UNIVERSITY FOR DEVELOPMENT STUDIES

SURVIVORSHIP ANALYSIS AND BREAST CANCER RISK FACTORS AMONG GHANAIAN WOMEN

ALICE CONSTANCE ABAKAH

THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS OF THE FACULTY OF MATHEMATICAL SCIENCES, UNIVERSITY FOR DEVELOPMENT STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF PhD DEGREE IN APPLIED STATISTICS

2013
UNIVERSITY FOR DEVELOPMENT STUDIES

SURVIVORSHIP ANALYSIS AND BREAST CANCER RISK FACTORS AMONG GHANAIAN WOMEN

BY

ALICE CONSTANCE ABAKAH (M.Sc in Statistics)

UDS/DAS/0005/09

THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS OF THE FACULTY OF MATHEMATICAL SCIENCES, UNIVERSITY FOR DEVELOPMENT STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF PhD DEGREE IN APPLIED STATISTICS

NOVEMBER, 2013
I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this University or elsewhere:

Candidate's Signature: .................................................. Date: .............. 6/11/13

Name: Alice Constance Abakah

Supervisors'

I hereby declare that the preparation and presentation of the thesis was supervised in accordance with the guidelines on supervision of thesis laid down by the University for Development Studies.

Principal Supervisor’s Signature: .................................. Date: .............. 06/11/2013

Name: Prof. K.S. Nokoe

Co-Supervisor’s Signature: .............................................. Date: .............. 06/11/2013

Name: Dr. Joel Yarney
Abstract

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in female’s worldwide. The objective of the study was to identify and describe breast cancer survival pattern in Ghana and factors that explain the disparity in survival rates for breast cancer by the use of Cox proportional hazard. The extent to which selected demographic, hormonal and reproductive factors influence the cause of breast cancer were also examined using the logistic regression technique to determine risk of getting the disease. Two thousand three hundred and ninety seven (2397) women were sampled for the study, of which 1022 (42.64%) were diagnosed with breast cancer between the periods 1st January 2002 to 31st December 2008. The overall mean age of patients was 43.51 years whiles that for breast cancer cases were 47.97 years. The highest number of cases being 351 (34.34%) was aged 40-49 years. Cumulative 5-year survival was 47.91%. Main mortality risk factors were, tumour size (logrank <0.0001); axillary node involvement (logrank < 0.0064); stage at diagnosis (p = 0.000) and BMI was 0.0015. Survival rate was better for early staged presentation (59.93%); obese women (50.26%); lymph node involvement of less than 25% (63.26%) and tumour size of less than 5cm was 52.30%. The gamma model was found to be the best fitted model for predicting survival following a diagnosis of breast cancer. Breast feeding, late menarche, contraceptive usage, and time interval between age at menarche and age at menopause decreased the risk of breast cancer development. Later age at menopause (OR = 2.306, < 0.0001) on the other hand increased the risk of breast cancer development. There existed some relationships among some of the risk factors using path analysis namely age, parity, breast feeding and age at first child on breast cancer development. It is recommended that governmental/nongovernmental organizations improve on health education/campaigns about breast cancer to create awareness and reduce mortality.
I would like to express my gratitude to my supervisors. I am deeply indebted to my principal supervisor, Professor K.S. Nokoe, for his thoughtful guidance, invaluable expertise, intellectual support and encouragement throughout the period of my study, which made this thesis possible. I am also indebted to my co-supervisor, Dr. Joel Yarney (Head of the National Centre for Radiotherapy and Nuclear Medicine Korle-Bu Teaching Hospital), who provided me with invaluable suggestions, comments and constructive discussion during the period of this study that significantly improved my work. I am very grateful again for the unique perspectives that they provided throughout this project, including their clinical insight and methodological guidance. I assume, of course, full responsibility for any remaining errors.

In addition, I am greatly indebted to Accra Polytechnic, which granted me financial support to pursue the programme.

I would like to express my sincere thanks to Messers E. Enchil, Philip Oduro, Isaac Aidoo, Isaac Appiah and Samuel Egyir and Miss Eva Abakah who provided me with assistance in the data collection. I am thankful to and the radiotherapy unit staff of the Korle-Bu Teaching Hospital, who kindly provided me with all the materials that I required and for giving me the permission to browse through the files for the collection of my data. I am also grateful to Dr. S.Y. Opoku for his assistance in seeking ethical clearance.

I remain grateful to all members of the Faculty of Mathematical Sciences for their support and encouragement during my PhD journey especially Prof. Gbolagade, Dr. Luguterah, Dr. Seini and Dr. Twum.

For my husband, B.B.Mensah, Mother, brother and sisters, no word can express my love for your understanding, tolerance and suffering through the duration of my studies. Special thanks to my dear children, for their patience towards their ‘absent mother’. Thanks to Pastor Henry K. Sarfo for his everlasting support and encouragement.
Dedication

I dedicate this thesis to my Children, Ewuradjoa Kofoa, Kweku Aidoo and Nana Ama Ampebah for their unconditional support and trust in me which enabled me to be confident in myself.
<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Acronyms</td>
<td>xiv</td>
</tr>
</tbody>
</table>

**CHAPTER 1 INTRODUCTION**

1.1 Background of the Study
   1.1.1 Symptoms of Breast Cancer

1.2 Statement of the Problem

1.3 Objectives of the Study

1.4 Significance of the Study

1.5 Organization of the Thesis

1.6 Limitation

**CHAPTER 2 LITERATURE REVIEW**

2.0 INTRODUCTION

2.1 STATE OF BREAST CANCER GLOBALLY
   2.1.2 Mortality and Survival Rates
2.2 State of Breast Cancer in Sub-Saharan Africa

2.2.1 Mortality and Survival Rates

2.3 STATE OF BREAST CANCER IN GHANA

2.4 BREAST CANCER RISK FACTORS

2.4.1 The Age Factor

2.4.2 Family and Personal History Factor

2.4.3 Gynecologic and Reproductive Factors

2.4.4 Body Mass Index Factor

2.4.5 Lifestyle, Culture, and Socioeconomic Status

2.4.6 Alcohol Consumption Factor

2.4.7 The Gender Factor

2.4.8 Contraceptive Usage Factor

2.5 BREAST CANCER PROGNOSTIC FACTORS

2.5.1 Axillary Node

2.5.2 Tumour Size

2.5.3 Histology / Grade

2.5.4 Hormone Receptors

2.5.5 Cancer Staging

2.5.5.1 Breast cancer stage grouping

2.5.5.2 Disease Presentation

2.5.6 Tumor Biology and Genetics

2.6 MODEL CLASSIFICATION AND EVALUATION

2.7 CONCLUSION
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0 INTRODUCTION</td>
<td>51</td>
</tr>
<tr>
<td>3.1 STUDY DESIGN</td>
<td>51</td>
</tr>
<tr>
<td>3.2 STUDY SETTING</td>
<td>51</td>
</tr>
<tr>
<td>3.3 TARGET AND POPULATION</td>
<td>54</td>
</tr>
<tr>
<td>3.4 STUDY VARIABLES</td>
<td>54</td>
</tr>
<tr>
<td>3.5 DATA COLLECTION, MANAGEMENT AND ANALYSIS</td>
<td>54</td>
</tr>
<tr>
<td>3.5.1 Data Collection</td>
<td>54</td>
</tr>
<tr>
<td>3.5.2 Data Management</td>
<td>55</td>
</tr>
<tr>
<td>3.5.3 Survival Analysis of the Data</td>
<td>56</td>
</tr>
<tr>
<td>3.5.4 Analyzing Survival Data</td>
<td>59</td>
</tr>
<tr>
<td>3.5.4.1 Methods of Analysis</td>
<td>59</td>
</tr>
<tr>
<td>3.6 SURVIVAL MODEL</td>
<td>60</td>
</tr>
<tr>
<td>3.6.1 The cumulative Distributive Function</td>
<td>60</td>
</tr>
<tr>
<td>3.6.2 The Probability Density Function</td>
<td>61</td>
</tr>
<tr>
<td>3.6.3 The Survival Function</td>
<td>61</td>
</tr>
<tr>
<td>3.6.4 The Hazard Function</td>
<td>62</td>
</tr>
<tr>
<td>3.7 KAPLAN-MEIER PRODUCT LIMITESTIMATE</td>
<td>65</td>
</tr>
<tr>
<td>3.7.1 Log-Rank Test</td>
<td>68</td>
</tr>
<tr>
<td>3.7.2 Cox Proportional Hazard Model</td>
<td>70</td>
</tr>
<tr>
<td>3.8 EMPIRICAL MODEL</td>
<td>73</td>
</tr>
<tr>
<td>3.9 LOGISTIC REGRESSION</td>
<td>75</td>
</tr>
<tr>
<td>3.9.1 Mathematical Background</td>
<td>76</td>
</tr>
</tbody>
</table>
3.9.2 Hypothesis Testing 81
3.9.3 Interpretation of the Logistic Regression Coefficients 83

3.10 DESCRIPTION OF VARIABLES 84

3.11 Path Analysis 86
3.11.1 Path Analysis with Logistic Regression 87

CHAPTER 4 ANALYSIS OF RESULTS

4.0 INTRODUCTION 89

4.1 PRELIMINARY RESULTS 89
4.1.1 Demographic Analysis 89
4.1.2 Analysis of Reproductive Factors 95
4.1.3 Hormonal, Lifestyle and Hereditary Factors 96
4.1.4 Social Characteristics 96

4.2 DESCRIPTIVE ANALYSIS OF PROGNOSTIC FACTORS 97
4.2.1 Tumour Characteristics 97
4.2.2 Clinical Staging Analysis 97
4.2.3 ER/PR Receptors 101
4.2.3 Analysis by Grading 103
4.2.4 Histology 103
4.3 Summary

4.4 SURVIVAL ANALYSIS

4.4.1 Kaplan-Meier Method

4.4.2 Log Rank Analysis of Tumour Size

4.4.3 Analysis of the clinical stage

4.4.4 Tumour Grade Analysis

4.4.5 Analysis of Body Mass Index

4.4.6 Proportion Hazard Model

4.4.7 AFT Model and Cox-PH Model

4.5 STEPWISE LOGISTIC REGRESSION ANALYSIS

4.6 PATH ANALYSIS

4.7 SUMMARY OF RESULTS

CHAPTER 5 DISCUSSIONS AND CONCLUSIONS

5.0 Introduction

5.1 Discussion

5.1.1 Survival Analysis

5.1.2 Prognostic Factors

5.1.3 Age

5.1.4 Breast feeding and Contraception Usage
5.1.5 Reproductive Factors

5.1.6 Family history

5.2 Conclusion

5.3 Contribution To Knowledge

5.4 Recommendation

References

Appendices

Appendix 1: Schedule for Breast Cancer Patients

Appendix 2: Schedule for Walk in Clinic

Appendix 3A: Frequency Distribution of all the Women by Age Group

Appendix 3B: Frequency Distribution of women with Breast Cancer by Age

Appendix 3C: Frequency Distribution of all the Women by Reproductive Factors

Appendix 3D: Frequency Distribution of Women with Breast Cancer by Reproductive Factors

Appendix 4A: No. of patients by Hormone Status and age group

Appendix 4B: No. of patients by Hormone Status and age group

Appendix 4C: Frequency Distribution of ER, PR and HER2
Table 2.1: 5-Year Breast Cancer Survival Rate Globally 18
Table 2.2 Age Distribution for Breast Cancer Incidence and Mortality in Ghana 26
Table 3.1 General Layout for Survival Analysis 68
Table 4.1 Descriptive Statistics of some Risk Factors for all the Women 88
Table 4.2 Descriptive Statistics of some Risk Factors for the Controls 89
Table 4.3 Descriptive Statistics of some Risk / Prognostic Factors for Breast Cancer Cases 90
Table 4.4 Description of Variables 95
Table 4.5 Death at Tumour Stage, Age and Menopausal Status 99
Table 4.6 ER/PR Receptors by Tumour Stage 101
Table 4.7 Histology Presentation by Age and Death 102
Table 4.8 Multivariate 5-year Survival Analysis of the Breast Cancer Patients 106
Table 4.9 5-year Survival by Age, Menopausal Status and Histology 113
Table 4.10 Cox Regression Analysis of 5 year Survival by Prognostic Factors 115
Table 4.11 Summary of Results of Fitting Parametric AFT Models to Survival 118
Table 4.12 Stepwise Logistic Analysis of Significant Risk Factors for all the Women 119
Table 4.13 Stepwise Logistic Analysis of Significant Risk Factors for Premenopausal 121
Table 4.14 Stepwise Logistic Analysis of Significant Risk factors for Postmenopausal 122
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The Structure of the Female Breast</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>Breast Cancer Death Rates between 1990 and 2006</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Incidence of Breast Cancer by Age</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Front View of Korle-Bu Teaching Hospital</td>
<td>52</td>
</tr>
<tr>
<td>3.2</td>
<td>Survival Curve in Theory</td>
<td>66</td>
</tr>
<tr>
<td>3.3</td>
<td>Survival Curve in Practice</td>
<td>67</td>
</tr>
<tr>
<td>3.4</td>
<td>Graph of Logistic Function</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Graph of Natural Log of the Logit Function against Odd Ratios</td>
<td>80</td>
</tr>
<tr>
<td>3.6</td>
<td>Hypothetical Path Model of Breast Cancer Risk Factors</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Distribution of the Women by Age Group</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>Distribution of Breast Cancer within Age Group</td>
<td>95</td>
</tr>
<tr>
<td>4.3</td>
<td>Percentage Staging of the Patients by Age Group</td>
<td>98</td>
</tr>
<tr>
<td>4.4</td>
<td>Distribution by Stage and Menopausal Status</td>
<td>99</td>
</tr>
<tr>
<td>4.5</td>
<td>ER/PR Receptors by Age Group</td>
<td>101</td>
</tr>
<tr>
<td>4.6</td>
<td>Survival Analysis for Breast Cancer cases within the Study Period</td>
<td>106</td>
</tr>
<tr>
<td>4.7</td>
<td>Kaplan Meier survival curves for Tumour Size</td>
<td>108</td>
</tr>
<tr>
<td>4.8</td>
<td>Kaplan Meier Survival Curves for Lymph Node</td>
<td>109</td>
</tr>
</tbody>
</table>
Figure 4.9: Kaplan Meier Survival Curves for Tumour Staging 110
Figure 4.10: Kaplan Meier Survival Curves for Tumour Grading 111
Figure 4.11: Kaplan Meier Survival Curves for Body Mass Index 112
Figure 4.12 Path Model of some Breast Cancer Risk Factors 124
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFC</td>
<td>Age at first Child</td>
</tr>
<tr>
<td>AFT</td>
<td>Accelerated Failure Time</td>
</tr>
<tr>
<td>AG</td>
<td>Age at first Visit</td>
</tr>
<tr>
<td>AIC</td>
<td>Arkaike Information Criterion</td>
</tr>
<tr>
<td>AJCC</td>
<td>American Joint Committee on Cancer</td>
</tr>
<tr>
<td>AL</td>
<td>Alcohol Intake</td>
</tr>
<tr>
<td>AN</td>
<td>Axillary Nodes</td>
</tr>
<tr>
<td>APM</td>
<td>Age interval between Menarche and Menopause</td>
</tr>
<tr>
<td>BC</td>
<td>Breast Cancer</td>
</tr>
<tr>
<td>BF</td>
<td>Breast Feeding</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CPHM</td>
<td>Cox Proportional Hazard Model</td>
</tr>
<tr>
<td>CTRP</td>
<td>Contraceptive Usage</td>
</tr>
<tr>
<td>DA</td>
<td>Dead or Alive</td>
</tr>
<tr>
<td>DCIS</td>
<td>Ductal Carcinoma Insitu</td>
</tr>
<tr>
<td>DG</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen Receptor</td>
</tr>
<tr>
<td>FH</td>
<td>Family History</td>
</tr>
<tr>
<td>GR</td>
<td>Tumour Grade</td>
</tr>
<tr>
<td>HRT</td>
<td>Hormone Replacement Therapy</td>
</tr>
<tr>
<td>HT</td>
<td>Height</td>
</tr>
</tbody>
</table>
IARC  International Agency for Research on Cancer  
IDC  Invasive Ductal Carcinoma  
ILC  Invasive Lobular Carcinoma  
IQR  Inter-Quartile Range  
KBTH  Korle-Bu Teaching Hospital  
LCI  Lobular Carcinoma In situ  
MN  Age at Menarche  
MP  Age at Menopause  
OC  Oral Contraceptives  
OR  Odds Ratio  
PR  Progesterone Receptor  
PT  Parity  
SD  Standard Deviation  
ST  Stage at Diagnosis  
TL  Tumour Location  
TS  Tumour Size  
UGMS  University of Ghana Medical School  
WT  Weight
CHAPTER ONE

1.1 BACKGROUND

Chronic or life-threatening illnesses can have a devastating impact on both the patient and the family. In today's new world of medicine, many consumers have come to realize that they are the ones who are primarily responsible for their own health care as well as for the health care of their loved ones.

When facing a chronic or life-threatening illness, one needs to become an educated consumer in order to make an informed health care decision. Essentially that means finding out everything about the illness - the treatment options, the doctors, and the hospitals – so that you can become an educated health care consumer and make the tough decisions.

When cancer starts in the breast, it is called breast cancer. Breast cancer is a malignant tumour in the glandular tissues of the breast. Such tumours, also called carcinomas, form when the processes that control normal cell growth breaks down, enabling a single abnormal cell to multiply at a rapid rate. Thus, Breast cancer starts when a single cell in the breast begins to divide and grow in an abnormal way. Breast cancer is not one single disease and there are several types of breast cancers. It can be diagnosed at different stages of development and can grow at different rates. Breast cancer is a complex disease and it is difficult to predict what course it will take, (Carey et al., 2006).
Breast cancer is the most common type of life-threatening cancer, and the second most common cause of cancer-related deaths of women in the Western world, American Cancer Society (2008).

In Ghana breast cancer is now the most common malignant disease in women and accounts for the majority of cancer related deaths (Wiredu et al., 2006).

In order to understand breast cancer, it helps to have some basic knowledge about the normal structure of the breasts. The female breast is made up mainly of lobules (milk-producing glands), ducts (tiny tubes that carry the milk from the lobules to the nipple), and stroma (fatty tissue and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels).

Figure 1.1 The Structure of the Female Breast
Most breast cancers begin in the cells that line the ducts (ductal cancers). Some begin in the cells that line the lobules (lobular cancers), while a small number start in other tissues American Cancer Society, 2008.

1.1.1 SYMPTOMS OF BREAST CANCER

Although widespread use of screening mammograms has increased the number of breast cancers found before they cause any symptoms, some breast cancers are not found by mammograms. Early breast cancer, for instance, usually does not cause symptoms. Thus, regular breast examination is important. The most common sign of breast cancer is a new lump or mass. A painless, hard mass that has irregular edges is more likely to be cancerous, but breast cancers can be tender, soft, or rounded.

According to the American Cancer Society (2009 – 2010), any of the following unusual changes in the breast can be a symptom of breast cancer:

- swelling of all or part of the breast (even if no distinct lump is felt)
- skin irritation or dimpling
- breast pain
- nipple pain or the nipple retraction (turning inward)
- redness, scaliness, or thickening of the nipple or breast skin
- a nipple discharge other than breast milk
- a lump in the underarm area
Sometimes a breast cancer can spread to underarm lymph nodes and cause a lump or swelling there, even before the original tumour in the breast tissue is large enough to be felt American Cancer Society 2008.

Symptoms of Advanced breast cancer may include:

- bone pain
- breast pain or discomfort
- skin ulcers
- swelling of one arm (next to breast with cancer)
- weight loss

These changes also can be signs of less serious conditions that are not cancerous, such as an infection or a cyst.

1.2 STATEMENT OF THE PROBLEM

To a woman, the breast is a significant expression of beauty, womanhood and an important piece of property. When it becomes diseased, many women become frightened, and Ghanaian women are no exception. It is unfortunate to note that cancer of the breast is responsible for the deaths of millions of women worldwide. Breasts have gone from representing fertility to that of a symbol of sexuality (Beckmann et al., 1983) and as such are highly symbolic of the processes that make up the construction of feminity (Carter, 1995). Early socialization and cultural views of the appearance and worth of women, together with social expectations of sexuality and feminity are unseen powerful forces which influence
the social construction of how women perceive their bodies (Kasper, 1995). The breast is often perceived by women to be an object for others to use, and is an important measure of a woman’s desirability and acceptance.

In all parts of the world today, cancer of the breast is a stark reality. For many years people have perceived breast cancer as a frightening disease surrounded by fears and myths.

According to Biritwum et al., (2000), in 1996 12.8% of all admissions for malignant neoplasms to the Korle Bu Teaching hospital were for breast cancer. This alarming difference is due to many factors. Myths and misconceptions about breast cancer include:

- Breast cancer is not curable
- Breast cancer is the result fate
- It is due to witchcraft or the devil
- It is a result of a person’s failures or faults
- It is the result of a curse on the patients family

Thus, the attitudes towards the disease ranged from fear which was linked to death in most cases; to denial and guilt; as well as the spiritual and supernatural attributes of the disease. These myths and misconceptions were supported by (Opoku et al., 2012). They reported that, respondents did not only link the disease with death but also the surgical treatment of the disease with death. This is because many women present themselves for treatment at the advanced stages of
the disease and as such died shortly after surgical intervention. It was also evident from their study that respondents were poorly informed about the risk factors and the signs and symptoms of the disease. Indeed the worst that a lump in the breast could be is that of cancer of the breast.

In Ghana breast cancer is a common cause of hospital admissions and mortality among women (Biritwum et al., 2000; Wiredu et al., 2006). Reported clinical cases from some sub Saharan African countries including Ghana, indicate that breast cancer in indigenous black African women population is often severe and with unfavourable prognostic features (Amir et al., 1994; Gukas et al., 2005). Some of these features include young age at presentation, advanced stage at diagnosis, large tumour size, high grade histologic subtypes and low rate of receptor positivity (Gukas et al., 2005; Gakwaya et al., 2008; Mbonde et al., 1998; Amir et al., 1997; Hassan et al., 1992; Clegg-Lamptey and Hodasi 2007; Mbonde et al., 2000). The features are believed to explain why African women are more likely than women from the developed countries to die from the disease (Gakwaya et al., 2008).

There are concerns about increasing rate of breast cancer among young people in Ghana. In addition to the fact that the incidence of the disease appears to be on the increase, late presentation with poor outcomes of treatment is the hallmark of breast cancer in most developing countries including Ghana (Opoku et al., 2012). It is also disturbing that the average age at diagnosis for breast cancer in Ghana is 46.29 years with a range of 26 to 80 years as compared to an average age of over 65 years in Europe and America (Anim 1993; NCRNM, 2007). In addition
(Ghartey, 2001) also reported that Ghanaian women tend to develop breast cancer at a younger age. Breast cancer is also an important contributor of mortality among women in developing countries like Ghana.

“The incidence of breast cancer in Ghana had been estimated at 50 to 70 cases in every 100,000 women. Health experts describe this as frightening, considering the fact that survival rate is very low due to low level of awareness, late detection and high cost of treatment which is estimated at between Gh¢2,000 and Gh¢3,000. Breast cancer accounts for fifteen per cent of all fatal cancer cases admitted to Korle-Bu, while half of all cancer patients also report to the hospital when the disease is in its advanced stage, making it difficult to manage.” These statements were made by Prof. Nii Oto Nartey, the Chief Executive Officer of the Korle-Bu Teaching Hospital (KBTH) at the opening of a Regional African Co-operation Research Agreement (AFRA) Training Course in Breast Cancer in Accra, 2011.

Prognosis and possible cure from cancer are important measures of lifetimes which can be assessed by analyzing the survival of cancer patients. Different statistical approaches are used for analyzing the cancer survival data. Cancer survival rates or survival statistics tells the percentage of people who survive a certain type of cancer for a specific amount of time. The results of survival analysis for cancer patients have been widely presented and reported for different human sub populations of the globe (Woolson, 1981; Kardaun, 1983; Beadle et al., 1984; Sedmak et al., 1989).

According to SEER, 80.4% of women with breast cancer survive after 5 years in the US 1983 – 90. The American cancer society cancer facts and figures (2009 –
reports that 98% of women survive breast cancer if it is detected while it is 2cm in diameter; 88% if it is detected while the tumour size is 2 – 5cm and has spread to axillary lymph nodes; 76% survive breast cancer if it is detected even over 5cm in diameter and has not spread to axillary lymph node.

However, very few survival results at national level are available for the female population of Ghana. The statistical evidence about the survival of the breast cancer patients at the KBTH is scanty if not available in the literature.

Moreover, the actual cause of breast cancer remains unknown. Studies have shown that certain factors are intimately associated with breast cancer some of which are amenable to change. For this study, established risk factors that are well described in the breast cancer literature (Chlebowski et al., 2005; Hall et al., 2005; Garcia-Closas et al., 2006), and commonly acknowledged by major cancer organizations, such as the American Cancer Society and the Susan G. Komen Foundation were considered. These factors include reproductive history, family history of breast cancer, menstrual history, hormone use, alcohol consumption, physical activity, height and body mass index. The influence of anthropometric measures on breast cancer risk has been the subject of many studies (Sasieni et al., 2011; Anyanwu 2000; Jatoi, 1999; Burke 2000; Tessaro et al., 2001; Kahlenborn et al., 2008; Casey et al., 2008; Lower 2008; Ansink and Burger 2007; Hoffman et al., 2000). These risk factors; age, oral contraceptive use, reproductive history, family history of breast cancer and menstrual history were linked to increase breast cancer risk. However, these findings were derived from studies conducted mainly in Western countries. The contribution of well-
established risk factors to breast cancer among developing countries remains unclear. Therefore, this thesis focused on breast cancer survival pattern in Ghana and then evaluated the prevalence of established breast cancer risk factors and their associations with breast cancer risk using logistic regression.

1.3 OBJECTIVES OF THE STUDY

A number of demographic and socio-economic variables have been speculated to be associated with breast cancer. There is the need to verify these suppositions by giving a rigorous empirical basis for the identification of variables that affect survival rates and also determine the likelihood of one getting breast cancer. To this end, the study seeks to:

- Identify and describe breast cancer survival pattern in Ghana.
- Identify and evaluate the factors that could explain the disparity in survival rates for breast cancer in Ghana.
- Examine the extent to which selected demographic, hormonal and reproductive factors influence the cause of breast cancer by using a logistic regression mode.
- Determine the likelihood of one’s risk of getting the breast cancer based on the logistic regression model.
- Use Path analysis to identify direct and indirect effects of the risk factors on breast cancer development.
1.4 SIGNIFICANCE OF THE STUDY

The following are identified notable significance of the study:

- Since breast cancer elicits many fears, findings of this study can provide useful information and knowledge to help the victims make the best decisions concerning their care.
- Heightened awareness of breast cancer risk will lead to increase in women undergoing mammography screening leading to the detection of cancers in earliest stages and results in improved survival rates.
- In treating cancer, doctors can look out for those factors identified in the study to determine where the cancer is most vulnerable and help determine how best to treat it.

1.5 ORGANISATION OF THE DISSERTATION

The rest of the dissertation is structured as follows: Chapter 2, provides a discussion of the literature on breast cancer survival, incidence and mortality. Chapter 3 focuses on the research methodology adopted in the empirical analysis as well as the data issues considered. The main findings are given in Chapter 4 while Chapter 5 presents the discussions, conclusions and contribution to knowledge.

1.6 LIMITATIONS

The study has a number of limitations. First, the national cancer registry in Ghana is not fully functional and local data on breast cancer stage distribution were derived from Korle Bu Teaching Hospital, Accra. As this is based on presenting
patients rather than on all patients, the data may not reflect reality and may be biased.

Secondly the overall survival rate may be deficient since some patients may have died from other causes other than breast cancer.

Another limitation was participants' recall of their symptoms and information. The primary exposures of interest include early-life events, and given that some women participants are older than 50 years, recall of exact events may have been poor for some women resulting in exposure of misclassification.
CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

Breast cancer starts when a single cell in the breast begins to divide and grow in an abnormal way. Breast cancer is a complex disease difficult to predict and to cure in its various types, and its behaviour depends on the histological and molecular subtype.

2.1 STATE OF BREAST CANCER GLOBALLY

Breast cancer is the most commonly diagnosed cancer worldwide with nearly 1,000,000 new cases diagnosed per year and the second leading cause of cancer deaths in women worldwide (Ferlay et al., 2001). In 2008, approximately 1.4 million women were diagnosed with breast cancer worldwide with corresponding 460,000 deaths. Of these, approximately 450,000 women were diagnosed with the disease in Europe with a corresponding 140000 deaths, while 68000 women were reportedly diagnosed with the disease in Africa with a corresponding 37000 deaths (Ferlay et al., 2010).

Forouzanfar et al., (2011), reported the following: The number of new cases of breast cancer has jumped dramatically worldwide, from about 640,000 in 1980 to more than 1.6 million in 2010. In the year 2010, 51 percent of new cases of breast cancer were reported in developing countries. One out of every 46 women had that risk. In addition, among women aged 15 to 49 there were twice as many cases of breast cancer in developing countries than in developed countries. Deaths from breast cancer were also higher in developing countries compared with
developed countries. However, around the world the increase in deaths from breast cancer has been slower than the increases in cases. This may be due possibly to early detection and treatment advances in developed countries. According to Forouzanfar et al., (2011), the risk has gone from one in 64 women dying to one in 35 in Zimbabwe. Again not only is the threat of breast cancer and cervical cancer shifting more heavily toward developing countries, it is also shifting to women of reproductive age. For example, in Bangladesh, more than 60 percent of women who died from breast cancer are under age 50.

According to the IARC report (2002), globally, each year, more than 1.1 million women will be diagnosed with breast cancer and more than 410,000 of them will die from the disease. In the USA, according to the report, nearly 240,000 will be diagnosed with breast cancer and nearly 40,000 of them will die from the disease. Over the next 25 years, the report said, 10 million people around the world could lose their lives to breast cancer. Globally a case of breast cancer is diagnosed every 29 seconds. A woman dies from breast cancer every 75 seconds worldwide.

Breast cancer is the most common cancer in the UK. More than 44,300 women and around 300 men were diagnosed with breast cancer in the UK in 2004, (Ferlay et al., 2008). Due to earlier detection and improved treatment survival rates for breast cancer are improving. It is estimated that around 172,000 women are living in the UK who have been diagnosed with breast cancer in the last 10 years.

In terms of number of new cases, breast cancer is the third most frequent cancer in the world (796,000 cases in 1990) and by far the most common malignancy of
women (21% of all new cases). In the UK, 85% of breast cancer is detected at the early stage. With early stage detection, breast cancer is 95% curable.

Worldwide, the ratio of mortality to incidence is about 61%. Incidence rates are high in all of the developed countries (except Japan, where it is third after stomach and colon and rectal cancer), with the highest age-standardized incidence in the United States (87.1 per 100,000). The incidence is more modest in North Africa, South America, and Eastern, Southeastern, and Western Asia, but it is still the most common cancer of women in these geographic regions. The rates are low (less than 30 per 100,000) in most of sub-Saharan Africa (except South Africa) and in Asia. The lowest incidence rate of 11.8 per 100,000 is in China. Age-adjusted breast cancer detection rates in North America, Northern Europe, Australia, and New Zealand average 95–100 cases for 100,000 persons. In contrast, incidence is quite low in Western Africa and Eastern Asia, at approximately 20 per 100,000 (Ferlay et al., 2008).

In Ghana, only 25% of breast cancers are detected at the early stage, 75% are presented at late stage.

2.1.2 MORTALITY AND SURVIVAL RATES

The death rate for breast cancer in women has decreased since 1990, according to the American Cancer Society, Surveillance Research (2009 - 2010). This is illustrated in Figure 2.1.
Figure 2.1: Breast Cancer Death Rates between 1990 and 2006 (American Cancer Society, Surveillance Research, 2009 – 2010)

The percentage decline was larger among younger age groups. From the year 1990-2006, death rates decreased by 3.2% per year among women younger than 50, and by 2.0% per year among women 50 or older. The decline in breast cancer mortality has been attributed to both improvements in breast cancer treatment and early detection. Generally, African American women and women of other racial and ethnic groups have benefited less than white women from these advances. From the years 1997-2006, female breast cancer death rates declined by 1.9% per year in non-Hispanic whites and Hispanics/Latinas, 1.6% in African Americans, 0.6% per year in Asian Americans/Pacific Islanders, and remained unchanged among American Indian/Alaska Natives. A striking divergence in long-term breast cancer mortality trends between African American and white women began
in the early 1980s; by the 2006, death rates were 38% higher in African Americans than in white women.

In 1975, the incidence rate for female breast cancer in the United States was 105 new cases diagnosed for every 100,000 women in the population; the mortality rate was 31 deaths for every 100,000 women. Among women diagnosed with breast cancer during the period from 1975 through 1977, about 75% survived their disease at least 5 years. Among white women, the 5-year relative survival rate was 76%; among African American women, it was 62%.

In 2007, according to the latest updated statistics, (US department of health and human services 2010), the U.S. incidence rate for female breast cancer was approximately 125 new cases for every 100,000 women in the population; the mortality rate was approximately 23 deaths for every 100,000 women. Although the incidence rate in 2007 was higher than that in 1975, this rate had been declining since the years 1998-1999, when it had peaked at a rate of 141 new cases for every 100,000 women in the population. The breast cancer death rate in the United States has been declining steadily since 1989-1990, when it peaked at a rate of 33 deaths for every 100,000 women.

Among women diagnosed with breast cancer during the period from 1999 through 2006, 90% were expected to survive the disease for at least 5 years. Among white women, the 5 year relative survival rate was 91%; among African American women, it was 78%. The increase in breast cancer survival rate, seen since the mid-1970s, has been attributed to both screening and improved treatment, (US department of health and human services 2009 - 2010).
Five-year relative survival rates, standardised to the International Cancer Survival Standard (Corazziari et al., 2004), were calculated for patients aged 15-99 years diagnosed during the year 1990-94. Breast cancer survival rates varied from over 80% in North America, Sweden, Japan, Australia and Finland to less than 60% in Brazil and Slovakia and below 40% in Algeria. Most European countries including Scotland, England, Ireland and Wales, had rates in the 70-79% range (Coleman et al., 2008). As with the deprivation gap, a variety of factors are likely to affect these outcomes.

In Europe, Office for National Statistics in 2005 reported that, breast cancer survival had improved over time and inter-country survival differences were reducing. However, survival in the UK is far from the best and much lower than reported in the US. The most recent breast cancer survival rates in England were for women diagnosed in 2001 to 2006. For this group of women five-year relative survival was 82% compared with only 52% thirty years earlier in the year 1971-75, states (Richard 2008).
Table 2.1: 5-Year Breast Cancer Survival Rate Globally

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>Survival Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic White</td>
<td>88.8%</td>
</tr>
<tr>
<td>Black</td>
<td>77.5%</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>85.6%</td>
</tr>
<tr>
<td>Asian</td>
<td>90.7%</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>85.4%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>83.8%</td>
</tr>
</tbody>
</table>


2.2 STATE OF BREAST CANCER IN SUB-SAHARAN AFRICA

Cancer is an increasingly important public health problem in developing countries, including Africa, (Farmer 2010). As public and professional awareness of the cancer problem has grown, so has interest in the pattern of disease presentation, its epidemiology and treatment outcome. To date, however, there has been limited research about breast cancer in Africa. Africa has the highest death rate from the disease in the world, (Friedlin 2005).

There has been a significant increase in the incidence of breast carcinoma in sub-Saharan African countries and in other low-resource countries (Fregene et al., 2005 and Ly et al., 2011). There is also substantial variation in breast cancer
incidence and mortality rates among regions within the continent of Africa. Estimates of age standardized incidence rates (per 100,000 women) are 20.2 in Eastern Africa, 13.5 in Middle Africa, 24.8 in Western Africa, and 31.8 in Southern Africa, (Globaccon 2008).

According to the Uganda Breast cancer working group (2003), in Uganda, breast cancer incidence has doubled from 11 per 100,000 in 1961 to 22 per 100,000 in 1995. This increase has been attributed to the adoption of western lifestyles; however, improvements in data collection and reporting may also be contributing factors.

Breast cancer is reported to rank second in global cancer incidence and is the most common cancer diagnosis among Nigerian women according to some studies, (Ries et al., 2008; Parkin, et al., 2005; Adebamowo and Adekunle 1999; Adebamowo and Ajayi 2000).

African breast cancer patients are also more likely to be premenopausal (Amir et al., 1994); Anyanwu, 2000 and Hassan et al., 1992). Breast cancer incidence peaks between the ages of 35 and 45 years, approximately 10–15 years earlier than the peak incidence for Western countries outside of the Western Africa region, (Amir et al., 1998 and Ihekwaba 1992). A study from South Africa by (Muthuphei 1994) and from Zimbabwe, (Muguti, 1993) reported a second, but smaller, increase in incidence occurring between the ages of 60 and 69. Data from South Africa's National Cancer Registry (NCR) show breast cancer as the leading cancer among women (NCR 2001). South African women have a 1 in 29 lifetime risk of developing breast cancer, with an age-standardized incidence rate
of 30.6 per 100,000 population. These rates vary by race group, with Black women having the lowest (16.3) and White women the highest (69.4) rates of breast cancer diagnosis.

Reports from other regions in Africa demonstrate a consistent decline in breast cancer risk following menopause, which is in sharp contrast to the rising incidence rates seen among postmenopausal women from North America and Europe. It has been postulated that the lower postmenopausal breast cancer incidence rates observed for Africans are a consequence of demographics, especially population age and overall life expectancy (Adebamowo and Ajayi, 2000). Compared with expected longevity of American women (79 years), Ghanaian women have a life expectancy of 58 years; Nigerian, 51 years; Kenyan, 50 years; and South African women, 52 years (WHO 2000). This unfortunate truncation of lifespan precludes the ability to make robust conclusions regarding risk of postmenopausal breast cancer. It has also been suggested that breast cancer case ascertainment is disproportionately low among older African women because of their lower literacy rates, poor socioeconomic status, and diminished awareness of breast cancer. Breast cancer mortality rates are disproportionately high among African women compared with their generally low incidence rates. Estimated age standardized mortality rate for women with breast cancer in the year 2000 was 15.6 –21.4 in Africa compared with 27.7 in the USA and Canada, (Sandelin et al., 2002).

Within Africa, age-standardized mortality rates were estimated at 9.18 for Eastern Africa, 6.18 for Middle Africa, and 14.45 for Southern Africa (Globocan
2008). The higher mortality risk among African breast cancer patients can be attributed to several different factors, including delayed presentation, limited therapeutic modalities, and, perhaps, a predisposition to biologically aggressive tumors.

Although the proportion of deaths from breast cancer is higher in developed countries, worldwide population densities and demographics are such that over 50% of all breast cancer deaths occur in underdeveloped countries; it is also estimated that by the year 2020, approximately 70% of new cancer cases will occur among individuals in developing countries, and a substantial fraction of these are likely to be breast malignancies, (Jones 1999, WHO report 2001)

A review of cancer morbidity in adults in Ibadan, Nigeria, showed prostate cancer as the most common in males, whilst breast cancer topped the list in females by (Ogunbiyi et al., 1999). Breast cancer is a relatively unusual malignancy in African countries. Several investigators such as (Ijaduola et al., 1998; Ihekwaba 1992 and Ikpatt et al., 2002) have documented a younger age distribution and a greater prevalence of high grade, estrogen-receptor-negative disease among breast cancer patients in the Ghanaian and Nigerian populations of Western Africa. Similar to the patterns of breast cancer reported among African-American women. According to (Yarney et al., 2008) and (Stark et al., 2010) breast cancer in African women tend to be the aggressive triple negative subtype, similar to those observed in African-American women in the US, which is non-responsive to commonly used therapeutic drugs.
2.2.1 MORTALITY AND SURVIVAL RATES

Breast cancer mortality rate is much higher among Sub-Saharan women as compared to women in Western countries, even though the incidence rate is much higher in Western women (Fregene et al., 2005 and Ly et al., 2011). Apart from the fact that African women develop a more aggressive form of breast cancer, the higher mortality rate has been attributed to a general lack of public awareness of the disease, coupled with limited screening programs which often results in late diagnosis of the disease even after it has already metastasized to other organs (Wadler et al., 2011) Breast cancer is the second leading cause of death among African women.

In Uganda, a study of cancer of the breast, the 5-year survival showed that 23% had early disease (Stage 0-IIb) and 26% had metastatic disease. From breast cancer diagnosis, poorly differentiated was the most common pathological grade (58%) followed by moderately differentiated (33%) and well differentiated (9%) tumors and this is explained by poor levels of awareness in the population, lack of resources, lack of access to early diagnosis and treatment options in the country and lack of national breast cancer screening policy. The 5-year survival for early disease were 74% and 39% for advanced disease, (Gondos et al., 2005)

Statistics from Kampala cancer registry, indicate that median age of a person diagnosed with breast cancer is 45 (annual incidence rate of 46.8 per 100,000), the trend increases with age, reaching highest at age group of 50 (77.6 per 100,000) (Parkin, et al., 2007). A study of cancer survival in Kampala, Uganda showed that, the differences in survival between Ugandan and black American patient
populations was particularly dramatic for those cancer types for which early diagnosis and effective treatment is possible. The very poor prognosis of Ugandan patients is most likely explained by the lack of access to early diagnosis and treatment options in the country (Gakwaya et al., 2008)

2.3 STATE OF BREAST CANCER IN GHANA

Breast cancer is the leading malignancy in Ghana and accounts for 15.4% of all malignancies and appears to be on the increase (Badoe and Baako, 2000). In 1996 12.8% of all admissions for malignant neoplasms to the Korle Bu Teaching hospital were for breast cancer, (Biritwum et al., 2000). Fifty percent or more of Ghanaians with breast cancer report to the hospital with advanced disease (Asumanu et al., 2000; Archampong, 1977). On the average the women report 8 months or more after first noticing a change in their breasts, (Asumanu et al., 2000).

The World Health Organization (2008), has forecast that about 3000 women in Ghana will develop breast cancer each year. The report says that presently about 2000 women are diagnosed of breast cancer each year in Ghana. The report further estimates, that incidence of mortality ratio of breast cancer in Ghana is 0.68 as compared to 0.2 in the United States indicating that its prevalence in Ghana is relatively high.

Although breast cancer may not be a priority to international aid organizations due to the enormity of other health concerns, 70% of women who are diagnosed with breast cancer in Ghana happen to be in the advanced stages of the disease,
resulting in a higher mortality rate compared to high-income countries, (Kirby, 2005). In addition, a study has shown that Ghanaian women are more likely to be diagnosed with high-grade tumors that are triple negative breast tumor.

Explanations for the delayed presentation among Ghanaian women have been traced to the cost of, and access to, routine screening mammography, lack of awareness, and cultural attitudes according to Stark (2010). Mayo et al., (2003) also cited in particular, harsh social stigma, fear of mastectomy or death, and the appeal of traditional healers over doctors as some of the cultural reasons for late presentation of cases in Ghana. Furthermore, women with breast cancer in Ghana describe a feeling of hopelessness and helplessness, largely due to their belief in fatalism, which contributes to denial as a means of coping. Ghana is said to rank 10th on the ladder of the breast cancer burden in Africa.

The International Agency for Research on Cancer (IARC) has recently published (Globocan, 2002) estimates of cancer incidence, mortality and prevalence for Ghana, with the acknowledgement that the degree of detail and quality of this source of data (both the accuracy of the recorded cause of death and the completeness of the registration) vary considerably. The (Globocan,2002) estimates for Ghana are therefore a mixture of real data, extrapolations from limited samples, and informed guesses. GLOBOCAN 2002 reported that the top 10 causes of cancer mortality in descending order in females in Ghana were cervical, breast, liver, haematopoietic organs, stomach, colorectal, ovarian and bladder, with pancreas and Kaposi sarcoma tied for the 9th and 10th positions.
Indeed, cancers accounted for 2.6% of all admissions, and 5.6% of all deaths at the KBTH in the year 1996, (Biritwum et al., 2000). Malignancies of the breast were seen as the leading cause of deaths in females, followed closely by the haematopoietic organs, liver and the cervix. In males, the highest mortality was from the liver, followed closely by the prostate, the haematopoietic organs and the stomach.

The top 10 causes of cancer mortality in descending order in females in Ghana were breast, haematopoietic organs, liver, cervix, ovary, pancreas, stomach, colorectal, gall bladder and brain, (Wiredu et al., 2006). About 2,062 women were diagnosed of the disease country-wide 2011 alone.
Table 2.2 Age Distribution of Breast Cancer Incidence and Mortality in Ghana

<table>
<thead>
<tr>
<th>Age Groups</th>
<th>Female population*</th>
<th>Incidence (100 000)</th>
<th>Number of incident cases (%)</th>
<th>Mortality (100 000)</th>
<th>Number of Deaths (%)</th>
<th>Mortality/incidence ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–14</td>
<td>4 605 974</td>
<td>0.1</td>
<td>5 (0.2%)</td>
<td>0.00</td>
<td>0.00 (0.0%)</td>
<td>n/a</td>
</tr>
<tr>
<td>15–29</td>
<td>3 145 512</td>
<td>1</td>
<td>31 (1.1%)</td>
<td>0.4</td>
<td>13 (0.7%)</td>
<td>0.4</td>
</tr>
<tr>
<td>30–44</td>
<td>2 013 112</td>
<td>31.7</td>
<td>638 (22.5%)</td>
<td>11.3</td>
<td>227 (12.3%)</td>
<td>0.36</td>
</tr>
<tr>
<td>45–59</td>
<td>1 231 140</td>
<td>80.1</td>
<td>986 (34.8%)</td>
<td>53.2</td>
<td>655 (35.3%)</td>
<td>0.66</td>
</tr>
<tr>
<td>60–69</td>
<td>482 535</td>
<td>104.5</td>
<td>504 (17.8%)</td>
<td>84.4</td>
<td>407 (22.0%)</td>
<td>0.81</td>
</tr>
<tr>
<td>70–79</td>
<td>247 858</td>
<td>169.3</td>
<td>420 (14.8%)</td>
<td>143.7</td>
<td>356 (19.2%)</td>
<td>0.85</td>
</tr>
<tr>
<td>80+</td>
<td>90 061</td>
<td>273.5</td>
<td>246 (8.7%)</td>
<td>218.3</td>
<td>197 (10.6%)</td>
<td>0.8</td>
</tr>
</tbody>
</table>

*Based on population Ghana in 2009.

2.4 RISK FACTORS

A risk factor is something that increases the chance that a person will develop a disease. It is not a guarantee and does not predict a future diagnosis. Studies have shown that, certain factors are intimately associated with breast cancer some of which are amenable to change.

There are different kinds of risk factors. Some factors, like a person's age or race, can't be changed. Others are linked to cancer-causing factors in the environment. Still others are related to one's personal behaviour, such as smoking, drinking, and diet. Some factors influence risk more than others, and one's risk for breast cancer...
cancer can change over time, due to factors such as aging or lifestyle. This means there are a number of demographic and socio-economic variables that have the potential to cause breast cancer. It is therefore worthwhile to find empirically a basis for the determination of variables that actually cause the disease.

2.4.1 Age Factor

As we grow older, one's risk of developing breast cancer increases. It is estimated that 80% of women diagnosed with breast cancer are 50 years or older. This doesn't mean that younger women are not at risk, except that the frequency is much less. This is illustrated in Figure 2.2.

![Breast Cancer Incidence By Age](image)

**Figure 2.2: Incidence of Breast Cancer By Age**

A study conducted in UK by (Sasieni et al., 2011) indicated that the older the woman, the higher the risk. In another study, although breast cancer risk clearly increased as a function of age, African-American women under the age of 45 years had a greater incidence of disease than Caucasian-American women in this young age range. These rates equalized during the fifth decade of life, and for women over the age of 50 years, incidence rates for Caucasian Americans surpassed those of African Americans, resulting in an overall higher lifetime risk for the Caucasian Americans. Although the absolute values of these population-based incidence rates may not appear very large in magnitude, this ethnicity-related variation in age distribution is far more striking in clinical practice, where 20% of Caucasian-American breast cancer patients were younger than 50 years of age, compared with 30%-40% of African-American breast cancer patients, National Cancer Data Base (1995 - 2000) report.

2.4.2 Family and Personal History Factor

Family is generally defined as biological relatives. Having one first-degree relative (mother, sister, or daughter) with breast cancer approximately doubles a woman's risk, and having two first-degree relatives increases one's risk for breast cancer by five-fold (American Cancer Society, 1999). Therefore, women who have one or more first-degree relatives with breast cancer have two to five times the risk of developing breast cancer than women who do not have familial risk factors. While family history can play a role in breast cancer development, women should not subscribe to the popular belief that without a family history of breast cancer one was not at risk. The American Cancer Society estimates that 70...
to 80% of women with breast cancer do not have a family history of breast cancer. The risk conferred by a family history has been assessed in both case-control and cohort studies. In a pooled analysis of 38 studies, the relative risk of breast cancer conferred by a first-degree relative with breast cancer was two percent (Pharoah et al., 1997). Even though the individual still had a ninety-eight percent chance of not getting breast cancer, a risk of two percent was relatively high.

Anyanwu (2000) compiled breast cancer data from 4 different hospitals in Nigeria and found a positive family history in 4% of cases; 12% of patients reported a prior history of benign fibrocystic breast changes. Women who had previously been diagnosed and treated for breast cancer were at a greater risk of developing breast cancer again. Risk varies with the age at which the affected relative was diagnosed. The younger the age of diagnosis, the greater the risk posed to relatives (Hemminki and Vaittinen, 1998; Yang et al., 1998; Negri et al., 1997; Pharoah et al., 1997; Colditz et al., 1993; Slattery and Kerber, 1993).

Up to 40% higher estimated risks of breast cancer have been found in studies of breast cancer patients with a strong family history of the disease (Easton et al., 1995 and Ford et al., 1998). The difference in these risk estimates suggests the existence of genetic or shared environmental factors within families that modify the risk of breast cancer (Begg, 2002). Based on previous research, a patient who has a relative with first degree breast cancer has a higher risk of having breast cancer (Rawal 2006, Gui 2001, Colditz 1993). In addition, the patients who have family members with other cancers, such as ovarian, cervical, and lymphomatic
cancer, had a higher risk of breast cancer. Positive family history of breast cancer is associated with an increase in risk both for pre- and post-menopausal women (Romieu et al., 1996; Lee et al., 1992; Helmrich et al., 1983).

2.4.3 Gynecologic and Reproductive Factors.

Reproductive factors that increase the duration and/or levels of exposure to ovarian hormones, which stimulate cell growth, have been associated with an increase in breast cancer risk. These factors include early onset of menstruation (early menarche), late onset of menopause, later age at first pregnancy (first full term pregnancy), and never having given birth, (nulliparous). Breast cancer risk increases with early menarche and late menopause, and it is reduced by early first full term pregnancy (Jatoi, 1999). This risk factor is based on the theory that exposure to estrogen increases breast cancer risk (Davis et al., 1997). Pregnancy and breastfeeding both reduce a woman’s lifetime number of menstrual cycles, and thus her cumulative exposure to endogenous hormones (Colditz et al., 2006). In addition, pregnancy and breastfeeding have direct effects on breast cells, causing them to differentiate, or mature, so they can produce milk. (Britt et al., 2007 and Russo et al., 2005) hypothesized that these differentiated cells are more resistant to being transformed into cancer cells than cells that have not undergone differentiation.

The gynecologic and reproductive patterns within African populations tend to result in fewer ovulatory cycles over a lifetime, and this contributes to a decrease in breast cancer risk. Although published studies have generally been small, the trends observed have included late menarche, multiparity, initiation of
childbearing at young ages, and prolonged lactation. All of these features lead to lower endogenous estrogen levels over a lifetime, thereby diminishing cumulative breast cancer risk.

In an international population-based study, parous women were found to have decreased risk of breast cancer with an OR of 0.36 among young twins (less than 50 years), (Swerdlow et al., 2002).

A study in New Zealand suggest that women aged 45 to 49 years who were parous have some reduced risk of breast cancer compared to nulliparous women (OR=0.58), (McCredie et al., 1998). Interestingly, a study by Gilani et al., (2004) among Pakistani women aged less than 45 years found that being parous actually increased their risk of developing breast cancer.

In Norway it has been found that high parity is associated with an overall reduced risk of breast cancer. Among women 20-29 years of age, however, the results suggested increased risk with increasing parity. The protective effect of high parity was particularly strong among women with first birth before the age of 20 years, and rather weak among those with first birth at the age of 30 years or more (Albrektsen et al., 1998). Jatoi (1999) and Johnson-Thompson and Guthrie (1999) found that full term pregnancy reduces breast cancer risk and that breast cancer risk decreases with each full term pregnancy.

Women, who have their first full-term pregnancy at an early age, less than 24 years, have a decreased risk of developing breast cancer later in life. For example, in women who have a first full-term pregnancy before age 20, the risk of developing breast cancer is about half that of women whose first full-term
pregnancy occurs after the age of 30 (Bernstein, 2002). Again (Jatoi, 1999 and Johnson-Thompson and Guthrie, 1999) found that full term pregnancy reduces breast cancer risk and that breast cancer risk decreases with each full term pregnancy. This risk reduction is limited to hormone receptor-positive breast cancer; age at first full-term pregnancy does not appear to affect the risk of hormone receptor-negative breast cancer (Lord et al., 2008 and Ma et al., 2006).

Childbearing begins at relatively younger ages among African women. One study reported a median age at first pregnancy of 19 years (Muguti, 1993), among 28 women surveyed. McCredie et al (1998) and Weiss et al., (1998) found no association, while Lambe et al., (1996) and Talamin et al., (1996) identified a trend of increasing risk with increase age at first child. Many studies (e.g. Kelsey et al., 1993) have established that women who had their first full-term pregnancy earlier than 25 years had a lower risk of breast cancer than nulliparous women or women who had their first full-term pregnancy when they were above 30 years; additional pregnancies are associated with even lower risks. In the general population, younger age at first birth is associated with decreased risk of breast cancer, (Kelsey et al., 1993). The association between high parity and a reduced risk of breast cancer has been found to be particularly strong among women who first gave birth before the age of 20 years (Albrektsen, 1995). According Collaborative Group on Hormonal Factors in Breast Cancer in the UK, the younger the woman is when she begins childbearing, the lower her risk of breast cancer. The relative risk of developing breast cancer was estimated by the study to increase by 3% for each year of delay.
Menopausal history among African women has not been well documented. Of 14 patients surveyed in one study from Zimbabwe, the median age at menopause was 50 years (Muguti, 1993). Collaborative Group on Hormonal Factors in Breast Cancer undertook a study in 1997 and reported that, women who had undergone menopause had a lower risk of breast cancer than pre-menopausal women of the same age and childbearing pattern. Risk increased by almost 3% for each year above the menopause age (natural or induced by surgery), so that a women who had menopause at age 55 rather than 45, had approximately 30% higher risk. A strong trend of increasing breast cancer risk with increasing age at menopause was observed (La Vecchuia et al., 1992; Negri et al., 1988) For both pre and post-menopausal females, an increased risk of breast cancer was also observed for late menopause (Paffenbarger et al., 1980). Also women who have a natural menopause after age 55 were at twice the risk of developing breast cancer as those with natural menopause prior to age 44, (Handerson, 1992).

Several studies have reported that African women experience menarche at older ages (median, 15 years) compared with western women (Anyanwu 2000; Muguti 1993; Rosenberg et al., 2002). Lee et al., (1992) observed an increase risk (not significant) with early menarche. Early menarche has been associated with small, nonsignificant increases in risk among pre-menopausal women in three of four studies (Johnson-Thompson and Guthrie, 1999; Jatoi, 1999; Davis et al., 1997).

Among West-African women, age at menarche ranges between 12 and 20 years but did not appear to be a risk factor. Age at menarche appears to be important among African-American women for development of breast cancer prior to
menopause (Dupont and Page, 1985; Schatzkin et al., 1987 and Muyberry and Stoddard-Wright, 1992).

Some studies have indicated a protective effect for late menarche (Hsieh et al., 1990; Brinton et al., 1997; Kvale and Heuch, 1988). An increased risk of breast cancer was observed for early menarche (Paffenbarger et al., 1980). Women, whose menarche begins at age 12 or earlier, have twice the risk of developing breast cancer as those whose menarche begins after age 13 (Choi et al., 1978; Trichopoulos et al., 1972). In a study of women aged 25-54 years there was a declining risk of breast cancer with increasing age at menarche, those reaching menarche at age of more than 15 years being at an approximately 20% lower risk than those who began menstruating when aged less than 12 years (McCredie et al., 1998). McCredie et al (1998) and Garland et al., (1998) reported of declining risk of breast cancer with older age at menarche compared with younger age. Relative risk for those over 13 years, is 0.66 (CI 0.44 – 0.99).

In UK, early age at menarche has been consistently associated with an increased risk of breast cancer. The estimated decrease in risk per five year delay in menarche is 22%, (Garcia-Closas et al., 2006).

Other studies have documented an increased parity history in African communities, averaging a reported 5–9 live births per woman, Parkin et al., 1963-1977; Strassmann 1999).

Rosenberg et al., (2002), evaluated breast cancer risk in relation to parity among South African women and found patterns similar to those observed in western populations. In the study (Rosenberg et al., 2002), women of mixed ethnic
backgrounds with first childbirth at age 30 years or older had a twofold increase in breast cancer risk compared with women who had their first child at age 16 years or younger. In another study from South Africa, women who developed breast cancer had an older age at first birth and a higher level of education compared with control groups. In contrast, another study of nearly 2000 African breast cancer patients, reported by a Nigerian medical center, failed to identify any correlation between breast cancer risk and parity, risk and age at menarche, or risk and lactation (Ihekwaba, 1992).

Extended postpartum lactation has also been reported among African women, and this reproductive pattern is likely to decrease lifetime incidence of breast cancer by decreasing the cumulative number of ovulatory menstrual cycles, in one study, 96% of the patients' breastfed for an average of 16 months, Muguti (1993).

Beral et al (2002) have observed that breast feeding is protective against breast cancer, which is consistent with the findings of this study. Lane-Claypon (1926) noted that the breast that has never lactated is more liable to become cancerous.

Coogan et al., (1999) evaluated decreased breast cancer risk related to lactation among South African women and postulated that lactation causes differentiation of ductal epithelial cells, which is protective against carcinogens. Some studies however did not find any relationship between breastfeeding and breast cancer risk, (McCredie et al., 1998; Katsouyanni et al., 1996; Freudenheim et al., 1997; Michels et al., 1996). Mayberry and Stoddard-Wright (1992) analyzed standard familial and gynecologic risk factors among breast cancer cases (3,934 Caucasian Americans, 490 African Americans) and controls (3,901 Caucasian Americans, 35
485 African Americans) and found that age at first live birth, parity, and surgical menopause had similar associations with breast cancer risk, but family history and age at menarche behaved differently as risk factors. For African Americans, first-degree and second degree family history of breast cancer had comparable strengths as risk factors (odds ratios, 1.61 and 1.71, respectively), whereas the association in Caucasian Americans was notably stronger in relation to the pattern of family history (odds ratio, 2.16 for first-degree relatives, and 1.44 for second-degree relatives). Palmer et al., (2003) demonstrated a dual effect of pregnancy on breast cancer risk: multiparity increased breast cancer risk prior to the age of 45 years but was protective against breast cancer risk after age 45.

Calculations based on breast cancer incidence rates during the 1990s suggest that the cumulative incidence of breast cancer in developed countries would be reduced by more than half, from 6.3 to 2.7 per 100, if woman had the average number of births (6.5 instead of 2.5 births) and lifetime duration of breastfeeding (breastfeed each child, on average, for 24 months instead of 8 months) as is typical in developing countries, (CGHFBC, 2002). Thus, childbearing reduces the risk of breast cancer and the higher the number of full-term pregnancies, the greater the protection.

A study done by Ma et al., (2006) and Ewertz et al., (1990), showed that risk of breast cancer reduced by 7% with each full-term pregnancy, and the overall women who have had children have a 30% lower risk than nulliparous women.
A study published in December 2011 estimated that, in 2010, around 3% of breast cancers in women in the UK were linked to women breastfeeding every child for fewer than six months (Parkin, 2011a).

2.4.4 Body Mass Index (BMI) Factor

BMI is a measure of body weight adjusted for height (U.S. Department of Health and Human Services, 2005). Overweight and obesity, as measured by high body mass index, moderately increases the risk of post-menopausal breast cancer and is one of the few modifiable risk factors for breast cancer. BMI is calculated by dividing weight in Kilograms by height in metres squared. A BMI under 18.5 is classified as underweight, 18.5-24.9 as healthy weight, 25-29.9 as overweight and 30 or over as obese) Compared to lean (BMI 22.5-24.9) women, overweight post-menopausal women have a 10-20% increased risk of breast cancer, and obese post-menopausal women, a 30% increased risk. Women with a BMI under 22.5 have a 15% reduction in risk compared to women with a BMI of 22.5-24.9. In contrast, obese pre-menopausal women have a 20% reduction in breast cancer risk, (Reeves et al., 2007). A study published in December 2011, (Parkin 2011b) estimated that around 9% of breast cancers in women in the UK in 2010 were linked to excess bodyweight.

Numerous studies, largely of White women, have assessed the relation of overall obesity, to risk of breast cancer. The relation differed by menopausal status: among postmenopausal women, those who were overweight or obese had an increased breast cancer risk compared with those of normal weight, whereas among premenopausal women, overweight women have a reduced risk (Hunter et
Several case-control studies have reported separately on the relation of obesity to breast cancer risk in African American women by their menopausal status. Among postmenopausal women, high BMI was associated with an increased risk (Schatzkin et al., 1987; Zhu et al., 2005). Also Adams-Campbell et al., (1996) and Hall et al., (2000) reported of a reduced risk among post menopausal women whiles no association was reported by Austin et al., (1979).

A case–control study from urban Nigeria, involving 234 breast cancer patients, suggested that postmenopausal obesity was associated with breast cancer risk when measured as a function of waist–hip ratios, (Adebamowo and Adekunle, 2003).

For pre-menopausal women, there is a general decreasing risk with increasing BMI, with indications of a reverse relationship for post-menopausal women (Zhang et al., 1997). Franceschi et al., (1997) reported from a case-control study that, BMI at diagnosis was inversely related to pre-menopausal risk and directly related to post-menopausal risk. This means that pre-menopausal women with a higher BMI were at less risk of developing breast cancer than do post-menopausal women.

2.4.5 Lifestyle, Culture, and Socioeconomic Status.

One study from South Africa by Hoffman et al., (2000), has reported a doubling of breast cancer incidence rates in women living in urban areas compared with those residing in rural areas. Urban areas frequently are characterized by western behaviors and lifestyles. More affluent women typically reside in urban areas, and
these differences in incidence rates, therefore, and at least partly, reflect differences in age of menarche, parity, and age at first live birth, and age at menopause.

2.4.6 Alcohol Consumption Factor

Epidemiological studies have consistently shown a significant association between alcohol consumption and breast cancer and a recent IARC report concluded that this association is causal (Baan et al., 2007). A study published by Parkin (2011) estimated that more than 6% of breast cancers in women in the UK in 2010 were linked to alcohol consumption.

2.4.7 The Gender Factor

Simply being a woman is a primary risk factor for developing breast cancer. Although women have many more breast cells than men, the main reason they develop more breast cancer is because their breast cells are constantly exposed to the growth-promoting effects of the female hormones estrogen and progesterone. Men can develop breast cancer, but this disease is about 100 times more common among women than men, (Wu et al., 2002)

2.4.8 The Contraceptive Usage Factor

Many studies have been carried out on the effects of oral contraceptives (OCs) and hormone replacement therapy (HRT) individually on breast cancer risk and generally found that (OCs) increase the risk of developing breast cancer, (Burke 2000; Tessaro et al., 2000; Kahlenborn et al., 2008; Casey et al., 2008; Lower,
2008; Ansink and Burger, 2007). In 1998, Brinton et al. conducted a case-control study on breast cancer risk in women under 55 years of age (1,031 cases and 919 controls). Women in this study who had taken OCs for over ten years had an increased risk of breast cancer, with a relative risk of 3.2 compared to non-users.

Some studies have found no significant increase of breast cancer with oral contraceptive use. These include (Chie et al., 1998; Brinton et al., 1997; Rosenberg et al., 1996). Norman et al. in 2003, found a negative interaction, with the risk of breast cancer higher in never-users of OCs than ever-users.

2.5 BREAST CANCER PROGNOSTIC FACTORS

A prognostic factor may be defined as a measurable variable that correlates with the natural history of the disease. In contrast, a predictive factor is one that is associated with response to a given therapy. Breast cancer prognostic factors include Axillary lymph node status, Tumor size, Lymphatic/vascular invasion, Patient age, Histologic grade, Histologic subtypes (eg, tubular, colloid [mucinous], papillary), Body Mass Index (BMI), Tumour Stage, Estrogen receptor/progesterone receptor status, Her2/neu gene amplification and/or over expression. In this study, only Axillary lymph node status, Tumor size, Patient age, Histologic grade, Histologic subtypes (eg, tubular, colloid [mucinous], papillary), Estrogen receptor/progesterone receptor status, BMI and Tumour Size are being considered. Since data is available for only these prognostic factors.
2.5.1 Axillary Node

The most significant prognostic factor in breast cancer is the presence or absence of axillary lymph node involvement. Lymph nodes are filters along the lymph fluid channels. They try to catch and trap cancer cells before they reach other parts of the body. If lymph nodes have some cancer cells in them, they are called positive, which is associated with an increased risk of the cancer spreading. "There is a direct relationship between the number of involved axillary nodes and the risk for distant recurrence" said Saez et al., (1989). The status of the axilla is usually assessed at the time of surgery using sentinel lymph node biopsy or axillary dissection. Axillary node status is the most consistent prognostic factor used in adjuvant therapy decision making. Adjuvant therapy is administered to patients with lymph nodes that are positive.

The vast majority of patients found to have lymph node metastases are candidates for adjuvant systemic therapy. However, determining which node-negative patients should receive adjuvant therapy is challenging, particularly because the majority are cured by surgical excision alone. The benefit to those patients who are destined to relapse is minimal, and the costs and toxicities of the treatments are significant. Thus, node-negative patients require further stratification using additional prognostic and predictive factors.

2.5.2 Tumour Size

Tumour size has long been recognized as an independent prognostic factor and as a predictor of axillary node status, with larger tumors being associated with a
worse prognosis and an increased likelihood of nodal metastasis. Thus, tumour size correlates with the presence and number of involved axillary lymph nodes. Distant recurrence rates increases with larger tumour size. For node negative patients, tumour size is the most powerful prognostic factor and is routinely used to make adjuvant treatment decisions. Tumor size is evaluated by radiologic examination.

2.5.3 Histology / Grade

Histology gives the information where the cancer starts from. Ductal denotes that the cancer begins in the milk duct, and Lobular starts inside the milk-making glands (called lobules). Certain histologic subtypes of breast cancer are generally associated with a favorable prognosis, including tubular, colloid (mucinous) and papillary carcinoma. The proliferation rate of the tumor as determined by the mitotic count is also of prognostic importance, but this information is included as a component of the tumor grade.

Tumour grade is based on how closely the biopsy sample resembles normal breast tissue. The grade helps predict a woman’s prognosis. In general, a lower grade number indicates a slower-growing cancer that is less likely to spread, while a higher number indicates a faster-growing cancer that is more likely to spread.

The most widely accepted grading system is the Scarff-Bloom-Richard (SBR) classification, (Bloom and Richardson, 1957). The classification is based on the arrangement of the cells in relation to each other: whether they form tubules; how
closely they resemble normal breast cells (nuclear grade); how many of the cancer cells are in the process of dividing (mitotic count). This system of grading is used for invasive cancers but not for in situ cancers.

- Grade 1 (well differentiated) cancers have relatively normal-looking cells that do not appear to be growing rapidly and are arranged in small tubules.
- Grade 2 (moderately differentiated) cancers have features between grades 1 and 3.
- Grade 3 (poorly differentiated) cancers, the highest grade, lack normal features and tend to grow and spread more aggressively.

Tumour grade is primarily used to make decisions for lymph node-negative patients with borderline tumour sizes. It is also important in patients who have small tumours and no lymph node involvement. Patients with small, well-differentiated tumours may require no further treatment after the tumour is removed, while patients with moderately or poorly differentiated tumours usually receive additional hormonal chemotherapy.

2.5.4 Hormone Receptors

All cases of invasive breast carcinoma are evaluated for Estrogen, Progesterone and HER2 receptor status using immune histochemistry, which has both predictive and prognostic value. Receptors are proteins on the outside surfaces of cells that can attach to certain substances, such as hormones, that circulate in the blood. Normal breast cells and some breast cancer cells have receptors that attach
to estrogen and progesterone. These hormones often fuel the growth of breast cancer cells.

An important step in evaluating a breast cancer is to test a portion of the cancer removed during biopsy or surgery for the presence of estrogen and progesterone receptors. Cancer cells may contain neither, one, or both of these receptors. Breast cancers that contain estrogen receptors are often referred to as "ER-positive" cancers, while those containing progesterone receptors are called "PR-positive" cancers. Women with hormone receptor-positive cancers tend to have a better prognosis and are much more likely to respond to hormone therapy than women with cancers without these receptors.

According to American Cancer Society, about 2 out of 3 breast cancers contain at least one of these receptors. This percentage is higher in older women than in younger ones.

As a prognostic factor, ER and/or PR positivity is associated with reduced mortality compared to women with ER and/or PR negative disease. However, the percentage of immune histochemically positive tumor cells used to classify a tumor as ER or PR positive varies among institutions. Studies support that carcinomas with greater than 1% ER positive cells have better survival rates compared to carcinomas that are completely devoid of ER.
2.5.5 Cancer Staging

A staging system is a standardized way for the cancer care team to summarize information about how far a cancer has spread. The most common system used to describe the stages of breast cancer is the American Joint Committee on Cancer (AJCC) TNM system. The stage of a breast cancer can be based either on the results of physical examination, biopsy, and imaging tests (called the clinical stage), or on the results of these tests plus the results of surgery (called the pathologic stage). The staging described here is the pathologic stage, which includes the findings after surgery, when the pathologist has looked at the breast mass and nearby lymph nodes. Pathologic staging is likely to be more accurate than clinical staging, as it allows the doctor to get a firsthand impression of the extent of the cancer. The TNM staging system classifies cancers based on their T, N, and M stages: T stands for tumor (its size and how far it has spread within the breast and to nearby organs). N stands for spread to lymph nodes (bean-shaped collections of immune system cells that help fight infections and cancers). M is for metastasis (spread to distant organs).

2.5.5.1 Breast Cancer Stage Grouping

Once the T, N, and M categories have been determined, this information is combined in a process called stage grouping. Cancers with similar stages tend to have a similar outlook and thus are often treated in a similar way. Stage is expressed in Roman numerals from stage I (the least advanced stage) to stage IV (the most advanced stage). Non-invasive cancer is listed as stage 0.
2.5.5.2 Disease Presentation

Breast cancers in African countries are typically characterized by a relatively advanced stage distribution. A study by Hassan et al., (1995), reported a mean primary tumor diameter of 10 cm in 129 Nigerian women, and matted axillary lymph nodes were reported in more than half of these cases. Other retrospective studies have reported that 70–90% of African women present with Stage III or IV disease, (Amir et al., 1997; Hassan 1995; Amir et al., 1994; Muguti, 1993; Hassan et al., 1992). In a recent report from Accra Ghana the average duration of symptoms before presentation was 10 months. As a result almost two-thirds (57.6%) of the patients who were diagnosed with breast cancer presented with Stage III-IV disease (Clegg-Lamptey and Hodasi, 2007) and in a report on the reasons for delay in reporting nearly a third (28.8%) of the respondents indicated that ignorance of the early signs of breast cancer accounted for the late presentation. Further, it can be reasonably conjectured that the clinical stage of many patients is underestimated because of the lack of proper diagnostic facilities.

The advanced stage distribution is at least partially explained by delayed presentation for medical evaluation, which has been reported to range from 2 weeks to 11 years following development of a self detected breast abnormality as commented on by Anyanwu (2000).

A study of 2033 black and white patients from South Africa demonstrated that these delays among black Africans resulted in significantly larger tumors and
more advanced nodal pathology compared with white South Africans (Dansey et al., 1988).

The possibility of inherently more aggressive tumor biology among African women also exists. Findings from Hassan et al. support this theory. The investigators found that 17 of 21 patients (81%) with a brief symptom duration (3 months or less) had Stage III or IV breast cancer. In a study in East Africa, more than 70% of the patients presented at stage III or IV, Hunter (1993).

The majority (approximately 85%) of tumors among African women are invasive ductal lesions, (Amir 1997 and Amir 1994). This is similar to the histopathologic predominance observed in western populations. Bjerregaard et al., commented on the difficulty of evaluating prevalence of the more unusual types of breast cancer because of the frequency of detecting advanced-stage, poorly differentiated tumors that have lost any distinguishing features.

2.5.6 Tumor Biology and Genetics

Unfortunately, there is a paucity of studies on sub-Saharan Africa detailing biology or genetics of breast cancer. One formidable hindrance involves obtaining and processing tissue samples (usually because of financial considerations). There is also difficulty obtaining family history due to decreased awareness of breast cancer in sub-Saharan Africa and due to the desire for secrecy that is sometimes found within families following a diagnosis of cancer, Anyanwu (2000).

Ikpatt et al., (2002) have published studies of several hundred patients defining the biology of breast tumors in African women. In these studies, comparisons of tumors from Finnish and Nigerian breast cancer patients showed that Nigerian
breast tumors had more extensive necrosis, more nuclear atypia, and a higher proliferative activity.

2.6 MODEL CLASSIFICATION AND EVALUATION

Some models used in analyzing cancers are cited and classified below.

Asma et al., (2008) used probit and logit regression to investigate the relation between demographic factors and type of gastrointestinal cancers.

Cox Proportional Hazard model has been used by many researchers, including Kakarala et al., (2010), who used it to estimate relative risks of breast cancer mortality and Palmer et al., (2007) to compute incidence rate ratio to assess the relation of body mass index and weight gain to breast cancer risk. Vigano et al., (2000) observed 227 cancer patients aged 18 years or older from July 1, 1996 through December 31, 1998 in Alberta, Edmonton, Canada Using Univariate Kaplan–Maier and Multivariate Cox regression analysis to establish the survival of patients with breast cancer.

Logistic regression and partial least square regression were employed to evaluate cervical cancer by Hailun Wang (2008); Kwan et al., (2009) described breast tumour subtypes by common breast cancer risk factors; Ingrid et al., (2003) estimated odds ratios for breast cancer risk factors.

non linear regression method to assess cancer risk factors. Pike et al., (1983) used Pike model based on the incidence rates to compute the observed relative risks for the population of interest.

Gail model was used by Amir et al., (2003), to evaluate breast cancer risk in the family history; Costantino et al., et al., (1999), to validate studies for models projecting the risk of invasive and total breast cancer incidence; etc.

Claus model was also employed by Claus et al., (1993), to calculate breast cancer risk for women with a first degree family history of ovarian cancer; McGuigan et al., (1996) to assess the agreement between breast cancer risk estimation methods; and McTiernan et al., (2001) also used the model to compare two breast cancer risk estimates in women with a family history of breast cancer.

Plevritis et al., (2006) used Monte Carlo Simulation to predict breast cancer mortality trends. A First Hitting Time (FHT) regression model was applied in a study (Hazelton et al., 2001) and also by (Lee et al., 2004) of lung cancer risk posed to Chinese tin miners by arsenic, radon and tobacco exposure.

2.7 CONCLUSION

As already highlighted in the literature review above, there is no single pattern but rather a multitude of factors that can influence breast cancer outcomes. A number of authors have used varying methodologies to predict breast cancer survival and failure. Some of the results from studies done affirmed what had been theoretically hypothesised. However, the outcomes of some breast cancer survival analyses were not consistent with both the theoretical literature and results of
empirical studies done in other parts of the world. From the literature, the main analytical methods used in cancer survival analyses comprised both parametric (logit, probit and ordinary least squares) and non-parametric (Cox proportional hazards) techniques. The use of a particular survival model was based on each individual or group of researcher's preference. Those who opted for the semi-parametric, cox proportional hazards model justified their choice by indicating that the use of other methods (parametric) leads to wastage of information and biases due to leaving out censored cases. In the case for patients’ failure, the standard regression analysis, such as the ordinary least squares are said to ignore patients or units that are outside the observation window and count them as if they have died and thus producing misleading results. Non parametric models on the other hand are reported to allow the researcher to control for both the occurrence of an event and the timing of the event, hence taking into account the evolution of the failure risk and its determinants over time. Based on its reported flexibility and accommodative nature, the Cox proportional hazard model will be estimated in this study (Kakarala et al., 2010; Vigano et al., 2000; Simon and Severson 1997). Logistic regression analysis will also be used to examine the extent to which selected demographic, hormonal and reproductive factors influence the cause of breast cancer.
CHAPTER THREE

METHODOLOGY

3.0 INTRODUCTION

The methodology which underpinned the study is explained in this chapter. The chapter provides, details of all the processes employed from study design, study settings, sampling procedures, and measurements through data collection, analysis and to the final summary of findings.

3.1 STUDY DESIGN

Case control study was used to identify factors that may contribute to a medical condition by comparing subjects who have that condition (cases) with subjects who do not have the condition but are otherwise similar (control). This is a retrospective design in which we "look into the past".

3.2 STUDY SETTING

The study was conducted in the Korle - Bu Teaching Hospital, Ghana’s National Referral and Teaching Hospital, see Figure 3.1 for a front view of the hospital
Korle Bu Teaching Hospital (KBTH) was established on October 9, 1923 with initial bed capacity of 200. At that time, Korle Bu was described as the finest hospital in Africa on account of its impressive array of fine buildings and a cadre of competent staff, who provided excellent medical care to the population of Ghana, in general, and the city of Accra, in particular. It is currently the third largest hospital in Africa and the leading national referral centre in Ghana. Korle Bu, which means 'the valley of the Korle lagoon', was established as a General Hospital to address the health needs of the indigenous people under Sir Gordon Guggisberg's administration, the then Governor of the Gold Coast. Korle Bu gained teaching hospital status in 1962, when the University of Ghana Medical School (UGMS) was established for the training of medical doctors. The UGMS and five other constituent schools are now subsumed under the College of Health Sciences to train an array of health professionals. All the

Figure 3.1: Front view of Korle-Bu Teaching Hospital

Korle Bu Teaching Hospital (KBTH) was established on October 9, 1923 with initial bed capacity of 200. At that time, Korle Bu was described as the finest hospital in Africa on account of its impressive array of fine buildings and a cadre of competent staff, who provided excellent medical care to the population of Ghana, in general, and the city of Accra, in particular. It is currently the third largest hospital in Africa and the leading national referral centre in Ghana. Korle Bu, which means 'the valley of the Korle lagoon', was established as a General Hospital to address the health needs of the indigenous people under Sir Gordon Guggisberg's administration, the then Governor of the Gold Coast. Korle Bu gained teaching hospital status in 1962, when the University of Ghana Medical School (UGMS) was established for the training of medical doctors. The UGMS and five other constituent schools are now subsumed under the College of Health Sciences to train an array of health professionals. All the

52
institutions of the College however, undertake their clinical training and research in the Hospital.

At the moment, the Hospital has 2,000 beds and 17 clinical and diagnostic Departments/Units. It has an average daily attendance of 1,500 patients and about 250 patient admissions. Clinical and diagnostic departments of the hospital include Medicine, Child Health, Obstetrics and Gynaecology, Pathology, Laboratories, Radiology, Anaesthesia, Surgery, Polyclinic, Accident Centre and the Surgical/Medical Emergency as well as Pharmacy. Other Departments includes, Pharmacy, Finance, Engineering, General Administration. The Hospital also provides sophisticated and scientific investigative procedures and specialization in various fields such as Neuro-surgery, Dentistry, Eye, ENT, Renal, Orthopaedics, Oncology, Dermatology, Cardiothoracic, Radiotherapy, Radio diagnosis, Paediatric Surgery and Reconstructive Plastic Surgery and Burns. The Reconstructive Plastic Surgery and Burn Centre, the National Cardiothoracic Centre and the National Centre for Radiotherapy and Nuclear Medicine in particular also draw a sizeable number of their clientele from neighbouring countries such as Nigeria, Burkina Faso and Togo. Korle Bu Teaching Hospital continues to blaze the trail when it comes to the introduction of specialized services. It recently carried out the first ever kidney transplant in Ghana. It is one of the few hospitals in Africa where DNA investigations are carried out. Other specialized services the Hospital provides include brachytherapy intervention for the treatment of prostate cancer and keyhole surgeries.
3.3 TARGET AND STUDY POPULATIONS

The population targeted for the study was all Ghanaian women.

The study population comprised all Ghanaian women who reported at the Breast Clinic of the Korle_Bu Teaching Hospital between January 2002 and December 2008, since KBTH is the leading national referral centre in Ghana.

3.4 STUDY VARIABLES

Nine key variables on breast cancer risk factors, which provided the required measures on data, were identified as: age at first visit; family history of breast cancer; age at menarche; age at menopause; age at first child; contraceptive usage; alcohol intake; parity and breastfeeding. Seven variables for prognostic factors were taken as: stage at diagnosis; tumour grade; tumour size, BMI, axillary node status, age at diagnosis and ER/PR status.

The time (measured in months) to the event (ie, death) was used for the survival analysis whiles the status of the respondent (whether or not had breast cancer) was used for the logistic regression analysis. These constituted the dependent variables for the study.

3.5 DATA COLLECTION, MANAGEMENT AND ANALYSIS

3.5.1 Data Collection

Secondary data was used for the study. Being a case control study of breast cancer in Ghana, both the cases and the controls were sampled from the Records Section of the Cancer unit and the Breast Clinic respectively of the KBTH. The records unit is a departmental registry where registered cancer cases for treatment are kept.
and the breast clinic is a walk-in one for screening. All Ghanaian women who visited the Korle-bu Teaching hospital, National Centre for Radiotherapy and Nuclear Medicine and the walk in clinic for breast screening between 1st January, 2002 and 31st December, 2008, with histologically confirmed primary breast cancer were the eligible cases. Sources for the data included, abstract of hospital records, (see Appendix 1 for some details), and in person interviews conducted by the female registered nurses. The nurse-interviewer elicited information on demographic and potential breast cancer risk factors, including first family history of breast cancer or other cancers, menstrual and reproductive history, socio-demographic and lifestyle characteristics (see Appendix 2 for sample questionnaire used). Cancer staging, grade and ER/PR status were obtained from the medical records for the cases sampled only. After the exclusion of 391 women, (on the grounds of: incomplete information on a number risk as well as prognostic factors of interest associated with them, being below twenty (20) years of age and not having been diagnosed of IDC, ILC, DCIS, LCIS and inflammatory breast cancer), the final data consisted of 1022 cases and 1375 controls.

3.5.2 Data Management

Data extracted from the records were cleaned and entered in the computer using Microsoft Office Excel application. The data was exported to the Minitab and SAS programmes subsequently for statistical analysis.
3.5.3 Survival Analysis of the Data

Survival analysis pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study period, or the study of the time between entry into observation and a subsequent event, (Allison, 1995). The analysis of survival data involves a collection of statistical procedures for which the outcome variable of interest is time until an event occurs.

Time can be referred to the number of years, months, weeks, or days from the beginning of follow-up of a subject until an event occurs. Allison (1995) defined an event as a qualitative change that can be situated in time. Qualitative change here refers to a transition from one discrete state to another at an instantaneous moment in time. Marriage for example is a transition from one state of being unmarried to the state of being married.

Ideally, the transitions occur virtually instantaneously and the exact times at which they occur are known. Some transitions may take a little time; however, the exact time of onset may be unknown or ambiguous.

Quantitative variables can be treated as events if the change is large and sudden compared to the usual variation over time. A stock market crash could be defined as any single day loss of more than 20% in the market index. Some researchers also define an event as occurring when a quantitative variable crosses a threshold. For example, a person is said to have gone into poverty when her income goes beyond some designated level.
In this work, the event of interest was death and the analysis consisted of following the subject until death. Survival Analysis is used mostly in the Medical and Biological Sciences, however the techniques are also widely used in the Social Sciences, Econometrics, and Engineering.

Survival analysis techniques allow for a study to start without all experimental units enrolled and to end before all experimental units have experienced an event, (Allison, 1995). This is extremely important because even in the most well developed studies, there will be subjects who choose to quit participating, who move too far away to follow, who will die from some unrelated event, or who will simply not have an event before the end of the observation period.

A common feature of survival data is censoring, it means that the exact failure times of a number of subjects are not known. This refers to incomplete information regarding units under observation. Perrigot et al., (2004) remarked that the term “censored” means that the exact length of the duration is ignored because the initial event date or the final event date is unknown. There are two main types of censoring in survival analysis studies, right and left censoring. Right censoring implies that the final date of the period separating the two events studied cannot be determined. According to Perrigot et al., (2004), when data are right censored, there is no ending time, but just a starting time because the event would have not ended. Left censoring presents a situation whereby there is data for which there is an ending point but no information regarding when the item studied was initially exposed to the risk, (Melnyk et al., 1995).

There are several reasons for censoring. A few are, (Allison, 1995):
Some patients may have exited the study early: they are said to be lost to follow up. (Reasons are examples: due to emigration, fatal accidents, financial difficulties, etc.).

The study ends when a fixed time is reached (right censoring of type I)

The study ends when a fixed number of failures occur (right censoring of type II)

In this study, right censoring of type I is being considered since the study ended with a fixed time (January 2011) and the survival time was in months. According to Melnyk et al., (1995), survival analysis possesses two aims: on the one hand, the aim is to estimate the time period during which an event can happen, while on the other hand the interest is to describe the time distribution of the event and estimate quantitatively the impact of independent variables (covariates) on the event. In utilising the techniques of Survival Analysis, Perrigot et al., (2004) identified the main concepts of the methodology as: Event, Measurement window and Censorship.

In survival analysis, a researcher seeks to model both the duration of time spent in the initial condition and the transition to a subsequent state, that is, the event. The period of time during which a researcher makes their observation defines the measurement window. The other important concept is censorship.

Censoring techniques enable one to analyze incomplete data due to delayed entry or withdrawal of a subject from a study. This is important in allowing each experimental unit to contribute all of the information possible to the model for the
amount of time the researcher is able to observe the unit. The problem of analyzing censored data is usually referred to as survival analysis.

In survival analysis terminology, patients who are observed until they reach the end point (e.g. death) are called uncensored cases while those who survive further than the end of the study or who are lost to follow-up at some point are called censored cases. Analyses of this type involve the amount of time that a subject is at risk while under observation. In survival analysis, the time variable is usually referred to as survival time, because it gives the time that an individual has "survived" over some follow-up period.

3.5.4 Analyzing Survival Data

Models for the analysis of data have three main characteristics:

1. The dependent variable or response is the waiting time until the occurrence of a well-defined event.

2. Observations are censored, in the sense that for some units the event of interest has not occurred at the time the data are analyzed.

3. There are predictors or explanatory variables whose effect on the waiting time one may wish to assess.

3.5.4.1 Methods of Analysis

Data that has censored observations are said to be incomplete and the analysis of such data requires special techniques. There are basically three methods that could be used in the analysis of survival data namely: fully parametric, non-parametric and semi-parametric.
Fully parametric methods assume the knowledge of the distributions of the survival time e.g. exponential, weibull, loglogistic, log-normal, gamma and gompertz. As the uses of survival analysis have grown, parametric model gave way to non-parametric and semi-parametric approaches due to their special individual appeal. Non-parametric models make no assumptions of the distribution of the survival times, e.g. the Kaplan-Meier estimators. The semi-parametric models assume a parametric form for the effects of the explanatory variables but make no assumptions on the distributions of the survival times. This study will highlight mainly only the non-parametric and semi-parametric approaches in analyzing the survival data.

3.6 SURVIVAL MODEL

The fundamental concepts of survival analysis involve the cumulative distributive function, probability density function, survival function and hazard function.

3.6.1 The Cumulative Distribution Function

Suppose $T$ is a non-negative random variable representing the waiting time until the occurrence of an event of interest. For simplicity we will adopt the terminology 'death' as the event of interest and the waiting time as 'survival' time. It is assumed that $T$ is a continuous random variable with a cumulative distribution function (cdf). The cumulative distribution function is very useful in describing continuous probability distributions of a random variable, such as time, in survival analysis. The cdf of a random variable $T$, denoted $F(t)$, ($t$ being time) is defined by:
This is interpreted as a function that will give the probability that the variable \( T \) will be less than or equal to any value \( t \) that we choose. Several properties of \( F(t) \) can be listed as a consequence of the knowledge of probabilities. Because \( F(t) \) is such that: \( 0 < F(t) < 1 \), it is a nondecreasing function of \( t \), and as \( t \) approaches infinity, \( F(t) \) approaches unity.

### 3.6.2 The Probability Density Function

The probability density function (pdf) is also very useful in describing the continuous probability distribution of a random variable. The pdf of a random variable \( t \), denoted \( f(t) \), is defined by

\[
 f(t) = \frac{dF(t)}{dt}
\]

That is, the pdf is the derivative or slope of the cdf. Every continuous random variable has its own density function, the probability \( P(a < T < b) \) is the area under the curve between the times \( a \) and \( b \).

### 3.6.3 The Survival Function

The survival function \( s(t) \) gives the probability of surviving or being event-free beyond time \( t \). Because \( S(t) \) is a probability, it is positive and ranges from 0 to 1. It is defined as \( S(0) = 1 \) and as \( t \) approaches infinity, \( S(t) \) approaches zero.

In survival analysis, it is convenient to work with the complement of the cdf. Suppose \( T > 0 \) have a pdf \( f(t) \) and cdf \( F(t) \). Then the survival function takes on the following form:

\[
 S(t) = P\{T > t\} = 1 - F(t)
\]
which gives the probability of being 'alive' at duration t, or the probability that the event of interest has not occurred by duration t.

3.6.4 The Hazard Function

The hazard function describes the concept of the risk of an outcome (e.g., death, failure) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual dies somewhere between t and \( t + \delta \), divided by the probability that the individual survived beyond time t. The hazard function \( \lambda(t) \) specifies the instantaneous rate of failure at \( T = t \) conditional upon survival to time \( t \) and is defined by the limit as \( \delta \to 0 \) of the following ratio:

\[
\lim_{\delta \to 0} \frac{P(t \leq T < t + \delta | T \geq t)}{P(T \geq t)} = \lim_{\delta \to 0} \frac{P(t \leq T < t + \delta)}{P(T \geq t)} = \frac{S(t) - S(t + \delta)}{\delta} \times \frac{1}{S(t)}
\]

\( \lambda(t) = \frac{f(t)}{S(t)} \)

The distribution of \( T \) is specified by its hazard function because the survival function is determined by the hazard function:

\[
\frac{d}{dt} \ln(S(t)) = -\frac{f(t)}{S(t)} = -\lambda(t)
\]

The hazard function seems to be more intuitive to use than the pdf because it attempts to quantify the instantaneous risk that an event will take place at time t given that the subject survived to time t.

If the distribution of \( T \), the failure time, is taken to be an exponential then the hazard function is a constant: \( \lambda(t) = \lambda \) (with \( \lambda > 0 \)). Suppose
\[ S(t) = \exp(-\lambda t) \]

then:

\[ F(t) = 1 - \exp(-\lambda t) \quad \text{and so} \quad f(t) = \lambda \exp(-\lambda t) \]

\[ \dot{\lambda}(t) = \lambda \]

So for the exponential distribution the instantaneous failure rate is independent of \( t \) so that the conditional chance of failure does not depend on how long the individual has been on trial.

The hazard function has two components. The first one is the concept of set at risk, which is the set of units, individuals, organizations etc. in a sample exposed to risk, in relation to an event occurring at a certain point in time (Perrigot et al., 2004). The second concept is the hazard rate, also referred to as conditional failure rate. The hazard rate gives the rate at which units under observation fail (or durations’ end) at time \( t \), given that the unit had survived until time \( t \) (Box-Steffensmeier et al., 2004).

An important generalization of the exponential distribution allows for a power dependence of the hazard function on time. This yields the two parameter Weibull distribution with hazard function

\[ \lambda(t) = \lambda p(\lambda t)^{p-1} = p\lambda^p \times t^{p-1} \]

This hazard function is monotone decreasing for \( p < 1 \), monotone increasing for \( p > 1 \), and reduces to a constant for \( p = 1 \).
From the Weibull distribution we obtain the survival function:

\[ S(t) = \exp\left( - \int_0^t p\lambda u^{p-1} du \right) = \exp\left( - (\lambda t)^p \right) \]  \[ \text{3.10} \]

\[ \ln(-\ln(S(t))) = \lambda (\ln(t) + \ln(\lambda)) \]  \[ \text{3.11} \]

In general the distribution of a failure or survival time is skewed. Skew distributions may be modeled by means of a lognormal distribution or a gamma distribution. If \( T \) has a lognormal distribution then this means that \( Y = \ln(T) \) has a normal distribution, described by an expectation \( \mu \) and a variance \( \sigma^2 \). The gamma distribution may be regarded as another generalization of the exponential distribution, its density function is

\[ f(t) = \frac{\lambda (\lambda t)^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \]  \[ \text{3.12} \]

where \( \Gamma(\alpha) \) is the well-known gamma function, given by:

\[ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx \quad (0 < \alpha < \infty). \]  \[ \text{3.13} \]

For \( \alpha = 1 \) the density reduces to the density of the exponential distribution, since \( \Gamma(1) = 1 \).

**3.7 KAPLAN-MEIER PRODUCT LIMIT ESTIMATE**

A popular survival analysis technique is the use of the product-limit estimate proposed by Kaplan and Meier in 1958. It provides quick, simple estimates of the
cumulative survival distribution. This estimator contains information from all of
the observations available, both uncensored and censored, by considering survival
to any point in time as a series of steps defined by the observed survival and
censored times.

The Kaplan-Meier product-limit estimate has an advantage over competing
estimates in that it does not depend upon a choice of intervals. For a given set of
survival time data, it is assumed that each observation is either a failure (death) or
a censor (living). It is further assumed that the data are ordered according to time
with failed observations occurring slightly before censored observations, if tied
observations occur.

The Kaplan-Meier estimate provides a simple way for computing the survival
curve. It involves computing the number of people who died at a certain time
point, divided by the number of people who were still in the study at that time and
multiplying their probabilities by any earlier computed probabilities.

Suppose that the survival times, including censored observations, of a
homogeneous group of \( n \) patients are represented by \( t_1, t_2, \ldots, t_n \). The assumption
is that the survival times (patients) are already ordered such that \( t_1 \leq t_2 \leq \ldots \leq t_n \).

For a given time \( t \), the largest value \( t_i \) is found such that \( t_i \leq t \) and the
probability \( S(t) \) is then estimated by:

\[
\hat{S}(t) = \frac{r_1 - d_1}{r_1} \times \frac{r_2 - d_2}{r_2} \times \ldots \times \frac{r_i - d_i}{r_i}
\]

3.12
The survival curve describes the relationship between the probability of survival where \( r_t \) is the number of subjects alive just before time \( t_k \) (the \( k \)th ordered survival time) and \( d_k \) denotes the number who died at time \( t_k \).

\[
S = \prod_{k} \left( \frac{r_{k-1}}{r_k} \right) 
\]

where \( r_k \) is the number of subjects alive just before time \( t_k \) (the \( k \)th ordered survival time) and \( d_k \) denotes the number who died at time \( t_k \).

S is based upon the probability that an individual survives at the end of a time interval, on the condition that the individual was present at the start of the time interval. S is the product of these conditional probabilities.

The survival curve describes the relationship between the probability of survival and time. The Kaplan-Meier survival curve is often illustrated graphically, looking like a poorly designed staircase, with vertical steps downward at the time of death of each individual subject.

![Figure 3.2 Survival Curve in Theory](https://www.udsspace.uds.edu.gh)
Theoretically, as \( t \) goes from 0 towards infinity, the survivor function (see Figure 3.2) goes from \( S(t) = 1 \) towards zero asymptotically. In practice (using data) we usually obtain estimated survivor curves that are step functions, as illustrated in Figure 3.3, rather than as smooth curves.

The general data layout for a survival analysis is given in Table 3.1 below. This layout is the basis upon which Kaplan–Meier survival curves are derived. The first column in the table gives ordered survival times from smallest to largest. The second column gives frequency counts of failures at each distinct failure time. The third column gives frequency counts, denoted by \( q_j \), of those persons censored in the time interval starting with failure time \( t(j) \) up to but not including the next failure time, denoted by \( t(j+1) \). The last column gives the risk set, which denotes the collection of individuals who have survived at least to time \( t(j) \).
Table 3.1: General Layout for Survival Analysis

<table>
<thead>
<tr>
<th>Ordered failure times</th>
<th>No. Of failures</th>
<th>No. censored in ( t_{d+j} )</th>
<th>Risk set</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t(j) )</td>
<td>( m_j )</td>
<td>( q_j )</td>
<td>( R_{t(j)} )</td>
</tr>
<tr>
<td>( t(0) )</td>
<td>( m_0 )</td>
<td>( q_0 )</td>
<td>( R_{t(0)} )</td>
</tr>
<tr>
<td>( t(1) )</td>
<td>( m_1 )</td>
<td>( q_1 )</td>
<td>( R_{t(1)} )</td>
</tr>
<tr>
<td>( t(2) )</td>
<td>( m_2 )</td>
<td>( q_2 )</td>
<td>( R_{t(2)} )</td>
</tr>
<tr>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
</tr>
<tr>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
</tr>
<tr>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
<td>( \cdot )</td>
</tr>
<tr>
<td>( t(k) )</td>
<td>( m_k )</td>
<td>( q_k )</td>
<td>( R_{t(k)} )</td>
</tr>
</tbody>
</table>

To estimate the survival probability at a given time, make use of the risk set at that time to include the information on a censored person up to the time of censorship. The actual computation of such a survival probability can be carried out using the Kaplan–Meier (KM) method.

### 3.7.1 The Log-Rank Test

The logrank test (Peto et al, 1977) is the most widely used method of comparing two or more survival curves. The method calculates at each event time, for each group, the number of events one would expect, since the previous event, if there were no difference between the groups. These values are then summed over all event to give the total expected number of events in each group, say \( E_i \) for group
i. The logrank test compares observed number of events, say $O_i$, for treatment group $i$, to the expected number by calculating the test statistic

$$
X^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}
$$

3.12

This value is compared to a chi-square distribution with $g - 1$ degrees of freedom, where $g$ is the number of groups. In this manner, a p-value may be computed to calculate the statistical significance of the differences between the complete survival curves. If the groups are naturally ordered, a more appropriate test is to consider the possibility that there is a trend in survival across them, for example, age groups or stages of cancer. Calculating $O_i$ and $E_i$ for each group on the basis that survival may increase or decrease across the groups results in a more powerful test. For the new $O_i$ and $E_i$ the test statistic for trend is compared with the chi-square distribution with one degree of freedom (Collett, 1994). When only two groups are compared, the logrank test is testing the null hypothesis that the ratio of the hazard rates in the two groups is equal to unity as against the alternative that the ratio of the hazard rates in the two groups is not equal to unity. There is no difference between the two survival curves for the two groups. The hazard ratio (HR) is a measure of the relative survival experience in the two groups and may be estimated by

$$
HR = \frac{O_1/E_1}{O_2/E_2}
$$

3.13

where $O_i/E_i$, ($i=1,2$) is the estimated relative (excess) hazard in group $i$. The HR has a similar interpretation in terms of the strength of effect as a risk ratio. A HR
of 1 indicates no difference in survival. In practice, it is better to estimate HRs using a regression modeling technique, such as Cox regression.

The Kaplan–Meier method with a log-rank test is useful for comparing survival curves in two or more group. Cox proportional-hazard regression, however, allows analyzing the effect of survival risk factors on survival. The hazard of Cox proportional regression is modeled as:

\[
H(t) = H_0(t) \times \exp(b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_kx_k) \tag{3.14}
\]

where \(x_1 \cdots x_k\) are a collection of predictor variables and \(H_0(t)\) is the baseline hazard at time \(t\), representing the hazard for a person with the value 0 for all the predictor variables. By dividing both sides of the above equation by \(H_0(t)\) and taking logarithms, we obtain:

\[
\ln \left( \frac{H(t)}{H_0(t)} \right) = b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_kx_k \tag{3.15}
\]

\(H(t)/H_0(t)\) is the hazard ratio. The coefficients \(b_1 \cdots b_k\) are estimated by the Cox regression, and can be interpreted in a similar manner as that of multiple logistic regression.

### 3.7.2 The Cox Proportional Hazard (PH) Model

The most widely used model in the analysis of survival data is the Cox proportional hazards model, (Allison 1985). This semi-parametric model is widely used in the analysis of survival data to explain the effects of explanatory variables on survival times. An important reason for the popularity of the Cox model is that even though the baseline hazard is not specified, reasonably good
estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations. In other words, the results from using the Cox model will closely approximate the results for the correct parametric model. Geroski et al., (2007) indicated that the Cox proportional hazards model enables the user to characterize the failure process more rigorously than is possible with conventional approaches such as Probit and Logit models. Aspelund et al., (2005) also argued that the method is advantageous because it avoids biases associated with censoring and that the method is “informationally efficient”.

So, when one is in doubt, the Cox model is a "safe" choice because it gives a reliable enough result and the user does not need to worry about whether the wrong parametric model is chosen. Hence, it can be said that one of the advantages of semi-parametric model is that it has relaxed assumptions on population distributions where it does not require any particular probability distribution to represent survival times.

The simple interpretation given by the Cox model as "relative risk" type ratio is very desirable in explaining the risk of event for a certain covariate.

By letting \( x_1 = (x_1, x_2, \ldots, x_p) \) denote a collection of \( p \) explanatory variables that affect survival time, the hazard function for the Cox proportional hazards model is represented by:

\[
h(t, X) = h_0(t) \exp \left( \sum_{i=1}^{p} \beta_i X_i \right)
\]

The \( \beta \)'s give the maximum likelihood estimates while \( h_0(t) \) represents the
baseline hazard function. The baseline function involves $t$, but not the $X$ variables. For the Cox proportional hazards model $h_0(t)$ is obtained by replacing all the variables in $h(t, X)$ by zeroes (i.e. $X_i = 0$). Hosmer et al., (1999) noted that $h_0(t)$ characterizes how the hazard function changes as a function of survival time while $\exp(\beta X_i)$ characterizes how the hazard function changes as a function of covariates.

The assumption of proportional hazard requires that the hazard rate is constant over time or similarly, the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time (Bekele and Worku 2008). For example, when there is a two-level covariate with a value of 1 or 2, the hazard ratio becomes $e^\beta$. If the value of the coefficient is $p=ln(3)$, then it is simply saying that the subjects labeled with a 2 are three times more likely to have an event than the subjects labeled with the number one (1). In this way, we have a measure of difference between our exposure cohorts instead of simply knowing whether they were different. Joslyn and West (2000), in another example, examined the effect of race on breast carcinoma. Cox proportions and multivariate analyses were utilized. Simon and Severson (1997) utilized Cox proportional hazards regression to evaluate the effect of socio demographic and clinical variables on survival rates of African-American women and white women with breast cancer, and they found that African-American women were more likely than white women to survive breast cancer with respect to these variables.
Cox regression analysis was used by Abbas et al., (2009) to investigate the association between survival, socio-demographic and pathological factors, distant metastasis at diagnosis and treatment options.

3.8 EMPIRICAL MODEL

The empirical model estimated for this study is given in this sub-section. The dependent variable \( h(t) \) represents the probability that a woman will die at time \( t \) having survived until a given time \( t' \), conditional on a vector of covariates represented by \( X \).

The empirical model is based on the standard proportional hazards regression:

\[
\hat{h}(t) = \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7)
\]

where \( \hat{h}(t) \) represents the hazard rate, \( \beta \)'s, the coefficient to be estimated and \( X \) the covariates represented by: \( X_1 \) (Stage at diagnosis), \( X_2 \) (Tumour size), \( X_3 \) (Axillary node), \( X_4 \) (Tumour Grade), \( X_5 \) (Body Mass Index), \( X_6 \) (Age at diagnosis) and \( X_7 \) (ER/PR Status). The duration in the model is defined by: \( t_0 \), \( t \) and \( d \) where, \( t_0 \) is the start date, \( t \) the end of the observation window and \( d \) date of event (death) represents failure. In this study, \( t_0 \) is the starting date for diagnosis, \( t \) the time at which observation ends and \( d \) the date that a woman dies. Because of the incomplete nature of the observations at the right tail of the time axis, such observations are said to be right censored. The censoring indicator, represented by \( d \) is defined as follows:

\[
d = \begin{cases} 
0, & \text{if right censored} \\
1, & \text{if death occurs}
\end{cases}
\]
The choice of covariates to be included in the analysis were determined by prior expectations based on theory and previous empirical studies. Their expected effects on patients survival or failure are outlined below: The first prior expectation is that outlook for breast cancer varies according to whether the cancer is diagnosed early or when it is more advanced. The earlier a breast cancer is diagnosed, the smaller it is likely to be and the lower the chance that it would have spread. Thus, breast cancer patients diagnosed earlier face a lower risk of death (failure) compared to those diagnosed at an advanced stage. The second expectation is that tumour size correlates with the presence and number of involved axillary lymph nodes and is also an independent prognostic factor. Patients with smaller tumour size, less than one centimeter will experience higher survival times. The third expectation is that there is a direct relationship between the number of involved axillary nodes and the risk of failure. Thus, the lower the number of involved axillary nodes the lower the risk of failure (higher survival rate). Fourth is that tumour grade does have prognostic significance which yields a greater probability of survival for a lower grade tumour. Fifth is that a woman with a high BMI (obese) will experience a higher probability of failure than one with a lower BMI. The Sixth is that younger aged patients have higher probabilities of failure compared to older aged patients (postmenopausal).

From the results of the empirical analysis presented in the next chapter, the prior expectations will either be rejected or supported.
3.9 LOGISTIC REGRESSION

Logistic regression involves procedures that are primarily used when the dependent variable is dichotomous and the independent variables are a mixture of categorical and continuous variables. There may be a large number of variables some of which are categorical and others continuous. For such a mixture of variables, the multivariate normality assumption will not hold. In such a case, one could use the logistic regression as it does not make any assumptions about the distribution of the independent variables. If the independent variables are only categorical, logistic regression reduces to a contingency table analysis. In this study, the logistic regression is used to determine the likelihood of a woman being diagnosed of breast cancer given some risk factors such as the respondents reproductive, demographic and lifestyle information.

Logistic Regression analysis is thus, a model building technique which determines the category to which an observation belongs by using the specified characteristics of the observation. It also examines the types of the characteristics that can significantly be used to determine the group to which the observation should be placed. the use of logistic regression has increased in the social sciences (e.g., Chuang, 1997; Janik & Kravitz, 1994; Tolman & Weisz, 1995) and in educational research—especially in higher education (Austin, Yaffee and Hinkle, 1992; Cabrera, 1994; Peng & So, 2002a; Peng, So, Stage, & St. John, 2002).

Logistic regression was also applied by Yusuff et al., (2012) and Aidan et al., (2005) in their studies.
3.9.1 Mathematical Background

Suppose $Y$ is the random binary variable whose value is zero or one; where zero and one represents two different classes to which a subject may belong. The probability that $Y = 1$ is denoted by $p$, that is, $P(Y = 1) = p$, then $P(Y = 0) = 1 - p$. The ratio of $p$ to $1 - p$ is referred to as the odds in favour of the event that $Y = 1$ and is given by:

$$\text{odds}(Y = 1) = \frac{p}{1 - p}$$  \hspace{1cm} (3.18)

Generally, the odds in favour of an event is defined as:

$$\frac{\text{Number of favourable choices}}{\text{Number of unfavourable choices}}.$$

The odds against an event is the reciprocal of the odds in favour of the event and is given be the ratio

$$\frac{\text{Number of unfavourable choices}}{\text{Number of favourable choices}}.$$

Odds and probabilities provide the same information, but in different forms. The odds in the above equation may be converted into probability as

$$p = \frac{\text{odds}(Y = 1)}{1 + \text{odds}(Y = 1)}$$  \hspace{1cm} (3.19)

It may be more relevant to consider the natural logarithm of the value of the odds obtained. Thus,
\[
\ln[\text{odds}(Y = 1)] = \ln \frac{p}{1 - p} \tag{3.20}
\]

which is a function of the probability, \( p \). This function of \( p \) may be expressed as a linear combination of \( k \) independent random variables \((X_1, X_2, X_3, \ldots, X_k)\). The function of \( p \) is thus known as the logit\((p)\). Thus,

\[
\text{logit}(p) = \ln \frac{p}{1 - p} . \tag{3.21}
\]

The logit\((p)\) equation is given as

\[
\ln \frac{p}{1 - p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k , \tag{3.22}
\]

or, in a matrix form

\[
\ln \frac{p}{1 - p} = \mathbf{\beta} \mathbf{X} \tag{3.23}
\]

where

\[
\mathbf{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_3, \cdots, \beta_k) \]

is a vector of \( k + 1 \) coefficients of the logistic regression model and \( \mathbf{X} = (1, X_1, X_2, X_3, \cdots, X_k) \) is a vector of independent variables that are a combination of categorical and continuous variables. In either of the forms given, the odds is then given as

\[
\text{odds} [Y = \frac{1}{(X_1, X_2, X_3 \cdots X_k)}] = \frac{p}{1 - p}
\]
where \( p \) is the probability of the event \( Y = 1 \) given the values of the independent variables, \((X_1, X_2, X_3, \cdots, X_k)\). As can be seen the logarithm of the odds can be modeled as a linear function of the independent variables, and is equivalent to a multiple regression equation with the logarithm of the odds as the dependent variables.

The "logistic" distribution is an S-shaped distribution function, see Figure 3.4, which is similar to the standard-normal distribution (which results in a probit regression model) but easier to work with in most applications (the probabilities are easier to calculate). The logit distribution constrains the estimated probabilities to lie between 0 and 1.

![Graph of the Logistic function](image)

Figure 3.4 Graph of the Logistic function.

The logistic function, with \( z \) on the horizontal axis and \( f(z) \) on the vertical axis, where \( z = \beta X \). The logistic function is useful because it can take as an input any value from negative infinity to positive infinity, whereas the output is confined to values between 0 and 1. The variable \( z \) represents the exposure to some set of
independent explanatory variables, while \( f(z) \) represents the probability of a particular outcome, given that a particular set of values of the explanatory variables occurs.

\[
z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k
\]

The intercept at 0.5 on the \( f(z) \) axis corresponds to the value of \( z \) when the values of all the independent variables are zero (e.g., the value of \( z \) in persons with no risk factors). Each of the regression coefficients \( \beta_i, (i = 1, 2, \ldots, k) \) describes the size of the contribution of a particular risk factor. A positive regression coefficient means that the corresponding explanatory variable increases the probability of the outcome, while a negative one means that the corresponding variable decreases the probability of that outcome; a large regression coefficient means that the risk factor strongly influences the probability of that outcome, while a near-zero regression coefficient means that that risk factor has negligible influence on the probability of that outcome. The interpretation of the \( \beta_j \) parameter estimates is the additive effect on the log of the odds for a unit change in the \( j \)th explanatory variable. In the case of a dichotomous explanatory variable, for instance gender, the exponential of \( \beta \), \( e^\beta \), is the estimate of the odds of having the outcome for, say, males compared with females.

The model has an equivalent formulation

\[
p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}}
\]

Where the probability, \( p_i \), is obtained from the equation:
\[ p = \frac{1}{1 + \exp(-\beta X)} \]  

Equation (3.26) is such that: if \( \beta X = 0 \), \( p = 0.50 \); as \( \beta X \to \infty \), \( p \to 1 \); and as \( \beta X \to 0 \), \( p \to 0 \).

An odds ratio of one means that there is a fifty-fifty chance that the event will occur with a small change in the value of the independent variable.

Figure 3.5: A Graph of natural log function.

Negative coefficients lead to odds ratios less than one. Odds ratio less than one (negative coefficients) tend to be harder to interpret than those greater than one (positive coefficients). It is noted that odds ratio for continuous independent variables tend to be close to one; this does not suggest, however, that the coefficients are insignificant. Figure 3.5 depicts a graph of the natural log of the logit function against the odds ratio.
If the function logit(p) is linearly related to X, then the relation between X and p is non-linear, and takes the form of the elongated S-shaped curve as shown in Figure 3.4.

3.9.2 Hypothesis Testing

The null (H₀) and the alternative (H₁) hypotheses for testing the significance of a particular variable in the model is given by

\[ H_0 : \beta_i = 0 \quad \text{and} \quad H_1 : \beta_i \neq 0 \]

Testing the hypothesis that the coefficient of an individual independent variable is significantly different from zero is similar to the test in ordinary least square models. The Wald statistic is given by

\[ \text{Wald} = \left( \frac{\beta_i}{se(\beta_i)} \right)^2 \quad i = 1,2,..k \]

which is distributed as chi-square with one degree of freedom. The Wald statistic is the squared ratio of the unstandardized logit coefficient estimate, \( \beta_i \), to its standard error (se). The statistic is simply the square of the (asymptotic) t-statistic. A large value, (based on the level of significance), of the statistic leads to the rejection of the null hypothesis.

The Wald statistic is widely considered to be less reliable (Menard, 2002; Agresti, 1996). This is because for large logit coefficients, the standard error is inflated which lowers the Wald statistic, leading to Type II errors (false negatives: suggesting the effect was not significant when it was). That is, there is a flaw in
the Wald statistic to the extent that very large effects may lead to large standard errors and small Wald chi-square values. This means that a variable may be considered as non-significant in the model when in fact it is.

An alternative to the Wald statistic is the likelihood ratio test. The likelihood test statistic compares the difference in the value \(-2\log\text{likelihood}\) for the overall model with that of a nested model which drops one of the independent variables. One can use the likelihood ratio test to drop one variable from the model to create a nested reduced model. In this case, the likelihood ratio test tests if the logistic regression coefficient for the dropped variable can be treated as zero, thereby justifying dropping the variable from the model. A non-significant likelihood ratio test indicates no difference between the full and the reduced models, hence justifying dropping the given variable so as to have a more parsimonious model that works just as well.

When the reduced model is the baseline model with the constant only, (ie \(\beta_0\) only), the likelihood ratio test tests the significance of one’s model as a whole. A well-fitting model is significant at the 0.05 level or better, meaning the model in question is significantly different from the one with the constant only. The likelihood ratio test gives the overall test of the model.

3.9.3 Interpretation of the Logistic Regression Coefficients

To understand the interpretation of the logistic regression coefficients, the logistic model can be written in terms of the odds of an event occurring. As defined earlier, the odds of an event occurring is the probability that it will occur to the
probability that it will not. According to Tabachnick and Fidell (1996), the estimated coefficients and the related statistics form the logistic regression. The coefficients indicate roughly the change in the log odds associated with one-unit change in the independent variable. The exp(β) given indicates that factor of change in the odds associated with one-unit change in the independent variables. It actually states the exact relationship. The positive increase in β shows the likelihood increase in relation to the reference category, while a negative β shows the likelihood decrease in the relation to reference category. Zero β indicates little or no change in relation to reference category. In this study, the odds ratio of being diagnosed with breast cancer is the relative risk in favour of or against the risk factors. Thus, the interpretation of any fitted model requires that practical inferences are drawn from the estimated coefficients in the model. For most models this will involve the estimated coefficients for the independent variables in the model. The estimated coefficients for the independent variables represent the slope or rate of change of the dependent variable per unit of change in the independent variable. This means, determining the functional relationship between the dependent variable and the independent variable, and the appropriately defining the unit of change for the independent variable.

In logistic regression model, the slope coefficient represents the change in the logit for a change of one unit in the independent variable. The odds ratio measure association as it approximates how much more likely (or unlikely) it is for the outcome to be present among those who have a particular characteristic and among those who do not have.
3.11 PATH ANALYSIS

Path analysis model has a long history. It started in the 1930s as a method of studying direct and indirect effects of variables while regression analysis model remains as a method of discovering causal relationships Pedhazur (1997). Also path analysis model is not a substitute of regression analysis, rather it is a complementary methodology to regression analysis. A set of additional regressions is added to the original regression analysis to trace out indirect effects. Because of this complexity, a path diagram is typically used to display all of the causal relationships.

Path analysis can be said to be an extension of multiple regression. Its aim is to provide estimates of the magnitude and significance of hypothesised causal connections between sets of variables. Thus, path analysis utilizes multiple regression techniques to model direct and indirect relationships and allows putative causal pathways of association to be explored. Path coefficients are obtained using regression analyses; thus either the standardized or unstandardized regression coefficients are used as path coefficients (Duncan, 1970). The statistical assumptions for the path model, therefore, are those for regression analysis. Path analysis also provides the opportunity to partition relationships (decomposition of effects) into their component parts. The total effect of a variable on a subsequent variable can be partitioned into the direct effect (the direct path between the two variables) and the indirect effect (the effect through all possible intervening variables). Partitioning of effects allows examination of
the adequacy of the proposed model in describing the observed data and identifies factors which have the greatest contribution (total effect) on an outcome.

Path analysis is a technique which can be useful to epidemiologists because it allows a hypothesized time sequence and biologic knowledge to enter into the analysis. Veterinary epidemiologists have used path analysis for deterministic modelling in a wide variety of situations ranging from the relationships of diseases in dairy cows to morbidity and mortality in feedlot calves.

3.11.1 PATH ANALYSIS WITH LOGISTIC REGRESSION

Most of the theoretical work for path analysis was done using ordinary least squares (OLS) regression methods (Duncan, 1970). However, when an outcome is dichotomous (no/yes), OLS methods are not appropriate because the error distribution is binomial rather than normal and the dichotomous outcome is bounded by 0 and 1 (Fienberg, 1980). Recently, however, researchers have begun to use the logistic model for path analysis (e.g., Curtis et al., 1985, 1988). The use of logistic regression introduces a problem with partitioning of effects since the method of estimation is maximum likelihood rather than OLS. There appears to be no "calculus" to path coefficients when logistic regression is used (Fienberg, 1980). Thus, traditional partitioning of effects is not possible. However, in the logistic path model the magnitude of the direct and indirect associations can be examined by using odds ratios (OR) as path coefficients.

Note that absence of an interaction implies that there is no direct effect. Likewise, the lack of an indirect effect signifies that there is no interaction. Thus, the OR for
the two variables is not dependent on each other. It is proposed that in the absence of interaction the combined effect of two variables on a common outcome is obtained by multiplying the direct ORs for that outcome.

Figure 3.4: Models with Independent variables
CHAPTER FOUR

ANALYSIS OF RESULTS

4.0 INTRODUCTION

In this chapter results from the analytical framework are presented. The relationship between the selected covariates and the outcome of interest, which is, “suffering from breast cancer” or “not having breast cancer” is analysed. A separate analysis for all women suffering from breast cancer is presented and the relationship between the selected covariates and the outcome of interest, which are either, failure or survival, is also investigated.

4.1 PRELIMINARY RESULTS

Two thousand three hundred and ninety seven (2397) women were sampled for the study, of which 42.64% were diagnosed with breast cancer between the periods 1\textsuperscript{st} January 2002 to 31\textsuperscript{st} December 2008. The outcome variable used for the study was binary. The value 1 represented the case and 0 the control. The data was collected for 1375 controls and 1022 cases, for all the risk factors of the study as noted earlier in the last chapter.
Table 4.1: Descriptive Statistics of some Risk Factors for All the Women

<table>
<thead>
<tr>
<th>All women = 2397</th>
<th>AG</th>
<th>AFC</th>
<th>MN</th>
<th>MP</th>
<th>PT</th>
<th>APM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>20</td>
<td>12</td>
<td>9</td>
<td>25</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Mean</td>
<td>43.51</td>
<td>23.83</td>
<td>15.17</td>
<td>47.55</td>
<td>2.47</td>
<td>32.38</td>
</tr>
<tr>
<td>Median</td>
<td>43</td>
<td>24</td>
<td>15</td>
<td>48</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>IQR</td>
<td>18</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Maximum</td>
<td>92</td>
<td>46</td>
<td>26</td>
<td>60</td>
<td>12</td>
<td>45</td>
</tr>
<tr>
<td>SD</td>
<td>12.72</td>
<td>4.7</td>
<td>2.03</td>
<td>3.55</td>
<td>2.31</td>
<td>4</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.34</td>
<td>0.72</td>
<td>0.65</td>
<td>-1.47</td>
<td>0.89</td>
<td>-1.21</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.17</td>
<td>1.6</td>
<td>3.17</td>
<td>5.71</td>
<td>0.8</td>
<td>3.78</td>
</tr>
</tbody>
</table>

From table 4.1, the ages of all the women at first visit (AG) ranged from 20 – 92 years, with a mean age of 43.51 years, a standard deviation (SD) of 12.72 and a median age of 43 years. The mean age at first child (AFC), menarche (MN) and menopause (MP), were 23.83, 15.17 and 47.55 years respectively. The corresponding median values were 24, 15 and 48 years. The minimum ages at first child, menarche and menopause were 12, 9, and 25, whiles their maximum values were 46, 26 and 60 years respectively. Parity (PT) ranged from 0 – 12 children with a (Mean ± SD) value of 2.47 ± 2.31 and a median of 2. Similarly, the age interval between menarche and menopause (APM), ranged from 8 – 45 years with a (Mean ± SD) value of 32.38 ± 4 and median 33 years.
Table 4.2 Descriptive Statistics of some Risk Factors for the controls

<table>
<thead>
<tr>
<th>Control Group = 1375</th>
<th>AG</th>
<th>AFC</th>
<th>MN</th>
<th>MP</th>
<th>PT</th>
<th>APM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>20</td>
<td>12</td>
<td>9</td>
<td>25</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Mean</td>
<td>40.91</td>
<td>23.77</td>
<td>15.18</td>
<td>47.99</td>
<td>2.17</td>
<td>32.81</td>
</tr>
<tr>
<td>Median</td>
<td>39</td>
<td>24</td>
<td>15</td>
<td>48</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>IQR</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Maximum</td>
<td>82</td>
<td>46</td>
<td>22</td>
<td>60</td>
<td>12</td>
<td>45</td>
</tr>
<tr>
<td>SD</td>
<td>12.38</td>
<td>4.62</td>
<td>1.89</td>
<td>2.48</td>
<td>1.99</td>
<td>3.07</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.43</td>
<td>0.78</td>
<td>0.07</td>
<td>-1.61</td>
<td>0.92</td>
<td>-1.04</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.38</td>
<td>1.93</td>
<td>0.09</td>
<td>18.04</td>
<td>0.81</td>
<td>7.16</td>
</tr>
</tbody>
</table>

Descriptive statistics for the control group in table 4.2 indicates that, the ages at first visit ranged from 20 – 82 years with a (Mean ± SD) value of 40.91 ± 12.38 years and a median of 39 years. The mean ages at first child, menarche and menopause were 23.77, 15.18 and 47.88 years respectively. The age range was 12 – 46 years for AFC, 9 – 22 years for MN and 25 – 60 years for MP. The number of children per the control group ranged from 0 - 12 children with a mean value of 2.17 children. Age interval between menarche and menopausal age also ranged from 8 – 45 with a mean and median value at 32.18 and 33 years respectively.

Ages of the woman at first visit, age at first child, age at menarche and parity had platykurtic data sets for controls,( that is, they had flatter peak around their
means), which caused thin tails within the distribution. The flatness resulted from the data being less concentrated around its mean, due to large variations within observations.

Table 4.3: Descriptive Statistics of some Risk and Prognostic Factors for Breast Cancer Cases

<table>
<thead>
<tr>
<th>Cases=1022</th>
<th>AG</th>
<th>AFC</th>
<th>MN</th>
<th>MP</th>
<th>PT</th>
<th>APM</th>
<th>BMI</th>
<th>TS</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum</strong></td>
<td>20</td>
<td>12</td>
<td>10</td>
<td>32</td>
<td>0</td>
<td>12</td>
<td>15.24</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>47.97</td>
<td>23.91</td>
<td>15.16</td>
<td>46.97</td>
<td>2.89</td>
<td>31.8</td>
<td>28.27</td>
<td>5.57</td>
<td>2.03</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>47</td>
<td>24</td>
<td>15</td>
<td>48</td>
<td>3</td>
<td>33</td>
<td>28</td>
<td>4.6</td>
<td>2</td>
</tr>
<tr>
<td><strong>IQR</strong></td>
<td>15</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3.79</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>92</td>
<td>45</td>
<td>26</td>
<td>60</td>
<td>12</td>
<td>42</td>
<td>76.1</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>11.77</td>
<td>4.79</td>
<td>2.21</td>
<td>4.56</td>
<td>2.02</td>
<td>4.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.43</td>
<td>0.7</td>
<td>1.04</td>
<td>-1.02</td>
<td>0.86</td>
<td>-0.97</td>
<td>7.33</td>
<td>2.2</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>0.35</td>
<td>1.24</td>
<td>4.64</td>
<td>1.73</td>
<td>0.84</td>
<td>1.42</td>
<td>8.63</td>
<td>8.35</td>
<td>2.5</td>
</tr>
</tbody>
</table>

From table 4.3 the age for cases ranged from 20 – 92 years with a (Mean ± SD) value of $47.97 \pm 11.77$ years and median of 47 years. The mean age at first child (AFC), menarche (MN) and menopause (MP), were 23.91, 15.16 and 46.97 years respectively. The corresponding median values were 24, 15 and 48 years. The minimum ages at first child, menarche and menopause were 12, 10, and 32, whiles their maximum values were 45, 26 and 60 years respectively. Parity (PT) ranged from 0 – 12 children with a (Mean ± SD) value of $2.89 \pm 2.02$ and a median of 3. Similarly age interval between menarche and menopause (APM),...
ranged from 12 – 42 years with a (Mean ± SD) value of 31.80 ± 4.93 and median 33 years. The mean value for Body Mass Index (BMI), Tumour Size (TS) and Grade (GR) were, 28.27, 5.57 and 2.03 respectively. BMI values ranged from 15.24 – 76.1, that of TS was 0.5 – 35 and GR values ranged from 1 – 3.

Age at menopause and the age interval between menarche and menopause had leptokurtic data set, (the distribution has higher peaks around the mean), compared to normal distributions, which led to thick tails on both sides. These peaks resulted from the data being highly concentrated around the mean, due to lower variations within observations.

With regards to the cases, age of the woman, age at first child, age at menopause, parity, age interval between menarche and menopause had platykurtic data sets, whiles age at menarche, BMI and tumour size, had leptokurtic data set.

Figure 4.1 illustrates the percentage distribution of all women by age group, whiles Figure 4.2 shows the percentage distribution of cases within age group in percentages. The population age ranged from 20 to 92 years.
Figure 4.1: Distribution of Women by Age Group

Figure 4.2: Distribution of Breast Cancer Within Age Group
4.1.2 Analysis of Reproductive Factors

The analysis revealed that, there was a history of breastfeeding in 35.46% of the women of which 17.18% had breast cancer. Women with no history of breastfeeding had 56.63% being diagnosed with breast cancer, of these 55.25% were below 50 years of age, see appendix 3C.

Mean age at menarche was found to be 15.16 ± 2.21 years for cases (median = 15) and 15.18 ± 1.89 for controls (median=15). Age at menarche for 9.43% of the women was below 13 years of age and 47.79% of them were suffering from breast cancer. Of the remaining 90.57% who had their first menstruation on or after 13 years 42.10% were diagnosed of having breast cancer.

With respect to parity, 22.53% of the women in the study were nulliparous, 61.87% had 1- 4 children and remaining 15.60% had more than 5 children. Out of these 30%, 43.76% and 56.42% were having breast cancer respectively. The median parity was 2 children for all the women in the study whiles it was 3 for women with breast cancer.

Age at first child ranged from 12 – 46 years among controls (median = 24) and 12 – 45 years among cases (median = 24). Of cases 41.10% and of controls 36.44% reported their age at first child below 24 years. The mean age at first child was 23.91± 4.79 years among cases and 23.77± 4.62 among controls.

The mean age at menopause was 46.97±4.56 years (median = 48) for cases and 47.99± 2.48 (median = 48) for controls. Menopausal status was categorized into
pre menopause or early menopause (the length of time before and one year after the final menstrual period during which ovarian hormonal patterns change) and post menopause.

4.1.3 Hormonal, Lifestyle and Hereditary Factors

Contraceptive use was found in 12.18% of the women. Of the cases and controls, 7.63% and 15.56% respectively used contraceptives. Of the remaining women who did not use contraceptive, 47.22% had the disease.

There was a history of alcohol usage in 1.91% of the women and all had breast cancer.

Family history of breast cancer was found in 6.14% of the women with 26.71% associated with first degree relatives. Family history of other cancers like colon, prostrate, etc. were found in 0.25% of the women. Thus, 6.39% of the women had a family history of cancers, out of these 67.11% were actually suffering from breast cancer.

4.1.4 Social Characteristics

The data also showed that 59.49% of all the women were married, 4.9% were widowed, 28.41% single, 4.09% were divorced and 3.09% had incomplete data. Again of the 1022 women diagnosed with breast cancer, 66.14%, 20.35%, 7.14% and 6.36% were married, single, widowed and divorced respectively.
4.1.5 DESCRIPTION OF VARIABLES

The table below presents the selected variables and the description of the categorical and continuous variables used in the analysis.

Table 4.4 Description of Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variables</strong></td>
<td></td>
</tr>
<tr>
<td>Patients Survival</td>
<td>Survived = 0, failed = 1</td>
</tr>
<tr>
<td>Patients Status</td>
<td>Having breast cancer = 1, Not having Breast Cancer = 0</td>
</tr>
<tr>
<td><strong>Independent Variables (Prognostic / Risk factors)</strong></td>
<td></td>
</tr>
<tr>
<td>Stage (ST)</td>
<td>Stage 0-II = 0, Stage III = 1, Stage = 2, Stage IV = 3</td>
</tr>
<tr>
<td>Body Mass Index (BM)</td>
<td>&lt;25 = 0, 25 = 1</td>
</tr>
<tr>
<td>Grade (GR)</td>
<td>Grade 1 = 0, Grade 2 = 1, Grade 3 = 2</td>
</tr>
<tr>
<td>Axillary Node (AN)</td>
<td>≤ 25% = 0, &gt; 25% = 1</td>
</tr>
<tr>
<td>Tumour Size (TS)</td>
<td>&lt; 2 cm = 0, 2 - 5 cm = 1, &gt; 5 cm = 2</td>
</tr>
<tr>
<td>Age (AG)</td>
<td>Age of the woman at first visit</td>
</tr>
<tr>
<td>Age at Menopause (MP)</td>
<td>≤ 48 = 0, &gt; 48 = 1</td>
</tr>
<tr>
<td>Age at Menarche (MN)</td>
<td>&lt; 13 years = 0, ≥ 13 years = 1</td>
</tr>
<tr>
<td>Parity (PT)</td>
<td>nulliparous = 0, 1 - 4 = 1, &gt; 4 = 2</td>
</tr>
<tr>
<td>Age at first child (AFC)</td>
<td>&lt; 24 years, ≥ 24 years = 1</td>
</tr>
<tr>
<td>Breast Feeding (BF)</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>Contraception Usage (CTRP)</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>Alcohol Intake (AL)</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>Family History (FH)</td>
<td>Yes = 1, No = 0</td>
</tr>
<tr>
<td>Age interval between Menarche and Menopause (AMP)</td>
<td>Age difference between menarche and menopause</td>
</tr>
</tbody>
</table>
4.2 DESCRIPTIVE ANALYSIS OF PROGNOSTIC FACTORS

This section reports on the outcome of the data analysis on tumour characteristics, tumour staging, hormone receptors, tumour grading and histology.

4.2.1 Tumour Characteristics

The site of the tumour on presentation was the left breast in 46.57% of cases and the right breast in 53.43% of cases.

Lump size on presentation (revealed by ultrasound, mammography and/or clinical palpation) in 71.28% of the women was 2-5 cm on first diagnosis, 4.46% was less than 2 cm and in 24.26% of the women was greater than 5 cm on diagnosis.

Axillary lymph node involvement was found in 90% of the women diagnosed with breast cancer when they were first seen by a physician; 89.20% had axillary lymph nodes of more than 25% involvement.

4.2.2 Clinical Staging Analysis

Data relating to the clinical stages of breast cancer on first diagnosis showed that 14.47% of the women presented were in stages 0 & I, 33.17% were in stage II, 47.16% were in stage III and 5.20% were in stage IV. In all, 52.35% presented were in the advanced stage (III and IV) while early stage presentation involved 47.65% of the women. In terms of age distribution, 57.76% aged less than 50 years were in the advanced stage as compared to 42.24% who were 50 years and above.
The study revealed that, 56.85% of the women diagnosed with breast cancer were below 50 years of which 7%, 30% 48% and 5% were staged 0&I, II, III and IV respectively. This is shown in Figure 4.3. The remaining cases were 50 years and above, 11%, 38%, 46% and 5% in this age group were also staged 0 & I, II, III and IV respectively as depicted in Figure 4.3. It can also be seen from the Figure that, more of the patients staged 0&I and III were less than 50 years as compared to those in stage II. In addition, 52.94% of the reported cases were in the advanced stage (III and IV) of which 57.12% were below 50 years of age. The remaining 47.06% reported were in the early stage (0 & II and III) of which 56.55% were below 50 years of age.
Figure 4.4: Distribution by Stage and Menopausal Status.

Figure 4.4 illustrates Distribution by stage and menopausal status.

In all 31.15% of the women with breast cancer cases were premenopausal with 9.63%, 36.33%, 47.83% and 6.21% in stages 0&I, II, III and IV respectively. For those with post-menopausal status, 68.49% were diagnosed of breast cancer with the following: 16.71% in stage 0 & I, 31.71% in stage II, 46.86% in stage III and 4.71% in stage IV.
Table 4.5: Death at Tumour Stage, Age, and Menopausal Status

<table>
<thead>
<tr>
<th>Age Group &lt;50yrs</th>
<th>0 &amp; I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premenopausal</td>
<td>2</td>
<td>14</td>
<td>47</td>
<td>14</td>
<td>77</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>2</td>
<td>30</td>
<td>93</td>
<td>15</td>
<td>140</td>
</tr>
<tr>
<td><strong>Age Group ≥50yrs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>-</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>2</td>
<td>36</td>
<td>92</td>
<td>16</td>
<td>146</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6</td>
<td>101</td>
<td>235</td>
<td>48</td>
<td>390</td>
</tr>
</tbody>
</table>

From the study, there were 390 cause-specific deaths due to breast cancer as depicted in Table 4.5. 55.64% of them occurred in the age group less than 50 years of which 64.52% was in stage III, followed by stage II with 20.28%, IV with 13.36% and 0&I 1.84%. Between the menopausal groups within the age group of less than 50 years, 64.52% of the death was among post menopausal women and the remaining in premenopausal.

Women who were aged 50 years and above suffered 44.36% of the deaths. Out of these deaths, 51.91%, 30.60%, 10.38% and 1.09% were in stages III, II, IV and 1 respectively. In this age group too, the highest death occurred in women with post menopausal status.
On the whole, the menopausal group that experienced the highest death was post menopausal 73.33%. Again 60.26% of the total deaths occurred at stage III, 25.90% at stage II, 12.31% at stage IV and 1.54% at stage I.

4.2.3 ER/PR Receptors

Hormone receptor information was available for 309 women of which 153 were below 50 years and 156 were 50 years or above. Estrogen and Progesterone positive receptors were found in 54 of the women; 206 had ER and PR negative; ER negative and PR positive in 11 women and 38 women showed ER positive and PR negative receptors.

![Percentage Distribution of Receptors by Age](image)

Figure 4.5: ER/PR Receptors by Age Group
Figure 4.5 depicts the percentage distribution of the hormone receptors. For the age group of less than 50 years, 22%, 61%, 12% and 5% respectively had $\text{ER}^+/\text{PR}^+$, $\text{ER}^-/\text{PR}^-$, $\text{ER}^+/\text{PR}^-$ and $\text{ER}^-/\text{PR}^+$ respectively. The percentage distribution for the age group 50 years or above were 13%, 72%, 12% and 3% respectively with $\text{ER}^+/\text{PR}^+$, $\text{ER}^-/\text{PR}^-$, $\text{ER}^+/\text{PR}^-$ and $\text{ER}^-/\text{PR}^+$.

Table 4.6: ER/PR Receptors by Tumour Stage

<table>
<thead>
<tr>
<th>Hormone Receptors</th>
<th>0&amp;I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ER}^+/\text{PR}^+$</td>
<td>4</td>
<td>28</td>
<td>19</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>$\text{ER}^-/\text{PR}^-$</td>
<td>11</td>
<td>92</td>
<td>99</td>
<td>4</td>
<td>206</td>
</tr>
<tr>
<td>$\text{ER}^+/\text{PR}^-$</td>
<td>2</td>
<td>15</td>
<td>20</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>$\text{ER}^-/\text{PR}^+$</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>17</td>
<td>142</td>
<td>142</td>
<td>8</td>
<td>309</td>
</tr>
</tbody>
</table>

From the Table 4.6, the 309 women with information on hormone receptors studied, showed that 45.95%, 5.20% and 2.59% were in stages II, III, 0&I and IV respectively. Majority (66.67%), of the patients were identified with $\text{ER}^-/\text{PR}^-$, with the highest figure found in stage III followed by stages II, 0&I and IV in that order. Within stage II, 28 patients were identified with $\text{ER}^+/\text{PR}^+$ receptors, 15 and 7 identified with $\text{ER}^+/\text{PR}^-$ and $\text{ER}^-/\text{PR}^+$ receptors. There were no counts in stages I and IV for $\text{ER}^-/\text{PR}^+$ receptors, however there were 7 and 4 counts in stages II and III. Furthermore, there was information on (116) 37.54% of the patients on HER2/neu protein receptors. Triple negative breast cancer was identified in 66.38% of these patients of which 51.95% were less than age 50, see appendix D.
4.2.3 Analysis by Grading

Tumor grade is a system used to classify cancer cells in terms of how abnormal they look under a microscope and how quickly the tumor is likely to grow and spread. Many factors are considered when determining tumor grade, including the structure and growth pattern of the cells. The specific factors used to determine tumor grade vary with each type of cancer. Data relating to tumour grading was high (80.72%) for grade 2 (moderately differentiated) followed by grade 3 (11.35%) (poorly differentiated) and the least was 7.93% for grade 1.

4.2.4 Histology

From the cases studied, 72.90% of the women were diagnosed with IDC, 1.37% with Invasive Lobular Carcinoma, 12.52% with other types of Breast Cancer. There was no data for 13.21% of the women.

Table 4.7: Histology Presentation by Age and Deaths

<table>
<thead>
<tr>
<th>Age group</th>
<th>IDC</th>
<th>ILC</th>
<th>Papillary</th>
<th>Others</th>
<th>No data</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;50 years</td>
<td>390</td>
<td>8</td>
<td>2</td>
<td>147</td>
<td>34</td>
<td>581</td>
</tr>
<tr>
<td>≥50 years</td>
<td>345</td>
<td>6</td>
<td>3</td>
<td>64</td>
<td>23</td>
<td>441</td>
</tr>
<tr>
<td>Death</td>
<td>330</td>
<td>6</td>
<td>4</td>
<td>26</td>
<td>24</td>
<td>390</td>
</tr>
</tbody>
</table>

From Table 4.7 Invasive/Infiltrating Ductal Carcinoma (IDC) was the most common breast cancer type diagnosed among the women, accounting for 38.16% in the age group of less than 50 years and 34.05% in the age group of 50 years or above. Other types of cancers were distributed between the two age groupings as
14.58% and 6.36%. ILC was the next breast cancer type identified in the study, representing 1.37% of the total diagnosis, 0.78% from age group below 50 years and the remaining were in the age group of 50 years or above.

Invasive ductal carcinoma diagnosed contributed 84.62% to the total deaths and 15.38% deaths were contributed by other breast cancer types.

4.3 SUMMARY

Following the results and analysis presented in this chapter, a few observations were made. The overall mean age of patients was 43.51 years whiles that for breast cancer cases were 47.97 years. The largest number of cases being 59.69% of the total was aged 40-49 years. History of breastfeeding was found in 35.46% of the women of which 17.18% had breast cancer. Women with no history of breastfeeding had 56.63% chance of being diagnosed with breast cancer. Majority of the women (90.57%), had their first menstruation after 12 years of age out of which 42.10% suffered the disease compared with 47.79% who had their first menstruation before 13 years of age. 61.87% had between 1 and 4 children of which 43.76% had breast cancer. The data also indicated that 37.80% of the women had their first child below the age of 24 years and 45.14% of these were diagnosed with breast cancer. Of those who had their first child after 24 years 47.42% had breast cancer. Contraceptive use was found in 12.18% of the women and 26.71% were diagnosed of breast cancer as compared to 47.22% who never used Contraceptive.
From the study population, 18.85% were in the pre-menopausal group of which 69.91% were diagnosed with breast cancer. The Post-menopausal group constituted 81.15% of the population and 36.30% of these had breast cancer.

Majority of the cases (72.28%) of the women had a tumour size 2 to 5 cm on first diagnosis. Data relating to the clinical stages of breast cancer on first diagnosis showed the highest stage of presentation was stage III, accounting for 47.16% of the cases. In all, 52.94% of the diagnosis was presented in the advanced stage (stage III and IV).

There were 390 cause-specific deaths as a result of breast cancer, 55.97% of the deaths occurred in the age group of less than 50 years, of which 49.44% was in stage III. 78.62% of the deaths were among post menopausal women.

Triple negative breast cancer was identified in 66.38% of the cases with complete information on ER, PR and HER2 status. Majority of cases (51.95%) were less than 50 years.

The data relating to tumour grading showed that grade 2 (moderately differentiated) was high (80.72%) followed by grade 3 (poorly differentiated) (11.35%) and the least was 7.93% for grade 1.

Invasive Ductal Carcinoma was the most common breast cancer type diagnosed among the women, accounting for 71.92% followed by other types of cancers (12.52%). ILC was the next breast cancer type identified in the study, representing 1.37% of the total diagnosis.
4.4 SURVIVAL ANALYSIS

4.4.1 Kaplan-Meier method

Application of the Kaplan – Meier method to the data resulted in the following information: survival analysis showed 390 breast cancer-related deaths which were 38.16% among the 1022 subjects in the study. The mean survival time was 4.59 years (55.13 months). The 5-year overall survival was 47.9%. The survival curve is shown in Figure 4.6.

Figure 4.6: Survival Analysis for Breast Cancer Cases within the study period
Table 4.8: Multivariate 5-Year Survival Analysis of the Breast Cancer Patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Breast Cancer cases</th>
<th>Breast Cancer survivors</th>
<th>Survival rate</th>
<th>Test of Equality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Log rank</td>
</tr>
<tr>
<td>Axillary Node (AN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = (≤25%)</td>
<td>110</td>
<td>86</td>
<td></td>
<td>7.4(0.0064)</td>
</tr>
<tr>
<td>1 = (&gt;25%)</td>
<td>912</td>
<td>546</td>
<td>46.03</td>
<td></td>
</tr>
<tr>
<td>Tumour Size (TS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = &lt;2cm</td>
<td>47</td>
<td>34</td>
<td>-</td>
<td>19.13(0.0001)</td>
</tr>
<tr>
<td>1 = 2-5cm</td>
<td>714</td>
<td>462</td>
<td>52.3</td>
<td></td>
</tr>
<tr>
<td>2 = &gt;5cm</td>
<td>261</td>
<td>136</td>
<td>33.31</td>
<td></td>
</tr>
<tr>
<td>Stage (ST)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = 0&amp;I</td>
<td>148</td>
<td>142</td>
<td>91.94</td>
<td>133.18(0.000)</td>
</tr>
<tr>
<td>1 = II</td>
<td>339</td>
<td>238</td>
<td>59.93</td>
<td></td>
</tr>
<tr>
<td>2 = III</td>
<td>482</td>
<td>247</td>
<td>33.95</td>
<td></td>
</tr>
<tr>
<td>3 = IV</td>
<td>53</td>
<td>5</td>
<td>15.09</td>
<td></td>
</tr>
<tr>
<td>Tumour Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>81</td>
<td>51</td>
<td>49.32</td>
<td>2.00(0.367)</td>
</tr>
<tr>
<td>2</td>
<td>825</td>
<td>515</td>
<td>48.83</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>116</td>
<td>66</td>
<td>40.87</td>
<td></td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 = (≤25)</td>
<td>230</td>
<td>122</td>
<td>36.74</td>
<td>10.12(0.0015)</td>
</tr>
<tr>
<td>1 = (&gt;25)</td>
<td>792</td>
<td>510</td>
<td>50.26</td>
<td></td>
</tr>
</tbody>
</table>
4.4.2 Log Rank Analysis of Tumour Size

Tumour size less than 2 cm was found in 47 of the cases with 92.59% censored. For cases with tumour size between 2 and 5 cm and greater than 5 cm, the rates were 52.30% and 33.31% respectively. The mean survival times were 56.64 months and 45.62 months for tumour size between 2 and 5 cm and greater than 5 cm respectively. The median months were respectively 68 months and 36 months, with a p-value of 0.0004. Survival analysis of tumour size is shown in Figure 4.7. The graph of tumour sizes, less than 2 cm, between 2 and 5 cm, and greater than 5 cm had a log rank test ($\chi^2 = 12.5, P = 0.0003$) indicating significant differences among the groups and that of the Wilcoxon test is 10.08 ($p = 0.001$) also indicating significance.
Figure 4.7: Kaplan Meier Survival Curves for Tumour Size

The analysis further showed that with lymph node involvement of less than or equal to 25%, cumulative survival was 63.26% for 52 months (4.33 years). Lymph node involvement of more than 25% at diagnosis gave 5-year cumulative survival of 46.03%. The Log rank test of 13.22 (P = 0.0003) indicates significant differences between the two groups. Similarly Wilcoxon test statistics also reports significance of (P = 0.0003) for the lymph node. The survival curves depicted in Figure 4.8 indicates a better survival for lymph node involvement of less than 25% than that greater than or equal to 25%.
4.4.3 Analysis of the Clinical Stage

Among the 148 cases diagnosed in stage 0&I, the cumulative survival was 91.94% for stage 0&I, 59.93% for stage II; stage III had a rate of 33.95% and stage IV had 15.09%. Breast cancer mortality was correlated to the stage at diagnosis. Testing equality among the groups, P value of 0.000 for both tests indicated significance, see Figure 4.8. These support the observation by Davis et al (1997) that the earlier breast cancer is diagnosed and treated, the more likely a woman could survive the disease.
4.4.4 Tumour Grade Analysis

From the data, tumour grade 1 had 5 year cumulative survival rate of 42.41%, that of grade 2 was 55.46% and grade 3, 48.40%. Differences among tumour grades was not significant as the p-values were 0.3057 and 0.7496 for Log rank and Wilcoxon respectively. From Figure 4.9, the three graphs intersected each other at various points, confirming the insignificant nature of tumour grade for survival.
The results for the analysis of BMI revealed that 230 women had BMI less than 25, and had 5 years cumulative survival of 36.74% whiles that of women with BMI of more than 25 was 50.26%. A test statistics value of 10.12 and a corresponding p value of 0.0015 used for the Log rank test indicated significance. The Wilcoxon test results showed a marginal significance of 0.026. The survival curve for the two groups is shown in Figure 4.10, indicating a better survival for those with BMI greater than 25.

Figure 4.10: Kaplan Meier Survival Curves for Tumour Grading

4.4.5 Analysis of Body Mass Index

The results for the analysis of BMI revealed that 230 women had BMI less than 25, and had 5 years cumulative survival of 36.74% whiles that of women with BMI of more than 25 was 50.26%. A test statistics value of 10.12 and a corresponding p value of 0.0015 used for the Log rank test indicated significance. The Wilcoxon test results showed a marginal significance of 0.026. The survival curve for the two groups is shown in Figure 4.10, indicating a better survival for those with BMI greater than 25.
Figure 4.11: Kaplan Meier Survival Curves for Body Mass Index
Table 4.9: 5-Year Survival by Age, Menopausal Status and Histology

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Survivors</th>
<th>Survival Rate</th>
<th>Test Statistics (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Log Rank</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50yrs</td>
<td>581</td>
<td>364</td>
<td>48.05%</td>
<td>0.263</td>
</tr>
<tr>
<td>≥50yrs</td>
<td>441</td>
<td>268</td>
<td>47.98%</td>
<td></td>
</tr>
<tr>
<td><strong>Menopausal status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-menopausal</td>
<td>322</td>
<td>218</td>
<td>55.19%</td>
<td>0.024</td>
</tr>
<tr>
<td>Post menopausal</td>
<td>700</td>
<td>414</td>
<td>44.59%</td>
<td></td>
</tr>
<tr>
<td><strong>Histology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDC</td>
<td>745</td>
<td>415</td>
<td>42.95%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ILC</td>
<td>14</td>
<td>8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>263</td>
<td>209</td>
<td>65.03%</td>
<td></td>
</tr>
</tbody>
</table>

The study revealed that, there is no significant difference between the age groupings as far as survival is concerned. The survival rate was 48.05% for age group less than 50 years and 47.98% for age group above 49 years which was supported by the test statistic with p-values above the 0.05 significance level.

On the other hand there was a significant difference among the menopausal groups, with post menopausal group being the worst, having a survival rate of 44.59% as against 55.19% for the premenopausal group.

A 5-year survival for histology indicated that, there was a significant difference in survival rates of 42.95% for patients diagnosed with IDC as compared to 65.03% for other breast cancer types.

The survival analysis indicated a mean survival time of 4.58 years (55.13 months), with a 5-year overall survival of 47.91%.
In conclusion, lymph node involvement of more than 25% at diagnosis, had a 5-year cumulative survival of 46.03%, with \( P = 0.0003 \) indicating significant differences between the two groups. The highest survival rate occurred in the group with tumour size less than 2cm. Tumour size is a significant prognostic factor. Further, breast cancer mortality is said to be correlated to the stage at diagnosis. Tumour grading was not found to be a significant prognostic factor, \( p = 0.367 \) from the study. It can also be concluded that the higher the BMI the better the survival. There was no significant difference between the age groupings as far as survival was concerned. There was a significant difference among the menopausal groups, with post menopausal being the worst having a survival rate of 44.59%. A 5-year survival for histology indicated that, other breast cancer types have a higher survival rate, 65.03% than patients diagnosed of IDC, which was 42.95%.

4.4.6 Proportional Hazard Model

The Cox proportional hazard model is estimated in this section and the effects of covariates on survival are presented in terms of hazard ratios. According to Box-Steffensmier and Jones (2004), when the CPH model estimates take the form of a hazard ratio (HR), interpretation of the results is such that: if the hazard ratio was less than one, an increase in the variable associated with the coefficient reduced the risk (or hazard), thus resulting in a longer survival time. In contrast, hazard ratio greater than one implies that the risk was rising with an increase in the variable associated with the given coefficient estimate, leading to a reduced survival time. For hazard ratios closer to one, the resulting implication is that the
hazard rate was non-responsive to changes to the covariate, or the variable of interest had no influence on the increase or decrease of the hazard.

Table 4.10 shows the main results obtained from estimating the CPH model and they are discussed below. All coefficient estimates are interpreted while controlling for other variables.

<table>
<thead>
<tr>
<th>Table 4.10 Cox regression Analysis of 5-year Survival for all Patients by Prognostic Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>TS</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>ST</td>
</tr>
<tr>
<td>GR</td>
</tr>
<tr>
<td>BM</td>
</tr>
<tr>
<td>AN</td>
</tr>
</tbody>
</table>

Four variables met the 0.05 criterion for statistical significance: Tumour size, Stage at first diagnosis, axillary node and Body Mass Index. The estimated hazard ratio $e^β$ is the expected survival time for the groups. Thus, controlling for other covariates, the expected time for one to die of breast cancer, who was diagnosed of an advanced stage cancer was, 63% greater than for those in the early stage. The expected time to die for those with tumour size greater than 5cm was 48% greater than those with tumour size less than or equal to 5cm.
The probability of surviving for obese women suffering from breast cancer is 32% greater than that for non-obese patients.

For Axillary node involvement, the risk of death was 154% higher for those with more than 25% node involvement than those with 25% or less involved.

Results from CPMH suggest that age at diagnosis has no effect on survival or death of the patient. This is due to the hazard ratio being equal to 1. Thus the risk of dying is assumed to be the same for all ages at diagnosis. This result is supported at 5% level of significance.

From the results women with higher tumour grading faces a 23% (p = 0.322) higher risk of dying than those with lower tumour grade.

In conclusion, controlling for other covariates, the expected time to die when diagnosed of advanced stage breast cancer was 63% greater than for those in the early stage. The expected time to die for tumour size greater than 5cm is 48% greater than those with tumour size less than or equal to 5cm. The probability of surviving for obese women suffering from breast cancer is 32% greater than that of non-obese patients. The risk of death is 154% higher for those with more than 25% node involvement than those with 25% or less involved. Age at diagnosis has no effect on survival or death of the patient. From the results women with higher tumour grading faces a 23% higher risk of dying than those with lower tumour grade.
4.4.7 AFT Model and Cox-PH Model

As mentioned, the most commonly used method for analysing censored data is survival analysis. To model the survival time of breast cancer patients, the accelerated failure time model and the cox proportional hazard model were applied. The major objective for applying these models was to identify which of the six attributable variables were significant by contributing to the survival of breast cancer patients. The six explanatory variables used in the models are size of tumour, tumour grade; stage; axillary node; BMI and Age (age of the patient in years).

The most commonly used AFT models such as exponential; Weibull and lognormal AFT models and Cox-PH model were applied. After running the models including all covariates and interactions between covariates, the number of parameters that drive the attributable variables were reduced using stepwise regression based on Arakaike Information Criterion (AIC) which gives a measure of the goodness of fit of an estimated statistical model.

It is given by:

$$AIC = -2\log (\text{likelihood}) + 2(p + k),$$

where $p$ is the number of parameter, and $k$ is the number of parameters in the distribution. Statistical models with lower AIC are preferred.
Table 4.11: Summary of Results of Fitting Parametric AFT Models to Survival

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lognormal</th>
<th>Weibull</th>
<th>Exponential</th>
<th>Gamma</th>
<th>Llogistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.332</td>
<td>0.58</td>
<td>0.545</td>
<td>0.399</td>
<td>0.49</td>
</tr>
<tr>
<td>Tumour size</td>
<td>0.034*</td>
<td>0.018*</td>
<td>0.041*</td>
<td>0.028*</td>
<td>0.020*</td>
</tr>
<tr>
<td>Axillary Node</td>
<td>0.006*</td>
<td>0.026*</td>
<td>0.025*</td>
<td>0.010*</td>
<td>0.016*</td>
</tr>
<tr>
<td>BMI</td>
<td>0.072**</td>
<td>0.038*</td>
<td>0.062**</td>
<td>0.067**</td>
<td>0.084**</td>
</tr>
<tr>
<td>Grade</td>
<td>0.529</td>
<td>0.312</td>
<td>0.332</td>
<td>0.454</td>
<td>0.377</td>
</tr>
<tr>
<td>Stage</td>
<td>0.0001*</td>
<td>0.0001*</td>
<td>0.0001*</td>
<td>0.0001*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-404.758</td>
<td>-405.97</td>
<td>-420.303</td>
<td>-386.37</td>
<td>-405.67</td>
</tr>
<tr>
<td>Scale Parameter</td>
<td>1.209</td>
<td>0.681</td>
<td>1</td>
<td>1.057</td>
<td>0.618</td>
</tr>
<tr>
<td>Shape parameter</td>
<td>-</td>
<td>1.468</td>
<td>-</td>
<td>0.277</td>
<td>-</td>
</tr>
</tbody>
</table>

* significant at 0.05 level; **significant at 0.10 level

Four different parametric models – Exponential, Weibull, Lognormal and Llogistic- were applied to the data and the results presented as in Table 4.11. Axillary node, tumour size and stage at diagnosis were found to be significant at 0.05 level for all the distributions, however at 0.10 level of significance, BMI was significant for all the distributions. Based on the log likelihood estimates, the gamma model was found to be the best fitted model for predicting survival following a diagnosis of breast cancer.
4.5 STEPWISE LOGISTIC REGRESSION ANALYSIS

Logistic regression is used in modeling incidence of breast cancer. The model assumes a linear relation between the log of odds and independent variables, $X_1, X_2, \ldots, X_k$ and can be written in the following form:

$$\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k.$$  \hspace{1cm} 4.1

Where $p=P(Y=1)$

The maximum likelihood estimation is used to obtain the estimates of the model parameters. After estimators of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are computed, it is easy to compute predicted probabilities using the following formula derived from the equation 4.1.

$$p = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k}}.$$  \hspace{1cm} 4.2

Table 4.12 Stepwise Logistic Analysis of significant Risk factors of Breast Cancer for all the Women

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{\beta}$</th>
<th>$e^{\hat{\beta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>-0.7352</td>
<td>0.479</td>
</tr>
<tr>
<td>AG</td>
<td>0.0407</td>
<td>1.042</td>
</tr>
<tr>
<td>MP</td>
<td>0.8353</td>
<td>2.306</td>
</tr>
<tr>
<td>CTRP</td>
<td>-0.6813</td>
<td>0.506</td>
</tr>
<tr>
<td>BF</td>
<td>-1.7081</td>
<td>0.181</td>
</tr>
<tr>
<td>APM</td>
<td>-0.1223</td>
<td>0.885</td>
</tr>
</tbody>
</table>
The model looked at nine variables but only the significant ones are shown in Table 4.12. The explanatory variables Parity, family history, alcohol intake and age at first child were found to be insignificant variables and as such were dropped. Family history had an odds ratio of 1.291 which indicated an increased risk of 29.1% of developing breast cancer for those with family history as compared to those who had no family relation suffering from breast cancer. Thus, there is an increased risk of breast cancer if the woman has a family history of cancer. For each increase in full term pregnancy, the odds of being at risk in developing breast cancer increased from 1 to 1.252, which means multiparity increased the risk of developing breast cancer according to this study.

The results from the stepwise logistic model indicated the following:

Age at Menopause with an odd ratio of 2.306 indicates an increased risk of breast cancer among those who attained menopause on or after 48 years than those whose menopausal age was less than 48 years in the study population. In other words, a woman was 130.6% more likely of developing breast cancer if she attains menopause on or after 48 years. Thus age at menopause affected breast cancer development.

The occurrence of breast cancer was 0.506 as frequent among those who used contraceptives as among those who did not in the study population. Contraceptive use had an impact on the risk of developing breast cancer. Women who used contraceptives were 49.4% less likely to develop breast cancer than women who did not use contraceptives.
There is an inverse relationship between the risk of developing breast cancer and breast feeding. Thus, the more one breast feeds the less likely she is to develop breast cancer by 18.1%.

The odds ratio for age at menarche of the women was 0.479 indicating a decrease in risk of breast cancer development by 52.1% in women with menarche on or after 13 years. Age at menarche was a significant variable in the study.

The period between age at menarche and menopause was seen from previous studies to be associated with a significant higher risk of breast cancer. From the study, the time interval between age at menarche and age at menopause indicated a positive relationship between the risk of developing breast cancer and the period. An odd ratio of 0.885 indicated a decreased risk of developing breast cancer by 11.5% per year.

Table 4.13 Stepwise Logistic Analysis of significant Risk factors of Breast Cancer for Pre menopausal status

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{\beta}$</th>
<th>P value</th>
<th>$e^{\hat{\beta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN</td>
<td>-0.636</td>
<td>&lt;0.0009</td>
<td>0.231</td>
</tr>
<tr>
<td>AG</td>
<td>-0.0253</td>
<td>&lt;0.0003</td>
<td>0.9433</td>
</tr>
<tr>
<td>CTRP</td>
<td>-0.0414</td>
<td>&lt;0.0185</td>
<td>0.909</td>
</tr>
<tr>
<td>BF</td>
<td>-1.291</td>
<td>&lt;0.0001</td>
<td>0.511</td>
</tr>
<tr>
<td>APM</td>
<td>-0.0374</td>
<td>&lt;0.0001</td>
<td>0.9174</td>
</tr>
</tbody>
</table>
From Table 4.13, the study revealed that premenopausal women who did breast feed, whose menarche started on or after 13 years and those who used oral contraceptives, had 0.511, 0.231 and 0.909 folds chance respectively of developing breast cancer. A year’s increase in age decreased the risk of developing breast cancer in premenopausal woman by 5.67%. A year’s increase in age between menarche and menopause also decreased the risk of developing breast cancer by 8.26%. Nevertheless, no association was found in the current cohort between premenopausal and age at first childbirth, parity and family history after multivariate analysis at p greater than 0.05.

**Table 4.14 Stepwise Logistic Analysis of significant Risk factors of Breast Cancer for Post menopausal status**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{\beta}$</th>
<th>P value</th>
<th>$e^\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHIT</td>
<td>0.208</td>
<td>&lt;0.0401</td>
<td>1.613</td>
</tr>
<tr>
<td>AG</td>
<td>0.027</td>
<td>&lt;0.0001</td>
<td>1.064</td>
</tr>
<tr>
<td>CTRP</td>
<td>-0.270</td>
<td>&lt;0.0006</td>
<td>0.538</td>
</tr>
<tr>
<td>BF</td>
<td>-0.597</td>
<td>&lt;0.0001</td>
<td>0.253</td>
</tr>
<tr>
<td>APM</td>
<td>0.051</td>
<td>&lt;0.0001</td>
<td>1.124</td>
</tr>
</tbody>
</table>

There was no evidence of lowered risk by having later age at menarche, parity and later age at child birth p greater than 0.05. However, strong protective effect of breast cancer risk was observed for breast feeding (OR = 0.253) for post
menopausal women. Women with family history of breast cancer in this study had 61.30% of increase in risk for postmenopausal breast cancer when compared with women without family history in the same cohort (OR = 1.613). Contraceptive use had an effect on breast cancer development. Post menopausal women who did use contraceptive decreased their risk of developing breast cancer by 46.20%. Increase in age by a year increased the risk of breast cancer by 6.4% whiles a years increase in the age interval between menarche and menopause increased breast cancer risk by 12.4%.

4.6 PATH ANALYSIS

A Path Analysis was conducted to determine the effects of:

- age on parity and on breast cancer development;
- breast feeding on parity and on breast cancer development;
- age at first child on parity and on breast cancer development;
- breast feeding on menopause and on breast cancer development.

The analysis found that age and parity were negatively associated ($p = .001$), there was a statistically significant relationship between age and parity. This signifies interaction, which means the OR for age on breast cancer and OR for parity on breast cancer were dependent on each other. Also the analysis showed a direct effect of breast feeding on parity, although the relationship was negatively associated ($p = 0.0093$). There was also an interaction between age at first child and parity on breast cancer development at a significance level of 0.0008. There
was also significant (p = 0.0001) interaction present between breast feeding and menopause on breast cancer development, there existed indirect effect.

Figure 4.12: Path Model of some risk factors of Breast Cancer

4.5 Summary Of Results

Age at Menopause affects breast cancer development, breast cancer occurs more among those who attain menopause on or after 48 years than those whose menopausal age is less than 48 years in the study population. Contraceptive use has an impact on the risk of developing breast cancer; women who used contraceptives are less likely than women not using contraceptives. Breast feeding was also a very significant risk factor of one getting breast cancer. Multiparity increases the risk of breast cancer from the study but the factor is seen to be insignificant. Age at menarche on or after 13 years showed a decreased in
Using the dichotomous logistic regression with multiple predictor variables, the model indicates that, risk factors associated with breast cancer, age at menopause, breastfeeding, age at menarche, age interval between menarche and menopause and contraceptive use (p < 0.05), contribute to breast cancer development. With the overall model deemed significant, it suggests that no one factor or group of factors can explain more than half of the cause of breast cancer.

Path analysis found that age and parity, breast feeding and parity are negatively associated and statistically significant on the risk of developing breast cancer. There was also an interaction between age at first child and parity on breast cancer development as well as breast feeding and menopause.

Among premenopausal women, breastfeeding and age at menarche have stronger protective effect on breast cancer development. However a year’s increase in age decrease the risk of breast cancer in premenopausal women and increases the risk amongst postmenopausal. Similarly an increase in age interval between menarche and menopause by one decreases the risk in the case of premenopausal women and increases the risk in post menopausal women.

There was a statistically significant relationship between age and parity, thus the two variables depend on each other. Also the analysis showed a direct effect of...
breast feeding on parity, although the relationship was negatively associated. There was also an interaction between age at first child and parity on breast cancer development.
Chapter Five

Discussion and Conclusion

5.0 Introduction

This study identified the relationship between selected prognostic variables on breast cancer survival as well as selected variables and the risk of developing breast cancer. It is a case-control study, which has provided a description and analysis of reproductive risk factors of breast cancer as well as the role family history plays, which are non-modifiable breast cancer risk factors. This chapter discusses the findings and concludes the study, while making reference to the specific contributions to knowledge resulting from the study.

5.1 Discussion

A sample of 2397 women was selected for the study, of which 42.64% were diagnosed with breast cancer between the periods 1st January 2002 to 31st December 2008. The mean age from the study was 47.97 years for the women diagnosed with breast cancer consistent with (Anyanwu, 2008), whiles that for all the women was 43.51 years. The age range was 20 – 92, with the highest incidence rate occurring in age group 40 - 49 years. This indicates that, that Ghanaian women present breast cancer at a significantly younger age as compared to women in the developed world, (Adebamowo and Ajayi 2000; Schatzkin et al., 1987; Muyberry and Stoddard-Wright 1992; Bowen et al., 2008). Again within the age groupings, breast cancer increases with an increase in

www.udsspace.uds.edu.gh
age as depicted in Figure 4.2. Thus the older one gets, the higher the risk of being diagnosed of breast cancer.

5.1.1 Survival Analysis

A survival analysis was conducted in relation to the following factors: tumour size, lymph node involvement, clinical stage, tumour grade body mass index and age group. Survival analysis showed 390 cancer-related deaths (38.16%) among the 1022 subjects in the study. The overall survival for all subjects was 47.91% after 5 years which is consistent with a previous study by (Gajalakshmi et al., 1997).

5.1.2 Prognostic factors

Tumour size less than 2 cm was found in 47 of the cases with 92.59% censored. For cases with tumour size between 2 and 5 cm and greater than 5cm, the rates were 52.30% and 33.31% respectively. The mean survival time were 56.64 months and 45.62 months for tumour size between 2 and 5 cm and greater than 5cm respectively. The median months respectively were 68 months and 36 months with a p-value of 0.0004. Tumour size was found to be a significant variable with a p value of 0.0001. The expected time to die for those with tumour size greater than 5cm is 48% greater than those with tumour size less than or equal to 5cm. Thus the smaller the tumour size the better the prognosis.

Of the 110 cases with lymph node involvement of less than or equal to 25% and 86 censored, cumulative survival was 63.26% for 52 months (4.33 years). Among
912 cases with lymph node involvement of more than 25% at diagnosis and 546 censored, cumulative survival was 46.03%. There was a significant difference between the two groups at \( p = 0.0003 \). The hazard ratio is 2.54, which means the risk of death was 154% higher for those with more than 25% node involvement as against those with 25% or less involvement.

Data relating to the clinical stages of breast cancer on first diagnosis showed that in all 52.35% of the women were presented at the advanced stages (III and IV) whiles early stage presentation involved 47.65%. This is consistent with some studies done in Ghana, (Clegg-Lamptey and Hodasi, 2007; Asumanu et al., 2000 and Anim, 1979) and in Africa, (Boder, 2011; Ikpatt et al., 2002).

Cumulative survival for stage 0&I was 91.94%, stage II 59.93%, stage III had a rate of 33.95% and stage, IV 15.09%. Breast cancer mortality was found to be correlated to the stage at diagnosis. Testing equality among the groups, at a \( p \) value of 0.000 indicated significance. Considering the staging in terms of early and advanced, gave a hazard ratio of 1.63. Thus, controlling for other covariates, the expected time to death for breast cancer patients who were diagnosed with advanced stage cancer was 63% greater than for those in the early stage. 5-year survival was 68.95% for early stage diagnosis and 32.09% for advanced stage.

From the study population 18.85% were in the pre-menopausal group, 69.91% diagnosed of breast cancer. Post-menopause constituted 81.15% of the population, 36.30% has breast cancer.
These findings are similar to other studies, (example, Schrijvers et al., 1995) which identified clinical stage at diagnosis as an important determinant of survival.

Data relating to tumour grading was highest for grade 2, 80.72%, and least for grade 1 of 7.93%. Meanwhile, tumour grade 2 had 5 year cumulative survival rate of 48.83%, with the least being grade 3 of 40.87%. Differences among tumour grading were not significant as the p value was 0.3667 from Log rank statistics. From the results, women with higher tumour grading faced a 22.50% higher risk of dying than those with lower tumour grade. However the results were not supported at 5% significant level.

There was evidence of survivorship with BMI of 50.26% for BMI more than 25, (P = 0.0015) for Log rank indicates significance. The probability of surviving for obese women suffering from breast cancer has 32% greater than that of non-obese patients.

5.1.3 Age at diagnosis

Results from Proportional Hazard ratio test suggest that age at diagnosis has no effect on survival or death of the patient. This was due to the hazard ratio being equal to 1. Thus the risk of dying was assumed to be the same for all ages at diagnosis. This result is supported at 5% level of significance. The study revealed that, there was no significant difference between the age groupings as far as survival is concerned. The survival rate was 48.05% for the age group less than 50
years and 47.98% for the age group above 49 years which was supported by the test statistic with P > 0.05.

5-year survival for histology indicated that, there was a significant difference in survival rates of 42.95% for patients diagnosed with IDC as compared to 65.03% for other breast cancer types. The implication is that, a woman diagnosed of a cancer type other than IDC, has a better chance of surviving five years or more.

Using the Cox proportional hazards model four variables met the 0.05 criterion for statistical significance: Tumour size, Stage at first diagnosis, axillary node and Body Mass Index. This means that these variables affects ones rate of surviving breast cancer. A tumour size greater or equal to 5cm, advanced stage of presentation, axillary nodes involvement of more than 25% and BMI less than 25 indicated poor or low survival.

5.1.4 Breastfeeding and Contraception Usage

There was a history of breastfeeding in 35.46% of the women of which 17.17% had breast cancer. It can be said that breastfeeding was a protective variable against breast cancer, since minority (17.18%) of those who breastfed had the disease as against the majority (70.90%) of those who did no breastfeeding.

From the logistic regression analysis, breast feeding had an odds ratio of 0.181 meaning the risk of breast feeding was 0.181 times less of developing breast cancer than those who did not breastfeed (P = 0.0001). Again between premenopausal and postmenopausal women, breastfeeding was found to be more
protective in the premenopausal women (OR=0.511) than the post menopausal women (OR=0.253). These findings are similar to some previously done studies, (Romieu et al., 1996; Newcomb et al., 1994; Yang et al., 1993; Yoo et al., 1992; McTiernan et al., 1986). However, other studies done were not consistent with the outcome of this study; see section 4.4.3 in chapter 2.

Contraceptive use was found in 12.18% of the women and 26.71% were diagnosed of breast cancer. 2105 of the women did not use contraceptive out of which 55.20% were not diagnosed of the disease. From the study, the occurrence of breast cancer was 0.506 less frequent among those who used contraceptives as among those who did not in the study population. Contraceptive use had a positive impact on the risk of developing breast cancer and women who used contraceptives were 49.6% less likely of developing the disease than women who did not. The 95% confidence interval of 0.374 – 0.685 used was significant for contraceptive use. This result is consistent with some previous studies done, Norman et al. 2003 and inconsistent with the following studies; Kahlenborn et al., 2008; Casey et al., 2008; Lower, 2008; Ansink and Burger, 2007 Burke 2000; Tessaro et al., 2000; Brinton et al., 1998.

5.1.5 Reproductive Factors

The Median age at menarche was 15 years which was in conformity with several studies as discussed in section 2.4.3 of chapter 2. Age at menarche for 9.43% of the women was below 13 years, and 47.79% of them were suffering from breast cancer. This means that majority (52.21%) of the women who had their menarche...
below age 13 were not suffering from breast cancer. Of the remaining 90.57% who had their first menstruation on or after 13 years, 42.10% were diagnosed with having breast cancer. The odds ratio for age at menarche was 0.479, indicating a decreased in risk of developing breast cancer by 52.1% for those women with late menarche, (p = 0.0001).

From the study 22.53% of the women were nulliparous, 1483 (61.87%) had 1-4 children and the remaining 374 (15.60%) had more than 5 children. Out of a total of 1483 nulliparous women, 162 (30%) had breast cancer. Women who had 1-4 children, 649 (43.76%) had breast cancer and 211 (56.42%) of women with more than five children had breast cancer. This reveals that as the number of children increases the proportion of women diagnosed of breast cancer also increases.

For each increase in full term pregnancy, the odds of being at risk in developing breast cancer increased from 1 to 1.252, which means multiparity increased the risk of developing breast cancer according to this study.

This outcome is consistent with some studies already mentioned in section 2.4.3 of chapter 2.

The data analysis indicated that 906 of the women studied had their first child below the age of 24 years; 45.14% of these were diagnosed of breast cancer. Similarly, 1491 of the women had their first child when they were 24 years or above of which 40.38% were suffering from the disease. The median age at first child was 24 years. The odds ratio for age at first child for the women was 0.95, indicating a decrease risk in breast cancer development by 5%, (p > 0.05).
Age at Menopause with an odd ratio of 2.306 indicates an increased risk of breast cancer among those who attained menopause on or after 48 years than those whose menopausal age was less than 48 years in the study population. In other words, a woman was 130.6% more likely of developing breast cancer if she attains menopause on or after 48 years. Thus age at menopause affected breast cancer development. This is consistent with studies by Talamin et al., (1996), reporting of an increasing risk of breast cancer with increasing age at menopause (OR = 1.8 for 53+ years menopause against <45 years).

The period between age at menarche and menopause is seen from previous studies to be associated with significant higher risk of breast cancer, the longer the interval between the two variables (early menarche and late menopause), the higher the risk of developing breast cancer. From the study, the time interval between age at menarche and age at menopause indicates a negative relationship between the risk of developing breast cancer and the period. An odd ratio of 0.917 indicates a decreased risk of breast cancer by 8.30% for a year.

5.1.6 Family History

Family history of breast cancer was found in 146 of the women with 39 being associated with first degree relatives. Family history of other cancers like colon, prostate, etc. were associated with 6 of the women. Thus a total of 152 women had a family history of cancers of which 67.11% actually had breast cancer.

Family history had an odds ratio of 1.291 which indicated an increased risk of 29.1% of developing breast cancer for those with family history as compared to
those who had no family relation suffering from cancer. Thus, there is an increased risk of breast cancer if the woman has a family history of cancer, although this variable was insignificant as far as this study was concerned.

5.2 CONCLUSION

The study took into account Tumour size, Stage at first diagnosis, axillary node age at diagnosis, tumour grade and Body Mass Index as predictors of breast cancer survival. The significant results show that women with bigger tumour size faced a higher risk of dying than the risk faced by those with smaller tumour size. Regarding BMI the claim that obese women faced a higher probability of surviving when compared to women with normal bodyweight was supported. BMI was seen as a significant variable to predict breast cancer survival. The study’s results indicate that early staged breast cancer patients suffer a lower risk of death than the risk suffered by advanced staged breast cancers. As a component of cancer staging, axillary node, proved to be a significant predictor of survival with less than 25% involvement having a higher risk of survival. Age failed to explain the survival differences of the women. Contrary to the view that tumour grade has a positive influence on survival, the results from this study showed that tumour grade was an insignificant variable. There was a significant difference in survival rates for patients diagnosed with IDC (who faced a lower risk of survival) compared to patients with other breast cancer types. There was a significant difference among the menopausal groups, with post menopausal being the worst surviving group. Again 47.91% of the women will be surviving after 5
years of being diagnosed with the disease. The gamma model was found to be the best fitted model for predicting survival following a diagnosis of breast cancer.

The study provides several important findings. First, the results indicated that there was a decreased risk of developing breast cancer among those who breastfed than those who did not. Secondly, the result indicated an increased risk with age of the women on their first visit. Thus, for a year increase in age, the risk of developing breast cancer increased by 4.2% and this variable was significant. From the results, the risk of developing breast cancer was higher among women who attained menopause on or after age 48 years, than those whose menopausal age was less than 48 years. Thus age at menopause affected breast cancer development positively. Furthermore age at menarche increased the risk in those women with early menarche than those with late menarche (>13 years). For each increase in full term pregnancy, the risk of developing breast cancer increased, which means multiparity increased the risk of breast cancer. Later age at first child indicated a decreased risk of breast cancer, however, age at first child was not a significant variable in the study. Among the significant variables, contraceptive use had a negative impact on the risk of developing breast cancer: women who used contraceptives were less likely to develop the disease than women who did not use contraceptives. Family history had an increased risk compared to those who had no family history of breast cancer, although this variable was insignificant as far as this study was concerned. The time interval between age at menarche and age at menopause indicated a negative relationship on the risk of developing breast cancer for all the women but showed positive
relationship for post menopausal women. From the analysis alcohol use was not seen as a significant variable in the study. Under path analysis it was found that age and parity were significantly and negatively associated. The effect of breast feeding on parity was negatively associated. There was also an interaction between age at first child and parity on breast cancer development.

5.3 Contribution To Knowledge

This work has analyzed and put together important findings, for example, Tumour size, Stage at first diagnosis, axillary node, age at diagnosis and Body Mass Index are significant predictors in relation to breast cancer survival based on real data on breast cancer patients collected from the period 2002 to 2008. The study which is one of the first of its kind in respect of Ghana, will serve as a useful data base for further studies.

Furthermore, the work examined breast cancer cases among women utilizing survival analysis technique and Cox proportional hazard model with time-varying prognostic factors, will contribute to existing breast cancer literature. The study further examined the influence of some prognostic factors on breast cancer patients' survival.

Additionally the study developed a model using logistic regression to predict one's risk of being diagnosed of breast cancer by extensively examining the various demographic, reproductive and social risk factors.

To the best of the researcher's knowledge, this is the first study to utilize path analysis in identifying the relationship and effects among some breast cancer risk factors in Ghana.
This study will contribute to the current body of knowledge on breast cancer survivorship and the factors that may improve a patient’s quality of life during the transition period from being a patient to being a survivor. Results from this study will inform counselors, counselor educators, and other health and medical professionals, who specialize in the biopsychosocial concerns of oncology populations.

5.4 Recommendation

There is a greater need for a national database registry for breast cancer diagnosis and other diseases. There is also the need for the government to absorb the cost of treating breast cancer fully under the NHIS to encourage early detection and reduction in mortality. Again there is the need for schools, parents, religious organizations, the media houses, governmental/nongovernmental organizations to improve on health education/campaigns about breast cancer.

Since early menarche and late menopause, which increase the risk of breast cancer, have genetic bases there is the need for information sharing among family members in order to reduce this risk. Again due to the desire for secrecy that is sometimes found within families, there is the need for awareness to decrease the risk of the disease. It is also recommended that, sexually active women take contraceptive especially those in the post menopausal stage to reduce the risk of the breast cancer.
References


SAS Institute, Inc., Cary, N.C.


American Cancer Society, Cancer Facts & Figures for African Americans 2009-2010


Friedlin J. Having cancer is not sinful. Marie Claire. October 2005;246-250.


Hailun Wang, 2008 Statistical Analysis Of The Spectroscopic Evaluation Of Cervical Cancer


Kakarala M, Rozek L, Cote M, Liyanage S and Brenner DE. Breast cancer histology and receptor status characterization in Asian Indian and Pakistani women in the U.S. *BMC Cancer* 2010, 10:191


National Cancer Database.


NCRNM. Annual Report National Center for Radiotherapy and Nuclear Medicine, Korle Bu Teaching Hospital. 2007; Accra Ghana.


Reeves, G.K. et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. Bmj 2007;335(7630)1134


<table>
<thead>
<tr>
<th>SN.ID</th>
<th>Patient ID</th>
<th>Age</th>
<th>Marital Status</th>
<th>MN</th>
<th>MP</th>
<th>AFC</th>
<th>PT</th>
<th>FH</th>
<th>CTRP</th>
<th>AL</th>
<th>BF</th>
<th>AMP</th>
<th>WT</th>
<th>HT</th>
<th>DG</th>
<th>Prognostic Factors</th>
<th>ER</th>
<th>PR</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[www.udsspace.uds.edu.gh](http://www.udsspace.uds.edu.gh)
Appendix 2: Schedule for Walk in Clinic

<table>
<thead>
<tr>
<th>S/N</th>
<th>AG</th>
<th>MN</th>
<th>MP</th>
<th>AFC</th>
<th>PT</th>
<th>BF</th>
<th>AL</th>
<th>CTRP</th>
<th>FH</th>
<th>DG</th>
<th>INVESTIGATION</th>
<th>TREATMENT PLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

169

www.udsspace.uds.edu.gh
Appendix 3A

Frequency Distribution of all the Women by Age Group

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG: 20-29</td>
<td>353</td>
<td>14.73</td>
</tr>
<tr>
<td>30-39</td>
<td>581</td>
<td>24.24</td>
</tr>
<tr>
<td>40-49</td>
<td>712</td>
<td>29.70</td>
</tr>
<tr>
<td>50-59</td>
<td>489</td>
<td>20.40</td>
</tr>
<tr>
<td>60-69</td>
<td>193</td>
<td>8.05</td>
</tr>
<tr>
<td>70+</td>
<td>69</td>
<td>2.88</td>
</tr>
</tbody>
</table>

Appendix 3B

Frequency Distribution of women with Breast Cancer by Age

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG: 20-29</td>
<td>41</td>
<td>4.01</td>
</tr>
<tr>
<td>30-39</td>
<td>196</td>
<td>19.18</td>
</tr>
<tr>
<td>40-49</td>
<td>351</td>
<td>34.34</td>
</tr>
<tr>
<td>50-59</td>
<td>266</td>
<td>26.03</td>
</tr>
<tr>
<td>60-69</td>
<td>122</td>
<td>11.94</td>
</tr>
<tr>
<td>70+</td>
<td>34</td>
<td>3.33</td>
</tr>
</tbody>
</table>
## Frequency Distribution of all the Women by Reproductive Factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>% Frequency</th>
<th>% with Breast cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN: &lt;13</td>
<td>226</td>
<td>9.43</td>
<td>47.79</td>
</tr>
<tr>
<td></td>
<td>2171</td>
<td>90.57</td>
<td>42.10</td>
</tr>
<tr>
<td>AFC: &lt;24</td>
<td>906</td>
<td>37.80</td>
<td>46.36</td>
</tr>
<tr>
<td></td>
<td>1491</td>
<td>62.20</td>
<td>40.38</td>
</tr>
<tr>
<td>MP: &lt;48</td>
<td>2028</td>
<td>84.61</td>
<td>38.91</td>
</tr>
<tr>
<td></td>
<td>369</td>
<td>15.39</td>
<td>63.14</td>
</tr>
<tr>
<td>PT: 0</td>
<td>540</td>
<td>22.53</td>
<td>30.00</td>
</tr>
<tr>
<td>1–4</td>
<td>1483</td>
<td>61.87</td>
<td>43.76</td>
</tr>
<tr>
<td></td>
<td>377</td>
<td>15.73</td>
<td>56.42</td>
</tr>
<tr>
<td>CTRP: Yes</td>
<td>292</td>
<td>12.18</td>
<td>27.71</td>
</tr>
<tr>
<td>No</td>
<td>2105</td>
<td>87.82</td>
<td>44.85</td>
</tr>
<tr>
<td>FH: Yes</td>
<td>138</td>
<td>5.76</td>
<td>44.92</td>
</tr>
<tr>
<td>No</td>
<td>2259</td>
<td>94.24</td>
<td>42.50</td>
</tr>
<tr>
<td>ALC: Yes</td>
<td>47</td>
<td>1.96</td>
<td>100.00</td>
</tr>
<tr>
<td>No</td>
<td>2350</td>
<td>98.04</td>
<td>41.49</td>
</tr>
<tr>
<td>BF: Yes</td>
<td>850</td>
<td>35.46</td>
<td>17.18</td>
</tr>
<tr>
<td>No</td>
<td>1547</td>
<td>64.54</td>
<td>56.63</td>
</tr>
</tbody>
</table>
## Frequency Distribution of Women with Breast Cancer by Reproductive Factors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN: &lt;13</td>
<td>108</td>
<td>10.57</td>
</tr>
<tr>
<td>≥13</td>
<td>913</td>
<td>89.33</td>
</tr>
<tr>
<td>AFC: &lt;24</td>
<td>420</td>
<td>41.10</td>
</tr>
<tr>
<td>≥24</td>
<td>602</td>
<td>58.90</td>
</tr>
<tr>
<td>MP: &lt;48</td>
<td>789</td>
<td>77.20</td>
</tr>
<tr>
<td>≥48</td>
<td>233</td>
<td>22.80</td>
</tr>
<tr>
<td>PT: 0</td>
<td>162</td>
<td>15.85</td>
</tr>
<tr>
<td>1 – 4</td>
<td>649</td>
<td>63.50</td>
</tr>
<tr>
<td>≥5</td>
<td>211</td>
<td>20.65</td>
</tr>
<tr>
<td>CTRP: Yes</td>
<td>78</td>
<td>7.63</td>
</tr>
<tr>
<td>No</td>
<td>944</td>
<td>92.37</td>
</tr>
<tr>
<td>FH: Yes</td>
<td>62</td>
<td>6.07</td>
</tr>
<tr>
<td>No</td>
<td>960</td>
<td>93.93</td>
</tr>
<tr>
<td>ALC: Yes</td>
<td>47</td>
<td>4.60</td>
</tr>
<tr>
<td>No</td>
<td>975</td>
<td>95.40</td>
</tr>
<tr>
<td>BF: Yes</td>
<td>146</td>
<td>14.29</td>
</tr>
<tr>
<td>No</td>
<td>876</td>
<td>85.71</td>
</tr>
</tbody>
</table>
Appendix 4A: No. of patients by Hormone Status and age group

<table>
<thead>
<tr>
<th>Status</th>
<th>&lt;50</th>
<th>≥50</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER⁺</td>
<td>52</td>
<td>40</td>
<td>92</td>
</tr>
<tr>
<td>ER⁻</td>
<td>103</td>
<td>116</td>
<td>219</td>
</tr>
<tr>
<td>PR⁺</td>
<td>40</td>
<td>25</td>
<td>65</td>
</tr>
<tr>
<td>PR⁻</td>
<td>113</td>
<td>131</td>
<td>244</td>
</tr>
</tbody>
</table>

Appendix 4B: No. of patients by Hormone Status and age group

<table>
<thead>
<tr>
<th>Status</th>
<th>&lt;50</th>
<th>≥50</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER⁺/PR⁺</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>ER⁻/PR⁺</td>
<td>33</td>
<td>21</td>
<td>54</td>
</tr>
<tr>
<td>ER⁻/PR⁻</td>
<td>94</td>
<td>112</td>
<td>206</td>
</tr>
<tr>
<td>ER⁺/PR⁻</td>
<td>19</td>
<td>19</td>
<td>38</td>
</tr>
</tbody>
</table>
Appendix 4C: Frequency Distribution of ER, PR and HER2

<table>
<thead>
<tr>
<th>Status</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER+/PR+/Her-</td>
<td>1</td>
</tr>
<tr>
<td>ER+/PR+/Her-</td>
<td>22</td>
</tr>
<tr>
<td>ER+/PR-/Her-</td>
<td>77</td>
</tr>
<tr>
<td>ER+/PR-/Her+</td>
<td>4</td>
</tr>
<tr>
<td>ER-/PR+/Her+</td>
<td>1</td>
</tr>
<tr>
<td>ER-/PR+/Her+</td>
<td>11</td>
</tr>
</tbody>
</table>