RISK FACTORS ASSOCIATED WITH MORTALITY OF BREAST CANCER PATIENTS AFTER SURGERY: THE CASE OF ZAMBIA.

By

MULOPE MULOPE

A dissertation submitted to The University of Zambia in fulfilment of the requirements for the Master of Science Degree in Statistics in the Department of Mathematics and Statistics, School of Natural Sciences.

THE UNIVERSITY OF ZAMBIA

LUSAKA

2024

AUTHOR'S DECLARATION

I, **Mulope Mulope**, do declare that this dissertation represents my own work and that it has neither in part nor in whole, been presented as substance for the award of any degree at this or any other institution of learning or research. Where other people's work has been used, acknowledgement has been made.

Candidates Name: …………………………………………………………………………

Signature of Author: ………………………………………………………………………

Date: ……………………………………………………………………………………….

Supervisor's Name: …………………………………………………………………………

Signature of the Supervisor: ………………………………………………………………...

Date: …………………………………………………………………………………………

COPYRIGHT

CERTIFICATE OF APPROVAL

The University of Zambia approves the dissertation of **Mulope Mulope** as fulfilling part of the requirements for the award of Master of Science Degree in Statistics.

Name Of Examiner	Signature	Date
1. …………………………………………	…………………	…………………
2. …………………………………………	…………………	…………………
3. …………………………………………	…………………	…………………

Supervisor's Name: …………………………….. …………………. …………………..

Chairperson Board of Examiners

Name: ……………………………………… ………………..… …….……………

ABSTRACT

Breast cancer (BC) has become a public health problem world-wide due to its high mortality rate among women in both developed and developing countries. Numerous studies have been conducted to determine prognostic factors for breast cancer mortality and projection of mortality in developed countries. However, little was known of how factors related to both patients care and personal attributes in the case of Zambia, with a different environmental and cultural setting. Therefore, this study identifies the factors that affects survival rates and estimates a 5-year mortality of BC patients after surgery.

We conducted a retrospective study encompassing 233 women who had undergone BC surgery between 2013 and 2018, and were followed-up to the end of 2019. The data were gathered from medical records of patients from the cancer center registry at Cancer Disease Hospital (CDH) in Lusaka, Zambia. We determined the prognostic factors of BC mortality after surgery using the Logistic regression model. The most commonly used statistic of comparison was the odds ratio (OR). Further, 5-year mortality rates for various age groups were estimated using the fitted model.

A total of 54 deaths were recorded during the study period. The results of the study showed that age at surgery, marital status, HIV status, BMI, BC stage, histologic grade, Progesterone Receptor (PR) status were statistically and significantly associated with mortality of BC patients after surgery, using a multiple logistic regression model. It was also established that factors which had two response categories (positive or negative) had higher mortality rates for subjects who were found to be positive on that factor compared to subjects who were negative. For other factors such as tumor size, lymph node status, BC stage and histologic grade, the odds of death increased with the degree of severity across the levels of the factor. Additionally, the odds of death were higher for both HIV positive and the single patients compared to the HIV negative and married ones, respectively. Further, the odds of death for obese or overweight patients were more than twice of those classified as not obese or overweight. Furthermore, the 5-year mortality of different age groups using multiple logistic regression stood at 74.1% for the younger group (≤ 35 years) and 98.9% for the older group (50 <= age). Therefore, the study recommends the need for early detection of breast cancer along with the availability and accessibility of appropriate treatment.

Keywords: Breast cancer, mortality, survival, logistic regression, prognostic, retrospective study

DEDICATION

This research project is dedicated especially to my Almighty GOD for His abundant blessings and protection, to my lovely wife Palisa Sonile Mulope for the sacrifice she made for me to complete this project, her love, care endless support, daily encouragement and enthusiasm that inspired me to achieve this goal, to my mother and late father; Mr and Mrs Nakushowa, brothers and sisters, relatives, friends and colleagues. It is also dedicated to my first-born son Mulope and other unborn lovely kids.

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my Heavenly Father, the Almighty GOD for His love, care and guidance and also seeing me through this work successfully. Through Him everything is possible.

I will again extend my deepest gratitude to my lecturer and supervisor, Dr. A Banda for his direction, coordination and critics from the first to the last page of this work. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. I would also like to thank all the members of staff at The University of Zambia, School of Natural Science especially those from the Department of Mathematics and Statistics who made it possible for me to complete this work. May the Almighty GOD bless you all.

Nobody has been more important to me in the pursuing of this programme than the members of my family. I would like to thank my parents, for their wise counsel and sympathetic ear, whose love and guidance are with me in whatever I pursue. More importantly, I wish to thank my loving and supportive wife, Palisa, and my wonderful son, Mulope, who have been my pillars of strength and support. You are always there for me.

I would also like to thank my fellow MSc students for the enjoyable learning environment and camaraderie. In addition, I would like to acknowledge my friends, who were of great support in deliberating over our problems and findings, as well as providing happy distraction to rest my mind outside of my research.

Further, my thanks are due to Ministry of Health especially the Management at Cancer Disease Hospital for the support rendered to me. Finally, special profound thanks go to all other concerned people who in one way or the other helped me in producing this work, I say thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	: American Cancer Society
Art	: Anti-retroviral treatment
BC	: Breast Cancer
BMI	: Body Mass Index
BSE	: Breast Self-Examination
CDH	: Cancer Disease Hospital
CI	: Confidence Interval
CK	: Cytokeratins
DFI	: Disease Free Interval
ER	: Estrogen receptor
GLMs	: Generalized Linear Models
HER2	: Human Epidermal Growth factor receptor 2
HIV	: Human Immunodeficiency Virus
JBCRG	: Japan Breast Cancer Research Group
MoH	: Ministry of Health
NGOs	: Non-Governmental Organizations
OR	: Odds ratio
PH	: Proportional Hazard
PR	: Progesterone receptor
S. E	: Standard errors
SPSS	: Statistical Package for the Social Sciences
TNM	: Tumor, Node and Metastasis
UK	: United Kingdom
UTH	: University Teaching Hospital
WHO	: World Health Organization

CHAPTER ONE: INTRODUCTION

1.1 Overview

This chapter explains the background of the study, statement of the problem, purpose of the study, objectives and research hypothesis. Further, significance of the study, delimitation, and operational definitions are also presented in this chapter.

1.2 Background

Breast cancer has become a source of concern worldwide due to its high mortality rate among women in both developed and developing countries [1, 2]. It is the most common severe disease and one of the main causes of death in women around the world [3, 4]. According to Fan [5], the estimates of the incidence rates and mortality increase annually in all regions of the world. It was estimated that 2.1 million new cases of breast cancer were identified and 0.6 million cases of deaths due to breast cancer occurred world-wide in 2018 [6, 7]. Mortality rate of breast cancer varies among different communities and countries, as shown in Figure 1.1.

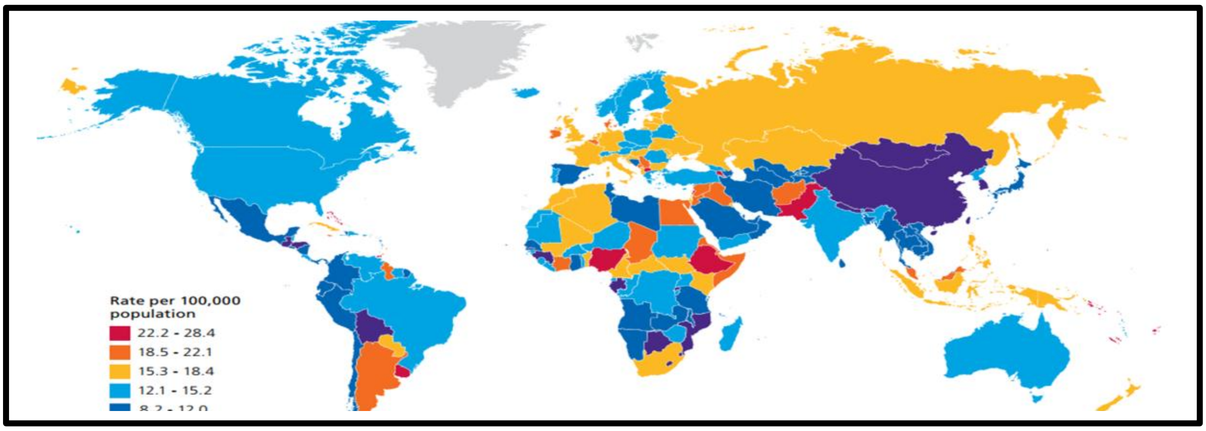

Figure 1.1: World-wide variation in female Breast Cancer mortality rate.
Source: American Cancer Society 2015.

The distribution of breast Cancer mortality rate varies significantly among different regions of the world, possibly due to different geographical, economic, social and cultural factors that affect the spread of breast cancer. The variation in mortality rate from region to region suggests that a one-model-fits-all, for predicting mortality, may not be possible and the need to adapt models to regional data is essential.

Breast cancer is one of the most prevalent public health problems in Zambia and unfortunately, diagnosis is often delayed, compounding the problem further since management of the cancer is difficult and expensive. According to World Health Organization (WHO) [8] report, breast cancer in women is the second most common cancer and cause of cancer-related deaths in Zambia among women after cervical cancer, accounting for 4.9% of cancer-related deaths. It is estimated that breast cancer in Zambia kills approximately 400 women each year [9]. Further, the two-year survival rate of breast cancer in women treated at Cancer Disease Hospital (CDH) remains below 50% [9]. Thus, breast cancer is an important health issue for Zambian women.

Breast cancer develops when cells in the breast tissue divide and grow without normal controls forming a mass called a tumor [10, 11]. This most commonly originates from the inner lining of milk ducts or the lobules that supplies the ducts with milk. While the overwhelming majority of cases occur in women world-wide, breast cancer could also occur in men [12, 13]. There are a number of ways of treating breast cancer depending on the stage of the cancer but these are mainly categorized into two types; local (Surgery and radiation) and systematic (chemotherapy and hormone therapy) [13]. Surgical treatment is aggressive, the goal is to remove as much of the breast and surrounding area as feasible. Surgical treatment is in two categories, breast-conserving surgery and mastectomy.

Several factors have been investigated to establish valid criteria for the objective evaluation of the prognosis of disease. A patient's survival is related to several prognostic factors including; number of positive lymph nodes, tumor size, ethnicity, socioeconomic status, type of surgery, histological grade, hormone receptor status and human epidermal growth factor receptor 2 (HER2) [14-17]. In addition, socio-demographic aspects, and those related to healthcare also influence, directly or indirectly, the survival of patients with breast cancer. Factors such as marital status, age at diagnosis, obesity, laterality (left or right) and family history of breast cancer are considered influential for predicting the survival of breast cancer patients [18, 17, 19, 20]. Although family history is a well-known etiological risk factor for breast cancer, its relationship with survival remains unclear. Studies that assessed the impact of family history on breast cancer mortality have yielded conflicting results [17, 21-23]. Marital status as a prognostic factor of breast cancer has, inconsistently been reported by several studies [24-26]. These gaps suggest that more investigations are required to ascertain the effect on the survival of patients especially after surgery.

HIV has always been associated with certain unusual cancers, especially Kaposi's sarcoma and cervical cancer [27]. However, due to the success of anti-retroviral treatment (ART), HIV is changing from being the major cause of death to a chronic condition with which the patients may survive into middle age and beyond. The change raises a question of what impact, if any, HIV may have on breast cancer patients after surgery with respect to survival. Identifying prognostic factors of breast cancer plays an important role in treatment and care of patients as it may improve delivery of health care to groups at risk. Identification of factors that predict survival after surgery might help to guide adjuvant treatment recommendations and help illuminate the natural history of the disease. However, the effects of factors such as age at diagnosis, stage of cancer, marital status, and family history are unclear and still challenging topics [28], thus there is a need for further investigation. Furthermore, the statistical significance of these factors in models vary due to a variety of issues, some of the issues may include environment and culture. Table 1.1 shows some of these variations in models that researchers have attempted.

Table 1.1: Risk factors of breast cancer mortality from some selected studies

		Momenyan [29] Logistic regression	Chang [30] Logistic regression	Alvarez-Banuelos [17] Cox regression
Factors in the model		**P-Value** (* indicates significant factors)		
1	Age at diagnosis	0.006*	0.003*	0.003*
2	Breast cancer stage	< 0.001*		
3	Breast feeding	0.3		
4	Education level		0.057	0.35
5	ER	0.008*	0.009*	0.035*
6	Family history	0.3	>0.05	0.045*
7	HER2 status		0.014*	0.029*
8	Histologic grade	< 0.001*		0.242
9	Lymph node status	< 0.001*	< 0.0001*	0.042*
10	Marital status	0.4	> 0.05	
11	Menopause			0.158
12	PR	0.006*	0.009*	0.053
13	Surgery type	< 0.001*	0.0001*	
14	Tumor grade		< 0.0001*	
15	Tumor size	< 0.001*	< 0.0001*	< 0.001*
16	Type of treatment			< 0.001*

Note: A blank P-Value for a particular model indicates that the factor was not in the model.

As alluded earlier, mortality rates of breast cancer vary among different communities and countries. In developed countries, breast cancer, often diagnosed at an early stage, allows for early treatment and reduction of mortality due to this disease [31]. In countries with limited resources however, locally advanced breast cancer is still common and has poor prognosis resulting into higher mortality rate [32]. The 5-year mortality rate of breast cancer patients after surgery tends to be higher in developing countries [18, 31]. A 5-year relative mortality for breast cancer patients was lower in United Kingdom, Canada, and Libya but higher in Algeria, South Africa, Gambia, Mali and Mongolia [19, 31, 32], as shown in Table 1.2.

Table 1.2: Mortality rates from different selected countries.

Reference	ACS [31]			Vanderpuye [32]			
Country	UK	Canada	Mongolia	Libya	Algeria	South Africa	Gambia
5-year mortality rates	19.0%	14.0%	43.0%	23.4%	41.2%	46.6%	88.1%

Mortality rate is the proportion of patients who have died after diagnosis or treatment within a given time period. Assessment of risk factors of mortality for breast cancer patients and calculating the mean life time in different groups of populations in order to improve management of patients is still a challenging topic for most epidemiologists and other researchers.

To investigate the risk factors associated with mortality of breast cancer patients, regression methods have become an integral component of any data analysis concerned with the explanation of relationship between a response variable and one or more explanatory variables [21, 24, 27, 29, 33]. One type of regression model in common use is the Cox regression model, used when the effect of covariates on the hazard ratio is of interest. However, the basis and the most important assumption underlying this model is the proportionality of hazard rates, which may not hold in some situations. Researchers have argued that when the proportional hazard (PH) assumption does not hold, it is improper to use the standard Cox PH model as it may entail serious bias and loss of power when estimating or making inference about the effect of a given prognostic factor on mortality [34]. Another regression model in common use is the logistic regression. The model is appropriate when the response variable is binary and the explanatory factors are either continuous or categorical.

As a statistical fitting model, it is widely used to model medical data because the methodology is well established and coefficients can have intuitive clinical interpretations. Generally, it is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables [35]. The central mathematical concept, on which the logistic regression is anchored, is the idea of the logit - the natural logarithm of the odds of an outcome.

The prognosis of breast cancer can differ according to geographic and ethnic factors [30]. Zambia has different environmental and cultural settings from the Western world, where a number of studies have been conducted. Zambian studies may give rise to different model results from those already tested. Studies regarding breast cancer epidemiology in Zambia are rare, and surveys conducted by the American Cancer Society and other studies have not included Zambia [31, 32]. Studies in Zambia regarding information based on mortality rate for breast cancer patients after surgery and its possible associated factors related with both patient's care and personal attributes are scanty. This motivated the need for this study, to examine factors which may play a role in predicting mortality for breast cancer patients after surgery and to assess their impact on survival five years after surgery or longer. The basis for choosing five years or longer is because five-year survival is a term doctors and researchers use as a benchmark. Further, improvements in 5-year survival are commonly regarded as a utile metric for measuring progress in combating cancer [29, 30]. Consequently, the investigation attempted to identify risk factors that affect survival of women with breast cancer after surgery with a hope of providing an insight into survival rate. The study used the logistic regression which allows an investigator to, simultaneously, examine the relationship between all possible risk factors. The list of possible risk factors included; demographic, clinical, HIV status, pathological factors, and the response factor was mortality status of breast cancer patients after surgery. The data utilized was accessed from the Lusaka Cancer Disease Hospital (CDH), Zambia.

1.3 Statement of the Problem

Breast cancer is the most frequently occurring disease among women, impacting over 1.5 million each year, and also causes the greatest number of cancer-related deaths worldwide [8, 36]. Breast cancer is the second most common cancer among women in Zambia after cervical accounting for about 361 deaths in 2018 [8, 37]. Studies that have examined prognostic factors of breast cancer mortality after surgery using statistical methods such as logistic and cox regression models evaluated both patient's care-related and personal attributes factors such as marital status, family history of breast cancer, tumor grade, tumor size, type of surgery, laterality, education level, hormone receptor status and lymph node status [5, 14, 17, 28]. The significance of these factors varied from model to model possibly due to variability in geographic characteristics. Zambia, having its own environmental and cultural settings, both personal and care-related factors such as histologic grade, breast cancer stage, alcohol consumption, marital status and others could present a different picture in terms of mortality, an aspect the researcher wanted to examine using the logistic regression model. Further, a 5-year mortality of breast cancer patients after surgery in Zambia was estimated using both a statistical model and the Actuarial life-tables.

1.4 Aim of the study

The aim of the study was to improve management of breast cancer patients by determining all possible factors associated with mortality of breast cancer after surgery from medical records at Lusaka Cancer Disease Hospital.

1.5 Objectives of the study

The objectives of the study were as follows:

- **1.5.1** To identify prognostic factors associated with mortality of breast cancer patients after surgery.
- **1.5.2** To estimate mortality rate of breast cancer patients after surgery over a 5-year period.

1.6 Research Hypothesis

Risk factors associated with mortality of breast cancer patients after surgery exist, and these include; age, lymph node status, HIV status and care-related factors and these factors vary in their impact from region to region.

1.7 Significance of the study

Prognostic factors for breast cancer mortality play an important role in predicting survival probabilities of breast cancer patients and are fundamental to the process of planning for cancer management as well as setting priorities for future research. Therefore, information generated from the present study could improve the management of breast cancer patients, and for strategic planning by the Ministry of Health (MoH), Non-Governmental Organizations (NGOs), donor agencies, and any other individuals concerned with breast cancer in Zambia. Further, the findings of the study may add to the existing body of knowledge and be used as a point of reference for future researchers.

1.8 Delimitation

The study was conducted at Lusaka Cancer Disease Hospital (CDH) in Zambia, hence accessed data on patients received treatment from CDH.

1.9 Operational Definitions

Cancer: uncontrolled growth of abnormal cells in the body.

Mortality: number of deaths occurring during a specific period of time.

Breast cancer: a malignant tumor that develops when cells in the breast tissue divide and grow without the normal controls on cell death and cell division.

Prognosis: a medical term for predicting the likely or expected development of a disease, including whether the signs and symptoms will improve or worsen or remain stable over time.

Diagnosis: identification of nature of illness or other problems by examination of the symptoms.

Prognostic factors: factors determined at the time of diagnosis that are associated with disease-free or overall survival and often used to predict the natural course of the disease.

Survival rate: the proportion of patients who survive for a specified period of time after diagnosis or treatment.

Surgery: directly removal of breast cancer tumor.

Chemotherapy: use of chemicals to kill breast cancer cells.

Radiotherapy: use of X-rays to kill breast cancer cells.

Hormone therapy: an adjuvant treatment to treat hormone receptor positive breast cancer.

Global level: term used to mean outside Africa.

Logistic regression model: a model used to examine the relationship between a binary outcome (dependent) variable and independent variables or factors.

Significant prognostic factors of death: factors associated with death caused by breast cancer, which are statistically determined from the multivariable analysis.

Environment: the surroundings or conditions in which a person, animal, or plant lives or operates.

Culture: the ideas, customs, and social behavior of a particular people or society.

1.10 Chapter Summary

This chapter gave the background to the study on breast cancer worldwide, Zambia, prognostic factors of mortality for breast cancer patients, and prognostic methods used to predict mortality and determine effective factors. The chapter also presented the statement of the problem, purpose of the study, research objectives and hypothesis, and significance of the study. Further, the chapter discussed delimitation and operational definition of terms used in the study.

CHAPTER TWO: LITERATURE REVIEW

2.1 Overview

The previous chapter highlighted the background to the study, statement of the problem, purpose of the study, research objectives, hypothesis, significance of the study and delimitation of the study. Further, the chapter also presented operational definition of terms used in the study. This chapter reviews some important literature on prognostic factors of mortality for breast cancer patients, observed female breast cancer mortality rate and methods used to predict mortality. The literature is based on Global, African and Zambian contexts.

2.2 Global trends of Breast Cancer Mortality

Empirical evidence from studies around the world have shown that the mortality rate of breast cancer is still high in both developed and developing nations [18, 31]. The American Cancer Society [31] reported that mortality of breast cancer for a 5-year period from 2005 to 2009 in some selected countries stood at 19% in the United Kingdom, 13% in Brazil, 13% in South Korea, 26% in Poland, 47% in South Africa, 40% in India and 40% in Algeria. Mortality rate for breast cancer patients vary widely and it is influenced by a number of factors which include; demographic, tumor size, hormone receptors status, human epidermal growth factor receptor 2 (HER2), factors related to both patient's care and personal attributes.

Studies around the world have been carried out to identify prognostic factors of mortality for breast cancer patients and further, to observe female breast cancer mortality rate over time. A retrospective study conducted in Iran by Momenyan [29], compared traditional statistical analysis and data mining techniques as research methods for identifying prognostic factors regarding survival time of patients with breast cancer after surgery. The study unveiled that age at diagnosis, histologic grade, axillary lymph node status and type of surgery were statistically significant with regard to the probability of death in patients with breast cancer after surgery. Another retrospective study by Zaid [19] in Saudi Arabia, assessed women who were diagnosed with breast cancer between the years 2011 and 2012 and whose follow-up was at least three years from the date of diagnosis. The study focused on determining overall observed 1 and 3-year survival rate of female breast cancer patients and to investigate the factors affecting survival rate. It was noted that histological type of tumor, metastases and type of treatment were statistically significantly associated with survival of patients diagnosed with breast cancer. Besides risk factors, the study also established that a 1-year and 3-year survival rates of breast cancer were 96% and 85%, respectively. In Waikato, New

Zealand, Seneviratne [33] examined prognostic factors of mortality for women diagnosed with breast cancer. A logistic regression model was used to link numerous factors and mortality. It was discovered that advanced stage, higher grade, absent hormone receptors (i.e., estrogen and progesterone) and HER-2 amplification were significantly associated with higher risks of breast cancer mortality. Nonetheless, the aforementioned studies utilized data of patients whose follow up was less than 5 years.

Chang [30] conducted a logistic model building study on mortality risk for Taiwanese women with breast cancer, to identify prognostic factors of breast cancer patients after surgery and to develop a more precise predictive mortality risk model. The findings of the study indicated that higher tumor grade, negative estrogen receptor (ER) status, negative progesterone receptor (PR) status, positive HER-2 status, increased tumor size, surgery type, age, menopause and axillary lymph nodes were positively associated with increased breast cancer mortality. In addition, the study noted that the 5-year survival probability for breast cancer was greater than 90% and for patients with stage III disease, the survival probability for 5 years was 84.33%. In Brazil, Balabram [14] evaluated the survival of patients with operable breast cancer stages I-III at a public hospital from 2001 to 2008 using Cox regression model. Factors such as tumor size, higher histologic grade, higher number of positive nodes and age older than 70 years were found to be associated with a shorter breast cancer specific survival. Moreover, the study revealed that a 5-year breast cancer specific survival for the entire cohort was 78.5%. However, the revealed studies excluded patients with stage IV, in this manner information on factors affecting mortality of these patients was not provided.

Lan [18] in Vietnam, examined the survival probability at 1-, 3-, and 5-years following diagnosis between 2001 and 2010 and determined prognostic factors for breast cancer mortality using Cox regression model. The study observed a decrease in survival probability over time following diagnosis. The study also reported that education level, marital status, stage at diagnosis and hormone therapy were prognostic factors of mortality for breast cancer patients. Further, observed overall survival rates at 1, 3 and 5 years were estimated to be 94%, 83% and 74%, respectively. Similarly, Fan [5] in Taiwan, developed a prognostic nomogram for identifying those factors which influence the 2 and 5-year survival chances of women who underwent breast cancer surgery from 2004 to 2007 using Cox regression model. The results showed that histological grade, tumor grade, lymph node status, metastasis (TNM staging), estrogen receptor status, progesterone receptor status and type of surgery were statistically significant predictive factors of overall survival. Further, the study observed that

2 and 5-year survival rates of breast cancer patients after surgery stood at 94.3% and 84.9%, respectively. Another important aspect from both studies was the determination of the observed overall 5-year survival rate, information which was not known in the case of Zambia. Further, patients of age less than 22 years and 35 years respectively were excluded from the said studies even though they are also at risk of breast cancer. Additionally, Lan's study found education level and marital status significant contrary to the findings of other reviewed studies [19, 29, 30].

A study done by Alvarez-Banuelos [17] in Mexico, aimed at identifying prognostic factors associated with survival in women with breast cancer using Cox regression, disclosed that tumor size, lymph node status, estrogen receptor status, HER2, disease stage, age at diagnosis, metastasis, treatment method and family history were important for survival as prognostic factors for breast cancer patients. Another critical feature from the abovementioned study was the determination of overall survival rate for women during a 5-year follow-up, which was determined to be 63%. Additionally, the study analyzed medical records of women who were diagnosed with breast cancer while our study utilized data of patients who have undergone surgery as primary treatment. More interestingly, family history was found to be significant contrary to other reviewed studies [29, 30].

In Japan, Kawaguchi et al. [38] conducted a study on factors associated with prolonged overall survival in patients with postmenopausal estrogen receptor-positive advanced breast cancer using real-world data: a follow-up analysis of the Japan Breast Cancer Research Group (JBCRG) Safari study. The study revealed that that younger age (< 60 years) and progesterone receptor negativity (PR-) were significantly correlated with prolonged overall survival. However, cox regression was used to identify the risk factors attributed to survival of breast cancer patients while in the present study, logistic regression will be used. Another point of variation was that patients who did not undergo surgery were also employed in the said study.

A summary of literature reviewed is given in Table 2.1.

Table 2.1: Summary of literature revealed

		Cox Regression (+ indicates factors in the model)					**Logistic Regression** (+ indicates factors in the model)		
No.	Studies	[5]	[14]	[17]	[18]	[19]	[29]	[30]	[33]
1	Age at diagnosis	+*	+*	+*	+	+	+*	+*	+
2	Histologic grade		+*	+			+*		
3	lymph node status	+*	+*	+*			+*	+*	
4	type of surgery		+*				+*	+*	
5	type of treatment				+*	+*			
6	metastases	+*		+*		+*			
7	tumor stage		+*			+			+*
8	tumor grade	+*				+		+*	+*
9	Estrogen receptor	+*		+*			+*	+*	+*
10	Progesterone receptor	+		+			+*	+*	+*
11	HER-2 status			+*				+*	+*
12	Menopause			+				+*	
13	Education level			+	+*	+		+	
14	Marital status				+*	+	+	+	
15	Stage at diagnosis				+*		+*		
16	Hormone therapy				+			+	
17	Family history			+*			+	+	
18	Laterality	+	+			+*			
19	Tumor size		+*	+*			+*	+*	
20	Neoadjuvant		+*	+*					
22	Ethnic			+	+				+*

Note: * indicates significant factors in the model; Fan [5], Lan [18], Balabram [14], Chang [30], Alvarez-Banuelos [17], Seneviratne [33], Zaid [19] and Momenyan [29].

2.3 Literature on Breast Cancer Mortality in Africa

In developed countries breast cancer is frequently diagnosed early allowing conservative treatment and reducing mortality due to this disease. In countries with limited resources, locally advanced breast cancer is still common and has poor prognosis. Africa and many low and middle-income countries have several challenges including poor health infrastructure, lack of population awareness, delayed health seeking behaviour and low levels of female education that have led to high mortality from breast cancer [39]. Breast cancer is the second most common cancer among African women after cervical cancer [40]. Breast cancer prevention and control in Africa is relatively limited owing, in part, to a lack of reliable epidemiologic risk factor data and information from which evidence-based interventions could have been developed [41]. Only a few epidemiologic studies on prognostic factors of mortality for breast cancer patients and on mortality rate prediction have been conducted in Africa. The potential risk factors that have been examined have mainly focused on incidence, risk factors and complications rather than mortality and its associated factors especially after surgery [41, 42].

In Guinea, Traore [20] conducted a study aimed at analyzing predictive factors associated with 5-year survival after breast cancer surgery using the Cox regression model. The outcome of the study revealed that body mass index (BMI), origin of patients (source of referral), radiotherapy and cancer relapse were risk factors independently associated with death due to breast cancer. The study also observed an overall 5-year mortality of 42.1% for breast cancer patients after surgery. This study replicated Traore's but used Logistic regression with additional factors which included HIV status and alcohol consumption.

A retrospective study conducted by Ngowa [43] in Cameroon aimed at estimating the survival rate of breast cancer in a group of patients followed up at the Yaounde General Hospital, showed a correlation between survival rates and factors such as cancer stage and the type of surgery. Further, a 5-year survival rate was found to be 30%. However, the aforesaid study used few factors in determining survival rates and its aim, different from the aim of this study, was not to identify risk factors which tend to influence survival rates of breast cancer patients. Another point of variance with this study was the exclusion of patients under 20 years of age who are also at risk of breast cancer.

Another retrospective study by Makanjuola [44] was carried out to examine 5-year survival from breast cancer cases diagnosed between 2005 and 2008 in Nigerian women using Cox regression. The study reported that cancer stage and treatment type were seen to be significantly associated with survival of breast cancer patients. In addition, the observed 5-year survival rate was estimated as 24.1%. Another study conducted by Gakwaya [12] in Uganda, aimed at investigating the 5-year overall survival of breast cancer patients, showed that stage at diagnosis, histological grade, age at diagnosis, availability and accessibility to treatment modalities were prognostic factors of mortality for patients with breast cancer following diagnosis. It was also observed that a 5- year overall survival rate, 56%, was lower compared with those of the United States 88%, Canada 86% and South African blacks 64%. Seedhom [45] in Egypt conducted a study aimed at determining breast cancer survival time and the association between breast cancer survival and socio-demographic and pathologic factors among women. The study included women who were diagnosed with breast cancer between 2005 and 2009 and followed up to 2010. The findings indicated that greater tumor size, higher grade, higher number of involved lymph nodes, type of treatment and type of surgery were significantly associated with mortality. Coghill et al. [46] investigated the role of HIV in cancer survival in Uganda and found that HIV-infected cancer patients experienced a more than two-fold increased risk of death during the year following cancer diagnosis compared to HIV-uninfected cancer patients [hazard ratio 2.28]. Nevertheless, the reviewed studies considered the population sample of patients diagnosed with breast cancer while the current study utilized data of patients who have undergone surgery as primary treatment.

In Morocco, Laamiri [47] evaluated the risk factors associated with breast cancer in different age groups and determined susceptible and protective factors. Using a Logistic regression model, early age of menarche, late menopause, family history of first degree, use of contraceptive pills were found to be positively associated with breast cancer. However, the study focused on risk factors of breast cancer rather than risk factors associated with mortality for breast cancer patients over a 5-year period after surgery.

In Tunisia, Bouzguenda et al [48] conducted a study aimed at identifying some determinants of survival in metastatic breast cancer. The results indicated that age less than 70 years, hormone-dependence of the tumor, less than two metastatic sites, no visceral metastases, disease free interval (DFI) greater than 24 months and surgery of the primary tumor were prognostic factors of survival. However, the study focused on breast cancer metastatic prognostic factors not on risk factors associated with mortality for patients with breast cancer after surgery.

2.4 Literature on Breast Cancer Mortality in Zambia

Studies carried out in Zambia have not paid much attention to identifying risk factors of mortality for patients with breast cancer after surgery. Mumba [49] conducted a study aimed at determining the knowledge, attitude and practice of breast self-examination (BSE), among women in Roan Township in Luanshya, Zambia. The findings of the study indicated that the proportion of women who had poor knowledge on BSE was 48.1% and those who had positive knowledge (how to carry out breast self-examination) were 51% of which 39.3% had average knowledge and only 12.5% had good knowledge. In addition, television was cited as the main source of information on BSE by 34% of the respondents. The study observed that the attitude toward BSE was good with 74.4% of respondents having a positive attitude. Similarly, Mulenga [50] conducted a cross sectional study aimed at determining the knowledge, attitude and practice of women attending gynecological clinic at Ndola Teaching Hospital. The study revealed that 70.7% of the participants had inadequate knowledge. It also found that most of the respondents (88.7%) had negative attitude towards breast cancer and breast self-examination. In addition, the study established that 84% of the participants had poor practices (did not know how to perform BSE). However, the focus of the cited studies was on BSE, not on risk factors of mortality. Thus, in Zambia little is known about prognostic factors of mortality for breast cancer patients and mortality projection over a 5-year period after surgery.

2.5 Synthesis of Literature Gap

While there has been much research regarding prognostic factors for breast cancer, the majority of these studies, however, were from developed countries. Numerous studies reviewed indicated that tumor characteristics are the major factors which tend to influence mortality of patients with breast cancer after surgery. Although several studies have been conducted to determine survival rate and its associated factors, many of the findings differed in a number of respects including study populations, which differ in composition and cultural practices from country to country, an issue that had received little attention. Interestingly, not every study identified the same set of risk factors leading to variation in prognosis following breast cancer surgery over a period of 5 years or longer. Consequently, more investigation was required to ascertain the impact of factors such as HIV status and alcohol consumption on mortality of breast cancer patients over a 5-year period after surgery. Further, it was observed that in Zambia, studies on mortality rate of breast cancer patients after surgery and its associated factors over a 5-year period had been inadequate, thereby creating a knowledge gap this study intended to address.

2.6 Summary of Literature Review

This chapter reviewed literature of studies at global, Africa and Zambia levels. From numerous studies reviewed, it was observed that factors affecting prognosis of patients with breast cancer varied from region to region. Consequently, survival rates of breast cancer patients after surgery or diagnosis also varied. It was a matter of interest to replicate the studies in the Zambian context.

CHAPTER THREE: METHODOLOGY

3.1 Overview

The previous chapter gave a review of some important literature on breast cancer studies from International, African and Zambian contexts with respect to mortality rate and its associated factors. The current chapter presents the methodology used in the study that includes the study design, study population, sample size, data collection, recorded variables and statistical analysis ending with ethical considerations.

3.2 Study Design

A retrospective study design was utilized to identify risk factors associated with breast cancer mortality after surgery and to estimate a 5-year mortality rate using data from the Lusaka Cancer Disease Hospital (CDH), Zambia.

3.3 Study Population

The sampling frame constituted all women who had surgery as the primary treatment for breast cancer and confirmed by a surgeon or specialists between 2013 and 2018 at CDH in Lusaka district, Zambia. In addition, women meeting the following criteria were included in the study; patients who were older than 18 years of age, underwent surgery and whose end-point result was either death or survival. Further, Patients who were followed up to 31st December 2019 from the date of surgery. Data for any subject lost to follow up or dying from other causes during this period were treated as having been censored.

3.4 Sample Size

A total of 466 patients were randomly selected from those that met the inclusion criteria given in Section 3.3. The sample size was determined as follows; let the end-point result be denoted by Y ($Y = 1$, indicating death; $Y = 0$, indicating survival) and X denote HIV status ($X = 1$, HIV positive; $X = 0$, HIV negative) as our primary covariate of interest, using Hsieh [51] formula given by:

$$n = \frac{\left(z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{\pi}} + z_{1-\beta}\sqrt{P_1(1-P_1) + \frac{P_2(1-P_2)(1-\pi)}{\pi}} \right)^2}{(P_1 - P_2)^2(1-\pi)(1-\rho^2)} \qquad (3.1)$$

Where

- $P_1 = Pr(Y = 1|X_1 = 0)$ and $P_2 = Pr(Y = 1|X_1 = 1)$ are event rates for a patient who is HIV negative and a patient who is HIV positive respectively,

- π is the proportion of the sample with $X_1 = 1$,
- P is the overall event rate given by $P = (1 - \pi)P_1 + \pi P_2$,
- α the probability of type I error,
- β is the probability of type II error,
- Z_r is the percentile value of the standard normal distribution and,
- ρ is the multiple correlation coefficient of X_1 with $X_2, X_3, \ldots, X_m$ representing other covariates.

Using a balanced sampling design and setting, π = 0.5, Type I Error at 5% and Type II Error at 20% we obtain $z_{1-\alpha/2} = 1.96$, a threshold value for Type I Error and $z_{1-\beta} = 0.845$, a threshold value associated with Type II Error. We set the odds of dying from breast cancer to increase by 70% for HIV infected patients than non-infected yielding an OR of 1.5. Now taking $\rho = 0.1$ and setting $P_1 = 0.4$, $P_2 = \frac{P_1 \times OR}{1 - P_1 + OR \times P_1} = 0.53$ and $P = (1 - 0.5)0.4 + 0.5(0.53) = 0.47$. Thus, substituting these values in the above equation yields $n = 466$ as the maximum sample size. The value of ρ was taking as 0.1 because the data to be collected from CDH was not much regarding patients who had undergone surgery as alluded by the Hospital Management. However, according to Hesieh [51], the value of ρ ranges from 0 to 0.7, depending on the availability of the data. If the data is not enough, the value of 0.1 can be used as in the case of our study. Regarding the choice of $\beta = 20\%$, a power of 80% is recommended when using Hsieh formula for sample size for logistic regression [54]. The higher the statistical power, the greater the chance of avoiding an error. It is often recommended that the statistical power should be set to at least 80% prior to conducting any testing [51, 54].

However, our study employed 233 breast cancer patients yielding 50% of the maximum sample size as calculated above due to the exclusion criteria. In addition, some patients were excluded from the study because their medical files had missing data in a number of variables.

3.5 Data Collection

The study analyzed data obtained from CDH in Lusaka district, Zambia. The CDH is the cancer treatment centre for all cancers in the nation. CDH has maintained a register of all cancer patients since its establishment in 2007. The CDH database provides detailed administrative data regarding healthcare services, including outpatient visits, prescriptions, hospitalizations and is comprehensive. The required data was retrieved from the hospital registry. A data collection form was designed (Excel sheet) to store patients records and information needed which included factors related to patient's care, personal attributes and other important factors for predicting the outcome of interest. For each patient, data was collected from the date of cancer surgery until 31st December 2019, the end of follow-up period, and whose end-point result was either death or survival. Further, time in months when they died or got lost or the last seen date was also noted from patients' medical records.

3.6 Recorded Variables

All relevant variables recorded by CDH that included socio-demographic data and clinic-pathological data were collected and assessed to identify significant ones with respect to mortality of women with breast cancer after surgery. Socio-demographic data collected included family history, contraceptive use, alcohol consumption which were classified as either yes or no, age, marital status (married or unmarried), number of children (Nulliparous or gave birth), date last seen or outcome of end-point result. Unmarried women included those who had never been married before, those who were divorced and the widowed.

In addition, clinic-pathological data such as laterality (left or right), histologic grade (grades I, II or III), type of surgery (mastectomy or lumpectomy), breast cancer stage (I/II or III/IV), tumor size (< 5 cm or ≥ 5cm) were collected. In addition, lymph node, HIV status, ER status, PR status and human epidermal growth factor receptor 2 (HER2) expression which were classified as either positive or negative were also collected from patients' medical records. Body weight, measured in kilograms, and height, measured in meters, were used to calculate Body Mass Index (BMI). In this study, BMI was used to determine obesity and it was categorized as not obese or obese/overweight.

3.7 Statistical Analysis

This section of the chapter outlines how analysis was carried out.

3.7.1 Categorical Response Data

In biomedical sciences, categorical scales measure outcomes such as whether a medical treatment is successful or not. Although categorical data are common in the social and biomedical sciences, they are by no means restricted to those areas. They frequently occur in epidemiology for measuring responses such as whether a patient survives or dies from a disease such as breast cancer [52, 53].

3.7.2 Multinomial Distribution

A categorical response variable with k possible outcomes such as severity of disease with the outcomes; mild, moderate, or severe are often assumed to follow a multinomial distribution. Specifically, we assume that a categorical response variable Y has k categories or outcomes and that in a sample of size n, Y_i responses fall in category i with probability π_i, so that $\underline{Y} = (Y_1, Y_2, \dots, Y_k) \sim Mult(n, \underline{\pi})$ where $\sum_{i=1}^{k} \pi_i = 1$, and $\sum_{i=1}^{k} Y_i = n$. When $k = 2$ the multinomial distribution is referred to as a binomial distribution. In that case $\underline{Y} = (Y_1, Y_2) \sim Bin(n, \{\pi, 1 - \pi\})$, where Y_1 is often referred to as a success and Y_2 as a failure. Further, $pr(Y = Y_1) = \pi$ and $pr(Y = Y_2) = 1 - \pi$. In this study Y_1 represents death of breast cancer patient after surgery while Y_2 represents survival.

Our interest was to link π with possible risk factors of breast cancer mortality after surgery using the logistic regression model, a transformation of π which is discussed later.

3.7.3 Contingency Table

If we have a dichotomous response variable Y and K categorical variables $X_1, X_2, \dots, X_K$, each with c_j categories for $j = 1, 2, \dots, K$, it is possible to display the data resulting from a classification of n subjects on Y and the K variables in a two-way contingency table as shown in Table 3.1.

Table 3.1: Hypothetical cross tabulation of a response variable Y and K categorical variables

Categorical variables					**Response Variable (Y)**	
X_1	X_2	. . .	$X_{(k-1)}$	X_k	Y_1	Y_2
C_{11}	C_{21}	. . .	$C_{(k-1\ 1)}$	$C_{(k\ 1)}$	N_{11}	N_{12}
C_{12}	C_{22}		$C_{(k-1\ 2)}$	$C_{(k\ 2)}$	N_{21}	N_{22}
. . .	. . .	. . .			. . .	. . .
C_{1k}	C_{2k}	. . .	$C_{(k-1\ k)}$	$C_{(k\ k)}$	N_{k1}	N_{k2}

Analysis of categorical data involves the use of data which can be displayed in a similar table. Suppose there are two categorical variables denoted by X and Y. Let r and c denote the number of categories of X and Y, respectively. A rectangular table having r rows and c columns for the categories of X and Y respectively, has cells that display the rc possible combinations of outcomes for X and Y [52]. A table that displays counts of outcome in the cells is called a contingency table [52]. In other words, a contingency table, sometimes called a two-way frequency table, is a tabular array of frequency counts. In modelling the response, a function of the *X* variables (predictors), such as in logistic regression, dummy variables are used in place of the predictors.

3.7.4 Generalized Linear Models (GLMs)

Generalized linear models (GLMs) extend ordinary regression models to encompass non-normal response distributions and modelling functions of the mean [52]. The logistic regression model is an example of a broad class of models known as Generalized Linear Models (GLMs). Among other examples include linear regression and Poisson regression. There are three components to a GLM, a random component, a systematic component and a link function described as follows:

- *Random component:* Identifies the response variable *Y* and its probability distribution.
- *Systematic component:* Specifies the explanatory variables as a linear model of a function of the mean vector,

$$\eta_i = \sum_j \beta_j x_{ij} \qquad (3.2)$$

- *Link Function,* η: Connects the random and the systematic component.

$$g(\mu_i) = \eta_i \qquad (3.3)$$

3.7.5 Cox Regression Model

The Cox proportional-hazards model is essentially a statistical regression model commonly used in medical research for investigating the association between the survival time of patients and one or more predictor variables. It builds a predictive model for time-to-event data. The model produces a survival function that predicts the probability that the event of interest occurs at a given time *t* for given values of the predictor variables [53]. The purpose of the model is to evaluate, simultaneously, the effect of several factors on survival. In other words, it allows to examine how specified factors influence the rate (hazard rate) of a particular event happening (e.g., death) at a particular point in time.

3.7.6 Logistic Regression Model

Regression techniques are versatile in their application to medical research because they can measure associations, predict outcomes, and control for confounding effects of variables. One such technique, logistic regression is an efficient and powerful tool for analyzing the effect of a group of independent variables on a binary outcome by quantifying each independent variable’s unique contribution. It builds a model between the probability of the binary response event and predictor variables by the logistic function [29]. In general, this model is used to model the outcome of a categorical dependent variable.

Logistic regression and linear regression model differ in that the outcome variable in logistic regression is binary or dichotomous. The difference is reflected both in the choice of parametric model and in the assumptions made [54]. If the response variable is discrete, it cannot be modeled directly by linear regression. Thus, instead of predicting point estimate of the event itself, it builds the model to predict the odds of its occurrence. While logistic regression is a very powerful modeling tool, it assumes that the response variable (the log odds) is linear in the coefficients of the predictor variables. It is the most popular multivariable method used in health sciences because it determines the impact of multiple independent variables simultaneously on the dependent variable [54].

3.7.7 Assumptions of Logistic Regression Model

According to Park [55] and Bewick [56]:

- Logistic regression does not need a linear relationship between the dependent and independent variables.
- The dependent variable must be a dichotomous (2 categories).
- Logistic regression requires each observation and the error terms to be independent.
- The independent variables are not normally distributed, nor linearly related, nor of equal variance within each group.
- Logistic regression assumes linearity of each of independent variables and log odds.
- Large samples are needed because maximum likelihood coefficients are large sample estimates.
- Error terms (the residuals) do not need to be multivariate normally distributed.

3.7.8 Multiple Logistic Regression

Like ordinary regression, logistic regression extends to models with multiple explanatory variables [56]. In a logistic regression model, the random component for the (success, failure) outcomes has a *binomial distribution*. The link function is the logit function, which is defined as the log odds of success and symbolized by "*logit(π)*." Logistic regression models are often called *logit models* [55]. Whereas π is restricted to the range [0, 1], the logit can be any real number.

The method of including variables in the model can be carried out in a stepwise manner going forward or backward, testing for the significance of inclusion or elimination of the variable at each stage. The tests for inclusion are based on the change in the likelihood resulting from including or excluding the variable [54].

3.7.8.1 Parameter coding for categorical variables

In our model, we used both continuous and categorical variables. Dummy variables were generated for categorical variables. If a categorical variable *X* has *k* levels (categories), then *k -1* dummy variables were generated to be included in the logistic regression model. The dummy variables take values *0* and *1*. Let $X_1, X_2, \ldots, X_k$ be the dummy variables of *X* corresponding to levels *1, 2, . . . , k*, respectively. If X_r is the reference group then only X_p such that $p \neq r$, will appear in the model but X_r will not appear. X_p takes the value 1 if *X* is in level *p*, it has the value zero if *X* is not in level *p*, this applies to all other dummy variables with the exception of X_r. When *X* is in level *r* then all the other $k-1$ dummy variables assume the value *0*.

3.7.8.2 The Model

Let us consider the general logistic regression model with multiple explanatory variables. Denote the *p* predictors for a binary response *Y* by $\underline{X} = (X_0, X_1)'$, where $\underline{X}_0' = (X_1, X_2, \ldots, X_k)$ represent factors that have been proven to be significantly associated with higher mortality for patients similar to those defined in our study, and $X_1' = (X_{k+1}, X_{k+2}, \ldots, X_p)$ represent new factors we intended to investigate to determine their significance in relation to mortality for patients in our study. Let $\pi(\underline{X})$ and $1 - \pi(\underline{X})$ be the probabilities that $Y = 1$ (indicating death due to breast cancer) and $Y = 0$ (indicating survival from breast cancer) respectively, then

$$\pi(\underline{X}) = Pr(Y = 1|X_0, X_1) \quad (3.4)$$

$$1 - \pi(\underline{X}) = Pr(Y = 0|X_0, X_1) \quad (3.5)$$

The ratio of the two probabilities is called the odds:

$$odds(\underline{X}) = \frac{\pi(\underline{X})}{1 - \pi(\underline{X})} \quad (3.6)$$

which is a positive quantity and large values of $odds(\underline{X})$ indicate a strong association between the response and predictors, the *X* variables or factors. The model for the log odds called the logistic regression is given as:

$$g(\underline{X}) = ln\left(\frac{\Pr(Y = 1|X_0, X_1)}{\Pr(Y = 0|X_0, X_1)}\right) = ln\left(\frac{\pi(\underline{X})}{1 - \pi(\underline{X})}\right) = \underline{X}_0'\underline{\beta}_0 + \underline{X}_1'\underline{\beta}_1 \quad (3.7)$$

Equation (3.7) yields:

$$odds(\underline{X}) = \frac{\pi(\underline{X})}{1 - \pi(\underline{X})} = e^{\underline{X}_0'\underline{\beta}_0 + \underline{X}_1'\underline{\beta}_1} \quad (3.8)$$

as model-based odds of how many more times we are likely to observe a death (Y = 1) as opposed observing a survival (Y = 0) when the X factors are taken into account. The probability of observing a death associated with the X factors can be worked out as:

$$\pi(x) = \Pr(Y = 1|X_0, X_1) = \frac{e^{\underline{X}_0'\underline{\beta}_0 + \underline{X}_1'\underline{\beta}_1}}{1 + e^{\underline{X}_0'\underline{\beta}_0 + \underline{X}_1'\underline{\beta}_1}} \quad (3.9)$$

The vector of parameters $\underline{\beta}_i$ refers to the effect of $\underline{X}_i$ on the log odds that *Y = 1* (a patient dies from breast cancer) controlling for the other $\underline{X}_j$ covariates.

3.7.8.3 Parameter Estimation

The goal of logistic regression is to estimate the *p + 1* unknown parameters $\beta' = (\underline{\beta}_0, \underline{\beta}_1)$ in equation (3.7). This is done through maximum likelihood estimation which entails finding the set of parameters for which the probability of the observed data is greatest [55, 57]. The maximum likelihood equation is derived from the probability distribution of the dependent variable.

For a set of observations in the data $(x_i; y_i)$, $\pi(x_i)$ and $1 - \pi(x_i)$ are the contributions to the likelihood function, where $y_i = 1$ and $y_i = 0$ respectively, where *i = 1, 2, 3, …, n*. Thus, the contribution equation to the likelihood function for only one set of observation $(x_i; y_i)$ is given as:

$$\omega(x_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad (3.10)$$

The likelihood function has the same form as the probability density function, except it expresses the values of β in terms of known, fixed values of *y*. Now, since the observations are assumed to be independent of each other, we can multiply their likelihood contributions to obtain the complete likelihood function as follows;

$$L(\beta|y) = \prod_{i=1}^{n} \omega(x_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \quad (3.11)$$

where $\pi(x_i) = \pi_i$. According to Czepil [58], the maximum likelihood estimates are the values for β that maximize the likelihood function in equation (3.11). Now, from equation (3.11), the equation to be maximized can be written as:

$$\prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1-\pi_i)^1 \qquad (3.12)$$

Taking exponential to both sides of equation (3.7) yields,

$$\left(\frac{\pi_i}{1-\pi_i}\right) = e^{\underline{X}_{i0}'\underline{\beta}_0 + \underline{X}_{i1}'\underline{\beta}_1} \qquad (3.13)$$

Here, $\underline{X}' = \left(\underline{X}_{i0}', \underline{X}_{i1}'\right) = \left(X_{i1}, X_{i2}, \ldots, X_{ik}, X_{ik+1}, X_{ik+2}, \ldots, X_{ip}\right)$, so that in equation 3.13, we have:

$$\left(\frac{\pi_i}{1-\pi_i}\right) = e^{\underline{X}_{i0}'\underline{\beta}_0 + \underline{X}_{i1}'\underline{\beta}_1} = e^{\beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{ik}\beta_k + X_{ik+1}\beta_{k+1} + \cdots + X_{ip}\beta_p}$$

$$= e^{\sum_{j=0}^{k} X_{ij}\beta_j + \sum_{j=k+1}^{p} X_{ij}\beta_j}$$

$$= e^{\beta_0 + \sum_{j=1}^{k} X_{ij}\beta_j + \sum_{j=k+1}^{p} X_{ij}\beta_j}$$

$$= e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}$$

Where $\sum_{j=1}^{p} X_{ij}\beta_j = \sum_{j=1}^{k} X_{ij}\beta_j + \sum_{j=k+1}^{p} X_{ij}\,\beta_j$

Thus,

$$\left(\frac{\pi_i}{1-\pi_i}\right) = e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j} \qquad (3.14)$$

Solving for π_i, gives

$$\pi_i = \frac{e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}} \qquad (3.15)$$

Thus, substituting equations (3.14) and (3.15) in equation (3.12) yields,

$$\prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1-\pi_i)^1 = \prod_{i=1}^{n} \left(e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}\right)^{y_i} \left(1 - \frac{e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}\right)^1 \qquad (3.16)$$

Therefore, simplifying equation (3.16) further yields,

$$\prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1-\pi_i)^1 = \prod_{i=1}^{n} \left(e^{y_i\beta_0 + y_i \sum_{j=1}^{p} X_{ij}\beta_j}\right) \left(1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}\right)^{-1} \qquad (3.17)$$

Thus, the maximum likelihood function becomes,

$$L(\beta) = \left(e^{\beta_0 \sum_{i=1}^{n} y_i + \left(\sum_{i=1}^{n} \sum_{j=1}^{p} X_{ij}\beta_j y_i\right)}\right) \prod_{i=1}^{n} \left(1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}\right)^{-1} \qquad (3.18)$$

In equation (3.18), β is the collection of parameters $\beta_0, \beta_1, \dots, \beta_p$ and $L(\beta)$ is the likelihood function of β. The maximum likelihood estimates (MLE's) $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ can be obtained by calculating the values of β which maximizes $L(\beta)$.

Thus, taking the natural log of equation (3.18) yields the log likelihood function:

$$l(\beta) = \ln\big(L(\beta)\big) = \ln\left[\left(e^{\beta_0 \sum_{i=1}^{n} y_i + (\sum_{i=1}^{n} \sum_{j=1}^{p} X_{ij}\beta_j y_i)}\right) \prod_{i=1}^{n} \left(1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}\right)^{-1}\right]$$

or

$$l(\beta) = \beta_0 \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \sum_{j=1}^{p} X_{ij}\beta_j \, y_i - \sum_{i=1}^{n} ln\left(1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}\right) \qquad (3.19)$$

The critical points of a function (maxima and minima) occur when the first derivative equals 0. If the second derivative evaluated at that point is less than zero, then the critical point is a maximum. Thus, finding the maximum likelihood estimates requires computing the first derivative and second derivatives of the likelihood function. Therefore, to find the critical points of the log likelihood function, we set the first derivative with respect to each β equal to zero. Thus, differentiating equation (3.19) with respect to β_0 yields;

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}$$

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^{n} (y_i - \pi_i) = 0 \qquad (3.20)$$

Also, differentiating equation (3.19) with respect to β_j, yields;

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} y_i - \sum_{i=1}^{n} x_{ij} \frac{e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j}}$$

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} y_i - \sum_{i=1}^{n} x_{ij} \pi_i$$

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} x_{ij} [y_i - \pi(x_i)] \qquad (3.21)$$

for *j = 1, 2, …, p.*

Therefore, the maximum likelihood estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ for $\beta_0, \beta_1, \dots, \beta_p$ can be found by setting each of the equations (3.20) and (3.21) equal to zero and solving for each $\beta_j, j = 0, 1, \dots, p$. Thus, the likelihood functions that result may be expressed as follows:

$$\sum_{i=1}^{n}(y_i - \pi_i) = 0$$

and

$$\sum_{i=1}^{n} x_{ij}[y_i - \pi(x_i)] = 0$$

However, solving a system of nonlinear equations is not easy, the solution cannot be derived algebraically as it can in the case of linear equations as noted by Bewick [54] and Czepiel [57]. The solution must be numerically estimated using an iterative process. Therefore, to find the MLE $\hat{\beta} = \left(\underline{\hat{\beta}_0}\,, \underline{\hat{\beta}_1}\right)$ for $\beta = \left(\underline{\beta_0}, \underline{\beta_1}\right)$, where $\underline{\hat{\beta}_0} = (\hat{\beta}_0,\ \hat{\beta}_1\,, \ldots, \hat{\beta}_k)$ and $\underline{\hat{\beta}_1} = (\hat{\beta}_{k+1},$ $\hat{\beta}_{k+2}\,, \ldots, \hat{\beta}_p)$, iterative computation methods such as Newton-Raphson are used [57].

3.7.8.4 Interpretation of Regression Coefficients

Each regression coefficient describes the size of contribution of the corresponding predictor variable to the outcome. This effect is commonly measured by the use of odds ratio of the predictor variable, which represents the factor by which the odds of an outcome change for a unit change in the predictor, keeping the other variables constant. Each estimated coefficient is the expected change in the log odds of dying from breast cancer for a unit increase in the corresponding predictor variable holding the other predictor variables constant.

3.7.9 Statistical Tests: Odds Ratio

In equation 3.8, we have the ability to test the significance of a single factor X_i keeping the rest constant. Let

$$Odds\left(X_i, \underline{X_{i-}}\right) = \frac{\pi(X_i, \underline{X_{i-}})}{1 - \pi(X_i, \underline{X_{i-}})} = e^{\beta_i X_i + \underline{X_{i-}'}\underline{\beta_{i-}}} \qquad (3.22)$$

where $\underline{X_{i-}}$ is a set of all other variables with the exception of X_i. If X_i is a 0, 1 variable as in the case with dichotomous factors, then we have:

$$Odds\left(X_i = 1, \underline{X_{i-}}\right) = \frac{\pi(X_i, \underline{X_{i-}})}{1 - \pi(X_i, \underline{X_{i-}})} = e^{\beta_i + \underline{X_{i-}'}\underline{\beta_{i-}}} \qquad (3.23)$$

For $X_i = 1$, representing the presence of the factor X_i. Similarly, we have:

$$Odds\left(X_i = 0, \underline{X_{i-}}\right) = \frac{\pi(X_i, \underline{X_{i-}})}{1 - \pi(X_i, \underline{X_{i-}})} = e^{\underline{X_{i-}'}\underline{\beta_{i-}}} \qquad (3.24)$$

The ratio of (3.23) and (3.24) is referred to as odds ratio with respect to X_i, large values indicate a significant association of Y with the presence of X_i. Specifically, we have:

$$Odds\ Ratio\ (X_i, \underline{X}_{i-}) = OR(X_i, \underline{X}_{i-}) = \frac{e^{\beta_i + \underline{X}_{i-}' \underline{\beta}_{i-}}}{e^{\underline{X}_{i-}' \underline{\beta}_{i-}}} = e^{\beta_i} \quad (3.25)$$

The significance of the factor X_i is determined through the asymptotic distributions of either the estimate of the odds ratio or the estimate of β_i yielding a test such as Wald tests.

If we let X_a be the age variable and $\underline{X}_{a-}$ be the collection of all other variables minus the age variable, the odds ratio that compares mortality between those patients aged X_a and those aged $X_a + 5$ for example, can be calculated as:

$$OR(X_a + 5, \underline{X}_{a-}) = \frac{e^{\beta_a (X_a+5) + \underline{X}_{a-}' \beta_{a-}}}{e^{\beta_a X_a + \underline{X}_{a-}' \beta_{a-}}} = e^{5\beta_a} \quad (3.26)$$

A 5-year mortality of those aged X_a now can be calculated as:

$$\pi(X_a + 5, \underline{X}_{a-}) = \Pr(Y = 1 | X_a + 5, \underline{X}_{a-}) = \frac{e^{\beta_a (X_a+5) + \underline{X}_{a-}' \beta_{a-}}}{1 + e^{\beta_a (X_a+5) + \underline{X}_{a-}' \beta_{a-}}} \quad (3.27)$$

3.7.10 Statistical Tests: Wald Statistic

The Wald test is used to assess the contribution of individual predictors or the significance of individual coefficients in logistic regression [55, 56]. The Wald statistic W_i is calculated as:

$$W_i = \frac{\hat{\beta}_i^2}{[\widehat{SE}(\hat{\beta}_i)]^2} \quad (3.28)$$

where $\hat{\beta}_i$ represents the estimated coefficient of β and $\widehat{SE}(\hat{\beta}_i)$ denotes a model-based estimator of the standard error of the respective parameter, which is given by

$$\widehat{SE}(\hat{\beta}_i) = \left[\widehat{Var}(\hat{\beta}_j)\right]^{1/2} \quad (3.29)$$

where $\widehat{Var}(\hat{\beta}_j)$ is the variance of estimators, $\hat{\beta}_j$.

Under the null hypothesis H_0: $\beta_i = 0$, the quantity (3.28) follows a chi-square distribution with one degree of freedom.

According to Hosmer-Lemeshow [54], the method of estimating the variances and covariances of the estimated coefficients follows from well-developed theory of maximum likelihood estimation. The theory states that the estimators are obtained from the matrix of second partial derivatives of the log likelihood function [54, 57]. These partial derivatives have the following general form:

$$\frac{\partial^2 l(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}^2 \pi_i (1 - \pi_i) \quad (3.30)$$

and

$$\frac{\partial^2 l(\beta)}{\partial\beta_j \partial\beta_l} = -\sum_{i=1}^{n} x_{ij} x_{il} \pi_i (1-\pi_i) \qquad (3.31)$$

for *j, l = 0, 1, 2, ..., p,* where π_i denotes $\pi(x_i)$. Now, let the $(p+1)\times(p+1)$ matrix containing the negative of the terms given in equation (3.30) and (3.31) be denoted as $I(\beta)$. The variances and covariances of the estimated coefficients are obtained from the inverse of this matrix denoted $Var(\beta) = I^{-1}(\beta)$. The notation $Var(\beta_j)$ is used to denote the j^{th}diagonal element of this matrix, which is the variance of $\hat{\beta}_j$. The estimators of the variance, denoted $\widehat{Var}(\hat{\beta})$, are obtained by evaluating $Var(\beta)$ at $\hat{\beta}$.

3.7.11 Statistical Tests: The Likelihood Ratio Test

A logistic regression model with *p* independent variables (full model or model with variables) is said to provide a better fit to the data if it demonstrates an improvement over the model with fewer variables (reduced model or null model or model with no variables) [55]. Improvement is determined by carrying a test utilizing the likelihood ratio test. Let *L(Full)* represent the likelihood of the full model and *L(Reduced*) represent the likelihood of a reduced model. The number of β's in the full model is *p*, while the number of β's in the reduced model (null model) is *r*. Thus, the number of β's being tested in the null hypothesis above is *p – r = p* (since *r = 0*). Now, we have the full model as:

$$ln\left(\frac{\pi(\underline{X})}{1-\pi(\underline{X})}\right) = \underline{X}_0'\underline{\beta}_0 + \underline{X}_1'\underline{\beta}_1 = \sum_{j=0}^{p} X_{ij}\beta_j = \beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j \qquad (3.32)$$

Where $\sum_{j=0}^{k} X_{ij}\beta_j + \sum_{j=k+1}^{p} X_{ij}\,\beta_j = \sum_{j=0}^{p} X_{ij}\beta_j$, and the reduced model or the null model is given as;

$$ln\left(\frac{\pi(\underline{X})}{1-\pi(\underline{X})}\right) = \beta_0 \qquad (3.33)$$

Further, let $\underline{\beta}$ represent the set of coefficients present in the full model but not in the reduced model, i.e, the coefficients of the extra variables added to the reduced or null model in order to achieve the full model. The full model is an improvement over the reduced or null model if the null hypothesis below is rejected.

$$H_0: \underline{\beta} = \underline{0} \text{ versus } H_1: \underline{\beta} \neq \underline{0}$$

To test the hypotheses above, a Chi-square test is carried out using the G statistic.

$$G = \chi^2(p) = -2ln\frac{L(Reduced)}{L(Full)} \quad (3.34)$$

where k is the number of parameters in $\underline{\beta}$.

In our study, there were three tests of interest stated through the hypotheses below.

(i) $H_0: \underline{\beta_0} = \underline{0}$ vs $H_1: \underline{\beta_0} \neq \underline{0}$

(ii) $H_0: \underline{\beta_1} = \underline{0}$ vs $H_1: \underline{\beta_1} \neq \underline{0}$

(iii) $H_0: \underline{\beta} = \underline{0}$ vs $H_1: \underline{\beta} \neq \underline{0}$

In (i), we tested the significance of factors, $\underline{X_0'} = (X_1, X_2, \ldots, X_k)$, that have been proven to have a significant association with breast cancer mortality after surgery. In (ii), we tested the significance of factors, $\underline{X_1'} = (X_{k+1}, X_{k+2}, \ldots, X_p)$, that have not been captured in literature as prognostic factors of breast cancer mortality after surgery. Finally, in (iii), the significance of the totality of the prognostic covariates of breast cancer mortality in (i) and (ii), $\underline{X}' = (\underline{X}_0, \underline{X}_1)$, was tested.

The distribution of "G" statistics given in equation (3.34) is a chi-square with k, $p - k$ and $p - r$ degrees-of-freedom for (i), (ii) and (iii) respectively, where degrees of freedom correspond to the number of covariates in the logistic regression equation for the appropriate model. This is a measure of how well all of the independent covariates affect the response variable [52, 54]. A good model is one that results in a high likelihood of the observed results.

3.7.12 Statistical Tests: Confidence Interval

The odds ratio (OR) with a 95% confidence interval (CI) can be used to test for the contribution of individual covariates at 5% level of significance [54, 55]. The 95% CI is used to estimate the precision of the OR. A wide confidence interval indicates a low level of precision, whereas a narrow confidence interval indicates a higher precision [56].

A $100(1 - \alpha)\%$ two-sided CI for Odds Ratio according to Hosmer and Lemeshow [54] and Szumilas [58] is given by:

$$\exp\left[\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \times \widehat{SE}(\hat{\beta}_j)\right] = e^{\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \times \widehat{SE}(\hat{\beta}_j)} \quad (3.35)$$

Where $\widehat{SE}(\hat{\beta}_j) = \left[\widehat{Var}(\hat{\beta}_j)\right]^{1/2}$, the standard error of the estimators, α is the level of significance and Z_r is the percentile value of the standard normal distribution.

3.7.13 Goodness-of-fit statistics: Hosmer-Lemeshow test

The Hosmer-Lemeshow statistics evaluates the goodness-of-fit by comparing the observed proportions of events to the predicted probabilities of occurrence in subgroups of the model population. It is assessed by dividing the predicted probabilities into deciles (10 groups based on percentile ranks) and then computing a Pearson chi-square that compares the predicted to the observed frequencies in a 2-by-10 table. Thus, the test statistic is a chi-square statistic with a desirable outcome of non-significance, indicating that the model prediction does not significantly differ from the observed. Hosmer-Lemeshow test is written as:

$$H = \sum_{k=1}^{10} \frac{(O_k - E_k)^2}{E_k} \qquad (3.36)$$

where O_k and E_k denote the observed and expected events respectively for the k^{th} risk deciles group. The statistics test, H, asymptotically follows a chi-square distribution with 8 (10 – 2, number of groups minus two) degrees of freedom, $H \sim \chi_8^2$.

The constructed multivariable logistic regression model would be considered reasonable if $p > 0.05$. Thus, the larger the p-*value* obtained using the Hosmer-Lemeshow test, the smaller the square of the distance between O_k and E_k, and hence, the better the fit of the model [30, 56].

The key quantities for further logistic diagnostic as in linear regression, are the components of the "residual sums of square". Central to the formation and interpretation of linear regression diagnostic are the "hat" matrix and the leverage values delivered from it. However, in logistic regression, the hat matrix is a little different from the ordinary linear regression hat matrix. In logistic regression, the hat matrix is given by

$$H = V^{1/2}X(X'VX)^{-1}X'V^{\frac{1}{2}} \qquad 3.37$$

Where V is a j x j diagonal matrix with general element $v_j = m_j\hat{\pi}(x_j)[1 - \hat{\pi}(x_j)]$. Due to the complex nature of the hat matrix, further diagnostic based on the hat matrix were beyond the scope of this study.

3.7.14 Mortality and Actuarial Life-Tables

Life tables are used to measure mortality, survival and the life expectancy of a population at varying ages. There are two types of life tables used in actuarial science. The period life table represents mortality rates during a specific time period of a certain population. A cohort life table, often referred to as generation life table, is used to represent the overall mortality rates of a certain population's entire lifetime.

3.7.14.1 Life-Table for Censored Data

Life tables are of particular value in the use of data that may be affected by censoring or loss to follow-up. The life table approach can make efficient use of partial data by including observations up to the time of censoring. Observations are considered to be censored if patients are lost to follow-up. Those who are censored during an interval are assumed to have been followed, on average, for half the interval. The information may be displayed as in Table 3.2.

Table 3.2 Hypothetical Life-Table for Censored data

j	Year after surgery (interval)	n_j	d_j	w_j	$\widehat{CI}_j$	$\widehat{CI}_{0,j}$
1	0-1	n_1	d_1	w_1	$\widehat{CI}_1$	$\widehat{CI}_{0,1}$
2	1-2	n_2	d_2	w_2	$\widehat{CI}_2$	$\widehat{CI}_{0,2}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
k	(k-1)-k	n_k	d_k	w_k	$\widehat{CI}_k$	$\widehat{CI}_{0,k}$

Where:

- j = follow-up time,
- n_j = # of subjects observed in interval j,
- d_j = # of subjects died from breast cancer in interval j,
- w_j = # of subjects withdrawals ("censored") in interval j.
- CI is the cumulative incidence.
- An estimate of the probability of dying due to breast cancer after surgery during the time interval j is given as;

$$\widehat{CI}_j = \frac{d_j}{n_j - w_j/2}$$

An estimate of the probability of dying from breast cancer after surgery for an accumulated period of time, $\widehat{CI}_{0,j}$ is given as;

$$\widehat{CI}_{0,j} = 1 - \prod_{j'=1}^{j}\left(1 - \widehat{CI}_{j'}\right) \qquad (3.38)$$

3.7.15 Statistical Analysis Summary

To investigate the association between death from breast cancer after surgery and each potential risk factor, odds ratios were computed and *p*-values were evaluated by using the multivariable logistic regression model. Odds ratios were used to evaluate the relative odds of death caused by breast cancer between the levels of the factor, and *p*-values were calculated to assess significance of results. A multivariable logistic regression analysis was used to measure the significance of several risk factors simultaneously and to predict the survival probability of breast cancer patients using SPSS version 20. Three models were fitted; first, a model which involved factors found to be significant in literature, this was done for the purpose of validation using our data, second, a model using factors found insignificant in literature including those which were not considered in literature. The second model was fitted to independently assess factors which were not considered in literature and to check the influence of factors in the first model. Finally, a model consisting of all the significant factors from the two models was fitted. Goodness-of-fit for the model was assessed using Hosmer-Lemeshow statistic test. Various tests, using the odds ratio as a test statistic were performed and a prediction of a five-year survival was done for various groups of study subjects using the model. Also, overall observed five-year breast cancer mortality was performed using Actuarial life tables.

3.8 Ethical Considerations

Ethical clearance was sought from the University of Zambia ethics committee and permission to retrieve data for the study was obtained from the Senior Medical Superintendent at CDH. Requirement for patient consent was waived since no direct contact with patients was planned and de-identified data was used for analysis and reporting.

3.9 Chapter Summary

This chapter discussed the methodology employed in the study. Under methodology the items discussed include; research design, study population, sample size, sampling procedures and data collection. Further, description of certain models, statistical analysis and ethical issues were also considered.

CHAPTER FOUR: ANALYSIS AND RESULTS

4.1 Overview

The previous chapter highlighted the methodology employed in the study. The current chapter describes the results of the study in a series of steps. The first step describes characteristics of the study subjects followed by a step-by-step fit of the logistic model and projection of the five-year mortality rate by group variables.

4.2 Characteristics of Patients

The data utilized in this study was obtained from the Cancer Disease Hospital (CDH) in Lusaka, Zambia. From 450 collected patients' files 217 were excluded because either the patients did not undergo any surgery or had data missing in a number of variables or due to lost to follow up. Thus, the total number of breast cancer patients included in the final data set was 233. Socio-demographic characteristics of the study sample (n = 233) are depicted in Table 4.1. The median age at surgery was 47.6 years and the age at surgery range was from 18 to 77. Most of the patients were married (59.2%). The probabilities of death due to breast cancer after surgery, over the study period, was estimated to be 13.8% for the married and 36.8% for the singles.

Obesity was one of the factors considered in the study and using body mass index (BMI) to measure it, 51.9% of the population were not obese while 48.1% were either obese or overweight. Approximately, 15.7% of the patients who were classified as not obese or overweight died within the study period. For those classified as obese or overweight, about 31.3% died as indicated in Table 4.1.

Almost 11.2% of the patients in the study had a history of breast cancer in the family from which 34.6% died within the study period. Among those whose alcohol intake status was determined, about 33.0% used to consume alcohol. The five-year mortality for those who took alcohol was 31.9% while for non-alcohol consumers was 18.6%. In addition, among those who provided information on contraceptive use, about 57.1% reported to have used contraceptives. Regarding the number of births, around 5.8% of the patients who provided information about number of births were nulliparous (never gave birth) and 94.2% gave birth.

The percentage of non-response varied from one factor to another, 42.9% on the use of contraceptive, 10.3% on the intake of alcohol and 48.1% on the number of births. A summary of this information is given in Table 4.1.

Table 4.1 Demographic and Social related factors of patients

Factors	Deceased (%)		Survived (%)		Total (%)	
Marital status						
Married	19	(13.8%)	119	(86.2%)	138	(100%)
Single	35	(36.8%)	60	(63.2%)	95	(100%)
BMI						
Not obese	19	(15.7%)	102	(84.3%)	121	(100%)
Obese/overweight	35	(31.2%)	77	(68.8%)	112	(100%)
Family history						
Yes	09	(34.6%)	17	(65.4%)	26	(100%)
No	45	(21.7%)	162	(78.3%)	207	(100%)
Alcohol consumption						
No	26	(18.6%)	114	(81.4%)	140	(100%)
Yes	22	(31.9%)	47	(68.1%)	69	(100%)
Unknown	06	(15.0%)	18	(75.0%)	24	(100%)
Contraceptive use						
Yes	17	(22.4%)	59	(77.6%)	76	(100%)
No	15	(26.3%)	42	(73.7%)	57	(100%)
Unknown	22	(22.0%)	78	(78.0%)	100	(100%)
Number of birth						
Nulliparous	02	(28.6%)	05	(71.4%)	07	(100%)
Gave birth	19	(16.7%)	95	(83.3%)	114	(100%)
Unknown	33	(29.5%)	79	(70.5%)	112	(100%)

In our study, patients with left breast tumors were more common making up 52.4% of the entire sample population. A five-year mortality stood at 27.9% for patients with right tumors and 18.9% for those with left tumors. Regarding surgery treatment for subjects in the study; 85.0% of patients underwent mastectomy and 15.0% had lumpectomy or breast conservation surgery. This information is summarized in Table 4.2 (a).

Table 4.2 (a) Clinic characteristics of patients

Factors	Deceased (%)		Survived (%)		Total (%)	
Surgery type						
Lumpectomy	06	(17.1%)	29	(82.9%)	35	(100%)
Mastectomy	48	(24.2%)	150	(75.8%)	198	(100%)
Laterality						
Right	31	(27.9%)	80	(72.1%)	111	(100%)
Left	23	(18.9%)	99	(81.1%)	122	(100%)

In Table 4.2 (b), we present the results on pathological characteristics. The distribution of subjects according to histologic grade was 52.8% for those with either grade I or II and 47.2% for those with III. The majority of the women who underwent surgery did so in late stages (III and IV) of breast cancer, accounting for 60.1% of the entire study population. Patients with early breast cancer stages (I and II) made up 39.9% of the study population as presented in Table 4.2 (b).

In general, factors which had two response categories (positive or negative) had higher mortality rates for subjects who were found to be positive on that factor compared to subjects who were negative. These includes; ER, PR, HER2 and HIV as displayed in Table 4.2 (b). For other factors such as tumor size and lymph node status, mortality rate increased with the degree of severity across the levels of the factor.

However, a significant number of patients had missing information with respect to HIV status (22.3%), HER2 (10.7%), PR status (6.4%). Additionally, about 6.4% of patients' hormonal receptors (ER status) was not classified. Also, about 9.9% and 8.6% of the patients had missing information on tumor size and lymph node status, respectively.

Table 4.2 (b) Pathological characteristics of patients

Factors	Deceased (%)		Survived (%)		Total (%)	
Breast cancer stage						
Stage I & II	10	(10.8%)	83	(89.2%)	93	(100%)
Stage III & IV	44	(31.4%)	96	(68.6%)	140	(100%)
ER status						
Negative	16	(17.6%)	75	(82.4%)	91	(100%)
Positive	36	(28.3%)	91	(71.7%)	127	(100%)
Unknown	02	(13.3%)	13	(86.7%)	15	(100%)
PR status						
Negative	15	(14.3%)	90	(85.7%)	105	(100%)
Positive	37	(32.7%)	76	(67.3%)	113	(100%)
Unknown	02	(13.3%)	13	(86.7%)	15	(100%)
HER2						
Negative	27	(20.6%)	104	(79.4%)	131	(100%)
Positive	24	(31.2%)	53	(68.8%)	77	(100%)
Unknown	03	(12.0%)	22	(88.0%)	25	(100%)
HIV status						
Negative	19	(15.4%)	104	(84.6%)	123	(100%)
Positive	25	(43.1%)	33	(56.9%)	58	(100%)
Unknown	10	(19.2%)	42	(80.8%)	52	(100%)
Histologic grade						
Grade I & II	19	(15.4%)	104	(84.6%)	123	(100%)
Grade III	35	(31.8%)	75	(68.2%)	110	(100%)
Tumor size						
Tis, 1 & 2	15	(16.9%)	74	(83.1%)	89	(100%)
T3 & T4	38	(31.4%)	83	(68.6%)	121	(100%)
Unknown	01	(4.3%)	22	(95.7%)	23	(100%)
Lymph node status						
N0 & 1	14	(13.3%)	91	(86.7%)	105	(100%)
N2 & N3	39	(36.1%)	69	(63.9%)	108	(100%)
Unknown	01	(5%)	19	(95%)	20	(100%)

4.3 Comparisons of patients' survival on continuous factors

Age at the time of surgery was recorded on the 233 subjects. Those who survived had an average of 45.98 years with a standard deviation of 11.598 and those who died had an average age of 52.94 with a standard deviation of 13.921 as presented in Table 4.3. On average, patients who died were slightly older than those who survived, for the period under study. In addition, using the date on which surgery was performed and the date the patient was last seen, we created a variable of how long the subject was under observation (time on study). Those who survived were observed for an average of 2.4 years with a standard deviation of 1.46, while those who died for an average of 2.0 years with a standard deviation of 1.36 years.

Table 4.3 Comparison of patients' age at surgery and time on study by survival status

Variables	Survival status	N	Mean	Std. deviation	Std. error mean
Age at surgery	Dead	54	52.94	13.921	1.894
	Alive	179	45.98	11.598	.867
Time on study	Dead	54	2.022	1.3598	.1850
	Alive	179	2.394	1.4575	.1089

The difference in the mean age was statistically significant at 5% level using a T test assuming equal variance, the P-value was less than 0.001 for a T value of 3.687. However, the difference in the means between those who survived and those who died with respect to time on study was not significant yielding a P-value of 0.097.

4.4 Significant characteristics on survival

Chi-square tests of association between individual factors and survival status were performed and the results are showed in Table 4.4. The results showed that marital status and alcohol consumption were significantly associated with mortality of breast cancer patients after surgery. Similarly, the analysis of the clinical data revealed that breast cancer stage, histologic grade, PR status, tumor size, lymph node, HIV status and BMI were also significantly associated with breast cancer mortality after surgery.

Relative comparison using odds ratio showed that the odds of death for patients who were single were about 3.65 times higher than for the married ones. The odds of death for breast cancer patients classified as obese or overweight was 2.44 times higher than those classified as not obese or overweight.

Table 4.4 A summary of tests of association of factors versus survival

FACTORS	LEVELS	OR	STATISTIC (Chi-square)	P-value
Marital status	Married Single	Ref 3.654	16.825	< 0.001*
BC stage	0, I & II III & IV	Ref 3.804	13.417	< 0.001*
BMI	Not obese Obese/overweight	Ref 2.44	7.897	0.005*
Alcohol consumption	No Yes	Ref 2.052	4.630	0.031*
PR status	Negative Positive	Ref 2.921	10.209	0.001*
HIV status	Negative Positive	Ref 4.147	16.384	< 0.001*
Histologic grade	I & II III	Ref 2.554	8.741	0.003*
Tumor size	Tis, 0, 1 & 2 T3 & T4	Ref 2.259	5.754	0.016*
Lymph node	N0 & N1 N2 & N3	Ref 3.674	14.778	< 0.001*
HER2	Negative Positive Unknown	Ref 1.744	2.921	0.087
Family history	No Yes	Ref 1.906	2.151	0.142
Contraceptive use	No Yes Unknown	Ref 0.774	0.393	0.531
Surgery type	Lumpectomy Mastectomy	Ref 1.547	0.842	0.359
Laterality	Right Left	Ref 0.600	2.689	0.101
ER status	Negative Positive Unknown	Ref 1.854	3.382	0.066
Number of birth	Nulliparous Gave birth Unknown	Ref 0.5	0.652	0.420

Note: * indicates significant factors at 5% level of significance and Ref means reference group

In addition, for study subjects whose HIV status was known, the odds of death given one was HIV positive were 4.15 times higher than the odds of dying given one was HIV negative. This means being HIV positive increases the likelihood of dying from breast cancer after surgery. The odds ratio of patients with positive PR (OR = 2.921) dying from breast cancer after surgery were about 3 times higher than those with negative PR. The study also established that consuming alcohol increases the likelihood of dying from breast cancer after surgery. It was found that the odds ratio of death due to breast cancer after surgery for patients who consumes alcohol were about 2 times higher than those who did not consume alcohol (OR = 2.052).

Also, patients who underwent surgery with breast cancer stages III or IV were at risk of death due to breast cancer than those in stages 0, I and II. It was established that the odds ratio of death among patients who underwent surgery in stages III or IV were about 4 times higher than those who were in stages 0, I or II. Further, the odds of death increased about 3 times when the histologic tumor grade was III versus grades I or II as indicated in Table 4.4. The increase in the number of involved lymph nodes increased the odds of death. The odds of death in the patients with 4-9 or above 9 involved nodes were about 3.7 times higher than for those with either 1-3 or no involved nodes. Regarding tumor size, patients with larger tumor size were highly susceptible to die from breast cancer after surgery. According to the results in Table 4.4, the odds of deaths for patients with T3 or T4 were about 2 times higher than those with tumor size less than 5cm (OR = 2.259).

On the other hand, the type of surgery had no significant association with mortality of breast cancer even though mortality inclined more to patients with mastectomy than lumpectomy (OR = 1.547). The major reason for this result could be due to a higher percentage of patients with advanced disease who underwent mastectomy. Also, family history showed no association with mortality although the odds of death for those with family history were about 2 times higher than those with no family history of breast cancer (OR = 1.906).

In addition, tumor laterality showed an insignificant association with survival status of breast cancer patients after surgery. This implies that the side of the tumor in terms of the breast in this study has no impact with respect to mortality of breast cancer patients after surgery. However, the odds of death due to breast cancer after surgery for patients with left tumor were 0.600 less than those who presented with right tumor. Even though the variables such as HER2, contraceptive use and number of births were found to be insignificant, the results may be due to a significant number of subjects whose status for the said factors were not recorded as indicated in Table 4.2. With respect to odds ratio, HER2 positive patients (OR = 1.744) were highly susceptible to die from breast cancer after surgery than those who were HER2 negative. Thus, mortality was slightly more with HER2 positive patients than with HER2 negative subjects after surgery. The odds of death due to breast cancer after surgery for contraceptive users were about 0.774 times less than for non-contraceptive user. Furthermore, giving birth (OR = 0.5) decreases the probability of dying from breast cancer after surgery. Also, regarding ER status, the results showed that patients who were positive on the factor were more likely to die from breast cancer than those who were negative (OR = 1.854).

4.5 Multiple Logistic regression model of breast cancer after surgery

4.5.1 The fitted model

Variables that were significant in univariate analysis were entered into a multiple logistic regression model. The model containing literature proven factors was fitted first (model 1). The model contained the following factors; age at surgery, breast cancer stage, histologic grade, BMI, PR status, tumor size and lymph node status. The results of the model showed that age at surgery, BMI, PR status, breast cancer stage and histologic grade were significantly associated with mortality of breast cancer patients after surgery.

Another model (model 2) containing other additional factors including those which were insignificant in literature was also fitted. The second model (model 2) included the following variables; marital status, HIV status and alcohol consumption. The results revealed that marital status and HIV status were significantly associated with mortality. The dependent variable in both models was survival status (dead or alive).

Further, the model containing all the independent variables which were significantly associated with death due to breast cancer in model 1 and 2 above was fitted. The variables were; Age at surgery (x_1), breast cancer stage (III or IV = x_2), BMI (obese/overweight = x_3), PR status (positive PR = x_4), histologic grade (grade III = x_5), HIV status (positive = x_6) and marital status (single = x_7) with the dependent variable being survival status (dead or alive).

In our study, stages I and II were combined into one group which was used as the reference group for the variable breast cancer stage. Other reference groups were as follows: not obese for variable BMI, HIV negative for HIV status, grade I and II were combined into one group for histologic grade, PR negative for PR status and married for marital status. For variables having two categorical levels, the level assumed disadvantageous to the subject assumed the value of 1 and the other the value of 0. For instance, HIV status, x_6, takes the value 1 if the status of HIV is positive and 0 if the status is negative. This also applied to the other dummy variables. The model fitted was as follows:

$$ln\left(\frac{\pi(\underline{x})}{1-\pi(\underline{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \quad (4.1)$$

where $\underline{x} = (x_1, x_2, \ldots, x_7)'$, is a vector of factors, $\pi(\underline{x})$ is the probability of a subject dying after surgery given the presence of the factors in vector $\underline{x}$ and $\beta_1, \beta_2, \ldots, \beta_7$ are coefficients associated with the factors and β_0 is a constant. The results of the fitted model are tabulated in Table 4.5.

Table 4.5 Estimated coefficients, standard errors (SE) and Wald statistics, P-values, adjusted odds ratio (AOR) and 95% CI for the preliminary model

Factors			$\hat{\beta}$	**S. E**	**Wald**	**P-value (p)**	**AOR**	**95% CI**
Age at surgery		x_1	0.068	0.021	10.916	0.001	1.070	(1.028, 1.114)
BC stage	III or IV	x_2	2.361	0.616	14.714	< 0.001	10.603	(3.173, 35.429)
BMI	Obese/overweight	x_3	1.090	0.497	4.814	0.028	2.973	(1.123, 7.869)
PR status	Positive	x_4	1.965	0.552	12.674	< 0.001	7.133	(2.418, 21.040)
Histologic grade	III	x_5	1.521	0.505	9.085	0.003	4.577	(1.702, 12.305)
HIV status	Positive	x_6	1.171	0.479	5.985	0.014	3.226	(1.262, 8.243)
Marital status	Single	x_7	1.410	0.496	8.066	0.005	4.096	(1.548, 10.838)
Constant			-9.803	1.682	39.980	< 0.001		

Using the estimated coefficients from Table 4.5 the fitted model was:

$$ln\left(\frac{\pi(\underline{x})}{1-\pi(\underline{x})}\right) = -9.803 + 0.068x_1 + 2.361x_2 + 1.090x_3 + 1.965x_4 + 1.521x_5 + 1.171x_6 + 1.410x_7 \quad (4.2)$$

4.5.2 Interpretation of the regression coefficients

As earlier alluded to, each regression coefficient describes the size of the relative contribution to the log odds of a subject dying when the corresponding predictor variable takes a particular level compared to the reference group. In our fitted model, the coefficient of breast cancer stage III or IV was 2.361. This means that the log odds of mortality from breast cancer after surgery for a woman with a BC stage of either III or IV increased by 2.361 when compared to a woman with a BC stage of either I or II while controlling for other variables (i.e., when other variables remained unchanged). Regarding BMI, the log odds of mortality after surgery of a patient who was obese or overweight increased by 1.090 compared to the one who was not obese while controlling for other variables. With respect to PR status, the log odds of breast cancer mortality after surgery for a woman who was PR positive increased by 1.965 as compared to a woman who was PR negative.

Additionally, the increase in log odds of mortality from breast cancer after surgery compared to the reference group were: 1.521 for histologic grade III, 1.171 for HIV positive and 1.410 for marital status. With respect to age, the log odds of mortality increased by 0.068 for 1 year increase in age of patients while holding the other predictor variables constant.

4.5.3 Influential Factors of breast cancer mortality after surgery

To assess the significance of each categorical variable in the model, we used the Wald test. The results, as presented in Table 4.5, showed that all the factors were significant at 5% level of significance. Age at surgery (P-value (p) = 0.001), breast cancer stage ($p < 0.001$), obese/overweight ($p = 0.028$), PR status ($p < 0.001$) and histologic grade ($p = 0.003$) were significant predictors of breast cancer mortality after surgery. In addition, HIV status ($p = 0.014$) and marital status ($p = 0.005$) were also identified as risk factors of breast cancer mortality after surgery.

The significance of the parameters indicated that the difference in log odds between the indicated level and the reference level of a factor was significant, this in turn, pointed out that the odds ratio was also significant. For instance, marital status was coded as married versus single, where single includes those never married, widowed, as well as divorced (x_7), in terms of marital status, we could express the model as:

$$log\left(\frac{\pi(\underline{x}_{-7},x_7)}{1-\pi(\underline{x}_{-7},x_7)}\right) = \underline{x}_{-7}\underline{\hat{\beta}}_{-7} + 1.410x_7, \tag{4.3}$$

where $\underline{x}_{-7}$ represents all other factors except x_7 and $\underline{\hat{\beta}}_{-7}$ stand for a vector of estimates of their parameters. The odds ratio between the married and the singles was derived as follows:

$$log\left(\frac{\pi(\underline{x}_{-7},x_7=1)}{1-\pi(\underline{x}_{-7},x_7=1)}\right) = \underline{x}_{-7}\underline{\hat{\beta}}_{-7} + 1.410, \tag{4.4}$$

as the log odds for those single, and

$$log\left(\frac{\pi(\underline{x}_{-7},x_7=0)}{1-\pi(\underline{x}_{-7},x_7=0)}\right) = \underline{x}_{-7}\underline{\hat{\beta}}_{-7}, \tag{4.5}$$

as the log odds for the married. The odds ratio was then expressed as;

$$\frac{\frac{\pi(\underline{x}_{-7}, x_7=1)}{1-\pi(\underline{x}_{-7}, x_7=1)}}{\frac{\pi(\underline{x}_{-7}, x_7=0)}{1-\pi(\underline{x}_{-7}, x_7=0)}} = \frac{e^{\underline{x}_{-7}\underline{\hat{\beta}}_{-7}+1.410}}{e^{\underline{x}_{-7}\underline{\hat{\beta}}_{-7}}} = e^{1.410} = 4.096. \qquad (4.6)$$

The odds that one dies of breast cancer after surgery, given they are single was 4.096 times higher than odds of dying given one was married. This simply says being married improves one's odds of survival and reasons for this may be many, care by spouse may be one of the possibilities. The 95% confidence interval for the true odds ratio was:

$$(e^{1.410-1.96\times 0.496}, e^{1.410+1.96\times 0.496}) = (1.549, 10.828), \qquad (4.7)$$

which does not include the null value of 1.

In addition, the adjusted odds ratio (AOR) of death for patients who underwent surgery in stages III or IV was 10.603 times higher than those who underwent surgery in stages 0, I or II. It was also noted that the obese or overweight patients were highly likely to die from breast cancer compared to non-obese or overweight patients (AOR = 2.973). Also, being PR positive increased the likelihood of dying from breast cancer after surgery. The adjusted odds ratio of death for PR positive patients was 7.133 times higher than those who were PR negative. Regarding histologic grade, patients who were in grade III were more likely to die from breast cancer after surgery than those who were in grades I or II (AOR = 4.577). additionally, the adjusted odds of death were more for HIV positive patients than for HIV negative (AOR = 3.226). Concerning marital status, the results showed that single patients were about 4 times more likely to die from breast cancer than married patients (AOR = 4.096).

Further, we categorized age in years into various age groups as showed in Table 4.6.

Table 4.6 Unadjusted odds ratios for mortality and age groups compared to those under 31 years

Factor	**Levels or categories of age**	**Odds Ratio (OR)**
Age at surgery	< 31	Reference age
	31 – 40	5.442
	41 – 50	1.385
	51 – 60	10.687
	≥ 61	13.714

According to Table 4.6, the odds of death for subjects aged between 31 and 40 years were about 5.4 times higher than the odds for those aged less than 31 years. The odds ratio for mortality increased with an increase in age of patients as indicated in Table 4.6 except for age group 41 to 50.

Further, in the literature, factors such as breast cancer stage, BMI, PR status, age and histologic grade were found to be significantly associated with mortality of breast cancer. Although marital status was an insignificant factor of mortality in literature, it was established as a significant factor of breast cancer mortality after surgery in this study. In addition, the present study has established that HIV status is a significant predictor of mortality of breast cancer patients after surgery.

4.5.4 The full model assessment

The significance of the model was tested via the likelihood ratio test of the full model versus a constant model. The -2 log likelihood obtained by fitting the constant model only was 192.866, and the -2 log likelihood for the overall model was found to be 112.770 as showed in Table 4.7.

Table 4.7: The likelihood ratio test

Model	**Model fitting Criteria**	**Likelihood Ratio Tests**		
	-2Log Likelihood	Chi-square	df	Sig.
Intercept only	192.866			
Final	112.770	80.096	7	0.000

To test the hypotheses above:

$$H_0: \underline{\beta} = \underline{0} \text{ versus } H_1: \underline{\beta} \neq \underline{0}$$

A Chi-square test was carried out using the G statistic given by:

$$G = 192.866 - 112.770 = 80.096$$

The -2 log likelihood statistics yielded a chi-square value of 80.096 with a p-value less than 0.001. The results showed that at least one of the covariates' regression coefficients was not equal to zero because of the small p-value ($p < 0.001$) which is less than 0.05. This led us to reject the null hypothesis (H_0) in favor of the alternative hypothesis (H_1). Therefore, we conclude that at least one and perhaps all beta's coefficients are different from zero. This result showed that our model was significant over the intercept (constant) model.

4.5.5 Goodness of fit assessment

The goodness of fit for the model was carried out using Hosmer-Lemeshow test and the results are presented in Table 4.8.

Table 4.8: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4.490	8	.810

Our model showed a good fit based on Hosmer-Lemeshow test. The Hosmer-Lemeshow test exhibited an excellent performance of the model (P = 0.810). This showed that our model's estimates fitted the data at an acceptable level. Further, the overall prediction accuracy increased from 74.9% to 84.2%.

4.5.6 Extended Logistic Regression model

Several interaction terms were examined but only the interaction between BMI and HIV remained significant in the model. The introduction of the interaction term rendered the individual factors, BMI and HIV insignificant in the model. The extended model includes these individual terms together with the interaction in keeping with the convention of the model building. The extended model contained the variables such as age in years (x_1), breast cancer stage (x_2), BMI (x_3), PR status (x_4), histologic grade (x_5), HIV (x_6), marital status (x_7) and BMIxHIV (x_8). The results of the extended model are tabulated in Table 4.9.

Table 4.9: The estimated coefficients, standard errors (SE) and Wald statistics, P-values, and adjusted odds ratios (AOR) for the final model

Variables in the model		$\hat{\beta}$	S.E.	Wald	P-value (p)	AOR
Age in years	x_1	.075	.022	11.759	0.001*	1.078
Breast cancer stage	x_2	2.360	.640	13.596	< 0.001*	10.589
BMI	x_3	-.048	.690	.005	0.945	.953
PR status	x_4	2.031	.569	12.746	< 0.001*	7.623
Histologic grade	x_5	1.573	.520	9.155	0.002*	4.821
HIV	x_6	-.313	.807	.151	0.698	.731
Marital status	x_7	1.839	.565	10.606	0.001*	6.293
BMIxHIV	x_8	2.535	1.115	5.174	0.023*	12.621
Constant		-9.856	1.756	31.502	0.000	.000

Note: * indicates significant factors at 5% level of significance.

Using results from Table 4.7, the following model was fitted:

$$ln\left(\frac{\pi(\underline{x})}{1-\pi(\underline{x})}\right) = -9.856 + 0.075x_1 + 2.360x_2 - 0.048x_3 + 2.031x_4 + 1.573x_5 - 0.313x_6 + 1.839x_7 + 2.535x_8 \quad (4.8)$$

where x_8 is the interaction term between BMI and HIV.

After including the interaction term, age in years, breast cancer stage, PR status, histologic grade and marital status remained significant factors of breast cancer patients after surgery. Even though BMI and HIV are insignificant, their joint significance is captured in the interaction. In general, BMI has four categories; Underweight = <18.5, Normal weight = 18.5–24.9, Overweight = 25–29.9, Obesity = 30 or greater. However, the data obtained for the study had two categories, obese/overweight and not obese, this could have led to some loss of information in the interaction of HIV and BMI.

4.6 Breast cancer mortality rates after surgery

In this section, we predicted a 5-year mortality rate for various age groups using the final model. Further, we determined an overall 5-year mortality for breast cancer patients after surgery using Actuarial life-tables as well.

4.6.1 Prediction of a 5-year mortality using logistic regression model

In order to estimate a five-year mortality, the time on study (in years) as a variable was also included in the model given in equation (4.8) to get the following model.

$$ln\left(\frac{\pi(\underline{x})}{1-\pi(\underline{x})}\right) = -9.259 + 0.079x_1 + 2.381x_2 - 0.027x_3 + 2.091x_4 + 1.585x_5 - 0.172x_6 + 1.665x_7 + 2.600x_8 - 0.373x_9 \quad (4.9).$$

where x_8 is the interaction term between BMI and HIV, and $x_9 =$ time on study.

When the variables were ordered by the level of significance we obtained; Stage, PR, Age, Grade, Marriage, interaction between BMI and HIV, Time, HIV and BMI. The first two and the fourth are cancer related factors and we grouped them together as level 1 variables. Marriage, interaction between BMI and HIV, HIV and BMI were social and lifestyle related factors which were classified as level II variables while Age and Time were used for stratification and time-dependent variables, respectively. We grouped the subjects by considering the combinations of the presence and absence of the two-level classified variables as shown in Table 4.10.

Table 4.10 Five-year mortality by age group using logistic regression model

	Pattern of variables									
	Age	26	40	70	26	40	70	26	40	70
Variables	Parameter estimates	Cancer + social and life style variables			Cancer related variables			Social and life style variables		
Stage	2.381	1	1	1	1	1	1	0	0	0
PR	2.091	1	1	1	1	1	1	0	0	0
Grade	1.585	1	1	1	1	1	1	0	0	0
Marriage	1.665	1	1	1	0	0	0	1	1	1
BMIxHIV	2.600	1	1	1	0	0	0	1	1	1
HIV	-0.172	1	1	1	0	0	0	1	1	1
BMI	-0.027	1	1	1	0	0	0	1	1	1
Age	0.079									
Time	-0.373	5	5	5	5	5	5	5	5	5
Constant	-9.259	1	1	1	1	1	1	1	1	1
Mortality probability		0.741	0.897	0.989	0.047	0.130	0.614	0.0067	0.0199	0.178

Note: 1 means the variables is included in the model and 0 means it is not included.

We categorized age into three groups, namely; younger group (35 years or less), middle group (35 < age <= 49) and older group (50 <= age). For younger age group, we took the average of 18 years and 35 years and fixed age at 26 years, while for middle and older groups, age was fixed at 40 and 70 years, respectively. Cancer related factors were classified as level I while social and life style related factors as level II. In total, we grouped subjects into 9 patterns, three for each age group where both classes were present; then where only one of them was present as shown in Table 4.10. In addition, age was used for stratification and "time on study" (Time) as a time-dependent variable. Further, the combinations of the presence and absence of the factors in the two levels were used to divide subjects into nine groups as displayed in Table 4.10.

The results of the study revealed that the five-year mortality rates of patients aged 26, 40 and 70 years, having both classes of variables were 74.1%, 89.7% and 98.9%, respectively. Mortality was rather high for patients with the two classes of variables present, pattern 1. On other hand, breast cancer mortality was low for patients with either one of the two classes present, pattern 2 and 3. Table 4.8 also shows that five-year mortality rate of breast cancer patients after surgery increased with age as may be observed within each pattern of the age groups. Further, the results also shows that cancer-related variables have far much more impact on mortality.

4.6.2 Prediction of 5-year overall mortality using the Actuarial Life-Tables

An actuarial life table is a table that provides a way of calculating the probability of a person at a certain age dying before their next birthday. Insurance companies often use such tables to calculate the remaining life expectancy for people at different ages and stages, without considering other factors as it is done in logistic regression. Logistic regression provides a superior method for estimating mortality because it takes into account other factors.

An Actuarial life table was used on our data to determine the overall mortality rate for patients with breast cancer after surgery. Patients were followed from the date of surgery until death or loss to follow-up. Time on study variable was used to create one-year intervals where the outcome of interest was death due to breast cancer. We estimated a 1, 3 and 5-year mortality of breast cancer patients after surgery using the spreadsheet-like table used in actuarial calculations. The results are shown in Table 4.11.

Table 4.11 Life-table probabilities of dying from breast cancer

Year	Interval (j-1, j)	$\mathbf{n_j}$	$\mathbf{d_j}$	$\mathbf{w_j}$	$\widehat{\mathbf{CI}}_{\mathbf{j}}$	$\widehat{\mathbf{CI}}_{\mathbf{0,j}}$
1	0, 1	233	18	36	0.0837	0.0837
2	1, 2	179	10	45	0.0639	0.1423
3	2, 3	124	11	51	0.1117	0.2381
4	3, 4	62	11	17	0.2056	0.3947
5	4, 5	34	4	22	0.1739	0.499998

Note:

- j = 1, 2, . . ., 5; follow-up time,
- n_j is the number of subjects at the start of interval j,
- d_j is the number of subjects died from breast cancer in interval j,
- w_j is the number of subjects withdrawals ("censored") in interval j.
- CI is the cumulative incidence.
- $\widehat{CI}_{0,j}$ is an estimate of the probability of dying from breast cancer after surgery for an accumulated period of time.

In Table 4.9, the probability of death due to breast cancer after surgery during the first year was estimated using equation (3.21) from section 3.7.13 as follows:

$$\widehat{CI}_1 = \frac{d_{(0,1)}}{n_{(0,1)} - \frac{w_{(0,1)}}{2}} = \frac{18}{233 - \frac{36}{2}} = \frac{18}{215} = 0.0837$$

Additionally, using equation (3.22) from section 3.7.13, we estimated a 5-year mortality rate of breast cancer patients after surgery as follows:

$$\widehat{CI}_{0,j} = 1 - \prod_{j'=1}^{j}\left(1 - \widehat{CI}_{j'}\right)$$

$$\widehat{CI}_{0,5} = 1 - [\left(1 - \widehat{CI}_1\right)\left(1 - \widehat{CI}_2\right)\left(1 - \widehat{CI}_3\right)\left(1 - \widehat{CI}_4\right)\left(1 - \widehat{CI}_5\right)]$$

$$= 1 - [(1 - 0.0837)(1 - 0.0639)(1 - 0.1117)(1 - 0.2056)(1 - 0.1739)]$$

$$= 1 - 0.5000247$$

$$= 0.49998$$

The 5-year mortality rate of breast cancer patients after surgery was 50% as indicated in Table 4.11. Furthermore, a 1 and 3-year mortality rate were estimated to be 8.4% and 23.8%, respectively.

4.7 Chapter summary

This chapter presented the results of the study. The results were presented in a series of steps starting with the description of the characteristics of the patients in the study. The results from the univariate analysis performed followed by model fitted using logistic regression were then presented. The chapter ended with a five-year mortality projection using both the model and the Actuarial Life-Tables.

CHAPTER FIVE: DISCUSSION

5.1 Overview

The previous chapter presented the results obtained in this study. This chapter presents the discussions and limitations of the study.

5.2 Discussion

Breast cancer is one of the most common causes of cancer deaths among women world-wide [1- 4, 6, 18, 45]. Analysis of mortality rate data and related outcomes is essential to evaluate cancer treatment programmes and to monitor the progress of regional and national cancer control programmes. The present study was based on the data obtained from the cancer registry at the Comprehensive Cancer Center at CDH in Lusaka. The hospital collects information about cancer incidence, type, location, stage, and the kinds of treatment received by patients from across the country. In a logistic model, demographic as well as clinical variables such as breast cancer stage, PR status, marital status, HIV status and age among others were employed.

Clinical stage is an important factor for prognosis and determining the appropriate cancer treatment [17, 18, 43]. In a multiple logistic regression model fitted, patients who underwent surgery at an advanced stages (stage III and IV) demonstrated a poor prognosis for survival time. It was found that patients who underwent surgery at advanced stages (stages III and IV) had a worse prognosis after surgery (OR = 3.804; $p < 0.001$) than those in early stages (stages 0, I and II). The reason for higher mortality rate in this study could be due to a huge number of patients who underwent surgery at advanced stage (60.1%). According to Traore et al. [20], patients who were diagnosed with advanced stage had shorter survival after surgery when compared with those with early stages of the disease ($p = 0.002$). Alvarez et al. [17] reported a decrease in survival rate for patients with advanced clinical stages III and IV after diagnosis ($p = 0.027$). However, the level of significance of the factor in our study ($p < 0.001$) differed from the studies conducted by Traore et al. [20] ($p = 0.002$) and Alvarez et al. [17] ($p = 0.027$). This shows that Zambia having different environment and cultural settings, the factor influences breast cancer mortality after surgery differently from other countries. The high mortality rate of breast cancer patients after surgery in the developing countries when compared to some developed countries has been suggested to be related to the advanced stage at diagnosis and limited availability of adequate staging and therapy [43, 45]. The late stage of the disease at diagnosis in developing countries like Zambia could be due to numerous factors such as poverty, lack of screening programmes, low availability of diagnostic facilities, cultural beliefs creating a barrier to early presentation of the disease [20, 37, 50]. On the contrary, Momenyan et al. [29] reported that breast cancer stage was not a prognostic factor of mortality after surgery in a multiple logistic regression.

Studying breast cancer mortality and its prognostic factors gives an insight into the natural history of the disease. Several predictive factors of breast cancer mortality have been identified in many studies over the years. The factor, tumor grade, has been established to have an impact on the mortality of breast cancer patients especially after surgery. Higher histological grade was highly correlated with mortality for patients of breast cancer after surgery. In our study, histologic grade III increased the odds of death from breast cancer after surgery compared to histologic grade I and II (OR = 4.577; $p = 0.003$). A number of studies have shown the importance of histologic grade in predicting the mortality of breast cancer patients especially after surgery [5, 29, 30]. This factor is one of the most important prognostic factor of breast cancer mortality [14, 33]. Balabram, Turra and Gobbi [14] found that histologic grade was statistically associated with three-year survivor of breast cancer patients by a univariate analysis using a log rank test ($p < 0.001$). In a retrospective study by Momenyan et al. [29], histologic grade was also established as a prognostic factor of breast cancer mortality after surgery ($p = 0.03$). These results shows that histologic grade as a factor influences mortality of breast cancer patients differently. We have found that the level of significance of histologic grade with respect to mortality differed from other countries. On the contrary, Alverez et al. [17] found no association between histologic grade and breast cancer mortality.

Body mass index (BMI) also play an important role with respect to mortality of patients with breast cancer [20, 29, 30, 33]. Our study showed that an increase in BMI was associated with mortality of breast cancer patients after surgery. This study established that the risk of death due to breast cancer was more in obese or overweight patients than in those who were not (OR = 2.973; $p = 0.028$). On the contrary, Traore et al. [20] noted insignificant association between BMI and breast cancer mortality after surgery ($p = 0.37$). The findings of the present study show that obesity is an important condition to take into account in the fight to reduce breast cancer mortality after surgery in developing countries like Zambia. This could, possibly be achieved by promotion of regular physical exercise and eating healthy, which have been shown to have positive effects in the prevention of obesity among women [20].

The HIV pandemic remains a public health crisis worldwide [27, 46]. Although the number and lifespan of individuals living with HIV have increased significantly with the scale-up of antiretroviral therapy (ART), in our study, HIV status was identified as a prognostic factor of mortality for breast cancer patients after surgery. During the course of the study, among those whose HIV status was determined, about 43.1% died among HIV positive patients compared to 15.4% among HIV negative patients. This study established that HIV positive patients were at higher risk of death due to breast cancer after surgery compared to patients who were HIV negative. The odds of death given HIV positive were 4.15 times higher than the odds of dying given HIV negative (OR = 4.147, $p < 0.001$). One possible reason for this significant result could be due to a substantial number of HIV positive patients (69%) who were diagnosed with late stages of cancer compared to about 54.5% for those with HIV negative. The difference in percentages was not significant ($p = 0.069$). However, data regarding the effect of HIV status on breast cancer mortality is scarce. In our study, the interaction between BMI and HIV status was significant. However, the introduction of the interaction term reduced the individual effect of BMI (OR = 0.953; $p = 0.945$) and HIV (OR = 0.731; $p = 0.698$) on the risk of death from breast cancer after surgery.

Marital status was another factor which we found to be significantly associated with breast cancer mortality after surgery and the cancer was more prevalent in married women (59.2%). However, single women had an increased risk of mortality from breast cancer after surgery. Being married appeared as a favorable prognostic factor for mortality for women with breast cancer post-surgery. The single women were found to be about 4.1 times more likely to die from breast cancer than the married ones (OR = 4.096; $p = 0.005$). Aizer et al [24] in United States observed that single patients were significantly at higher risk of presentation with metastatic cancer, under treatment, and death resulting from breast cancer ($p < 0.001$). although marital status was significant in Aizer’s study but the level of significance was different from our study. This could be due to different environmental and cultural settings between the two countries. In some studies, they observed that benefits of emotional support, good lifestyle, and stable economics were likely to be protective factors on survival for married women with cancer [24]. In addition, Lan [18] observed that social support and social network had a more important role in reduced mortality for breast cancer in single women.

In Zambia, single women, especially those having many children, generally encounter harsh economic conditions. The financial drain of the treatment course of breast cancer could be an obstacle for seeking care and appropriate treatment compliance especially when patients are required to pay for health care services. This could have contributed to the higher mortality of breast cancer patients after surgery among single patients in this study. Although the government supports vulnerable groups in health care, expenses should also include breast cancer patients to whom the cost exceeds their ability to pay. In so doing, patients with breast cancer would be encouraged to adhere to their long-term treatment and hence reducing mortality rate from this disease. However, further research is required to establish the reasons for higher mortality among single women compared to the married ones.

Although family history is a well-established risk factor for breast cancer, its association with mortality remains vague. In the present study 34.6% of patients with at least one family member who had a history of breast cancer died from the disease. Some studies observed a reduction in survival for patients with a positive family history of breast cancer [17]. However, there are still, studies that have found no difference in mortality rates between patients with or without a family history of breast cancer [22, 29, 30]. In our study, family history was not significantly associated with mortality of breast cancer after surgery.

In our study, age was found to be significantly associated with mortality of breast cancer patients after surgery. The study showed that the odds of death (OR = 1.070; p = 0.001) from breast cancer after surgery increased with age. Previous studies have reported different outcomes about the significance of age on breast cancer [2, 14, 17, 18]. In a retrospective study by Momenyan et al. [29], age between 40 and 70 years was significantly associated with mortality of breast cancer patients after surgery (p = 0.04). In addition, Bouzguenda [48], identified age of more than 70 years as a depreciatory factor influencing significantly the survival of breast cancer patients using both univariate (p = 0.006) and multivariate (p = 0.001) analyses. Further, Chang and Kuo [30] revealed that mortality of breast cancer patients after surgery was more in patients aged less than 35 years. Alvarez et al. [17] revealed that patients younger than 40 years of age had a lower survival rate when compared with patients older than 40 years of age. On the contrary, some studies found no significant difference in survival between the different age groups [17-19, 43].

The effect of the surgical method used on the survival of breast cancer patients is one of the most addressed issues in recent studies. In this study, univariate analysis showed that some variables such as: ER status, HER-2, type of surgery, tumor laterality, oral contraceptives and number of births were not prognostic factors of mortality for breast cancer patients after surgery. This is consistent with some previous studies [17, 30]. Our study showed that the odds of death in women who underwent lumpectomy surgery for breast cancer was less than for patients who underwent mastectomy (OR = 2.053), though not significant, the outcome is consistent with the results of similar studies [14, 29, 30]. In the present study, most of patients who underwent either lumpectomy or mastectomy type of surgery did so with higher tumor grades. This could explain the reason why the difference in mortality with respect to surgery type was not significant. However, other studies have identified these factors as predictors of death in breast cancer [5, 17, 19, 20, 30, 48]. According to Ngowa et al. [43], survival among patients who had breast-conserving surgery was significantly higher than those who had radical mastectomy. Furthermore, some studies suggest that reproductive factors may contribute unfavorably to the development of breast cancer [6, 42].

In our study, some of the variables like lymph nodes, tumor size and alcohol consumption, which were statistically significant in univariate analysis lost their significance in the multiple logistic regression analysis. One of the possible reasons for this could be that the presence of the more significant factors overshadowed their significance. In contrast, a study by Momenyan et al. [29] and Chang and Kuo [30] revealed that lymph node status was statistically associated with mortality due to breast cancer after surgery. Seedhom and Kamal [45], also showed that tumor size and lymph node were important predictors of death for breast cancer patients after diagnosis. However, in the present study, these factors were not significant in a multiple logistic regression analysis despite being significant in the univariate analysis. Similar results were also noted by Momenyan et al. [29] in which tumor size was not significantly associated with mortality in multiple logistic. On the other hand, Chang and Kuo [30] showed that tumor size was an important predictor of death after surgery.

Alcohol consumption is often considered to be an important factor in a number of epidemiological studies. Although several studies have shown that breast cancer survival is associated with alcohol consumption [6, 19, 31, 42], in this study, no significant relationship was observed in breast cancer mortality after surgery among alcohol consumers and non-consumers in a multiple logistic regression. Conflicting results may be due to number of things including; age range of the study groups used in those studies, stage of disease could also have varied among others. Misclassification leads to patients being placed in wrong categories leading to differences in parameter estimates in the attempted models. In the present study, possible reasons as to why these variables lost their significance in the multivariate analysis could be that their significance was surpassed by the more significant factors. In addition, misclassification or measurement error could have led to their effect not be captured accurately especially that our data were secondary and no improvement could be made to it.

The final prognostic factor of mortality after surgery found to be significant was the presence of PR. This factor increased the risk of death among breast cancer patients in the study population (OR = 7.133, $p < 0.001$). Kawaguchi et al. [38], in a multivariate analysis, showed that PR negativity was significantly correlated with prolonged overall survival. On the other hand, other studies found that patients with negative PR were at higher risk of death than positive ones [29, 30]. Baozguenda et al. [48] revealed that PR positive breast cancer patients presented better survival rate as compared to PR negative patients. A study by Alvarez et al. [17] reported no significant relationship between PR status and improved survival.

In the present study, we also compared the 5-year mortality among different age groups using a fitted multiple logistic regression model. Our study showed that the 5-year mortality rates for patients aged 26, 40 and 70 years having both classes of variables were 74.1%, 89.7% and 98.9%, respectively. Our results were similar to those reported in numerous previous studies in other African countries in which the 5-year overall mortality varied from about 40% to 88% except for older groups [20, 43, 44, 48]. Since the cancer hospital is only in Lusaka, patients aged above 50 years tend to live in the rural and far from the Cancer Hospital and may find it difficult when it comes to after surgery reviews due to distance. This could have contributed to higher mortality rate for the older patients in our study.

In addition, the 5-year overall observed mortality of the study subjects after surgery using actuarial life tables was 50% and this was higher compared to other studies that reported varied proportions from 26% to 44% [12, 14, 17, 18, 31, 32]. On the other hand, the overall 5-year mortality observed in our study was lower compared to 88.1% in Gambia, 75.9% in Nigeria and 88% in Tunisia [32, 44, 48]. Ngowa et al. [43] showed a 5-year overall breast cancer mortality of 70% in Cameroon. However, the preferred method was the logistic because life tables are ideal for follow-up studies with specific time period and require large samples to obtain good estimates of mortality or survival. Further, actuarial estimates in our case did not make a distinction in terms of age, as does the logistic model.

5.3 Limitations

Despite the importance of this study in Zambia and its prospective effect on the quality of care for patients with breast cancer, especially after surgery, it may have some limitations and these may include;

- In this study, misclassification may be possible particularly with categorical factors having several levels like I, II, III and IV, patients could have been placed into wrong categories, affecting the sample size in those categories. This has a bearing in parameter estimation. In addition, ascertainment of age may not be possible for some patients, particularly the older ones. Further, the effect of variables such as HIV may be influenced by the presence of ART which was not measured in the present study.
- The main facility with capacity to treat cancer is located in Lusaka; hence the poor in far-flung areas may have no resources to travel to Lusaka. Therefore, this segment of the population may not be represented in the data obtained and there was no credible measure of distance to desegregate the population by distance. Place of residence, as a variable, did not work out well as a measure of distance in this study.
- Another possible source of bias in this study was due to loss to follow-up such as patients relocating to other parts of the country there by making it difficult for them to continue with their routine treatment or may have died but such information was not available. Thus, the sample size was reduced considerably due to exclusion of patients whose final status could not be ascertained.
- Information regarding factors such as lymph node, alcohol consumption, tumor size and number of births was missing for a number of patients. This affected the analysis of those variables both in the univariate as well as in multiple regression.

Nevertheless, the results of the current study provide some knowledge on the risk factors of breast cancer mortality in Zambia. Further, since the study population was reasonably representative of the whole country, the results of the study present a credible outlook on breast cancer mortality for the country.

5.4 Chapter summary

This chapter presented the discussion and limitations of the study. The discussion was done in accordance with studies reviewed in literature review. The next Chapter provides the conclusion and recommendations of the study.

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

6.1 Overview

The previous chapter presented the discussion and limitations of the study. This chapter provides the conclusions of the findings and recommendations of the study.

6.2 Conclusions

Multiple factors are involved in mortality from breast cancer and there is still much to be learned about the biological, medical and socioeconomic mechanisms responsible for improving survival. The current study has shown that age, marital status, PR status, breast cancer stage, histologic grade and obesity were prognostic factors of mortality for breast cancer patients after surgery. The study also demonstrated that HIV-positive women with breast cancer appear to have decreased survival compared with HIV-negative women.

On the other hand, our study showed that there was no association between mortality of breast cancer patients after surgery with factors such as ER status, Her2 status, number of births, family history, contraceptive use, tumor laterality and surgery type. Although significant in univariate analysis, lymph node status, tumor size and alcohol consumption were not significantly associated with breast cancer mortality in the multiple logistic models.

In our study, 5-year mortality rate of breast cancer after surgery was similar to other rates estimated in many studies conducted world-wide especially in Africa. Comparisons with estimates for high-income countries showed that our results of mortality were much higher, which may be due to differences in the socioeconomic, demographic, environment, culture and health characteristics of the patients analyzed in each study. In addition, our study showed that patients older than 70 years had shorter survival time followed by middle age group in comparison with younger patients [35 years less], as one would expect. Early diagnosis of breast cancer and early treatment can increase survival based on this finding.

These results challenge the country to improve implementation of breast cancer screening programmes in the communities in order to reduce late detection of the disease. However, educational campaigns should be the first step for the reason that early detection cannot be successful if the population is unaware of the problemme or has adverse misconceptions about the value of early detection. Therefore, an early detection program and early treatment with support policies could reduce breast cancer mortality in Zambia.

6.3 Recommendations

Based on the findings from this study, the following recommendations were formulated:

- Social factors may point to behaviors such as physical exercise and social support, thus, BMI can be managed by promotion of regular physical exercise and eating healthy which has been shown to have positive effects in the prevention of breast cancer and in improving the daily life of breast cancer patients.
- Factors such as alcohol consumption, age at first child, smoking, contraceptives use, breast feeding, number of children, menopause, tumor size, lymph node status and education level did not work well in this study due to missing data in some patient's medical files. Thus, this study recommends that CDH improves collection of information in such a manner that there are no missing data in these variables to enable researchers assess the effect of the same on breast cancer mortality.
- In Zambia, medical care for breast cancer is not free and health insurance is not universal. Thus, the financial drain of the treatment course of breast cancer could be an obstacle for seeking care and appropriate treatment compliance especially when patients are required to pay for health care services. This could have contributed to the higher mortality of breast cancer patients after surgery among single patients in this study. Therefore, our study recommends that there is need to carry out research to compare employment status versus marital status with respect to breast cancer mortality after surgery. It also suggests that total coverage of health insurance should be promoted in Zambia in order to increase care accessibility to all women across the country.
- Further research should be conducted to ascertain the impact of distance between place of residence and CDH with respect to breast cancer mortality.
- Survival of breast cancer patients after surgery could be improved by moderating the effects of the factors. This could be done by examining both the clinical and social factors that have been proven to be significant in the model. Clinical factors may point to early screening while social factors may point to behavior modification such as physical exercise and social support. Since survival is better for women who had early stage of the disease, the study recommends that early detection of breast cancer along with the availability and accessibility of appropriate treatment should be made available to improve life expectancy for women with the disease in Zambia.

REFERENCES

[1] Ngatali C. F. S., Boumba L. M. A., Mabiala Y., Moukassa D and Nkoua M. J. B., 2019. Survival of Women with Breast Cancer at Pointe Noire (Congo Brazzaville). *Journal of Cancer and Oncology*. Vol.3, no.2: 1-8. DOI: 10.23880/oajco-16000152. ISSN: 2578-4625.

[2] Yazdani A., Yaseri M., Haghighat S., Kavani A and Zeraati H., 2019. Investigation of Prognostic Factors of Survival in Breast Cancer Using a Frailty Model: A Multicenter Study. *Breast Cancer: Basic and Clinical Research.* Vol.13: 1-10. Doi: 10.1177/1178223419879112. PMID: 31632048; PMCID: PMC6769229.

[3] Phung M. T., Tin S. T and Elwood M. J., 2019. Prognostic models for breast cancer: A systematic review. *BioMed Central Cancer.* Vol.19, no.230: 1-18. https://doi.org/10.1186/s12885-019-5442-6

[4] Zamani A., Nourelahi M., Talei A and Tahmasebi S., 2019. A Model to predict breast cancer survivability using logistic regression. *Middle East Journal of Cancer*. Vol.10, no.2: 132-138. Doi: 10.30476/mejc.2019.78569.

[5] Fan Y. P., Liu C. L., Chiangi I. J and Lin C. Y., 2011. Development of a prognostic nomogram for identifying those factors which influence the 2- and 5- year survival chances of Taiwanese women diagnosed with breast cancer. *European Journal of Cancer Care*. Vol.20: 620-626. Doi: 10.1111/j.1365-2354.2011.01240.x

[6] Bray F., Ferlay J. M. E., Soerjomataram M. D., Siegel L. R., Torre A. L and Jemal A., 2018. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. Vol.68: 394-424. Doi: 10.3322/caac.21492. PMID: 30207593.

[7] WHO, 2018. Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018. International Agency for Research on Cancer. Press release no 263.

[8] WHO, 2018. Globocan: Zambia global cancer observatory. World Health Organization. [Accessed on 20/03/20].

[9] National Cancer Control Strategic Plan 2016 – 2021. Ministry of Health, Zambia.

[10] American Cancer Society. *Breast Cancer Facts and Figures 2017-2018*. Atlanta: American Cancer Society, Inc. 2017.

[11] Jerez-Aragone J. J. M., Gomez-Ruiz J. A., Ramos-Jimenez G., Munoz-Perez J., Alba-Conejo E., 2003. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial Intelligence in Medicine.* Vol. 27: 45-63. Doi: 10.1016/s0933-3657(02)00086-6. PMID: 12473391.

[12] Gakwaya A., Mugambe J. B. K., Kavuma A and Kanyike D et al., 2008. Cancer of the breast: 5-year survival in a tertiary hospital in Uganda. *British Journal of Cancer*. Vol.99: 63-67. Doi: 10.1038/sj.bjc.6604435. PMID: 18577991.

[13] Delen D., Walker G and Kadam A., 2004. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*. Vol.34: 113-127. Doi: 10.1016/j.artmed.2004.07.002. PMID: 15894176.

[14] Balabram D., Turra M. C and Gobbi H., 2013. Survival of patients with operable breast cancer (stages I-III) at a Brazilian public hospital- a closer look into cause-specific mortality. *BioMed Central cancer of Journal,* Vol.13, no. 434: 1-10. Doi: 10.1186/1471-2407-13-434. PMID: 24063763.

[15] Khan H. M. R., Saxena A., Gabbidon K., Shauna-Jenne T. S and Bhatt C., 2014. Survival Analysis for white non-Hispanic female breast cancer patients. *Asian Pacific Journal of Cancer Prevention.* Vol.15, no.9: 4049-4054. Doi: 10.7314/apjcp.2014.15.9.4049. PMID: 24935595.

[16] Khan R. M. H., Saxena A., Gabbidon K., Rana S and Ahmed U. N., 2014. Model-based survival estimates of female breast cancer data. *Asian Pacific Journal of Cancer Prevention.* Vol.15, no.6: 2893-2900. Doi: 10.7314/apjcp.2014.15.6.2893. PMID: 24761921.

[17] Álvarez-Bañuelos M. T., Rosado-Alcocer L. M and Carvajal-Moreno M et al., 2016. Prognostic Factors Associated with Survival in Women with Breast Cancer from Veracruz, Mexico. *Journal of Cancer Science and Therapy.* Vol.8, no.4: 92-98. DOI: 10.4172/1948-5956.1000398.

[18] Lan H. N., Laohasiriwong W and Stewart F. J., 2013. Survival probability and prognostic factors for breast cancer patients in Vietnam. *Global Health Action.* Vol.6, no.18860: 1-9. Doi: 10.3402/gha.v6i0.18860. PMID: 23336619; PMCID: PMC3549066.

[19] Zaid L. Z. A., Nuzhat A and Rafiqe M., 2017. Factors affecting survival of women with breast cancer in King Fahad Medical City, Saudi Arabia. *International Journal of Community Medicine and Public Health.* Vol.4, no.4: 910-915. https://doi.org/10.18203/2394-6040.ijcmph20171306

[20] Traore B., Toure A., Sy T., Dieng M. M., Conde M., Deme A and Keita N., 2015. Prognosis of breast cancer patients underwent surgery in a developing country. *Journal of Cancer Therapy.* Vol.6, no.9: 803-810. DOI: 10.4236/jct.2015.69088.

[21] Chang E. T., Milne R. L., Phillips K. A., Figueiredo J. C., Sangaramoorthy M., Keegan M. H. T et al., 2009. Family history of breast cancer and all-cause mortality after breast cancer diagnosis in the breast cancer family registry. *Breast Cancer Research and Treatment.* Vol.117, no.1: 167-176. Doi: 10.1007/s10549-008-0255-3. PMID: 19034644; PMCID: PMC2728159.

[22] Malone K. E., Daling J., Doody D. R., O'Brien C. A., Resler A. J., Ostrander A. E and Porter L. P., 2011. Family history of breast cancer in relation to tumor characteristics and mortality in a population-based study of young women with invasive breast cancer. *Cancer Epidemiology, Biomarkers and Prevention.* Vol.20, no.12: 2560-2571. Doi: 10.1158/1055-9965.EPI-11-0781. PMID: 21960690; PMCID: PMC3242640.

[23] Melvin C. J., Wulaningsih W., Hana Z., Purushotham D. A and Hemelrijck V. M et al., 2016. Family history of breast cancer and its association with disease severity and mortality. *Cancer Medicine*. Vol.5, no.5: 942-949. Doi: 10.1002/cam4.648. PMID: 26799372; PMCID: PMC4864823.

[24] Aizer A. A., Chen M. H., McCarthy P. E., Mendu L. M and Nguyen L. P et al., 2013. Marital status and survival in patients with cancer. *Journal of Clinical Oncology*. Vol.31, no.31: 3869-3876. Doi: 10.1200/JCO.2013.49.6489. PMID: 24062405; PMCID: PMC4878087.

[25] Martinez M. E., Unkart T. J., Tao L., Kroenke H. C., Schwab R., Komenaka I and Gomez, L. S., 2017. Prognostic significance of marital status in breast cancer survival: A population-based study. *PLOS One*. Vol.12, no.5: 1-14. Doi: 10.1371/journal.pone.0175515. PMID: 28475579; PMCID: PMC5419505.

[26] Hinyard L., Wirth L. S., Clancy M. J and Schwartz T., 2016. The effect of marital status on breast cancer-related outcomes in women under 65: A SEER database analysis. *The Breast*. Vol.32: 13-17. Doi: 10.1016/j.breast.2016.12.008. PMID: 28012410.

[27] Spano J. P., Lanoy E., Mounier N., Katlama C., Costagliola D and Heard I., 2012. Breast cancer among HIV infected individuals from the ONCOVIH study, in France: Therapeutic implications. *European Journal of Cancer*. vol.48, no.18: 3335-3341. Doi: 10.1016/j.ejca.2012.05.019. Epub 2012 Jul 3. PMID: 22766516.

[28] Escarela G., Balandra A. J., Antonio G. N and Mosccoso A. G., 2017. Long-Term cause-specific mortality after surgery for women with breast cancer: A 20-year follow-up study surveillance. Epidemiology, and end results cancer registries. *Breast cancer: Basic and Clinical Research*. Vol.11: 1-11. Doi: 10.1177/1178223417711429. PMID: 28615951; PMCID: PMC5459512.

[29] Momenyan S., Baghestani A. R., Narges M., Naseri P and Akbari M. E., 2018. Survival prediction of patients with breast cancer: Comparisons of decision tree and logistic regression analysis. *International Journal of Cancer Management*. Vol.11, no.7; 1-8. doi: 10.5812/ijcm.9176.

[30] Chang W. T and Kuo L. Y., 2010. A model building exercise of mortality risk for Taiwanese women with breast cancer. *BMC Medical Informatics and Decision Making*. Vol.10, no.43: 1-8. https://doi.org/10.1186/1472-6947-10-43. ISSN: 1472-6947

[31] American Cancer Society (ACS), 2015. *Global Cancer facts and figures* 3rd edition. Atlanta, Georgia.

[32] Vanderpuye V., Hammad N. G., Prabhakar P., Simonds H., Olopade F and Stefan C. D., 2017. An update on the management of breast cancer in Africa. *Infectious Agents and Cancer*. Vol.12, no.13: 1-12. Doi: 10.1186/s13027-017-0124-y. PMID: 28228841; PMCID: PMC5307840.

[33] Seneviratne S. A., Campbell I. D., Scott N., Lawrenson R. A., Shirley R and Elwood J. M., 2015. Risk factors associated with mortality from breast cancer in Waikato, New Zealand: a case-control study. *Public Health*. Vol.129, no.5: 549-554. Doi: 10.1016/j.puhe.2015.02.008. PMID: 25753278.

[34] Karimi A., Delpisheh A., Sayehmiri K., Saboori H and Rahimi E., 2014. Predictive factors of survival time of breast cancer in Kurdistan province of Iran between 2006-2014: A cox regression approach. *Asian Pacific Journal of Cancer Prevention*. Vol.15, no.19: 8483-8488. doi: 10.7314/apjcp.2014.15.19.8483. PMID: 25339051.

[35] Peng J. C., 2002. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*. Vol.96, no.1: 3-14. ISSN: ISSN-0022-0671.

[36] Huang S. H., Loh J. K., Tsai J. T., Houg M. F and Shi H. Y., 2017. Predictive model for 5-year mortality after breast cancer surgery in Taiwan residents. *Chinese Journal of Cancer*, Vol.36, no.23: 1-9. Doi: 10.1186/s40880-017-0192-9. PMID: 28241793; PMCID: PMC5327555.

[37] Banda L, Nyirongo T, Muntanga M., 2019. Cancer – An Emerging Health Problem: The Zambian Perspective. *Health Press Zambia Bull*. Vol.3, no.2: 2-4.

[38] Kawaguchi H., Masuda N., Nakayama T., Aogi K and Ohno S et al., 2019. Factors associated with prolonged overall survival in patients with postmenopausal estrogen receptor-positive advanced breast cancer using real-world data: a follow-up analysis of the JBCRG-C06 Safari study. *Breast Cancer*. Vol.27; 389–398. Doi: 10.1007/s12282-019-01029-3. PMID: 31811519; PMCID: PMC7196081.

[39] Adeloye D., Sowunmi Y. O., Jacobs W., Misra S., Gadanya M and Chan K. Y et al., 2018. Estimating the incidence of breast cancer in Africa: a systematic review and meta-analysis. *Journal of Global Health*. Vol.8, no.1: 1-21. Doi: 10.7189/jogh.08.010419. PMID: 29740502; PMCID: PMC5903682.

[40] Obrist M., Osei-Bonsu, E., Merajver D. S., Schmid K., Awuah B., Watanabe-Galloway S and Soliman A. S., 2015. Factors related to incomplete treatment of breast cancer in Kumasi, Ghana. *The Breast*. Vol.23, no.6: 821-828. Doi: 10.1016/j.breast.2014.08.014. PMID: 25282667; PMCID: PMC4274250.

[41] Brinton A. L., Awuah B., Clegg-Lamptey J. N., and Nyarko M. K. et al., 2017. Design considerations for identifying breast cancer risk factors in a population-based study in Africa. *International Journal of Cancer*. Vol.140, no.12: 2667-2677. Doi: 10.1002/ijc.30688. PMID: 28295287; PMCID: PMC5926189.

[42] Amin M. S., Ewunonu S. A., Oguntebi E and Liman M. I., 2017. Breast cancer mortality in a resource-poor country: A 10-year experience in a tertiary institution. *Sahel Medical Journal*. Vol.20, no.3: 93-97.

[43] Ngowa J. D. K., Kasia M. J., Yomi J., Sando Z., and Ndom P. et al., 2015. Breast cancer survival in Cameroon: Analysis of a cohort of 404 patients at the Yaounde General Hospital. *Advances in Breast Cancer Research*. Vol.4: 44-52. DOI: 10.4236/abcr.2015.42005

[44] Makanjuola B. L. S., Popoola O. A and Oludara A. M., 2014. Radiation therapy: A major factor in the five-year survival analysis of women with breast cancer in Lagos, Nigeria. *Radiotherapy and Oncology*. Vol.111: 321-326. Doi: 10.1016/j.radonc.2014.03.013. PMID: 24746579.

[45] Seedhom E. A and Kamal N. N., 2011. Factors affecting survival of women diagnosed with breast cancer in El-Minia Governorate, Egypt. *International Journal of Preventive Medicine.* Vol.2, no.3: 131-138. PMID: 21811654; PMCID: PMC3143525.

[46] Coghill A. E., Newcomba P. A., Madeleinea M. M., and Casper C et al., 2013. Contribution of HIV infection to mortality among cancer patients in Uganda. *National Institutes of Health (NIH) Public Access.* Vol. 27; no.8; 2933–2942. Doi: 10.1097/01.aids.0000433236. 55937.cb.

[47] Laamiri F. Z., Bouayad, A., Hasswane, N., Ahid, S., Mrabet, M and Amina, B., 2015. Risk factors for breast cancer of different age groups: Moroccan data? *Open Journal of Obstetrics and Gynecology*. Vol.5: 79-87. DOI: 10.4236/ojog.2015.52011.

[48] Bouzguenda R., Khanfir, A., Lahiani, F., Ayedi, I., Daoud, J and Frikha, M., 2013. Prognostic factors and survival in metastatic breast cancer: A single institution experience. *Reports of Practical Oncology and Radiotherapy*. Vol.18: 127-132. Doi: 10.1016/j.rpor.2013.01.001. PMID: 24416543; PMCID: PMC3863182.

[49] Mumba L. R., 2017. Knowledge attitude and practice of breast self-examination for early detection of breast cancer among women in roan constituency in Luanshya, Copperbelt province, Zambia. *Asian Pacific Journal of Health Sciences*. Vol.4, no.3: 74-82. DOI: https://doi.org/10.21276/apjhs.2017.4.3.12.

[50] Mulenga K, Bwalya M., Sichilima A., 2018. Knowledge, attitudes and practices towards breast cancer among women attending Obsteteric and Gynaecology clinic at Ndola Teaching Hospital in Ndola, Zambia. *Health Press Zambia Bulletin*. Vol.2, no.8: 7-15.

[51] Hsieh F. Y., Bloch D. A and Larsen D. M., 1998. A simple method of sample size calculation for linear and logistic regression. *Statistical medicine*. Vol.17: 1623-1634.

[52] Agresti A., 2002. *Categorical Data Analysis.* 2nd Edition, New Jersey, USA: John Wiley and Sons, Inc.

[53] Jewell N. P., 2004. Statistics for Epidemiology. Chapman and Hall/CRC. New York City.

https://doi.org/10.1201/9781482286014

[54] Hosmer D, Lemeshow S., 2000. *Applied logistic regression*, 2nd edition. New York, USA: John Wiley and Sons.

[55] Park H. A., 2013. An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy Nursing.* Vol.43, No.2: 154-64. doi: 10.4040/jkan.2013.43.2.154. PMID: 23703593.

[56] Bewick V., Cheek, L and Ball, J., 2005. Statistics review 14: Logistic regression. *Critical Care (London, England).* Vol.9, no.1: 112-118.

[57] Czepiel A. S., 2002. Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. http://czep.net/contact.html. Accessed 15-04-24 at https://czep.net/stat/mlelr.pdf

[58] Szumilas M., 2010. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*. Vol.19, no.3: 227-229. PMID: 20842279; PMCID: PMC2938757.

APPENDICES

Appendix A: Ethical clearance approval from NASREC

THE UNIVERSITY OF ZAMBIA

DIRECTORATE OF RESEARCH AND GRADUATE STUDIES

Great East Road | P.O. Box 32379 | Lusaka 10101 | **Tel:** +260-211-290 258/291 777
Fax: +260-1-290 258/253 952 | **Email:** director@drgs.unza.zm | **Website:** www.unza.zm

Approval of Study

15th August, 2019

REF NO. NASREC: 2019-JUN-010

Mr. Mulope Mulope
The University of Zambia
School of Natural Sciences
Department of Mathematics and Statistics
Box 32379
LUSAKA

Dear Mr. Mulope,

RE: "RISK FACTORS ASSOCIATED WITH MORTALITY OF BREAST CANCER PATIENTS AFTER SURGERY: THE CASE OF ZAMBIA"

Reference is made to your resubmission. The University of Zambia Natural and Applied Sciences Research Ethics Committee IRB resolved to approve this study and your participation as Principal Investigator for a period of one year.

Review Type	**Ordinary /Expedited Review**	**Approval No. REF No. NASREC: 2019-JUN-010**
Approval and Expiry Date	Approval Date: 15th August, 2019	Expiry Date: 14th August, 2020
Protocol Version and Date	Version-Nil	14th August, 2020
Information Sheet, Consent Forms and Dates	• English.	To be provided
Consent form ID and Date	Version	To be provided
Recruitment Materials	Nil	Nil

Excellence in Teaching, Research and Community Service re are specific conditions that will apply to this approval. As Principal Investigator your responsibility to ensure that the contents of this letter are adhered to. If these not adhered to, the approval may be suspended. Should the study be suspended, dy sponsors and other regulatory authorities will be informed.

nditions of Approval

- No participant may be involved in any study procedure prior to the study approval or after the expiration date.
- All unanticipated or Serious Adverse Events (SAEs) must be reported to the IRB within 5 days.
- All protocol modifications must be IRB approved by an application for an amendment prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address or methodology and methods. Many modifications entail minimal risk adjustments to a protocol and/or consent form and can be made on an Expedited basis (via the IRB Chair). Some examples are: format changes, correcting spelling errors, adding key personnel, minor changes to questionnaires, recruiting and changes, and so forth. Other, more substantive changes, especially those that may alter the risk-benefit ratio, may require Full Board review and approval. In all cases, except where noted above regarding subject safety, any changes to any protocol document or procedure must first be approved by the IRB before they can be implemented.
- All protocol deviations must be reported to the IRB within 5 working days.
- All recruitment materials must be approved by the IRB prior to being used.
- Principal investigators are responsible for initiating Continuing Review proceedings. Documents must be received by the IRB at least 30 days before the expiry date. This is for the purpose of facilitating the review process. Any documents received less than 30 days before expiry will be labelled "late submissions" and will incur a penalty.
- Every 6 (six) months a progress report form supplied by The University of Zambia Natural and Applied Sciences Research Ethics Committee IRB must be filled in and submitted to us. There is a penalty of K500.00 for failure to submit the report.
- The University of Zambia Natural and Applied Sciences Research Ethics Committee IRB does not "stamp" approval letters, consent forms or study documents unless requested for in writing. This is because the approval letter clearly indicates the documents approved by the IRB as well as other elements and conditions of approval.

Should you have any questions regarding anything indicated in this letter, please do not hesitate to get in touch with us at the above indicated address.

On behalf of The University of Zambia Natural and Applied Sciences Research Ethics Committee (IRB), we would like to wish you all the success as you carry out your study.

Yours faithfully,

Dr. E. Mwanaumo

CHAIRPERSON
THE UNIVERSITY OF ZAMBIA NATURAL AND APPLIED SCIENCES RESEARCH ETHICS COMMITTEE IRB

: Director, Directorate of Research and Graduate Studies
Assistant Director (Research), Directorate of Research and Graduate Studies
Assistant Registrar (Research), Directorate of Research and Graduate Studies
Senior Administrative Officer (Research), Directorate of Research and Graduate Studies

Appendix B: Permission letter from CDH Management to access the data

espondence should be addressed to the
ledical Superintendent
-260 211 257706

In reply please

MH/CDH/101/14/1

No. ....

REPUBLIC OF ZAMBIA

MINISTRY OF HEALTH

CANCER DISEASES HOSPITAL

P.O. Box Rw 5
LUS

5th September, 2019

Mr. Mulope Mulope
University of Zambia
Department of Natural Sciences
P.O. Box 50110
LUSAKA

Dear Mr. Mulope

Re: Approval to conduct research – Yourself

Reference is made to the above.

I wish to inform you that the Cancer Diseases Hospital (CDH) has no objection to your request to conduct research at our institution, entitled: **"Risk Factors associated with mortality of breast cancer patients after surgery: The case of Zambia."** However, this permission is subject to you obtaining ethical clearance for your research.

You are required to come with a copy of the ethical clearance letter, and a copy of this letter in order for you to proceed to conduct data collection.

Yours sincerely,

Dr. Lewis Banda
SENIOR MEDICAL SUPERINTENDENT