
Jimma University

College of Graduate Studies

Department of Statistics

# Statistical Analysis Of Breast Cancer Types and Associated Factors: Application of Bayesian approach.

M.Sc. Thesis

By

Yoseph Yohannes

A Thesis Submitted to Department of Statistics, College of Natural Sciences,Jimma University for a Partial Fulfillment for the Requirements of the Degree of Master of Science in Biostatistics.

February, 2023

Jimma, Ethiopia

# Statistical Analysis Of Breast Cancer Types and Associated Factors: Application of Bayesian approach.

M.Sc. Thesis

By

Yoseph Yohannes

Advisor: Dr.Sisay Wondaya

Co-Advisor: Reta Habtamu(MSc)

February, 2023

Jimma, Ethiopia

**Approval Sheet-1**

This is to certify that the thesis entitled **"Statistical Analysis Of Breast Cancer Types and Associated Factors: Application of Bayesian approach"** submitted in partial fulfillment of the requirements for the degree of Master of Science in Biostatistics with the graduate program of the department of Statistics, Jimma University and is a record of original research carried out by **Yoseph Yohannes Dase,Id.No RM0702/13** under my supervision and no part of the thesis has been submitted for any other degree or diploma. The assistance and the help received during the course of this investigation have been duly acknowledged. Therefore, we recommend that the thesis would be accepted as partial fulfillment of the requirement for Master of Science in Biostatistics.

| | | |
|---|---|---|
| Dr.Sisay Wondaya | _____ | _____ |
| Advisor | Signature | Date |
| Mr.Reta Habtamu(MSc) | _____ | _____ |
| Co-advisor | Signature | Date |

**Approval Sheet-2**

We, the undersigned, members of the board of examiners of the final open defense by **Yoseph Yohannes Dase** have read and evaluated his thesis entitled **"Statistical Analysis Of Breast Cancer Types and Associated Factors: Application of Bayesian approach"** and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in statistics with specialization of Biostatistics.

| | | |
|---|---|---|
| Dr.Megersa B | _____ | _____ |
| Name of External Examiner | Signature | Date |
| Mr.Samuel Fikadu(MSc) | _____ | _____ |
| Name of Internal Examiner | Signature | Date |
| Mr.Kibralem Sisaye(MSc) | _____ | _____ |
| Name of chair person | Signature | Date |

## Declaration

I declare that this thesis is my original work, has not been presented for Degrees in any other University and all sources of materials used for the thesis have been duly acknowledged. This thesis has been submitted in partial fulfillments of the requirements for M.Sc. Degree in Biostatistics at Jimma University.

| **Yoseph Yohannes Dase** | _____ | _____ |
|---|---|---|
| Name of the student | Signature | Date |

# ACKNOWLEDGEMENT

First, and foremost glory to Almighty God for giving me full of health, peace, the opportunity to pursue my graduate study at Jimma University and for being with me in all aspects elsewhere.

My deep gratitude goes to my advisor Dr. Sisay Wondaya for his professional and deep comments, advice and guidance at any stage of this thesis. I appreciated his friendly contact and his punctuality for commented this paper Simply, without his frequent supervision, this thesis may not reach at the stage it is now.

I sincerely appreciate my co-advisor Mr. Reta Habtamu (MS.c) for his valuable suggestions and comments for the successful realization of this thesis. It is a great pleasure to express his commitment to share his experience, knowledge and his smooth contacts is also appreciable.

I would like to acknowledge Jimma University and all my instructors and staffs of Statistics Department at Jimma University for their cooperation from starting title selection up to the completion of this thesis. I would like to be grateful for my sponsor Sidama Regional State Court Bureau for the contribution in my study. Finally, I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future.

# List of Abbreviations

| | |
|---|---|
| BC | Breast Cancer |
| BMI | Body Mass Index |
| DDT | dichlorodiphenyltrichloroethane |
| FMoH | Federal Ministry of Health |
| GY | Gray (is the standard unit of absorbed inonizing-radiation dose) |
| HPV | Human papillomavirus |
| HRT | Hormone Replacement Therapy |
| IARC | International Agency for Research on Cancer |
| JU | Jimma University |
| JUMC | Jimma University Medical Center |
| LMICs | Low- and Middle-income countries |
| MCMC | Markov Chain Monte Carlo |
| MLE | Maximum Likelihood Estimation |
| MLR | Multinomial Logistic Regression |
| NCC | National Cancer Commitee |
| NCDs | Non-Communicable disease |
| NCCP | National Cancer Control Plan |
| SSA | Sub-Saharan Africa |
| UK | United Kingdom |
| WHO | World Health Organization |

# Contents

# List of Figures



# List of Tables

# Abstract

***Introduction:*** *Breast cancer is a form of the disease characterized by cells in the breast becoming abnormal and multiplying uncontrollably, resulting in a tumor. Breast cancer(BC) is the most common public health problem and the main cause of cancer-related death worldwide. In 2020, there were about 2.3 million new cases of breast cancer globally and about 685,000 deaths from this disease, with large geographical variations observed between countries and world regions.*

***Objectives:*** *This study has aimed to determine significant predictors of breast cancer for women by applying Bayesian approaches to multinomial logistic regression analyses. It is also designed to determine the risk factors and see the diagnosis of breast cancer patients at Jimma University Medical Center.*

***Methods:*** *The study is conducted at Jimma University Medical center and the data is secondary, which is obtained from Jimma University Medical center, Oncology Department. The response variable is the histologic type of breast cancer disease, whether it is ductal carcinoma, lobular carcinoma or it is other types of carcinoma. The multinomial unordered category of breast cancer cases has been analyzed with factors like age of patients, residence area, marital status, stages, treatments that a patient received, family history, HRT(hormone replacement therapy) and oral contraceptives. Bayesian Multinomial logistic regression is employed for posterior model fitting by using the MCMC Metropolis algorithm made for parameters.*

***Results:****Approximately 63.48%, 17.06%, and 19.46% of breast cancer women were attacked by ductal, lobular, and other histologic types of carcinoma respectively, at Jimma medical center. Women from urban areas were (RRR= 0.036; 95%CrI(0.005,0.162)) and (RRR=0.138;95%CrI(0.0444, 0.432)) less likely to have lobular and others histologic type of carcinoma respectively compared to women from rural areas. The patients who received treatment surgery and radiotherapy were (RRR= 6.706; 95%CrI(1.887,24.532)) and (RRR=7.924;95%CrI(1.896,36.162))respectively, more likely to have lobular carcinoma compared to the patients who received chemotherapy.*

***Conclussion:****Based on the study results we can conclude that age, residence area, treatments,oral contraceptives and releated factors had a significant effect on breast cancer patients.*

***Key Words:****Bayesian, Breast cancer, multinomial logistic regression, Markov chain, Monte Carlo.*

# CHAPTER ONE

# 1 INTRODUCTION

## 1.1 Background of the Study

Breast cancer(BC) is the most common public health problem and the main cause of cancer-related death worldwide (Grywalska et al., 2019; Sung et al., 2021). Breast cancer is a type of cancer characterized by uncontrolled development and the spread of abnormal breast cells (Rebner & Pai, 2020). In 2020, there were about 2.3 million new cases of breast cancer globally and about 685,000 deaths from this disease, with large geographical variations observed between countries and world regions.

Breast cancer incidence rates are highest in countries that have undergone economic transition, but transitioning countries carry a disproportionate share of breast cancer deaths (Arnold et al., 2022). Breast cancer is not a transmissible or infectious disease. Unlike some cancers that have infection-related causes, such as human papillomavirus (HPV) infection and cervical cancer, there are no known viral or bacterial infections linked to the development of breast cancer. More than half of the incidence of breast cancer and 60% of deaths occur in low- and middle-income countries (LMICs)(Grywalska et al., 2019; Mutebi et al., 2020). For example, African countries had the highest age-standardized mortality rate (17.3 deaths per 100,000 annually) associated with breast cancer (Azubuike et al., 2018; Hamdi et al., 2021). As we go through different stages in life, our bodies are subjected to many negative things. One of such negative things is a disease called cancer. Cancer is a generic term for a large group of diseases that affect any part of human body. Breast cancer is a form of disease that characterized by cells in the breast, becoming abnormal, and multiply uncontrollably, resulting in a tumor.

According to the World Health Organization (WHO) and American Cancer society, they recognized several risk factors for breast cancer, including early menarche, late menopause, null parity, late age at first pregnancy (>30 years), non-breastfeeding mothers, hormonal contraceptives, hormone therapy after menopause, alcohol, cigarette smoking, obesity, physical inactivity, diet, and family history of breast cancer(Tolessa et al., 2021). Several risk factors are assumed to increase disease occurrence, which incorporates modifiable and non-modifiable risk factors(Arifin et al., 2022). However, several epidemiological studies conducted in different populations indicate that the incidence and mortality of breast cancer greatly vary between countries. This indicates that environmental and lifestyle fac-

tors may contribute to the development of breast cancer(Torre et al., 2015; Arifin et al., 2022; Tan et al., 2018; Liu et al., 2017; Sepandi et al., 2014). In African countries, including Ethiopia, the health care system cannot handle care and treatment for the expanding incidence rates of breast cancer due to inappropriate treatment strategies in place and economic barriers. Therefore, focusing on preventive strategies through appropriate identification and exposure reduction to established risk factors is important(Azubuike et al., 2018; Tolessa et al., 2021; Kathrikolly et al., 2020).

In Ethiopia, breast cancer incidence is rising and become the foremost common cancer, causing high rates of morbidity and mortality(Timotewos et al., 2018). The incidence of breast cancer accounts for 15,244 (22.6%) all cases of cancer and 8,159 (17%) cancer mortality annually (Torre et al., 2015; Memirie et al., 2018). The age standardized incidence rate(ASR) of breast cancer was 40.6 per 100,000 females Timotewos et al. (2018). Although breast cancer overwhelmingly occurs in high-income countries, recently, the burden has extended in LMICs due to factors such as a westernized lifestyle and urbanization (Grywalska et al., 2019; Sung et al., 2021; Tolessa et al., 2021). Evidence indicates that breast cancer has several influences, including psychological distress to the patients, family members, reduced productivity, and increased cost of the health care system (Al-Azri et al., 2014).

Based on 2013 data from the Addis Ababa Cancer Registry, breast cancer accounted for 31.4%, cervical cancer for 14.3% and ovarian cancer for 6.3% of all cancer cases. Despite the fact that non-communicable diseases (NCDs), such as cardiovascular diseases, cancers, diabetes and chronic respiratory diseases, are on the increase in Ethiopia, the health systems in the country have traditionally concentrated on the Control of communicable diseases. As a result, health and development plans have not adequately invested in the Control of NCDs. The silent epidemic of NCDs now imposes a 'double burden of disease' to the country, that unless addressed, will overwhelm it in the near future. A study by Pham (2014)on survival analysis of breast cancer patients, using Weibull model, revealed that age, stage of cancer, treatment by only surgery, treatment by both surgery and radiation has effect on patients' survival time and similarly,study by Bacha et al. (2021) on risk factors affecting the survival time of breast cancer patients by using different shared frailty survival models revealed that the factors, such as age, place of residence, treatment taken, stage, histologic grade, tumor size, smoking habit, and oral contraceptives were significantly influencing breast cancer patients.

According to the only oncology centre in the country (the Tikur Anbessa (Black Lion) Specialized Hospital), about 80% of reported cases of cancer are diagnosed at advanced stages, when very little can be done to treat the disease. This is largely due to the low awareness of cancer signs and symptoms, inadequate screening and early detection and treatment services, inadequate diagnostic facilities and poorly structured referral. The country has very few cancer specialists (only 4 qualified oncologist for the entire population). This makes it difficult for a great majority of the population to access cancer treatment services, which results in long waiting times and cause many potentially curable tumors to progress to incurable stages (Haileselassie et al., 2019).

## 1.2 Statement of the Problem

According to the World Health Organization(WHO), breast cancer is the most common cancer among women worldwide, claiming the lives of hundreds of thousands of women each year and affecting countries at all levels of modernization. In 2020, there were 2.3 million women diagnosed with breast cancer and 685,000 deaths globally. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer. There are more lost disability-adjusted life years by women to breast cancer globally than any other type of cancer. Breast cancer occurs in every country of the world in women at any age after puberty but with increasing rates in later life(Ferlay et al., 2021).

Breast cancer is not a transmissible or infectious disease. Unlike some cancers that have infection-related causes, such as human papillomavirus(HPV) infection and cervical cancer, there are no known viral or bacterial infections linked to the development of breast cancer. The fight against this disease is given priority all over the world. Statistics regarding the issue tell the need for intensive efforts towards addressing the shock of this disease in Africa in particular because this chronic disease also affects this continent. Different techniques have been used such as classical and non-parametric statistics, but they still have important limitations and need to be improved.

Furthermore, several studies in Ethiopia on breast cancer health services were not using advanced statistical models. Therefore, this study mainly concentrated on multinomial logistic regression by applying Bayesian analysis to identify significant factors associated with breast cancer. Thus, the study have attempted to answer the basic research questions on.

✓How to explain the different patterns of breast cancer cases in different socio-economic and demographic,and Medical factors?

✓How to demonstrate the application of multinominal logistic regression with histological types of breast cancer,and in which type the women more attacked either ductal, lobular, and or other/unknown?
✓Which risk factor is more associated with Histologic types of Breast cancer?

## 1.3 Objective of the study

### 1.3.1 General Objectives

The main objective is to determine the significant predictors of breast cancer for women's by applying Bayesian approaches on multinomial logistic regression analyses.

### 1.3.2 Specific Objectives

- To explain different patterns of breast cancer cases in different health-related and socioeconomic and demographic statuses.
- To demonstrate the application of multinominal logistic regression with histological types of breast cancer,and in which type the women more attacked either ductal, lobular, and or other/unknown.
- To assess the association between breast cancer patients and some socio-demographic and Medical factors.

## 1.4 Significance of the Study

The results of this study may help the organization as well as individuals who work in this area to get a clue as to what extent of BC disease is at Jimma University Medical Center in the future. It may also be an input to see the strength of BC prevalence by comparing the result of this study with previous studies. The other basic significance of the study is that it may also further assist other researchers interested in this area and they may use it as a guideline for their future work. In determining the posterior distribution, the MCMC simulation technique is the most applicable method used for a long period of time. This has more advantages for health professionals(physicians) in order to give good treatment for breast cancer on identified risk factors. Also, the thesis will help academicians or readers in the selection of appropriate statistical approaches for further study on risk factors among breast cancer patients and the result of this study will also be expected to help those make a policy of any BC, concerned agendas and strategies.

# CHAPTER TWO

# 2 LITERATURE REVIEW

The problem of breast cancer is not a new scientific problem across many fields and among researchers. Most researchers from medical sciences, sociology, epidemiology, and other disciplines have studied the problem for many years, and investigations are still ongoing. This chapter aims to provide a synthesis of the literature about the specific points of interest defined in our objectives. We aim to review the literature on breast cancer prevalence, its explanatory factors, as well as their mechanisms of influence.

## 2.1 Global estimates of Breast Cancer

Breast cancer is currently the most common cancer globally, accounting for 12.5% of all new annual cancer cases worldwide. Here are the American Cancer Society estimates for breast cancer in the United States for 2022. About 13% (about 1 in 8) of U.S. women are going to develop invasive breast cancer in the course of their life. In 2022, an estimated 287,850 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 51,400 new cases of non-invasive (in situ) breast cancer. About 2,710 new cases of invasive breast cancer are expected to be diagnosed in men in 2022. A man's lifetime risk of breast cancer is about 1 in 833. As of January 2022, there are more than 3.8 million women with a history of breast cancer in the U.S. This includes women currently being treated and women who have finished treatment. Breast cancer is the most commonly diagnosed cancer among American women (Giaquinto et al., 2022).

In 2022, it's estimated that about 30% of newly diagnosed cancers in women are going to be breast cancers. Breast cancer incidence rates in the United States began decreasing in the year 2000, after increasing for the previous two decades. They dropped by 7% from 2002 to 2003 alone. One theory is that this decrease was partially due to the reduced use of hormone replacement therapy (HRT) by women after the results of a large study called the Women's Health Initiative were published in 2002. These results suggested a connection between HRT and increased breast cancer risk. In recent years, incidence rates have increased slightly by 0.5% per year. A woman's risk of breast cancer nearly doubles if she has a first-degree relative (mother, sister, daughter) who has been diagnosed with breast cancer. Approximately 15% of women who get breast cancer have a family member diagnosed with it.

About 85% of breast cancers occur in women who have no family history of breast cancer (Arzanova & Mayrovitz, 2022; Giaquinto et al., 2022). These occur due to genetic mutations that happen as a result of the aging process and life in general, rather than inherited mutations. The most significant risk factors for breast cancer are being a woman and getting older. If you are a trans man or a trans woman, it's important you speak with your doctor about your personal level of risk so you can make sure to get screened as often as makes sense for you. About 43,250 women in the U.S. are expected to die in 2022 from breast cancer. Breast cancer death rates have been decreasing steadily since 1989, for an overall decline of 43% through 2020. These decreases are thought to be the result of treatment advances and earlier detection through screening. However, the decline has slowed slightly in recent years. Breast cancer is one of the leading causes of cancer-related death in women in the United States, second only to lung cancer Giaquinto et al. (2022); Siegel et al. (2021); Arzanova & Mayrovitz (2022).

In developed countries such as Canada, reports showed that fewer of their women are dying from breast cancer. In the same vein, the Canadian Cancer Society reported a reduction by 42% since the peak in 1986 (Kim et al., 2014). It was indicated further that Canadian women who are diagnosed with breast cancer are living longer than ever before, based on a 5-year survival rate of 88% (Kim et al., 2014). A significant increase in the incidence of breast cancer has been noticed (Ferlay et al., 2021). This increment cuts across all continents of the world, and particularly Africa. Cancer of the breast is now the most dangerous disease common to women globally, and is a significant cause of cancer-related mortality in women across the world (Agboola et al., 2012). Therefore, it has increasingly become a focal point of research across the globe. On comparing the latest versions of the World Health Organization (WHO) and International Agency for Research on Cancer (IARC) reports on breast cancer(Ferlay et al. (2021)), it was discovered that the number of new cases increased from 12.7 million in 2008 to 14.1 million cases in 2012.

Breast cancer stages have been shown to be one of the major prognostic factorsMøller et al. (2016); McCormack et al. (2020), particularly in an African setting. But, Ethiopia does not have a national BC screening program, and over 80% of patients are diagnosed with advanced stage disease. Globally, (Ferlay et al., 2021) as reported by Bray et al. (2018) showed that the impact of breast cancer has been rising in most continents of the world. In the same vein, there are wide gaps between rich and poor countries as noticed by(Murugesan et al., 2017). Among the developed countries, the incidences remain highest, while mortality rates are relatively much higher in less-developed countries. Findings

show that the burden of breast cancer will increase in the years to come, not only because of the steep increase in incidence, but because of the increase in population in these countriesTaib et al. (2014) found that an increase in life expectancy is believed to be an outcome of a reduction in mortality from infectious diseases by 2020. The higher breast cancer mortality rate for women in less developed countries noticed in the literature is partly because clinical advances to combat the disease are not available for women(Forman et al., 2012). This is attributed to the lack of early detection and poor access to treatment as a result of lack of awareness, lack of education, and deficient infrastructural and healthcare facilities. According toYoulden et al. (2014), the report of Ferlay et al. (2021) indicated an alarming disparity in breast cancer incidence and mortality between the United States and the rest of the world.

In addition,the threat that breast cancer poses to human health, particularly Africa,Vineis & Wild (2014); Hamdi et al. (2021) observed that few countries in this region have breast cancer-related data. For instance, most of the breast cancer incidence data in Sub-Saharan Africa (SSA) in recent times were based on reports from registries (Parkin et al., 2010; Jedy-Agba et al., 2012). On the African continent, a serious challenge is the lack of cancer registries by most countries. Ethiopia is one of the them which has no most registered cancer at National level. The World Health Organization (WHO) estimates that the incidence rate of breast cancer in Africa has increased steadily over the years (Sung et al., 2021). As reported byAnderson et al. (2015) ,there is also a higher mortality rate among breast cancer patients in most African countries.

In Africa, breast cancer is responsible for 28% of all cancers and 20% all cancer deaths in women. (16% & 11% both sexes). Incidence rates are still generally low in Africa, estimated below 35 per 100,000 women in most countries(compared to over 90–120 per 100,000 in Europe or North America). Precise incidence figures in Africa are lacking given the absence of cancer registration in most countries. Recent incidence data from registries in Kampala, Harare, the Gambia and Mali-Bamako provide substantial support for the notion of an increasing breast cancer incidence in Sub-Saharan Africa. The Gambia and Mali reported the greatest rate of increase for women under age 55 years.The average age of diagnosis of breast cancers among African women tends to be women 50 years or younger a considerably younger age than seen in Caucasian populations and >70% of patients had node positive (Stage III tumors) (Clegg-Lamptey, 2017).

## 2.2 Socio-Economic and Demographic Factors

- **Age:** In all countries of the world age is strongly related to breast cancer. It is extremely rare below the age of 20 years, but thereafter, the incidence steadily rises so that by the age of 90 years about 20% of women are affected by this disease. The risk for breast cancer increases with age. Most breast cancers are diagnosed after age 50. Since breast cancer biology differs from patient to patient with respect to factors like age, variations in response to treatment, and substantial competing risks of mortality Biganzoli et al. (2021); Mieog et al. (2012); Derks et al. (2018), the exclusion of some patients might have led to invalid/unreliable results. This implies that the exclusion of elderly women in trials probably led to an untrue representation for the general older population (Ebenezer & Lougue, 2019). The Breast cancer patients and disease burden across various age groups have been reported from different parts of the world. According to study Ogunsakin & Lougue (2019) the proportion of respondents in each age group decreases with increasing age. Approximately 41% of the respondents were in the 20-34 years age group and approximately 27% were from the 35-49 years age group. Relatively smaller proportions of respondents were drawn from the older age groups.

- **Personal history of breast cancer or certain non-cancerous breast diseases:** Women who have had breast cancer are more likely to get breast cancer a second time. Some non-cancerous breast diseases such as a typical hyperplasia or lobular carcinoma in situ are associated with a higher risk of getting breast cancer. An individual who has already been diagnosed with breast cancer has a higher risk of developing it again, either in the same breast or the other breast, than if they never had the disease (Sun et al., 2017).

- **Race and geographical areas:** There are countless factors contributing to breast cancer which vary with respect to socio-economic, area, race, and life style differences. The role of lifestyle factors is well-illustrated in migrant studies, where breast cancer risks were compared between female migrants from low-to high-incidence countries and their offspring, showing that the risk increases in the following generations as a change in lifestyle is adopted (Shimizu et al., 1991).

  From the literature, it can be observed that white women are slightly more likely to develop breast cancer than African American women. Asian, Hispanic, and Native American women have a lower risk of developing and dying from breast cancer. In the United States, the rural health disparity has been a recent focus of attention and made a priority for improvement(Subrahmanian et al., 2018). While urban health is an emerging disciplineKrefis et al.

(2018) ,many standardized definitions of urban and rural exist and are used by social scientists and demographers, they are found in sources unfamiliar to health researchers and have largely not been used in public health studies.

- **Family history of breast or ovarian cancer:** A woman's risk for breast cancer is higher if she has a mother, sister, or daughter (first-degree relative) or multiple family members on either her mother's or father's side of the family who have had breast or ovarian cancer. Having a first-degree male relative with breast cancer also raises a woman's risk (Sun et al., 2017).

## 2.3 Medical factors

- **Oral Contraceptive:** Pills that control conception can also be a factor that increases chances of breast cancer. These pills work on the organs and hormones, some increasing blood flow for days, which only occurs when some hormones have been over produced. This abnormal multiplication can result in cancerous cells(Urban et al., 2013).

- **Hormone Replacement Therapy (HRT)**

  The International Agency for Research on Cancer (IARC) found that the use of combined oestrogen-progestogen hormone replacement therapy (HRT) for menopausal symptoms is as a highly probable cause of breast cancer. IARC has classified the use of oestrogen-only HRT as a possible cause of breast cancer, based on limited evidence (Chi et al., 2015). An estimated 3% of female breast cancers in the UK are linked to HRT use(Román et al., 2016). A report from Beral et al. (2019)has shown that breast cancer risk is also higher in oestrogen-only HRT users. These findings are supported by the findings of Chlebowski et al. (2020), which indicated that 5 years of combined HRT was associated with a 26% increased risk of invasive breast cancer in post-menopausal women.

- **Radiation therapy:** Radiation therapy for breast cancer uses high-energy X-rays, protons or other particles to kill cancer cells. Rapidly growing cells, such as cancer cells, are more susceptible to the effects of radiation therapy than are normal cells. Radiation therapy may be used to treat breast cancer at almost every stage. Radiation therapy is an effective way to reduce your risk of breast cancer recurring after surgery. In addition, it is commonly used to ease the symptoms caused by cancer that has spread to other parts of the body (metastatic breast cancer) (Stewart & Boorjian, 2015).

- **Exposure to estrogen:** Because the female hormone estrogen stimulates breast cell growth, exposure to estrogen over long periods of time, without any breaks,can increase the risk of breast cancer. Some of these risk factors are not under the individual's control, such as: starting menstruation (monthly periods) at a young age (before age 12) going through menopause (end of monthly cycles) at a late age (after 55) exposure to estrogens in the environment (such as hormones in meat or pesticides such as DDT, which produce estrogen-like substances when broken down by the body).

## 2.4 Overview Of Bayesian Modeling

The Bayesian statistical methodology presents a well-established framework for making an inference from observed data for quantities of interest by using an underlying probability model for a comprehensive overview of modern Bayesian statistical analysis. The Bayesian methodology differs from the classical frequentest approach in that all of the unknown parameters in the underlying probability model are treated as random variables, as opposed to unknown constants in the classical frequentest approach. As such, the unknown parameters are assigned prior distributions which are based on a prior subjective beliefs or scientific knowledge about the unknown parameters. In other words, prior distributions serve as probabilistic descriptions of what is known about the unknown parameters before observational data are collected and analyzed (Vehtari et al., 2017).

The classical statistics fit the logistic regression by means of an iterative procedure like maximum likelihood. In many situations, as a result of the assumptions underlying this iterative procedure, the estimation in classical statistics may result in non-convergence. These shortcomings as a result of non-convergence can be addressed using Bayesian inference as an alternative approach. Markov chain Monte Carlo(MCMC) algorithm approach can be used to provide a very general recipe for estimating properties of complicated distributions in Bayesian statistics without any difficulties.

**Why use Bayesian analysis?** There are many reasons to use Bayesian analysis instead of frequentest analytics. Bayesian analysis is really flexible in that: we can include information sources in addition to the data, we can make any comparisons between groups or data sets, Bayesian methods allow us to directly the question we are interested in: How plausible is our hypothesis given the data? This allows us to quantify uncertainty about the data and avoid terms such as "prove", and Models are more easily defined and are more flexible, and not susceptible to things such as separation.

Bayesian and classical statistics both have advantages and disadvantages, as well as some similarities. When the sample size is large, Bayesian statistics often provides results that are equivalent to those obtained by classical statistics. Bayesian statistics uses a single tool, Bayes Theorem, which is applicable in all situations. This contrasts to classical statistics that require many tools. Bayesian statistics are unaffected by the overfitting of a model, unlike classical statistics where the overfitting of model is a serious problem. Bayesian statistics allows for the incorporation of prior information in addition to the data that helps in strengthening inferences about the unknown parameters and can help in reducing necessary sample sizes. Bayesian inference via MCMC is unbiased with the small of sample size but classical statistics is biased when the sample size is small and the sample is sparse. Bayesian statistics via MCMC algorithms have a theoretical guarantee that the MCMC algorithm will converge if it is run long enough, unlike classical statistics where there are no guarantees for the convergence of the MLE, and so on (Bolstad & Curran, 2016).

In this paper, we apply multinomial logistic regression using Bayesian approach. As a requirement of the Bayesian approach, several diagnostics tests were performed to ensure convergence of the Markov chain Monte Carlo and the true reflection of the posterior distribution.

# CHAPTER THREE

# 3 DATA AND METHODOLOGY

## 3.1 Study Area

The study would be conducted at Jimma University Medical Center. Jimma University Medical Center was one of the oldest public hospitals in Ethiopia. It was established in 1930 E.C by Italian invaders for the service of their soldiers. After the withdrawal of the colonial occupants, it has been governed by the name of "Ras Desta Damtew Hospital" and later "Jimma Hospital" during the Dergue regime and currently JUMC. This time the hospital provides services for more than 20 million patients with 800 bedded. The hospital was located in Jimma city and, Jimma is the largest city in southwestern of Oromia Region at a distance of 352 Km from Addis Ababa, the capital city of Ethiopia. It has latitude and longitude of $7^0$40'N $36^0$50'E. Jimma has a relatively cool tropical monsoon climate. The temperatures are in a comfortable range, with the daily mean being from 20 to 25 degrees Celsius.

Jimma Hospital, which was established 60 years ago during the Italian Invasion, is the only specialized referral Hospital in Southwest Ethiopia situated to the east of the town at about 3km from the town, Jimma Municipality. Though it was built to serve the invading soldiers, after the defeat of the Italian forces, the Ethiopian Ministry of Health started to run it & give service to the people and, as of 1984, it became a training center for health workers (Medical Doctors, Nurses, Pharmacy Technicians, Laboratory Technicians & Environmental Health Experts) without infrastructure change or major renovation. Jimma University started an organized oncology service on 11 Oct 2008 on the Ethiopian calendar. Accordingly, Jimma Medical Centre's pediatric oncology unit was launched some five years ago, whereas the adult oncology unit was opened two years ago. The unit has received more than 2500 visitors as well as regular follow-ups for chemotherapy. More than a quarter of them are waiting for the commencement of the radiotherapy, which was under construction.

## 3.2 Source of data

The data for this study is mainly based on secondary data that is obtained from Jimma University Medical Center in the Oncology Department in Jimma, Ethiopia. The population of this study would include all women with breast cancer who have been registered at the Oncology Department of Jimma University Medical center for three years starting from September 2019 up to 2021. The data has

been carefully reviewed from the registration log book and patients' registration cards; any inadequate information encountered is checked from the file and excluded from analysis if proven to be inadequate. The data would include information on health-related demographic characteristics and socio-economic factors, and other information about the number of cases of women diagnosed with histologically and pathologically confirmed cancer of the breast.

## 3.3 Inclusion and Exclusion Criteria

**Inclusion Criteria:** All women with breast cancer registered with full information including the study. The variables of interest in the registration book or in the chart were considered to be eligible for the study.

**Exclusion Criteria:** Women with insufficient information regarding study variables on the registration book or on the card were not eligible.

## 3.4 Variable of study

Depending on the demonstrated related literature reviews, the variables that would be included in this study are listed as follows. As discussed in the literature review, the risk factors of breast cancer were socio-economic and demographic factors and medical factors.

### 3.4.1 Response variable

The outcome or response variable of interest in this study is the histologic types of breast cancer, which are either ductal, lobular, or others/unknown carcinoma. A woman is considered to have breast cancer if she has been checked by a health professional on which type is more attacked. Therefore, the outcome is a categorical variable that is measured as ductal, lobular, and others/unknown service care.

Table 3.1: Response variable

| Variable | Description | Category and Coding |
|---|---|---|
| Histologic | histologic types of BC | ductal=1<br>lobular=2<br>others/unknown=3 |

### 3.4.2 Independent variables

Independent variables are those variables that are presumed to affect or determine a dependent variable. They can be changed as required, and their values do not represent a problem requiring expla-

nation in analysis but are taken simply as given. The variables given below could be the major factors that may affect the histologic type of breast cancer.

Table 3.2: Explanator variables

| NO | Variables | Description | Category and Coding |
|---|---|---|---|
| 1 | age | age of the patients | 0=20-34<br>1=35-49<br>2=50-69<br>3=70+ |
| 2 | Residence | Patient's residence area | 0=Urban,1=Rural |
| 3 | Oral contraceptives | birth control pills | Not used=0<br><br>used=1 |
| 4 | Marital Status | Marital Status | married=0<br>single=1<br>separated=2 |
| 5 | stages | stages of breast cancer | I=0, II=1<br>III=2, IV=3 |
| 6 | Family history | Family history of breast cancer | No=0<br>Yes=1 |
| 7 | Treatments | Treatments that a patients' received | Chemotherapy=0<br>Radiotherapy=1<br>Surgery=2<br>Hormone Therapy=3<br>Combination of<br>the two or more=4 |
| 8 | HRT | HRT that a patients' received | Estrogen only=0<br>Cyclical or sequential=1<br>Continuous=2<br>Local estrogen=3 |

## 3.5 Methods of Analysis

Statistical analysis is carried out using the Bayesian multinomial logistic regression. The effect of explanatory variables on the dependent variable can be investigated using multinomial logistic regression that can be formulated under the Bayesian setup. We considered this approach in this study.

### 3.5.1 Introduction to Multinomial Logistic Regression

The logistic model, as a non-linear regression model, is a special case of a generalized linear model where the assumptions of normality and constant variance of residuals are not satisfied (McCullagh & Nelder, 2019). This model is a statistical technique for predicting the probability of an event, given a set of predictor variables. Logistic regression is used to predict the probability of a dependent variable on the basis of independent variables and to determine the effect size of the independent variables on

the dependent, to assess interaction effects, and to understand the impact of covariate control variables. The impact of predictor variables is usually explained in terms of the odds ratio and hence the name logistic regression also called the log-odds function. This model applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent variable occurring or not). There are three main types of logistic regression: binary, multinomial and ordinal. They differ in execution and theory. Binary regression deals with two possible values, essentially: yes or no. Multinomial logistic regression deals with three or more values. And ordinal logistic regression deals with three or more classes in a predetermined order (Hosmer Jr et al., 2013).

In this study, we used multinomial logistic regression by applying Bayesian analysis, since the dependent variable has 3 categories. The MLR model is a simple extension of the binomial logistic regression model. The multinomial response could be ordinal (ordered categories) or nominal (unordered categories). It is used when the dependent variable has more than two nominal or unordered categories, in which dummy coding of independent variables is quite common. The MLR does necessitate careful consideration of the sample size and examination for outlying cases. While a binary logistic regression model compares one dichotomy, a multinomial logistic regression model compares a number of dichotomies. A multinomial approach outputs a number of logistic regression models that make specific comparisons of the response categories. Multinomial logit models are multiequation models. A response variable with J categories will generate J-1 equations. Each of these J-1 equations is a binary logistic regression comparing a group with the reference group. Multinomial logistic regression simultaneously estimates the J-1 logits. Further, it is also the case, that the model tests all possible combinations among the k groups although it only displays coefficients for the J-1 comparisons. The probability of a categorical variable in a multinomial model is estimated using maximum likelihood estimation (Bayaga, 2010).

### 3.5.2 Model Formulation and Parameter Estimation in Multinomial Logistic Regression

**Model Formulation**

The simplest approach to multinomial data is to nominate one of the response categories as baseline or reference cell, calculate log-odds for all other categories relative to the baseline, and then let the log-odds be a linear function of the predictors. MLR models pairs each outcome category with a reference category. When there are J categories of the response variable, the model consists of J-1 logit equations which are fit simultaneously. The last category or probably the most common cate-

gory is assumed to be picked as reference (Agresti, 2016). There are different strategies for choose the reference category in dummy coding. From those, a few common options for choosing a category are normative category,largest category and the category whose mean is in the middle, or conversely, at one of the ends. In our case we take the most or largest category as a reference or baseline category. Suppose $x_i = (x_{i0}, x_{i1}, ....., x_{ip})^T$ denote the explanatory variables for individual $1 \leq i \leq n$ and $\beta_j = (\beta_{j0}, \beta_{j1}, ...., \beta_{jp})$, $1 \leq j \leq J-1$, a row vector, represent the regression parameters for the Jth reference category. Suppose yi = (yi1, yi2 ... yiJ ) denote a multinomial trial for individual $1 \leq i \leq n$. The trial yij is equal to one whenever a trial occurs in category j. Let $\pi_j(xi) = Pr((yij = 1)|xi)$ be the probability that the $i^{th}$ trial occurs in category j given a set of covariates xi. Then the multinomial logistic regression model

$$log\frac{\pi_j(xi)}{\pi_J(xi)} = \beta_j^T x, where j = 1, 2, ..., J-1 \tag{1}$$

With the logit link we can interpret the coefficients. In multinomial logistic regression the exponentiated coefficients are not odds ratios per se. The exponentiated coefficients, $exp(\beta_j)$, can be interpreted as relative risk ratios (RRR) or risk ratio, represents the odds of a trial falling into the category j against category J, all other things equal. An RRR is greater than one indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group increases as the variable increases. In other words the comparison outcome is more likely. On the other hand, if the value of relative risk ratio (RRR) is less than one, the relative risk of the outcome falling in the comparison group relative to the relative risk of the outcome falling in the reference group decreases as the variable increases. In other words the outcome is more likely to be in the reference group.

Using the logit link we have response probabilities

$$\pi_j(x) = \frac{exp(\beta_j^T x)}{1 + \sum_{k=1}^{J-1} exp(\beta_k^T x)} \tag{2}$$

In order to fit the MLR model we need to derive the log likelihood function for regression parameters. For the log likelihood of the regression parameters we use notation from (Agresti, 2016). The likelihood of the regression coefficients is derived from the multinomial likelihood function. For n independent observations

$$L(y|\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} \pi_j(x_i)^{y_{ij}} \tag{3}$$

by taking the log on both sides, we have expressed log likelihood as follow:

$$\ell(\beta) = log(L(y|\beta)) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} log\pi_j(x_i) \tag{4}$$

Putting the response probabilities in equation (2) into equation( 4) results in the log likelihood for our regression parameters as represented follow:

$$\ell(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} log \frac{exp(\beta_j^T x)}{1 + \sum_{k=1}^{J-1} exp(\beta_k^T x)} \tag{5}$$

$$\ell(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J} \left\{ y_{ij} \left( log \left[ exp(\beta_j^T x) \right] - log \left[ 1 + \sum_{k=1}^{J-1} exp(\beta_k^T x) \right] \right) \right\} \tag{6}$$

By taking the first partial derivatives of $\ell(\beta)$ with respect to each of the unknown parameters and setting these equations equal to zero,we can obtained the maximum likelihood estimators.

That means,

$$\frac{\partial \ell(\beta)}{\partial \beta} = 0 \tag{7}$$

As nonlinear equations, we use similar iterative procedures like Newton-Raphson iterative method. The Hessian matrix is calculated to obtain the estimator of the covariance matrix of the ML estimator, which is the inverse of the observed information matrix. Again, the estimates of the parameters and variance covariance matrix can be obtained by any standard statistical computer packages like SPSS, and R (nnet package).

Suppose, let Y is a categorical response variable with three categories, in our case it represented as 1, 2 and 3. Since the outcome variable has three categories, with two logit models, as the logistic regression model uses a binary outcome variable which parameterizes in terms of the logit y= 1 against y =0. Assume there are p explanatory variables, $x = (x_1, x_2, ...., x_p)$, in our model we have 8 explanatory variables,$x = (x_1, x_2, ...., x_8)$. The logit models for nominal responses pair each response category to a baseline category and the choice is arbitrary. If we set the last category as the baseline, then the baseline category logits are represented as follows:

$$log(\frac{Pr(y=2|x)}{Pr(y=1|x)}) = \beta_{20} + \beta_{21} x_1 + ..... + \beta_{28} x_8 = \beta_2^T x \tag{8}$$

$$log(\frac{Pr(y=3|x)}{Pr(y=1|x)}) = \beta_{30} + \beta_{31} x_1 + ..... + \beta_{38} x_8 = \beta_3^T x \tag{9}$$

From those two logit equations,we have a response probabilities as:

$$Pr(y=2|x) = \frac{exp(\beta_2^T x)}{1 + exp(\beta_2^T x) + exp(\beta_3^T x)} \tag{10}$$

$$Pr(y=3|x) = \frac{exp(\beta_3^T x)}{1 + exp(\beta_2^T x) + exp(\beta_3^T x)} \tag{11}$$

$$Pr(y=1|x) = \frac{1}{1+exp(\beta_2^T x)+exp(\beta_3^T x)} \tag{12}$$

with unknown parameters $\beta = (\beta_2, \beta_3)$. Thus, the two coefficients, $\beta_2$ and $\beta_3$ represent the log odds of being in the target groups relative to the reference group.

Now, we recode the outcome variables as the following

$Y_1 = 1, Y_2 = 0, Y_3 = 0 for Y = 1$

$Y_1 = 0, Y_2 = 1, Y_3 = 0 for Y = 2$

$Y_1 = 0, Y_2 = 0, Y_3 = 1 for Y = 3$

From the requirded outcome variables, Since, the value Y takes, the sum of these outcome variables is $\sum_{i=1}^{3} y_i = 1$.

The conditional likelihood function, given the covariates for independent observations of sample n, is expressed as

$$L(y|\beta) = \prod_{j=1}^{n} \left\{ \left( \frac{exp(\beta_2^T x)}{1+exp(\beta_2^T x)+exp(\beta_3^T x)} \right)^{y_{1j}} \left( \frac{exp(\beta_3^T x)}{1+exp(\beta_2^T x)+exp(\beta_3^T x)} \right)^{y_{2j}} \left( \frac{1}{1+exp(\beta_2^T x)+exp(\beta_3^T x)} \right)^{y_{3j}} \right\} \tag{13}$$

by taking the log on both sides, the expression reduces to

$\ell(\beta) = ln(L(y|\beta))$

$= \sum_{j=1}^{n} \{ y_{1j} \left( \beta_2^T x - ln \left[ 1 + exp(\beta_2^T x) + exp(\beta_3^T x) \right] \right) + y_{2j} \left( \beta_3^T x - ln \left[ 1 + exp(\beta_2^T x) + exp(\beta_3^T x) \right] \right)$

$- y_{3j} ln \left( 1 + exp(\beta_2^T x) + exp(\beta_3^T x) \right) \}$

Or

$$\ell(\beta) = \sum_{j=1}^{n} \left\{ y_{1j}(\beta_2^T x) + y_{2j}(\beta_3^T x) - (y_{1j} + y_{2j} + y_{3j}) ln \left( 1 + exp(\beta_2^T x) + exp(\beta_3^T x) \right) \right\} \tag{14}$$

But,from the above definition ,$\sum_{i=1}^{n} yij = 1$ for each j.Then equation 14 reduced to

$$\ell(\beta) = \sum_{j=1}^{n} \left\{ y_{1j}(\beta_2^T x) + y_{2j}(\beta_3^T x) - ln \left( 1 + exp(\beta_2^T x) + exp(\beta_3^T x) \right) \right\} \tag{15}$$

Hence, by maximizing the above equation 15 we can theoretically estimate the parameters $\beta$. However, the resulting equation obtained by talking the first order derivatives do not have an analytical solution. Therefore, $\beta$ can be obtained by maximizing the equation using the most effective and well known Newton-Raphson iterative method (Agresti, 2016).

## 3.6 Bayesian multinomial Logistic Regression

In classical statistics, the parameters in the models are considered to be fixed, while in Bayesian analysis, the parameters are treated as random variables. The parameters being random variables, they are

given distributions. Prior distribution of the parameters is one before data is collected while posterior distribution is one realized after scaling the prior distribution with new information obtained. The posterior can be interpreted as the summary (in a probabilistic sense) of the available information on $\beta$, once x is observed. The Bayesian approach realizes somewhat the updating of prior information by observation of x, through $\pi(\beta|x)$ (Frühwirth-Schnatter & Frühwirth, 2012). Bayesian inference for logistic regression models is derived applying a Markov Chain Monte Carlo algorithm to simulate from the joint posterior distribution of the regression and the link parameters. In the Bayesian framework, there are three key components associated with parameter estimation: the prior distribution, the likelihood function, and the posterior distribution.

### 3.6.1 Likelihood Function

The one part to a Bayesian analysis are the likelihood function, which reflects information about the parameters contained in the data, and the prior distribution, which quantifies what, is known about the parameters before observing data. We can see the likelihood function of y for $i^{th}$ individual with $j^{th}$ category is expressed above equation(3).

### 3.6.2 Bayesian Prior Distribution

One of the prior conditions in any Bayesian estimation is the choice of a prior for all the unknown parameters, thus this is main important feature of any Bayesian estimation. With small samples this choice can be critical, but with larger samples the choice is less crucial, since information in the data be more important than information in the prior. If the posterior is highly dependent on the prior, then the data(likelihood function) may not contain sufficient information. Bayesian statistics differs from classical statistics in the sense that parameters are regarded as random variables in the former, while a prior distribution has to be specified in order to make inference in the latter.

The major challenge in Bayesian statistics is the correct specification of a Bayesian prior distribution, because appropriate prior specification is key in Bayesian modeling. Gelman et al. (2017)indicated that the prior distribution is an important part of Bayesian inference, representing information about an uncertain parameter $\beta$ which is combined with the probability distribution of the likelihood of new data to produce the posterior distribution. This is then used for future inference on $\beta$. Therefore, necessary precaution should be taken in selecting priors because inappropriate choices of priors may result to wrong inference. In specifying priors, a number of points need to be considered. A key point among them is the fact that priors can be not sure. Because inference is assumed to be dependent

on prior choice, alternative priors are examined to explore how sensitive the main conclusions are to alterations in the prior.

There are two types of prior distribution namely, informative and non-informative prior distributions can be used. Informative prior distributions are applied if something is known about the likely values of the unknown parameters. On the other hand, non-informative priors are employed if either little is known about the coefficient values or if one wishes to ensure that prior information plays very little role in the analysis. That is, the data is allowed to remain influential in the analysis. As a result of the objectivity of non-informative priors, the majority of researchers in statistics prefer to make use of it compared to informative priors. The most common choice of non-informative priors is the flat prior, which assigns equal likelihood to all possible values of the parameters. In our case, there is no enough information or little prior knowledge about the value of parameters, therefore, we take non informative prior distribution. Non informative priors are employed since we want prior information to play a very little role in our analysis which makes the data to remain influential in the analysis. For this purpose, the most common priors for logistic regression parameters is normal distributed, and we assume a multivariate normal prior on $\beta$ with a large variance. In the case of Bayesian approach, prior distributions were assigned to all the parameters and the prior distribution of $\beta$ is assumed to be multivariate normal with parameter $\beta_j \sim N(\mu_\beta, \Sigma_\beta)$.

$$p(\beta) \propto \left[\frac{1}{2\pi|\Sigma_\beta|}\right]^{p/2} exp\left\{-\frac{1}{2}\left(\left(\beta-\mu_\beta\right)' \Sigma_\beta^{-1}\left(\beta-\mu_\beta\right)\right)\right\} \tag{16}$$

Where,p is the size of the design matrix X.

The most common choice for $\mu_\beta$ is zero vectors, and $\Sigma_\beta$ is usually chosen to be a diagonal matrix($\Sigma = diag(\sigma_0^2, \sigma_1^2, ...., \sigma_k^2)$) with large variances that to be considered as non-informative prior. In our case, we can common choices for the variances $\sigma_j^2$is 10000. Therefore,for the regression coefficients, we assumed non-informative independent priors to follow a multivariate normal distribution with mean zero vector and large variance, that means $\beta_j \sim N(0, 10^4)$ (Dijk, 2011; YILMAZ & ÇELİK, n.d.).

### 3.6.3 Posterior Distribution via Bayes Theorem

The Bayesian approach is first introduced by the Reverend Thomas Bayes. Today the Bayesian concept has gained popularity among many researchers across different fields as a result of its ability to handle complex models. Bayesian methods have also been embraced in other fields of science due to their ability to handle complexity in real-world problems. In addition, Bayesian inference has means of incorporating prior knowledge about the parameters under consideration since they influence the

posterior inference. In the case of Bayesian methods, parameters are assumed to follow a probability distribution while model parameters are considered as random variables. The main object of interest in Bayesian inference is the posterior distribution. Classical inference estimation depends solely on approximations as well as asymptotic results.

The posterior distribution is derived by multiplying the prior distribution over all parameters by the full likelihood function, it contains all the available knowledge about the parameters in the model. Gill (2002) details the discussion on the concept of the Bayesian analysis. A posterior distribution build prior and observed data, and the likelihood. That means,according to Baye's rule, we can express a posterior distribution, as presented by Bolstad & Curran (2016), as:

$$p(\beta|y) = \frac{L(\beta,y)}{L(y)} = \frac{L(y|\beta)p(\beta)}{\int_\beta L(\beta,y)p(\beta)d_\beta} \propto L(y|\beta)p(\beta) \tag{17}$$

where $L(y|\beta)$ is likelihood function, $p(\beta)$ is the prior distribution, and $p(\beta|y)$ is the posterior distribution. To obtain the posterior distribution we substitute the equation (3 and 16)in to equation (17),then it becomes:

$$p(\beta|y) \propto \prod_{i=1}^{n}\prod_{j=1}^{J} \pi_j(x_i)^{y_{ij}} \times \left[\frac{1}{2\pi|\Sigma_\beta|}\right]^{p/2} exp\left\{-\frac{1}{2}\left(\left(\beta-\mu_\beta\right)'\Sigma_\beta^{-1}\left(\beta-\mu_\beta\right)\right)\right\} \tag{18}$$

The posterior distribution has multivariate normal distribution.

- Given X,Y,$\Sigma$ and $\beta$ sample size $Z_{ij}$ for all i=1,2,...,n and j=1,2,.....,J-1, $Z_{ij} \sim N\left(d_{ij}, \tau_{ij}^2\right)$
- Given X,$\Sigma$ and Z

  $$\beta|X,Z,\Sigma \sim N(\mu_1, \bigwedge_1)$$

  Where, $\bigwedge_1 = \left(\bigwedge_0^{-1} + X_0'X_0\right)^{-1}$,$\mu_1 = \bigwedge_1\left(\bigwedge_0^{-1}\mu_0 + X_0'Z_0\right)$ N MVN
  $X_0 = C'X$, $Z_0 = C'Z$, $Z_i = X_i\beta + \varepsilon_i$ ,$\varepsilon \sim N(0,I)$ and $\mu_0$ and $\bigwedge_0$ are prior mean and covariance matrix of $X\beta$ and $\Sigma^{-1} = C'C$
- Given X,$\beta$ and Z sample $\Sigma$ such that $\Sigma|X,Z,\beta \sim W\left(v+n,(v_0+S)^{-1}\right)$ ,where $v_0$ is wishart distribution and $S = \varepsilon'\varepsilon$

The Wishart distribution is the sampling distribution of the matrix of sums of squares and products of normal distributional assumption.

## 3.7 Bayesian Estimation for Multinomial Logistic Regression

Estimation of $\beta$ on the posterior distribution may be difficult, for this reason we need to use non-analytic method. The most popular method of simulation technique is Markov Chain Monte Carlo (MCMC) methods. MCMC is a class of methods in which we can simulate draws that are slightly dependent and are approximately from a (posterior) distribution. The most commonly used MCMC techniques are Metropolis-Hasting and Gibbs sampler techniques. The latter part of the above expression being recognized as normal distributions for the $\beta$ parameters, then we can use the Metropolis-Hasting as implemented by R software to solve approximate the properties of the marginal posterior distributions for each parameter (Martin et al., 2022).

### 3.7.1 Markov Chain Monte Carlo Methods

Bayesian inference is solved by randomly drawing a very large sample from the posterior distribution. The idea of drawing a large sample from the posterior distribution is called Markov Chain Monte Carlo. Using MCMC techniques such as Gibbs sampling or the Metropolis–Hastings algorithm, we can directly sample sequences of values from the posterior distribution of interest, giving up the need for analytic solutions. MCMC methods have transformed Bayesian inference to a practical area of modern statistics (Craiu et al., 2022).

### 3.7.2 Metropolis-Hastings algorithm

Metropolis–Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution. The Metropolis–Hastings algorithm works by generating a sequence of sample values. In such a way that, as more and more sample values are produced, the distribution of values more closely approximates the desired distribution. In this thesis the posterior doesn't look like any distribution we know (no Conjugacy) and some (or all) of the full conditionals do not look like any distributions we know (no Gibbs sampling for those whose full conditionals we don't know). The metropolis-Hastings algorithm does not need availability of full conditionals. Rather, it generates a sequence of samples from a probability distribution by using the full joint density function and proposal distribution.

The basic MH algorithm can be described by the following steps:
Initialize $\theta_0$ Start b=0
Set B number of iterations:in this study we use 100,000

Iterate as follow

While b<B

do

Set $\theta = \theta^b$ select a component i

Propose new variable $\theta_i$ for component i from proposal distribution q($\theta_i|\theta^b$)

Set $\theta^{b+1}$

Accept i.e. Set $\theta_i^{b+1} = \theta_i$ with the probability $\alpha = min\left\{1, \frac{\pi(\theta)q(\theta_i|\theta^b)}{\pi(\theta^b)q(\theta_i|\theta^b)}\right\}$

Otherwise Set $\theta^{b+1} = \theta^b$

Set b = b + 1, end while

Once convergence is reached, all simulation values are from the target posterior distribution and a sufficient number will be drawn so that all areas of the posterior will be also explored.

### 3.7.3 Model Diagnostic

**Test of Convergence of the algorithm**

The empirical results from a given MCMC analysis are not viewed as reliable until the chain has reached its stationary distribution. To account this, the term convergence of an MCMC algorithm refers to whether the algorithm has reached its equilibrium (target) distribution. If this is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. Among several convergence assessment methods, basically, the most popular approaches used to determine convergence for Markov chains are discussed below.

**I. Autocorrelation:** High autocorrelation between the parameters of a chain tends to give slow convergence, whereas high autocorrelation within a single parameter chain leads to slow mixing and possibly individual non-convergence to the limiting distribution because the chain will tend to explore less space with much time. In analyzing Markov chain autocorrelation, it is helpful to identify lags in the series in order to calculate the longer run trends in correlation, and in particular whether they decrease with increasing lags (Papastamoulis & Ntzoufras, 2022).

**II.Time series plots or trace plots:** Iteration numbers on x-axis and parameter value on y-axis are commonly used to assess convergence(Papastamoulis & Ntzoufras, 2022). If the plot looks like a horizontal band, with no long upward or down ward trends, then we have evidence that the chain has converged. The posterior distribution is obtained by sampling toward the end of this longer iteration

sequence when the posterior distribution is stationary, as determined by an examination of trace plots of the iteration history of selected model quantities.

**III.Density plot:** This is another technique for identifying convergence and a classic sign of non-convergence is multimodality of the density estimate (Papastamoulis & Ntzoufras, 2022)

### 3.7.4 Software

The statistical software used in this study are SPSS version 20 and R version 4.2.1. SPSS is used for the descriptive and R software is used for the Bayesian multinomial logistics analysis. The Bayesian multinomial logistic regression analysis using the Metropolis–Hastings algorithm in the R software is considered. The Bayesian model uses normal in which the dependent variable is expected to follow a normal distribution with the prior of the coefficients normally distributed as non-informative priors. It is assumed that the regression parameters follow a normal distribution with a mean of zero vectors and large variance, which means $\beta_j \sim N(0, 10^4)$. Evaluation of Model Convergence of MCMC is checked by trace or time-series, density plot, and autocorrelation plot, with all plots showing that the convergence of the model is achieved. The Bayesian approach usually reports either the mean or median of the posterior samples for each parameter of concern as a point estimate.

# CHAPTER FOUR

# 4 Results and Discussion

The aim of this Chapter is to describe and conduct an analysis of the effect of major socio-economic, demographic, and medical factors on breast cancer, the target area at Jimma University Medical Center. In this section, the data introduced earlier is analyzed, and the results of the analysis based on descriptive analysis, multinomial logistic analysis, and Bayesian multinomial analysis techniques are presented. Recall that the aim of this study is to determine significant predictors of breast cancer using a multinomial logistic regression model. To achieve this, we set up a multinomial logit model using the Bayesian approach and the final outcome is to find predictors of individuals' risks of having breast cancer at Jimma University medical center. Various predictors, including medical factors and socio-demographic factors in the model, are intercepts, age group, marital status, stages, family history, treatments, HRT, and oral contraceptives. In order to fit the model to the data, the parameters of the model have to be estimated.

## 4.1 Descriptive Data Analysis

Descriptive analysis is a process of describing a given set of data via tables, graphs, or summary calculations in a meaningful way. It is a set of brief descriptive coefficients that summarizes a given dataset, which can either be a representation of the entire population or a sample. The measure used to describe the data set is frequency and percentages. We used descriptive statistics to get some information about the distribution of the variables. There are many variables considered in this study. The variables and the major characteristics of the variables are summarized in the following tables.

### 4.1.1 Summary descriptive results for the histologic type of breast cancer

Table 4.1: Summary of BC histologic type.

| histologic type | number | percentage(%) |
|---|---|---|
| ductal | 186 | 63.48 |
| lobular | 50 | 17.06 |
| others | 57 | 19.46 |

Of all 293 breast cancer patients, 186 (63.48%) of respondents were ductal carcinoma, which is the commonest histologic type of breast cancer, followed by 57(19.46%) were other (unknown) histologic types of carcinoma, and 50(17.06%) were lobular carcinoma is the least one of histologic type

was found in the Breast cancer patients (Table 4.1). In addition, we employed to describe predictor variables of major medical, demographic and socioeconomic factors presented as follows.

### 4.1.2 Summary descriptive results for Socio-Economic and Demographic Factors

Table 4.2 indicates that the socio-economic and demographic characteristics of breast cancer have been recorded. The results in the table indicate that, out of 293 BC patients, 136(46.42%) of the respondents are at age 20-34,75(25.93%) at age35-49, 48(16.39%) at age50-69, and 34(11.60%) of respondent are at age70+. Out of 293, about (218)74.40% of breast cancer patients are residing in a rural area and had a higher percentage compared with breast cancer women who lived in urban areas, 75(25.60%). Based on marital status, 200(68.27%) of women are married, 37(12.62%) are separated,and 56(19.11%) are single. However, this indicated that, from the total 293 cases of breast cancer in women, most of them were married, the followed are single, and the fewest are separated.

Table 4.2: Summary for Breast cancer histology type with Socio-Economic and Demographic Factors.

| | histologic type | | | | |
|---|---|---|---|---|---|
| Varibles | ductal(%) | lobular(%) | others(%) | Count | Percentage(%) |
| Age | | | | | |
| 20-34 | 81(27.65%) | 23(7.85%) | 32(10.92%) | 136 | 46.42 |
| 35-49 | 41(13.99%) | 15(5.12%) | 19(6.48%) | 75 | 25.59 |
| 50-69 | 40(13.65%) | 3(1.02%) | 5(1.72%) | 48 | 16.39 |
| 70+ | 24(8.18%) | 9(3.07%) | 1(0.34%) | 34 | 11.60 |
| Residence | | | | | |
| rural | 122(41.64%) | 46(15.70%) | 50(17.06%) | 218 | 74.40 |
| urban | 64(21.84%) | 4(1.37%) | 7(2.39%) | 75 | 25.60 |
| Marital Status | | | | | |
| married | 137(46.76%) | 27(9.21%) | 36(12.29%) | 200 | 68.27 |
| separated | 16(5.46%) | 11(3.75%) | 10(3.41%) | 37 | 12.62 |
| single | 33(11.26%) | 12(4.10%) | 11(3.75%) | 56 | 19.11 |

### 4.1.3 Summary descriptive results for Medical factors

Out of 293 BC patients, 35 (11.95%) are early stage, 87 (29.69%) are presented with advanced stage at the first hospital visit, 100(34.12%) are stage III and 71(24.24%) are stage IV. This indicates that the BC diagnosis for the women was highest at stage (III) and lowest at the early stage. The other one

is for patients diagnosed with breast cancer who have a family history of 16(5.46%) and who don't have a family history was 277(94.54%).

Table 4.3: Summary for Breast cancer histology type with medical factors.

| | histologic type | | | | |
|---|---|---|---|---|---|
| Varibles | ductal(%) | lobular(%) | others(%) | Count | Percentage(%) |
| Stages | | | | | |
| I | 20(6.83%) | 11(3.75%) | 4(1.37%) | 35 | 11.95 |
| II | 56(19.11%) | 15(5.12%) | 16(5.46%) | 87 | 29.69 |
| III | 63(21.50%) | 12(4.10%) | 25(8.52%) | 100 | 34.12 |
| IV | 47(16.04%) | 12(4.10%) | 12(4.10%) | 71 | 24.24 |
| family history | | | | | |
| No | 172(5.70%) | 49(16.73%) | 56(19.11%) | 277 | 94.54 |
| Yes | 14(4.78%) | 1(0.34%) | 1(0.34%) | 16 | 5.46 |
| treatments | | | | | |
| chemotherapy | 91(31.06%) | 23(7.85%) | 31(10.58%) | 145 | 49.49 |
| combination | 46(15.70%) | 1(0.34%) | 2(0.68%) | 49 | 16.72 |
| hormone | 21(7.17%) | 1(0.34%) | 1(0.34%) | 23 | 7.85 |
| radiotherapy | 11(3.75%) | 10(3.41%) | 2(0.68%) | 23 | 7.84 |
| surgery | 17(5.80%) | 15(5.12%) | 21(7.17%) | 53 | 18.10 |
| HRT | | | | | |
| continues | 29(9.90%) | 13(4.44%) | 21(7.17%) | 63 | 21.51 |
| cyclic | 130(44.37%) | 20(6.83%) | 22(7.51%) | 172 | 58.71 |
| estrogen | 20(6.83%) | 1(0.34%) | 2(0.68%) | 23 | 7.85 |
| local.est | 7(2.38%) | 16(5.45%) | 12(4.10%) | 35 | 11.93 |
| contraceptive | | | | | |
| not used | 155(52.90%) | 49(16.73%) | 55(18.76%) | 259 | 88.38 |
| used | 31(10.58%) | 1(0.34%) | 2(0.68%) | 34 | 11.62 |

From the total sample, the patients who have received the treatments of chemotherapy, radiotherapy, hormone, surgery, and the combination of two or more treatments are (145)49.15%,(23)8.19%, (23)7.85%,(53)14.33% and (49)20.48% respectively. This indicates the highest percentage of the treatments of the patients who received the commonest treatment is chemotherapy and the rarest treatment is hormone and radiotherapy. Again, also, the patients who received Hormone Replacement Therapy (HRT) are continuous, cyclic/sequential, estrogen-only, and local estrogen is 63(21.51%), 17(58.71%),23(7.85%), and 35(11.93%) respectively. However, this indicates that the HRT that the patient has received, cyclic/sequential, is the highest percentage. Regarding oral contraceptives, the patients who have used oral contraceptives are 34(14.67%) showing a smaller percentage than the patients who have not used oral contraceptives 259(81.33%)(Table 4.3).

## 4.2 Results of the Bayesian multinomial logit Model

A Bayesian model assumes parameters as random variables. The bayesian perspective of estimation allows us to combine the observed data with prior information. Multinomial logistic regression is a conditional probability model parameterized by a matrix of coefficients of the explanatory variable. Multinomial logistic regression is a generalization of binary logistic regression. For this study the response variable is trinomial. Therefore our parameters are a matrix of $2\times$p, where p is the number of explanatory variables. We have two linear combinations of explanatory variables for log odds, which are log odds of lobular carcinoma and the other is a linear combination for other types of carcinoma with respect to the reference category ductal carcinoma. As with any statistical model, we must avoid overfitting for multinomial logistic regression to make an accurate prediction of unobserved data. One approach to estimation is using the Bayesian method. In our case, we assign prior distribution for coefficients by assigning a high probability that most of the coefficients $\beta's$ will have a value at 0 or near 0. In the Bayesian approach, multinomial logistic regression has been fitted by using MCMC in R software(Bayesian Inference using Metropolis-Hastings algorithm) (Martin et al., 2022).

The posterior summary estimates by the MCMC algorithm (metropolis hasting algorithm), like posterior mean, standard deviation, Monte Carlo error, and credible interval were estimated using the R software of the MCMCmnl function. This function generates a sample from the posterior distribution of a multinomial logistic regression model using either Metropolis-Hastings sampling, random walk Metropolis sampling, or slice sampling. The user supplies data and priors, and a sample from the posterior distribution is returned as an mcmc object, which can be subsequently analyzed with functions provided in the MCMCpack/coda package. The Bayesian models were fitted using noninformative prior. The results for the Bayesian model with noninformative prior are given in (Table 4.4). The findings also provided an approximation of the standard deviation of the posterior distribution and computational accuracy of the mean. Furthermore, they show some selected 95% percentile credible intervals and MC errors of the posterior distribution. From the literature, it was established that noninformative prior should not have an effect on the posterior distribution. Approximately 2.5% and 97.5% percentiles of the posterior samples for each parameter provide a 95% posterior credible interval (interval within which the parameter lies with a probability of 0.95). We can report posterior summaries for the relative-risk ratios(RRR) instead of the regression coefficients. This is equivalent to applying an exponential transformation, $exp(\beta)$, to the simulated values of each of the regression coefficients, $\beta$, and then summarizing them.

Table 4.4: The summary result of Posterior Distribution for parameters from MCMC Bayesian multinomial logit Model

Posterior distribution of the model parameters under outcome category **lobular** Vs **ductal** carcinoma

| variables | mean($\hat{\beta}$) | sd | MCerror | 0.05*sd | 95%CrI for $exp(\hat{\beta})$ | | RRR |
|---|---|---|---|---|---|---|---|
| | | | | | **2.5**% | **97.5**% | |
| (intercept) | 2.084 | 0.936 | 0.009 | 0.047 | 1.120 | 51.676 | 8.037 |
| **Age** | | | | | | | |
| 20-34(ref) | | | | | | | |
| 35-49 | 0.285 | 0.614 | 0.006 | 0.031 | 0.405 | 4.104 | 1.330 |
| 50-69 | -2.342 | 0.858 | 0.009 | 0.043 | 0.018 | 0.485 | 0.096 |
| 70+ | 0.925 | 0.731 | 0.007 | 0.037 | 0.572 | 10.412 | 2.522 |
| **Residence** | | | | | | | |
| rural(ref) | | | | | | | |
| urban | -3.332 | 0.872 | 0.009 | 0.044 | 0.005 | 0.162 | 0.036 |
| **Marital status** | | | | | | | |
| Married(ref) | | | | | | | |
| separeted | 2.649 | 0.905 | 0.009 | 0.045 | 2.042 | 73.479 | 14.14 |
| single | 0.418 | 0.621 | 0.006 | 0.031 | 0.445 | 4.953 | 1.520 |
| **family history** | | | | | | | |
| No(ref) | | | | | | | |
| Yes | -1.088 | 1.426 | 0.014 | 0.071 | 0.018 | 5.212 | 0.337 |
| **Stage** | | | | | | | |
| I(ref) | | | | | | | |
| II | -2.323 | 0.808 | 0.008 | 0.040 | 0.021 | 0.522 | 0.098 |
| III | -2.739 | 0.881 | 0.009 | 0.044 | 0.013 | 0.334 | 0.065 |
| IV | -3.270 | 1.005 | 0.010 | 0.050 | 0.006 | 0.234 | 0.038 |
| **treatments** | | | | | | | |
| chemotherapy(ref) | | | | | | | |
| combination | -3.913 | 1.533 | 0.015 | 0.077 | 0.000 | 0.205 | 0.020 |
| hormone | -3.831 | 1.475 | 0.015 | 0.074 | 0.001 | 0.260 | 0.022 |
| radiotherapy | 2.070 | 0.766 | 0.008 | 0.038 | 1.896 | 36.162 | 7.924 |
| surgery | 1.903 | 0.625 | 0.006 | 0.031 | 1.887 | 24.532 | 6.706 |
| **HRT** | | | | | | | |
| continues(ref) | | | | | | | |
| cyclic | -1.607 | 0.544 | 0.005 | 0.027 | 0.069 | 0.562 | 0.200 |
| estrogen | -4.141 | 1.396 | 0.014 | 0.070 | 0.001 | 0.180 | 0.016 |
| local.estrogen | 1.836 | 0.905 | 0.009 | 0.045 | 1.283 | 37.826 | 6.271 |
| **contraceptive** | | | | | | | |
| notused(ref) | | | | | | | |
| used | -3.563 | 1.294 | 0.013 | 0.065 | 0.002 | 0.242 | 0.028 |

Posterior distribution of the model parameters under outcome category of **others(Unknown)** type of carcinoma Vs **ductal** carcinoma

| variables | mean($\hat{\beta}$) | sd | MCerror | 0.05*sd | 95%CrI for $exp(\hat{\beta})$ | | RRR |
|---|---|---|---|---|---|---|---|
| | | | | | **2.5**% | **97.5**% | |
| (intercept) | 1.380 | 0.894 | 0.009 | 0.045 | 0.701 | 20.739 | 3.975 |
| **Age** | | | | | | | |
| 20-34(ref) | | | | | | | |
| 35-49 | -0.214 | 0.527 | 0.005 | 0.026 | 0.295 | 2.401 | 0.807 |
| 50-69 | -1.906 | 0.714 | 0.007 | 0.036 | 0.030 | 0.568 | 0.149 |
| 70+ | -2.934 | 1.304 | 0.013 | 0.065 | 0.003 | 0.483 | 0.053 |
| **Residence** | | | | | | | |
| rural(ref) | | | | | | | |
| urban | -1.979 | 0.585 | 0.006 | 0.029 | 0.0444 | 0.432 | 0.138 |
| **Marital status** | | | | | | | |
| Married(ref) | | | | | | | |
| separeted | 1.652 | 0.736 | 0.007 | 0.037 | 1.140 | 20.045 | 5.217 |
| single | -0.214 | 0.571 | 0.006 | 0.029 | 0.244 | 2.232 | 0.807 |
| **family history** | | | | | | | |
| No(ref) | | | | | | | |
| Yes | -1.568 | 1.376 | 0.014 | 0.069 | 0.006 | 1.966 | 0.208 |
| **Stage** | | | | | | | |
| I(ref) | | | | | | | |
| II | -0.952 | 0.807 | 0.008 | 0.040 | 0.083 | 2.266 | 0.386 |
| III | -0.064 | 0.780 | 0.008 | 0.039 | 0.204 | 4.768 | 0.938 |
| IV | -0.889 | 0.816 | 0.008 | 0.041 | 0.085 | 2.396 | 0.411 |
| **treatments** | | | | | | | |
| chemotherapy(ref) | | | | | | | |
| combination | -2.807 | 1.021 | 0.010 | 0.051 | 0.007 | 0.352 | 0.060 |
| hormone | -3.843 | 1.306 | 0.013 | 0.065 | 0.001 | 0.227 | 0.021 |
| radiotherapy | -0.870 | 1.000 | 0.010 | 0.050 | 0.044 | 2.843 | 0.419 |
| surgery | 1.637 | 0.509 | 0.005 | 0.029 | 1.921 | 13.423 | 5.140 |
| **HRT** | | | | | | | |
| continues(ref) | | | | | | | |
| cyclic | -1.748 | 0.500 | 0.005 | 0.025 | 0.070 | 0.499 | 0.174 |
| estrogen | -2.288 | 0.937 | 0.009 | 0.047 | 0.015 | 0.555 | 0.101 |
| local.estrogen | 1.640 | 0.844 | 0.008 | 0.042 | 1.027 | 31.343 | 5.155 |
| **contraceptive** | | | | | | | |
| notused(ref) | | | | | | | |
| used | -2.747 | 1.071 | 0.011 | 0.054 | 0.004 | 0.457 | 0.064 |

**ref**=reference category,**CrI**=Credible interval,**RRR**=Relative Risk Ratio

The results of Table 4.4 indicate that the variables such as; age group50-69, place of residence area urban, and marital status of patients were separated, stages(II-IV), treatments that the patient's who have received radiotherapy, hormone, surgery, and the combination of two or more treatments, Hormone Replacement Therapy (HRT) that the patient's who received cyclic/sequential, estrogen-only and local estrogen, and Oral contraceptive used were the significant predictor of BC patients exposed to lobular carcinoma vs ductal carcinoma at 5% level of significance. However, the age group(age group50-69 and age group70+), place of residence area urban, and marital status of the patient's group was separated, treatments the patient who has received hormone, surgery, and the combination of two or more treatments, Hormone Replacement Therapy (HRT) that the patient's who received cyclic/sequential, estrogen-only and local estrogen, and Oral contraceptive used was the significant predictor of BC patients exposed to others type of carcinoma vs ductal carcinoma at 5% level of significance.

In Bayesian analysis, to have accurate posterior estimates, the simulation should be run until the Monte Carlo (MC) error for each parameter of interest is less than about 5% of the sample standard deviation(see table 4.4). The Bayesian multinomial logistic analysis procedure was used to make inferences about the parameters of a multinomial logistic model. In our model,2000 burn-in terms were discarded and the Metropolis hasting algorithm was implemented with 100,000 iterations. The researcher used a non-informative normal prior distribution with a mean of zero vector and large variance.

### 4.2.1 Interpretation of Bayesian multinomial logistic regression model

The findings of the Bayesian multinomial logistic regression model, by using non-informative prior, indicate that the patients' age,Residence area,Marital status,Oral contraceptives,treatments and HRT are a significant predictor for breast cancer patients with histologic types. Considering the effects of age on the risk of breast cancer patients with histologic type, findings show that the age group 50-69 years are (RRR= 0.096; 95%CrI(0.018, 0.485 )) less likely to have lobular carcinoma compared to 20-34 age group. Our results also indicate that place of residence area is another significant variable for the BC patients and the risk of breast cancer patients who reside in urban places are (RRR= 0.036; 95%CrI(0.005, 0.162 ))less likely to have lobular carcinoma compared to that the patients who live in rural. This means that women who live in urban areas have less risk of lobular carcinoma type of breast cancer as compared to women who live in rural areas. The risk of breast cancer for separated

women are (RRR=14.14;95%CrI(2.042, 73.479)) more likely to have lobular carcinoma compared to the patients who are married. Another finding of this study indicates that the stage of the disease is significantly associated with breast cancer. The risk of BC patients with stage II, stage III, and stage IV were (RRR= 0.098;95%CrI (0.021, 0.522)), (RRR= 0.065;95%CrI(0.013, 0.334)), and (RRR= 0.038;95%CrI(0.006, 0.234)) were less likely to have lobular carcinoma compared to that the patients with stage I respectively. This implies that Breast cancer patients with stageI are more likely to have lobular carcinoma compared to that patients with another type of stage. Contraceptives are also another variable that is considered in this study. The risk of breast cancer patients who use contraceptive methods are (RRR=0.028;95%CrI(0.002, 0.242)) less likely to have lobular carcinoma compared to the patients who do not use it and contraception is a significant contribution to breast cancer patients according to this study.

In addition, the risk of breast cancer patients who received treatment hormone and the combination of two or more are (RRR=0.022;95%CrI(0.001, 0.260)) and (RRR=0.020;95%CrI(0.000, 0.205)) respectively less likely to have lobular carcinoma compared to that the patients who received chemotherapy, and the patients who received treatment surgery and radiotherapy are (RRR= 6.706;95%CrI(1.887, 24.532)) and (RRR=7.924;95%CrI(1.896, 36.162)) respectively more likely to have lobular carcinoma compared to that the patients who received chemotherapy. From this, we conclude that the breast cancer patients who have received the hormone and the combination of two or more are less likely to have lobular carcinoma compared to the patients who received chemotherapy, and the patients who have received treatment for surgery and radiotherapy are more likely to have lobular carcinoma compared to the patients who received chemotherapy in Jimma University Medical Center. Besides, HRT(hormone replacement therapy) is also a highly significant factor for BC patients. The risk of breast cancer patients who received HRT of estrogen and cyclic/sequential were (RRR=0.016;95%CrI(0.001, 0.180)) and (RRR=0.200;95%CrI(0.069, 0.562)) respectively less likely to have lobular carcinoma compared to that the patients who received HRT continuously, and the patients who received HRT of local estrogen are (RRR=6.271;95%CrI(1.283, 37.826)) more likely to have lobular carcinoma compared to that the patients who received HRT continuously.

However, in Table 4.4, the findings of Posterior distribution of the model parameters under the outcome category of other histologic type of carcinoma shows that the age group 50-69 years and age group 70+ years are (RRR=0.149;95%CrI(0.030, 0.568)) and (RRR=0.053;95%CrI(0.003, 0.483)) respectively were less likely to have others histologic type of carcinoma compared to 20-34 age group.

Our results also indicate that place of residence is another significant variable for BC patients and the risk of breast cancer patients who reside in urban places are (RRR=0.138;95%CrI(0.0444, 0.432)) less likely to have others type of carcinoma compared to patients who live in rural. This means that women who live in urban areas have less risk of other types of carcinoma as compared to women who live in rural areas with breast cancer. The risk of breast cancer for the marital status of separated women are (RRR=5.217;95%CrI(1.140, 20.045)) more likely to have another histologic type of carcinoma compared to the patients who are marred. Another finding of this study indicates using oral contraceptives a patient uses or not using. The risk of breast cancer patients who use the oral contraceptive method are (RRR=0.064;95%CrI(0.004, 0.457)) less likely to have other types of carcinoma compared to the patients who do not use it. The other variables are interpreted similarly(Table 4.4).

## 4.3 Results of Bayesian Model Diagnostics

### 4.3.1 Assess the convergence of Markov chain Monte Carlo (MCMC)

There are a lot of commonly used methods to assess the convergence of MCMC output, but in this study, only some of them are used. In this part of the requirement for the Bayesian statistics, it is important to check for the convergence of the Markov chain Monte Carlo (MCMC). Before proceeding to examine the results of the model, it is essential to do some diagnostics to assess whether the Markov chain has converged to its stationary or posterior distribution. There are several different methods to check for convergence. These are Trace or Time series plots, Autocorrelation Plots, and Density plots. The algorithm converged after 100,000 iterations. In order to remove autocorrelation and the burn-in period, a lag of 500 is considered which requires iterations up to 100,000 iterations. The first 2,000 iterations are removed to cater to the burn-in period. The result of auto-correlation plots for some selected parameters is given below. The autocorrelations for all parameters become smaller after a lag equal to 500.

Figure 4.1 blow showed that the highlights for the auto-correlation plots diminished or reduced slowly; therefore, this provides evidence of the convergence of the Markov chain and suggests that it may be appropriate to average Markov chain output. Figure 4.2 blow showed the highlights of the posterior density plot and trace plot, which provides a graphical representation of the posterior density estimate for each parameter. So, the density plot shows that the posterior distribution of each parameter is almost approximated to normally distributed and the trace plot also looks like a horizontal band and there are no long upward or downward trends. For all simulated parameters, trace

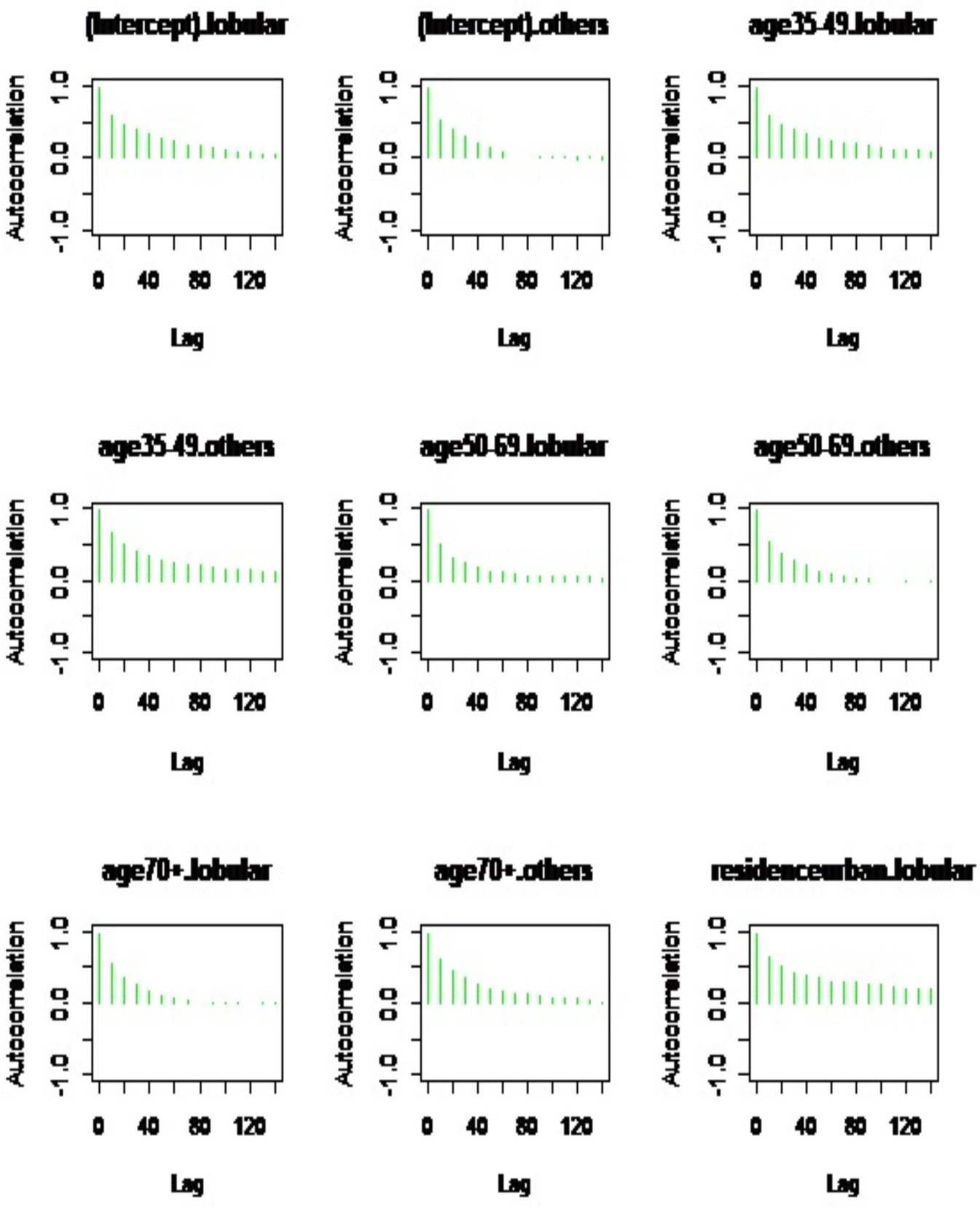



Figure 4.1: Posterior probability Auto-correlation plots for the Histologic types of Breast Cancer Patients.

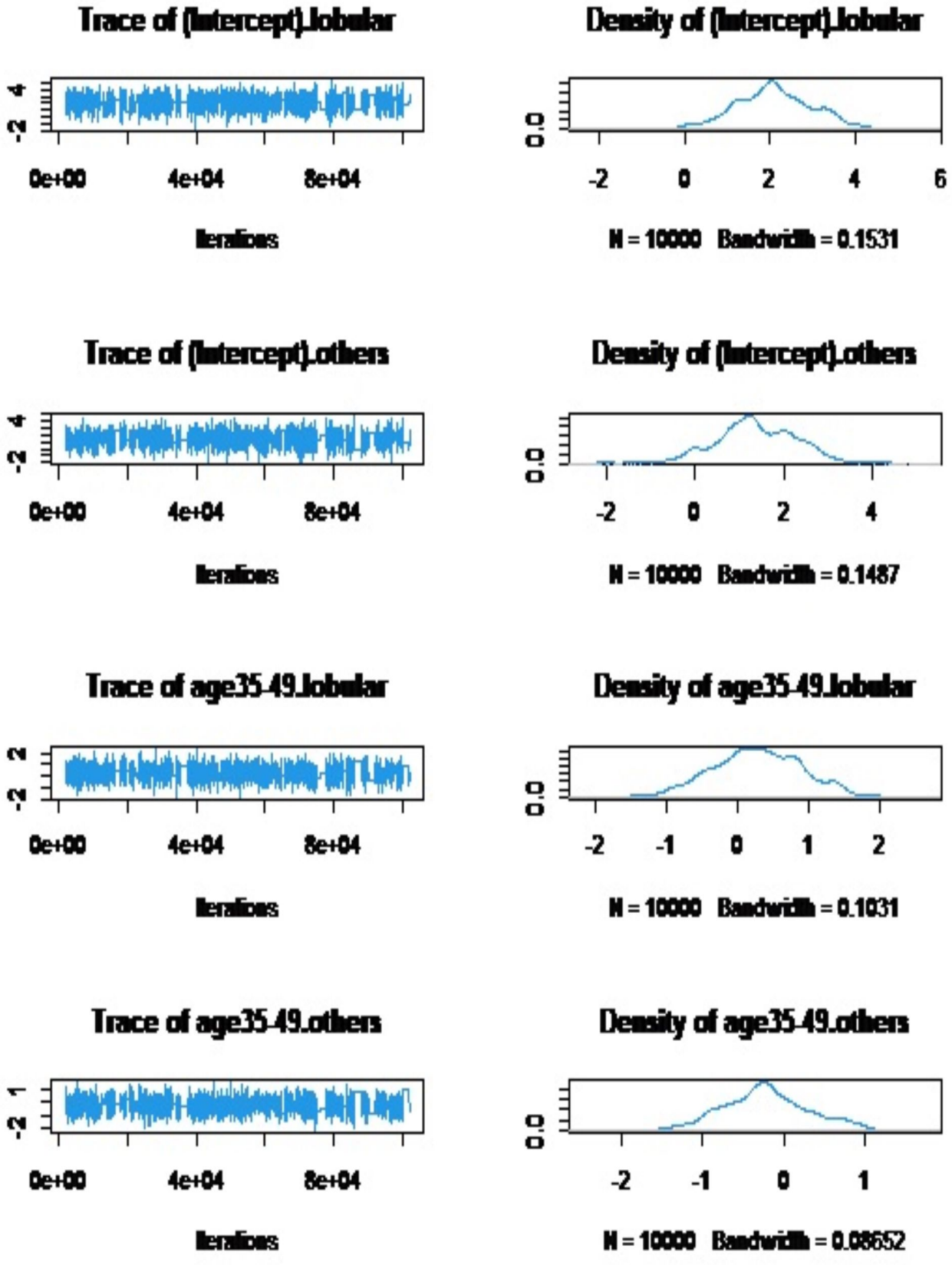



Figure 4.2: Posterior probability density and trace plots for the Histologic types of Breast Cancer Patients.

and density plots indicate a good convergence. The following plot in the figure 4.2 shows the simulations in trace and density plots for intercept, age35-49, age50-69, age70+, and urban residence areas for lobular carcinoma. The trace and density plots shows that all of the coefficients are almost converged. Therefore, Markov chain convergence was almost performed. The other remaining predictors of characteristics, we can see APPENDEX C( 5.2).

### 4.3.2 Assessing Accuracy of the Bayesian estimation

In addition to the above graphical methods of checking the convergence of the chain to its posterior distribution, the Monte Carlo standard error of the posterior mean (which is an estimate of the difference between the estimate of the posterior mean for each parameter and the true posterior mean) is another way to assess the accuracy of Bayesian logistic regression. Here we can use a Monte Carlo error for each parameter. If the MC error value is less than 5% of its posterior standard error, then the posterior density is estimated with accuracy. Table 4.4 contains the estimated coefficients: mean, standard deviation (sd), Monte Carlo (MC) errors, 5% of the standard deviation (0.05*sd), and 95%credible Interval. Findings indicates that, the MC errors for all parameters were less than 5% of its posterior standard error.

From the findings of significant predictors, we can interpret the MC error by comparing it with 5% of its posterior standard error. The study showed that the age group 50-69 and age group 70+ were risk factors for BC patients. The MC error for breast cancer patients at aged50-69 who have breast cancer type of lobular carcinoma is 0.009 and this is less than 0.05*sd = 0.047, and the MC error for breast cancer patients at aged50-69 who have the other histologic type of carcinoma is 0.007, it is also less than 0.05*sd=0.036. And, MC error for breast cancer patients aged 70+ who have the other histologic type of carcinoma is 0.013. It is also less than 0.05*sd=0.065.

Besides, the MC error for breast cancer patients residing in urban areas who have lobular carcinoma is 0.009 and this is less than 0.05*sd = 0.044, and the MC error for breast cancer patients residing in urban areas who have the other histologic type of carcinoma is 0.006, it is also less than 0.05*sd=0.029, etc. This implies that the MC error for each significant predictor variable is less than 5% of its posterior standard deviation and hence, this approach suggests that the convergence and accuracy of the posterior estimates are achieved. It indicates that convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics.

## 4.4 Discussion

This studies attempted to identify some socioeconomic, demographic, and medical factors of breast cancer patients at Jimma University Medical Center. The main objective of this study is to determine the significant predictors of breast cancer using a multinomial logistic regression model by applying Bayesian analysis to estimate parameters. The study included eight predictor variables that were categorized under socioeconomic, demographic, and medical factors. From the result of this study, we observe that age group50-69 and age group70+, residence area urban, marital status being separated, treatments that patients who have received(radiotherapy, hormone, surgery, and combination), Stage(II-IV), HRT(Hormone replacement therapy) that patients who have been received were cyclic/sequential,estrogen-only and local estrogen, and oral contraceptive used were the significant predictors for Breast cancer patients.

The descriptive results of the study indicated that, out of 293 causes,186 (63.48%) of respondents have the ductal carcinoma type of breast cancer, which indicates the commonest histologic type of breast cancer. The following 57(19.46%) causes were other (unknown) histologic types of carcinoma, and 50(17.06%) causes of lobular carcinoma is the least one histologic type found in breast cancer patients. This study is similar to the finding byOgunsakin & Lougue (2019) in Nigeria. In this study and Ogunsakin & Lougue (2019)study, the highest percentage of breast cancer histologic type is ductal carcinoma.

The findings indicated that the patient's residence area is a highly significant effect on breast cancer. This implies that the risk of breast cancer patients who resided in urban areas are less likely to have lobular carcinoma compared to that of the patients who live in rural areas and the risk of breast cancer patients who resided in urban areas are less likely to have other histologic types of carcinoma compared to areas who live in rural. This study supports previous literature. In the United States, rural health disparity has been a recent focus of attention and has been made a priority for improvement (Subrahmanian et al., 2018). While urban health is an emerging disciplineKrefis et al. (2018) ,many standardized definitions of urban and rural areas exist and are used by social scientists and demographers. They are found in sources unfamiliar to health researchers and have largely not been used in public health studies and Bacha et al. (2021) indicated that the breast cancer patients who lived in rural areas had lower survival times than those who lived in urban areas. Another finding showed that age is an important determinant of breast cancer patients. Several studies have established that breast cancer risk increases with an increase in age. This study is also supported by the previous study

(Ogunsakin & Lougue, 2019; Sharma & Abebe, 2019). From the results, the stages of breast cancer have a significant effect on women. Moreover, the significance of the variables in this study is supported by the finding of different researchers( Allemani et al. (2018); Sharma & Abebe (2019); Pham (2014)) also showed that the stages of breast cancer at diagnosis significantly affected the survival time of women with breast cancer patients. The results of this study also showed that Contraceptives is also another significant variable on breast cancer patients. The study done by study( (Tolessa et al., 2021; Bacha et al., 2021; Sharma & Abebe, 2019)) showed that the survival time of women with breast cancer patients significantly affected by oral contraceptives and the expected survival time of women using oral contraceptives was less as compared to women did not use oral contraceptives. This study is supported by the previous study. Besides, marital status was another significant factor for breast cancer patients. In this, we observed that separated women are an important determinant for breast cancer and had more significant effects as compared to others.

Furthermore,the findings in this study showed that the treatments that a patient's have received are significant factors in breast cancer patients. In this study, the risk of breast cancer patients who received the treatment hormone and the combination of two or more are less likely to have lobular carcinoma compared to the patients who received chemotherapy, and the patients who received the treatment surgery and radiotherapy are more likely to have lobular carcinoma compared to the patients who received chemotherapy. Similarly, the risk of breast cancer patients who have received hormone treatment, radiotherapy, and a combination of two or more are less likely to have another histologic type of carcinoma compared to the patients who received chemotherapy, and the patients who have received treatment or surgery are more likely to have other histologic carcinoma compared to at the patients who received chemotherapy in Jimma University Medical Center. This finding is in line with earlier studies by (Allemani et al., 2018; Bacha et al., 2021; Sharma & Abebe, 2019).

Moreover, the Bayesian multinomial logistic regression analysis also revealed that all independent variables are statistically significant. Using Bayesian simulation MC error for each significant predictor was found to be less than 5% of its posterior standard error. This implies that convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics.

# CHAPTER FIVE

# 5 CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Conclusions

In order to meet the objectives of the study, the technique of multinomial logistic regression can be used to model categorical response variables using MCMC algorithms by applying Bayesian approaches. In this study, the highest percentage of breast cancer women are exposed to cancer disease due to ductal carcinoma. Based on this study, the results showed that the age of the patients, marital status, residence area, treatments, oral contraceptives, and related factors are a significant effect on breast cancer patients.

The result of MCMC methods for the Bayesian approach showed that all parameters in the model are significant factors for breast cancer patients under the response category of lobular carcinoma with reference to ductal carcinoma except age group35-49 and age group70+, marital status being single, and family history. Besides, there are also all parameters in the model that are significant factors for the breast cancer patients under the response category for other histologic types of carcinoma with reference to ductal carcinoma except age35-49, marital status being single, stage(II-IV), the treatment that a patient's who has been received of radiotherapy, and family history.

From the findings of posterior estimation, the imaged exploration of; autocorrelation, trace, and density plots shows that the Markov chain is converged and lastly reached a stable situation and corresponding equilibrium distribution. Similarly, the MC error is less than 5% of the posterior standard deviation for a parameter in both categories of the response variable. Therefore, we have concluded that the sample from the posterior distribution of the parameters of the model is valid.

Finally, by using Bayesian multinomial logistic regression analysis, we suggest that the convergence and accuracy of the posterior estimates are achieved. It indicates that convergence and accuracy of posterior estimates are attained and the model is appropriate to estimate posterior statistics.

## 5.2 Recommendations

Depending on the above foremost findings, the researcher recommended the following points for researchers, Jimma University Medical Center and mainly, Oncology Departments interested in any sub-work of this study.

• All factors, except family history, in this study, were significant factors for breast cancer patients. Thus, Jimma University Medical Center and mainly, the Oncology Departments should have to focus on minimizing Breast cancer cases with a particular focus on members that have a high severity of the disease.

•This study is limited to a few variables recorded at the Jimma University Medical Center and mainly, Oncology Departments. In that case, researchers are recommended to include clinical diagnostic-related variables, and the medicines take responsibility for registering biased free and cleared variables in the patient books, to carry out quality healthcare services on healthcare workers, healthcare facilities; medicines, devices, and other technologies; information systems; and financing.

•We recommend that governments, non-governmental organizations, and other sectors involved in policy-making put in place policies, strategies, and excitation that target on those risk factors for women to enhance their assistance of breast cancer screening in health facilities, so as to access appropriate management health assessment as well as providing financially supported treatments for breast cancer patients.

• Finally, further researchers should check the significant predictors of breast cancer by Bayesian approaches on multinomial logistic regression analyses of both men and women breast cancer patients.

**Limitation of the study**

The data is retrospective, obtained from the Jimma Medical Center, Oncology Department and the information gathered by this survey doesn't have full information. Some women were especially eligible under some covariates, therefore there is no full information about breast cancer patients. There are many risk factors affecting breast cancer patients that are studied by different researchers, but in Jimma Medical Center, Oncology Department, some important factors that may affect breast cancer patients are not gathered. For instance, the main expected predators of breast cancer patients' likes; education level,late or no pregnancy, late menopause, etc, were not included.

# References


Agboola, A., Musa, A., Wanangwa, N., Abdel-Fatah, T., Nolan, C., Ayoade, B., et al. (2012). Molecular characteristics and prognostic features of breast cancer in nigerian compared with uk women. *Breast cancer research and treatment*, *135*(2), 555–569.

Agresti, A. (2016). Some issues in generalized linear modeling. In *International workshop on matrices and statistics* (pp. 75–88).

Al-Azri, M., Al-Awisi, H., Al-Rasbi, S., El-Shafie, K., Al-Hinai, M., Al-Habsi, H., et al. (2014). Psychosocial impact of breast cancer diagnosis among omani women. *Oman medical journal*, *29*(6), 437.

Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., et al. (2018). Global surveillance of trends in cancer survival 2000–14 (concord-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet*, *391*(10125), 1023–1075.

Anderson, B. O., Ilbawi, A. M., & El Saghir, N. S. (2015). Breast cancer in low and middle income countries (lmic s): a shifting tide in global health. *The breast journal*, *21*(1), 111–118.

Arifin, H., Chou, K.-R., Ibrahim, K., Fitri, S. U. R., Pradipta, R. O., Rias, Y. A., et al. (2022). Analysis of modifiable, non-modifiable, and physiological risk factors of non-communicable diseases in indonesia: Evidence from the 2018 indonesian basic health research. *Journal of Multidisciplinary Healthcare*, *15*, 2203–2221.

Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., et al. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, *66*, 15–23.

Arzanova, E., & Mayrovitz, H. N. (2022). The epidemiology of breast cancer. *Exon Publications*.

Azubuike, S. O., Muirhead, C., Hayes, L., & McNally, R. (2018). Rising global burden of breast cancer: the case of sub-saharan africa (with emphasis on nigeria) and implications for regional development: a review. *World journal of surgical oncology*, *16*(1), 1–13.

Bacha, R. H., Jabir, Y. N., Asebot, A. G., & Liga, A. D. (2021). Risk factors affecting survival time of breast cancer patients: The case of southwest ethiopia. *Journal of Research in Health Sciences*, *21*(4), e00532.

Bayaga, A. (2010). Multinomial logistic regression: Usage and application in risk analysis. *Journal of applied quantitative methods*, *5*(2).

Beral, V., Peto, R., Pirie, K., & Reeves, G. (2019). Menopausal hormone therapy and 20-year breast cancer mortality. *The lancet*, *394*(10204), 1139.

Biganzoli, L., Battisti, N. M. L., Wildiers, H., McCartney, A., Colloca, G., Kunkler, I. H., et al. (2021). Updated recommendations regarding the management of older patients with breast cancer: a joint paper from the european society of breast cancer specialists (eusoma) and the international society of geriatric oncology (siog). *The Lancet Oncology*, *22*(7), e327–e340.

Bolstad, W. M., & Curran, J. M. (2016). *Introduction to bayesian statistics*. John Wiley & Sons.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *68*(6), 394–424.

Chi, A. C., Day, T. A., & Neville, B. W. (2015). Oral cavity and oropharyngeal squamous cell carcinoma—an update. *CA: a cancer journal for clinicians*, *65*(5), 401–421.

Chlebowski, R. T., Anderson, G. L., Aragaki, A. K., Manson, J. E., Stefanick, M. L., Pan, K., et al. (2020). Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up of the women's health initiative randomized clinical trials. *Jama*, *324*(4), 369–380.

Clegg-Lamptey, J.-N. (2017). Epidemiology of breast cancer in africa. *Accra-Ghana: School of Medicine and Dentistry, University of Ghana*.

Craiu, R. V., Gustafson, P., & Rosenthal, J. S. (2022). Reflections on bayesian inference and markov chain monte carlo.

Derks, M. G., Bastiaannet, E., Kiderlen, M., Hilling, D. E., Boelens, P. G., Walsh, P. M., et al. (2018). Variation in treatment and survival of older patients with non-metastatic breast cancer in five european countries: a population-based cohort study from the eurecca breast cancer group. *British journal of cancer*, *119*(1), 121–129.

Dijk, H. van. (2011). *The oxford handbook of bayesian econometrics*. Oxford University Press.

Ebenezer, O. R., & Lougue, S. (2019). Bayesian generalized linear mixed modeling of breast cancer. *Iranian Journal of Public Health*, *48*(6), 1043.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: An overview. *International journal of cancer*, *149*(4), 778–789.

Forman, D., Martel, C. de, Lacey, C. J., Soerjomataram, I., Lortet-Tieulent, J., Bruni, L., et al. (2012). Global burden of human papillomavirus and related diseases. *Vaccine*, *30*, F12–F23.

Frühwirth-Schnatter, S., & Frühwirth, R. (2012). Bayesian inference in the multinomial logit model. *Austrian Journal of Statistics*, *41*(1), 27–43.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., et al. (2022). Breast cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(6), 524–541.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. Chapman and Hall/CRC.

Grywalska, E., Sobstyl, M., Putowski, L., & Roliński, J. (2019). Current possibilities of gynecologic cancer treatment with the use of immune checkpoint inhibitors. *International Journal of Molecular Sciences*, *20*(19), 4705.

Haileselassie, W., Mulugeta, T., Tigeneh, W., Kaba, M., & Labisso, W. L. (2019). The situation of cancer treatment in ethiopia: challenges and opportunities. *Journal of cancer prevention*, *24*(1), 33.

Hamdi, Y., Abdeljaoued-Tej, I., Zatchi, A. A., Abdelhak, S., Boubaker, S., Brown, J. S., et al. (2021). Cancer in africa: the untold story. *Frontiers in oncology*, *11*, 650117.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., Oga, E., Fabowale, T., Igbinoba, F., et al. (2012). Cancer incidence in nigeria: a report from population-based cancer registries. *Cancer epidemiology*, *36*(5), e271–e278.

Kathrikolly, T. R., Shetty, R. S., & Nair, S. (2020). Opportunities and barriers to breast cancer screening in a rural community in coastal karnataka, india: a qualitative analysis. *Asian Pacific Journal of Cancer Prevention: APJCP*, *21*(9), 2569.

Kim, C. A., Spratlin, J. L., Armstrong, D. E., Ghosh, S., & Mulder, K. E. (2014). Efficacy and safety of single agent or combination adjuvant chemotherapy in elderly patients with colon cancer: a canadian cancer institute experience. *Clinical colorectal cancer*, *13*(3), 199–206.

Krefis, A. C., Augustin, M., Schlünzen, K. H., Oßenbrügge, J., & Augustin, J. (2018). How does the urban environment affect health and well-being? a systematic review. *Urban Science*, *2*(1), 21.

Liu, L.-Y., Wang, F., Cui, S.-D., Tian, F.-G., Fan, Z.-M., Geng, C.-Z., et al. (2017). A case-control study on risk factors of breast cancer in han chinese women. *Oncotarget*, *8*(57), 97217.

Martin, A. D., Quinn, K. M., Park, J. H., & Park, M. J. H. (2022). *Package 'mcmcpack'*. Univ. of California Berkeley, CA.

McCormack, V., McKenzie, F., Foerster, M., Zietsman, A., Galukande, M., Adisa, C., et al. (2020). Breast cancer survival and survival gap apportionment in sub-saharan africa (abc-do): a prospective cohort study. *The Lancet Global health*, *8*(9), e1203–e1212.

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.

Memirie, S. T., Habtemariam, M. K., Asefa, M., Deressa, B. T., Abayneh, G., Tsegaye, B., et al. (2018). Estimates of cancer incidence in ethiopia in 2015 using population-based registry data. *Journal of global oncology*, *4*, 1–11.

Mieog, J. S. D., Kruijf, E. M. de, Bastiaannet, E., Kuppen, P. J., Sajet, A., Craen, A. J. de, et al. (2012). Age determines the prognostic role of the cancer stem cell marker aldehyde dehydrogenase-1 in breast cancer. *BMC cancer*, *12*(1), 1–9.

Møller, H., Henson, K., Lüchtenborg, M., Broggio, J., Charman, J., Coupland, V. H., et al. (2016). Short-term breast cancer survival in relation to ethnicity, stage, grade and receptor status: national cohort study in england. *British journal of cancer*, *115*(11), 1408–1415.

Murugesan, R., Nobes, A., & Wild, J. (2017). A mooc approach for training researchers in developing countries. *Open Praxis*, *9*(1), 45–57.

Mutebi, M., Anderson, B. O., Duggan, C., Adebamowo, C., Agarwal, G., Ali, Z., et al. (2020). Breast cancer treatment: A phased approach to implementation. *Cancer*, *126*, 2365–2378.

Ogunsakin, R., & Lougue, S. (2019). Socio economic determinants of breast cancer revalenc e in southwestern, nigeria. *Asian Journal of Epidemiology*, *12*, 9–16.

Papastamoulis, P., & Ntzoufras, I. (2022). On the identifiability of bayesian factor analytic models. *Statistics and Computing*, *32*(2), 1–29.

Parkin, D. M., Ferlay, J., Curado, M.-P., Bray, F., Edwards, B., Shin, H.-R., et al. (2010). Fifty years of cancer incidence: Ci5 i–ix. *International journal of cancer*, *127*(12), 2918–2927.

Pham, M. H. (2014). Survival analysis-breast cancer. *Undergraduate Journal of Mathematical Modeling: One+ Two*, *6*(1), 4.

Rebner, M., & Pai, V. R. (2020). Breast cancer screening recommendations: African american women are at a disadvantage. *Journal of Breast Imaging*, *2*(5), 416–421.

Román, M., Graff-Iversen, S., Weiderpass, E., Vangen, S., Sakshaug, S., Hofvind, S., et al. (2016). Postmenopausal hormone therapy and breast cancer prognostic characteristics: A linkage between nationwide registrieshormone therapy use and breast cancer characteristics. *Cancer Epidemiology, Biomarkers & Prevention*, *25*(11), 1464–1473.

Sepandi, M., Akrami, M., Tabatabaee, H., Rajaeefard, A., Tahmasebi, S., Angali, K. A., et al. (2014). Breast cancer risk factors in women participating in a breast screening program: a study on 11,850 iranian females. *Asian Pacific Journal of Cancer Prevention*, *15*(19), 8499–8502.

Sharma, M., & Abebe, E. (2019). Determinants of survival time of women with breast cancer. *Archives of Oncology and Cancer Therapy ISSN*, 2638–5074.

Shimizu, H., Ross, R., Bernstein, L., Yatani, R., Henderson, B., & Mack, T. (1991). Cancers of the prostate and breast among japanese and white immigrants in los angeles county. *British journal of cancer*, *63*(6), 963–966.

Siegel, R. L., Miller, K. D., Fuchs, H. E., Jemal, A., et al. (2021). Cancer statistics, 2021. *Ca Cancer J Clin*, *71*(1), 7–33.

Stewart, S. B., & Boorjian, S. A. (2015). Radical prostatectomy in high-risk and locally advanced prostate cancer: Mayo clinic perspective. In *Urologic oncology: Seminars and original investigations* (Vol. 33, pp. 235–244).

Subrahmanian, K., Swamy, P., et al. (2018). *Global child health: A toolkit to address health disparities*. Springer.

Sun, Y.-S., Zhao, Z., Yang, Z.-N., Xu, F., Lu, H.-J., Zhu, Z.-Y., et al. (2017). Risk factors and preventions of breast cancer. *International journal of biological sciences*, *13*(11), 1387.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209–249.

Taib, N. A., Yip, C. H., & Low, W. Y. (2014). A grounded explanation of why women present with advanced breast cancer. *World journal of surgery*, *38*(7), 1676–1684.

Tan, M.-M., Ho, W.-K., Yoon, S.-Y., Mariapun, S., Hasan, S. N., Lee, D. S.-C., et al. (2018). A case-control study of breast cancer risk factors in 7,663 women in malaysia. *PloS one*, *13*(9), e0203469.

Timotewos, G., Solomon, A., Mathewos, A., Addissie, A., Bogale, S., Wondemagegnehu, T., et al. (2018). First data from a population based cancer registry in ethiopia. *Cancer epidemiology*, *53*, 93–98.

Tolessa, L., Sendo, E. G., Dinegde, N. G., & Desalew, A. (2021). Risk factors associated with breast cancer among women in addis ababa, ethiopia: Unmatched case–control study. *International Journal of Women's Health*, *13*, 101.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, *65*(2), 87–108.

Urban, M., Banks, E., Egger, S., Canfell, K., O'connell, D., & Beral, V. (2013). Injectable and oral contraceptive use and cancers of the breast, cervix, ovary, and endometrium in black south african women.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*(5), 1413–1432.

Vineis, P., & Wild, C. P. (2014). Global cancer patterns: causes and prevention. *The Lancet*, *383*(9916), 549–557.

YILMAZ, A., & ÇELİK, H. eray. (n.d.). A bayesian approach to binary logistic regression model with application to oecd data. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *26*(2), 94–101.

Youlden, D. R., Cramb, S. M., Yip, C. H., & Baade, P. D. (2014). Incidence and mortality of female breast cancer in the asia-pacific region. *Cancer biology & medicine*, *11*(2), 101.

# Appendix A:Bayesian Multinomial logit Equations and Estimated Response Probability

• Model parameters equation under outcomes category lobular carcinoma versus ductal carcinoma.

$$Y_1 = log\left[\frac{pr(histology=\textbf{lobular})}{pr(histology=\textbf{ductal})}\right] = 2.084 + 0.285(age35-49) - 2.342(age50-69)$$
$$+ 0.925(age70^{+}) - 3.332 residenceurban + 2.649 Marital.Sseparated + 0.418 Marital.Ssingle$$
$$- 2.323 stageII - 2.739 stageIII - 3.270 stageIV - 1.088 fam.histYes - 3.913 treatcomb - 3.831 treathormone$$
$$+ 2.070 treatradio.th + 1.903 treatsurgery - 1.607 HRTcyclic - 4.141 HRTestrogen$$
$$+ 1.836 HRTlocal.est - 3.563 contracptused \quad (19)$$

• Model parameters equation under outcomes category of others type of carcinoma versus ductal carcinoma.

$$Y_2 = log\left[\frac{pr(histology=\textbf{others})}{pr(histology=\textbf{ductal})}\right] = 1.380 - 0.214(age35-49) - 1.906(age50-69)$$
$$- 2.934(age70^{+}) - 2.979 residenceurban + 1.652 Marital.Sseparated - 0.214 Marital.Ssingle$$
$$- 0.952 stageII - 0.064 stageIII - 0.889 stageIV - 1.568 fam.histYes - 2.807 treatcomb - 3.843 treathormone$$
$$- 0.870 treatradio.th + 1.637 treatsurgery - 1.748 HRTcyclic - 2.288 HRTestrogen$$
$$+ 1.640 HRTlocal.est - 2.747 contracptused \quad (20)$$

• By using equations 18 and 19,the estimated response probabilities appeared that the case number 293 has probability as follow:

$$Pr(y = lobular) = \pi_1 = \frac{exp(Y_1)}{1 + exp(Y_1) + exp(Y_2)} \quad (21)$$

$$Pr(y = others) = \pi_2 = \frac{exp(Y_2)}{1 + exp(Y_1) + exp(Y_2)} \quad (22)$$

$$Pr(y = ductal) = \pi_3 = \frac{1}{1 + exp(Y_1) + exp(Y_2)} \quad (23)$$

According to the definition,

$$\pi_1 + \pi_2 + \pi_3 = 1 \quad (24)$$

This implies,

$$\frac{exp(Y_1)}{1 + exp(Y_1) + exp(Y_2)} + \frac{exp(Y_2)}{1 + exp(Y_1) + exp(Y_2)} + \frac{1}{1 + exp(Y_1) + exp(Y_2)} = 1 \quad (25)$$

# Appendix B:The summary results for Classical multinomial logistic regression

Table 5.1: Summary statistics of the classical multinomial model excluding reference category.

Parameter Summaries of classical multinomial model under outcomes category of lobular carcinoma

| Variables | coef | st.error | \|z\| | p-value | 95%CI of B | exp(B) |
|---|---|---|---|---|---|---|
| (Intercept) | 1.79 | 0.89 | 2.08 | 0.04 | (0.08, 3.51) | 6.36 |
| Age | | | | | | |
| 20-34(ref) | | | | | | |
| 35-49 | 0.23 | 0.58 | 0.40 | 0.69 | (-0.90,1.36) | 1.26 |
| 50-69 | -2.04 | 0.85 | -2.44 | 0.02 | (-3.69,-0.40) | 0.13 |
| 70+ | 0.76 | 0.69 | 1.14 | 0.26 | (-0.57, 2.09) | 2.19 |
| Residence | | | | | | |
| rural(ref) | | | | | | |
| urban | -2.96 | 0.77 | -3.83 | 0.00 | (-4.44 ,-1.48) | 0.05 |
| Marital Status | | | | | | |
| married(ref) | | | | | | |
| separated | 2.41 | 0.77 | 3.04 | 0.00 | (0.92, 3.90) | 10.73 |
| single | 0.41 | 0.61 | 0.60 | 0.55 | (-0.78 ,1.60) | 1.44 |
| family.hist | | | | | | |
| No(ref) | | | | | | |
| Yes | -1.10 | 1.18 | -0.93 | 0.35 | (-3.40,1.21) | 0.33 |
| Stages | | | | | | |
| I(ref) | | | | | | |
| II | -2.06 | 0.76 | -2.74 | 0.01 | (-3.54, -0.59) | 0.13 |
| III | -2.44 | 0.78 | -3.11 | 0.00 | (-3.94, -0.94) | 0.09 |
| IV | -2.89 | 0.87 | -3.34 | 0.00 | (-4.56 ,-1.21) | 0.06 |
| treatments | | | | | | |
| chemo.trp(ref) | | | | | | |
| combination | -2.96 | 1.16 | -2.56 | 0.01 | (-5.22,-0.70) | 0.05 |
| hormone | -3.06 | 1.22 | -2.46 | 0.01 | (-5.42, -0.69) | 0.05 |
| radiotherapy | 1.85 | 0.76 | 2.45 | 0.01 | (0.39, 3.31) | 6.44 |
| surgery | 1.68 | 0.57 | 3.04 | 0.00 | (0.58, 2.79) | 5.63 |
| HRT | | | | | | |
| continues(ref) | | | | | | |
| cyclic | -1.41 | 0.55 | -2.65 | 0.01 | (-2.47 ,-0.35) | 0.24 |
| estrogen | -3.27 | 1.24 | -2.69 | 0.01 | (-5.68, -0.86) | 0.04 |
| local.estrogen | 1.63 | 0.78 | 2.05 | 0.04 | (0.14, 3.12) | 4.91 |
| contracept | | | | | | |
| notused(ref) | | | | | | |
| used | -2.91 | 1.21 | -2.42 | 0.02 | (-5.28, -0.54) | 0.05 |

Parameter Summaries of classical multinomial model under outcomes category others type of carcinoma

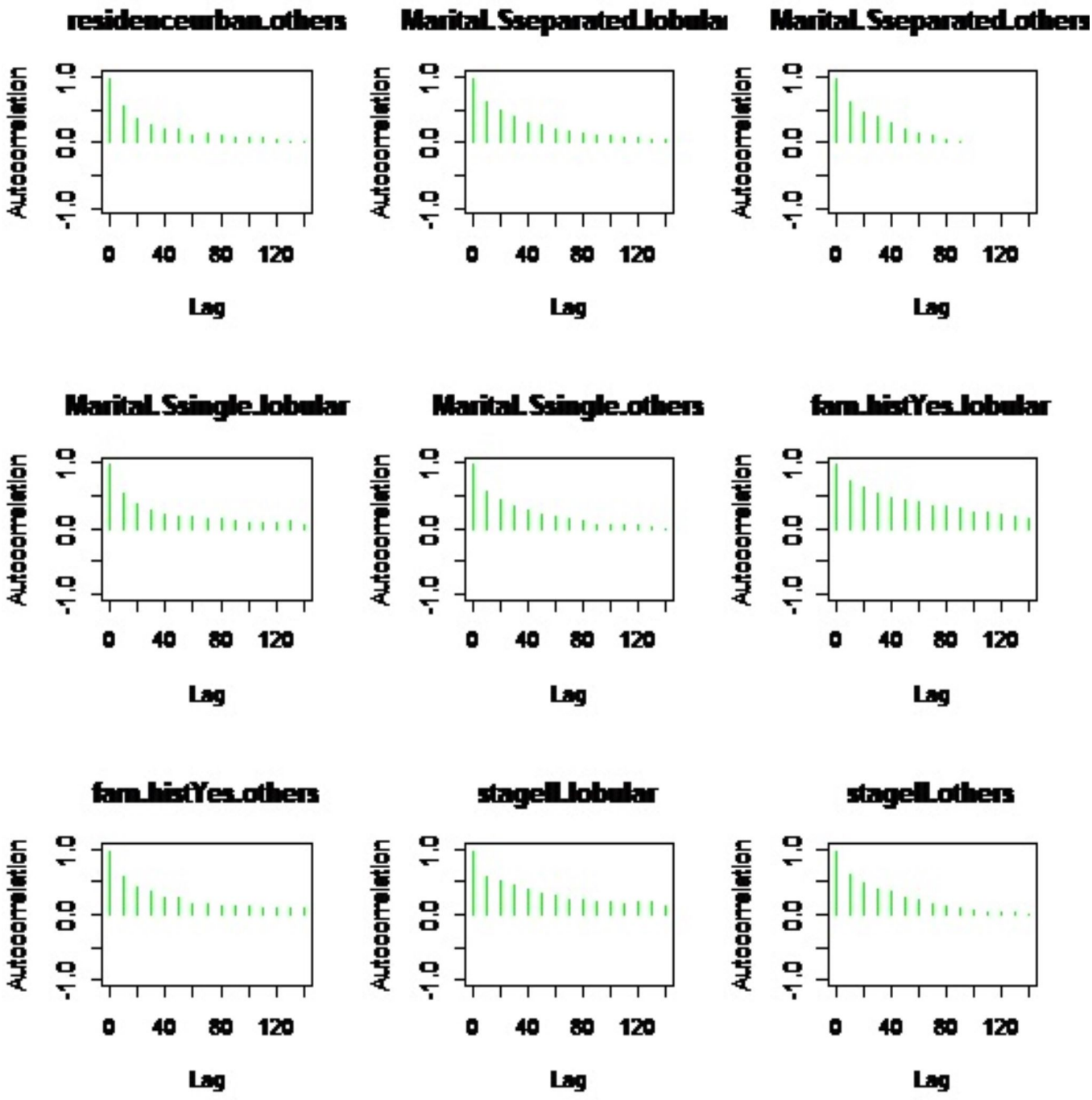
| Variables | coef | st.error | \|z\| | p-value | 95%CI of B | exp(B) |
|---|---|---|---|---|---|---|
| (Intercept) | 1.23 | 0.87 | 1.45 | 0.15 | (-0.45,2.97) | 3.54 |
| Age | | | | | | |
| 20-34(ref) | | | | | | |
| 35-49 | -0.22 | 0.49 | -0.51 | 0.61 | (-1.20,0.70) | 0.78 |
| 50-69 | -1.68 | 0.70 | -2.54 | 0.01 | (-3.15,-0.40) | 0.17 |
| 70+ | -2.26 | 1.131 | -2.01 | 0.04 | (-4.49,-0.06) | 0.10 |
| Residence | | | | | | |
| rural(ref) | | | | | | |
| urban | -1.83 | 0.56 | -3.25 | 0.00 | (-2.91,-0.72) | 0.16 |
| Marital Status | | | | | | |
| married(ref) | | | | | | |
| separated | 1.53 | 0.67 | 2.28 | 0.02 | (0.22,2.84) | 4.61 |
| single | -0.20 | 0.56 | -0.35 | 0.73 | (-1.30,0.904) | 0.82 |
| family.hist | | | | | | |
| No(ref) | | | | | | |
| Yes | -1.10 | 1.18 | -0.93 | 0.35 | (-3.40,1.21) | 0.33 |
| Stages | | | | | | |
| I(ref) | | | | | | |
| II | -0.86 | 0.78 | -1.11 | 0.27 | (-2.39,0.66) | 0.42 |
| III | -0.06 | 0.75 | -0.08 | 0.94 | (-1.53,1.42) | 0.94 |
| IV | -0.80 | 0.93 | -0.98 | 0.33 | (-2.39,0.80) | 0.45 |
| treatments | | | | | | |
| chemo.trp(ref) | | | | | | |
| combination | -2.40 | 0.85 | -2.81 | 0.00 | (-4.07,-0.73) | 0.09 |
| hormone | -3.21 | 1.17 | -2.72 | 0.01 | (-5.52,-0.90) | 0.04 |
| radiotherapy | -0.72 | 0.81 | -0.78 | 0.44 | (-2.55,1.10) | 0.49 |
| surgery | 1.44 | 0.51 | 2.81 | 0.00 | (0.44,2.44) | 4.21 |
| HRT | | | | | | |
| continues(ref) | | | | | | |
| cyclic | -1.61 | 0.49 | -3.31 | 0.00 | (-2.56,-0.65) | 0.14 |
| estrogen | -1.94 | 0.92 | -2.12 | 0.03 | (-3.74,-0.15) | 0.20 |
| local estrogen | 1.34 | 0.73 | 1.85 | 0.07 | (-0.08,2.77) | 3.83 |
| contracept | | | | | | |
| notused(ref) | | | | | | |
| used | -2.27 | 0.92 | -2.48 | 0.01 | (-4.07,-0.48) | 0.10 |

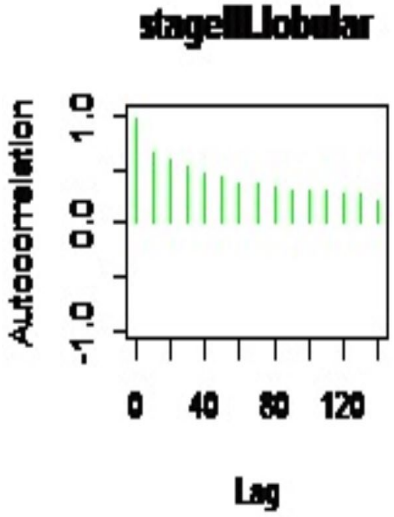
Table 5.2: Model Fitting Information with predictor variables.

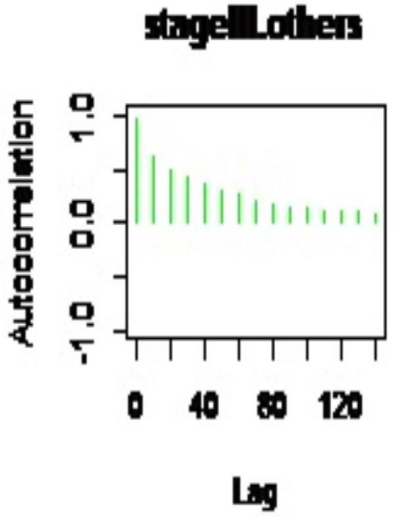
| model | Resid.df | Resid.Dev | AIC | BIC | Chi-square | df | sig |
|---|---|---|---|---|---|---|---|
| intercept | 584 | 532.493 | 536.493 | 543.854 | | | |
| full | 548 | 326.368 | 403.095 | 542.941 | 205.4 | 36 | 0.000 |

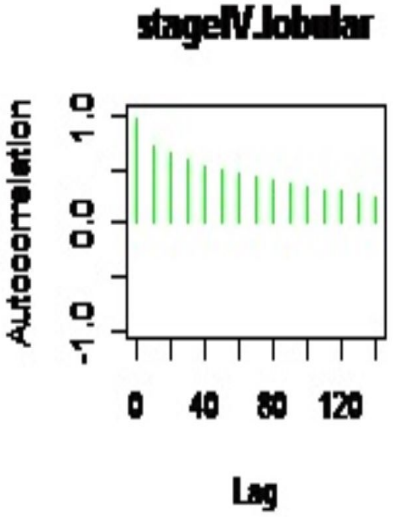
# Appendix C:Statistical Figures for Bayesian analysis

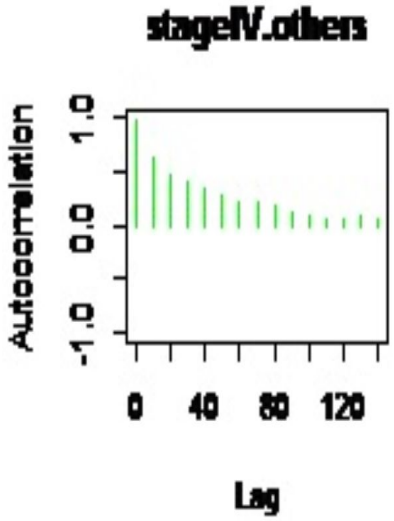
**I. Posterior probability Auto-correlation plots for the Histologic types of Breast Cancer Patients.**

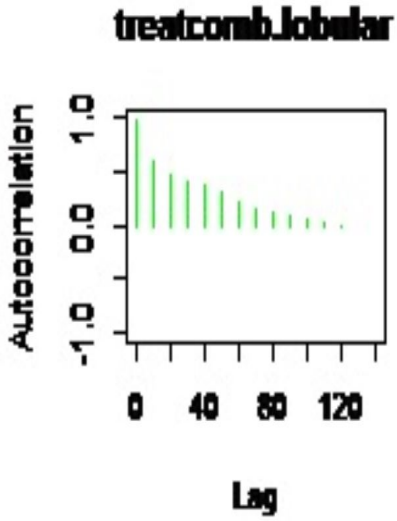


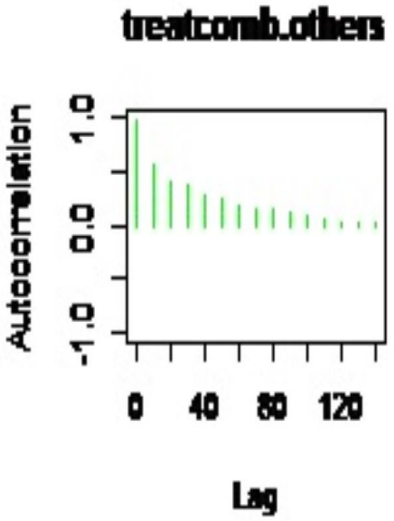

stageIII.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

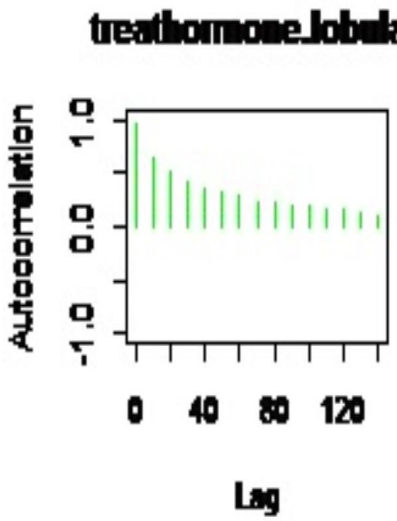

stageIII.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

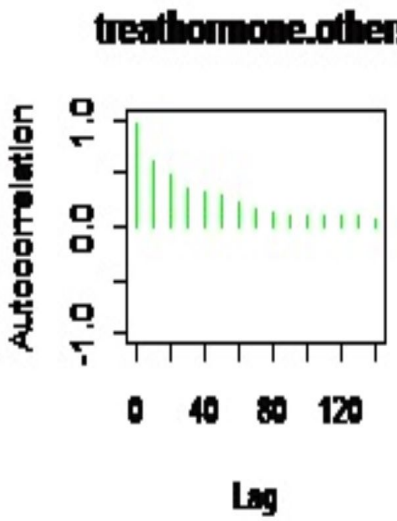

stageIV.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

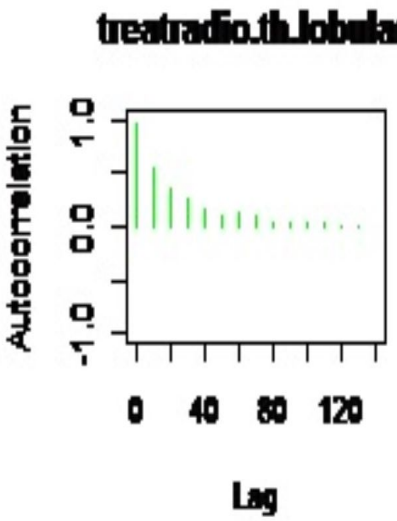

stageIV.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

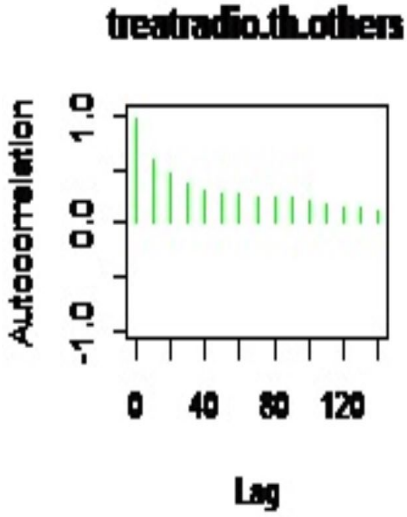

treatcomb.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

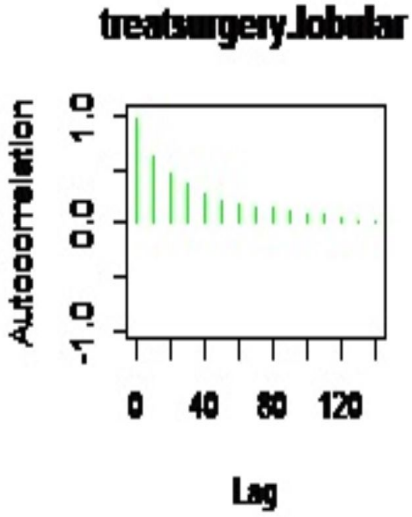

treatcomb.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

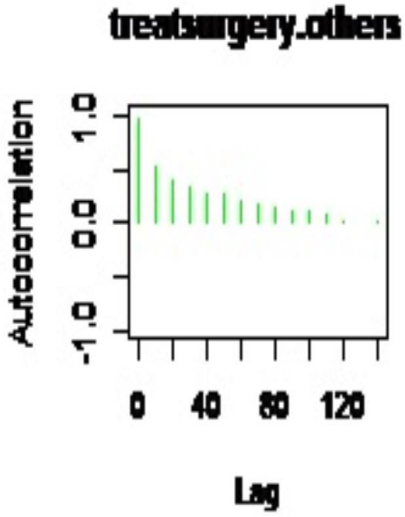

treathormone.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

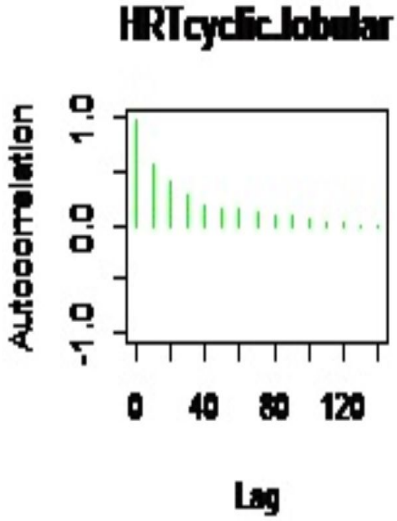

treathormone.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

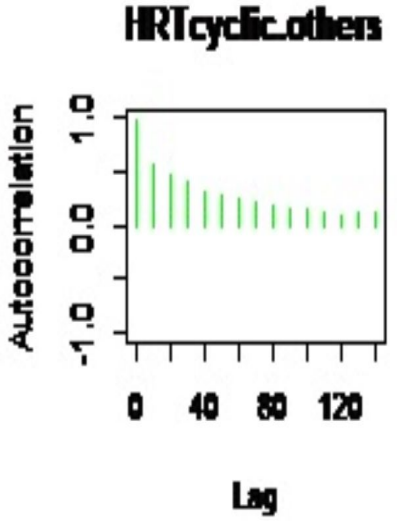

treatradio.th.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

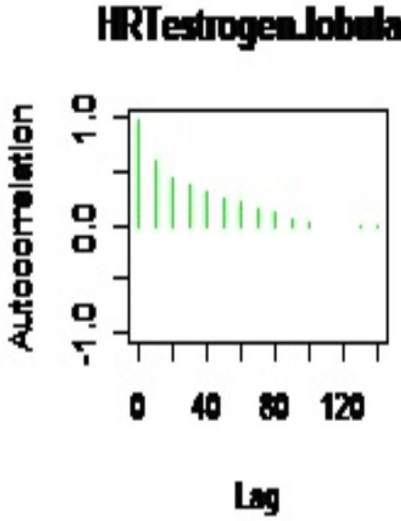

treatradio.th.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

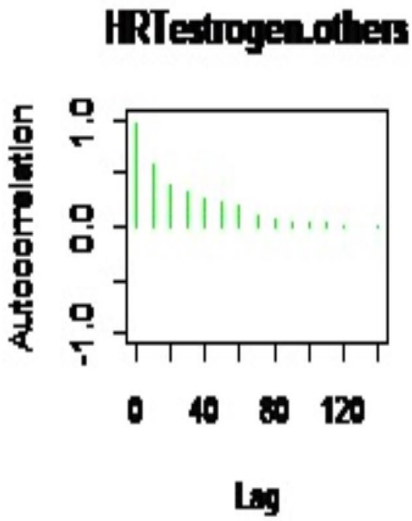

treatsurgery.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

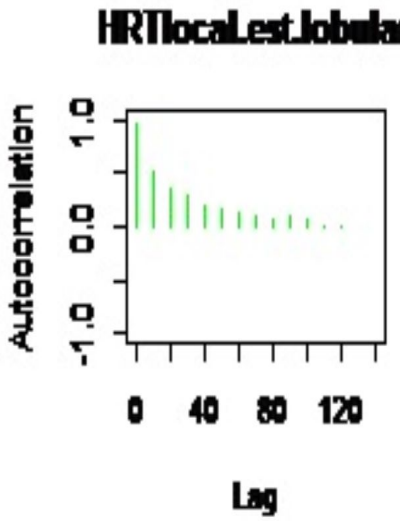

treatsurgery.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

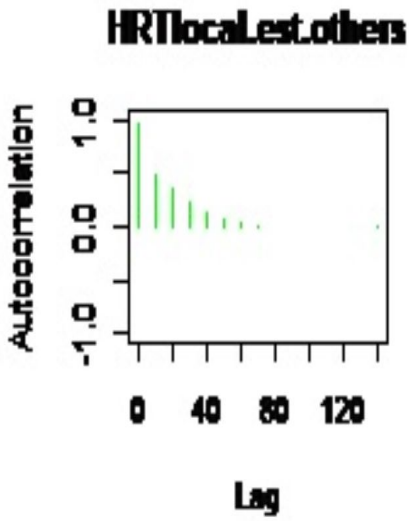

HRTcyclic.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag


HRTcyclic.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

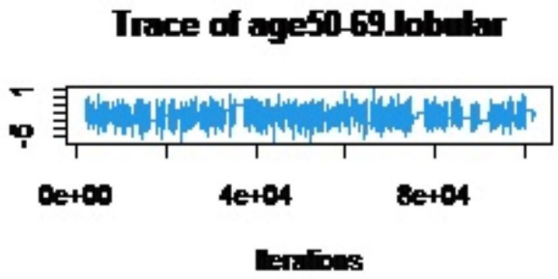

HRTestrogen.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

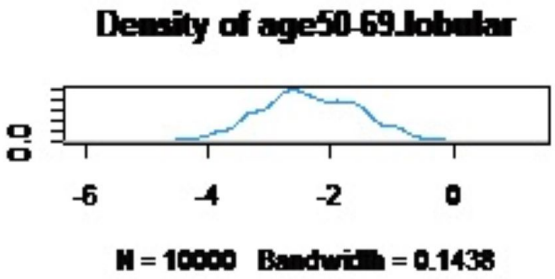

HRTestrogen.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

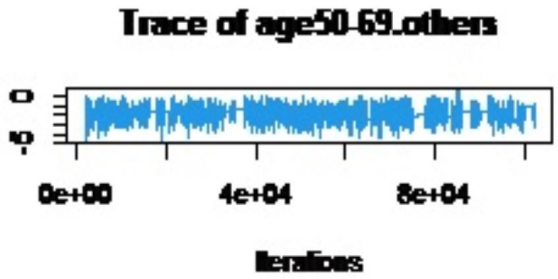

HRTlocalest.lobular
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

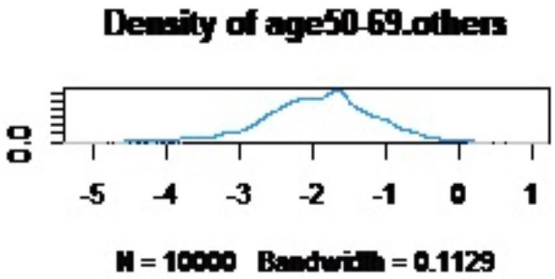

HRTlocalest.others
Autocorrelation
-1.0
0.0
1.0
0
40
80
120
Lag

**II. Posterior probability density and trace plots for the Histologic types of Breast Cancer Patients.**

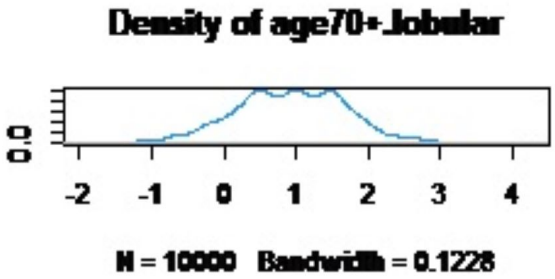


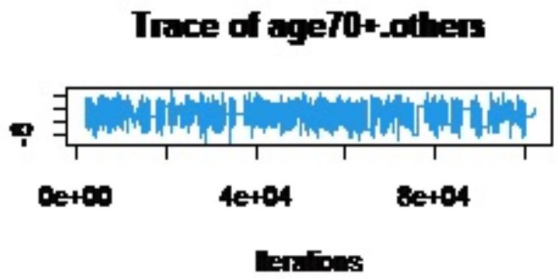



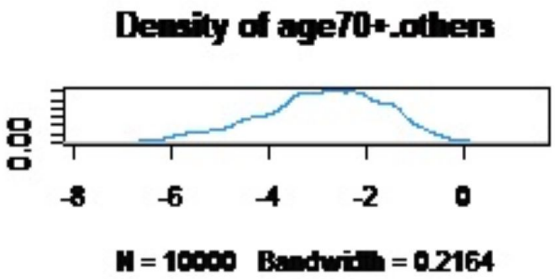



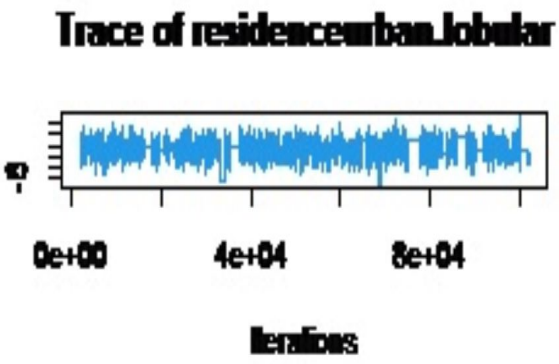



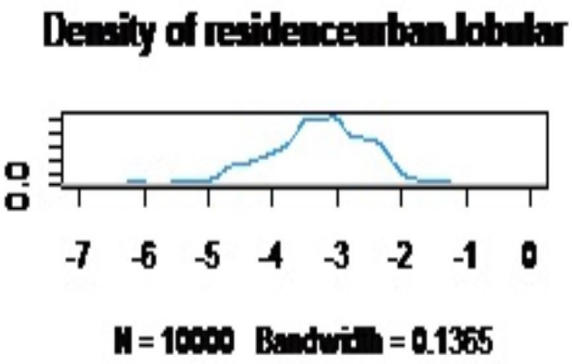


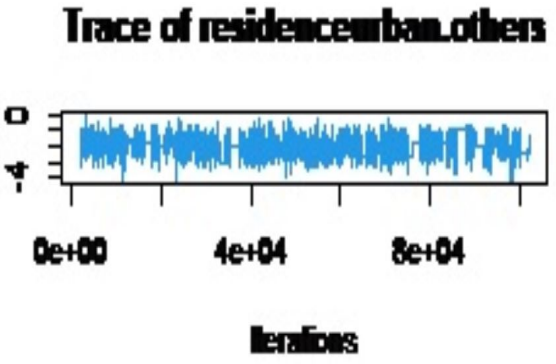


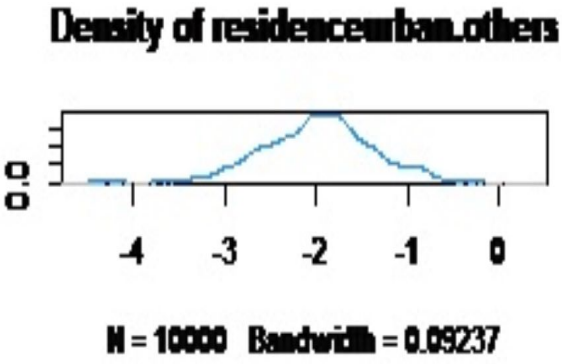



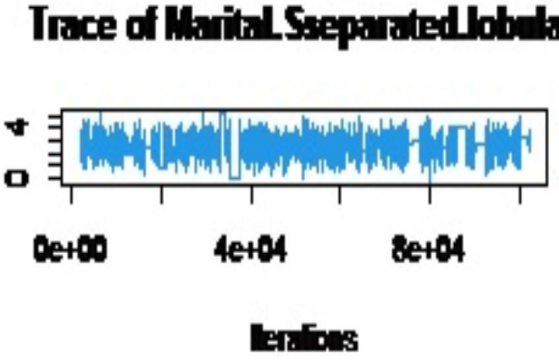



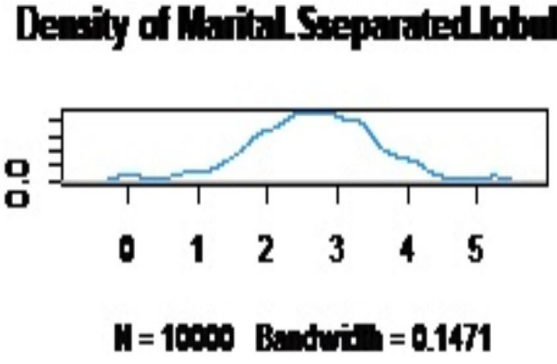

Trace of residenceurban.lobular
-6
0e+00
4e+04
8e+04
Iterations

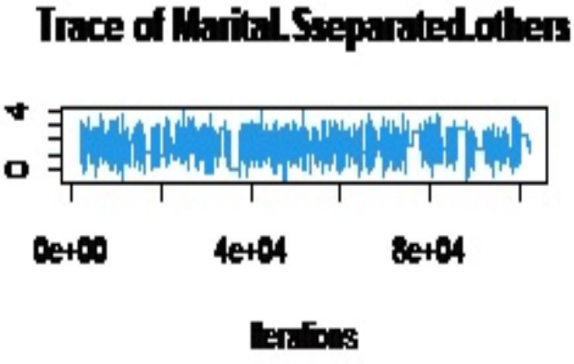

Density of residenceurban.lobular
0.0
-7
-6
-5
-4
-3
-2
-1
0
N = 10000 Bandwidth = 0.1365

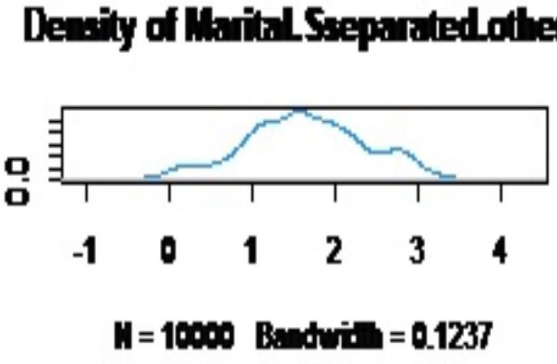

Trace of residenceurban.others
0
-4
0e+00
4e+04
8e+04
Iterations

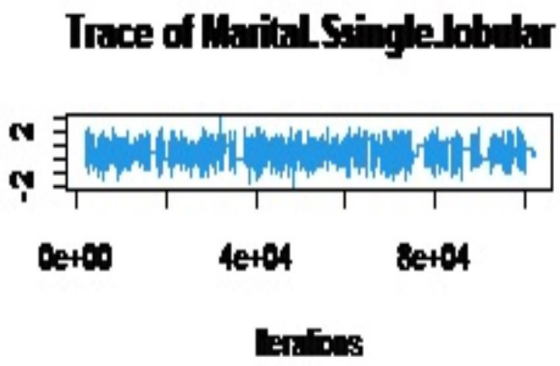

Density of residenceurban.others
0.0
-4
-3
-2
-1
0
N = 10000 Bandwidth = 0.09237

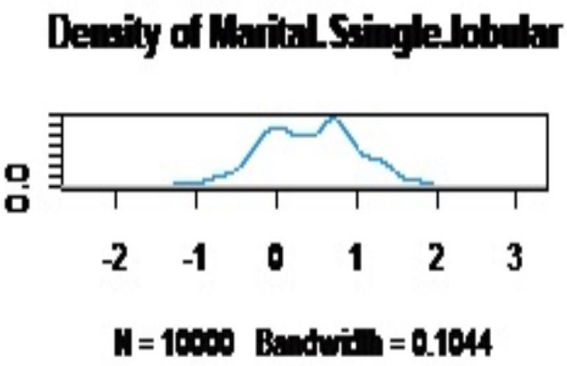

Trace of Marital.Sseparated.lobular
0
4
0e+00
4e+04
8e+04
Iterations

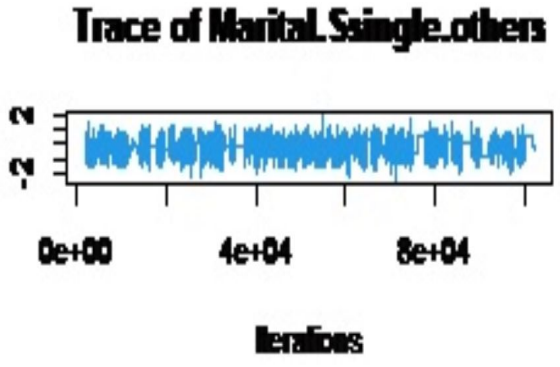

Density of Marital.Sseparated.lobular
0.0
0
1
2
3
4
5
N = 10000 Bandwidth = 0.1471

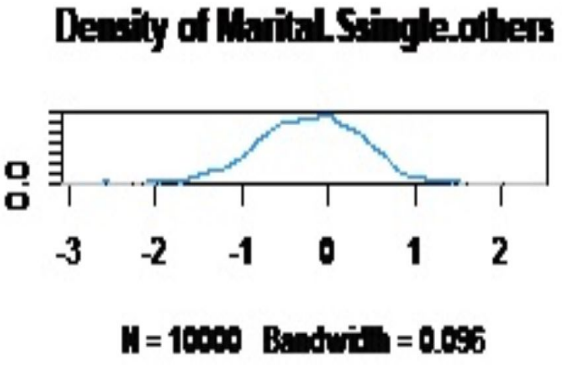

Trace of Marital.Sseparated.others
0
4
0e+00
4e+04
8e+04
Iterations

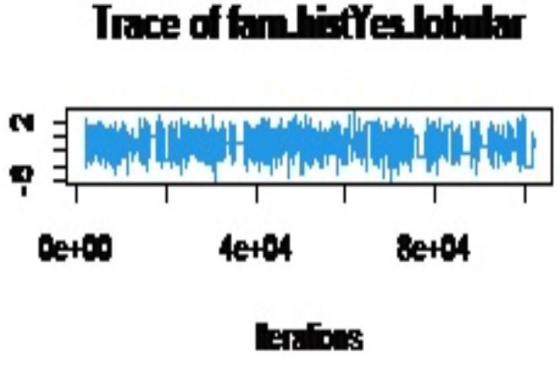

Density of Marital.Sseparated.others
0.0
-1
0
1
2
3
4
N = 10000 Bandwidth = 0.1237

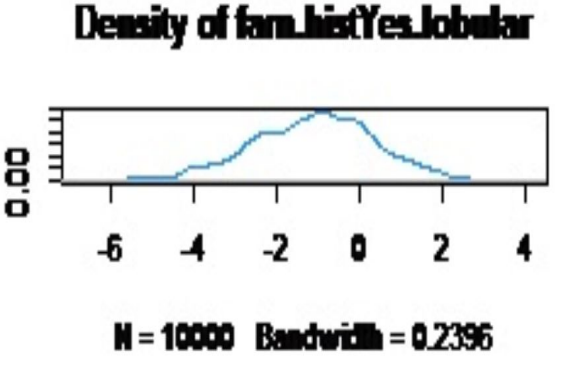

Trace of Marital.Ssingle.Jobular
-2
2
0e+00
4e+04
8e+04
Iterations

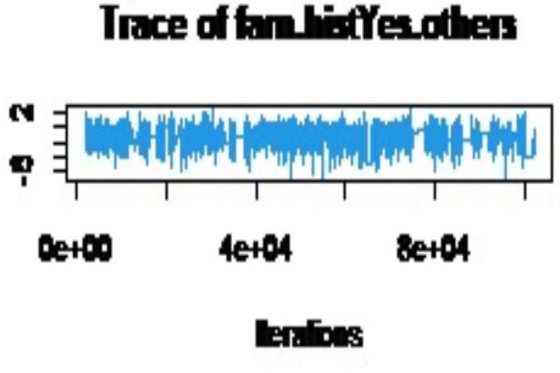

Density of Marital.Ssingle.Jobular
0.0
-2
-1
0
1
2
3
N = 10000 Bandwidth = 0.1044

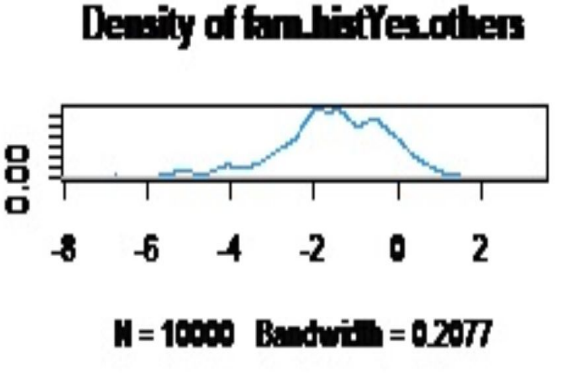

Trace of Marital.Ssingle.others
-2
2
0e+00
4e+04
8e+04
Iterations

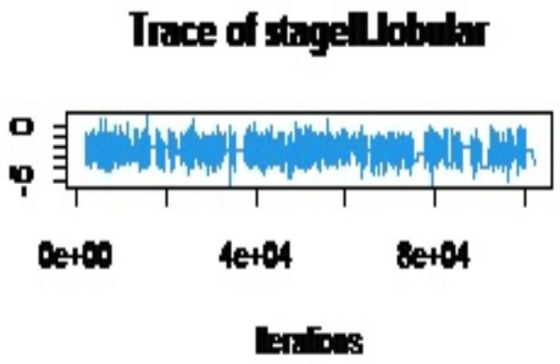

Density of Marital.Ssingle.others
0.0
-3
-2
-1
0
1
2
N = 10000 Bandwidth = 0.096

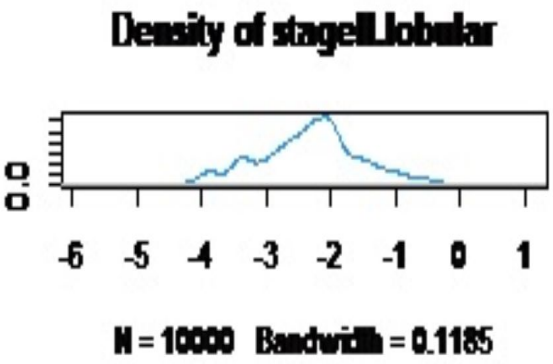

Trace of famJistYes.Jobular
-6
2
0e+00
4e+04
8e+04
Iterations

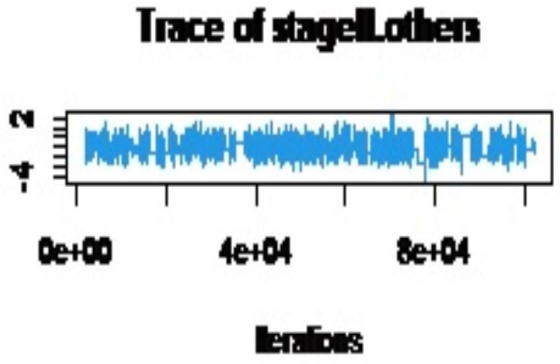

Density of famJistYes.Jobular
0.00
-6
-4
-2
0
2
4
N = 10000 Bandwidth = 0.2396

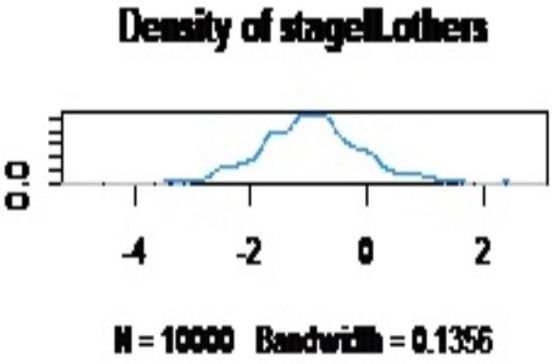

Trace of famJistYes.others
-6
2
0e+00
4e+04
8e+04
Iterations

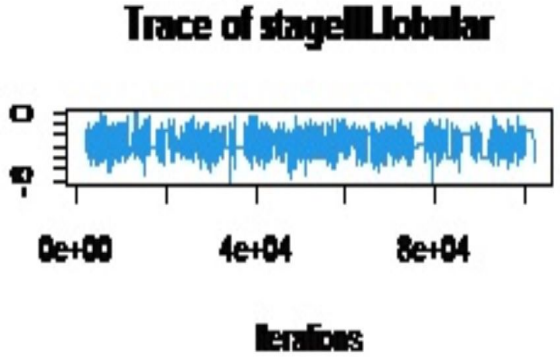

Density of famJistYes.others
0.00
-8
-6
-4
-2
0
2
N = 10000 Bandwidth = 0.2077

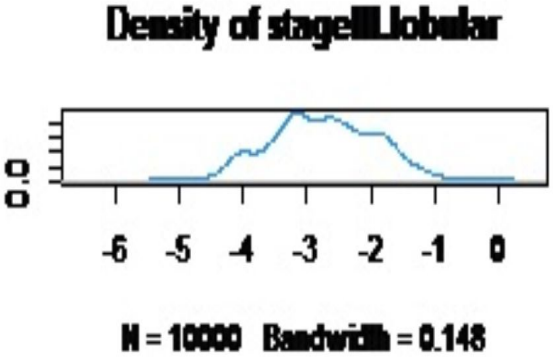

Trace of stage1Llobular
-6
0
0e+00
4e+04
8e+04
Iterations

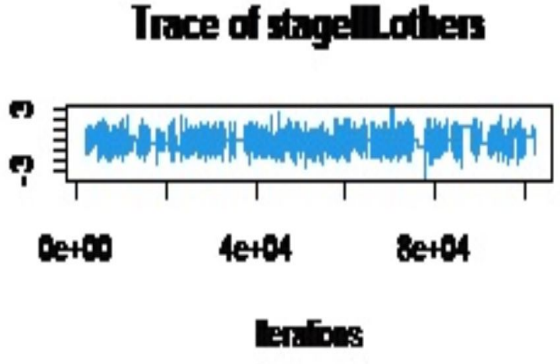

Density of stage1Llobular
0.0
-6
-5
-4
-3
-2
-1
0
1
N = 10000 Bandwidth = 0.1185

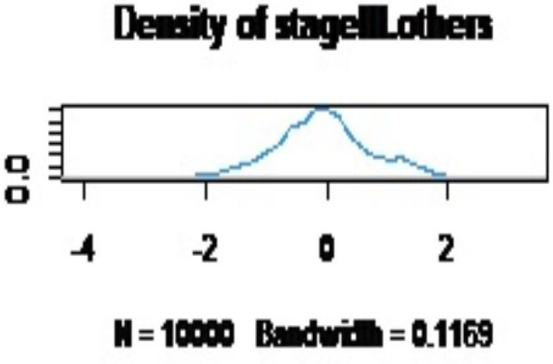

Trace of stage1Lothers
-4
2
0e+00
4e+04
8e+04
Iterations

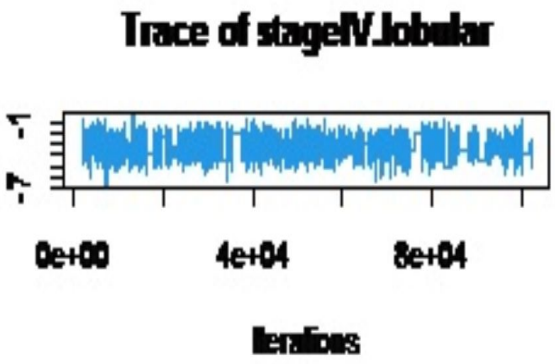

Density of stage1Lothers
0.0
-4
-2
0
2
N = 10000 Bandwidth = 0.1356

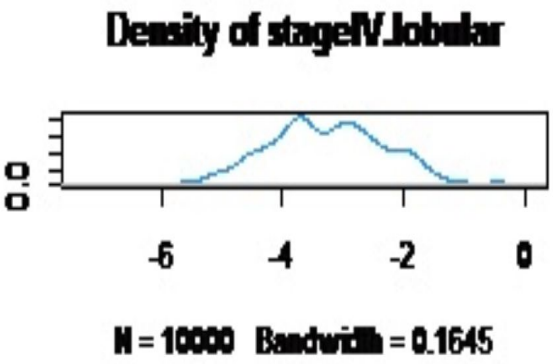

Trace of stage1Llobular
-6
0
0e+00
4e+04
8e+04
Iterations

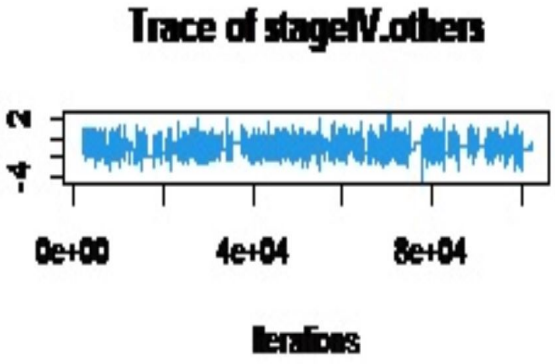

Density of stage1Llobular
0.0
-6
-5
-4
-3
-2
-1
0
N = 10000 Bandwidth = 0.148

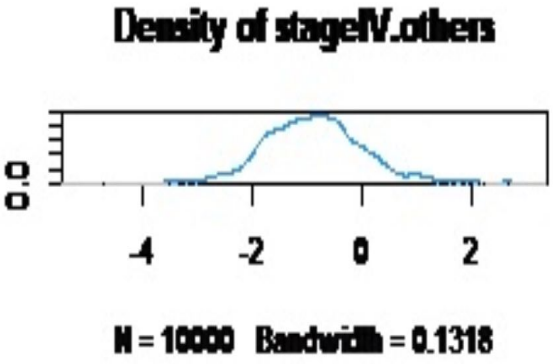

Trace of stage1Lothers
-3
3
0e+00
4e+04
8e+04
Iterations

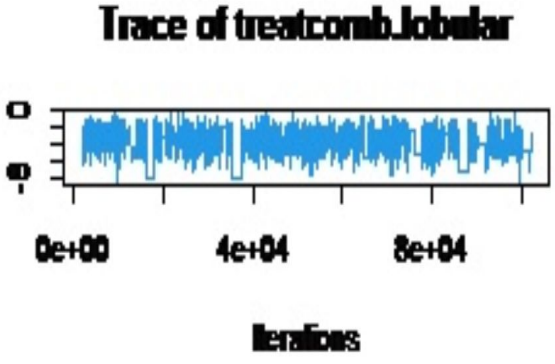

Density of stage1Lothers
0.0
-4
-2
0
2
N = 10000 Bandwidth = 0.1169

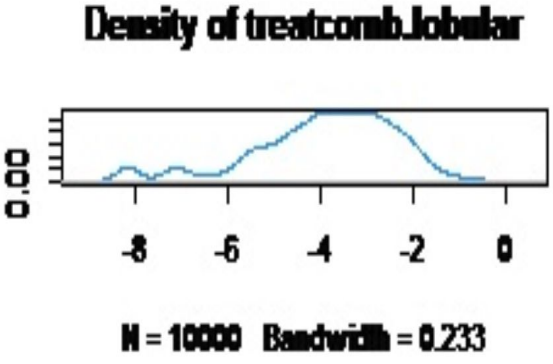

Trace of stageIV.lobular
-7
-1
0e+00
4e+04
8e+04
Iterations

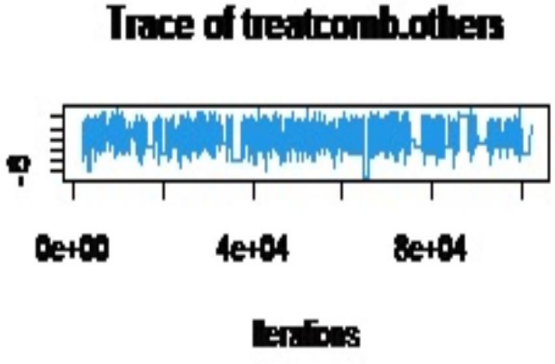

Density of stageIV.lobular
0.0
-6
-4
-2
0
N = 10000 Bandwidth = 0.1645

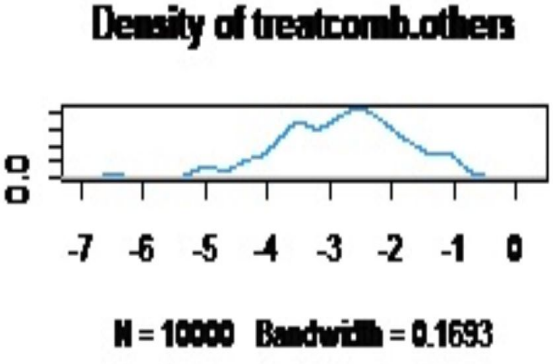

Trace of stageIV.others
-4
2
0e+00
4e+04
8e+04
Iterations

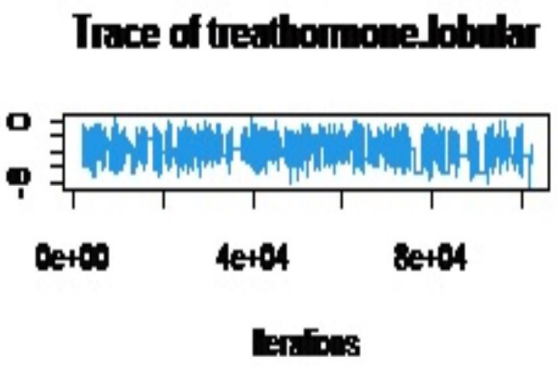

Density of stageIV.others
0.0
-4
-2
0
2
N = 10000 Bandwidth = 0.1318

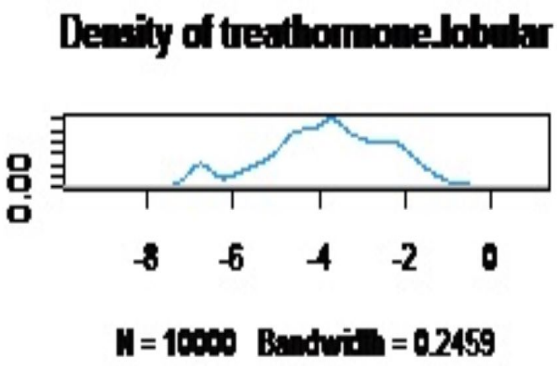

Trace of treatcomb.lobular
-9
0
0e+00
4e+04
8e+04
Iterations

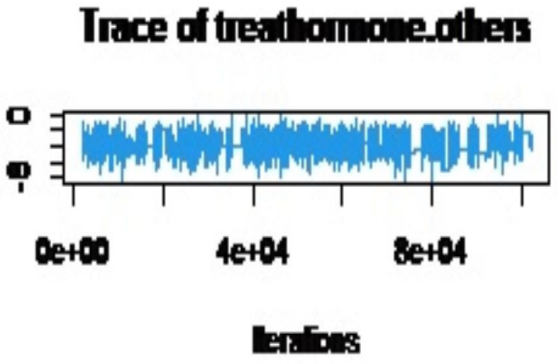

Density of treatcomb.lobular
0.00
-8
-6
-4
-2
0
N = 10000 Bandwidth = 0.233

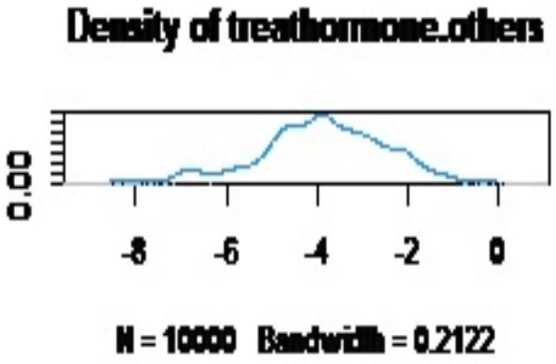

Trace of treatcomb.others
-9
0e+00
4e+04
8e+04
Iterations


Density of treatcomb.others
0.0
-7
-6
-5
-4
-3
-2
-1
0
N = 10000 Bandwidth = 0.1693

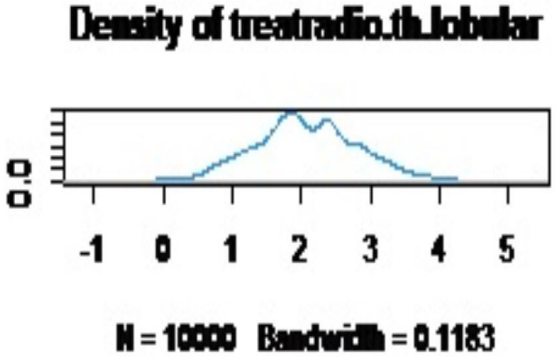

Trace of treathormone.lobular
0e+00
4e+04
8e+04
Iterations

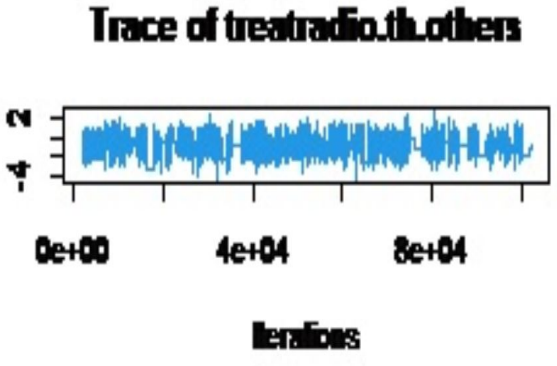

Density of treathormone.lobular
-8
-6
-4
-2
0
N = 10000 Bandwidth = 0.2459

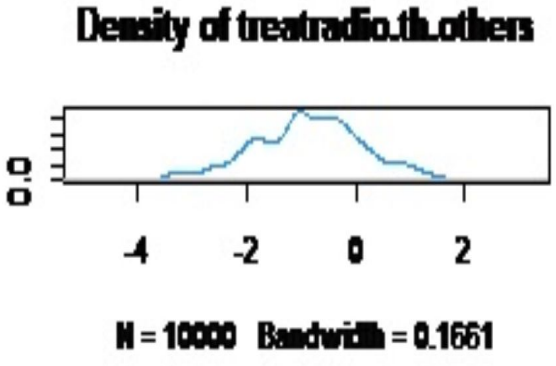

Trace of treathormone.others
0e+00
4e+04
8e+04
Iterations

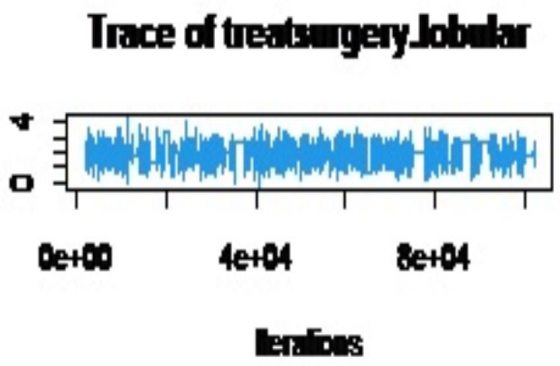

Density of treathormone.others
-8
-6
-4
-2
0
N = 10000 Bandwidth = 0.2122

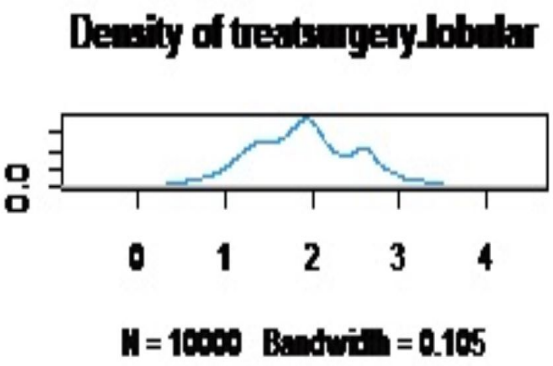

Trace of treatradio.th.lobular
0e+00
4e+04
8e+04
Iterations

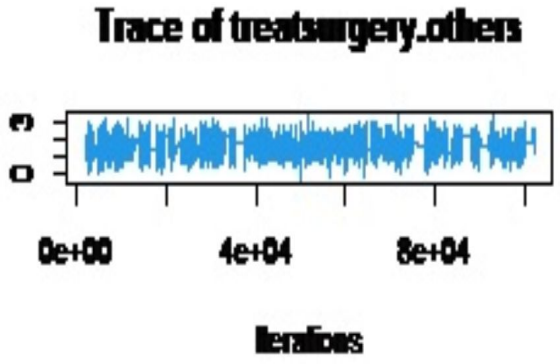

Density of treatradio.th.lobular
-1
0
1
2
3
4
5
N = 10000 Bandwidth = 0.1183

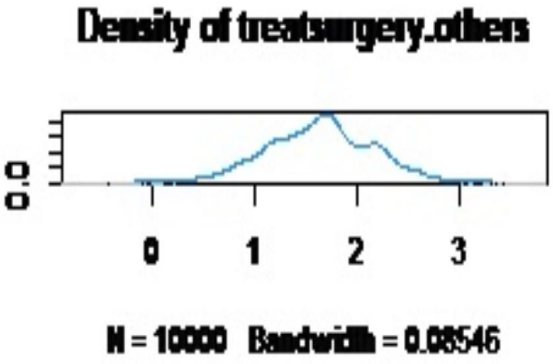

Trace of treatradio.th.others
0e+00
4e+04
8e+04
Iterations

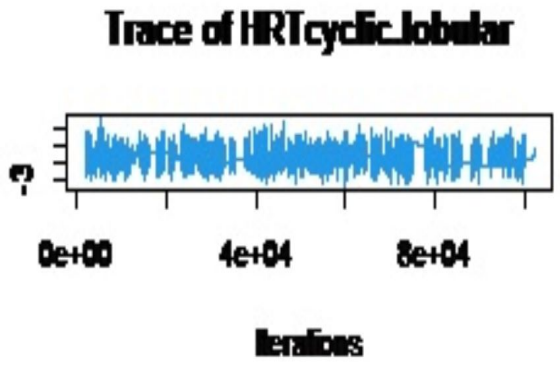

Density of treatradio.th.others
-4
-2
0
2
N = 10000 Bandwidth = 0.1661

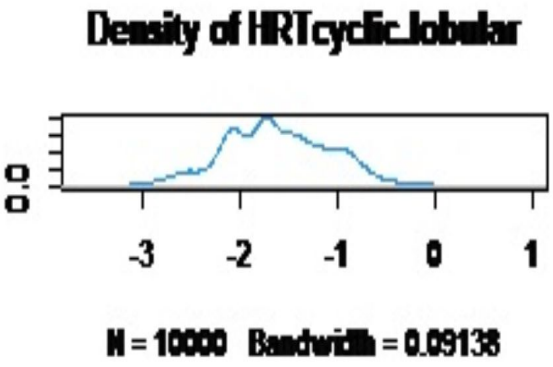

Trace of treatsurgery.lobular
4
0
0e+00
4e+04
8e+04
Iterations

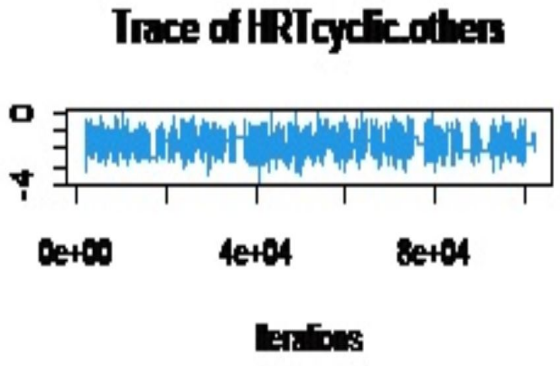

Density of treatsurgery.lobular
0.0
0
1
2
3
4
N = 10000 Bandwidth = 0.105

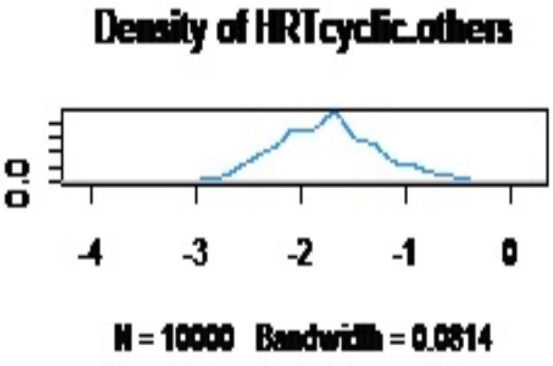

Trace of treatsurgery.others
3
0
0e+00
4e+04
8e+04
Iterations

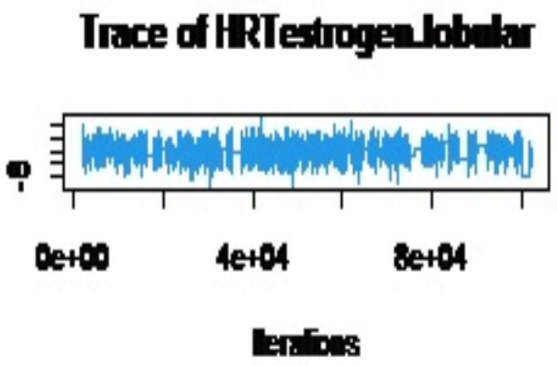

Density of treatsurgery.others
0.0
0
1
2
3
N = 10000 Bandwidth = 0.08546

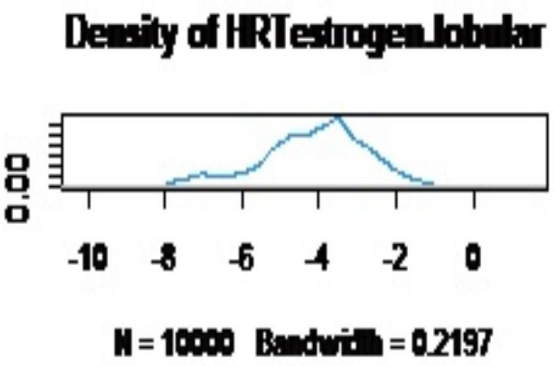

Trace of HRTcyclic.lobular
-3
0e+00
4e+04
8e+04
Iterations

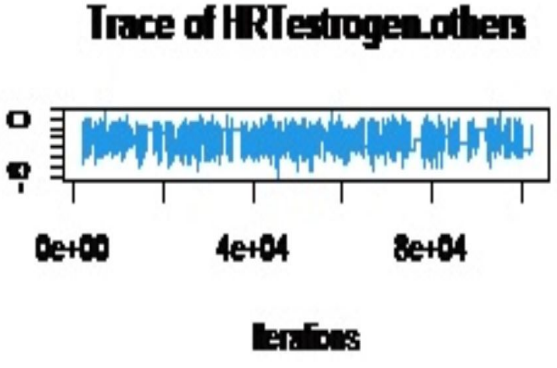

Density of HRTcyclic.lobular
0.0
-3
-2
-1
0
1
N = 10000 Bandwidth = 0.09136

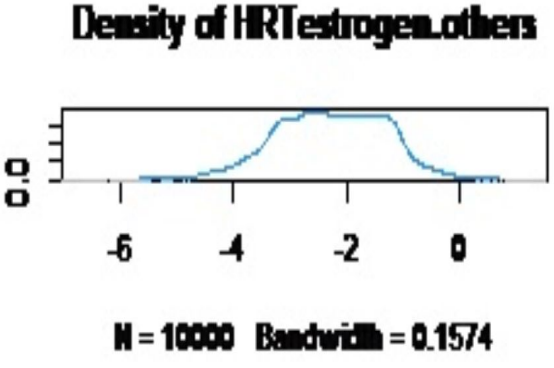

Trace of HRTcyclic.others
0
-4
0e+00
4e+04
8e+04
Iterations

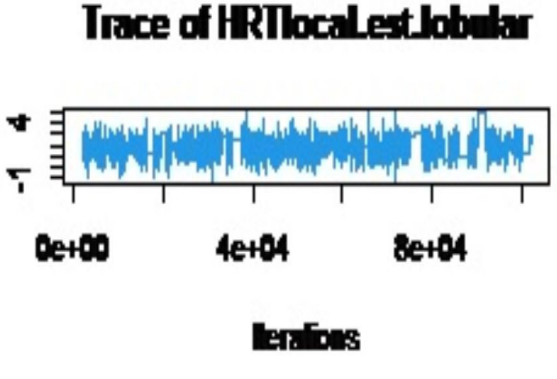

Density of HRTcyclic.others
0.0
-4
-3
-2
-1
0
N = 10000 Bandwidth = 0.0814

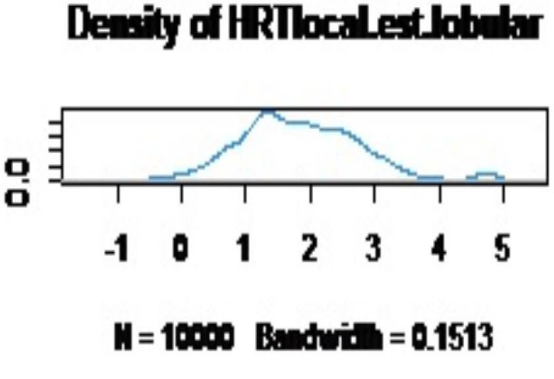

Trace of HRTestrogen.Jobular
-9
0e+00
4e+04
8e+04
Iterations

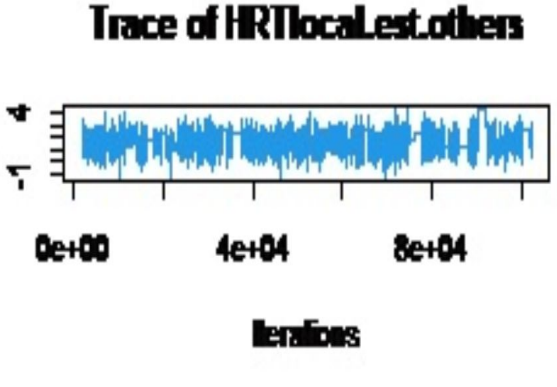

Density of HRTestrogen.Jobular
0.00
-10
-8
-6
-4
-2
0
N = 10000 Bandwidth = 0.2197

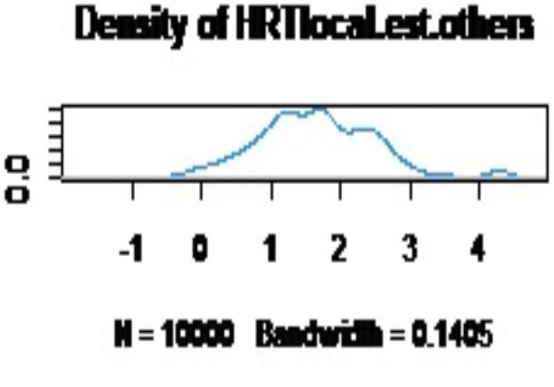

Trace of HRTestrogen.others
-9 0
0e+00
4e+04
8e+04
Iterations

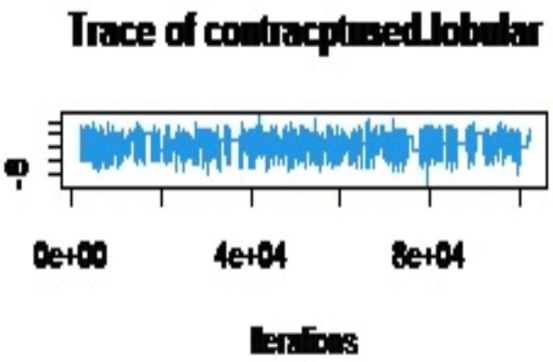

Density of HRTestrogen.others
0.0
-6
-4
-2
0
N = 10000 Bandwidth = 0.1574

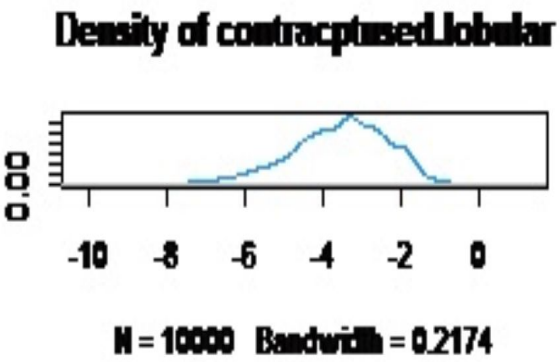

Trace of HRTlocaLest.Jobular
-1 4
0e+00
4e+04
8e+04
Iterations

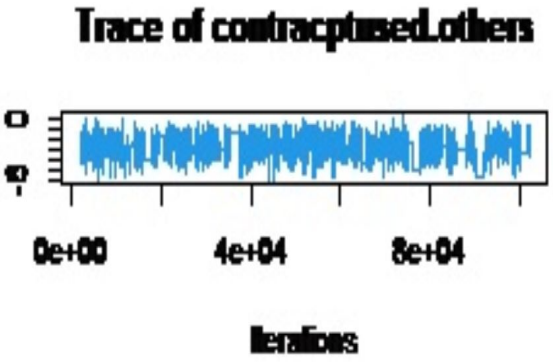

Density of HRTlocaLest.Jobular
0.0
-1
0
1
2
3
4
5
N = 10000 Bandwidth = 0.1513

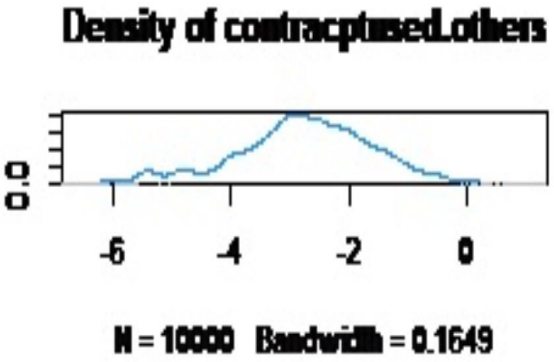

Trace of HRTlocaLest.others
-1 4
0e+00
4e+04
8e+04
Iterations


Density of HRTlocaLest.others
0.0
-1
0
1
2
3
4
N = 10000 Bandwidth = 0.1405


Trace of contraceptiveUlobular
-9
0e+00
4e+04
8e+04
Iterations


Density of contraceptiveUlobular
0.00
-10
-8
-6
-4
-2
0
N = 10000 Bandwidth = 0.2174


Trace of contraceptiveUothers
0
-9
0e+00
4e+04
8e+04
Iterations


Density of contraceptiveUothers
0.0
-6
-4
-2
0
N = 10000 Bandwidth = 0.1649