavior, 44*(3), 531-569. doi:10.1007/s10508-014-0374-3

Hinde, A. (2014). *Demographic Methods*. 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN: Routledge.

Indongo, N., & Pazvakawambwa, L. (2012). Determinants of fertility in Namibia. *African Journal of Reproductive Health / La Revue Africaine de la Santé Reproductive, 16*(4), 50-57.

Islam, A., Hossain, T., Sarwar, G., Kawsar, L., Smrity, L., Alam, A., . . . Bhuia, M. (2016). Structural Equation Modeling to Assess the Impact of Socio-Demographic Variables on Fertility of Ethnic Manipuri Women. *Biom Biostat Int J, 3*(5), 00078.

Johnson, K., Abderrahim, N., & Rutstein, S. O. (2011). *Changes in the direct and indirect determinants of fertility in Sub-Saharan Africa*. DHS Analytical Studies No. 23. Calverton, Maryland, USA: ICF Macro.

Majumder, N., & Ram, F. (2015). Explaining the role of proximate determinants on fertility decline among poor and non-poor in Asian countries. *PloS one, 10*(2), e0115441.

Marston, M., Nakiyingi-Miiro, J., Kusemererwa, S., Urassa, M., Michael, D., Nyamukapa, C., . . . Eaton, J. W. (2017). The effects of HIV on fertility by infection duration: evidence from African population cohorts before antiretroviral treatment availability. *AIDS, 31*.

Marston, M., Zaba, B., & Eaton, J. W. (2017). The relationship between HIV and fertility in the era of antiretroviral therapy in sub-Saharan Africa: evidence from 49 Demographic and Health Surveys. *Tropical medicine & international health : TM & IH, 22*(12), 1542-1550. doi:10.1111/tmi.12983

Motsima, T., & Malela-Majika, J.-C. (2016). The Effects of Early first Sexual Intercourse amongst Lesotho Women Evidence from the 2009 Lesotho Demographic and Health Survey. *African Journal of Reproductive Health / La Revue Africaine de la Santé Reproductive, 20*(2), 34-42.

NSA. (2014a). *Namibia 2011 Census Fertility Report*. Retrieved from Windhoek: https://d3rp5jatom3eyn.cloudfront.net/cms/assets/documents/Namibia_2011_Census__Fertility_Report.pdf

NSA. (2014b). *Namibia Population Projections 2011-2041*. Retrieved from Windhoek: https://cms.my.na/assets/documents/p19dn4fhgp14t5ns24g4p6r1c401.pdf

Palamuleni, M. E. (2017). Determinants of fertility decline in Namibia: An analysis of the proximate determinants. *Bangladesh e-Journal of Sociology, 14*(2), 41-63.

Sánchez, B. N., Budtz-Jørgensen, E., Ryan, L. M., & Hu, H. (2005). Structural Equation Models. *Journal of the American Statistical Association,* *100*(472), 1443-1455. doi:10.1198/016214505000001005

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*(6), 323-338. doi:10.3200/JOER.99.6.323-338

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling*. New York: Taylor and Francis Group.

Shemeikka, R., Notkola, V., & Siiskonen, H. (2005). Fertility decline in North-Central Namibia
an assessment of fertility in the period 1960-2000 based on parish registers. *Demographic Research, 13*, 83-116.

Shinyemba, T. (2014). *Determinants of Adolescent Fertility in Namibia: Evidence from 2006-07 Namibia Demographic and Health Survey.* (Bachelor of Science Honours), University of Namibia, Windhoek. Retrieved from https://www.researchgate.net/publication/273203942_Determinants_of_Adolescent_Fertility_in_Namibia_Evidence_from_2006-07_Namibia_Demographic_and_Health_Survey

Shinyemba, T., Nickanor, N. N., & Kazembe, L. N. (2019). Linking Food Insecurity to Quality of Life Using Structural Equation Models. In N.-B. Kandala & L. N. Kazembe (Eds.), *Statistical Modelling of Complex Correlated and Clustered Data Household Surveys in Africa* (Vol. 1, pp. 103-136). New York: Nova Science Publishers.

UN-DESA. (2019). *World Population Prospects 2019* (Vol. Volume II: Demographic Profiles (ST/ESA/SER.A/427)). New York, 10017, USA: United Nations.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Household poverty levels in Namibia and their associated sociodemographic factors: An empirical investigation of the 2015/16 Namibia household income and expenditure survey

Opeoluwa Oyedele[1*]

[1]Department of Computing, Mathematical & Statistical Sciences, University of Namibia, Windhoek, Namibia
*Corresponding author: OpeoluwaOyedele@gmail.com

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Poverty levels,
Ordered probit model,
household poverty,
Namibia

ABSTRACT

Despite the intervention strategies that have been put in place to fight poverty, Namibia continues to experience prevalence of poverty with large numbers of households still living in poverty conditions and unable to afford the minimum daily essentials for a decent life. In this quantitative cross-sectional study design, the impact of sociodemographic characteristics of households on their poverty levels was statistically analysed using an ordered probit regression on data from the 2015/16 Namibia household income and expenditure survey. Results showed that sociodemographic characteristics such as the types of household dwelling unit, highest education attainment of the head of household, household main language, household tenure and household main source of income had a significant impact on the household's poverty levels. Households living in a mobile home dwelling unit, whose heads had secondary education as their highest educational attainment as well as households that were mortgaged and whose main source of income were from other sources were less likely to be severely household poor and more likely to be household poor. Furthermore, households living in a single-quarters dwelling unit and whose main language were Setswana were more likely to be severely household poor and less likely to be household poor. It is therefore recommended that the Namibian government and policy makers put more efforts in improving the sociodemographic characteristics of households, particularly those living in a single quarter dwelling unit and whose main language were Setswana.

1. Introduction

Household poverty is a state in which a household lacks the adequate financial, physical and social resources necessary for a (minimum) standard of living acceptable within the society in which the household lives (Maslen et al., 2013). According to the 2016 World Bank Annual report, household poverty has been a widespread recurrent challenge in Africa and Namibia is not excluded from this challenge. Using the international poverty line of US$1 per person per day, this report showed that Sub-Saharan Africa had the highest ratio (close to 50%) among all world regions. Here, poverty line was defined as the level of income or expenditure required by an individual to purchase or satisfy a minimum basket of consumption goods and services for him/her not to be in poverty (Chaudhry, Malik & Hassan, 2009). In Namibia, the upper bound poverty line estimated at N$520.80 was defined as households/persons that are considered to be poor while the lower bound poverty line estimated at N$389.30 was defined as households/persons that are food-poor since their total consumption expenditures are insufficient to meet their daily survival requirement (Namibia Statistics Agency, 2018).

Over the years, several studies have been done on poverty in general and its contributing factors globally, with factors such as education, migration, source of income, employment status, household indebtedness and marital status identified as the significant ones (Bulatao & Anderson, 2004; Wan, 2010; Hartfree & Collard, 2014; Yang, 2014; Mupetesi et al., 2015; Devaraj, 2017; Biyase & Zwane, 2018; Omoniyi, 2018; Trading Economics, 2020). Although there are policies and intervention strategies put in place by some institutions to reduce household poverty in societies, many countries, including Namibia, continue to experience high prevalence of household poverty. In the 2015/16 Namibia household income and expenditure survey report by Namibia Statistics Agency (2018), each household was classified as poor or severely poor based on their costs of basic needs compared to the national poverty lines. Here, severe poverty was defined as a condition characterized by severe deprivation of basic human needs, including food, safe drinking water, sanitation facilities, health, shelter, education and information.

As an educated nation with vast natural resources and an approximated population size of 2.5 million people, 28% of households in Namibia were classified as poor in 2004 and this figure decreased to 17% in 2016, while 14% and 11% of the households were classified as severely poor in 2004 and 2016 respectively (Namibia Statistics Agency, 2021). Notwithstanding, the incidence of poverty in Namibia stands at 43.3% with an average intensity value of 44.0%. This loosely means that poor people in Namibia experience 44.0% of weighted deprivations such as education, health and living standards (Namibia Statistics Agency, 2021). For this reason, the aim of this study was to examine the sociodemographic factors contributing to household poverty levels in Namibia. Findings from this study may provide further assist in the development of policy recommendations that can guide relevant organizations and governmental ministries to examine ways of re-allocating resources for the reduction of household poverty in the country.

2. Methodology

The data used in this study were extracted from the 2015/16 Namibia household income and expenditure survey (the latest thus far in the country) obtained from the Namibia Statistics Agency. This survey data is freely available to the public on the agency's website at www.nsa.org.na. All households with incomplete, non-response or missing information were excluded from this study.

2.1 Data Analysis

Consider a set of centred predictor variables $\mathbf{X}: \mathrm{N} \times \mathrm{P}$ and a set of centred response variable $\mathbf{y}: \mathrm{N} \times 1$, regression analysis measures the effect of $\mathbf{X}$ on $\mathbf{y}$ via the linear equation model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}, \tag{1}$$

where $\mathbf{e}$: $\mathrm{N} \times 1$ is the error term and $\mathbf{b}$: $\mathrm{P} \times 1$ is the unknown (regression) coefficient vector estimated through the least squares method as

$$\hat{\mathbf{b}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}. \tag{2}$$

Here, equation (1) assumes that $\mathbf{y}$ is a continuous variable and follows a normal distribution with mean μ and constant variance σ^2 (Oyedele & Ntusi, 2021).

More often $\mathbf{y}$ is dichotomous or binary in nature and non-normally distributed. In such situation, the modelling of $\mathbf{y}$ can be done through the usage of generalized linear models such as the probit model (Oyedele & Lubbe, 2018). However, for a non-binary $\mathbf{y}$ with at least 2 ordered categories, an ordered probit model is more appropriate. Consider $\mathbf{y}$ with $\mathrm{M} > 1$ ordered categories. The ordered probit model for $\mathrm{Y_n}$, with $\mathrm{n} = 1,2,\cdots,\mathrm{N}$, can be obtained as

$$\mathrm{Y_n^*} = \mathbf{Xb} + \mathbf{e} \tag{3}$$

where

$$\mathrm{y_n} = \begin{cases} 1 & \mathrm{y_n^*} \leq \mu_1 \\ 2 & \mu_1 < \mathrm{y_n^*} \leq \mu_2 \\ 3 & \mu_2 < \mathrm{y_n^*} \leq \mu_3 \\ \vdots & \qquad \vdots \\ \mathrm{M} & \mu_{\mathrm{M}-1} < \mathrm{y_n^*} \end{cases}$$

(Della-Lucia et al., 2013).

The predictor variables in this study were the sociodemographic characteristics (age of head, types of dwelling unit, composition, size, tenure, highest education of head, land ownership, main source of income, main language, region, sex of head and location) of the households, while the response variable was the households' poverty levels. In addition, the household poverty levels were determined using the Namibia Statistics Agency (2018) poverty line estimated at N$293.10 (per month), with a lower and upper bound estimate of N$389.30 and N$520.80 respectively. Each household was classified into three categories, namely poor (if spending is N$389.30 - N$520.80), severely poor (if spending is < N$389.30) and not poor (if spending is > N$520.81). All data analysis aspects of this study were performed using the R programming language (version 4.1.2).

3. Results

As per the inclusive criteria of this study, a total of 22,026 households were considered. To identify the best fit model to use in identifying the impact of sociodemographic characteristics of households on their poverty levels, all the predictor variables were used in model I. Afterwards, all the significant (explanatory) variables from model I were later used as (explanatory) variables in model II, and then the resulting significant variables from model II were used as variables in model III. This continued until there were no more significant variables left to use. The best fit model was identified as model I because it had the lowest Akaike information criterion value of 4969.305

and highest log-likelihood value of -2415.652. The resulting output of model I is shown in Table 1.

From Table 1, with a significant probability value (p-value) at a 5% level of significance, sociodemographic characteristics such as the types of household dwelling unit, highest education attainment of the head of household, household main language, as well as household tenure and household main source of income at a 10% level of significance, can be concluded to have a significant impact on the household poverty levels.

Table 1: Output from the fitted ordered probit model

	Estimate (adjusted)	Standard error	P-value
Age of household head	-0.002	0.002	0.135
Types of household dwelling unit			
Semi-detached house/Town house	-0.021	0.116	0.856
Apartment	-0.529	0.367	0.149
Guest flat	-0.163	0.416	0.695
Part commercial/Industrial building	-0.855	0.399	0.032*
Mobile home (caravan/tent)	-1.172	0.593	0.048*
Single quarters	0.400	0.192	0.037*
Traditional dwelling	-0.092	0.071	0.196
Improvised housing unit	-0.081	0.070	0.249
Others	-0.395	0.353	0.262
Detached house (Ref)			
Household composition			
With head and spouse(s) only	0.023	0.115	0.839
With 1 child, no relatives/non-relatives	0.086	0.091	0.345
With 2+ children, no relatives/non-relatives	-0.043	0.083	0.602
With relatives, no non-relatives	-0.063	0.076	0.412
With domestic worker(s)	0.155	0.148	0.296
With non-relatives	-0.144	0.095	0.130
With head alone (Ref)			
Household size	0.002	0.009	0.852
Household tenure			
Owned with mortgage	-0.172	0.097	0.075**
Not stated (Ref)			
Highest education of household head			
Primary	-0.049	0.056	0.387
Secondary	-0.169	0.062	0.006*
Tertiary	-0.141	0.097	0.146
Not stated	-0.246	0.241	0.306
No formal education (Ref)			
Household land ownership			
No	0.017	0.046	0.715
Yes (Ref)			
Household main source of income			
Subsistence farming	-0.022	0.072	0.761
Commercial farming	-0.110	0.276	0.691
Business activities, non-farming	0.105	0.073	0.151
Employment and/or annuity funds pensions	0.061	0.186	0.743
Cash remittances (exclude alimony/child support)	-0.089	0.088	0.311
Rental money	0.203	0.289	0.483
Interest from savings/investments	-0.161	0.540	0.765
State old age pension	-0.008	0.077	0.916
War veterans/ex-combatants grant	-0.054	0.238	0.820
Disability grants for adults (over 16 years)	0.086	0.151	0.570
State child maintenance grants	0.173	0.177	0.327
State foster care grants	0.007	0.353	0.984
State special maintenance grants (disability under 16 years)	-0.158	0.645	0.806
Alimony and similar allowances	0.019	0.420	0.965
Drought relief assistance	-0.032	0.153	0.834
In-kind receipts	-0.007	0.146	0.960
Other sources	-0.291	0.162	0.072**
Salaries & wages (Ref)			
Household main language			
Zambezi languages	-0.198	0.235	0.398
Otjiherero	0.114	0.183	0.534
Rukavango	-0.185	0.189	0.327
Nama/Damara	0.085	0.184	0.643

Oshiwambo	-0.008	0.183	0.964
Setswana	0.748	0.338	0.027*
Afrikaans	0.111	0.203	0.584
German	0.327	0.386	0.397
English	0.112	0.322	0.729
Other European	-0.779	0.551	0.157
Other African	0.090	0.305	0.766
Others	-0.021	0.252	0.934
Khoisan (Ref)			
Region			
Erongo	0.116	0.130	0.371
Hardap	-0.074	0.131	0.573
Kavango East	0.082	0.164	0.618
Kavango West	0.130	0.167	0.437
Khomas	-0.099	0.128	0.436
Kunene	0.161	0.144	0.264
Ohangwena	-0.117	0.136	0.390
Omaheke	0.107	0.135	0.428
Omusati	-0.075	0.137	0.586
Oshana	-0.032	0.135	0.811
Oshikoto	-0.057	0.134	0.669
Otjozondjupa	0.193	0.130	0.138
Zambezi	0.195	0.203	0.336
!Karas (Ref)			
Sex of household head			
Male	-0.028	0.042	0.506
Female (Ref)			
Household location			
Rural	0.007	0.063	0.910
Urban (Ref)			
Household poor | Not Household poor	-2.164	0.237	<0.001*
Not Household poor | Severely Household poor	1.270	0.235	<0.001*

(Ref) = Reference category,
*Significant at a 5% level of significance,
**Significant at a 10% level of significance

Furthermore, at a 5% level of significance and keeping all other variables constant, households living in a part commercial/industrial building (p-value=0.032) and in a mobile home (p-value=0.048) dwelling units were significantly and negatively associated with household poverty levels, suggesting that households who were living in these dwelling units were less likely to be severely household poor and more likely to be household poor as shown in Table 1. On the other hand, households living in a single quarter dwelling unit (p-value=0.037) and whose main language was Setswana (p-value=0.027) were significantly and positively associated with household poverty levels, suggesting that households who were living in this dwelling unit and spoke Setswana were more likely to be severely household poor and less likely to be household poor.

Moreover, households whose heads had secondary education (p-value=0.006) as their highest educational attainment were significantly and negatively associated with household poverty levels, suggesting that these households were less likely to be severely household poor and more likely to be household poor. Although at a 10% level of significance, mortgaged households (p-value=0.075) and those whose had other sources of main income (p-value=0.072) were significantly and negatively associated with household poverty levels, suggesting that these households were less likely to be severely household poor and more likely to be household poor as shown in Table 1.

4. Discussion

In this study, the ordered probit modelling technique was used to statistically examine the sociodemographic factors contributing to household poverty levels in Namibia using data obtained from the 2015/16 Namibia household income and expenditure survey.

Sociodemographic characteristics such as the types of household dwelling unit, highest education attainment of the head of household, household main language, household tenure and household main source of income had a significant impact on the household's poverty levels. These findings are similar to those found in Chaudhry et al. (2009), Wan (2010), Mupetesi et al. (2015) and Biyase & Zwane (2018). Mupetesi et al. (2015) concluded that the higher the crop production for households whose source of income were from other sources such as staple maize crop farming, the better improved their household poverty levels, while Wan (2010) concluded that as the number of years of education increases, the proportionate number of persons living below the poverty line decreases. In addition, Biyase & Zwane (2018) concluded that households living in urban type of dwelling units were less likely to be poverty stricken compared to those in the traditional/rural types.

Furthermore, households living in a part commercial/industrial building dwelling unit, living in a mobile home dwelling unit, whose heads had secondary education as their highest educational attainment as well as households that were mortgaged and whose main source of income were from other sources were less likely to be severely household poor and more likely to be household poor. Moreover, households living in a single quarter dwelling unit and whose main language were Setswana were more likely to be severely household poor and less likely to be household poor. This study findings are not startling, since most potential employers in Namibia require a higher or specific class of qualifications from their employees and new potential job candidates, while having higher education attainment can serve as an investment that improves the economic worth of individuals which in turns can lower the likelihood of such individuals living in severe poverty. Also, quite a lot of households in Namibia have at least six living-in members to cater for, which requires more cost on food & essential services on a daily basis. As a result of such financial burden to bear on a daily basis, household heads or breadwinners are driven to obtain loans and/or mortgage their homes for the upkeep of their households, in addition to their household income, thereby increasing their likelihood of living in poverty. Additionally, households whose main language were not English or any of the country's official language(s) tend to experience the highest incidence, depth and severity of poverty. This can be due to the fact that a lot of employers in the non-profit institutions, parastatals, government institutions and privately-owned enterprises require their employees and new applicants to be well conversant in internationally-friendly languages such as English, Afrikaans, German, Chinese and French.

5. Conclusions

With sociodemographic characteristics such as the types of household dwelling unit, highest education attainment of the head of household, household main language, household tenure and household main source of income having a significant impact on the household's poverty levels, it is therefore recommended that the Namibian government and policy makers put more efforts in improving the sociodemographic characteristics of households, particularly those living in a single quarters dwelling unit and whose main language were Setswana. Additionally, relevant organizations and governmental ministries in Namibia should continue to strengthen the national poverty eradication measures to achieve the national objectives as set out in the United Nations' sustainable development goals 1-6 and 8, and as per the national development plans. Further studies on this topic is recommended with a multidimensional household poverty definition using the next Namibia household income and expenditure survey tentatively planned for 2022/23 that would be incorporating a multidimensional poverty concept.

Conflicts of interest

The author of this paper has no competing interests to declare.

Funding

This research study did not receive any specific grant from funding agencies in the public, commercial, or non-profit organization sectors.

References

Biyase, M. & Zwane, T. (2018). An Empirical Analysis of the Determinants of Poverty and Household Welfare in South Africa. The Journal of Developing Areas, 52(1), 116 -130.

Bulatao, R., & Anderson, N. (2004). Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda of the National Research Council Panel on Race, Ethnicity, and Health in Later Life. National Academies Press. Washington, D.C.

Chaudhry, I., Malik, S., & Hassan, A. (2009). The Impact of Socioeconomic and Demographic Variables on Poverty: A Village Study. The Lahore Journal of Economics, 14 (1), 39-68.

Della-Lucia, S.M., Minim, V.P.R., Silva, C.H.O., Minim, L.A., & Cipriano, P.D. (2013). Ordered Probit Regression Analysis of The Effect of Brand Name on Beer Acceptance by Consumers. Food Science and Technology, 33(3), 586-591.

Devaraj, T. (2017). Impact of Poverty on School Children in Malawi. International Journal of Business and Economics Research, 3(3), 328-332.

Hartfree, Y. & Collard, S. (2014). Poverty, Debt and Credit: An Expert-Led Review. Personal Finance Research Centre, University of Bristol.

Maslen, C., Raffle, A., Marriott, S. & Smith, N. (2013). Food Poverty Report: What Does the Evidence Tell Us? Bristol City Council for Bristol's Food Policy Council.

Mupetesi, T., Francis, J., & Gomo, R. (2015). Poverty Rates in a Rural District of Zimbabwe: A Case Study of the Guruve District. Journal of Social Sciences, 43(1), 25-37.

Namibia Statistics Agency (2018). Namibia Household Income and Expenditure Survey 2015/2016 Report. Ongwediva.

Namibia Statistics Agency (2021). Namibia Multidimensional Poverty Index (MPI) Report 2021. Windhoek.

Omoniyi, B. (2018). An Examination of the Causes of Poverty on Economic Growth in Nigeria. Africa's Public Service Delivery and Performance Review, 6(1).

Oyedele, O.F & Lubbe, S. (2018). Partial Least Squares Biplot for Generalized Linear Models. International Journal of Statistics: Advances in Theory and Applications, 2(1), 1-20.

Oyedele, O.F. & Ntusi, N.S. (2021). A Multinomial Modelling of the Factors that Impact Viral Load Levels in Adults on Anti-Retroviral Therapy in Namibia: A Case Study of Nankudu Health District. International Science and Technology Journal of Namibia, 14, 73-82.

Trading Economics (2020). Namibia Youth Unemployment Rate 2012-2018. Retrieved from https://tradingeconomics.com/namibia/youth-unemployment-rate. (Accessed on July 10, 2021).

Wan, T. (2010). Effects of Social and Demographic Factors on Poverty Among Adult Males In The USA. Department of Sociology and International Population Program, Cornell University, New York. Retrieved from https://ejournals.epublishing.ekt.gr/index.php/ekke/article/viewFile/6935/6655.pdf. (Accessed on July 10, 2021).

World Bank (2016). World Bank Annual Report 2016 (English). Washington, D.C.: World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/763601475489253430/World-Bank-annual-report-2016. (Accessed on July 10, 2021).

Yang, R. (2014). The Determinants of Poverty and Vulnerability in Small-Scale Fisheries Communities in Vietnam. MA. Thesis, University of Ottawa.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

An application of survival analysis on the determinants of employment longevity in Namibia: Evidence from 2018 Labour Force Survey

Opeoluwa Oyedele[1,*], Liina E.M. Angula [1], Klemens T. Mutorwa[1]

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: ooyedele@unam.na

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Employment longevity,
Cox Proportional Hazard,
Kaplan Meier,
Namibia

ABSTRACT

Employment longevity shapes the total work experience of an employee's career, in addition to providing additional information about the employment stability of the employee and employer. However, in recent years, employment stability has decreased globally, with Namibia being amongst the highly affected countries where workers are having a high number of short-term employment records over the course of their careers. In this paper, the Kaplan-Meier and Cox Proportional Hazard techniques were used to estimate the survival of employment longevity for employed adults in Namibia using the 2018 Namibia Labour Force Survey. Results showed that majority of the employees were working in the private enterprises and government institutions, attained junior and senior secondary education, never married, worked for a paying job and were from the Khomas, Erongo and Otjozondjupa regions. Likewise, majority of the employees employed for less than 1 year and for 1-2 years were aged 20-29 years, while majority employed for 3-5 years and 6-10 years were aged 30-39 years. The employed adults' characteristics such as age group, type of employer, highest education attained, marital status, region, current schooling status and sex had a significant association with their survival of employment longevity. In addition, employees aged 30-39 and 40-49 years, employed in non-profit institutions, parastatals and government institutions, and from the Oshikoto, Omaheke, Oshana, Khomas, Erongo and Otjozondjupa regions had a high survival of employment longevity, while employees employed in privately owned informal enterprises and had already attained a technical or vocational certificates/diplomas, junior and senior secondary education had a low survival. It is therefore recommended that all relevant organizations and governmental ministries that deals with employment and labour matters should frequently engage all employers through their respective human resources departments, to further assist in the creation and implementation of favourable employment contracts that best suits their respective employees, especially, for those employed in privately owned informal enterprises, employed for less than 1 year, 1-2 years and 3-5 years, who are in their 20s and 30s (age-wise), and already attained a technical or vocational certificates/diplomas, junior and senior secondary education.

1. Introduction

Employment longevity can be defined as the number of years/months/days a person is employed in the service they are currently employed in. It shapes the total work experience of an employee and can provide (more) information about the work stability of both the employees and employers (Ignaczak & Voia, 2011; Noon, Blyton & Morrell, 2013). Most often, employers use limited duration contracts that lasts for a specified period, to create certainty and limit legal risk in respect of staffing solutions, as these contracts will expire on a certain date or upon termination by the employer with valid reasons, which in turns allows an employer to plan for the employee's exit in advance (McKenzie, 2020). These contracts are usually regulated by the labour laws of each country, to ensure that employers still fulfill basic labour rights regardless of a contract's form, and most often they have a minimum duration of one or three months and a maximum term of two to three years, depending on the employer's policies.

An employee can be kept on successive limited duration contracts or can become a permanent employee, as long as it does not violate the employer's policies as well as the labour laws of the country. However, with the increased incorporation of Information and Communications Technologies (ICT) as well as Artificial Intelligence (AI) innovations at workplaces by employers, concerns have been raised with regards to their effects on human employment and comparative demand for skills at workplaces, as well as the segmentation of the labour market as a result of the changes in ICT and AI innovations (Castro-Silva & Lima, 2017). These concerns are fathomable as these innovations (most often) displaces low-skilled, entry-level and low-level workers, which increases the chances of being unemployed or offered very short-term (non-renewable) employment contracts. Moreover, the uncertainty of a country's economy and unforeseen corporate downsizings have also impacted in the retention of employees and the loss of critical employees (Sinha & Sinha, 2012) and Namibia is not immune to this debacle. For this reason, the aim of this study was to examine the factors associated with employment longevity in Namibia for (employed) adults, in addition to estimating their survival of employment longevity.

2. Methodology

The data used in this paper were extracted from the Namibia Labour Force Survey (NLFS), administered by the Namibia Statistics Agency in 2018. The NLFS is conducted yearly from 2012 to 2018, to provide labour force information on the employment, socio-demographic and educational characteristics of all persons living in households in Namibia. In this study, the inclusion criteria were all employed adults aged 18 years and above living in households during the reference period of the 2018 survey. All NLFS reports and datasets are freely available online at www.nsa.org.na, and for more information about the 2018 survey, refer to the NLFS report of 2018.

2.1 Statistical Analysis

Survival analysis is a collection of statistical methods that can be used for analyzing data whose outcome variable is measured as the time until an event occurs (Kleinbaum & Klein, 2015). It analyses the rates of occurrence of events over time, without assuming the rates are constant, through the usage of a survival function, hazard function and cumulative hazard function. A survival function is defined as the probability that an individual will survives longer than time t, that is,

$$S(t) = P(T > t)$$

with T being the survival time. The hazard function, on the other hand, is defined as the probability of failure during a very small-time interval, assuming that the individual has survived to the beginning of the interval, while the cumulative hazard function is defined as the total number of failures or deaths over an interval of time and it is obtained as

$$H(t) = \int_0^t h(u)du$$

where $h(u)$ is the hazard risk and u is the accumulated risk (Petrus & Oyedele, 2021). To estimate the survival function, the non-parametric Kaplan Meier (KM) method can be used. This estimation method is the easiest way to determine survival over time in spite of all the problems associated with either subjects or situations, and uses curves to determine events, censoring and the survival probability (Etikan, Abubakar, & Alkassim, 2017). The survival probability is obtained as

$$\widehat{S(t)} = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where t_i, for $i = 1, 2, \cdots, n$, represents the time at which failures occurs, n_i represents the number of individuals at risk at time t_i, and d_i represents the number that failed at time t_i (Harrell, 2015). To estimate the hazard function, the Cox Proportional Hazard (CPH) approach can be used. This approach is a semi-parametric regression model that simultaneously assesses the effects of several (risk) factors on survival time of an event of interest (Petrus & Oyedele, 2021) and is given as

$$h(t, X) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i}$$

where $X = (X_1, X_2, \cdots, X_p)$ is the set of predictors, $h_0(t)$ is the baseline hazard function, X_i, for $i = 1,2, \ldots, p$, is the i^{th} predictor, β_i is the coefficient for the i^{th} predictor and e^{β_i} is the hazard ratio that measures the effect of the i^{th} predictor on the survival time.

The response variable in this paper was the employment longevity (in years) of the employed adults, measured using their respective length of employment, while the predictors were their age group (in years), type of employer, highest education attained, marital status, working for payment job, region, (current) schooling status and sex. All the data analyses of this paper were performed using the R software.

3. Results

Out of the 6,658 employed adults aged 18 years and above considered in this study, majority (2,139) were employed for less than 1 year, followed by those employed for 3-5 years (1,323), 1-2 years (1,164) and for at least 11 years (1,110), while only 922 (13.85%) were employed for 6-10 years as shown in Table 1. Majority of the employees employed for less than 1 year and for 1-2 years were aged 20-29 years, followed by those aged 30-39 years, while majority of those employed for 3-5 years were aged 30-39 years, followed by those aged 20-29 years. On the other hand, majority of the employees employed for 6-10 years were aged 30-39 years, followed by those aged 40-49 years, while majority of those employed for at least 11 years were aged 40-49 and 50-59 years. Furthermore, majority of the 6658 employees were working in the private enterprises (32.91%) and government institutions (20.94%), attained junior secondary education (32.14%) and senior secondary education (22.84%), and were never married (57.40%). In addition, majority were from the Khomas (15.76%), Erongo (9.91%) and Otjozondjupa (9.12%) regions, worked for a paying job (94.56%), and were males (54.19%) as shown in Table 1.

Table 1: Distribution of the employed adults' characteristics and their length of employment

Characteristic	<1 year	1-2 years	3-5 years	6-10 years	≥11 years		P-value
	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)	Total (%)	
Age Group							
<20	128 (1.92)	24 (0.36)	13 (0.20)	1 (0.02)	0 (0.00)	166 (2.49)	<2e-16*
20-29	996 (14.96)	484 (7.27)	411 (6.17)	126 (1.89)	18 (0.27)	2035 (30.56)	
30-39	594 (8.92)	391 (5.87)	519 (7.80)	384 (5.77)	201 (3.02)	2089 (31.38)	
40-49	283 (4.25)	175 (2.63)	259 (3.89)	287 (4.31)	429 (6.44)	1433 (21.52)	
50-59	111 (1.67)	75 (1.13)	95 (1.43)	106 (1.59)	422 (6.34)	809 (12.15)	
60-69	21 (0.32)	12 (0.18)	22 (0.33)	17 (0.26)	33 (0.50)	105 (1.58)	
≥70	6 (0.09)	3 (0.05)	4 (0.06)	1 (0.02)	7 (0.11)	21 (0.32)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Employer Type							
Unspecified	20 (0.30)	7 (0.11)	7 (0.11)	3 (0.05)	5 (0.08)	42 (0.63)	<2e-16*
Cooperative	35 (0.53)	18 (0.27)	25 (0.38)	12 (0.18)	13 (0.20)	103 (1.55)	
Government	77 (1.16)	103 (1.55)	321 (4.82)	325 (4.88)	568 (8.53)	1394 (20.94)	
Non-profit institution	10 (0.15)	7 (0.11)	8 (0.12)	9 (0.14)	12 (0.18)	46 (0.69)	
Parastatal	71 (1.07)	60 (0.90)	90 (1.35)	85 (1.28)	112 (1.68)	418 (6.28)	
Private enterprise	644 (9.67)	497 (7.46)	522 (7.84)	300 (4.51)	228 (3.42)	2191 (32.91)	
Private enterprise (informal)	288 (4.33)	98 (1.47)	58 (0.87)	23 (0.35)	24 (0.36)	491 (7.37)	
Private household (commercial farm)	201 (3.02)	106 (1.59)	91 (1.37)	53 (0.80)	51 (0.77)	502 (7.54)	
Private household (non-farm)	396 (5.95)	128 (1.92)	114 (1.71)	62 (0.93)	43 (0.65)	743 (11.16)	
Private household (subsistence farm)	397 (5.96)	140 (2.10)	87 (1.31)	50 (0.75)	54 (0.81)	728 (10.93)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Highest Education attained							
Postgraduate Certificate/Diploma/ Degree	14 (0.21)	21 (0.32)	40 (0.60)	60 (0.90)	91 (1.37)	226 (3.39)	<2e-16*
Junior secondary	845 (12.69)	414 (6.22)	414 (6.22)	226 (3.39)	241 (3.62)	2140 (32.14)	

None	302 (4.54)	136 (2.04)	116 (1.74)	78 (1.17)	101 (1.52)	733 (11.01)	
Primary	481 (7.22)	185 (2.78)	172 (2.58)	115 (1.73)	148 (2.22)	1101 (16.54)	
Senior secondary	364 (5.47)	265 (3.98)	346 (5.20)	269 (4.04)	277 (4.16)	1521 (22.84)	
Technical or Vocational Certificate/Diploma	25 (0.38)	32 (0.48)	42 (0.63)	28 (0.42)	18 (0.27)	145 (2.18)	
Undergraduate Certificate/ Diploma/Degree	108 (1.62)	111 (1.67)	193 (2.90)	146 (2.19)	234 (3.51)	792 (11.90)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Marital Status							
Separated	16 (0.24)	6 (0.09)	10 (0.15)	3 (0.05)	8 (0.12)	43 (0.65)	
Consensual union	271 (4.07)	139 (2.09)	156 (2.34)	95 (1.43)	94 (1.41)	755 (11.34)	<2e-16*
Divorced	6 (0.09)	7 (0.11)	17 (0.26)	11 (0.17)	27 (0.41)	68 (1.02)	
Married traditionally/customary	139 (2.09)	72 (1.08)	81 (1.22)	53 (0.80)	59 (0.89)	404 (6.07)	
Married with certificate	181 (2.72)	166 (2.49)	282 (4.24)	280 (4.21)	555 (8.34)	1464 (21.99)	
Never married	1503 (22.57)	759 (11.40)	767 (11.52)	464 (6.97)	329 (4.94)	3822 (57.40)	
Unspecified	4 (0.06)	3 (0.05)	0 (0.00)	1 (0.02)	0 (0.00)	8 (0.12)	
Widowed	19 (0.29)	12 (0.18)	10 (0.15)	15 (0.23)	38 (0.57)	94 (1.41)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Payment Working							
No	105 (1.58)	72 (1.08)	75 (1.13)	43 (0.65)	67 (1.01)	362 (5.44)	
Yes	2034 (30.55)	1092 (16.40)	1248 (18.74)	879 (13.20)	1043 (15.67)	6296 (94.56)	0.500
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Region							
Zambezi	107 (1.61)	45 (0.68)	56 (0.84)	43 (0.65)	39 (0.59)	290 (4.36)	
Erongo	190 (2.85)	121 (1.82)	157 (2.36)	98 (1.47)	94 (1.41)	660 (9.91)	
Hardap	151 (2.27)	94 (1.41)	84 (1.26)	63 (0.95)	81 (1.22)	473 (7.10)	
!Karas	154 (2.31)	105 (1.58)	101 (1.52)	78 (1.17)	86 (1.29)	524 (7.87)	
Kavango East	178 (2.67)	78 (1.17)	86 (1.29)	61 (0.92)	84 (1.26)	487 (7.31)	
Kavango West	118 (1.77)	39 (0.59)	46 (0.69)	19 (0.29)	26 (0.39)	248 (3.72)	
Khomas	226 (3.39)	196 (2.94)	261 (3.92)	190 (2.85)	176 (2.64)	1049 (15.76)	
Kunene	119 (1.79)	46 (0.69)	57 (0.86)	48 (0.72)	37 (0.56)	307 (4.61)	5e-11*
Ohangwena	131 (1.97)	51 (0.77)	56 (0.84)	33 (0.50)	53 (0.80)	324 (4.87)	
Omaheke	146 (2.19)	76 (1.14)	84 (1.26)	40 (0.60)	63 (0.95)	409 (6.14)	
Omusati	148 (2.22)	75 (1.13)	71 (1.07)	36 (0.54)	74 (1.11)	404 (6.07)	
Oshana	136 (2.04)	72 (1.08)	85 (1.28)	67 (1.01)	105 (1.58)	465 (6.98)	
Oshikoto	149 (2.24)	69 (1.04)	77 (1.16)	58 (0.87)	58 (0.87)	411 (6.17)	
Otjozondjupa	186 (2.79)	97 (1.46)	102 (1.53)	88 (1.32)	134 (2.01)	607 (9.12)	

Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Current Schooling Status							
Attending school	108 (1.62)	64 (0.96)	98 (1.47)	69 (1.04)	68 (1.02)	407 (6.11)	3e-08*
Left school	1754 (26.34)	976 (14.66)	1119 (16.81)	784 (11.78)	961 (14.43)	5594 (84.02)	
Never attended	277 (4.16)	124 (1.86)	106 (1.59)	69 (1.04)	81 (1.22)	657 (9.87)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	
Sex							
Female	938 (14.09)	526 (7.90)	620 (9.31)	417 (6.26)	549 (8.25)	3050 (45.81)	0.004*
Male	1201 (18.04)	638 (9.58)	703 (10.56)	505 (7.58)	561 (8.43)	3608 (54.19)	
Total	2139 (32.13)	1164 (17.48)	1323 (19.87)	922 (13.85)	1110 (16.67)	6658 (100)	

* Significant at a 5% level of significance

Looking at the KM curves, it can be observed that the cumulative survival probability was higher for (adult) employees aged 50-59 years and lower for the 20-29 and less than 20 years age groups, while the probability was the same for both employees who worked for payment and those who did not, as shown in Figure 1. Similarly, the probability was higher for employees in the Khomas, Oshana, Erongo, !Karas and Otjozondjupa regions, but lower for the remaining regions, while it was higher for the female employees and lower for the males as shown in Figure 2.

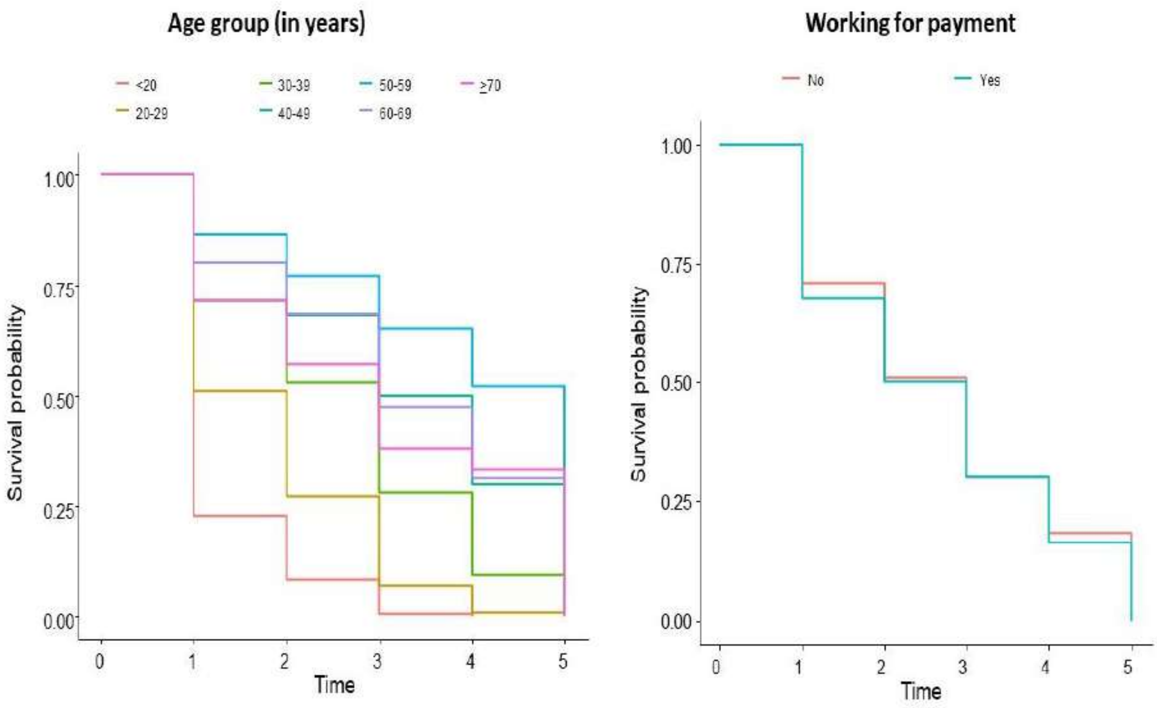


Figure 1: KM curves for employed adults' age group and working for payment.

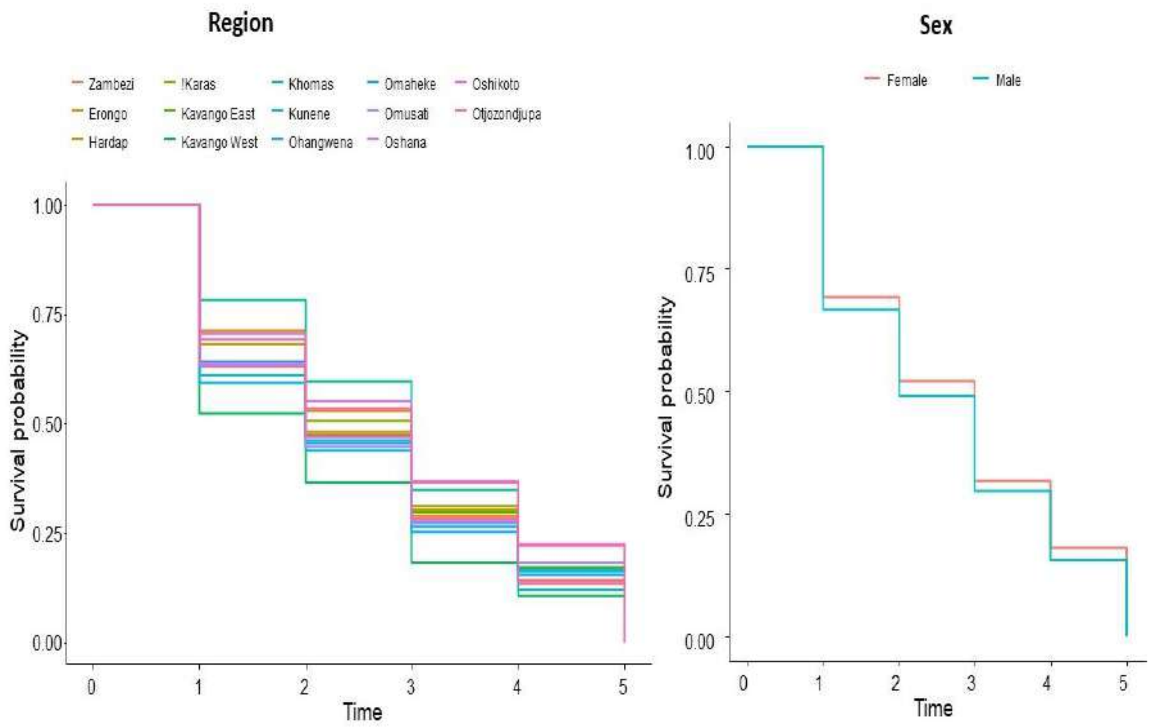


Figure 2: KM curves for employed adults' region and sex.

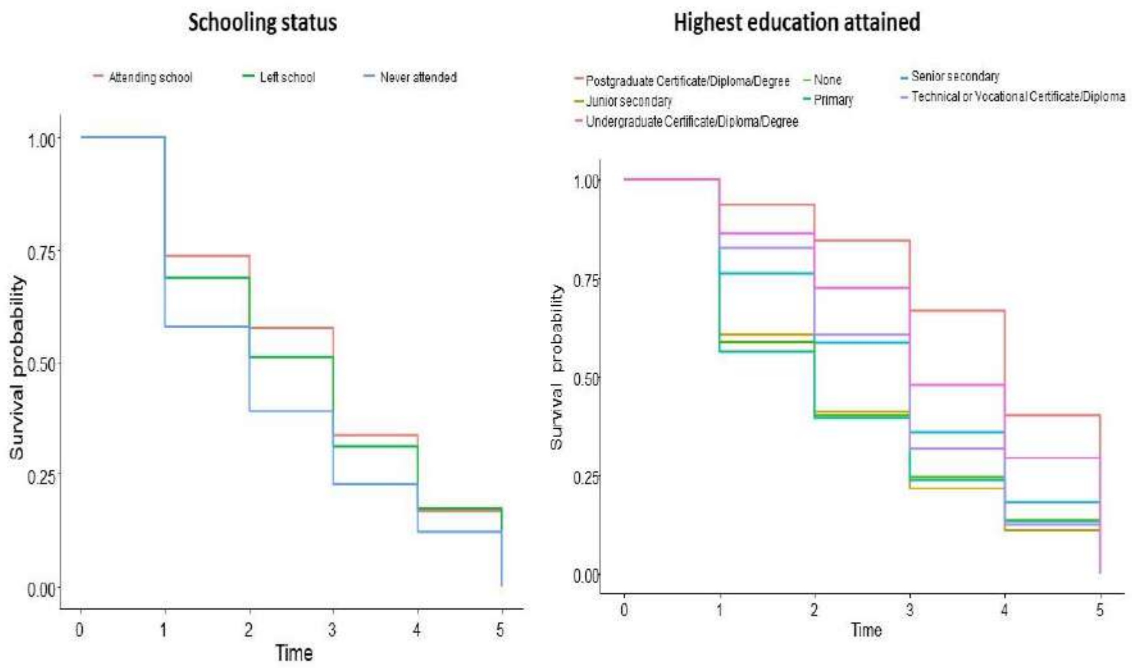


Figure 3: KM curves for employed adults' schooling status and highest education attainment.

From Figure 3, it can be observed that the cumulative survival probability was higher for employees who were currently attending school and those that left school, but lower for those who never attended school, while the probability was higher for postgraduate certificate/diploma/degree holders and lower for junior secondary and primary school certificate holders as well as uneducated employees. Likewise, the probability was higher for employees employed at government institutions, but lower for those employed at privately owned informal enterprises, non-farm and subsistence farm households, while it was higher for employees who were divorced, married with certificate and widowed, but lower for the employees with unspecified marital status as shown in Figure 4. Thus, it can be concluded that the employed adults' age group, region, sex, current schooling status, highest education attainment, marital status and type of employer were associated with their survival of employment longevity, while working for payment or not working for payment was not associated.

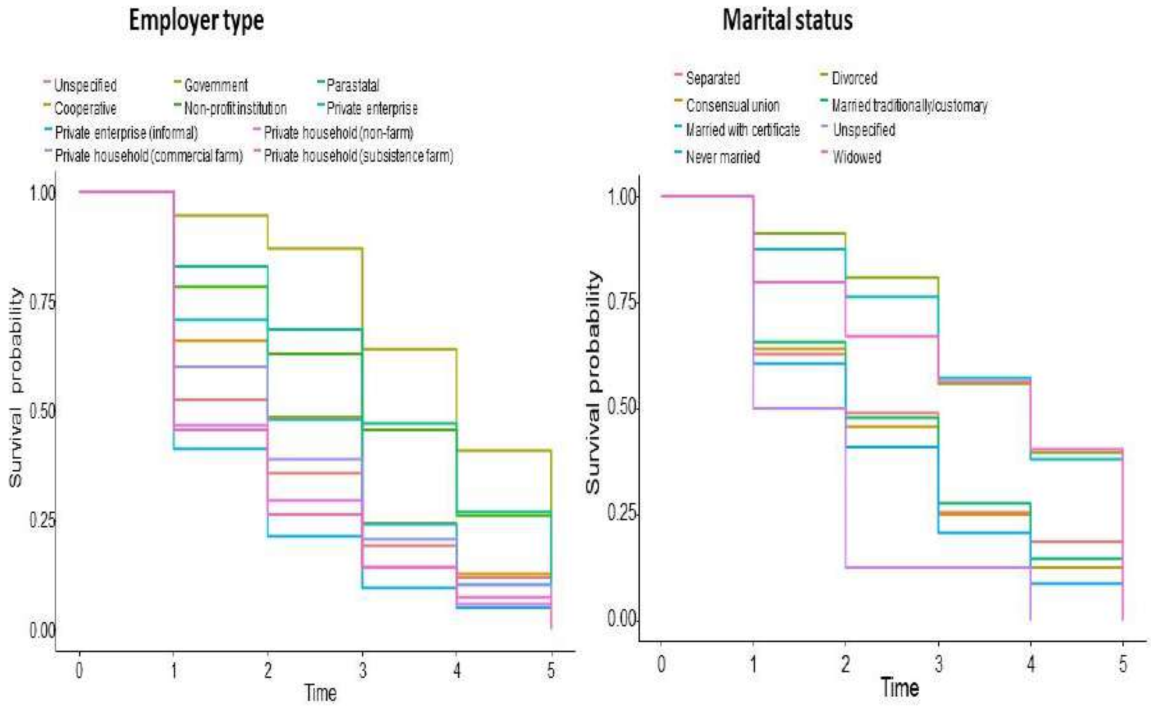


Figure 4: KM curves for employed adults' employer type and marital status.

Moreover, testing for the equality of survival distribution using the log rank (Mantel-Cox) test and with a significant p-value at a 5% level of significance, the employed adults' age group (p-value<2e-16), type of employer (p-value<2e-16), highest education attained (p-value<2e-16), marital status (p-value<2e-16), region (p-value=5e-11), current schooling status (p-value=3e-08) and sex (p-value=0.004) can be concluded to have a significant association with their survival of employment longevity, while the working for payment variable (p-value=0.500) does not have a significant association, as shown in Table 1. Hence, all the associated employed adults' characteristics were considered in the fitted CPH regression model and the subsequent results shown in Table 2.

Table 2: Output from the fitted CPH regression model

Characteristic	Adjusted estimate	Hazard ratio	Standard error	P-value	95% Confidence Interval for hazard ratio	
					Lower	Upper
Age Group						
<20 (Ref)						
20-29	-0.418	0.659	0.083	<0.001*	0.560	0.774
30-39	-0.988	0.372	0.084	<2e-16*	0.316	0.439

40-49	-1.432	0.239	0.087	<2e-16*	0.202	0.283
50-59	-1.673	0.188	0.092	<2e-16*	0.157	0.225
60-69	-1.614	0.199	0.130	<2e-16*	0.154	0.257
≥70	-1.614	0.199	0.235	<0.001*	0.126	0.316
Employer Type						
Unspecified (Ref)						
Cooperative	-0.135	0.874	0.184	0.465	0.609	1.255
Government	-0.729	0.482	0.159	<0.001*	0.353	0.658
Non-profit institution	-0.388	0.679	0.215	0.049*	0.445	1.034
Parastatal	-0.445	0.641	0.163	0.006*	0.466	0.883
Private enterprise	-0.161	0.851	0.157	0.304	0.626	1.157
Private enterprise (informal)	0.276	1.318	0.162	0.040*	0.960	1.810
Private household (commercial farm)	-0.077	0.926	0.162	0.635	0.674	1.272
Private household (non-farm)	0.221	1.247	0.160	0.168	0.911	1.706
Private household (subsistence farm)	-0.058	0.943	0.161	0.717	0.688	1.293
Highest Education attained						
Postgraduate Certificate/Diploma/Degree (Ref)						
Junior secondary	0.187	1.206	0.074	0.012*	1.042	1.395
None	0.112	1.119	0.137	0.412	0.855	1.464
Primary	0.192	1.212	0.079	0.015*	1.039	1.414
Senior secondary	0.088	1.092	0.073	0.230	0.946	1.260
Technical or Vocational Certificate/Diploma	0.215	1.240	0.108	0.046*	1.004	1.533
Undergraduate Certificate/Diploma/Degree	0.067	1.069	0.076	0.377	0.922	1.241
Marital Status						
Separated (Ref)						
Consensual union	-0.155	0.857	0.158	0.329	0.628	1.169
Divorced	-0.193	0.825	0.197	0.328	0.561	1.213
Married traditionally/customary	0.005	1.005	0.162	0.977	0.731	1.381
Married with certificate	-0.246	0.782	0.157	0.118	0.575	1.064
Never married	-0.061	0.940	0.156	0.694	0.693	1.277
Unspecified	0.633	1.883	0.387	0.101	0.883	4.017
Widowed	-0.073	0.929	0.187	0.694	0.645	1.339
Region						
Zambezi (Ref)						
Erongo	-0.225	0.798	0.075	0.003*	0.689	0.925
Hardap	-0.072	0.931	0.079	0.360	0.798	1.086
!Karas	-0.087	0.917	0.077	0.262	0.788	1.067
Kavango East	0.061	1.063	0.076	0.421	0.916	1.234
Kavango West	-0.054	0.948	0.090	0.550	0.795	1.130
Khomas	-0.224	0.799	0.071	0.001*	0.696	0.918
Kunene	0.060	1.062	0.085	0.483	0.899	1.254
Ohangwena	-0.012	0.988	0.084	0.890	0.838	1.165
Omaheke	-0.153	0.858	0.081	0.047*	0.733	1.005

Omusati	-0.019	0.981	0.081	0.813	0.838	1.149
Oshana	-0.196	0.822	0.078	0.012*	0.705	0.959
Oshikoto	-0.134	0.874	0.081	0.049*	0.747	1.024
Otjozondjupa	-0.231	0.794	0.075	0.002*	0.685	0.920
Current Schooling Status						
Attending school (Ref)						
Left school	-0.079	0.924	0.053	0.141	0.832	1.027
Never attended	-0.060	0.942	0.127	0.638	0.735	1.207
Sex						
Female (Ref)						
Male	0.009	1.009	0.026	0.745	0.958	1.062

* Significant at a 5% level of significance (Ref) = Reference category

From Table 2, with significant p-values<2e-16 and <0.001, (adult) employees who were aged 20-29 years (HR=0.659, p-value<0.001) had a higher survival of employment longevity, compared to the survival for employees who were less than 20 years old, while employees aged 30-39 (HR=0.372, p-value<2e-16) and 40-49 years (HR=0.239, p-value<2e-16) had a high survival of employment longevity. In addition, employees aged 50-59 (HR=0.188, p-value<2e-16), 60-69 (HR=0.199, p-value<2e-16) and at least 70 years old (HR=0.199, p-value<0.001) had a slightly fair survival of employment longevity as shown in Table 2. With regards to the type of employers, employees who were employed in non-profit institutions (HR=0.679, p-value=0.049) and parastatals (HR=0.641, p-value=0.006) had a higher survival of employment longevity, compared to the survival for those who had unspecified employers, while employees who were employed in government institutions (HR=0.482, p-value<0.001) had a high survival of employment longevity. However, employees who were employed in privately owned informal enterprises (HR=1.318, p-value=0.040) had a lower survival of employment longevity as shown in Table 2.

Furthermore, with significant p-values between 0.012 and 0.046, employees who had a technical or vocational certificates/diplomas (HR=1.240, p-value=0.046), senior secondary education (HR=1.212, p-value=0.015) and junior secondary education (HR=1.206, p-value=0.012) had a lower survival of employment longevity, compared to the survival for those who had postgraduate certificates/diplomas/degrees as shown in Table 2. Moreover, looking at the region characteristic, employees from the Oshikoto (HR=0.874, p-value=0.049), Omaheke (HR=0.858, p-value=0.047), Oshana (HR=0.822, p-value=0.012), Khomas (HR=0.799, p-value=0.001), Erongo (HR=0.798, p-value=0.003) and Otjozondjupa regions (HR=0.794, p-value=0.002) had a higher survival of employment longevity, compared to the survival for employees who were from the Zambezi region as shown in Table 2.

4. Discussion

From this study, majority of the employees were employed for less than 1 year, while majority of the employees employed for less than 1 year and for 1-2 years were aged 20-29 years. On the other hand, majority of the employees employed for 3-5 years and 6-10 years were aged 30-39 years. However, majority of the employees employed for at least 11 years were aged 40-49 and 50-59 years. Additionally, majority of the employees were working in the private enterprises and government institutions, attained junior and senior secondary education, never married, worked for a paying job and were from the Khomas, Erongo and Otjozondjupa regions.

In this study, it was revealed that the employed adults' age group, type of employer, highest education attained, marital status, region, current schooling status and sex had a significant association with their survival of employment longevity. This study key findings are similar to the conclusions made in Ignaczak & Voia (2011), Madden et al. (2014), Jendrossek et al. (2019), ten Berge et al. (2020) and Selwaness & Krafft (2020). Ignaczak & Voia (2011) and Madden et al. (2014) concluded that employee's sex was related to their employment longevity, with the longevity decreasing abruptly among the male employees, while for the female employees a miscellaneous pattern was observed (Ignaczak & Voia, 2011). Likewise, Selwaness & Krafft (2020) concluded that the employees' marital status and sex were related to their employment longevity, with individuals forestalling marriage and those getting married having a strong relation with women's employment longevity outcomes. Jendrossek et al. (2019) concluded that the employees' age as well as the type of their employers were related to their employment longevity, with older aged employees and those working in government institutions having fair to high employment longevity. Madden et al. (2014) and ten Berge et al. (2020) concluded that the attainment of higher education by the employee was associated with lower likelihood of job ending at the company.

Moreover, compared to the less than 20 years old age group, employees aged 30-39 and 40-49 years had a high survival of employment longevity, while employees aged 50-59, 60-69 and at least 70 years old had a slightly fair survival. This is not surprising as the older the employees get, the closer they are to the early or universal retirement age, which in turns lowers their chance of getting long-term employment contracts. In addition, some employers often give critical and scarce-skilled employees who have passed the universal retirement age of 65 years short-term (renewable) employment contracts of a duration between 6 months to 1 year. This is often done with the understanding that the critical and scarce-skilled employees train and pass on their skills to the low-level and junior employees in the company. This finding is in line with findings reported by Jendrossek et al. (2019) where it was concluded that the older the employees gets, the fairly to high their employment longevity in the company. Employees who were employed in non-profit institutions, parastatals and government institutions had a high survival of employment longevity. However, employees who were employed in privately owned informal enterprises had a low survival of employment longevity. This is not startling as most privately owned informal enterprises in Namibia are often micro, small and medium-sized enterprises that are non-subsidiary, independent and employs fewer than 10 employees, with a low to medium turnover and capital basis. Additionally, majority of these businesses experience difficulties during the first 12 to 24 months of their existence due to factors such as lack of proper planning, poor financial management, lack of management skills, inability to manage growth, lack of financial support, and lack of capital and access to finance. On the other hand, enterprises such as non-profit institutions, parastatals and government institutions most often do not lack financial support and access to institutional credit and financing, which micro, small and medium-sized enterprises most frequently lack resulting in their high financing costs and eventual failure. This finding somewhat echoes the observations made by Jendrossek et al. (2019) that the type of employer an employee has was associated with the employees' longevity at the company, especially, for those working in government institutions having higher longevity.

Furthermore, employees from the Oshikoto, Omaheke, Oshana, Khomas, Erongo and Otjozondjupa regions had a high survival of employment longevity, compared to the survival for employees who were from the Zambezi region. This may be due to the high number of employment opportunities within these developed and industrialized regions. Compared to the survival for those who had postgraduate certificates/diplomas/degrees, employees who had a technical or vocational certificates/diplomas, as well as junior and senior secondary education had a low survival of employment longevity. This is not surprising as most employers require a (higher) degree or specific classes of qualification from their employees before they can get promoted or be considered for a promotion at work, and an employee having a higher education attainment for his/her job makes him/her a more attractive candidate for work promotion opportunities within the company. Also, having higher education attainment can serves as an investment that improves the economic worth of individuals which in turns can increase a country's overall productivity and economic competitiveness, since an economy's productivity increases as the number of educated workers increases due to the prospect that skilled workers can perform tasks more efficiently. This finding is in line with findings reported by ten Berge et al. (2020) and Madden et al. (2014) where it was concluded that the attainment of higher education by the employee was associated with lower likelihood of job ending/loss.

5. Conclusion

It is recommended that all relevant organizations and governmental ministries that deals with employment and labour related matters should frequently engage all employers through their respective human resources departments, to further assist in the creation and implementation of favourable employment contracts that best suits their respective employees, especially, for those employed in privately owned informal enterprises and those who have already attained a technical or vocational certificates/diplomas, as well as junior and senior secondary education. This in turns will minimize the dislodgment of low-skilled, entry-level and low-level employees within the company, as well as minimize the loss of critical and scarce-skilled employees in the company. Additionally, such engagements may result in the reduction in the chances of being unemployed, especially, for employees employed for less than 1 year, 1-2 years and 3-5 years who are in their 20s and 30s (age-wise).

References

Castro-Silva, H., Lima, F. (2017). Technology, employment and skills: A look into job duration. Research Policy Elsevier, 46(8), 1519-1530.
Etikan, I., Abubakar, S., Alkassim, R. (2017). The kaplan meier estimate in survival analysis. Biometrics & Biostatistics International Journal, 5(2), 55-59.

Harrell, F.E. (2015). Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis, 2nd Ed. Springer Series in Statistics. Switzerland.

Ignaczak, L., Voia, M. (2011). A retrospective analysis of employment duration: Evidence from the second half of the twentieth century. Labour, 25 (1), 97-125.

Jendrossek, M., Edmunds, W.J., Rohan, H., Clifford, S., Mooney, T.A., Eggo, R.M. (2019). Health care worker vaccination against Ebola: Vaccine acceptance and employment duration in Sierra Leone. Vaccine, 37(8), 1101-1108.

Kleinbaum, G.D., Klein, M. (2015). Survival analysis - A self learning text. Springer.

Madden, E.E., Scannapieco, M., Painter, K. (2014). An examinantion of retention and length of employment among public child welfare workers. Children and Youth Services Review, 41, 37-44.

McKenzie, B. (2020). South Africa: limited duration employment contracts - Unlimited hassles when not properly managed.
https://www.lexology.com/library/detail.aspx?g=b97407e0-1104-4a9b-9012-c3cd45c64d6e
Accessed date: 31 January 2021.

Noon, M., Blyton, P., Morrell, K. (2013). The realities of work: Experiencing work and employment in contemporary society, 4th Ed. Red Globe Press. London, England.

Petrus, A., Oyedele, O. (2021). An application of survival analysis on the prevalence and risk factors of breast cancer in Namibia. International Science and Technology Journal of Namibia,14:25-33.

Selwaness, I., Krafft, C. (2020). The dynamics of family formation and women's work: What facilitates and hinders female employment in the Middle East and North Africa? Population Research and Policy Review, 1-55.

Sinha, C., Sinha, R. (2012). Factors affecting employee retention: A comparitive analysis of two organizations from heavy engineering industry. European Journal of Business and Management, 4(3), 145-163.

ten Berge, J., Lippenyi, Z., van der Lippe, T., Goos, M. (2020). Technology implementation within enterprises and job ending among employees. A study of the role of educational attainment, organizational tenure, age and unionization. Research in Social Stratification and Mobility, 69, 100548.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Customer relationship management (CRM) and passenger loyalty in delivering high quality service at Air Namibia: A structural equations approach

Albert J. Isaacs[1]*, S. B. Lwendo [1], Lawrence N. Kazembe[2]

[1]School of Education, University of Namibia, Windhoek, Namibia
[2]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: aisaacs@unam.na

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Customer relations management,
Passenger loyalty,
Structural equation modelling,
Namibia

ABSTRACT

Air Namibia, like any other airline, faces challenges as it operates in the global economy. Extreme scrutiny and debate about Air Namibia's viability has highlighted some of the airline's major issues of strategic, operational inefficiency and inability to create customer value. The study's aim was to establish the impact of passenger loyalty on customer relationship management (CRM) in delivering high quality service to passengers and value creation. Structural equation modelling (SEM) was used to analyze survey data collected from 181 international, regional and domestic passengers using Air Namibia for passengers travelling through the Hosea Kutako International and Eros Airports. The study further identified the primary factors affecting passenger satisfaction with Air Namibia's value chain activities, these factors included: customer orientation, operational specialties, domain expertise, and service recovery and information technology. While, factors contributing to passenger retention and loyalty towards Air Namibia's products and services included Marketing and Promotional Activities, Loyalty aspects, Value for Money and Comfort Issues. In exploring the relationship between customer relationship management (CRM) and passenger loyalty, the study found that interpersonal relationships between staff and the customers are crucial to CRM initiatives as they result in a better understanding of customer needs, which in turn leads to passenger loyalty.

1. Introduction

The airline industry has experienced extraordinary growth since World War II, with approximately 2,000 airlines operating from a global fleet of over 23,000 aircraft, carrying above 2.2 billion passengers annually, for business and leisure (IATA, 2012). The air transport industry also includes about 3750 airports, spread through a route network of several million kilometres currently being served, managed by highly sophisticated air navigation service providers (IATA, 2012).

The size and complexity of the air transport industry, makes it extremely volatile and highly competitive. Moreover, the industry faces numerous challenges and building customer relationships has become a fundamental instrument for improving passenger retention, satisfaction, and value-added services, essential for survival. Many airlines have recognized the importance of building relationships with their passengers, as they attempt to lower passenger acquisition costs, as well as increase profitability and customer satisfaction (Kim and Cha, 2002). While customer satisfaction may have a direct impact on passenger loyalty, having satisfied customers is not sufficient to keeping them loyal to the organization's products and services (Bowen and Chen, 2001).

Wilson (2005) argues that the airline industry faces the challenge of finding the right strategic balance between the provision of efficient service and satisfying individual customer needs. In order to improve service and retain customers, management of customer relationships should synthesize customer touch points. As indicated by Yu (2001), good customer relationship management, (CRM), means presenting a single image of the company across all the channels, a customer may use to interact with the firm, and keep a single image of the customer, that is shared across the enterprise.

Deregulation led the airlines to efficiently allocate resources in an attempt to reduce prices, especially for long haul flights and improve service quality due to increased competition within the industry (Grant, 2008). Furthermore, the delivery of a high level of service quality by airline companies became a marketing requisite in the early 1990s, as competitive pressures continued to increase, thus, forcing a majority of the airlines to offer various incentives, such as the frequent flyer programs, in an effort to build and maintain the loyalty of customers (Miller, 1993). Moreover, this liberalization of the airline industry, prompted the increased need for privatisation of airlines, and most state-owned airlines were challenged by the increasing rivalry. As such, state owned airlines or traditional flag carriers like Air Namibia, increasingly faced with operational challenges and competition from emerging low-cost and low-fare carriers, were forced to either transform or to die. Additionally, state owned airlines, in most countries have seen a reduction in government subsidy, which has pushed them to become more competitive and customer oriented (Doganis, 2006).

Air Namibia

The airline, Air Namibia, is categorized as a full-service carrier (FSC), and is a state owned flag carrier with the GRN being the sole shareholder and operating a Hub-and-spoke network with its major objective to provide full coverage of as many demand categories as possible through the optimization of connectivity in the hub. Air Namibia makes a meaningful economic contribution to the Republic of Namibia. Its services, provide vital domestic and international connectivity to the 560,000 people who flew with the airline in 2015/16. This report explores how the domestic economy benefits from its flag carrier's presence.

This study quantifies the airline's economic contribution through two main channels;
The first is Air Namibia's core contribution to the economy of Namibia. This encompasses the activity sustained by the airline's operations and capital spending, and is quantified in terms of its contribution to the Namibian GDP, the employment it supports and the tax revenues it generates. The second stage captures the wider 'catalytic' economic impact it generates, through the broader activity enabled and stimulated by its services.

In 2015/16, Air Namibia's operations and aviation-related capital spending made a N$704 million contribution to the Namibian economy and sustained 4,550 jobs. In addition to the airline's own operations, Air Namibia spent over N$1 billion on goods and services supplied by local companies. These purchases supported activity in businesses throughout Namibia, as did the spending of wages by those employed by Air Namibia and by firms within its supply chain. These benefits are not only retained within the aviation or tourism sectors, but rather 'ripple out' throughout the economy. Of the nine broad sectors in the Namibian economy, five of them enjoy activity in excess of N$50 million as a result of Air Namibia's operations.

Air Namibia's core business operations, primarily involve the maintenance and provision of air transport services to passengers, and cargo. The airline also provides ground handling services to passengers and aircraft at both the local and international airports. Air Namibia plays a global role by servicing domestic, international, and intercontinental markets with short, medium and long haul flights from hubs to different continents.

Customer relationship management

Customer Relationship Management (CRM) concepts seek to establish long term, committed, trusting and cooperative relationships with customers. This is characterised by openness, genuine concern for the delivery of high quality services, responses to customer suggestions, fair dealing and the willingness to sacrifice short term advantage for long term gains (Bennett, 1996). In applying CRM, the goal of the organizations is to identify their own profitable customers and to provide personalized services, to enhance customer satisfaction and loyalty.

Airliners have to exert more effort and resources to be competitive because competitors offer products and services of similar or superior quality and price, making it difficult for Air Namibia to secure passengers, based on customer satisfaction alone. According to Kotler and Keller (2006), CRM enables organizations to provide excellent real-time customer service through the effective use of individual account information. In the context of airlines, account information can be created from information that passengers provide the airline when booking a ticket or merely inquiring about the airline's service. While taking privacy issues into consideration, the airline can investigate and anticipate customer needs and build relationships with existing and potential customers. Thus, the management of customer relationships ought to be viewed as the next paradigm shift in modern marketing and a potential source of creating sustainable competitive advantage, (Payne and Holt, 2001).

Passenger loyalty

Airlines have over the years acknowledged the importance of building relationships with their customers, with the resulting factor leading to lower customer acquisition costs as well as increased profitability (Kim and Cha, 2002). The management of customer relationships is viewed as the next paradigm shift in modern marketing and a potential source of creating sustainable competitive advantage for airliners (Payne and Holt, 2001). The idea of Relationship Marketing was introduced by Berry (1991) in the Service marketing literature. The ultimate principles of relationship marketing are aimed at establishing and advancing value transactions into cooperatives and profitable relationships that are continually nurtured over the lifetime of a customer. Establishing long-term relationships is a necessity for competitive advantage (Jüttner and Wehrli, 1994). Passenger loyalty, built towards a business is defined as "a profoundly held commitment to rebuild or re-patronize a favoured product/service consistently, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts having the potential to cause switching behaviour", (Oliver, 1999). Gaining loyal customers is a strategy aimed at building mutual rewards to benefit businesses and customers (Reichheld and Detrick, 2003). Passenger loyalty is the extended and uninterrupted retention of the relationship by offering services that meet and even go beyond the customer needs (Acuner, 2001: 89).

Theories of consumer behaviour

The theory of consumer behaviour is essential in in an effort to understand why customers choose to build loyalty towards a specific product, brand and service. Particularly in the airline industry, the study of consumer behaviour involves understanding passenger behaviour, attitudes, values, motivations, perceptions, expectations, preferences, and choices from repurchase to post-purchase (Robinson, 2012). The consumer purchase decision process continues after the initial purchase, after which consumers evaluate their experience, thereafter, coming up with new thoughts and plans on the next purchase decision, based on the previous consumption experience.

Consumer behaviour is dictated by the willingness of a consumer to purchase the same product and keep the same profitable relationship with a particular company (Inamullah, 2012). This study will look into the essence of consumer behaviour theories, which indicate that people learn from experience and, the outcomes of experience will modify their actions on future events. Brand loyalty, is defined as a deeply held commitment to rebuy or use a preferred product or service consistently in the future, resulting in repetitive purchasing behaviour (Oliver, 1999).

Chaudhuri and Holbrook (2001), suggests that behavioural, or purchase loyalty consists of repeated purchases of the brand, whereas attitudinal brand loyalty includes a degree of dispositional commitment in terms of some unique value associated with the brand. Oliver (1999), has proposed four ascending brand-loyalty stages according to the cognition–affect–conation pattern. The first stage is cognitive loyalty, where customers are loyal to a brand, based on their information on that brand. The second phase is affective loyalty, which refers to a customer liking or positive attitudes toward a brand. The third step is conative loyalty or behavioural intention. This is regarded as a deeply held commitment to buy or is seen as a "good intention". This desire may result in unrealized action. The last stage is action loyalty, where customers convert intentions into actions. Customers at this stage experience action, inertia, coupled with a desire to overcome obstacles to make a purchase. Despite action, loyalty is ideal, it is difficult to observe and is often equally difficult to measure. As a compromise, most researchers tend to employ the conative or behavioural-intention model to measure the importance of loyalty. The relationship existing between the airline and its potential retention makes learning theory relevant for this study.

2. Methodology

The study employed the quantitative survey method, where a questionnaire was designed to collect primary data. The target population for this research is the international, regional and domestic passengers using Air Namibia's flights. Air Namibia operates regional, domestic and international flights. The passenger population included selected passengers from the different routes, both departing and returning. The selected routes include Ondangwa, Luderitz, Walvisbay, Katima Mulilo, Johannesburg, Capetown, Harare, Luanda and Frankfurt. Air Namibia brings in about 490 000 passengers a year (NAC annual report, 2012/2013). The airliner deploys its capacity seats of 6048 seats per week to South Africa: Cape Town and Johannesburg. Germany accounts for 1946 seats per week, while Angola accounts for 1512 seats per week. The domestic routes account for 2442 seats per month (CAPA, 2015). These passengers bring in a total population of regional routes to 30240, international to 7784 and domestic to 2442 passengers. The total passenger population is given as 40466 per month giving a total of 10 117 passenger per week.

The quantitative survey uses a stratified convenient sampling. For this research, a stratified random sample was considered appropriate, considering the varied nature of the types and profiles of passengers flying through Hosea Kutako and Eros Airport. The study uses the plane destinations as strata, where the survey will be carried out on passengers boarding domestic, regional and international flights. The total sample size of the passenger sample was 300. Survey questionnaires with close-ended questions were used in this research. The responses were structured using a five point Likert scale. The questionnaire had three sections. Quantitative data collected from the survey was analysed using SPSS version 23. In achieving the study objectives, univariate descriptive analyses, cross tabulation, bivariate analysis and inference statistics were used to interpret the findings of the study.

3. Results

3.1. Inferential Analysis of Research Variables

The study used One-Way ANOVA to examine whether there are significant differences in factorial means from different control variable groups. As a result, the ten factors from CRM initiatives and Value-Added services factorial analysis were treated as dependent variables and the control variables treated as independent variables. In addition, the Scheffe Post Hoc test was used to indicate which group means were different. The results are presented through the control variables as follows:

3.1.1. Loyalty to Air Namibia

The 10 research factors were subjected to a One-Way Analysis of Variance (ANOVA) analysis, with the reason for choosing Air Namibia, as the independent variable. The results are presented in Table 1. It shows how the *reason for choosing Air Namibia* is compared to the factors contributing to passenger retention and loyalty towards Air Namibia's products and services. The results show that those who considered service quality, as their reason for choosing Air Namibia were generally satisfied with the CRM initiatives and value-added services, with means ranging from 3.42 - 4.42. Table 3 shows that the passenger satisfaction means for Loyalty aspects (M=2. 88, S. D=1. 44), Marketing and Promotion Activities (MPA) (M=2. 77, S. D=1. 04), Interpersonal relationships (M=2. 68, S. D=1. 55) and Comfort issues (M=2. 67, S. D=1. 12) are significantly different among the six reasons.

Table 1 Air Namibia Passenger Loyalty

Factor	**N**	Random decision	Flight availability	Price	Business policy	frequent flier membership points	**Service Quality**	**Total**	**F**	**Sig**
Loyalty aspects (Lo)	**163**	**2.62**	**2.52**	**2.90**	**2.48**	**2.99**	**4.42**	**2.88**	**6.99**	**0.00**
domain expertise (DE)	174	3.44	3.84	3.98	3.90	4.02	4.13	3.89	0.90	0.48
customer orientation (CO)	177	3.33	3.84	3.86	3.88	3.57	4.09	3.81	1.32	0.26
service recovery (SR)	166	3.22	3.53	3.50	3.83	3.03	4.08	3.54	2.24	0.05
Operational Specialties (OS)	173	2.78	3.40	3.31	2.94	3.35	4.01	3.39	1.57	0.17
Marketing and Promotion Activities (MPA) Factor	**161**	**2.24**	**2.77**	**2.51**	**2.85**	**2.46**	**3.66**	**2.77**	**5.23**	**0.00**
interpersonal relationships (IF)	178	2.46	2.50	2.37	3.00	3.48	3.58	2.68	2.93	0.01
Value for Money (VM)	163	2.67	2.52	2.50	2.32	2.76	3.54	2.66	2.06	0.07
Comfort Issues (CI) Factor	160	2.04	2.65	2.58	2.44	2.81	3.45	2.67	3.25	0.01
information technology (IT)	159	2.58	3.04	2.92	3.19	2.48	3.42	3.00	1.80	0.12

Table 2 shows that those who chose *service quality* were having statistically significant means of satisfaction to those, who chose Air Namibia, due to *price* ($t = -1.52$, $p < 0.05$) *flight availability* ($t = -1.89$, $p < 0.05$) or just a *random decision* ($t = -1.79$, $p < 0.05$) are significantly different to those passengers who chose Air Namibia due to service quality. The service quality passengers are closely related to passengers who chose Air Namibia because of the frequent flier points or business policy.

Table 2. ANOVA Post Hoc Test

Post Hoc Test -Scheffe	Random decision	Flight availability	Price	Business policy	frequent flier membership points
Marketing and Promotion Activities (MPA) Factor	**-1.42226***	**-.88922***	**.-1.16***	-0.81	-1.20
Comfort Issues (CI) Factor	**-1.41897***	-0.80	-0.88	-1.01	-0.64
Value for Money (VM)	-0.88	-1.03	-1.04	-1.22	-0.78
Operational Specialties (OS)	-1.23	-0.62	-0.70	-1.07	-0.67
Loyalty aspects (Lo)	**-1.79479***	**-1.89329***	**-1.51810***	-1.94	-1.43
customer orientation (CO)	-0.76	-0.25	-0.23	-0.22	-0.52
domain expertise (DE)	-0.69	-0.29	-0.15	-0.23	-0.11
information technology (IT)	-0.83	-0.37	-0.49	-0.23	-0.94
interpersonal relationships (IF)	-1.12	-1.08	-1.21	-0.58	-0.09
service recovery (SR)	-0.86	-0.55	-0.58	-0.25	-1.05

Table 2 also presents the Post hoc test results. These results show that passengers who chose Air Namibia because of service quality, frequent flier points and business policy, could be regarded as loyal customers. While those who chose Air Namibia due to price, flight availability and random decisions, could be regarded, as a prospect. As such, the *reason for choosing Air Namibia* variable was transformed to the *Loyalty to Air Namibia* (LAN) variable, where prospective customers (1) and loyal customers (2) were derived from Table 1 and Table 2 with the area-shaded grey representing the means of loyal customers.

3.2.1. Customer relationship management (CRM) Factors

This section assesses the impact of the (CRM) and passenger loyalty on the airline's value chain. The study used EFA analysis to determine the CRM factors for further analysis. Table 3 presents the findings. Table 3 shows that when the passengers' perceptions are analysed only two factors emerge. These factors are categorized as Satisfaction and CRM factors or Retention and Loyalty factors. In addition, the two factors represent the two extremes of the passenger loyalty ladder, as well as the two extremes of the value chain models under review (Hines and Porter, in press). According to Godson (2009), the passenger loyalty ladder describes the relationship that customers and the organization will have over time. A climb in the ladder can be achieved by understanding the exact need of customers in order to be able to offer them additional value and satisfaction. The passenger loyalty ladder offers a good synopsis that classifies customer values at different stages of the relationship and enhances the chances to get customers that are most loyal. These include partner (high), advocate, supporter, client, customer, and prospect (low).

Table 3 PCA Results for CRM Factors

Item	Retention & Loyalty	CRM
Value for Money (VM)	.792	
Loyalty aspects (Lo)	.759	
Comfort Issues (CI) Factor	.737	
Marketing and Promotion Activities (MPA) Factor	.714	
Interpersonal relationships (IF)	.619	
Operational Specialties (OS)	.571	
Domain expertise (DE)		-.923
Customer orientation (CO)		-.904
Service recovery (SR)		-.776
Information technology (IT)		-.539
Eigenvalue	4.881	1.415
% of Variance	48.805	14.153
Cumulative %	48.805	62.958

3.2.2. Passenger Satisfaction and Loyalty Factors

An analysis was carried out for prospective customers that only chose Air Namibia due to price, flight availability and random decisions. Table 4 shows multiple factor cross loadings from key items like MPA (0.695), Loyalty aspects (0.509), Value for Money (0.675) and Comfort Issues (0.484). However, this does not overshadow the two factors extracted. The assumption, based on the first factor is the satisfaction and the second factor is loyalty.

The critical factors contributing to passenger retention and loyalty towards Air Namibia's products and services include Marketing and Promotional Activities (MPA), Loyalty aspects (Lo), Value for Money (VM), and Comfort Issues (CI). With, the remaining factors contributing to the passenger satisfaction. These include customer orientation (CO), operational specialities (OS), domain expertise (DE), service recovery (SR), and information technology (IT).

Table 4. Passenger Satisfaction and Loyalty PCA (Client/Customer/Prospect)

Item	Satisfaction	Loyalty
customer orientation (CO)	**.949**	
domain expertise (DE)	**.949**	
service recovery (SR)	**.782**	
information technology (IT)	**.707**	
Marketing and Promotion Activities (MPA) Factor	.695	*.305*
Operational Specialties (OS)	**.566**	
Comfort Issues (CI) Factor	.505	*.484*
interpersonal relationships (IF)		**.806**
Value for Money (VM)	*.311*	.675
Loyalty aspects (Lo)	*.421*	.509
Eigenvalue	5.194	1.255
% of Variance	51.94	12.55
Cumulative %	51.94	64.49

3.3. Confirmatory Factor Analysis

This section presents confirmatory factorial analysis of passenger survey questionnaire items.

3.3.1. Control Variables

Section A of the questionnaire covered the control variable factors. From the nine-item control variables, three factors were extracted using a Principal Axis Factoring, with Varimax rotation. Table 5 details the individual factors, eigenvalues, variance explained, and factor loadings associated with each factor.
Table 5 shows that the three factors explained a cumulative variance of 46.4% in control variables. The *frequency of route and travel* explains 19.35% of the variance, while items such as *reasons for choosing an airline* (14.2%), *demography* (12.86%) for variability of the passengers. Atalik and Ozdemir (2015) note that items related to these control variables such as flexibility, providing services for frequent fliers, after-sales services for tickets, and the companies name and reputation influence the satisfaction of the customers. However, *destination*, *purpose and class of travel* items had insignificance coefficient values (<0.3).

Table 5 Control Variables

Items	**Frequency**	**Reason**	**Demography**
Frequent_route	**.588**		
How_frequent_do_you_fly	**-.577**		
Are_you_a_member_of_RFFP	**.426**		
Purpose_of_travel			
Reason_for_choosing_airline		**.666**	
Class_of_travel			
Age			**.415**
Gender			**.360**
Destination			
Eigenvalue	1.741	1.278	1.157
% of Variance	19.35	14.20	12.86
Cumulative %	19.35	33.54	46.40

3.3.2. Passenger Satisfaction with CRM initiatives

The confirmatory factor analysis was conducted to test the proposition that a relationship between the observed variables and their underlying latent constructs exists. This section tests the following hypothesis:

- H_0: There is no correlational association between passenger satisfaction and customer relationship management (CRM).
- H_1: There is a correlational association between passenger satisfaction and customer relationship management (CRM).

Section B of the questionnaire covered the Passenger Satisfaction with CRM initiatives. The latent constructs in Section B include customer orientation (CO), domain expertise (DE), interpersonal relationships (IR), service recovery (SR) and information technology (IT) factors. As a result, Principal Axis Analysis, with Varimax rotation extracted three factors. Table 6 details the individual factors, eigenvalues, variance explained and factor loadings associated with each construct. The three extracted factors are Passenger Satisfaction (PS), Interpersonal relationship (IR) and information technology (IT). The customer orientation, domain expertise and service recovery define the Passenger Satisfaction (PS) factor. The pattern matrix in Table 4.6 was further analysed in SPSS AMOS and Figure 1 presents the results.

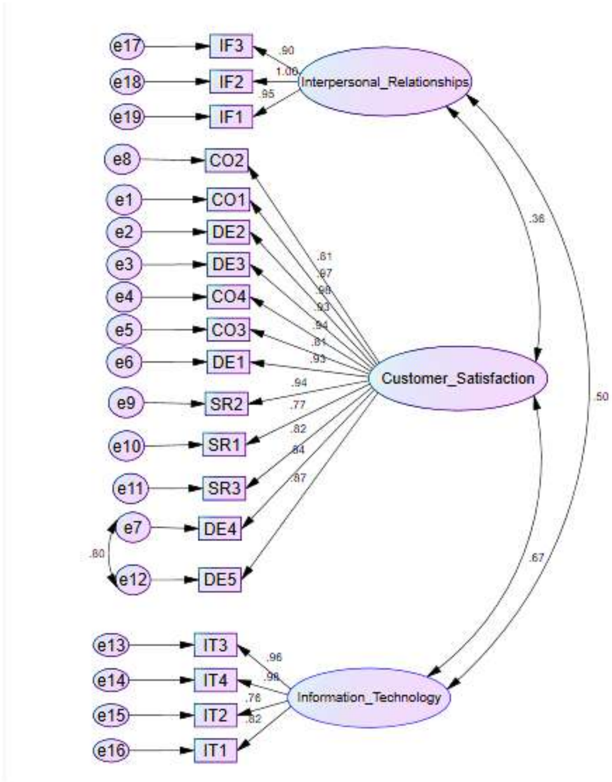


Figure 1 CFA/SEM for Passenger Satisfaction using CRM constructs

Figure 1 above presents the standardized estimate values and relationship pathways for the three Satisfaction levels with CRM factors by presenting the standardized estimates of the relationships between the CRM constructs. The covariant relationship between all constructs with the DE4 (training) and DE5 (highly qualified) particularly strong, which implies the need for continuous training. The CFA analysis results show significant regression weight of 7.676 (0.268) between "Understanding customer needs' (CO2) and the Interpersonal relationship (IR) factor. This means that if we repeat the analysis treating the regression weight for using Interpersonal Relationships to predict CO2 as a free parameter, the discrepancy will fall (become larger) by at least 7.676 (approximately 0.268). This direct relationship between CO2 and IR factors implies that Interpersonal relationships between the staff and the customers are crucial to CRM initiatives as they result in a better understanding of the customer's needs.

3.3.3. Passenger loyalty through Value Added Services

The passenger loyalty factor is used to aggregate loyalty constructs. The data was collected from Section C of the questionnaire, which covered questions on the passengers' satisfaction with value added services, passenger loyalty and overall service quality. The section includes an overall satisfaction rating on services offered after a buying the ticket. As well as the passenger's perceptions of value for money, passenger loyalty and repurchase intention questions. The overall satisfaction rating is assumed to be at par with passenger loyalty (loyalty = satisfaction).

4. CONCLUSION

4.1. Establishing the impact of passenger loyalty on customer relationship management (CRM) in delivering high quality service to passengers and value creation

The management of customer relationship strategies with the sole objective of creating passenger loyalty (Bozgeyik, 2005), enables the synchronization and cooperation between all the crucial elements: customers, business associates and all those involved in the airline value chain. This study established several important findings on the relationship between customer relationship management (CRM) and passenger loyalty, with respect to delivering high quality services and value creation. These findings provide airline management with information regarding the importance of CRM, value creation and the type of strategy intervention needed. Particularly, that interpersonal relationships critically impact passenger loyalty more than the price or value for money. These findings are in line with Jones et al.

(2000) three-item interpersonal relationships (IR) scale. Therefore, interpersonal relationships between the staff and the customers are crucial to CRM initiatives as they result in a better understanding of the customer's needs.

The study compared the *reason for choosing Air Namibia* to the factors contributing to passenger retention and loyalty towards Air Namibia's products and services. The results show that those who considered service quality, as their *reason for choosing Air Namibia*, were generally satisfied with the CRM initiatives and value-added services. The study found that service quality was also closely related to frequent flier points and clients' business policy. The study concludes that service quality, frequent flier points and the client's business policy affect passengers' loyalty to Air Namibia. While, price, flight availability and random decisions, are key factors in attracting new customers.

The findings of the study indicated that the critical elements to passenger retention and loyalty towards Air Namibia's products and services include Marketing and Promotional Activities (MPA), Loyalty aspects (Lo), Value for Money (VM), and Comfort Issues (CI). The study also identified the passengers' perceptions of value for money, passenger loyalty and repurchase intention questions.

On the contrary, Gómez et al., (2006), asserts that loyal customers are highly attractive to businesses, as they are less price sensitive and communicating with them requires a lower effort. In their study, on the factors affecting purchase decisions of domestic airline passengers and their preference priorities, Atalik and Ozdemir (2015), premised that the airline business is a substantially dynamic service-based business that needs to develop value added service factors.

As such, an airline's competitive position is to retain passengers as loyal users of their airline, which results in repeat business from less price sensitive passengers, who require a lower effort to communicate with (Gómez et al., 2006).

In this study, factors were used to assess the priorities of the passenger with regards to loyalty and retention. Interpersonal relationships between staff and the customers, were also considered crucial to CRM initiatives as they result in a better understanding of the customer's needs, which then leads to passenger loyalty.

2. The impact of passenger loyalty on customer relationship management (CRM) in delivering high quality service to passengers and value creation

The study concludes that there is a relationship between passenger loyalty and CRM, and that this relationship can be leverage to deliver high quality services to passengers through value creation. The study also established an association between customer/passenger retention with loyalty. The study establishes that contributory factors affecting passenger retention and loyalty towards Air Namibia's products and services, were:

- Marketing and Promotional Activities (MPA),
- Loyalty aspects (Lo);
- Value for Money (VfM), and
- Comfort Issues (CI).

Additionally, the study also established that contributory factors affecting passenger satisfaction with Air Namibia's products and services, and these were:

- customer orientation (CO);
- operational specialties (OS);
- domain expertise (DE);
- service recovery (SR), and;
- information technology (IT)

The study concludes that passenger retention, passenger loyalty and passenger satisfaction factors are critical to understanding the impact of passenger loyalty on customer relationship management (CRM) in delivering high quality service to passengers and value creation. Air Namibia needs to upgrade and update their information technology infrastructure and systems to allow them to give customers live updates and mail the latest messages. They need to invest heavily in databases and computer networks to manage customer requirements. As they will need to update customer information periodically, integrate information systems and processes that enhance the customer experience.

These secondary factors affect passenger satisfaction with Air Namibia's value chain activities. These include marketing and promotion activities, value for money, operational specialities, loyalty aspects and comfort issues. Air Namibia performed poorly in these areas, with most of their loyal customers disappointed in the value for money and operational specialities.

References

Acuner, S. (2001). Customer Satifisfaction and Measurement. *Mert Matbaasi*(2).

Atalik, O., & Ozdemir, E. (2015). A Hybrid Method Using Factor Analysis and AHP on Passenger Purchase Decisions: The Case of Domestic Airlines in Turkey. *International Business Research, 8*(1), 14-23. doi:10.5539/ibr.v8n1p14

Berry, L. L. (1991). Relationship marketing of services—growing interest, emerging perspectives. *Journal of the Academy of marketing science, 23*(4), 236-245.

Bowen, J. T., & Chen, S. L. (2001). The Relationship Between Customer loyalty and customer satisfaction. *International journal of contemporary Hospitality Management, 13*(5), 213-217.
Bozgeyik, A. (2005). *Why is CRM important*. Retrieved from biligiyonetimi: http://www.biligiyonetimi.org/cm/pages/mkl_gos.php?nt=462

Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing, 65*(2), 81-93.
Doganis, R. (2006). *The airline business in the 21st Century* (2 ed.). London: Routledge.
Godson, M. (2009). *Relationship marketing.* Oxford: Oxford University Press.
Gómez, G. B., Arranz, G. ,., & Cillán, G. J. (2006). The role of loyalty programs in behavioral and affective loyalty. *Journal of consumer marketing, 23*(7), 387-396.
Grant, R. (2008). The Resource –Based Theory of Competitive Advantage: Implications for strategy Formation. *Carlifornia Management Review, 33*(3), 113-134.
Gravetter, F., & Forzao, L. (2012). *Research Methods for Behaviour Sciences.* Wadsworth: Cengage Learning.
Gronholdt, L., Martensen, A., & Kristensen, K. (2000). The relationship between customers Satisfaction and Loyalty. *Total Quality Management, 11*(4-9), 509-514. Retrieved from http://dx.doi.org/10.1080/09544120050007823
Grönroos, C. (1984). A service quality model and its marketing implications. European. *Journal for Marketing, 18*(4), 36-44.
Grönroos, C. (1994). From marketing mix to relationship marketing. Toward a paradigm shift in marketing. *Management Decision, 32*(2), 4-32.
Gurses, F. (2006). Havayolu İşletmeciliğinde Müşteri Tercihlerini Etkileyen Faktörler ve Yerli Yolculara Yönelik Bir Araştırma. *Masters Thesis* . Balıkesir, Turkey: Balıkesir University, Institute of Social Sciences.
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2013). *Multivariate Data Analysis* (7th ed.). USA: Pearson Prentice Hall.
Heinrich, B. (2008). African Infrastructure Country Diagnostic. *Air Transport:Challenge To Growth;Summary of Background Paper*, 16.
Helms, M. M., & Mayo, D. T. (2008). Assessing poor quality service: perceptions of customer service representatives. *Managing Service Quality: An International Journal, 18*(6), 610-622.
Hennig, T. T., & Hansen, U. (2000). Relationship marketing—some reflections on the state-of-the-art of the relational concept. *In Relationship Marketing*, 3-27.
Heskett, J., Sasser, W., & Schlesinger, L. (1997). *The service profit chain. How leading companies link profit and growth to loyalty and value.* New York: Free Press.
Hines, P. (1993). *Integrated Materials Management: The Value Chain Redefined.*
Hines, P., & Rich, N. (1997). The seven value stream mapping tools. *International journal of operations & production management, 17*(1), 46-64.
Hobby, J. (1999). Looking after the one who matters. Accountancy Age. *MoU Press Briefing*, 28-30. Retrieved from http://www.staralliance.com/int/press/media_library/presentation
Holloway, S. (2008). *Straight and level: practical airline economics.* Ashgate Publishing, Ltd.
Hsu, C., & Shih, H. (2008). Small-world network theory in the study of network connectivity and efficiency of complementary International Alliances. *Journal of Air Transport Management, 14*(3), 123-129.
Iacovou, C. L., Benbasat, I., & Albert, S. D. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS quarterly*, 465-485.
IATA. (2011). *Airs@t survey of airline passengers.* Geneva, Switzerland: International Air Transport Association.
IATA. (2012). *The 2012 Annual Review.* International Air Transportation Association.
IATA. (2015). *IATA Global Passenger Survey 2013.* International Air Transport Association.
Jones, M. A., Mothersbaugh, D. L., & Beatty, S. E. (2000). Switching barriers and repurchase intentions in services. *Journal of retailing, 76*(2), 256-274.
Kim, W., & Cha, Y. (2002). Antecedents and consequences of relationship quality in hotel industry.
Kotler, P., & Keller, K. (2006). *Marketing Management* (12th ed.). New Jersey: Pearson Prentice Hall.
Miller, C. (1993, July 5). Rewards for the best Customers: Loyalty Programs Increase as Marketers try to Build Relationships. *Markerting News*, 1-6.
Payne, A., & Holt, S. (2001). Diagnosing Customer Value: Integrating the value process and relationship marketing. *British Journal of Management*, 159-182.
Wilson, S. (2005). *Human Resource Management in the Airline Industry - The Example of star Alliance.* Munich: GRIN Verlag.
Yu, L. (2001). Successful Customer-Relationship Management. *MIT Sloan Management Review, 42*(4), 81-86.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Application of Longitudinal Analysis to Crime Data: Windhoek Case study (2011-2016)

Lazarus Unandapo[1,*], Augustinus Ntjamba[1]

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: lunandapo@unam.na

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Longitudinal data analysis,
Crime analysis,
Windhoek,
Namibia

ABSTRACT

Crime in the Windhoek municipal area continue to be on an increase trend as the population grow over time. Despite past effort done to reduce crime, crime seems to be on a continuously increasing trend; mostly in area regarded as crime hotspots by Windhoek municipal Police. Past study done to analyse the crime record have concentrated mostly on cross-sectional analysis, which does not take correlation into account, thus makes it difficult to compare snapshots of crime over time. The main aim of this research was to analyse reported crime data for the period (2011-2016), using a more robust method known as longitudinal data analysis. This method helped us to visualise the different crime frequencies at different time points (month, day or time of day) in all identified police zones. Furthermore, the use of Generalised Estimating Equations (GEE) was also done, to model these crime data, where the best correlation structure was identified to be the exchangeable correlation structure, which assume constant correlation over time.

1. Introduction

Windhoek, the capital city of Namibia, is developing and the locations are expanding due to people migrating from several rural areas and other towns in the country to seek for better living. This is mainly due to the fact that the city is a big attraction to foreign investment and tourist which attribute to better living condition and employment opportunities as compared to many other towns in the country. As the population increases, so is the unequal distribution of wealth, which then leads to higher poverty rate. As a developing country, crime in Namibia fluctuates overtime. According to Neema and Böhning (2012) developing countries tend to have quite a high rate of crime due to unfavourable prevailing socio-economic conditions, high unemployment levels, and lack of organised policy and justice systems among others. The crime in Namibia especially in the capital city appears alarming. It is hard to note if crime is on increase, decrease or constant.

Research conducted on crime in Namibia have often bear unsatisfactory results due to the inability to compare types of crime on a longitudinal platform. The Namibian Police Force (Nampol) reported an upward trend in overall crime statistics for Wanaheda police station for a period of five year (2014/15-2018/19) (Lilungwe 2020).

Lilungwe (2020) further augured that the most prevalent crime reported in Khomas region are theft, assault, common assault, housebreaking and robbery. Crime such as theft, housebreaking, and robbery are referred to as property crime while assault, common assault is referred to as violent crime in the study. It was found that bad economies at times lead to more property crime as criminals steal popular items that they cannot afford (Kathena and Sheefeni, 2017). They further argue that hard economic times results in more domestic violence and greater consumption of mind- altering substances such as drugs and alcohol and in return this result even in more crime. The study of crime data is not only unique to Namibia, as an example, Tang (2011) researched on the dynamic relationship between tourist arrivals, inflation, unemployment and crime rates in Malaysia. Throughout almost two decades since 1990, reducing crime in Malaysia has been viewed as an urgent task for the policymakers and the royal Malaysia policy (RMP), because of the deep impact on socio-economic development.

It has been a challenge to measure change of a particular variable of interest over time. Longitudinal research employs continuous or repeated measures over time to follow a group of subjects or samples from the same population. In most cases longitudinal studies, time can be measure as calendar time, day, months, years, decades, whichever suit the situation better. Two important characteristics of change need to be captured in longitudinal study, namely within-unit change across time, or growth trajectory, and inter-unit differences in change. It is vital to be precise about variables that are expected to change and the reason for changing overtime (Ployhart and Vandenberg 2010).

To study how crime would relate to the hotspot overtime, using repeated measurement, the substantive meaning of change and the time need to be theorised, prior to the actual crime analysis. Furthermore, it is clear that crime does not change, revolve, or develop because of time; rather it does so overtime. The longitudinal approach attempts to describe the form of change overtime, and this can be either linear or nonlinear. Some features distinguishing longitudinal studies includes correlated observations (due to the variable measurements at multiple time points), high possibility of missing data (due to the rigorous follow-up needed for each subject) and the existence of multiple covariates. Generalised Estimating Equation (GEE) is a general statistical approach to fit a marginal model for longitudinal/clustered data analysis (Wang and Carey, 2014). The method is popular into the field of medicine and psychology, for instance, orthodontic measurements on children of different ages and response is the measurement of the distance from the centre of the pituitary to the pterygomaxillary fissure, measure at a different age repeatedly. The primary goal was to understand the patterns of crime over the study period and investigate how crime has change. The GEE model makes use of a population- averaged estimates where the quasi-likelihood function approach applies. The method develops as a means of testing hypothesis about the effect of factors on binary ad other exponentially distributed response variables.

2. Methods

A quasi-likelihood estimator, as defined by Zeger and Liang (1986) is a solution to the score-likelihood equation system given below:

$$\boldsymbol{S}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\frac{\partial \mu_i}{\partial \beta}\right]^T V(\hat{\alpha})_i^{-1}\big(y_i - \mu_i(\boldsymbol{\beta})\big) = 0 \qquad (1)$$

where $\boldsymbol{y} = (y_1, \dots, y_n)$ is a vector of outcomes (y_i) variable decomposed into n strata with μ_i an expected value of y_i given as:

$$\mathrm{E}(y_i) = h^{-1}\big(X_i(\boldsymbol{\beta})\big) \qquad (2)$$

X = (X_1, ..., X_n) is an $n \times p$ design covariate matrix of predictor variables decomposed into n strata and $\boldsymbol{\beta}$ its an $k \times 1$ vector of regression parameters. Here p is the dimension of each of the strata and k is the dimension of the vector of regression parameters. According to Zeger and Liang (1986), Zorn (2001), $\boldsymbol{h}$ is a link function, which specifies the relation- ship between $E(\boldsymbol{y}_i)$ and the X_i. This function transforms the expectation of the response variable μ_i to linear predictors, e.g. $h(\mu_i) = X_i(\boldsymbol{\beta})$. $V(\hat{\alpha})_i$ is the variance of $\boldsymbol{y}_i$ given as a known function g of $E(\boldsymbol{y}_i)$, e.g. $V(\hat{\alpha})_i = g(\mu_i)\varphi$ where φ is a scale parameter and $\hat{\alpha}$ is a consistent estimate of α (Zorn 2001). The solution to Equation (1) can be obtained by the method of iteratively re-weighted least squares (IRWLS) as stated by Zorn (2001), Zeger and Liang (1986), and Millar (2011). According to Crowder (1995) specifications of the correlation between the $\boldsymbol{y}_i$ can be avoided by assuming a prior working correlation matrix (working correlation structure) $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ when re- peated measurements are analysed using GEE models. Here $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ is a fully specified vector of unknown regression parameters (Weiss, 2005). The choices of working correlation matrix include independent working correlation matrix, exchangeable working correlation matrix, first order auto-regressive (AR1) working correlation matrix and unstructured working correlation matrix among others. Each $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ has its own assumptions, for example, the independent $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ assumes zero correlation between the subsequent measurements, exchangeable $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ assumes constant correlation across all observations in a strata (in this case seasons), while AR1 $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ assumes that two measurements taken one time point away within a strata tend to be highly correlated than two observations taken far apart in the same strata. See Weiss (2005), Pan and Connett (2002), Zorn (2001), Cui and Qian (2007), Wang and Carey (2003) for more choices of $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$.

Given $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ for response vector $\boldsymbol{y}$, Pan and Connett (2002), Zorn (2001), Zeger and Liang (1986), expressed the covariance matrix $\boldsymbol{V}(\hat{\boldsymbol{\alpha}})$ in terms of the correlation matrix $\boldsymbol{V}(\hat{\boldsymbol{\alpha}})$ as:

$$V = V(\hat{\alpha}) = A^{1/2} R(\hat{\alpha}) A^{1/2} \qquad (3)$$

where A = $diag(V(y_1), V(y_2), \dots, V(y_p))$ better link A and $V(y_i)$ is a diagonal matrix with $V(y_i) = V(\mu_i)$. The extension of Equation (1) to longitudinal data is expressed as:

$S(\beta) = \sum_{i=1}^{n} D_i^T \, V(\alpha)_i^{-1}(y_i - \mu) = 0$ (4)

with $D_i = D_i(\beta)$ the partial derivative of μ_i with respect to β. When $n = 1$, Zeger and Liang (1986) note that Equation (4) reduces to the quasi-likelihood estimation. They further state that when the link function h is correctly specified, the GEE (4) give consistent regression coefficients. Equation (4) is a score equation for $\boldsymbol{\beta}$, and depends on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (Zorn, 2001; Zeger and Liang, 1986).

Zeger and Liang (1986) replaced $\boldsymbol{\alpha}$ with some $K^{1/2}$ consistent estimator, $\hat{\boldsymbol{\alpha}}(y, \boldsymbol{\beta}, \varphi)$, in Equation (3) and (4) to express the two equations as functions of $\boldsymbol{\beta}$ only. They also replaced the scale parameter φ in $\hat{\boldsymbol{\alpha}}$ by $K^{1/2}$ consistent estimator, $\varphi(y, \boldsymbol{\beta})$, so that the estimate $\hat{\beta}$ of $\boldsymbol{\beta}$ is expressed as a solution to:

$$\sum_{i=1}^{n} \mathrm{m}U_i\{\beta, \hat{\alpha}[\beta, \hat{\phi}(\beta)]\} = 0. \qquad (5)$$

with $U = D^T \; V^{-1} \; S$ as a function of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. When K increases to infinity, $\hat{\beta}$ becomes a consistent estimator of β *and* $K^{1/2}(\hat{\beta} - \beta)$ becomes a multivariate Gaussian with covariate matrix V_β, which consistently estimate the variance (Zeger and Liang 1986; Oh et al. 2008):

$$V_\beta = \underbrace{K(\Omega)^{-1}\left[\sum_{i=1}^{n} D_i^T \, V_i^{-1} cov(y_i) V_i^{-1} D_i\right](\Omega)^{-1}}_{limit\ as\ K\to\infty}$$

solving the GEE for $\hat{\beta}$, one first has to solve for the regression coefficients, the correlation $\boldsymbol{\alpha}$ and scale parameter φ. If we are given an estimate of working correlation matrix $\boldsymbol{R}(\hat{\boldsymbol{\alpha}})$ and scale parameters φ, then $\boldsymbol{\beta}$ can be calculated by IRWLS method. If the V_i is reasonably approximated, then the estimates of $\boldsymbol{\beta}$ is efficiently relative to ML estimates.

3. Results

Figure 1 shows the crime trends across each Zone for the combined period 2011-2016. Overall, property crime was highest in Zones 1 and 14, while Zones 1 and 6 showed an increasing trend in violent crime as shown in figure 1. Moreover, there remaining crime types recorded low values across all 19 Zones, with the exception of property crime that had low values for Zones 2, 8, 12, 15, and 19.

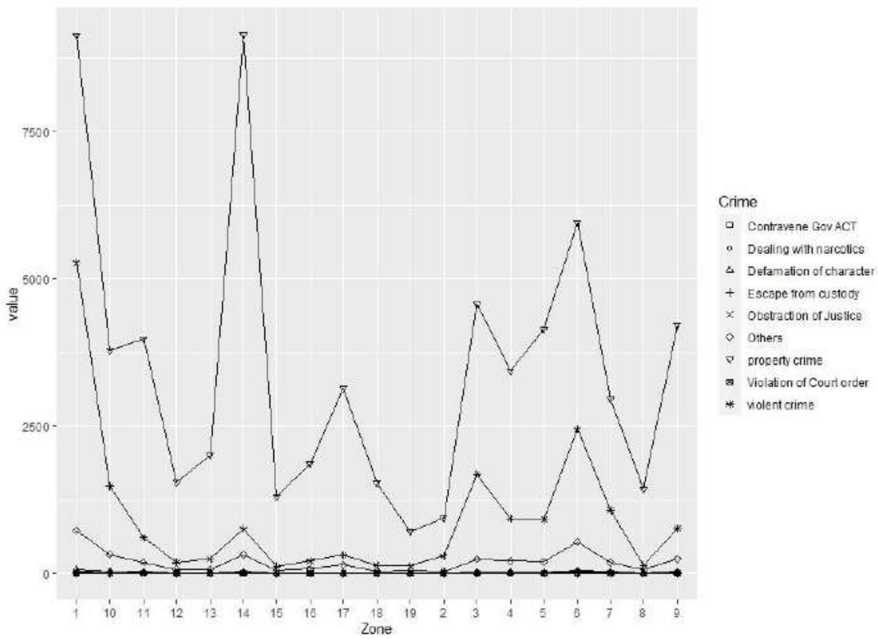


Figure 1: Crime trends across each Zone for the combined period 2011-2016

Figure 2 on the other hand, shows the overall crime trends for each hour of the day for the combined period 2011-2016. It can be observed that crime mostly happened from 09 AM to midnight, most probably due the fact that the city gets busy during these hours and criminals take advantage of this to commit crime. It can be noted that the frequencies of violent crimes and property crime is at its highest during this time period when most people are at work. The graph also showed that from midnight to early hour in the morning (01:00- 09:00) crime were very low. This may be due to police patrol and shutting down shebeens and other public gathering where crime is likely to occur and also everyone else being home with families.

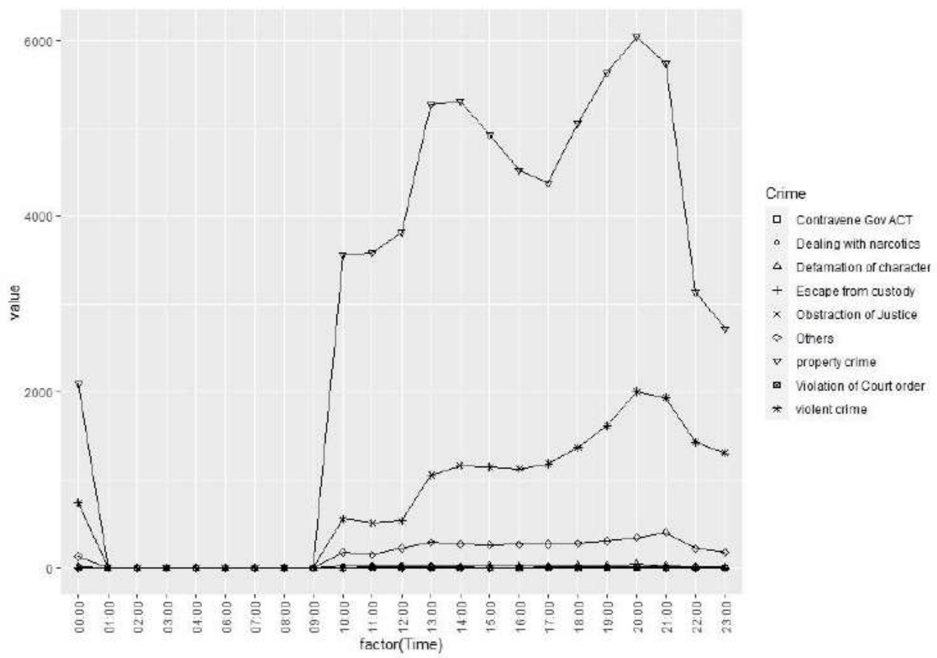


Figure 2: Overall crime trends for each hour of the day for the combined period 2011-2016

Finally the figure 3 shows the snapshots of this crime behaviours over months for the period 2011-2016. From this figure it can be seen that both property crime and violent crime have been fluctuating across the months but remained high in December, compared to the remaining types of crime reported in 2011-2016. During this vacation period, a lot of houses stay unoccupied hence a perfect opportunity for perpetrator. A lot of social events also take place during this time period, making it easier for things to be stolen from vehicles while individuals are at gatherings, thus leading to high increase of violent and property crime. The 'others' crime category remained flat from January to April and then peaked up from there and remained a bit high until December.

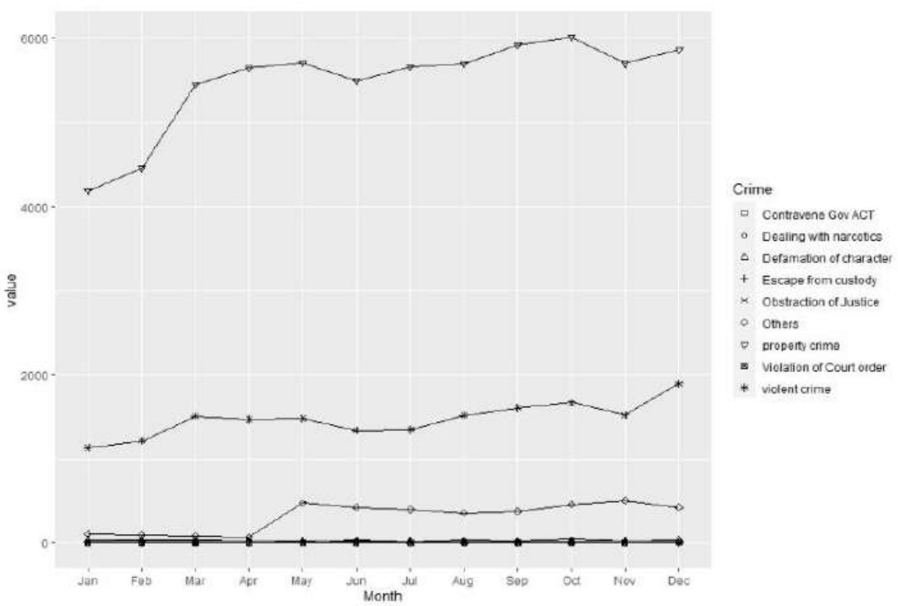


Figure 3: Snapshots of same crime behaviours over months for the period 2011-2016

Table 1 shows the output summaries of the fitted model for different correlation structure. The presents output from model fit with intercept and slope parameter only for crime data. The same results were produced for all the three (Exchangeable, Independent, and Unstructured) working correlation structure. This occurred, perhaps, due to the redundant in the data. Consequently, the researcher was unable to select the best model for the data since the QIC value were the same as well as the standard errors.

Table 1: Output summaries of the fitted model for different correlation structure

	Working correlation structures					
	Ind		Exch		Unstr	
Parameter	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**
Intercept	7.1673	0.0139(0.0142)	7.1673	0.0139(0.0142)	7.1673	0.0139(0.0142)
Zone	-0.0059	0.0005(0.0005)	-0.0059	0.0005(0.0005)	-0.0059	0.0005(0.0005)
Month	-0.0045	0.0009(0.0009)	-0.0045	0.0009(0.0009)	-0.0045	0.0009(0.0009)
Time	0.0146	0.0007(0.0007)	0.0146	0.0007(0.0007)	0.0146	0.0007(0.0007)

In theory the best working correlation structure is the unstructured one but the challenge with that is that if the data set is too big, then it estimates as many parameters as we can accommodate in any report. hence the exchangeable correlation structure could be adopted for this research. A different data set on medicare longitudinal data was used to test the problem with results obtain in this report and confirm that the theory is working fine for the medicare data, hence the conclusion that out crime data need serious revisitation to figure out the error.

4. Discussion

This study has demonstrated the analysis of longitudinal crime data in Windhoek municipality using Generalised Estimating Equations model fit. The study also looked at some factors that lead to crime. It was found that there are some similarities in the literature and the study results.

During the study period, the number of property and violent crime was high in Windhoek; these could be due to unemployment factor. This result was similar to the findings of Lilungwe (2020) on the relationship between youth and unemployment, which showed that there was a direct relationship between youth and unemployment. The number of property crime remained higher.

The number of property and violent crime were high in the month of January to April, while the remaining crimes were low. During December, property and violent crime increased, because residents get to celebrate social event such as Christmas leaving their houses unattended while some get drunk and start to have conflicts.

It was found that there was a correlation between variables (crime, Months, Zone, and Time), it is important to note that correlation does not mean interconnection (cause). This would not mean that there more people in a certain location during the day the more crime being committed, but also this could be the possibility that there is lack of police patrolling in the locations at that time. Furthermore, not all crime incidents that took place in various locations were reported and this lead to bias. It was found that for any change in Zone, crime decrease while for every hour increase in time, crime increased.

References

M. Crowder. On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika, 82(2):407–410, 1995.

J. Cui and G. Qian. Selection of working correlation structure and best model in gee analyses of longitudinal data. Communications in Statistics Simulation and Computation, 36 (5):987–996, 2007.

Neema I and Böhning D. Monitoring murder crime in Namibia using Bayesian space-time models. Journal of Probability and Statistics, 2012, 2012.

Ignatius N Kathena and Johannes PS Sheefeni. The relationship between eco- nomic growth and crime rates in Namibia. European Journal of Basic and Applied Sciences Vol, 4(1), 2017.

Mayumbelo Lilungwe. Youth unemployment and property crime in Namibia: A case of Samora Machel constituency. PhD thesis, 2020.

R. B. Millar. Maximum likelihood estimation and inference: with examples in R, SAS and ADMB, volume 111. John Wiley & Sons, 2011.

S. Oh, K. C. Carriere, and T. Park. Model diagnostic plots for repeated measures data using the generalized estimating equations approach. Computational Statistics & Data Analysis, 53(1):222–232, 2008.

W. Pan and J. E. Connett. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. Statistica Sinica, 12(2): 475–490, 2002.

Robert E Ployhart and Robert J Vandenberg. Longitudinal research: The theory, design, and analysis of change. Journal of management, 36(1):94–120, 2010.

Chor Foon Tang. An exploration of dynamic relationship between tourist arrivals, inflation, unemployment and crime rates in Malaysia. International Journal of Social Eco- nomics, 2011.

Y. G. Wang and V. Carey. Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. Biometrika, 90(1):29–41, 2003.

R. E. Weiss. Modeling Longitudinal Data: With 72 Figures. Springer Science & Business Media, 2005.

S. L. Zeger and K. Y. Liang. Longitudinal data analysis for discrete and continuous outcomes. Biometrics, pages 121–130, 1986.

C. J. W. Zorn. Generalized estimating equation models for correlated data: A review with applications. American Journal of Political Science, pages 470–490, 2001.

Namibian Journal for Research, Science and Technology
Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

Socio-Demographic Variations on Age-Sex Mortality in Namibia: An Analysis of the 2016 Civil Registration and Vital Statistics Data

Immanuel Shipanga[1*], Opeoluwa Oyedele[1], Tarcy C. Matengu[1]

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: ishipanga@unam.na

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Mortality, Mortality rates, Age-sex variation, Namibia

ABSTRACT

Mortality studies are important for the effectiveness of subsystem of health services in a country. Before the Coronavirus pandemic outbreak, there has been a gradual decline in the global mortality indicators, which can be linked to the improving economic development and social wellbeing of the global population, especially, in developing regions such as East Asia and the Pacific, Latin America and the Caribbean, Eastern Europe and the Middle East. This decline in mortality and the high fertility in developing countries are the contributing factors to the increase on global population. Apart from the aggregated frequencies of deaths and crude death rates reported in population census reports, little to no attention has been paid to detailed inference mortality analysis with respect to the age-sex variation perspective in Namibia. Thus, this paper used the negative binomial regression modelling technique to perform an inference mortality analysis across all ages and both sexes in the country as well as across regions and marital status using the 2016 Civil Registration Vital systems data from the Ministry of Home Affairs and Immigration. Results showed that there was a significant relationship between mortality and the individuals' age, sex, marital status and region. In addition, Oshana, Kavango East, Khomas, Hardap and Omaheke regions had high mortality rates, while infants and elderly individuals had a high probability of dying. Furthermore, the study revealed that individuals who were single and aged 15-59 and 5-14 years had less expected death count. Hence, it is recommended that interventions (such as affordable and proper health care and well-being services) targeted at the (most) vulnerable age groups, marital group and regions be made a priority, in order to meet Sustainable Development Goal 3.

1. Introduction

Mortality can be defined as the deaths in a given population (Weeks, 2014) or the deaths recorded in a particular population due to a specific cause and at a specific period of time (Porta, 2014). There are two biological components of mortality: (i) the lifespan component - the oldest age to which human beings can survive; (ii) the longevity component - the ability of people to resist death (Weeks, 2014). Weeks (2014) further stressed that the associated mortality impact can be caused by multiple infectious diseases such as communicable diseases (like Tuberculosis and Human Immuno-deficiency Virus/Acquired Immune Deficiency Syndrome (HIV/AIDS)), non-communicable diseases (like cancer and degenerative diseases) and deaths due to accidental and intentional injuries (like vehicle accidents, homicide and suicide). There are six types of mortality indicator, depending on the time death has occurred. The first indicator is infant mortality which is the number of deaths of children under 1 year per 1000 live births in a population (United Nations, 2019). This indicator can also be used to examine the risk of infant death at the vulnerable stage of life (i.e., less than 1 year old). The second indicator is child mortality which is the number of children dying under the age of 5 years per 1000 live births in a population (United Nations, 2012), while the third indicator is the older children death for children aged 5 to 14 years (United Nations, 2019). The fourth indicator is the youth mortality for persons aged 15 to 34 years old, while the fifth and sixth indicators are the adult mortality for persons aged 35 to 64 years old and the old age mortality for persons over 65 years respectively (United

Nations, 2019).

Mortality measures are important as they can provide information that are useful for development levels of the country in terms of healthy lives and well-being for all at all ages. From the Millennium Summit of 2000 to the Sustainable Development Summit of 2015, the United Nations as well as the global community made it their agenda to promote good health and fight against several leading causes of death and diseases. The global burden of disease studies of 1990 and 2000 estimated that Sub-Saharan Africa has the highest burden of disease in the world (WHO, 2000). These and other related studies revealed that the high levels of mortality were resulted from multiple diseases. When looking at the age and sex specific mortalities globally, as life expectancy increased and people survived in large numbers to older ages, most societies experienced less variability in the ages at which the people died, while women generally live longer than men due to several factors including the basic biological superiority in the ability of females to survive longer than males (Weeks, 2014). In developed countries the life expectancy of people is high as compared to the people of developing countries of the world (Khan & Khan, 2016). Males had a global average life expectancy of 70 years and females had a global average life expectancy of 75 years in 2020 (Statista, 2021). Males lived an average of 79 years and females lived an average of 82 years in more developed countries (United Nations, 2019). The situation is different for developing countries, males had a life expectancy of 63 years and females had a life expectancy of 67 years (United Nations, 2019). Closer to home, males in South Africa were 60 years old and females were 67 years (Statistics South Africa, 2014).

In Namibia (Republic of Namibia, 1994), the age specific mortality pattern had a less severe W-shape in 2011, like many other countries in Sub-Saharan Africa. This was because mortality rose among younger and middle aged adults due to the high prevalence of HIV/AIDS and these age groups were more at risk of contracting HIV infections (Namibia Statistics Agency (NSA), 2014). To be precise, the overall mortality pattern and rates had changed over time from 2001 to 2011, with the W-shape worsened between 2002 and 2006, and less severe in 2011. This was because of the decreasing impact of HIV/AIDS on mortality due to the introduction of health intervention programmes in the country (NSA, 2014). Apart from the aggregated frequencies of deaths and crude death rates reported in population census reports, little to no attention has been paid to detailed inference mortality analysis with respect to the age-sex variation perspective in Namibia. Thus, this paper examined the socio-demographic disparities in mortality in Namibia, as well as the difference in age-sex specific mortality. Findings from this study can provide a better understanding of how aggregating mortality by age and sex can help reveal the implausible mortality distributions from what would normally be expected with links to other socio-demographic factors in Namibia.

2. Methodology

2.1 Data source

The data used in this paper were extracted from the 2016 Civil Registration Vital Systems (CRVS) from the Ministry of Home Affairs and Immigration's Department of Civil Registration which is responsible for death records and issuance of death certificates. The CRVS data is collected by the Ministry of Home Affairs and Immigration (MHAI) as prescribed under the Births, Marriages and Deaths Registration Act 81 of 1963. This Act makes it mandatory for all events (deaths) to be registered in the country, and making an illegal act to bury a lifeless body without a death certificate and burial order. Once death has occurred, the MHAI is notified and the death is recorded on the registry. The information required on the death notification includes the date and place of death (region and constituency), name, and sex, date of birth, marital status and citizenship of the deceased. Further, the death notice indicates whether the death occurred in a hospital (and the name of medical practitioner) or some other places. Where the event happened out of state facilities (hospitals) or unnatural death is involved, the Namibian police ascertain the cause of death and recorded as such. The data are then captured in the system that is integrated with all regional offices. In addition to the CRVS data, data from the Namibia Intercensal Demographic Survey (NIDS) conducted in 2016 by NSA was used to estimate the population size in 2016 as well as for the mortality rates calculation.

2.1 Statistical analysis

In this paper, the socio-demographic differences in mortality were evaluated for four measures of mortality (life expectancy rates, crude death rates, age/sex specific death rates and probability of dying). The Crude Death Rate (CDR) can be defined as the total number of deaths in a year divided by the average total population expressed per a thousand people (Weeks, 2014). That is,

$$CDR = \frac{Total\ Number\ of\ deaths\ in\ a\ year}{Total\ Population} \times 1000$$

The Age/Sex Specific Death Rate (ASDR), used to account for the differences in dying by age and sex, can be defined as the number of deaths of people aged x years divided by the population aged x years expressed per a thousand people (Weeks, 2014). That is,

$$ASDR = \frac{{}_nd_x}{{}_nP_x} \times 1000$$

where ${}_nd_x$ is the number of deaths in a year of people of a particular age group (typically a five year age group) in the interval x to $x+n$, with x being the lower limit of the age interval and n representing the width of the interval, while ${}_nP_x$ is the average number of people of that age group in the midyear population. A life table, consisting of ages groups (grouped into five-year categories instead of single years of age), population of each age group and the number of deaths of each age group, estimates the probability of dying and the survivorship of individuals within a given population (Weeks, 2014). To be precise, it estimates the probability that persons at each age group die before their next birthdays and number of years that they are expected to live before dying.

Furthermore, a negative binomial regression model was fitted, in order to estimate the selected socio-demographic variables (age, sex, region and marital status) specific risk of mortality. Negative binomial regression model is a generalization of a Poisson regression model which loosens the restrictive assumption of the Poisson model that the variance is equal to the mean (NCSS, 2021). Both models are similar to the regular multiple regression model (Oyedele & Lubbe, 2018) except, here, the response variable (y) is an observed count with nonnegative integer values (Zwilling, 2013). With the negative binomial modelling, y follows the negative binomial distribution which is defined in terms of the number of trials until the r^{th} success (NCSS, 2021). This is slightly comparable to the binomial distribution, except in this distribution the number of successes is fixed while the number of trials is counted, whereas in the binomial distribution the number of trials is fixed while the number of successes is counted. Generally, the negative binomial regression model is specified as

$$ln(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

for a set of predictor variables $x_1, x_2, \dots, x_p$, where μ is the mean of y (Oyedele & Lubbe, 2018), β_0 and $\beta_1, \dots, \beta_p$ are the intercept and unknown parameters that are estimated from a set of data respectively (NCSS, 2021). In this paper, the predictor variables were the individuals' age group in years (under 1, 1-4, 5-14, 15-59 & 60+), sex (male & female), region (!Karas, Erongo, Hardap, Kavango East, Kavango West, Khomas, Kunene, Ohangwena, Omaheke, Omusati, Oshana, Oshikoto, Otjozondjupa & Zambezi) and marital status (single, married, divorced, widowed & unknown), while the response variable was the number of mortality cases. Thus, y_{ijkl} being the number of mortality cases across sex i, region j, marital status k and age group l for $i = 1, 2,$ $j = 1,2, \dots, 14,$ $k = 1, 2, \dots, 5,$ and $l = 1, 2, \dots, 5$ can be modelled as

$$\begin{aligned} ln(\mu_{ijkl}) = \beta_0 &+ \beta_1(Deaths\ by\ sex\ i) \\ &+ \beta_2(Deaths\ by\ regions\ j) \\ &+ (Deaths\ by\ marital\ status\ k) \\ &+ \beta_4(Deaths\ by\ age\ group\ l) \end{aligned}$$

where μ_{ijkl} is the mean of y_{ijkl}.

3.Results

3.1 Mortality distribution in Namibia

In 2016, a total of 19225 deaths were recorded in Namibia, out of which 10475 (54%) deaths were male deaths, while 8750 (46%) were female deaths as shown in Table 1. Looking at the mortality distribution by regions, it can be observed that there were more deaths in the highly populated and urbanised regions of Khomas (22.5%) and Oshana (18.9 %), while the lowest deaths were recorded in the Kavango West region (1.9%). Out of the 19225 recorded deaths, majority were recorded among the individuals who were single (73%), followed by those who were married (15%). In addition, there were more single and married males dying compared to their female counterparts as shown in Table 1. Moreover, the highest mortality recorded in 2016 was among the 15-59 age group (40.4%), followed by the 60+ years old age group (35.7%). In addition, there were slightly more male babies dying (8.8%) compared to female babies (7.9%)) among the infants (under 1 year) category, while the opposite is observed for the 60+ years age group.

Table 1: Mortality distribution across region, marital status, ages and sex in 2016

	Deaths by Sex			
	Female	Male	Both sexes	Chi-square & P-value
Region				
!Karas	258 (1.3%)	406 (2.1%)	664 (3.5%)	
Erongo	416 (2.2%)	598 (3.1%)	1014 (5.3%)	
Hardap	374 (1.9%)	507 (2.6%)	881 (4.6%)	
Kavango East	896 (4.7%)	983 (5.1%)	1879 (9.8%)	
Kavango West	176 (0.9%)	191 (1.0%)	367 (1.9%)	

Khomas	1920 (10.0%)	2403 (12.5%)	4328 (22.5%)	74.811 & <0.001
Kunene	222 (1.2%)	227 (1.2%)	449 (2.3)	
Ohangwena	628 (3.3%)	642 (3.3%)	1270 (6.6%)	
Omaheke	320 (1.7%)	432 (2.2%)	752 (3.9%)	
Omusati	814 (4.2%)	894 (4.7%)	1708 (8.9%)	
Oshana	1717 (8.9%)	1924 (10.0%)	3641 (18.9%)	
Oshikoto	166 (0.9%)	302 (1.6%)	468 (2.4%)	
Otjozondjupa	458 (2.4%)	572 (3.0%)	1031 (5.4%)	
Zambezi	385 (2.0%)	394 (2.0)	779 (4.1%)	
Marital status				
Single	6317 (32.9%)	7653 (39.8%)	13970 (72.7%)	
Married	950 (5.0%)	1936 (10.1%)	2886 (15.0%)	
Divorced	62 (0.3%)	57 (0.3%)	119 (0.6%)	582.896 & <0.001
Widowed	651 (3.4%)	198 (1.0%)	849 (4.4%)	
Unknown	16 (0.1%)	36 (0.2%)	52 (0.3%)	
Ages				
Under 1	1528 (7.9%)	1698 (8.8%)	3226 (16.8%)	
1-4	443 (2.3%)	486 (2.5%)	929 (4.8%)	
5-14	187 (1.0%)	237 (1.2%)	424 (2.2%)	551.794 & <0.001
15-59	3102 (16.1%)	4673 (24.3)	7775 (40.4)	
60+	3490 (18.2%)	3381 (17.6%)	6871 (35.7%)	

3.2 Mortality rates

Table 2 shows the crude death rate estimated for each region and for the whole country. The Oshana region can be seen to have the highest crude death rate of 19.2 per 1000 people followed by the Kavango East, Khomas, Hardap and Omaheke regions with 12.66, 10.4, 10.1 and 10.08 deaths per 1000 people respectively.

Table 2: Crude Death Rates per region

Regions	Population	Number of deaths	CDR
!Karas	85 759	664	7.74
Erongo	182 402	1014	5.56
Hadarp	87 186	881	10.10
Kavango East	148 466	1879	12.66
Kavango West	89 313	367	4.11
Khomas	415 780	4323	10.40
Kunene	97 865	449	4.59
Ohangwena	255 510	1270	4.97
Omaheke	74 629	752	10.08
Omusati	249 885	1708	6.84
Oshana	189237	3641	19.24
Oshikoto	195 165	468	2.40
Otjozondjupa	154 342	1030	6.67
Zambezi	98 849	779	7.88
Namibia	**2324388**	**19225**	**8.27**

3.3 Age-specific mortality rates

The age specific death rate was calculated for all the age groups and are shown in Table 3. It can be seen that the under 1 year age group had the highest age specific death rate of 47.6% followed by the 60+ years old age group with 47.1%, both of which are the most vulnerable to conditions that are major causes of death. The lowest age specific death rate was for the age group 5-14 years (0.8%), followed by the 1-4 years (3.7%), both of which are the less vulnerable to conditions that are major causes of death.

Table 3: Age Specific Mortality Rates, Namibia, 2016

Age Group	*Female Deaths (Population size)*	*Male Deaths (Population size)*	*Total Deaths (Population size)*	*ASDR*
Under 1	1528 (33319)	1698 (34417)	3226 (67735)	47.6
1-4	443 (125826)	486 (128365)	929 (254189)	3,7
5-14	187 (260436)	237 (263834)	424 (524270)	0,8
15-59	3102 (688781)	4673 (630687)	7775 (1332331)	5,8
60+	3490 (31964)	3381 (59589)	6871 (145861)	47.1
Namibia	**8750 (1194634)**	**10475 (1129754)**	**19225 (2 324 388)**	**8.27**

3.4 Life Table

The life tables used in this paper were the abridged life tables showing the population of each age group and the number of deaths of each age group. Additionally, the age sex death rate for both females and males in 2016 were computed and shown in column 5 of Annexure 1 and 2, in order to get the probability of dying before reaching the next birthday (shown in column 6 of Annexure 1 and 2). From both Life Tables (Annexure 1 and 2) the probability of dying is slightly higher in the first ages of life compared to the middle-aged children and then progressively rises from young adult ages until the last age group which had the highest probability of dying (1.00). The expectation of life at birth for males in Namibia was 67.9 years while for females it was 70.1 years in 2016. That is, if all persons born in the year 2016 had the same risks of dying throughout their lives as those indicated by the age specific death rates in 2016, then their average age at death would be 69.7 years if males and 70.1 years if females. Furthermore, females were expected to live longer than males by 2.2 years as seen in Annexure 1 and Annexure 2.

3.5 Negative binomial regression results

With a significant p-value at a 5% level of significance, the individuals' age (p-value<0.001), region (p-value<0.001) and marital status (p-value<0.001) can be concluded to have a significant association with their risk of mortality, as shown in Table 1. Hence, all the associated variables were considered in the fitted negative binomial regression model and the resulting results shown in Table 4.

Table 4: Binomial regression output

	Estimate	Standard Error	P-value	Expected count	95% Confidence Interval for Expected count	
					Upper	Lower
(Intercept)	9.721	.0509	<0.001	16658.799	15078.287	18404.981
Sex						
Female	.012	.0148	.407	1.012	.983	1.042
Male	0			1		
Region						
!Karas	-.020	.0530	.706	.980	.884	1.087
Erongo	-.025	.0478	.607	.976	.888	1.072
Hardap	-.019	.0493	.705	.981	.891	1.081
Kavango East	.001	.0427	.989	1.001	.920	1.088
Kavango West	.061	.0634	.334	1.063	.939	1.204
Khomas	.007	.0391	.865	1.007	.932	1.087
Kunene	.010	.0593	.860	1.011	.900	1.135
Ohangwena	-.005	.0456	.919	.995	.910	1.088
Omaheke	-.026	.0512	.614	.975	.881	1.077
Omusati	.041	.0433	.347	1.042	.957	1.134
Oshana	.002	.0395	.962	1.002	.927	1.083
Oshikoto	-.025	.0586	.671	.975	.870	1.094
Otjozondjupa	.021	.0476	.656	1.021	.930	1.121

Zambezi	0			1		
Marital Status						
Divorced	.009	.0983	.927	1.009	.832	1.224
Married	-.033	.0399	.404	.967	.895	1.046
Other	.009	.0443	.837	1.009	.925	1.101
Single	-.128	.0378	.001	.880	.817	.948
Unknown	.098	.1437	.495	1.103	.832	1.462
Widowed	0			1		
Age						
<1 year	-2.221	.0239	<0.001	.109	.104	.114
1-4 years	-1.392	.0365	<0.001	.249	.231	.267
5-14 years	-1.222	.0511	<0.001	.295	.267	.326
15-59 years	-.577	.0183	<0.001	.561	.542	.582
60+ years	0			1		
(Scale)	1					
(Negative binomial)	1					

From Table 4, it can be concluded that the expected death count for individuals who were single (p-value=0.001) was 0.880 less compared to the expected count for the widowed individuals. Furthermore, the expected death count for individuals who were aged 15-59 years (p-value<0.001), 5-14 years (p-value<0.001) and 1-4 years (p-value<0.001) were 0.561, 0.295 and 0.249 respectively less compared to the expected count for the individuals aged 60+ years. In addition, the expected death count for the less than 1-year individuals (p-value<0.001) was 0.109 less compared to the expected count for the individuals aged 60+ years as shown in Table 4. Moreover, from Table 4, the estimate of the fitted negative binomial regression model was 1 which was greater than 0, and therefore suggest a present of over-dispersion (i.e., the variance was greater than the mean).

4. Discussion

From this study, there were more males dying in 2016, compared to the females. This mortality differential is quite usual because according to Weeks (2014) males were more likely to suffer from conditions that were major causes of death such as chronic diseases, car accidents, etc., and another interpretation of the mortality difference was the biological superiority in the ability of females to survive longer than males (Weeks, 2014). Similar patterns of females living longer than males have been observed since the 1991 and 2001 census, which led to females reporting less deaths compared to males (NSA, 2014). In addition, Oshana, Kavango East, Khomas, Hardap and Omaheke regions had high mortality rates, while infants and elderly individuals had a high probability of dying. Khomas and Oshana regions recorded the highest of number of deaths due to the fact that the two regions house the main referral hospitals in the country and patients are transferred from their respective regions for better treatment because some hospitals in most regions lack the adequate medical equipment needed for the treatment of their patients. This situation also explains the results from the Kavango East region because of the Rundu Intermediate State Hospital being the only nearby referral hospital within the surrounding geographical areas. Also, according to NSA (2016), these regions house the highest population percentage, especially the age group 15-59 years, who are reportedly more susceptible to different causes of deaths such as HIV/AIDS, car accidents, suicide, homicide, and etc. Furthermore, the study revealed that individuals who were single and aged 15-59 and 5-14 years had less expected death count. This is not surprising as infants and elderly individuals are the most vulnerable and fragile to different types of illnesses and conditions that are major causes of death.

5. Conclusions and recommendations

In conclusion, mortality variation was examined by the socio-demographic factors such as age, sex, marital status and region, with the number of deaths differing for each sex and males having reported more deaths compared to females. Mortality also differed for each age group and regions, with high percentages among infants (<1 year), and older ages (60+ years), as well as in the Oshana, Kavango East, Khomas, Hardap and Omaheke regions. Furthermore, the study revealed that individuals who were single and aged 15-59 and 5-14 years had less expected death count. Hence, it is recommended that interventions (such as affordable and proper health care and well-being services) targeted at the (most) vulnerable age groups, marital group and regions be immediately made available, in order to meet the third goal under the health-related sustainable development goals of the United Nations. In addition, further research is needed focusing on the specific cause of death in the population and the quality of health care in Namibia.

Annexes

Annex 1 Abridged Life Table of males for Namibia 2016

Age x	Number of males	Number of deaths	Size of cohort	Death rate	Probability of dying	Number of survivors to age x	Number of deaths at age x	Number of years lived between lx-0 and lx	Total years lived between 0 & x	Life Expectancy
under 1	34417	1698	1	0.0493	0.0474	100 000	4735	95975	6787333	67.9
0 - 4	128363	486	4	0.0038	0.0150	95 265	1430	377628	6691357	70.2
5 -9	143495	132	5	0.0009	0.0046	93 835	431	468100	6313729	67.3
10-14	120339	105	5	0.0009	0.0044	93 405	407	466007	5845629	62.6
15-19	120327	148	5	0.0012	0.0061	92 998	570	463565	5379623	57.8
20-24	114754	357	5	0.0031	0.0154	92 428	1427	458573	4916058	53.2
25-29	102474	453	5	0.0044	0.0219	91 001	1989	450032	4457486	49.0
30-34	81978	595	5	0.0073	0.0356	89 012	3173	437127	4007453	45.0
35-39	68080	690	5	0.0101	0.0494	85 839	4242	418589	3570326	41.6
40-44	55781	705	5	0.0126	0.0613	81 597	4998	395487	3151737	38.6
45-49	42449	631	5	0.0149	0.0717	76 598	5489	369268	2756250	36.0
50-54	33595	531	5	0.0158	0.0760	71 109	5406	342030	2386982	33.6
55-59	24109	563	5	0.0234	0.1103	65 703	7248	310394	2044952	31.1
60+	59589	3381		0.0567	1.0000	58 455	58455	1030242	1734558	29.7

Annex 2 Abridged Life Table of Females for Namibia 2016

Age x	No. of females	No. of death s	Size of cohort	Death rate	Probabil ity of dying	No. of survivors to age x	No. of deaths at age x	No. of years lived between lx-0 and lx	Total years lived between 0 & x	Life Expectancy
under 1	33319	1528	1	0.0459	0.0441	100 000	4414	96248	7013012	70.1
1--4	125826	443	4	0.0035	0.0140	95 586	1335	379141	6916764	72.4
5--9	141152	97	5	0.0007	0.0034	94 251	323	470448	6537623	69.4
10--14	119284	90	5	0.0008	0.0038	93 928	354	468756	6067175	64.6
15-19	122491	140	5	0.0011	0.0057	93 574	533	466538	5598420	59.8
20-24	119344	231	5	0.0019	0.0096	93 041	896	462965	5131881	55.2
25-29	106322	302	5	0.0028	0.0141	92 145	1299	457476	4668916	50.7
30-34	86875	441	5	0.0051	0.0251	90 846	2277	448535	4211440	46.4
35-39	72053	421	5	0.0058	0.0288	88 569	2550	436467	3762905	42.5
40-44	60720	385	5	0.0063	0.0312	86 018	2684	423381	3326438	38.7
45-49	48349	412	5	0.0085	0.0417	83 334	3477	407978	2903057	34.8
50-54	40663	366	5	0.0090	0.0440	79 857	3515	390500	2495079	31.2
55-59	31964	404	5	0.0126	0.0613	76 343	4677	370021	2104579	27.6
60+	84470	3490		0.0413	1.0000	71 666	71666	1734558	1734558.116	24.2

References

Khan, A. &, & Khan, S. (2016). Factors effecting life expectancy in developed and developing countries of the world (An approach to available literature). International Journal of Yoga, Physiotherapy and Physical Education, 1(1), 4–6.

Namibia Statistics Agency (NSA). (2014). Namibia 2011 Census Mortality Report.

Namibia Statistics Agency (NSA). (2016). Namibia Inter-censal Demographic Survey 2016 Report Namibia Inter-censal Demographic Survey 2016 Report.

NCSS. (2021). Regression Analysis Software | Regression Tools | NCSS Software. https://www.ncss.com/software/ncss/regression-analysis-in-ncss/#NegativeBinomial

Oyedele, O.F & Sugnet Lubbe (2018). Partial least squares biplot for generalized linear models. International Journal of Statistics: Advances in Theory and Applications, 2(1), 1-20.

Republic of Namibia. (1994). 1991 Population and Housing Census.pdf. Cetral Statistics Office, NPC.

Statista. (2021). • Life expectancy in developed and developing countries | Statista. https://www.statista.com/statistics/274507/life-expectancy-in-industrial-and-developing-countries/#:~:text=In 2020%2C the average life,and 82 years for females.

Statistics South Africa. (2014). Mid-year Population estimates. Www.Statssa.Gov.Za/Publications/.../P03022014.Pdf, July, 1–18. https://doi.org/Statistical release P0302

United Nations. (2012). Changing Levels and Trends in Mortality : the role of patterns of death by cause. United Nations Publications, 1–81.

United Nations. (2019). World Mortality 2019. In World Mortality 2019. https://doi.org/10.18356/f6cceefe-en

Weeks, J. R. (2014). Population: An Introduction to Concepts and Issues. In Contemporary Sociology. Cengage Leaening.

Zwilling, M. (2013). Negative Binomial Regression. The Mathematica Journal, 15, 1–18. https://doi.org/10.3888/tmj.15-6

Namibian Journal for Research, Science and Technology
Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

A Logistic Regression Model to Assess Factors Influencing Schizophrenia Symptoms in Namibia

Nakunipa, E. N.N.[1], Pazvakawambwa, L.[1*], Iiyambo, P.[1]

[1]Department of Statistics and Population Studies, University of Namibia, Windhoek, Namibia
*Corresponding author: lpazvakawambwa@unam.na

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:
Schizophrenia, Logistic regression, Demographic and Health Survey, Namibia

ABSTRACT

Schizophrenia is a common mental disorder in Namibia and affects about 20 million people worldwide. Risk factors for schizophrenia in Namibia have not been extensively investigated. The objective of the study was to establish prevalence and factors influencing schizophrenia symptoms based on secondary data from the 2013 Namibia Demographic and Health Survey (NDHS). Descriptive statistics were computed to profile the background characteristics of the sample. Chi-square tests were conducted to assess association between socio-demographic variables and schizophrenia symptoms. Binary logistic regression was performed to establish determinants of schizophrenia symptoms. The prevalence of schizophrenia symptoms was 12.4% (13.6% among females and 11.0% among males). Regression results indicated that females (OR=1.159, 95% CI: 1.022-1.314, p=0.021) were more likely to have schizophrenia symptoms compared to their male counterparts. Those who resided in urban areas (OR=0.699, 95% CI: 0.609-0.803, p<0.001) were less likely to have schizophrenia symptoms compared to those who resided in rural areas. Those with no formal education (OR=0.378, 95% 0.273-0.523, p<0.001); those with primary education (OR=0.646, 95% CI: 0.501-0.834, p<0.001) and those with secondary education (OR=0.619, 95% CI: 0.495-0.775, p<0.001) were less likely to have schizophrenia symptoms compared to those with higher education. Results also showed that those who had never married (OR=0.275, 95% CI: 0.225-0.335, p<0.001); the married (OR=0.229, 95% CI: 0.184-0.284, p<0.001); and those living with a partner (OR=0.283, 95% CI: 0.225-0.355, p<0.001) were less likely to have schizophrenia symptoms compared to those who were on separation. Respondents who did not consume alcoholic drinks (OR=0.597, 95% CI: 0.526-0.677, p<0.001) were less likely to have schizophrenia symptoms compared to those who consumed alcoholic drinks. Schizophrenia symptoms were not significantly influenced by wealth index (p>0.05). There is need to step up gender-specific mental health programs especially in rural areas. Efforts to stabilize marital relationships at national level should be strengthened. Mental health could also be improved through drug abuse prevention and rehabilitation programs.

1. Introduction

Mental health is one of the key human rights (Dhaka, Musese, Kaxuxuena, Bakare, & Janik, 2017). The goal of the national policy for mental health is to achieve and maintain a high standard of mental health and wellbeing in the Namibian population MoHSS (2005). Mental health disorders are placed in discrete categories such as major-depressive disorder, bipolar disorder, schizophrenia and obsessive–compulsive disorder (OCD) based on theory and subjective symptoms (Adam, 2013). Schizophrenia is a common mental disorder in Namibia and affects about 20 million people worldwide. People with schizophrenia have 40% - 60% greater chances of dying than the general population

According to Srinivas, Neetha, Nair, Allencherry & Banerjee (2013), schizophrenia is a severe and disabling mental illness, that affects the brain, whereby a person sees or hears things that are actually not there. People suffering from schizophrenia struggle with differentiating between what is real and what is unreal, they find it difficult to think clearly and to behave normally. The delusions and hallucinations, despite obvious evidence that it is not true, clash with reality of life and are not understandable to others but are very real to the person experiencing them

(Smith, 2019).

A study done by Molina & Forastero (2015) was aimed at finding the percieved needs of schizophrenia patients and collecting enough data needed for effective therapies to improve the quality of life of patients. A qualitative analysis was done by interviewing 9 schizophrenia patients. Eight out of the nine participants believed that depression and anxiety are symptoms of schizophrenia and affects the quality of life negatively. This lead to an increase consumption of ciggarettes and junk food, among other negative things, which had a negative impact on their already poor health. Participants expressed the negative side effects that the antipsychotic medication had on their physical health. Furthermore, the participants indicated that social stigma and lack of social resources to create new social networks also affected their quality of life negatively. They concluded that anxiety and depression are the two factors that mostly affects the quality of life of schizophrenia patients negatively. They also found out that establishing social links was necessary.

Gibson et al. (2013) conducted a study aimed at contributing to the understanding of treatment choices and the support services schizophrenia patients need in order to maximise the benefits from their medication. The study was a mixed methods questionnaire, applying quantitative and qualitative analyses. Thirty five people diagnosed with schizophrenia and on psycho-phamarceutical treatment for schizophrenia answered online and telephone questions about whether, how and why they changed or stopped their treatment recommendation, and what support they currently have or would like to have. Over half of the participants said that they were non adherent, however, when they were asked about intentional and unintentional adherence, 77% reported deviating from treatment recommendation. Alarmingly, 29% were non-adherent and were satisfied with being so. The participants' statisfaction with their support was positively correlated with their satisfaction with their medication. The study suggests that non-adherence, either intentional or unintentional is common amongst people diagnosed with schizophrenia and that it often occurs without health professionals' knowledge or support. The access to more information and emotional support, could help patients make treatment choices that will refelct the long term risks of non-adherence.

People with schizophrenia face many emotional, mental and physical challenges. The quality of life of schizophrenia patients is mostly characterized by depression and annxiety (Molina & Forastero, 2015). Buizza, Schulze, Bertocchi, Rossi, Ghilardi & Pioli (2007) articulated that, the living condition of people with schizophrenia depends on how severe the illness is, and on their accaptence in the communities. Schizophrenia patients and their relatives faces stigmazation and this often hinders the treatment process. With proper treatment, people with schizophrenia can lead productive lives, though some patients have to live with the symptoms for the rest of their lives. Schizophrenia treatment includes therapy and medication.

Factors influencing schizophrenia can be genetic, environmental, psychological or drug abuse related. John, Deshpande, Nimgaonkar, & Thelma (2013) states that schizophrenia is a weakening neuropsychiatric disorder that has 80% chances of being inherited. Stress during pregnancy or at a later stage of development are major environmental factors and socioeconomic status are also enviromental factors. Psychological factors include physical or emotional abuse as a child and any traumatic experience. The overuse of drugs is also a contributing factor of schizophrenia. Buadze, Stohler, Schulze & Liebrenz (2010) investigated if the use of cannabis causes schizophrenia or if the use of cannabis by schizophrenia patients have any influence on the relapses. Schizophrenia patients were interviewed, and the results indicated that none of the interviewed patients, described a link between cannabis and schizophrenia.

Jajodia, Baghel, Kaur, Jain, & Kukreti (2013) evaluated the association between the genetics diverseness of the neurodevelopment gene and the risk of schizophrenia. The study consisted of 482 schizophrenia cases and 401 age, sex matched controls. Genotypic tests were done, and multivariate logistic regression was used to analyze the data. The study reported that, there was an association between a gene named SLC1A3 (15p13) and the risk of schizophrenia development.

Burns, Tomita, & Kapadia (2013) performed a multilevel mixed-effects Poisson regression to investigate the relationship between Gini coefficients and incidence rates of schizophrenia controlling for covariates. The results from the 26-country systematic review showed that there was a significant positive relationship between incidence rate of schizophrenia and Gini coefficient. Countries that were characterized by a large rich-poor gap were at increased risk of schizophrenia and measures of income inequality were associated with schizophrenia incidence.

Nyer, et al. (2010) established that married or cohabitating participants who were 40 or older had a later age of onset of first psychotic episode than those who were single. Furthermore, single participants rated their quality of life lower than those who were married. Married participants had less suicidal ideation than those who were divorced, widowed or separated. They concluded that, in middle-aged and older individuals with schizophrenia, marriage appeared to enhance quality of life.

Chun-Tea et al. (2018) examined the relationship between low-income and schizophrenia among the Taiwanese population based on a Cox proportional hazard regression analysis. The prevelance of schizophrenia was 1.23% in low income and 0.26% in non low income individuals. Evidence also showed that higher incidence rates were present in the 18-64 age category of lower income individuals.

Correlation analysis and binary logistic regression were used to examine the effect of age of onset of schizophrenia spectrum disorders on demographic and clinical variables. The age of onset had a significant relationship with the cignitive component of the Positive and Negative Syndrome Scale (PANSS) and Barratt Impulsive Scale (BIS) score. They concluded that, age of onset influenced illness course in patients with schizophrenia spectrum disorders (Yu-Chen & Yia-Ping, 2010).

According to Ashipala, Wilkinson, & Van Dyk (2016), mental health makes up five of the ten leading causes of health disability and out of the 3.1% disability rate in Namibia, 15% (7360) consisted of people registered as living with mental health problems.The 2013 Namibia Demographic and Health Survey (NDHS) reported that, schizophrenia is the common disorder of mental health in Namibia, followed by depression. The number of patients with schizophrenia increases every year.

Despite the increase in the number of patients with mental disorders, specifically schizophrenia, a study by Dhaka et al. (2017) reported that, mental health received low priority because of limited resources which are often directed to communicable and life-threatening diseases. Sankoh, Sevalie and Weston (2018) indicated that out of all clinical trials conducted in low income and middle-income countries in Africa, only 3% were of mental health.

NDHS (2013) reported that there has been an increase in mental health problems in Namibia, with schizophrenia being the common disorder. Even so, there is a lack of understanding of mental health problems among the public (WHO, 2017). Moreover, previous studies have shown that Namibia lacks scientific data collection systems, thus there was a challenge on epidemiological data on mental health (Ndjaleka, 2017). Therefore, the factors of schizophrenia in Namibia have not been extensively investigated. Thus, the study aimed at determining the factors that are associated with schizophrenia in Namibia.

The findings of this study are vital to the policy makers because identifying the factors associated with schizophrenia in Namibia will help them implement policies that are geared towards preventing schizophrenia. Moreover, knowing the factors would be essential to achieve the Sustainable Development Goal 3, which is designed to promote mental health and well-being. The objective of the study was to establish prevalence and factors associated with schizophrenia in Namibia.

2. Methods

The study was based on secondary data from the 2013 Namibia Demographic and Health Survey (NDHS) with 9,519 respondents. The sample for the 2013 NDHS was a two-stage stratified sample. In the first stage, 554 enumeration areas (EA) were selected with a stratified probability proportion to size selection from the sampling frame, of which, 269 in urban areas and 285 in rural areas. All the 13 regions were divided into urban and rural areas, thus every region was stratified into 26 sampling strata (13 urban strata and 13 rural strata). In the second stage, in every urban and rural cluster, a fixed number of 20 households were selected according to equal probability systematic sampling (NDHS Report, 2013).

The outcome variable of the study was based on whether respondents had symptoms of schizophrenia, thus, a respondent responding "Yes" to the question, *"Have you ever seen or heard things that are actually not there?" (Yes, No)*, was assumed as having symptoms of schizophrenia. The independent variables of the study were: age (15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50+) in years; gender (Male, Female); place of residence (Urban, Rural); employment Status (Yes, No); marital status (Never in a union, Married, Living with partner, Widowed, Divorced, No longer living together/separated); level of education (No education, Primary, Secondary, Higher); wealth quintile (Poorest, Poorer, Middle, Richer, Richest); and consumption of alcoholic drinks (Yes, No).

The Statistical Packages for Social Sciences (SPSS) version 26 was used to analyze the data. Firstly, descriptive statistics were calculated to describe various characteristics of the respondents, Chi-Square tests of association were used to analyze if there were any associations between the socio-economic & demographic factors and schizophrenia symptoms. The binary logistic regression was then used to establish factors influencing schizophrenia symptoms. Logistic regression allows categorical and continuous variables to predict categorical response variable (Jason, 2008; McDonald, 2008). Suppose that, there is a sample of n independent observations of the pair(x_i, y_i), $i = 1,2, \dots, n$, where x_i is the value of the independent variable, and y_i is the value of a dichotomous outcome variable for the ith subject. Let Y the outcome variable be coded as 0 and 1, where $Y = 1$ represents the presence of the outcome variable (schizophrenia symptoms) and $Y = 0$ represents the absence of the outcome variable (schizophrenia symptoms). Then, $Y \sim \text{Bernoulli}(\pi_i)$ where π_i is the probability of the observed set of data, where, $p(Y=1)=\pi$ and $p(Y=0)=1-\pi$.

An appropriate model is the logistic model, with the likelihood function given by:

$$f(y_i,\pi_i)=\prod_{i=1}^{n}\pi_i{}^{y_i}(1-\pi_i)^{1-y_i},$$

where $y_i=0,1$. The mean value of the outcome variable, given the value of the independent variable is the key quantity in any regression problem. This quantity can be expressed as $E(Y|x)$ and is called the conditional mean (Hosmer Jr, Lemeshow, & Sturdivant, 2013). Let $\pi=E(Y|x)$, the specific form of the logistic regression model is:

$$\pi=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

A transformation of $\pi(x)$ is the logit transformation and it is defined in terms of π as:

$$logit(\pi)=log\left[\frac{\pi}{1-\pi}\right]=\beta_0+\beta_1 x,$$

where, β_0 is the intercept and β_1 is the coefficient of x

Equivalently, consider a collection of p independent variables denoted by a vector $x=\begin{bmatrix}x_1\\ \vdots\\ x_p\end{bmatrix}$. The logit transformation of the multiple logistic regression is given by:

$$logit(\pi_i)=log\left[\frac{\pi_i}{1-\pi_i}\right]=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p.$$

And the multiple logistic regression model is given by:

$$\pi_i=\frac{e^{\beta_0+\beta_1x_1+\cdots\beta_px_p}}{1+e^{\beta_0+\beta_1x_1+\cdots\beta_px_p}}$$

Fitting a multiple logistic model requires that we obtain values of the vector $\beta=\begin{bmatrix}\beta_0\\ \vdots\\ \beta_p\end{bmatrix}$.

Regression coefficients are interpreted using odds ratio as a measure of association. The odds ratio denoted OR is given by e^{β} (Hosmer Jr et al., 2013). If OR is equal to 1, then the increase in the independent variable has no influence on the probability of the response variable, OR greater than 1 indicates an increase in the probability of the outcome variable to occur and OR less than 1 indicates a decrease in the probability of the outcome variable to occur.

3. Results

3.1 Descriptive Statistics

The sample consisted of 9519 respondents. Table 1 shows that, 53% were females and 47% were males. Most of the respondents were aged between 20 and 24 with 20.9% and the least were between the ages 45-49 with 4.3%. About 53% of the respondents were from rural areas, whereas, 47.4% were from urban areas. Furthermore, 59.4% of the respondents had secondary education as their highest level of education and the least had higher education with 6%, and most of them (22.4%) were from middle class, in terms of the wealth index, whereas, the least which is 16.4% where from the richest class. Furthermore, the majority (51.5%) were never in a union, whereas, 0.7% were widowed. Fifty one percent of the respondents were working and 52.5% consumed alcohol. Out of 9519 respondents, (1176) 12.4% indicated having experienced schizophrenia symptoms.

Table 1: Socio-demographic information of the respondents

Characteristics		Number of respondents	Percentage (%)
Sex	Male	5043	53
	Female	4476	47
Age	15-19	1201	12.8
	20-24	1993	20.9
	25-29	1944	20.4
	30-34	1498	15.7
	35-39	1182	12.4
	40-44	760	8
	45-49	410	4.3
	50+	531	5.6
Type of residence	Urban	4508	47.4
	Rural	5011	52.6
Higher education level	No education	917	9.6
	Primary	2374	24.9
	Secondary	5657	59.4
	Higher	571	6
Wealth index	Poorest	1758	18.5
	Poorer	1978	20.8
	Middle	2129	22.4
	Richer	2095	22
	Richest	1559	16.4

Marital status	Never in union	4905	51.5
	Married	2120	22.3
	Living with partner	2045	21.5
	Widowed	69	0.7
	Divorced	91	1
	Separated	289	3
Is the respondent currently working?	No	4878	51.2
	Yes	4641	48.8
Does the respondent Consume alcoholic drinks	No	4525	47.5
	Yes	4994	52.5
Have you ever seen/heard things that are not there?	No	8343	87.6
	Yes	9519	12.4

3.2. Results of Chi-Square Tests of Association between socio-economic variables symptoms of schizophrenia

The results of Chi-square tests of association at 5% level of significance are summarized in Table 2. Schizophrenia symptoms were significantly associated with sex (Chi-square with (Chi-square = 14.959, p value < 0.001); type of place of residence (Chi-square = 15.411, p value < 0.001); highest educational level attained Chi-square = 27.956, p value < 0.001); wealth index Chi-square =39.123, p value < 0.001); marital status Chi-square = 28.057, p value < 0.001); and consumption of alcoholic drinks Chi-square = 56.043, p value < 0.001). However, there was no significant association between schizophrenia symptoms and respondent age (p=0.425) and working status (p=0.508). Therefore, the variables respondents' age and respondents' working status did not qualify to be input into the binary logistic regression model to establish the determinants of schizophrenia symptoms.

Table 2. Results of Chi-Square Tests of Association between socio-economic variables symptoms of schizophrenia

Independent Variable	Have you ever seen/heard things that are not there? No (%)	Yes (%)	Chi-Square Statistic	p-value
Sex				
Females	86.4	13.6	14.959	<0.001
Males	89.0	11.0		
Age in 5-year groups				
15-19	88.8	11.2	7.033	0.425
20-24	87.7	12.3		
25-29	87.4	12.6		
30-34	87.2	12.8		
35-39	86.6	13.4		
40-44	86.8	13.2		
45-49	87.8	12.2		
50+	90.4	9.6		
Type of Place of Residence				
Urban	89.0	11.0	15.411	<0.001
Rural	86.4	13.6		
Highest Educational Level				
No Education	91.2	8.8	27.956	<0.001
Primary	85.8	14.2		
Secondary	87.4	12.6		
Higher	91.9	8.1		
Wealth Index				
Poorest	85.7	14.3	39.123	<0.001
Poorer	87.4	12.6		
Middle	86.3	13.7		
Richer	87.5	12.5		
Richest	92.2	7.8		
Marital Status				
Never in Union	87.6	12.4	28.057	<0.001
Married	89.7	10.3		
Living with partner	87.2	12.8		
Widowed	76.8	23.2		
Divorced	83.5	16.5		
Separated	81.3	18.7		
Respondent Currently working?				
No	87.9	12.1	0.437	0.508

Yes	87.5	12.5		
Does the respondent consume alcoholic drinks?				
No	90.3	9.7	56.043	**<0.001**
Yes	85.2	14.6		

3.3. Results of Binary Logistic regression to establish the determinants of schizophrenia symptoms.

The results of binary logistic regression to establish the determinants of schizophrenia symptoms are presented in Table 3. The results indicated that females (OR=1.159, 95% CI: 1.022-1.314, p=0.021) were more likely to have schizophrenia symptoms compared to their male counterparts. Those who resided in urban areas (OR=0.699, 95% CI: 0.609-0.803, p<0.001) were less likely to have schizophrenia symptoms compared to those who resided in rural areas. Those with no formal education (OR=0.378, 95% 0.273-0.523, p<0.001) those with primary education (OR=0.646, 95% CI: 0.501-0.834, p<0.001) and those with secondary education (OR=0.619, 95% CI: 0.495-0.775, p<0.001) were less likely to have schizophrenia symptoms compared to those with higher education. Results also showed that those who had never married (OR=0.275, 95% CI: 0.225-0.335, p<0.001); the married (OR=0.229, 95% CI: 0.184-0.284, p<0.001); and those living with a partner (OR=0.283, 95% CI: 0.225-0.355, p<0.001) were less likely to have schizophrenia symptoms compared to those who were on separation. Respondents who did not consume alcoholic drinks (OR=0.597, 95% CI: 0.526-0.677, p<0.001) were less likely to have schizophrenia symptoms compared to those who consumed alcoholic drinks. Schizophrenia symptoms were not significantly influenced by wealth index (p>0.05).

Table 3: Results of Binary Logistic regression to establish the determinants of schizophrenia symptoms

			95% Confidence Interval for Odds Ratio	
Variable	p-value	Odds Ratio	Lower	Upper
Sex				
Females	**0.021**	1.159	1.022	1.314
Males		1.000		
Type of Place of Residence				
Urban	**<0.001**	0.699	0.609	0.803
Rural (Ref)		1.000		
Highest Educational Level				
No Education	**<0.001**	0.378	0.273	0.523
Primary	**0.001**	0.646	0.501	0.834
Secondary	**<0.001**	0.619	0.495	0.775
Higher (Ref)		1.000		
Wealth Index				
Poorest	0.291	1.148	0.889	1.482
Poorer	0.520	1.082	0.852	1.374
Middle	0.116	1.197	0.956	1.497
Richer	0.241	1.135	0.919	1.402
Richest (Ref)		1.000		
Marital Status				
Never in Union	**<0.001**	0.275	0.225	0.335
Married	**<0.001**	0.229	0.184	0.284
Living with partner	**<0.001**	0.283	0.225	0.355
Widowed	0.076	0.583	0.322	1.057
Divorced	**0.003**	0.407	0.226	0.731
Separated (Ref)		1.000		
Does the respondent consume alcoholic drinks?				
No	**<0.001**	0.597	0.526	0.677
Yes		1.000		

4. Discussion

Results indicated that schizophrenia symptoms were influenced by gender, place of residence, level of education, marital status and alcohol consumption. The socioeconomic status (measured by wealth index) did not influence schizophrenia symptoms.

The study found that, in Namibia, women had a higher chance of having schizophrenia symptom compared to men in 2013. However, these findings contradict Chun-Tea, et al. (2018) who found that schizophrenia was common in men than in women. Riecher-Rossler, Butler & Karlkani (2018), also reported gender disparities schizophrenic psychoses citing differences in terms of symptomatology, comorbidity and neuro cognition. On the other hand, when focusing on functional rather than symptomatic outcomes of schizophrenia, women outperformed men in terms of educational achievement occupational functioning and interpersonal functioning (Seeman, 2019).

Respondents that were living in rural areas were found to be more likely to show schizophrenia symptoms compared to those who were living in urban areas. The results were in agreement with Solmi, Dykxhoorn, and Kirkbride (2017) who found that the most consistent evidence of rural-urban gradients in mental health risks existed in schizophrenia and suicide, with more mixed evidence in relation to common mental disorders.

Respondents with no formal education, those with primary education and those with secondary education were less likely to have schizophrenia symptoms compared to those with higher education. Improving educational level did not decrease the risk of schizophrenia as suggested by Luo, Pang, Zhao, Guo, Zhang & Zheng (2020).

The study found that there was no relationship between income (wealth) and showing or not showing of schizophrenia symptoms. This was in contrast with the findings from the study done by Burns, Tomita, & Kapadia (2013), which concluded that income was associated with schizophrenia. Hudson (2005) found that socioeconomic status impacted directly on the rates of mental illness as well as indirectly through the impact of economic hardship on low- and medium-income groups in the US. In China, Ran, Huang, Mao, Lin, Li, & Chan (2017) also established that low family economic status is a predictive factor of poor long-term outcomes of persons with schizophrenia and recommended that individual's family economic status should be taken into account when making mental health policy and providing community-based mental health services.

Results also showed that those who had never married; the married, and those living with a partner were less likely to have schizophrenia symptoms compared to those who were on separation. In rural China, being married was predictive of more favorable 14-year outcomes of persons with schizophrenia probably due to the fact that marriage can be instrumental for enhancing family-based support and caregiving, as well as improving the community tenure of persons with schizophrenia. The authors stressed the importance of developing programs to enhance opportunity for persons with schizophrenia to get and stay married (Ran, Wong, Yang, Ho, Mao, Li, & Chan, 2017). Negative symptoms, such as depression and suicidal thoughts were higher in schizophrenic patients who were not married compared to those who were married. Those married or cohabiting had a later age of onset of first psychotic episode or hospitalization compared to those who were single Nyer, Kasckow, Fellows, Lawrence, Golshan, Solorzano & Zisook (2010).

Respondents who consumed alcoholic drinks had higher chances of showing schizophrenia symptoms compared to those who did not. Buadze et al., (2010) concluded that clinicians should not let patients who do not have a link with cannabis use go through a treatment meant for patients that have a link with cannabis.

5. Conclusions

Results indicated that females were more likely to have schizophrenia symptoms compared to their male counterparts. Those who resided in urban areas were less likely to have schizophrenia symptoms compared to those who resided in rural areas. Those with no formal education, primary education and secondary education were less likely to have schizophrenia symptoms compared to those with higher education. Results also showed that those who had never married, the married and those living with a partner were less likely to have schizophrenia symptoms compared to those who were on separation. Respondents who did not drink alcohol were less likely to have schizophrenia symptoms compared to those who consumed alcoholic drinks. Schizophrenia symptoms were not significantly influenced by socio-economic status.

6. Recommendations

There is need to step up gender-specific mental health programs especially in rural areas. Efforts to stabilize marital relationships at national level should be strengthened. Mental health could also improve through drug abuse prevention and rehabilitation programs.

References

Ashipala, D. O., Wilkinson, W., & Van Dyk, A. (2016). Mental health policy impllementation as an intergral part of primary health care services in Oshana region, Namibia. *Journal of Nursing Education and Practice*, 55.

Buadze, A., Stohler, R., Schulze, B., & Liebrenz, M. (2010). Do patients think cannabis causes schizophrenia? - A qualitative study on the casual beliefs of cannabis using patients with schizophrenia. *Harm Reduction Journal*, 1-8.

Buizza, C., Schulze, B., Bertocchi, E., Rossi, G., Ghilardi, A., & Pioli, R. (2007). The stigma of schizophrenia from patients' and relatives' view: A pilot study in an Italian rehabilitation residential care unit. *BioMed Central*, 1-8.

Burns, J., Tomita, A., & Kapadia, A. (2013). Income inequality and schizophrenia: Increased schizophrenia incidence in countries with high levels of income inequality. *The International journal of social psychiatry*.

Chun-Tea, L., Chi-Yueh, H., Jia-Fu, L., Yi-Chyan, C., Oswald Ndi, N., Jing-Yang, H., et al. (2018). Relationship between Schizophrenia and Low-Income BAsed on Age and Sex: Results from a Nation-wide Population-Based Longitudinal Study. *NUEROPSYCHIATRY*.

Dhaka, P., Musese, A., Kaxuxuena, T., Bakare, K., & Janik, M. (2017). Mental Health Services in Namibia: Challenges and Prospects. *International Journal of Public Mental Health and Neurosciences*, 1-6.

Gibson, S., Sarah Brand, S. B., Boden, Z., & Benson, O. (2013). Understanding treatment non-adherence in schizophrenia and bipolar disoder: a survey of what survey users do and why. *BMC Psychiatry*, 1-12.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression.* Canada: John Wiley & Sons, Inc.

Jajodia, A., Baghel, H. K., Kaur, G., Jain, S., & Kukreti, R. (2013). Glumate Transporter Genes are Associated with Schizophrenia in South Indian Population. *Molecular Cytogenetics*, 1-5.

Jason, E. K. (2008). *Best Practices in Quantitative Methods.* California: Sage Publications Ltd.

John, J., Deshpande, S., Nimgaonkar, V., & Thelma, B. (2013). Role of miRNA Binding Sites SNPs in Candidate Genes in a North Indian Schizophrenia cohort. *Molecular Cytogenetics*, 1-3.

McDonald, H. J. (2008, July 20). *Handbook of Biological Statistics.* Delaware: Lulu.com.

MoHSS, M. o. (2005). *National Policy for Mental Health.* Windhoek: Ministry of Health and Social Services.

Molina, A. M., & Forastero, M. C. (2015). Assessing emotional health needs of patients with schizophrenia in order to apply more effetive therapies to improve their quality of life. *BMC Nursing*, 1-3.

NDHS. (2013). *Namibia Demographic and Health Survey.* Windhoek: MoHSS, NSA.

Ndjaleka, H. (2017). The state of Phychology in Namibia. *American Pysychological Association*, 1-3.

Nyer, M., Kasckow, J., Fellows, I., Lawrence, E., Golshan, S., Solorzano, E., et al. (2010). The relationship of marital status and clinical characteristics in middle-aged and older patients with schizophrenia and depressive symptoms.

Rakesh, R., & Richa, S. (2015). Chi-square test and it's application in hypothesis testing. *Journal of the Practice of Cardiovascular sciences*, 69-71.

Sankoh, O., Sevalie, S., & Weston, M. (2018). Mental health in Africa. *The Lancent Global Health*, 1-2.

Smith, K. (2019, July 17). *PSYCOM*. Retrieved March 3, 2020, from Psycom.net: https://www.psycom.net/schizophrenia-hallucinations-delusions/

Srinivas, L., Neetha, N., Nair, C., Allencherry, P., & Banerjee, M. (2013). Association of Pro-Inflammatory Cytokine Gene Polymorphisms with Schizophrenia in South Indian Population. *Molecular Cytogenetics*, 1-5.

WHO (2019, October 4). *World Health Organisation*. Retrieved November 21, 2020, from World Health Organisation: http://www.who.int

WHO. (2017). Mental health, human rights & legislation: A global human rights emergency in mental health. *World Health Organization.* https://www.who.int/mental_health/policy/legislation/en/. *Accessed, 9 December 2020.*

Yu-Chen, K., & Yia-Ping, L. (2010). Effects of age of onset on clinical characteristics in schizophrenia spectrum disorders. *BMC Psychiatry*, 1-11.

Adam, D. (2013). Mental health: On the spectrum. *Nature News*, *496*(7446), 416.

Riecher-Rössler, A., Butler, S., & Kulkarni, J. (2018). Sex and gender differences in schizophrenic psychoses—a critical review. *Archives of women's mental health*, *21*(6), 627-648.

Seeman, M. V. (2019). Does gender influence outcome in schizophrenia?. *Psychiatric Quarterly*, *90*(1), 173-184.

Solmi F., Dykxhoorn J., Kirkbride J.B. (2017) Urban-Rural Differences in Major Mental Health Conditions. In: Okkels N., Kristiansen C., Munk-Jørgensen P. (eds) Mental Health and Illness in the City. Mental Health and Illness Worldwide. Springer, Singapore. https://doi.org/10.1007/978-981-10-2327-9_7.

Luo, Y., Pang, L., Zhao, Y., Guo, C., Zhang, L., & Zheng, X. (2020). Gender difference in the association between education and schizophrenia in Chinese adults. *BMC psychiatry*, *20*(1), 1-7.

Ran, M. S., Yang, L. H., Liu, Y. J., Huang, D., Mao, W. J., Lin, F. R., ... & Chan, C. L. W. (2017). The family economic status and outcome of people with schizophrenia in Xinjin, Chengdu, China: 14-year follow-up study. *International Journal of Social Psychiatry*, *63*(3), 203-211.

Hudson, C. G. (2005). Socioeconomic status and mental illness: tests of the social causation and selection hypotheses. *American journal of Orthopsychiatry*, *75*(1), 3-18.

Nyer, M., Kasckow, J., Fellows, I., Lawrence, E. C., Golshan, S., Solorzano, E., & Zisook, S. (2010). The relationship of marital status and clinical characteristics in middle-aged and older patients with schizophrenia and depressive symptoms. *Ann Clin Psychiatry*, *22*(3), 172-179.

Ran, M. S., Wong, Y. L. I., Yang, S. Y., Ho, P. S., Mao, W. J., Li, J., & Chan, C. L. W. (2017). Marriage and outcomes of people with schizophrenia in rural China: 14-year follow-up study. *Schizophrenia research*, *182*, 49-54.

Namibian Journal for Research, Science and Technology

Volume 4 Issue 1
May 2022
P-ISSN: 2026-8548

Original Research Article

An ecological adjusted random effect model for property crime in Windhoek, Namibia (2011-2016)

J. Amunyela[1], I. Neema[2], L. Pazvakawambwa[3*]

[1]Department of Mathematics & Statistics, Namibia University of Science and Technology, Windhoek, Namibia
[2]Namibia Statistics Agency, Windhoek, Namibia
[3]Department of Computing, Mathematical & Statistical Sciences, University of Namibia, Windhoek, Namibia

*Corresponding author: lpazvakawambwa@unam.na

ARTICLE INFO

Received: March 2021
Accepted: May 2022

Keywords:
Crime intensity; Random effects; Windhoek; Zero-inflated; ZINB-GLMM

ABSTRACT

Count data that are zero inflated are often analysed using Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) when observations are correlated in ways that require random effects. This study investigated ecological factors influencing the number of property crimes in Windhoek by using data obtained from the Windhoek police over the period of six consecutive years (2011 to 2016). The ecological concepts were measured at different levels of aggregation. Limited studies in Windhoek have considered analysing crime data on Generalized Linear Mixed Model via Template Model Builder (TMB) R-package. Crimes were counted with respect to Month, Season, Year, Location and Density. Property crime data contained more zeros than expected. When comparing models fitted, it was found that the Relative Risks (RR) were highly significant for models fitted via Negative Binomial distribution. By adopting a ZINB-GLMM, the study attempted to address the potential covariates for Property crimes. The study showed that most of the variation property crimes was due to locations. Crime was high during spring and winter time during the study period. The study further discovered that areas with high population densities had high crime intensity. Security patrols and surveillance should be stepped up in Windhoek in high density suburbs especially during winter and spring seasons.

1. Introduction

Crime, no matter how small, has an enormous negative impact on the people’s lives. A rise in crime rate can negatively affect a country’s economy as it discourages investors and pushed skilled workers elsewhere. Despite all the interventions and procedures put into place by policy makers, governments, and non-profit organisations to curb crime in Namibia, the crime rate continues to increase unabated, especially in the capital city, Windhoek.

An individual is less likely to be involved in criminal activities when there are substantial rewards and when he or she enjoys respect from the society in which they belong. In addition, if a young person is gainfully engaged, either in education or employment, he or she is less likely to turn to criminal activities (Khan, Ahmed, Nawaz, & Zaman, 2015). In most European and African countries population density was found to be appropriate in crime prediction (Hanley, Lewis, & Ribeiro, 2016). Tonry (2014) argued that the rate for property crimes rose recently in all wealthy Western countries. Property crimes in Western countries were declining as a result of improved security technologies in motor vehicles, residences, and retail stores (Tonry, 2014). This argument is also supported by the collaborative online database (NUMBEO, 2018). The NUMBEO crime index indicates that in the Caribbean alone, the average crime index is currently standing at 71% while the safety index is only at 29%.

Crime has negatively affected several African countries and this led to the identification of three main phases of crime due to tangible shifts in the prevailing social, political and economic environment (Shaw & Reitano, 2013). Crime and insecurity are major challenges in African countries, threats to national development and individual quality of life (Wambua, 2015). In addition to that, only 11 African countries ranked in the top hundred countries worldwide in terms of safety and security with Benin being the top-ranked African country at number 50 (Legatum Institute, 2014). Subsequently, 38% of Africans say they have felt unsafe walking in the neighbourhoods. Study by Asongu and Kodila-Tedika (2016) found that in most African countries the wave of crime could be addressed if the fight against corruption is taken seriously by governments.

Drug trading, kidnapping, embezzlement, the large scale theft of minerals, or other plainly criminal activities exist in many parts of Africa (Ellis & Shaw, 2015). Palmary, Rauch, and Simpson (2014) contend that Johannesburg is believed to be the "crime capital" of South Africa. Besides that, a total of 267 arrests related to the illegal rhino horn trade were made in South Africa in 2012. This was linked to the arrest figures of 165 for 2010 and 232 for 2011 (Ayling, 2013). It was found that the illegal wildlife networks operating in South Africa and Namibia had no one distinct profile but both poachers and traffickers of rhino horn tended to be informal groups or predominantly individuals (Ayling, 2013).

Namibia is situated in Sub-Saharan Africa, a region that has one of the highest crime rates in the world (Neema & Böhning, 2012). According to the Overseas Security Advisory Council (Council, 2010), Namibians have regularly fallen victim to street crime.

Motor vehicle theft remains a major concern in Namibia. This type of crime usually involves smash-and-grab patterns and is sometimes associated with violence, especially when the occupants in the vehicle refuse to freely surrender their belongings to the perpetrators. Notably, Windhoek City Police (WCPOLS, 2006), observed that ATM card skimming, purse-snatching, vehicle breaks-ins and vehicle theft are among the most frequently recorded Property crime types in Namibia.

The introduction of Operation Kalahari and the installation of CCTV cameras was aimed at crime reduction and prevention in Windhoek. The Operation Kalahari and its predecessor Operation Horncranz involved Namibian Police force, Windhoek City Police and Namibian Defence Force working together to achieve a common goal. Personal robberies and residential break-ins and thefts remain prevalent in Namibia as well.

According to City of Windhoek (CoW), property crime's goal is to obtain money, property, or any other benefit. This may involve force, or the threat of force, in cases like robbery or extortion. This category includes, among other crimes, burglary, larceny, motor vehicle, arson, shoplifting and vandalism. High rates of Property crime were recorded recently at 74.06% in Windhoek (NUMBEO, 2018).

Scholarly opinion within the geography of crime and spatial criminology studies concurs that crime is highly concentrated in certain areas due to specific factors within those areas (Breetzke & Pearson, 2014; de Melo, Matias, & Andresen, 2015). Moreover, spatial patterns of these concentrations differ across crime types. Among youth, factors that influence crime are unemployment and lack of education (Dore, 2013), own house mortgage (Jones & Pridemore, 2012), crime-specific detection rate and prison population (Han, Bandyopadhyay, & Bhattacharya, 2013).

The basic model for count data like the number of property crimes is the Poisson model but most data does not satisfy the model assumptions. The study by Schielzeth and Nakagawa (2013), established that if the assumption made by random effects models are correct, then random effect would be the preferred choice because of its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at the high level.

In contrary, findings by Brooks et al. (2017a) were that count data can be analysed using a generalized linear mixed model when observations are correlated in ways that require random effects. They further claimed that attempting to fit the GLMM models via a **glmmTMB** package with a log Normal-poisson model and covariate-dependent zero-inflation led to convergence failure, and hence they substituted with a similar model (a Negative-binomial model). A comparison of the results showed that the deterministic methods (gam, glmmTMB and inla) were all fast; gam was fastest, because gam fitted a simpler model. While on the other hand the stochastic methods (**MCMCglmm** and **brm**) were about an order of magnitude slower (Brooks et al., 2017a).

The Generalized Linear Mixed Model is of significance. GLMM include a random effect model which may be considered in this study based on its ability to estimate covariates both within and between clusters, its ability to partition variance at multiple levels, its ability to examine variation in effects across cluster and it is parsimonious. For completeness, the Poisson regression, Negative Binomial (NB) model and the Zero-Inflated Negative Binomial (ZINB) model were explored. These models were worth being explored since they are widely used on count data and produce reliable results (Harrison, 2014).

Bolker et al. (2009) acknowledge that count data with so many zero values cannot be made normal by transformation. Even when one succeeds to transform the data, this transformed data might violate some statistical assumptions or limit the scope of inference (one cannot extrapolate estimates of fixed effects to new groups). GLMMs combine the properties of two statistical frameworks that are widely used in ecology and evolution study. These are the linear mixed models (which incorporate random effects) and generalized linear models (which handle non-normal data by using link functions and exponential family e.g normal, Poisson or binomial distributions) (Bolker et al., 2009). Bolker et al. (2009) further concluded that GLMMs are the best tool for analysing non-normal data that involve random effects.

2. Zero Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM)

Count data has been analysed using generalized linear mixed models especially when observations are correlated in a way that require random effects. However, count data are often zero-inflated, containing more zeros than would be expected from the typical error distributions used in GLMMs (Brooks et al., 2017b). For example, crime counts may be exactly zero for some months based on effective policing but may vary according to the negative binomial distribution for an area with poor policing. In addition, the Zero-Inflation model which is part of ZINB-GLMM estimates the probability of an extra zero such that a positive contrast indicates a higher chance of absence (e.g. $\beta_i < 0$ means fewer absences in a variable i unaffected by the fixed variable) while $\beta_i > 0$ means higher abundances in a variable i unaffected by the fixed variable.

Bolker (2018); Brooks et al. (2017b) established a new R package, generalized linear mixed model Template Model Builder ($glmmTMB$), that increases the range of models that can easily be fitted to count data using maximum likelihood estimation. The interface is simply developed to be familiar to users of the $lme4$ R package, a widely used tool for fitting GLMMs. All one must do, in principle, is to specify a distribution, link function, and structure of the random effects (Bolker et al., 2009).

The Template Model Builder (TMB) was known for maximising speed and flexibility through utilising automatic differentiation to estimate model slopes and the Laplace approximation for handling random effects. The strength of the R package $glmmTMB$ lies on the number of benefits it poses. Among others, $glmmTMB$ is more flexible than other packages available for estimating zero-inflated models via maximum likelihood, and faster than packages that use Markov chain Monte Carlo sampling for estimation (Brooks et al., 2017b).

Furthermore, it is also rated high in terms of flexibility for zero inflated modelling than INLA even though speed comparisons vary with model and data structure. Study results by Bolker et al. (2009) were that repeated measurements on the same individual, the same location, or observations taken at the same point in time are often correlated and this correlation can be accounted for using random effects in a GLMM. The $MCMCglmm$ and $brms$ packages can fit zero-inflated GLMMs with predictors on zero-inflation, but they are relatively slow. Based on these limitations and challeges, the researcher preferred using the newly developed R package glmmTMB that can easily estimate zero-inflated GLMMs using maximum likelihood. The ability to fit these types of models quickly, using a single package made it easier to find the best model to explain patterns in the data. According to Berridge (2011), in GLMMs, the explanatory variables and the random effects (for a two level model, x_{ij}, z_j and μ_{0j}) affect the response (for a two-level model, y_{ij}) via the linear predictor (θ_{ij}), where:

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \mu_{0j}. \quad (2.1)$$

Let $i = 1, 2, \dots, m$ denote the index of the observational unit while $j = 1,2,3, \dots, m_i$ denotes the index for the within observation in this unit. In this context $\theta = [\beta^T, \sigma^T]^T (d \times 1)$ denotes the vector of model parameters, where $\beta_{(p\times1)}$ represents the parameters for the fixed effect, and $\sigma_{(s\times1)}$ includes the parameters for the random effects while $d = p + s$. The observed outcome y_{ij} is assumed to be independently drawn from exponential family of distribution when conditioned on a vector $X_{ij}(p \times 1)$ and the random effects vector $\gamma_i(q \times 1) \sim N_q(0, \Delta_\sigma)$. In this case Δ_σ is a positive-definite symmetric covariance matrix. For simplicity, the study considered reparametrization $\gamma_i = D_\sigma u_i$ resulting from the Cholesky decomposition of $\Delta_\sigma = D_\sigma {D^T}_\sigma$ where u_i denotes the multivariate standard normal vectors. The GLMM is of the form:

$$g(u_{ij}) = \eta_{ij} = {X^T}_{ij}\beta + {Z^T}_{ij} D_\sigma u_i. \quad (2.2)$$

where u_{ij} denotes the conditional expectation of the outcome, $Z_{ij}(q \times 1)$ a design vector for the random effects and η_{ij} the linear predictor. Furthermore, $g(\,)$ represents a link function which maps the linear predictor and the conditional expectation of the outcome (Flores-Agreda & Cantoni, 2019).

The GLMM is obtained by specifying some function of the response (y_{ij}) conditional on the linear predictor and other parameters, i.e.

$$f(y_{ij}|\theta_{ij}, \emptyset) = \exp\left\{\frac{[y_{ij}\theta_{ij} - b(\theta_{ij})]}{\emptyset} + c(y_{ij}, \emptyset)\right\}, \quad (2.3)$$

where $\emptyset$ is the scale parameter, $f(\,)$ denotes the Probability Density Function (PDF), $b(\theta_{ij})$ is a function that gives the conditional mean (μ_{ij}) and variance of y_{ij}, namely:

$$E[y_{ij}|\theta_{ij}, \emptyset] = \mu_{ij} = b'(\theta_{ij}), \quad (2.4)$$
$$Var[y_{ij}|\theta_{ij}, \emptyset] = \emptyset b''(\theta_{ij}), \quad (2.5)$$

while $c()$ is a funtion that is automatically determined once the other functions have been chosen (or simply denotes a specific function), so that the entire distribution is normalized.

In generalized linear mixed models, the mean and variance are related so that:

$$Var[y_{ij}|\theta_{ij}, \emptyset] = \emptyset b''(b'^{-1}(\theta_{ij})) = \emptyset V[\mu_{ij}]. \quad (2.6)$$

where $V[\mu_{ij}]$ is refered to as the variance function, $b'^{-1}(\theta_{ij})$ is a link function which expresses θ_{ij} as a function of μ_{ij}, and $b'(\theta_{ij})$ is the inverse link function. The functions $b(\theta_{ij})$ and $c(y_{ij}, \emptyset)$ differ for different GLMMs. The distribution that works well in modelling the ZINB-GLMM is **nbinom2** by Magnusson et al., (2017), which returns an overdispersion parameter.

The expressions of the marginal PDF are obtained after integrating the random effects from the joint distribution $[y_{ij}, u_i{}^T]^T$.

Since the study data were counts which were zero-inflated, the Poisson, Negative Binomial, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial models were fitted as alternatives to the Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM). The ZINB-GLMM could be more reasonable for this study because of its ability to handle multiple random effects components together with the zero-inflation and dispersion components. ZINB-GLMM performed well among other models for non-normal count data involving non-structural zeros due to its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at high level.

The model selection was done using AIC and Residual plots. The AIC is a popular method for comparing the adequency of multiple, possibly non-nested models. The current practice is to accept a model with a small AIC value (Wagenmakers, 2004). The equation of the AIC is:

$$AIC = -2LL + 2r \qquad (2.6)$$

where LL is a log likelihood value, r indicates number of parameter and n is a sample size (Posada & Buckley, 2004). In assessing the best fit model, the study further used the R package DHARMa by Walker (2018). DHARMa stands for "Diagnostics for HierAchical Regression Models". DHARMa includes support from glmmTMB and it is suitable for testing whether a GLMM is in harmony with the data (Dunn & Smyth, 1996). However, there are still a few limitations such as misspecifications in GLMMs which cannot be reliably diagnosed with standard residual plots.The expected distribution of the data changes with the fitted values and that makes GLMM residual harder to interpret. The current standard practice is to eye ball the residual plots for major misspecifications, potentially have a look at the random effect distribution and then run the test for overdispersion, but this approach still possesses a number of problems. The scaled (quantile) residuals are computed with the ***simulateResiduals( )*** function in R. The default number of simulation (n =250) was considered to be a reasonable deal between computation time and precision. What the function does is to create n-new synthetic datasets by simulating from the fitted model, compute the cummulative distribution of simulated values for each observed value and then return the scaled value that corresponds to the observed values (Walker, 2018).

The main objective of this research was to investigate ecological factors influencing property crime in Windhoek, to provide information and insight leading to better and improved prevention strategies of this category of crime. The specific objectives were to assess the ecological characteristics of property crime in Windhoek with a view to create understanding of its root cause; evaluate property crime in Windhoek and determine locations with high crime or crime hot spots for possible interventions; determine the season of the year in which property crime happen more often; and to model factors influencing property crimes in Windhoek and asses their impact.

3. Methods

A quantitative design was adopted in this study based on secondary data on daily reported Property crimes obtained from the Windhoek City Police department (2011 to 2016). The City Police Chief authorised the researcher to use the data. The daily crime data were recorded in the pocketbooks by the City police officers who attended to crime scenes.

After the pocketbooks were fully completed, they were submitted to the immediate supervisors or to the City Police Statistics Department for recording. In this study, the response variable was the Number of Property crimes (Y_i). The independent variables X_i, for *i =1, 2….5*, refer to the *Month*, *Year*, *Location*, *Season* and *Density*. The variable *Month* represented a specific month from January to December in which crimes were committed in a specific year within a location.

The variable *Location* represented fifty-nine Windhoek geographical residential areas namely, Academia, Babylon, Brakewater, Cimbebasia, Damara Location, Dolam, Donkerhoek, Dorado Park, Dorado Valley, Eehambo NdaNehale, Eros, Eros Airport, Freedom Square, Freedomland, Gemeente, Golgota, Goreangab, Green-Well Matongo, Grysblock, Hakahana, Havana, Herero Location, Hockland Park, Independence Arena, Khomasdal, Kilimajaro, Kleine Kuppe, Klein Windhoek, Lafrenz, Ludwigsdorf, Marua Mall, Malaka Draai, Maroela, Mix, Mukwanangombe, Mukwanekamba, Northern Industrial, Nubuamis, Okahandja Park, Okuryangava, Olympia, Ombili, One-Nation, Ongulumbashe, Oshitenda, Otjomuise, Pionierspark, Properita, Rocky Crest, Shandumbala, Single Quarter, Southern Industrial, Soweto, Suiderhof, Vambo Location, Wanahenda, Windhoek Central, Windhoek North, and Widhoek West.

Prior to fitting the models for this study, a new variable *Season,* which represents Summer (January-March), Autumn (April-June), Winter (July-September), and Spring (October-December), was created (Kemper & Roux, 2005).

The variable *Density* was obtained as the population per kilometer square of the area. The density was then scaled per 10 000 people. A good rule of thumb is that input variables should be small values, probably in the range of 0-1 or standardized with a zero mean and a standard deviation of one. The dataset was cleaned and re-coded prior to the data analysis.

To obtain the first overview of the dependent variable (Number of Property crimes), a histogram, boxplot and the normal Q-Q plot of the observed count frequencies were presented. Multiple plots of the Number of Property crimes were displayed to check the crime pattern on a yearly, seasonal and monthly basis via boxplots.The study further assessed the patterns of these crimes across all the fifty-nine locations.

4. Results

Cross tabulation were computed using the Statistical Package for Social Sciences (SPSS) version 25 (Green & Salkind, 2016). This was done to obtain the final counts of reported cases across the Property crime. Table 1 below shows a count summary of seasonal reported cases under Property crime, reported from 2011 to 2016.

Table 1: Summary of Property crime statistics across seasons and years

		No of Property crimes	
		Count	percent
Season	Summer	17671	21.7
	Autumn	20848	25.5
	Winter	21248	26.0
	Spring	21830	26.8
Year	2011	12922	15.8
	2012	15105	18.5
	2013	637	0.8
	2014	17343	21.3
	2015	17970	22.0
	2016	17620	21.6

There were 81,597 Property crimes in total over the study period: 12922 in 2011, 15105 in 2012, 637 in 2013, 17343 in 2014, 17970 in 2015, and 17620 in 2016. The percentages of Property crimes were calculated by dividing Property crimes for each year by the total number of crimes for all years. Notably, of the 81,597 Property crime cases during the study period: 17,671 were counted in summer, 20,848 in autumn, 21,248 in winter and 21830 in spring.

The number of reported Property crimes ranged from 0 to 258 with an average of 41 cases per year. Findings showed that for the study locations, the minimum number of people per square kilometer was 41 while the maximum is 21 812. In addition, the average number of people per study location was found to be 4611.

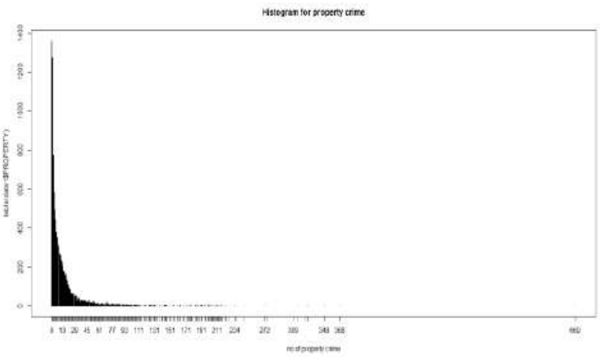


Figure 1: Histogram for the Number of Property crimes

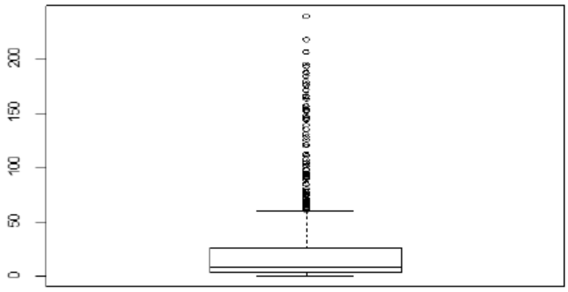
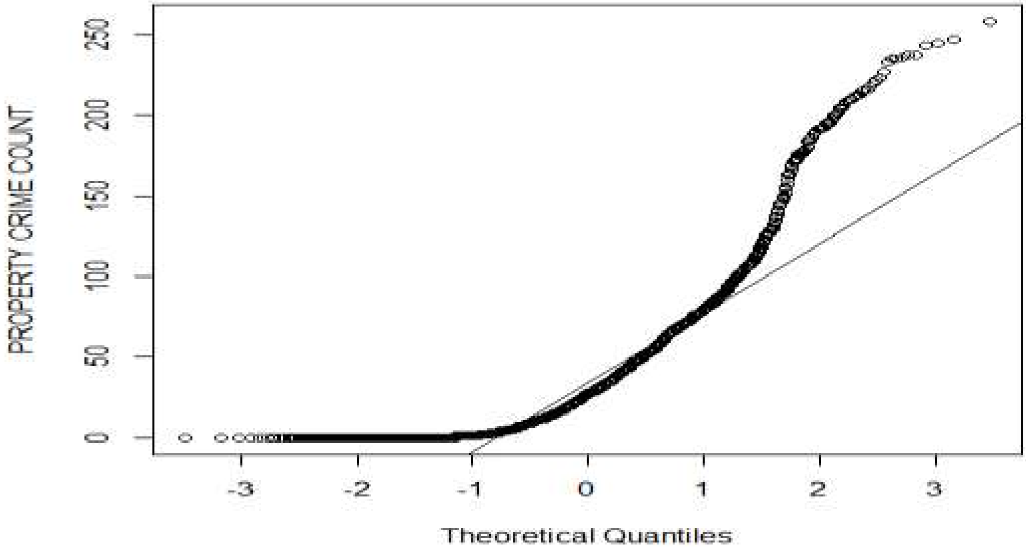


Figure 2: A box plot for the number of property crimes

The histogram (Figure 1) illustrates that the marginal distribution exhibits both substantial variation and a rather large number of zeros. It is clearly evident that the number of property crimes appears to be positively skewed, as indicated by the relative position of the median within the box plot (Figure 2) that contains half the data. However, there are some few outliers as shown in the Figure 2.

There are two distinct processes driving the zeros, one is non-structural zeros (sampling zeros) which occur by chance and can be assumed to be a result of a dichotomous process. The other one is structural zeros (true zeros) which are part of the counting process. Based on this concern, the choice should be based on the model providing the closest fit between observed and predicted values. The choice of the zero-inflated model in this paper is guided by the researcher's belief about the source of the zeros.



$$Shapiro\ Wilk\ normality\ test\ W = 0.80677, p - value < 0.001$$

Figure 3: Number of Property crime Q-Qplot.

The Q-Q plot (Figure 3) shows that the Number of Property crimes are positively skewed, as the points fall above the line as x-values increases. This violates a very important assumption for the linear mixed effect model and rather supports generalized linear mixed models. Literature outlined that, linear models are not approriate in some situations where the response is restricted to binary and count. In addition, linear models fail when the variance of the response depends on the mean. Finally, Shapiro's test result indicates that that the distribution is not normal.

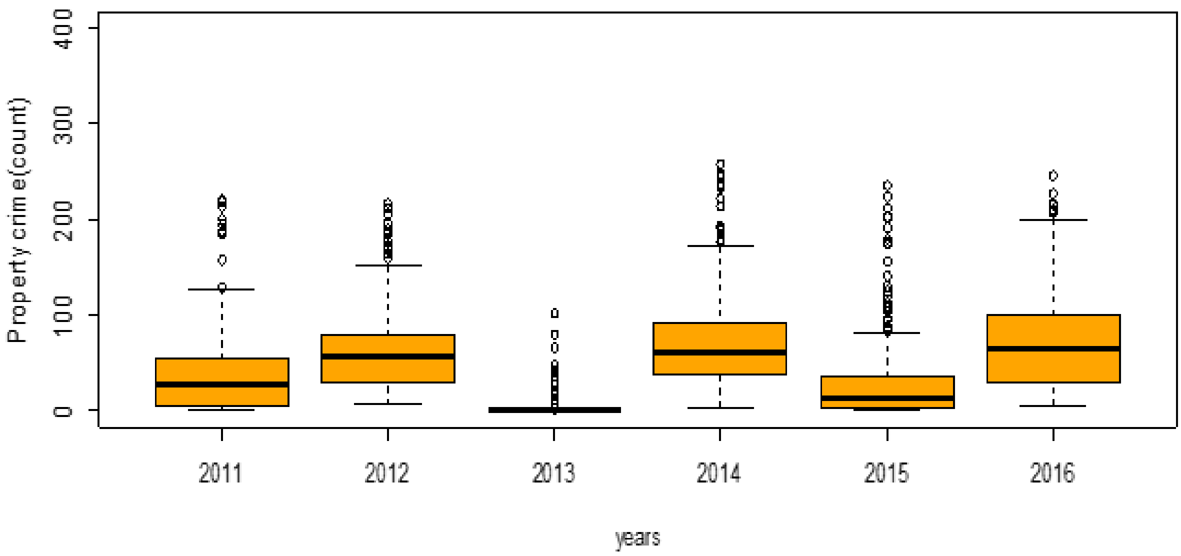


Figure 4: Number of Property crime distribution for each year

The boxplot presented in Figure 4, shows the variation in the median Number of Property crimes across the study periods, with the years 2012, 2014 and 2016 showing very similar median of the number of cases. This indicates that the variation across the years needs to be taken into account when fitting the model.

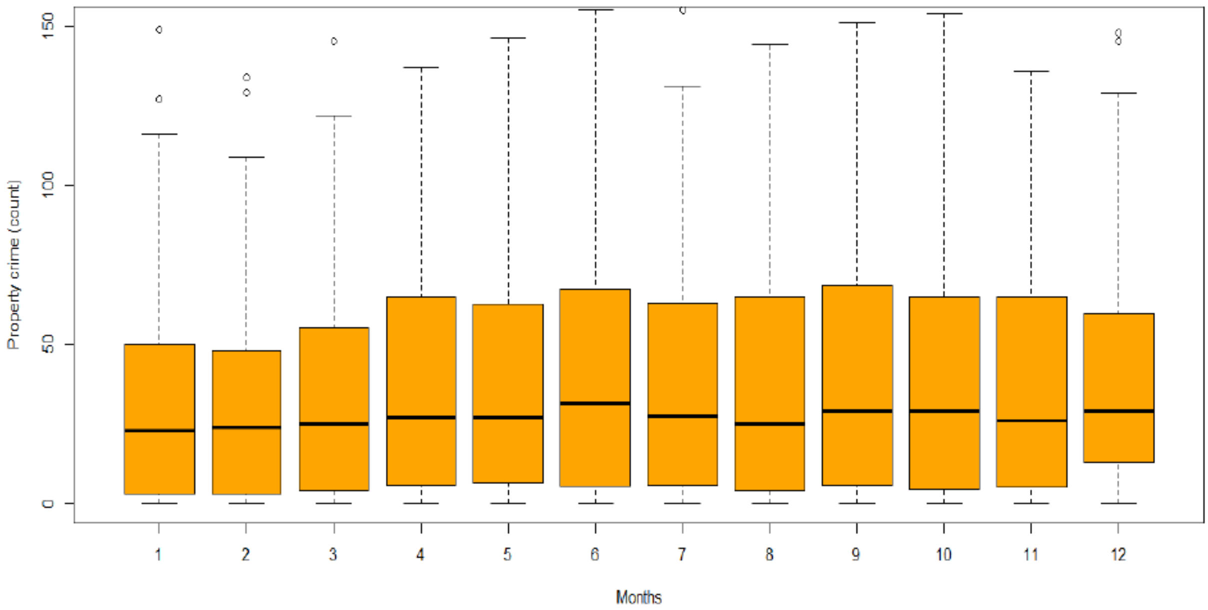


Figure 5: The Number of Property crime distribution for each month (January to December).

The boxplot shows that the median for the property crime is slightly different. It also shows that each month presents a different amount of variation in Property crime so that there is an overlap of values between some months. The variation seems to be high as from April to December. There are still noticeable differences and hence the study better accounts for them in the model.

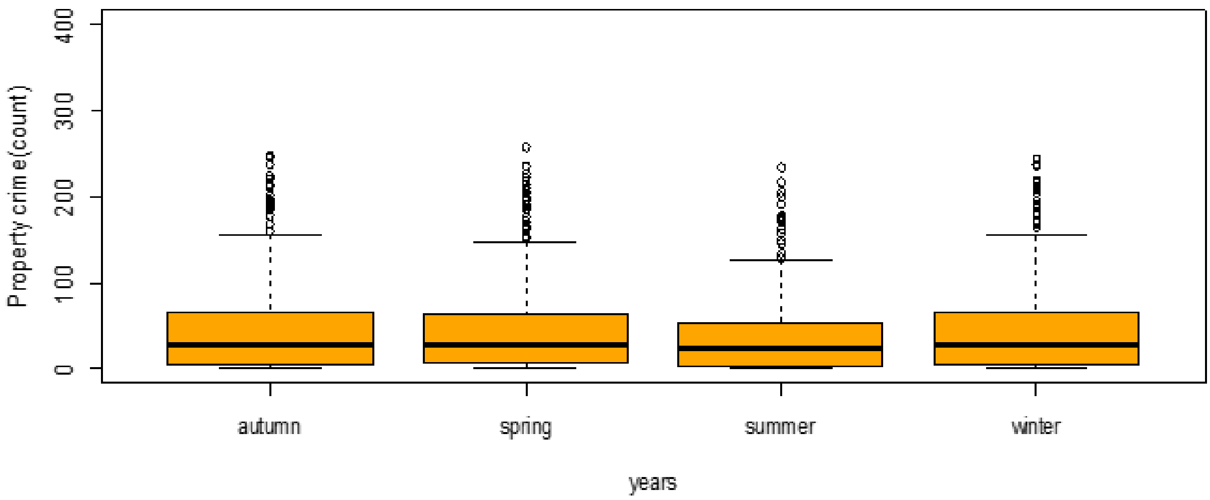


Figure 6: Number of Property crimes for each season

The above Figures 4, Figures 5 and Figures 6, demonstrate the possible presence of outliers in the study of property crimes. Hence, suggesting that the study should not heavily rely on the average of property crimes as they may distort the results. It was noted that in all the box plots, the average number of property crimes were above the median since the data was positively skewed.

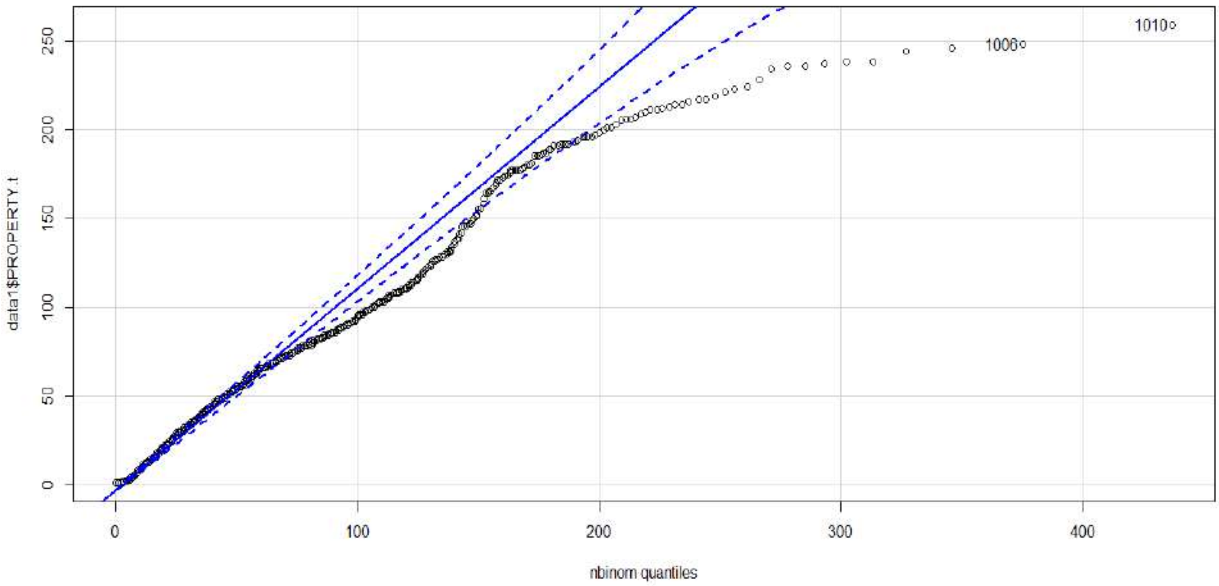


Figure 7: Asessing if the Number of Property crime data follows a negative binomial distribution using confidence interval

Considering the plots generated using Q-Q plot in Figure 7, the y-axis represents the observations, and the x-axis represents the quantiles modelled by the distribution. The solid blue line represents a perfect distribution fit, and the dashed blue lines are the confidence intervals of the perfect distribution fit. The aim is to see if the data follows a normal distribution or other distributions. In this case, it is the negative binomial distribution, in which only a few observations fall outside the dashed lines. This suggests that a negative binomial probability distribution best fits the Property crime data.

4.1 Model selection

In this study, ZINB-GLMM was reasonable in modelling the Number of Property crimes because of its small AIC values as compared to ZINB, ZIP, NB, and Poisson (Table 2). ZINB-GLMM was the best model for this study based on the benefit that it accounts for within variation through random effects and captures the non-structural zero counts in the dataset.

Table 2: Comparing AIC for five Property crime models.

MODELS	PROPERTY CRIME
ZINB-GLMM	14292.6
POISSON	95365
NB	18133
ZIP	74613
ZINB	17972

4.2 Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Property crime.

For the ZINB-GLMM, the response variable of interest was the Number of Property crimes while the independent variables were *Month*, *Season*, *Location*, *Year* and *Density*. The first four variables (Month, Season, Location and Year) were the random effects. These effects were chosen to be "random" because the crime commited in one month is independent of the crime commited in the next month. This is also applicable to Location, Season and Year. However, Density was chosen to be the fixed effect since the number of people per square kilometer could be measured during this study period. Moreover, the random effects were nested in this study, because each police officer recorded a certain number of cases, and no two officers recorded the same case.

The full Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) was modelled on the r output format of lme4. The model fitted was:

$$Mixedglmm1 < -glmmTMB(PROPERTY \sim Density2 + (1|Month) + (1|Year) + (1|Location) + (1|Season), zi = \sim Year + disp = \sim Season, data1, family = nbinom2) \quad (4.1)$$

Following the arguments in the function (Mixedglmm1), this model allows the conditional mean to depend on Density and vary randomly by Location (l), Month (m), Year (y) and Season (s). It further allows the number of structural (extra) zeros to depend on years. Additionally, it allows the dispersion parameter to depend on the season of the year. This model can be represented by the following specific equations (Brooks et al., 2017b):

$$\mu = E(count|a_{lmys}, NSZ) = \exp\,(\beta_0 + \beta_{Density2} + a_{lmys}) \quad (4.1)$$

$$a_{lmys} \sim N(0, \sigma^2_{a_{lmys}}) \quad (4.2)$$

$$\sigma^2 = var(count|a_{lmys}, NSZ) = \mu(1 + \frac{\mu}{\theta}) \quad (4.3)$$

$$Logit(p) = \beta_0^{z_i} + \beta_{year}^{z_i} \quad (4.4)$$

$$\log(\theta) = \beta_0^{(disp)} + \beta_{season}^{(disp)} \cdot Season \quad (4.5)$$

where a_{lmys} are Location, Month, Year and Season specific random effects, NSZ is the event "non-structure zero", $p = 1 - p_r(NSZ)$ is the zero-inflation probability, and β's are regression coefficients with the subscript denoting the covariate/ level (with 0 denoting intercept).

Table 3: Summary of the ZINB-GLMM for the Number of Property crimes

	Parameters	Est.	SE	RR	95% CI	p-value
Random effect	$\sigma_{Location}$	1.38				
	σ_{Month}	0.05				
	σ_{Year}	1.05				
	σ_{Season}	0.10				
Conditional model	Intercept	2.91	0.30	18.31	17.72; 18.90	***
	Density2	-0.79	0.35	0.45	-0.23; 1.13	*
Zero-Inflation model	Intercept	-0.50	0.36	0.60	-0.11; 1.31	***
	Ref (2011)					
	2012	-17.48	2387.26	0	-4679.03; 4679.03	
	2013	5.72	0.43	304.9	304.06; 305.74	***
	2014	-17.53	2385.93	0	-4676.42; 4676.42	
	2015	-0.77	0.50	0.46	-0.52; 1.44	
	2016	-17.53	2385.93	0	-4676.42; 4676.42	
Dispersion model	Intercept	1.96	0.10		-0.196; 0.196	***
	Ref(Summer)					
	Autumn	0.17	0.14	1.19	0.92; 1.46	
	Winter	-0.11	0.14	0.9	0.63; 1.17	
	Spring	-0.18	0.14	0.84	0.57; 1.11	

Estimate (Est.), Standard Error (SE), Relative Risk ratio (OR), Reference category (Ref), Significant level (***=0.001; *=0.05)

The model summary can be broken down into five sections. The first section includes the general overview containing a description of the model specification (family, formula, zero inflation, dispersion, data) together with the information criterion (AIC and BIC).

The second section describes the variability of the random effects. In this model, we only had random effects in the conditional model (equation 4.1). The estimated standard deviations; $\boldsymbol{\sigma_{Location} = 1.38}$, $\boldsymbol{\sigma_{Month} = 0.05}$, $\boldsymbol{\sigma_{Year} = 1.05}$, $\boldsymbol{\sigma_{Season} = 0.10}$ corresponding to $\sqrt{\sigma^2_{a_{lmys}}}$ in equation (4.2). This indicates how much of the variation in the study of property crimes can be attributed to each random term. The variability of the Number of Property crime according to Month and Season was smaller when compared to that of the Location and Year. This was caused by the high dispersion parameter computed under the two variables. The smaller variation in months and seasons, an indication of less time, is always expected between Property crime cases. To support the view, on the monthly and seasonal perspective, the study has shown that Property crime happens frequently.

The third section describes the relative risk ratio of the conditional model ($\beta_0, \beta_{Density2}$) including a 95% confidence interval and p-values. For the confidence interval, the null value is one since it is estimated on a natural scale. In most cases where a 95% confidence interval does not include the null value the findings are statistically significant. Alternatively, parameters that are statistically significant in the model have a p-value below 0.05 as shown in **Table 3**. Both the intercept ($\beta_0 = 18.31$) and the relative risk ratio $\beta_{Density2} = 0.45$ are statistically significant. Based on this, it means without considering the population density, approximately 18 Property crimes can be anticipated in Windhoek per annum. The expected counts are conditional on every other value being held constant. That is, including the random Location, Month, Year, and Season effects, population density is expected to have a 45 percent increase on property crime. In other words, considering the population density, the Number Property crimes will increase by 45%.

The fourth section describes the zero-inflation model which is like the conditional model except that this model has a logit link. The estimates in this section correspond to $\beta_0^{z_i}$and $\beta_{year}^{z_i}$ relative risk ratio from equation (4.4). The Zero-Inflation model estimates the probability of an extra zero. The baseline odds of no property crime reported in Windhoek is 0.6. In addition to that, $\beta_{2013} > 0$ means higher occurrences in year 2013 unaffected by population density. This essentially means that during year 2013, the number of property crimes that were not recorded (but there was an intention) was not due to the number of residents per square kilometre in the area. In contrast, the exploratory study has proven that an area with a high population density experienced high number of property crimes during each season. The confidence intervals were estimated given by:

$$e^{\widehat{\beta_i} \pm Z_{\frac{\alpha}{2}}(S.E(\widehat{\beta}_i)} \qquad (4.6)$$

In this case $\hat{\beta}_i$ represents the model parameters estimated, S.E is the standard error for the corresponding parameter and Z corresponds to the critical value associated with a 95% degree of confidence. The second component of the model [$Z_{\frac{\alpha}{2}}(S.E(\hat{\beta}_i)$] is called the margin of error.

Since $\hat{\beta}_i = \log(mean) = \log(\mu)$, hence $e^{\hat{\beta}_i} = (mean) = \mu$.

From **Table 3**, the confidence interval for the estimate $\hat{\beta}_{2013}$ on the conditional model is $304.06\,;\,305.74$. It was concluded that this confidence interval provided the study with plausible values for the parameter. If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain that population parameter.

5. Discussion

During the study period, an average of 68% of the recorded crime was Property crime in Windhoek. This result is not very different from the statistics by NUMBEO (2018) which indicated that Property crimes stood at 74.06%, during the same period.

Specifically, the Number of Property crimes was slightly high during Spring and Winter time. During Spring (October-December) crime increased probably because large numbers of residents normally travel for holidays with their families leaving their houses unattended or with no security. This motivates the offenders to commit Property crimes such as house breaking. Whereas in Winter (July-September), it was assumed that criminals take advantage of the windy and cold weather to commit property crime as most people prefer to stay indoors due to the cold weather. However, significant to note is that correlation does not mean causation. A possible explanation is not that there are more people on the streets committing crimes during the holidays and Winter times, but also that the possibility of arresting the offenders is lessened by the fact that there could be less cops patrolling in the streets at that time. Also, it is possible that the Property crime rate during these times was higher than what the data shows because of a reporting bias. In general, the number of property crimes has slightly increased in Windhoek between 2011 and 2016.

It was found that there was a direct relationship between the Number Property crime and population densities in Windhoek in line with the study results by (Hanley et al., 2016). The current study results show that most areas with a high population density have a higher crime intensity (with an exemption to Windhoek Central area).

Furthermore, those locations with high population densities were considered as overcrowded areas with people of generally low socio-economic status. Residents in these areas often move around leaving their houses unguarded and this attracts criminals. This is also in agreement with findings by (Cohen & Felson, 2016) that crimes result from the convergence of some elements such as, suitable target, motivated offender, and the absence of capable guardians. Even though Windhoek Central has a small population density more crimes were recorded there as people gather in the area for employment, school, and shopping purposes. Using the same logic, affluent areas attract more criminals for theft and robberies due to the opportunities available to them (Justus & Kassouf, 2013). Besides that, the study results strengthen findings that local crime rate are influenced by income deprivation and housing tenure structures (Livingston, Kearns, & Bannister, 2014).

Although Property crimes were found to be zero inflated, regardless of population density, 18 counts of Property crimes were expected every year. These data revealed that the number of property crimes have been constantly increasing as from 2011 to 2016. The zeros (non- structural zeros) obtained were results of no crime recorded within some months or seasons of the study period due to effective policing or lack of suitable targets and this contributed to the choice of ZINB-GLMM. The study indicated that crime data can be modelled using ZINB-GLMM as an alternative to spatial temporal patterns that were used by other researchers. Perhaps, an important factor to note, is that several other variables were not significant in the model and were not interpreted. However, this is not to say that insignificant findings signal no impact of the indicator on property crimes, since individual level driving factors cannot be investigated using police data. More complex models that include more variables influencing the Number of Property crime and interaction among these variables at different levels of aggregation would be preferable.

The crime data considered for this study was secondary data received from Windhoek police, focusing only on Windhoek reported crimes. However, due to the sensitive nature of this study, the researcher was given limited information and access to the crime data received from Windhoek police station.

6. Conclusion

This research investigated the ecological factors influencing the Number of Property crimes in Windhoek to provide information and insights leading to more effective and improved prevention strategies of these crimes. Premised on this background, the study adopted the ZINB-GLMM approach, which evaluated uncertainty in the random effects contributing to the variation in the Number of Property crimes. The random effects evaluated were based on Month, Year, Season and Location. The results indicated that most of the variation in the study of Property crimes was due to Location while the effect of Month, Year and Season was not as pronounced.

Density was one of the major contributing factors of the high crime rate in Windhoek. Overcrowded areas tend to attract more criminals for theft and robberies due to the opportunities available to them. Okuryangava location was among locations with very high population densities. Even though Hakahana has the highest population density, it was found that the number of reported crimes were quite few.

The study recommends more effective policing in Windhoek during Spring and Winter time, specifically in the areas with high population densities.

Windhoek community members should team up in neighbourhood watch interventions to avoid highly diverse population, with little of the social “glue” that binds communities together, in order to effectively reduce crime. Community members should avoid non-essential mobility and make security arrangements if they have to travel. The Windhoek police could record other important background variables such as the employment status, level of education, age and tribe etc. of the criminal when recording crime to improve data quality in future. It would also be advisable for the Windhoek police to geo-code crime data by location so that future researchers will analyse the spatial aspect of crimes.

References

Asongu, S., & Kodila-Tedika, O. (2016). Fighting African conflicts and crimes: which governance tools matter? *International Journal of Social Economics, 43*(5), 466-485.

Ayling, J. (2013). What sustains wildlife crime? Rhino horn trading and the resilience of criminal networks. *Journal of International Wildlife Law & Policy, 16*(1), 57-80.

Bolker, B. M. (2018). Getting started with the glmmTMB package.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

Breetzke, G. D., & Pearson, A. L. (2014). The fear factor: Examining the spatial variability of recorded crime on the fear of crime. *Applied Geography, 46*, 45-52.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017a). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal, 9*(2), 378-400.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017b). Modeling zero-inflated count data with glmmTMB. *bioRxiv*, 132753.

Cohen, L. E., & Felson, M. (2016). Social Change and Crime Rate Trends: A Routine Activity Approach (1979). In *Classics in Environmental Criminology* (pp. 203-232): CRC Press.

Council, O. S. A. (2010). South Africa 2010 Crime & Safety Report. *United States Department of State Bureau of Diplomatic Security*.

de Melo, S. N., Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography, 62*, 314-324.

Dore, P. (2013). Urbanisation, Inequality and Youth. *State of the Urban Youth, India 2012*, 75.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics, 5*(3), 236-244.

Ellis, S., & Shaw, M. (2015). Does organized crime exist in Africa? *African Affairs, 114*(457), 505-528.

Flores-Agreda, D., & Cantoni, E. (2019). Bootstrap estimation of uncertainty in prediction for generalized linear mixed models. *Computational Statistics & Data Analysis, 130*, 1-17.

Green, S. B., & Salkind, N. J. (2016). *Using SPSS for Windows and Macintosh, Books a la Carte*: Pearson.

Han, L., Bandyopadhyay, S., & Bhattacharya, S. (2013). Determinants of violent and property crimes in England and Wales: A panel data analysis. *Applied Economics, 45*(34), 4820-4830.

Hanley, Q. S., Lewis, D., & Ribeiro, H. V. (2016). Rural to urban population density scaling of crime and property transactions in English and Welsh parliamentary constituencies. *PloS one, 11*(2), e0149546.

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ, 2*, e616.

Jones, R. W., & Pridemore, W. A. (2012). The Foreclosure Crisis and Crime: Is Housing-Mortgage Stress Associated with Violent and Property Crime in US Metropolitan Areas? *Social Science Quarterly, 93*(3), 671-691.

Justus, M., & Kassouf, A. L. (2013). Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. *Economía, 14*(2), 88-101.

Kemper, J., & ROUX, J. P. (2005). Of squeezers and skippers: factors determining the age at moult of immature African Penguins Spheniscus demersus in Namibia. *Ibis, 147*(2), 346-352.
Khan, N., Ahmed, J., Nawaz, M., & Zaman, K. (2015). The socio-economic determinants of crime in Pakistan: New evidence on an old debate. *Arab Economic and Business Journal, 10*(2), 73-81.
Livingston, M., Kearns, A., & Bannister, J. (2014). Neighbourhood structures and crime: the influence of tenure mix and other structural factors upon local crime rates. *Housing Studies, 29*(1), 1-25.
Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., . . . Brooks, M. M. (2017). Package 'glmmTMB'.
Neema, I., & Böhning, D. (2012). Monitoring murder crime in Namibia using Bayesian space-time models. *Journal of Probability and Statistics, 2012*.
NUMBEO. (2018, January 2018). Crime in Windhoek, Namibia. Retrieved from https://www.numbeo.com/crime/in/Windhoek
Palmary, I., Rauch, J., & Simpson, G. (2014). Violent crime in Johannesburg. In *Emerging Johannesburg* (pp. 117-138): Routledge.
Posada, D., & Buckley, T. R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology, 53*(5), 793-808. doi:10.1080/10635150490522304
Schielzeth, H., & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution, 4*(1), 14-24.
Shaw, M., & Reitano, T. (2013). The evolution of organised crime in Africa-towards a new response. *Institute for Security Studies Papers, 2013*(244), 28.
Tonry, M. (2014). Why crime rates are falling throughout the Western world. *Crime and justice, 43*(1), 1-63.
Walker, S. (2018). *Crisis-Opportunity, Liability-Asset: Governing Vacant Land Reuse in Cleveland, Ohio.*
Wambua, P. M. (2015). Call the police? Across Africa, citizens point to police and government performance issues on crime.