زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

Predicting Risk Factors of Bypass Graft data in Erbil using the Application of Bayesian and Non-Bayesian methods

A Thesis
Submitted to the Council of the College of Administration and Economics at Salahaddin University-Erbil in Partial Fulfillment of the Requirements for the Degree of Master in Statistical Sciences

By

Azhin Mohammed Khudhur

BSc. Informative and Statistics

Supervised by

Asst. Prof. Dr. Dler Hussein Kadir

Erbil, KURDISTAN
October 2022

DECLARATION

I declare that the Master Thesis entitled (**Predicting Risk Factors of Bypass Graft data in Erbil using the Application of Bayesian and Non-Bayesian methods)** is my original work, and hereby certify that unless stated, all work contained within this thesis is my own pendent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgment is made in the text.

Signature:

Student Name: **Azhin Mohammed khudhur**

Date: / / 2022

Supervisor Certificate

This thesis has been written under my supervision and has been submitted for the award of the degree of Master of Science in statistics with my approval as supervisor.

Asst. Prof. Dr. Dler Hussein Kadir

Name Signature

/ / 2022

Date

I confirm that all requirements have been fulfilled.

Signature:
Name: **Asst. Prof. Dr. Bekhal Samad Sedeeq**
Head of the Department of Statistics
Date: / / 2022

I confirm that all requirements have been fulfilled.

Postgraduate Office
Signature:
Name: **Prof. Dr. Kurdistan Ibrahim Mawlood**
Date: / / 2022

Examining Committee Certification

We certify that we have read this thesis: (**Predicting Risk Factors of Bypass Graft data in Erbil using the Application of Bayesian and Non-Bayesian methods)** and as an examining committee examined the student (**Azhin Mohammed Khudhur**) in its content and what related to it. We approve that it meets the standards of a thesis for the degree of MSc. in Statistics.

Signature
Name:
Member:
Date: / / 2022

Signature
Name:
Member:
Date: / / 2022

Signature
Name:
Supervisor:
Date: / / 2022

Signature
Name:
Chairman:
Date: / / 2022

Dedication

Dedicated to:

- ❖ My dearest parents for all the unconditional love, guidance, and support that they have always given me.
- ❖ My Brothers and Sisters.
- ❖ To My teachers, who taught me at any education stages.
- ❖ My beloved friends for their efforts and encouragement.

Acknowledgments

I would like to express my gratitude to the following people for their support throughout conducting my M.Sc. study.

First, I would like to express my sincere gratitude to the dean of the College of Administration and Economics Dr. Ahlam Ibrahem Wali., and the Head of the department of statistics and informatics Dr. Bekhal Samad Sedeeq for giving me the opportunity to do my Master and being so helpful during that process.

A special thanks to my supervisor **Dr. Dler Hussein Kadir** for his permanent support of my M.Sc. study, his patience, motivation, and immense knowledge. His guidance helped me think critically throughout the process of writing this thesis. I could not have imagined having a better supervisor for my thesis study.

A special thanks go to Dr. Goran Qader Othman, Lecturer at the Department of medical laboratory Technology, who helped me collect the data and gave me detailed information about the content of the data. Many thanks also go to Erbil Cardiac Center for providing me with the data.

I would like to thank the staff and my lecturers in my M.Sc. study for their continuous support and help during my courses.

Finally, a special thank goes to my parents, sisters, and brothers for being helpful and motivated during my tough times. They have supported me spiritually throughout writing this thesis and my life in general. This project would not have been possible without their support.

Abstract

Coronary heart disease can be defined as a disease that plaque (a waxy substance) constructs inside the coronary arteries. The purpose of this study was to determine risk factors affecting Bypass Graft surgery. In this research, 100 adult patients underwent coronary angiography at Erbil Cardiac Center, and data was obtained from them. Patients were followed up from January 2016 until the end of December 2020. Data was also obtained from 60 healthy people who underwent the same coronary angiography.

Multiple logistic regression was used to determine factors affecting Bypass Graft surgery. Variables were selected using the forward selection method. We have found that HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, and T3 variables have a relatively interesting connection with the Bypass Graft using visual inspection. The significant variables contributing to the prediction of for Bypass Graft operation using multiple logistic regression are DBP, Age, MCH, WBC, and Blood Sugar. The risk factors associated with Bypass Graft surgery using Bayesian logistic regression were Age, WBC, and MCH.

In conclusion, there are different results we have achieved between classical and Bayesian logistic models. Further study needed using larger sample size and including informative prior distribution into the Bayesian model.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Number	Symbol	Name
1	CABG	Coronary Artery Bypass Graft
2	CAD	Coronary Artery Disease
3	PCI	Percutaneous Coronary Intervention
4	FREEDOM	Future Revascularization Evaluation In Patients With Diabetes Mellitus: Optimal Management Of Multivessel Disease
5	LMCA	Left Main Coronary Artery
6	RCA	Right Coronary Artery
7	EF	Ejection Fraction
8	HbA1c	A Hemoglobin A1c
9	BUN	Blood Urea Nitrogen
10	CHO	Cholesterol
11	TG	Triglyceride
12	HDL	High Density Lipoprotein Levels
13	LDL	Lower Density Level
14	AST	Aspartate Transaminase
15	ALT	Alanine Transaminase
16	ALP	Alkaline Phosphatase
17	TSB	Total Serum Bilirubin
18	T3	Triiodothyronine Tests
19	T4	Thyroxine Tests
20	TSH	Thyroid-Stimulating Hormone
21	TNT	Troponins Test
22	CK-MB	Creatine Kinase
23	WBC	White Blood Cell
24	HGB	Hemoglobin
25	HCT	A Hematocrit Test
26	MPV	Mean Platelet Volume
27	MCV	Mean Corpuscular Volume
28	MCH	Mean Corpuscular Hemoglobin
29	RDW	A Red Cell Distribution Width
30	PDW	Platelet Distribution Width
31	PLT	Platelets In The Blood
32	SBP	Systolic Blood Pressure
33	DBP	Diastolic Blood Pressure
34	SCTS	The Society Of Cardio-Thoracic Surgeons Of Great

Number	Symbol	Name
		Britain And Ireland
35	ICU	Intensive Care Unit
36	EuroSCORE	European System For Cardiac Operative Risk Evaluation
37	AusSCORE	The Australian System For Cardiac Operative Risk Evaluation
38	CSCs	Cardiac Surgical Centres
39	POAF	Predictors Of Postoperative Atrial Fibrillation
40	AF	Atrial Fibrillation
41	CHA2DS2-VASc	Congestive Heart Failure Or Left Ventricular Dysfunction Hypertension, Age ≥75 (Doubled), Diabetes, Stroke (Doubled)-Vascular Disease, Age 65–74, Sex Category
42	MCMC	Markov Chain Monte Carlo
43	MLP	Multilayer Perceptron Neural Networks
44	AIC	Akaike Information Criteria
45	ROC	Receiver Operating Characteristic
46	PCI	Percutaneous Coronary Intervention
47	GLM	Generalized Linear Model
48	OLS	Ordinary Least Squares
49	VIF	Variance Inflation Factor
50	AUC	Area Under The ROC Curve
51	BNN	Bayesian Neural Network

CHAPTER ONE

LITERATURE REVIEW AND THEORETICAL DESIGN

CHAPTER ONE

1. LITERATURE REVIEW AND THEORETICAL DESIGN

1.1. Introduction

Coronary heart disease can be defined as a disease that plaque (a waxy substance) constructs inside the coronary arteries. Oxygen blood is supplied to the heart by these arteries. Bypass Grafting surgery is a kind of surgery that enhances the amount of blood flow to the heart. A surgeon can Bypass multiple coronary arteries in just one surgery (Alexander & Smith, 2016).

One method of treating blocked or narrowed arteries is to Bypass the blocked piece of the coronary artery with a portion of a \healthy blood vessel from another part of the body. Blood vessels, or Grafts, used for the Bypass operation, could be parts of a blood vessel from the leg or an artery in the chest. It is also possible that an artery from the wrist will be used. The doctor attaches one end of the Graft above the blocking and the other below. Blood flows through the new Graft to reach the heart muscle, Bypassing the blockage. This is referred to as coronary artery Bypass surgery.

The Bayesian technique has been used to reduce mortality in diabetic individuals with multivessel coronary artery disease using Bypass data (CAD). The primary goal of the research was to assess death rates in patients with diabetic multivessel CAD after receiving PCI or coronary artery Bypass Graft (CABG) therapy for five years or longer. The authors updated preliminary information from eight clinical trials (n=2024) using a Bayesian technique. The likelihood distribution was derived from a diabetes patient's future revascularization examination. Data on optimal management of multivessel disease (n=1460) was also acquired to see if evidence from clinical trials supported

the basic concept is that CABG is superior to PCI for diabetics with multivessel CAD.

From several of each specific trial or meta-analysis of various trials identifying whether CABG is better to PCI and larger result for diabetic patients with multivessel CAD at 5 years or longest follow up. Incorporating of evidence from several studies, in diabetic patients with multivessel CAD, the basic notion that CABG revascularization improves survival over PCI is instantly supported by Bayesian approaches. In diabetic patients with multivessel CAD, CABG is recommended that advantage clearly exceed risk is most, if not all, diabetic patients with multivessel CAD. A suggestion of positive result in general survival in FREEDOM declares that CABG appeared to negotiate greater survival than did PCI in a single supporter of patients at an individual point in time (Lang, et al., 2015).

Guillermo Marshall et al. have used the Bayesian logit model for the risks associated with coronary disease of Bypass. He stated that the Bayesian approach is more popular nowadays among other approaches to predict mortality between patients undergoing Bypass surgery. They proposed a new model using data from 12,712 patients undergoing Bypass Grafting procedures collected between 1987 and 1990. the authors have provided a comparative analysis between Bayes theorem, logistic regression, discrimination analysis, and Bayesian logit models. Their data is divided into a train data set and test data set. The collected data were from various factors measured during Bypass surgery, and there were missing data in some variables. The data were obtained from more than 6,000 CABG surgery between 1987 and 1992, and they found outstanding risk factors that were significant.

The authors provided the results of logistic regression, discriminant analysis, Bayes theorem, and Bayesian analysis. They found that Bayesian analysis can be a constrained model of a mixture of discrimination and logistic analysis. However, it has been found that the Bayes theorem is more complex in the model equation. Thus, the interpretation of clinical models' results is not clear regarding predictor variables.

The benefits of predicting the outcome of patients experiencing cardiac surgery have been illustrated as continuous rather than constant intervals. Furthermore, applying the Bayesian-logit model approach, the study results have been compared with different statistical methods such as discrimination and logistic regression analysis.

Researchers gathered data at several sites across the United States depending on the number of CABG cases performed each year and surgeon experience with beating-heart procedures at each location. The study was limited to three years, from 1999 to 2001, to avoid any bias resulting from the centers' early learning experiences. The database comprised patients who had re-operative or original surgery. First, the authors compared various groups to determine mortality and morbidity selection criteria. Then, to eliminate selection bias, computer programming was utilized. Data on heart surgery patients was eventually acquired, and the data was entered into a logical database.

The data were evaluated using multivariate logistic regression to predict risk factors affecting mortality in 17401 isolated CABG patients. The data includes 7283 (41.9%) off-pump coronary arteries and 10118 (58.1%) routine CABG with cardiopulmonary Bypass procedures. Missing data was also considered as it accounted for more than 10% (Mack, et al., 2004).

Researchers used the multivariate logistic regression approach to find predictive risk factors for mortality. Sub-groups that are almost certain to benefit were identified. The Parsonnet risk model is a logistic regression model that was frequently used to characterize variation among treatment groups, which can affect patient selection and results, with 47 potential risk factors investigated preoperatively to detect risk. This model is used to compare anticipated and observed death rates. Moreover, using cardiopulmonary Bypass as an independent variable, backward elimination techniques were employed to find significant predictors of operation mortality from a collection of 20 preoperative risk factors. Finally, a series of alternate regression models were investigated with mortality and extensive morbidity as outcome variables to discover subgroups that may benefit from the beating-heart procedure (Mack, et al., 2004).

1.2. Variable Explanation

Artery: The large muscular vessels that carry blood away from the heart, their walls consist of three layers (endothelium, smooth muscle, and connective tissue). They lack valves.

Coronary arteries: are the vessels that carry the rich-oxygenated blood to the heart muscle. If the blood flow in coronary arteries is reduced or cut off, the muscle cell will die. This can happen when an artery is blocked by a blood clot or atherosclerosis, a disease characterized by the buildup of fatty materials on the interior walls of coronary arteries. Either type of blockage can lead to a heart attack.

1 **Left main coronary artery (LMCA):** The left main coronary artery carries blood to the left side of the heart muscle (the left ventricle and left atrium). The left main coronary divides into branches:

● The left anterior descending artery branches of the left coronary artery and carried blood to the front of the left side of the heart.

● The circumflex artery branches off the left coronary artery and encircles the heart muscle. This artery carries blood to the outer and backside of the heart.

2 **Right coronary artery (RCA):** The right coronary artery carries blood to the right ventricle, the right atrium, and the SA (sinoatrial) and AV (atrioventricular) nodes, which coordinate the heart rhythm. The right coronary artery divides into smaller branches, including the right posterior descending artery and the acute marginal artery. Together with the left anterior descending artery, the right coronary artery helps supply blood to the middle or septum of the heart.

Ejection fraction (EF): A measurement measuring how much blood comes out from the heart's lower chamber with each contraction, which is expressed as a percentage. The usual range of heart ejection might be between 50-70 percent.

Blood Sugar: The main sugar in human blood cells is glucose. You can get it from the food you eat, which is the main energy source. Human blood cells carry glucose to all parts of your body's cells for energy. A normal Blood Sugar level is less than 140 mg/dL (7.8 mmol/L). The range between 140 and 199 mg/dL (7.8 mmol/L and 11.0 mmol/L) indicates prediabetes. However, when it is over 200 mg/dL (11.1 mmol/L) after two hours indicates diabetes.

A hemoglobin A1c (HbA1c): A test that measures the quantity of Blood Sugar (glucose) attached to hemoglobin. The average amount of glucose attached to hemoglobin can be determined by the HbA1c test during the past three months. The usual range of hemoglobin A1c is below (5-7) percent. If the range is between (5.7-6.4) percent, it is prediabetes, but 6.5 percent or higher shows diabetes.

Blood Urea nitrogen (BUN): A test that measures the quantity of nitrogen in blood from the waste product urea. In general, around 6 to 24 mg/dL (2.1 to 8.5 mmol/L) is considered normal.

A Creatinine blood test: Is a test that measures the amount of Creatinine inside the blood. If the kidneys are healthy, they filter Creatinine out of the blood. Creatinine is considered a waste product, which is found in the muscles and breaks down. The usual range of Creatinine blood tests is between 0.74 and 1.35 mg/dL for males and 0.59 and 1.04 mg/dL for females.

Cholesterol: Is a waxy, fat-like substance located in the blood and every cell of the human body. Humans need some Cholesterol to perform their functions correctly.

Types of Cholesterol in the blood:

1 Total Cholesterol:

The combination of low-density lipoprotein (LDL) Cholesterol and high-density lipoprotein (HDL) Cholesterol in human blood. The normal range should be between (140-200) mg/dl.

2 Triglycerides:

It is considered one of the fats found in human blood. According to some studies, high levels of triglycerides in human blood are shown as one of the factors that might raise the risk of heart disease, especially in women. The normal range should be less than 150 mg/dl.

3 High-density lipoprotein levels (HDL):

Known as "good" Cholesterol, HDL helps to eliminate "bad" LDL Cholesterol. The normal range is between 35-75 mg/dl for females and 30-60 mg/dl for males.

4 Lower density level (LDL):

Also familiar as the "bad" Cholesterol, LDL is the most common factor of source artery blockages. The normal range should be less than 130 mg/dl.

Aspartate transaminase (AST): Also known as serum glutamic oxaloacetic transaminase (SGOT), an enzyme in the liver that helps metabolize amino acids. Any deficiency of (AST) level causes muscle damage, and liver damage, sometimes leading to liver disease. The usual range of (SGOT) is between 10-40 IU/L.

Alanine transaminase (ALT): Is also known as serum glutamic pyruvic transaminase (SGPT), an enzyme in the liver that is needed by the body to convert proteins into energy for the liver cells. Whenever the liver is damaged, (ALT) levels increase and are released into the bloodstream. The usual range of (SGPT) is between 10-40 IU/L

Alkaline phosphatase (ALP): Is an enzyme in the liver and bone that is essential for breaking down proteins. Above the expected levels of (ALP) may designate liver damage or diseases, such as a blocked bile duct or certain bone diseases. The usual range of (ALP) is between 40-112 U/L.

Total serum bilirubin (TSB): Is a test that measures the number of levels of Bilirubin in your blood. Bilirubin is a yellowish substance made during the body's normal process of breaking down red blood cells, a yellowish substance fluid produced in bile called Bilirubin, which assists in digesting food. The usual range of (TSB) is between 0-1mg/dl.

Triiodothyronine tests (T3): This helps to diagnose hyperthyroidism or to predict the seriousness of hyperthyroidism. The usual range of (T3) is between 100 - 200 ng/dL.

Thyroxine tests (T4): A thyroid test helps diagnose thyroid disorders, a small, butterfly-shaped gland found near the throat. The thyroid produces hormones that coordinate the way that the body uses energy. As a result, it significantly regulates body temperature, mood, muscle strength, and weight. Thyroxine, also known as T4, is one of the thyroid hormones. The usual range of the (T4) test is between 0.9 - 2.3 ng/dL.

Thyroid-stimulating hormone (TSH): This is a stimulating hormone that signals to the thyroid gland to regulate the release of thyroid hormones, which help maintain heart rate, blood pressure, and body temperature. The usual range of (TSH) is between 0.5 -5.0 mIU/L.

The troponins test (TNT): Is a measurement that measures the amount of cardiac-specific troponin in the blood. Troponins are a bunch of proteins located in skeletal and heart muscle fibers, which coordinate muscles' contraction and help, detect heart injury. The normal range for troponin I is 0 - 0.04 ng/mL.

Creatine kinase (CK-MB) test: A test that measures the quantity of creatine kinase in the blood.CK is the kind of protein that is known as an enzyme. It is found in skeletal and heart muscles. The amount of CK that is found in skeletal and heart muscles is more than the amount of CK in the brain. The normal range is between 5 to 25 IU/L.

White Blood cells (WBC): Also known as leukocytes, they are formed in red marrow. This type of blood cell help defend the body against the disease, and they may function for years if they lack hemoglobin. The average amount of (WBC) in the blood is 4,500 to 11,000 (WBC) per microliter (4.5 to 11.0 × 109/L).

Eosinophils: A type of white blood cell that Functions as a disease-fighting cell. This condition usually determines a parasitic infection, an allergic reaction, or cancer**.** Eosinophils usually make up 0-6 percent of the white blood cells**.**

Hemoglobin (HGB): It is an iron-containing protein molecule that transports respiratory gasses found in red blood cells. The usual range of (HGB) is between 13.5-17.5 g/dL for males and between 12.0-15.5 g/dL for females.

A hematocrit test (HCT): It is known as a packed-cell volume (PCV) test that measures the number of red blood cells inside the blood. Red blood cells are responsible for carrying oxygen in the body. Red blood cells above or under the normal range can be considered a symptom of certain diseases. The usual range of (HCT) is between 41-50 percent for males and 36-48 percent for females.

Mean platelet volume (MPV): Is a blood test that measures the average platelet size, which are small blood cells necessary for blood clotting. MPV test can help diagnose bleeding disorders and bone marrow diseases. The normal MPV range is between 7-12 fl (fl= one femtoliter, a minimal unit of blood).

Mean corpuscular volume (MCV): An (MCV) blood test that measures red blood cells' average volume and size. Corpuscles (blood cells) are divided into three forms: red blood cells, white blood cells, and platelets. The usual range of (MCV) is between 80-95 fl.

Mean corpuscular hemoglobin (MCH): Is the average quantity in each red blood cell of a protein called hemoglobin, which transports oxygen through the body.
The normal range for (MCH) test is between 27.5 -33.2 picograms (pg) per cell.

A red cell distribution width (RDW): Is a test that shows the variation in the size from the smallest to most significant red blood cells (RBC) in a sample. A typical range for (RDW) test is between 12.2 -16.1 percent in females and 11.8 - 14.5 percent in males.

Platelet distribution width (PDW): This is a systematic parameter in blood routine examination that follows the variation of the size of platelet distribution. The usual range of (PDW) test is between 8.3 -56.6 percent.

PLT: A test that measures the number of platelets in the blood and is usually in the general health examination. This test is helpful in identifying the normal range of platelet count for patients taking medicines, which could seriously affect the platelet count. A typical range for the number of platelets in the blood is between 150,000 - 450,000 platelets per microliter.

Systolic Blood Pressure (SBP): The heart pressures blood into the arteries when the heart beats. The average systolic blood pressure is 120 mmHg or less.

Diastolic Blood Pressure (DBP): The heart relaxes pressure among beats. Thus, it may refill with blood. The average diastolic blood pressure is 80 mmHg or less.

1.3 Aim of the Thesis

The aim of this thesis is to examine factors affecting undergoing bypass surgery. The objective of this thesis can be reached by using multiple logistic regression and Bayesian Logistic Regression, by assessing the contribution of factors that are linked to the probability of being into surgeries between healthy people and patients

1.4 Literature Review and Theoretical Design

Melin et al. (1985) used the conditional probability analysis of alternative hypothetical strategies in women for diagnostic coronary artery disease. The current study aimed to compare the sufficiency of hypothetical diagnostic strategies in coronary artery disease among women. They selected two groups of women who underwent stress electrocardiographic, thallium scintigraphic, and coronary angiographic examination. The researchers used statistical techniques as a conditional probability analysis based on Bayes' theorem to diagnose coronary artery disease. Accuracy score with the coronary angiogram was used as a standard accuracy diagnostic to estimate individual probability accuracy and later tested statistically with the contrast method of Scheffé. Finally, they used McNemar's test to compare diagnostic strategies with correctly classified, misclassified patients with an intermediate probability.

The researchers found that diagnostic tests suggest superior balances of identification of accuracy, the comfort of the patient, and reduced financial cost. In addition, physicians' prior knowledge of probability estimate treatment decreases the number of useless diagnostic coronary angiographic techniques.

Edwards et al. (1988) used a statistical model based on the Bayesian theorem on coronary Bypass Grafting data that improved mortality prediction in patients who underwent isolated CABG. The present study aimed to compare predicted and observed risk factor models for the mortality rate among patients. The patients were stratified reliably into two risk categories in the same hospital for three years. The researchers used the statistical risk model derived from the Bayesian model to predict the mortality rate among patients. The first group of patients was applied to their model as the initial database, and they prospectively evaluated each group from the remaining second part of patients incorporated into the database to update the model. The researchers found that the most reliable model of operative outcome between predictive and observed results was Bayesian theory, which includes flexible attributes, such as multiple risk factors, determining individual rather than group, prediction, and updated with time. These attributes are fundamental in light of current changes in patient profile in the coronary artery Bypass Graft.

Khan et al. (1990) used the Bayesian approach to Bypass data to explore whether the mortality rate of women in coronary artery surgery increased. In addition, the study aimed to discover whether divergence in women's preoperative status calculates higher surgery mortality rate for women having coronary artery Bypass surgery. All selected patients had coronary artery Bypass at five years. The researchers used multivariate analysis using a logistic regression model. In addition, they used maximum likelihood stepwise logistic regression to choose the significant variables of presurgery for the prediction of in-hospital death.

The researchers found that functional class and Age divergence calculates women's higher surgery mortality rate in coronary Bypass operations. In addition, women are referred for coronary Bypass surgery next in the course of their disease than men, and later referral may significantly increase their opportunities for surgery death.

Marshall et al. (1994) have applied the Bayes' theorem model on Bypass data to estimate mortality for operative risk assessment between patients undergoing CABG. The study aimed to compare the predictive power of the classic statistical alternative model of mortality risk factors among patients who underwent CABG. The patient was divided randomly into two samples with the ignored correlation between them at three years.

The researchers used different classic statistical procedures, including logistic regression, which was used for predicting mortality, combined cluster analysis with clinical judgment followed by logistic regression, and a model including a principal component used to eliminate the number of predictors, followed by logistic regression analysis, classification tree model, subjectively created sickness score. Further, they reformulated the Bayes' theorem into a more straightforward and familiar form as the Bayesian-logit model is a mixture of logistic regression and linear discriminant analysis.

The researchers found that the Bayes model had the highest predictive power among all the models. In addition, the Bayesian logit model could discriminate between deaths and survival operative is comparable with that of logistic regression and linear discriminant analysis.

Kurki and Kataja (1996) used the Bayesian approach for undergoing Coronary Artery Bypass Grafting data for the preoperative prediction of postoperative morbidity from patients. The current study aimed to improve the scoring procedure for postoperative morbidity prediction of individual patients undergoing Bypass Grafting. The patients who underwent CABG were retrospectively collected in a single center for one year.

The researchers used univariate analysis and logistic regression analysis to describe risk factors for perioperative morbidity and mortality; Higgins and associates studied a massive CABG group of patients. They used a stepwise Bayesian approach to determine posterior probabilities and likelihood ratios and to conform to the sensitivity. The CABDEAL model was used to select individual patients with a high risk of postoperative difficulties. The researcher showed that selecting patients for CABG according to their risk factor of postoperative difficulties can be stratified based on preoperative information for individual and group patients.

Lippmann et al. (1997) discussed the neural networks in their study as robust, nonparametric, and pattern recognition procedures that can be utilized to model complex relationships. The medical data came from 80,606 patients in The Society of Thoracic Surgeons database in 1993. The authors employed the normalization procedure to get a zero mean and unit variance for the three predictor variables. They used non-Bayesian logistic and Bayesian analysis, and both approaches were compared with single-layer without a hidden layer, two-layer with one hidden layer, and three-layer with two hidden layers of MLP neural networks.

The author concluded that the best model calibration was obtained using a committee classifier that combined a logistic model with the best neural network; nonetheless, the area under the ROC was 76 percent regardless of which predictive model was used.

TU (1997) used clinical data from 5,517 patients who had isolated CABG in Ontario in 1993 to develop 12 increasingly thorough risk-adjustment models to assess the Outcomes. The authors used logistic regression modeling and created 12 complicated risk-adjustment models for predicting in-hospital mortality following CABG. Each risk-adjustment model's parameter estimation using regression analysis was used to calculate a predicted death rate based on the prevalence of patient attributes at that institution. After various adjustment levels, the Spearman rank correlation coefficient and the Pearson correlation coefficient of risk-adjusted mortality rates of relative hospital rankings were determined.

In a risk-adjustment model, adding new covariates outside a core set had no significant impact on the results. However, in-hospital mortality rates may change due to random variance, uncounted case-mix variations, or disparities in quality of care. Additional research will be needed to assess the generalizability of their findings and to identify key risk indicators for assessing the short-term mortality hazards of CABG.

Abramov et al. (2000) used statistical techniques to Bypass data to trends in results of coronary artery surgery in a recent, nine years study. The study aimed to determine the contemporary risk factor for isolated CABG to analyze the trends in operation during nine years and could be associated with varying mortality of surgery and the combination of mortality and morbidity (M + M). All patients underwent CABG surgery and were divided into three-time groups.

The researchers used univariate and multivariate analysis techniques; the authors compared continuous variables using analysis of variance for three-time groups and categorical using Fisher's specific test or Chi-square. In addition, they used stepwise multiple logistic regression analysis to identify independent predictors of surgery mortality and early nonfatal difficulties.

The researchers found a trend toward surgery in adult patients with more difficulties. However, hospital mortality has been steady, and risk-modified (M + M) has constantly decreased. This decrease was incorporated with elevated use of left internal mammary artery Graft, multiple arterial conduits, and warm blood cardioplegia during the recent years of study.

Resnic et al. (2004) updated the Bayesian approach on rare interventional cardiology process data for investigation methods to monitor the safety in interventional cardiovascular procedures of medical devices were introduced into clinical application. The current study aimed to compare the clinical application result for monitoring medical device safety after investigating Bayesian updating in interventional cardiovascular to frequent classical statistics. The selected patients underwent a rare interventional cardiology process, rotational atherectomy, randomized into subgroups. The researchers used the conjugate binomial distribution, the beta distribution as the prior distribution. They estimated prior probability distribution for each subgroup of patients and then modified these estimates via the Bayesian updating method as empirical data were collected.

The researchers found that the practicability of Bayesian updating practiced to the safety assessment of medical devices and identified that the method is efficient in generating fixed estimates of risk in different groups of patient risk.

Mack et al. (2004) gathered 17401 isolated patients across the United States depending on the number of yearly CABG cases and surgeon experience with beating heart procedures at each location. The study was conducted over three years. The trial aimed to determine whether off-pump CABG is connected with better early outcomes than traditional CABG. Researchers used the multivariate logistic regression approach. The Parsonnet risk model is a logistic regression model frequently used to characterize variation among treatment groups. The authors have used alternative models with mortality and extensive morbidity as outcome variables to discover subgroups that may benefit from the beating-heart procedure.

The capacity to have superior outcomes in high-risk patients in terms of mortality and a variety of morbidity characteristics demonstrates the surgery's potential clinical benefit. Furthermore, even though it is not randomized, this study contributes to the growing body of patients by bridging the gap between sample sizes with sufficient statistical power to detect the variation.

Ugolini and Nobilio (2004) used collected data from the clinical administration that involved 6457 patients discharged from one of the six CSCS after CABG, containing elective and emergency admissions for two limited years. The study's goal was to see if administrative data could be used to predict mortality in patients undergoing CABG in the hospital. Between 2000 and 2001. The method used for the current study was multiple logistic regression to assess the efficiency with which mortality for patients can be predicted by conforming to each various risk adjustment procedure used. EuroSCORE was used to estimate mortality and assess hospital implications for patients in the hospital. In addition, the authors used the c-statistic to evaluate the model's discrimination. Finally, the Hosmer - Lemeshow statistic was used to investigate the model's calibration.

These findings suggest that it has a particular impact when considering the possibility of applying and adjusting administrative procedures to illnesses for which a clinical risk index such as EuroSCORE has not yet improved.

Rees and Dineschandra (2006) show the necessity of risk factors in monitoring heart operation training and reviews Bayesian and frequentist techniques to the issue. The data from medical records dating back over 25 years was obtained from SCTS. The main point of the present study was to predict the outcome of the patient affected by different factors, including hardness of illness, location of surgery, and efficiency of medication. The current study used statistical modeling procedures, including multivariate regression analysis on the demographic status of the patient, and the Parsonnet scoring system was utilized to adjust mortality rates for the first time.
In conclusion, there appears to be a slight difference in discrimination among the EuroSCORE and Bayes systems as observed by the area under the ROC curve, and the calibration plot implies that the Bayes system was more settled throughout time. Furthermore, Bayesian networks could be used instead of Bayes tables to model these approaches, making it a brighter and potentially more accessible procedure.

Cevenini et al. (2007) approved predictive models to compare and critically analyze the characteristic of several popular systems for predicting morbidity from patients and estimating morbidity probability after cardiac operation in an Intensive Care Unit (ICU), including the entire structure for 1090 patients who passed CABG for two years. The objective of the research was to practically test the fulfillment of several suitable predictive models when restrictedly constructed for an individual scenario. The authors have predicted eight models depending on the k-nearest neighbors and Bayes rule. In addition, the Higgins and direct scoring systems, the quadratic model, logistic regression, integer score systems, and two feed-forward artificial neural networks with one and two layers were also investigated from a theoretical perspective.
In conclusion, because Bayes and logistic regression had good generalization and calibration, both models pretended to be better than the other models. The Bayes

quadratic model proved to be a viable alternative to the Byes linear and logistic regression models, which are significantly more regular.

Reid et al. (2009) show the involvement of preoperative risk with cardiac operation can be identified due to different risk prediction models, not one thing which is particular to the Australian population used 7709 adult patients who underwent cardiac surgical operations were recorded in 6 public hospitals in Victoria for four years. The goal of this study was to highlight the risk factors connected with 30-day mortality after CABG in an Australian cohort and to enhance a preoperative model for predicting one-month mortality risk. The authors used logistic regression analysis with a P-value of Preoperative variables, and testing was structured using the bootstrap approach. Wherein six candidate models were identified. The candidate models group's last model (Aus SCORE) was identified using prediction mean square error and the Akaike Information Criteria (AIC). The author concluded that the AusSCORE model's fulfillment is better than the EuroSCORE model's for the same data and when compared to Australian data.

Lang et al. (2015) used the Bayesian technique to reduce mortality in diabetic individuals with multivessel coronary artery disease using Bypass data (CAD) from eight clinical trials for 2024 patients. The primary goal of the research was to assess death rates in patients with diabetic multivessel CAD after receiving PCI or coronary artery Bypass Graft (CABG) therapy for five years or longer. The authors showed that both models of conjugate prior to comparing mortality rates and weighted models of drug-eluting states favor the use of CABG, and Bayesian hierarchical meta-analysis utilizing non-informative prior distribution verified the findings.
In conclusion, several of each specific trial or meta-analysis of various trials identified whether CABG is better than PCI and more significant result for diabetic

patients with multivessel CAD at five years or most extended follow-up. A suggestion of favorable results in general survival in FREEDOM declares that CABG appeared to negotiate more remarkable survival than PCI in a single supporter of patients at an individual point in time.

Perrier et al. (2017) have used Bayesian analysis after isolated coronary artery Bypass Grafting data to determine postoperative predictors of atrial fibrillation.
The current study aimed to identify predictors of patients' risk of POAF with the necessary to create relevant groups to improve the efficient preventive strategies. They selected patients who isolated CABG at five years. The researchers have used two Markov Chain Monte Carlo with Gibbs sampling to compute the Gelman and Rubin Convergence selection. The AF risk was modeled with univariate and multivariate logistic regression. Although, which used normal prior distribution to estimate logistic regression.
The researchers have found five independent risk factors after isolated CABG of preoperative predictors of POAF was determined as CHA2DS2-Vasc score, severe obesity, renal failure, preoperative B-blocker use, and preoperative antiplatelet therapy play an increasing risk role of stroke in the AF or after surgery. Although it includes these five independent predictors of POAF is necessary to test preventive strategies using the medication, such as non-antiarrhythmic and anti-arrhythmic medications.

Othman et al. (2019) used statistical analysis on Bypass Graft data to investigate the linkage of angiography to hematological and some biochemical outcomes in coronary artery patients. The study aimed to examine the relation among harshness of coronary artery disease due to angiography and alteration of some necessary biochemical, hormonal and hematological variables in CABG patients. They selected eighty adults who underwent coronary angiography patients.

The researchers used statistical techniques, such as unpaired t-tests and one-way ANOVA for comparison among treatments. In addition, they used the Turkey test as a post hoc test for ANOVA results and considered all data as the mean and standard error of the mean.

The researchers found that the superiority of Cholesterol and its connected indexes, especially LDL rather than TG, refers to the harshness of coronary artery disease. Although, blood glucose and HbA1c were not connected to the degree of coronary artery disease. Decreasing T3 hormone and platelets significantly and increasing MPV were registered in patients suffering from three-vessel occlusion. They suggest rigid cooperation among the harshness of coronary artery disease and LDL, MPV, and T3 in CABG patients.

Rechardi et al. (2020) have applied the Bayes Factor approach to Choose surgical factors that negatively affect the duration of diagnosis of CABG postoperative patients using medical data collected from 673 patients who underwent CABG at Fundacardio Foundation, Valencia, Venezuela, for four years. The study aimed to predict the risk factors that predict postoperative morbidity in patients who experienced CABG using Naïve Bayes and Bayes Factor methods. The authors utilized the Nave Bayes classification and Bayes factor, which addict the postoperative morbidity.

According to the findings, 99 percent of patients with specific risk factors spent more than four days in the hospital after surgery. In addition, the patient's Age was greater than 80 years old, and they were female the severity of the threat of EuroSCOR, period of surgery more significant than 4 hours. The findings revealed that patient morbidity predictions could be used to determine the amount of warranty premium a customer must pay.

Adhikari et al. (2020) used a nonparametric Bayesian approach on patients undergoing PCI to investigate the heterogeneous effects of different coronary arterial access site strategies. The present study aimed to estimate heterogeneous effects of the radial arterial access and the femoral arterial access strategies between the female and males relying on hospitalization charges. The patients undergoing PCI were selected at least eighteen of Age.

The researchers used a potential continuous non-normal distribution outcome. Metropolis hasting within a Gibbs Markov chain Monte Carlo (MCMC) algorithm was used to sample from the posterior distribution of the parameters given observed data. Under latent index modeling, a hierarchical Bayesian likelihood relying on instrumental variable analysis is shown to jointly model outcome and treatment status. Unobserved heterogeneity is measured due to latent factor structure. Further, the Dirichlet process mixture model permitted nonparametric error distribution.

The researchers found that hospitalization charges were decreased from artery access in the wrist compared to artery access in the groin, with a higher decrement for male patients.

1.5 Layout

This thesis includes four chapters as follows:

Chapter one: Provides an introduction, variable explanation, literature review, objective of the thesis, and layout of the thesis.

Chapter two: Includes an overview of classical and Bayesian logistic regression

Chapter three: Contains the descriptive data analysis, checking for Multicollinearity, model building, and results of classical and Bayesian logistic regression models, as well as a discussion of the main findings.

Chapter four: Provides the conclusion and main recommendations of the study.

CHAPTER TWO

METHODOLOGY AND RESEARCH DESIGN

CHAPTER TWO

2. METHODOLOGY AND RESEARCH DESIGN

2.1. Generalized linear model

An efficient expansion of the fundamental Normal Linear Model is provided by Generalized Linear Models (GLMs). Examples of data that can be directly described as Bernoulli, Binomial, or Poisson distributions are data regarding binary analysis, proportions, or counts. The data do not require adjustments, and approaching Normality is not assumed. This enables the outcome variable not to have a Gaussian distribution (normal distribution). Two presumptions concerning the distribution of the response outcomes are necessary for GLMs. First, observations of the response variable y_i are presumed to be independent of one another given the explanatory variable. Second, the y_i distribution has to belong to the exponential family. The exponential family's general form is as follows:

$$f(\theta_i, \phi_i, \omega_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi}\omega_i + c(y_i, \phi_i, \omega_i)\right) \text{ ... } 2.1$$

Where is known as the dispersion parameter, is the canonical parameter, is a weight depending on whether the data are grouped, and b and c are defined as a function of the exponential family distribution types (Dobson & Barnett, 2018) and (McCullagh & Nelder, 2019).

2.2. Logistic Regression

One of the essential topics in regression analysis, which describes the shape of the relationship between one variable as from Figure 1, its shape is seen as S-Shape, and most of the time, it is recognized as S-Shape, the dependent variable, and one or more variables, the independent variables, by determining a mathematical equation that connects these variables. This model can be linear, in which case it is referred to as a linear regression model, or non-linear, as is the situation with most natural phenomena, in which situation it is referred to as a non-linear regression model, in which the dependent variable's values affect the quality of the regression model, which can be used. We can utilize sorts of regression models when the values are quantitative, but the logistic regression model is suited for this type of data when the values are descriptive (binary, multiple, monotonous) (Noel, et al., 2020).

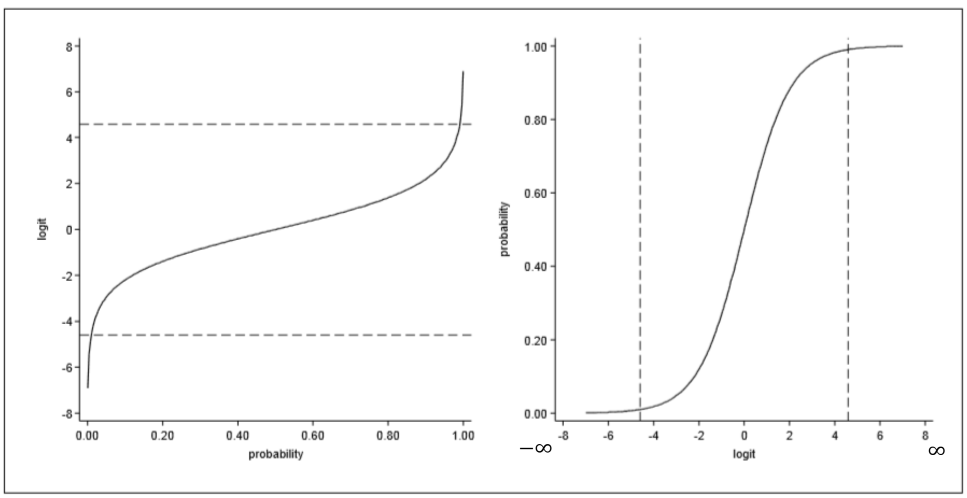

Fig. 2.1 Logit and inverse logit functions

Before delving into the logistic regression model in depth, it is critical to recognize that the purpose of any analysis using.

this model is the same as any other regression model used in statistics. The goal is to develop the best fitting and most efficient, clinically explainable model to represent the connection between a result (dependent or responder) variable and a group of independent (predatory explanatory) factors. The factors are frequently referred to as covariates. The most famous example of modeling, which readers of this article are likely to be familiar with, is the classic linear regression model in which the outcome variable is considered continuous (Le & Eberly, 2016).

The outcome variable in logistic regression is binary or dichotomous, which separates it from the linear regression model. The model's design and assumptions represent this distinction between logistic and linear regression. Once this difference is considered, the procedures utilized in a logistic regression study follow the same fundamental principles as linear regression. Thus, the approach to logistic regression is motivated by the techniques utilized in linear regression research (Jr, et al., 2013). For analyzing a dichotomous outcome variable, many distribution functions have been proposed. Some of them are discussed by Cox and Snell (1989). There are two main reasons for selecting logistic distribution. First and foremost, from a mathematical standpoint, it is very versatile and straightforward to use the function. Second, its model parameters are the foundation for clinically significant impact estimations. We utilize the quantity to simplify the country ($\frac{\pi(x)=E(Y)}{x}$) when the logistic distribution is employed to express the conditional mean of Ygiven, X the logistic regression model that we apply is of the following type:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_i x}} \quad i = 1,2,\dots,n \quad \dots\ 2.2$$

The logit transformation is a transformation of $\pi(x)$ that is fundamental to our study of logistic regression. This transformation is expressed in terms of $\pi(x)$ as follows:

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \dots 2.3$$

$$= \beta_0 + \beta_i x$$

This modification is significant because it exhibits several desirable qualities of a linear regression model. First, the logistic, g(x), has linear parameters that might be continuous or range from to $-\infty$ to $+\infty$ depending on the range of x. The conditional distribution of the outcome variable is the second significant difference between the linear and logistic regression models. We consider in the linear regression model that an observation of the outcome variable may be written as:

$$y = E\left(\frac{Y}{x+\varepsilon}\right)$$

The error(ε) is a number that reflects an observation's divergence from the conditional mean. The most prevalent assumption follows a normal distribution with mean zero and variance constant across independent variable levels. As a result, the conditional distribution of the outcome variable given x is normal, with a mean $E(\frac{Y}{x})$ and a constant variance. With a dichotomous result variable, this is not the case. In this situation, we may demonstrate the value of the outcome variable given x as $y = \pi(x) + \varepsilon$. Here the quantity (ε) may assume one of two possible values. If $y = 1$ then $\varepsilon = 1 - \pi(x)$ with probability $\pi(x)$, and if $y = 0$ then $\varepsilon = -\pi(x)$ with probability $1-\pi(x)$.thus, (ε) has a distribution with a mean zero and variance equal to $\pi(x)[1-\pi(x)]$. That is, the conditional distribution of the outcome variable follows a binomial distribution, with the conditional mean determining the probability $\pi(x)$ (Jr, et al., 2013).

In summary, we have demonstrated that the result variable in a regression analysis is dichotomous:

1. The regression equation's conditional mean of the regression model must be confined between zero and one. This condition is satisfied by the logistic regression model $\pi(x)$ in equation (2.3).

2. The binomial distribution, not the normal distribution, represents the error distribution and is the statistical distribution used in the analysis.

3. The ideas that lead to linear regression analysis also guide us in logistic regression.

2.2.1. Fitting the Logistic Regression Model

Assume we have independent observations of the pair (x_i,y_i),$i = 1,2,\dots,n$, where y_i signifies the value of a dichotomous outcome variable and x_i represents the value of the independent variable for the ith individual. Furthermore, assume that the outcome variable has been coded as a 0 or 1, indicating the presence or absence of the attribute, accordingly. This dichotomous outcome coding is utilized across the text. To fit the logistic regression model in equation (2.2) to a collection of data, we must first estimate the values of the unknown parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ (Kleinbaum, et al., 2002).

If Y is coded as 0 or 1, then the interpretation for (x) in equation (2.2) offers the conditional probability that Y is equal to 1 given x (for an arbitrary value of $\beta = (\beta_0, \beta_i)$, the vector of parameters). This is denoted by the symbol (x). It describes that the quantity $1{-}\pi(x)$ given the conditional probability that Y is equal to zero given x, $pr(Y = 0/x)$ Thus, for those pairings (x_i, y_i), where $y_i = 1$, the contribution to the likelihood function $1 - \pi(x_i)$, where $\pi(x_i)$ reflects the value of

$\pi(x_i)$ calculated at x_i (Faraway, 2016). The equation is a useful approach to represent the pair (x_i, y_i) contribution to the probability function.

$$pr(Y_i = y_i) = \binom{n_i}{y_i} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{n_i - y_i} \quad \ldots 2.4$$

2.2.2. Estimation and Hypothesis Testing

The recently created logistic model is an extended linear model with binomial errors and link logistics. This section summarizes the most critical findings required by the applications.

2.2.3. Maximum Likelihood Estimation

Because logistic regression predicts probabilities instead of classifications, we can fit it using likelihood; for each modeling data point, we have a vector of feathers, x_i and an observed class was either $\pi(x)$, if $y_i = 1$ or $1 - \pi(x)$, if $y_i = 0$ (Czepiel, 2002). The likelihood is then:

$$L(\beta_0, \beta_i) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \quad \ldots 2.5$$

The log-likelihood method converts products into sums:

$$L(\beta_0, \beta_i) = \sum_{i=1}^{n} [\,(y_i \, log\pi(x_i) + (1 - y_i)log(1 - \pi(x_i))] \quad \ldots 2.6$$

$$= \sum_{i=1}^{n} \left[y_i log\ \pi(x_i) + y_i log \frac{\pi(x_i)}{\left(1 - \pi(x_i)\right)} \right] \ \ldots\ 2.7$$

$$= \sum_{i=1}^{n} \log(1) - \pi(x_i) + \sum_{i=1}^{n} y_i(\beta_0 + \beta x_i) \ \ldots\ 2.8$$

$$= \sum_{i=1}^{n} -\log(1) + e^{(\beta_0 + \beta x_i)} + \sum_{i=1}^{n} y_i(\beta_0 + \beta x_i) \ \ldots 2.9$$

Where we finally employ the equation in the next-to-last step (2.4)

Commonly, we would differentiate the log-likelihood concerning the parameters, set the derivatives to zero, and then solve to get the greatest likelihood estimates (Queen, et al., 2002).To begin, compute the derivative concerning one component of β, say β_j.

$$\frac{\partial l}{\partial \beta_j} = -\sum_{i=1}^{n} \frac{1}{1 + e^{(\beta_0 + \beta x_i)}} e^{(\beta_0 + \beta x_i)} x_i + \sum_{i=1}^{n} y_i x_{ij} \quad \ldots\ 2.10$$

$$= \sum_{i=1}^{n} \left(y_i - \pi(x_i; \beta_0, \beta) x_{ij} \right) \quad \ldots\ 2.11$$

The method is analogous to repeatedly re-weighted squares (IRLS). For the procedure, see (Nelder & Wedderburn, 1972).

We cannot set this to zero and solve it precisely. (This is a transcendental equation with no closed-form solution). However, we may estimate the solution numerically.

2.2.4. Tests of Hypotheses

Consider the problem of hypothesis testing in logistic models. As is customary, we may compute Wald tests that are nearly normal with mean and variance-covariance matrix, as shown below.

$$var(\hat{\beta}) = (XWX')^{-1} \quad \dots\ 2.12$$

W is a weighted diagonal matrix with entries. Look at this (Hosmer, et al., 2013)

For example, we can put the hypothesis to the test.

$$H_0: \beta_j = 0$$

Concerning the importance of a single coefficient, the ratio of the estimate to its standard error is calculated;

$$Z = \frac{\hat{\beta}_j}{\sqrt{v\hat{a}r(\hat{\beta}_j)}} \quad \dots\ 2.13$$

In large samples, this statistic has a conventional normal distribution. Therefore, we may also consider the square of this statistic to be about a confidence interval for β_j can be calculated using the Wald test. We can say with $100(1-\alpha)$ percent confidence that the correct parameter is located in the interval with bounds.

$$\hat{\beta}_j \pm z_1 - \frac{\alpha}{2}\sqrt{v\hat{a}r(\hat{\beta}_j)} \quad \dots\ 2.14$$

Confidence intervals for impacts in the logit scale may be transformed into confidence intervals for odds ratios by exponentiating the bounds, where $(z_1 - \frac{\alpha}{2})$ is the typical critical value for two-sided tests of size α.

2.2.5. Model selection

Rather than R^2, the deviation is used as the metric for overall model fit in logistic regression. Chi-square is claimed to be a measure of the "**goodness of fit"** similarly, we utilize Chi-square as a measure of model fit here. It is the relationship between the actual value (Y) and the predicted value (Y). Pigeon and Heyse (1999) demonstrate that the more significant the difference (or "deviance") between actual and anticipated values, the worse the model's fit. Therefore, if at all feasible, we desire a minor deviation. The deviation should decrease as we add additional variables to the equation, showing a modeling in fit (Queen, et al., 2002).

Assume we have just fitted a model and want to see how well it matches the data, and it is computed as follows:

$$D = 2\sum_{i=1}^{k}\left[y_i\ log\left(\frac{y_i}{\hat{y}_l}\right) + (n_i - y_i)log(\frac{n_i - y_i}{n_i - \hat{y}_i})\right] \quad \sim x^2{}_{(n-p)} \quad ...2.15$$

Once the deviation for all models has been computed, the next stage is to see if the model with fewer variables outperforms the complete model. Statistical methods that can aid the procedure have been applied in this case. At various stages of the method, variable predictors are added and removed from a baseline or saturated model. In logistic regression, numerous strategies can be used, including stepwise, forward, and reverse selection (Field, et al., 2012).

This section covers all steps of model selection and the fitting of appropriate models. As previously stated, the forward selection approach was used to choose a relevant model to understand the link between the result and other independent variables. As a result, we will use a likelihood ratio test and the Akaike information criterion (AIC) to see if the model fits this process. We started with a simple null model (intercept alone) and worked our way up to more complicated models (Pan, 2001).

The first statistical computation, the likelihood ratio test, was performed to test the assumption of

H_0: The reduced model is superior to the entire model

H_1: The full model is superior to the reduced model

The process allows for the comparison of two models. The first model (reduction model) does not contain an extra predictor variable, whereas the second model (full model) adds an extra parameter and tests the importance of an extra predictor variable. Moreover, the AIC test benefits "automatically including a penalty for over-fitting". A test has also been employed since it allows the comparison of models that are not nested (Agresti, 2018).

In addition, we may derive the AIC test, which is represented by:

$$AIC = 2\big(K - Ln(L)\big) \quad \dots\ 2.16$$

Where K the number of estimated parameters in the model and L is is a maximum value of the likelihood function for the model.

2.2.5.1. Forward (Step-Up) Selection:

When there is an extensive collection of variables, this approach is frequently used to offer an initial screening of the potential variables. For example, assume you have fifty to one hundred variables, which is much outside the scope of the feasible regressions technique. A fair strategy would be to employ this forward selection technique to acquire the top ten to fifteen variables, after which the all-possible algorithm would be applied to the variables in this subset. When Multicollinearity is an issue, this technique is an excellent choice. The forward selection technique is straightforward to define. The model starts with no candidate variables. Choose the variable with the greatest R-Squared. At each phase, choose the candidate variable with the most significant effect on R-Squared. When none of the remaining variables is significant, stop adding variables. It is important to note that once a variable is added to the model, it cannot be removed.

2.2.5.2. Backward (Step-Down) Selection:

This strategy is less common since it starts with a model that includes all candidate variables. Because it works its way down rather than up, you always keep a high R-Squared value. The issue is that the models chosen through this approach may include variables that are not indeed required. The user determines the significance level at which variables can be introduced into the model.

The backward selection model begins with all of the model's candidate variables. At each phase, the least significant variable is eliminated. This step is repeated until no nonsignificant variables are left. Finally, the user specifies the degree of significance at which variables can be eliminated from the model.

2.2.5.3. Stepwise selection:

It is a mix of forwarding and backward selection approaches. Stepwise selection is a forward selection process in which predictors can be deleted using backward selection at each stage of refining the model. Predictors included early in the process can be removed later, and vice versa. Stepwise regression is a forwarding selection variation in which all model candidate variables are examined after each step to determine whether their significance has decreased below the stated tolerance threshold. If a variable is discovered insignificant, it is eliminated from the model. Stepwise regression necessitates two levels of significance: one for adding variables and one for eliminating variables. To avoid an endless loop, the cutoff probability for adding variables should be less than the cutoff probability for deleting variables. In stepwise selection. Significance values larger than 0.05, or small F statistics, are frequently suggested (and are the default settings in most regression software's stepwise selection processes) because they result in more predictors remaining in the model and lower the chance of deleting crucial variables. However, as is customary, the selection of significance levels is arbitrary (Queen, et al., 2002).

2.2.6. Model Diagnostics

In addition to evaluating the overall fit of the model, it is critical to analyze the contribution of each observation, or set of data, to the fit and deviations from the fit. The significance of residuals in OLS linear models are stressed, the difference between each observed and fitted or forecasted value (Queen, et al., 2002). It is commonly denoted as a standardized residual (e_i):

$$e_i = \frac{y_i - n\hat{y}_i}{\sqrt{[n_i\hat{y}_i(1-\hat{y}_i)]}} \quad \text{... } 2.17$$

Then plot it with the predictors to determine any outliers; if there are, suggest refitting the model.

2.3. Bayesian Inference

Bayesian statistics require a significantly different method of considering statistical Inference when compared to the traditional school, such as confidence intervals, P-values, hypothesis testing, etc. A significant difference between the Bayesian framework and frequentist lies in introducing the prior information in the framework of probability distributions (Dunson, 2001). The prior distribution gives a summary of everything that is obtained about parameter θ, except the data, moreover, in the Bayesian approach; conclusions are normally reached by the probability statement, i.e., parameter estimates cannot be indicated as point estimates only instead are statistical distributions (Dunson, 2001).

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \quad \text{... 2.18}$$

The use of Bayesian methods has significantly increased in statistical analysis and has been applied to a wide range of study areas and scientific research. Bayesian data analysis includes analyzing statistical models by integrating prior knowledge about the model parameters (Spiegelhalter, et al., 2002). Furthermore, indicating the model for the identified quantities $Y = y_1, y_2, \ldots, y_k$ is given a vector of unknown parameter θ normally in the form of a probability density function $p(y|\theta)$, where it is supposed that θ is a random quantity, having a prior distribution $p(\theta)$. Therefore, inference about θ is based on its posterior distribution $p(y|\theta)$, given by:

Where, $p(y) = \sum p(\theta)p(y|\theta)$ is a discrete random variable, and $p(y) = \int p(\theta)p(y|\theta)d\theta$ in the continuous covariates. Equation (2.18) can then be written in a proportional form

$$p(\theta|y) \propto p(\theta)p(y|\theta) \qquad \text{... 2.19}$$

All Bayesian estimates peruse from this posterior distribution because it contains all the relevant information about the parameters.

2.3.1. The Likelihood

The likelihood $p(y|\theta)$ can describe every particular value of the parameter $\theta \in \boldsymbol{\Theta}$. The data prediction shape is formed when the parameter has a particular value given by θ. The likelihood function is normally indicated as $I(\theta; y)$. It is the observed data with a function of θ serving as parameters of that function. Statisticians who use the frequentist method will normally select the value that gives them the maximum likelihood estimation of θ.

To have the likelihood function, an appropriate probability distribution should be selected. This selected likelihood distribution should match the sort of data that are detected (seen) and should have the ability to cover the estimated model within it. It should be able to discredit the estimated model when it is contradictory to the evidence. The likelihood should not be seen as fixed; instead, it is presented to explore variations in the Inference over a group of likelihoods, representing the hypothesis. It is suggested that it is preferred when the selected likelihoods are relatively unlimited (Lancaster, 2004).

2.3.2. The Prior Distribution

The challenge of Bayesian formulation frequently existed in complex equations for posterior until the modern computer age emerged. To avoid the challenge named conjugate priors, the distribution of the priors can be accurately selected to perform well with the likelihood function. However, over the past few decades, methodology and software have proven that the posterior distribution function can be directly sampled by using simulation techniques. The current challenge with the Bayesian approach is a description of the prior distribution, and choosing a proper prior is probably considered the most indispensable aspect of Bayesian modeling (Congdon, 2005).

The prior distribution is a significant part of the Bayesian method and constitutes (represents) information about an unclear parameter θ, which is integrated with the recent data to acquire the posterior distribution. Furthermore, it is used for making decisions involving θ and future Inference (Gelman, et al., 2008). Therefore, great care should be taken when selecting priors and must be backed (supported) well by providing documentation because unsuitable selections regarding prior may result in inaccurate posterior Inference.

The presence of the prior distribution has a strength and a weakness of the Bayesian approach; it has strength because it permits information beyond the observed data at hand to be utilized in making inferences, and it has a weakness because the Inference inevitably relays, at least to some extent, on the selection of prior. For example, (Gelman, 2002) demonstrates that having well-identified parameters and large sample sizes, fine selections of prior distributions will slightly affect posterior inferences. This denotes that with the availability of extensive data, the effect of the prior will sound insignificant, providing the exact outcomes as entirely data-driven inference.

2.3.3. The Posterior Distribution

The posterior distribution reflects the information in someone's belief about the parameter (prior distribution) and the likelihood function (observed data). The posterior is considered the conclusion of the empirical analysis in numerous applications. Illustrating the results of the empirical analysis requires a demonstration of the posterior distribution, to which the model and observed data have led and can be carried out in numerous ways (Lancaster, 2004).

- Drawing it: the preferable way of illustrating the content of the posterior distribution, when it is a scalar, is by drawing it. This is utilized as well when θ is a vector; however, the interesting parameter is a one-dimensional function of θ.
- Illustrating its moments: notifying an estimation of θ together with an estimation of the standard deviation of its repeated sampling is regarded as a traditional method. It often notifies a confidence interval θ. The Bayesian method attempts to find an interval in Θ, from the posterior distribution, such that a probability of 0.95θ lies within it.
- Calculating the marginal, from this mathematical perspective, is considered challenging due to calculation involvement in formatting the posterior of the parameter of interest. Fortunately, two solutions are available; the first is using approximations to posterior distributions. The second solution is to use the method of computer-aided sampling.

2.4. Bayesian logistic regression

A logistic equation is used in Bayesian logistic regression, which utilizes Bayesian Inference and contains both continuous and categorical predictor variables. Furthermore, the shapes of the explanatory covariate coefficients are compatible with other linear models thanks to the logistic equation's transformation of categorical response variables into logarithmic forms.

The logistic equation has the following generic form:

$$g(x) = logit\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta^T x \quad \dots 2.20$$

In Bayesian logistic regression, parameter estimates are obtained using Bayesian Inference, a practical way for defining the probability distribution for explanatory variables, as opposed to maximum likelihood approaches used in classical logistic regression β. Bayes' Rule, which encapsulates the essence of Bayesian inference, is used to obtain a posterior probability distribution for these unknown parameters β.

Bayesian Inference begins by formulating a prior probability distribution over the unknown parameters β, which summarizes a set of beliefs of knowledge in hand before any observations are taken. Since it does not assume a predictor variable distribution, it is a flexible technique. An indicator variable called Y_i is introduced in the chapter one (takes 1 if there is CABG surgery and 0 otherwise).

Then:

$$Y_i \sim Bernoulli(p_i) \qquad\qquad i = 1, \dots, N$$

The linear predictor η_i is:

$$\eta_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} \quad \dots 2.21$$

This is connected to the fitted probability through the logit function,

$$logit\left(\frac{p_i}{1-p_i}\right) = \eta_i \quad \dots 2.22$$

The Bayesian logistic model is formulated by specifying prior distributions on the logistic regression coefficients:

$$\boldsymbol{\beta_j} \sim p(\theta_j) \qquad\qquad j = 1, \ldots, K$$

This $\beta_j$in Bayesian method is directly derived from the posterior distribution of the unknown parameters. For Bayesian estimates, the moments, most excellent posterior density region, and quartiles are good aspects of the posterior probability distribution (Wilhelmsen, et al., 2009).

Between Bayesian Inference and traditional statistics, there are three key differences:

1. Bayesian Inference is based on the concept of prior probability, which is absent from classical statistics.
2. There are subtle differences between Bayesian reasoning and traditional statistics regarding how conclusions are drawn and interpreted. For example, traditional statistics fit a distribution to the observed data and use that information to estimate the parameters. Beginning with a proper distribution of the parameters, Bayesian Inference creates an updated posterior distribution of the parameters based on the previous assumptions.
3. Whereas Bayesian reasoning use credibility intervals, classical statistics constructs confidence intervals.

A confidence level is established by managing the $Type\ I$ (or alpha) error level, and the confidence interval for the parameters based on that confidence level is then created. The credibility interval of parameters can be obtained to have a natural meaning in terms of probabilities because the posterior probability of parameters is calculated in Bayesian Inference (Genkin, et al., 2007).

2.4.1. MCMC methods

MCMC methods focus on simulating direct draws from some complicated distribution of interests, usually a posterior distribution. MCMC approaches are so-named as one employs the previous sample values to produce the next sample value, producing a Markov chain. Initiated in the early 1990s (Gelfand & Smith, 1990), a particular MCMC method, the Gibbs sampler, is very extensively applicable to a broad class of Bayesian problems, and it has sparked a significant rise in the application of Bayesian analysis and this interest may continue expanding for some time to come. MCMC methods have their origins in the Metropolis algorithm (Metropolis, et al., 1953); An attempt by physicists to compute complicated integrals by considering them as expectations for some distribution and then evaluating this expectation by drawing samples from that distribution. The Gibbs sampler mentioned above (Geman & Geman, 1984) has its roots in image processing.

Markov Chain Monte Carlo (MCMC) approaches play a vital role in Bayesian statistics due to their simulation-based function. MCMC permits inferences to be formulated from rather complicated complex posterior distributions where methods such as numerical or analytical integration seem very difficult to be applied (Sorensen, et al., 2002). The primary application of the MCMC methods is to formulate Markov Chain (sample) with the aid of iterating Monte Carlo simulation for estimation features of probability distributions that seem impossible to be applied analytically. Hence, this sample is essential for estimating various distribution properties, for example, quintiles, modes, and other statistical issues.

2.4.1.1. Gibbs Sampler

Gibbs sampling, proposed by (Geman & Geman, 1984), is a specific case of the Metropolis-Hasting algorithm within Gibbs sampling, a sample (random number or draws) is generated from all conditional distributions, hence sampling the posterior joint distribution. This is simple if we can ascertain that each of these distributions is one of the well-known distributions that we can sample from it.

The Gibbs sampler aims to apply a distribution for a random variable by iteration conditioning on impermanent (temporary) starting values of the others on a continuing cycle until samples from this process empirically approximate the desired marginal distribution (Gill, 2014). Thus, random samples can be drawn from a conditional distribution for θ_1 and usage θ_1 to draw θ_2 from the conditional distribution and so forth. The iterative nature of the Gibbs sampling algorithm can be simplified by the requirement, which cycles during these full conditionals drawing parameter values based on the most recent version of all of the parameters already in the list. The order is not necessary; however, the utilization of the most recent draw from the other samples is substantial.

Let us consider that it is intended to draw samples for the set of random variables for the marginal posterior $\theta_1, \theta_2, \ldots, \theta_k$, but the marginal distribution cannot be obtained from the joint posterior analytically. However, the full conditional distribution, which can easily be sampled, is available. The procedure can be outlined as follows:

1. Determine initial values for the parameters θ, i.e., $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots);$
2. Following this, draw from each complete conditional distribution consecutively, employing the current updated values achieved from the earlier and recent steps. That is, for i = 1,..., N: y = $(y_1, \ldots, y_n)$

(a) Draw $\theta_1{}^{(i)} \sim \pi(\theta_1|\theta_2{}^{(i-1)}, \theta_3{}^{(i-1)}, \ldots, \theta_p{}^{(i-1)}, y)$

(b) Draw $\theta_2{}^{(i)} \sim \pi(\theta_2|\theta_1{}^{(i)}, \theta_3{}^{(i-1)}, \ldots, \theta_p{}^{(i-1)}, y)$

$\downarrow$

(c) Draw $\theta_p{}^{(i)} \sim \pi(\theta_p|\theta_1{}^{(i)}, \theta_2{}^{(i)}, \ldots, , y)$

3. Repeat step 2 until your chain n converges.

When convergence is obtained, total simulation values are from the target posterior distribution, and a sufficient number has to be drawn to analyze all areas of the posterior. The significant feature throughout every iteration of the cycling of the vector θ is conditioning on the values of θ that have already been sampled for that cycle; otherwise, the last cycle can take the θ values. As a result, the sample value for the k parameter has to be examined on all the i-step values in the final stage of a given cycle (Gill, 2014). The statements above show that having the complete set of conditional distributions is needed to run the Gibbs sampling algorithm. Sometimes running Gibbs sampling could be very inadequate. However, it decreases multidimensional concerns to a chain of univariate conclusions by considering parameters one by one, and as a result, it will be uncomplicated for the program (Casella & George, 1992). If the Gibbs sampler has been put to run for a relatively long period, a whole sample of the values in the θ vector will be generated by a complete cycle of the algorithm.

2.4.1.2. The Metropolis-Hasting Algorithm

One problem with implementing Monte Carlo integration is achieving samples from some complicated probability distribution $p(x)$. Attempting to tackle this problem is at the roots of MCMC methods. In particular, mathematical physicists integrate very complicated functions by random sampling (Metropolis, et al., 1953) and (Hastings, 1970) the resulting Metropolis-Hastings algorithm traced to attempts. Chib & Greenberg (1995) provided a clear and distinct review of this method. We aim to extract samples from some distribution $p(\theta)$, where $p(\theta) = f(\theta)/k$, where the normalizing constant k is unlikely to be known and difficult to compute (because it involves intractable integrals). The Metropolis algorithm (Metropolis, et al., 1953) results in a sequence of draws from this distribution as follows:

1. Use any initial value θ_0 satisfying $f(\theta_0) > 0$.

2. Working with the current θ value, sample a candidate point θ^* from jumping distribution $q(\theta_1, \theta_2)$ (Robert, et al., 2010), which is the probability of returning a value of θ_2 given a previous value of θ_2. This distribution is also denoted as the proposal or candidate-generating distribution. The only limitation on the jump density in the Metropolis algorithm must be symmetric, that is, $p(\theta_1, \theta_2) = q(\theta_1, \theta_2)$.

3. With the candidate point being θ^* , calculate the ratio of the density at the candidate θ^* and the current $\theta_{(t-1)}$

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})} \quad \ldots 2.23$$

Since we are looking at the ratio of $p(\theta)$ under two different values, the normalizing constant k cancels out.

4. If the jump causes a rise in the density $p(\alpha > 1)$, accept the candidate point $(\theta t = \theta^*)$ and return to step 2. If the jump causes a fall the density $p(\alpha <$

1) then with probability α accept the candidate point, else reject it and go back to step 2.

We can outline the Metropolis sampling as first computing and then accepting a candidate point with probability α (the probability of a move). This produces a Markov chain $(\theta_0, \theta_1, \ldots, \theta_k)$, as the transition probabilities from θ_t to θ_{t+1} rely only on θ_t and not $(\theta_0, \ldots, \theta_{t-1})$. Then, following an adequate burn-in period of k step, the chain reaches its stationary distribution, and samples from the vector $(\theta_{k+1}, \ldots, \theta_{k+n})$ are samples from $p(x)$.

2.4.2. MCMC Convergence

The Markov chain must converge to utilize the MCMC iterations to obtain an adequate representation of the actual posterior distribution. Obtaining convergence poses a significant implementation challenge connected to any MCMC-based approach. This means that the MCMC outputs no longer rely on the condition of the chain. This measure for utilizing MCMC approaches to investigate samples $f(.)$ can be summed up as follows:

1. Produce a Markov chain $(\theta_1, \ldots, \theta_k)$ employing the Gibbs sampling algorithm
2. Wait for the Markov chain to reach convergence; assume this happens at a time T
3. The values $(\theta_{T+1}, \ldots, \theta_{T+k})$ in the Markov chain can then be a sample from $f(.)$

The period before T is defined as burn-in, as the Markov chain obtained convergence; regrettably, how T is calculated cannot be calculated theoretically; instead, informal diagnostics can be utilized to determine when the chain has obtained convergence. Various diagnostic approaches have been presented in the literature to monitor the convergence of MCMC chains. The most familiar and most usable diagnostic methods can be briefed as follows.

2.4.3. Graphical Inspection

Graphically examining a trace of iterations is a widely used approach for observing the convergence of MCMC to estimate whether the sequence is mobilizing around the parameter space or is stuck somewhere. Informal convergence checks are based on the visual method first introduced by ((Gelfand & Smith, 1990). They run several independent chains with variable starting values, and each parameter of interest is analyzed.

“Trace Plot" is a primary tool for observing and mixing the chain's behavior. Samples of the parameter of interest can be diagrammed as the use of iterated numbers from the duplicated chains started from over-scattered values. Naturally, a wavy pattern proposes strong auto-correlations with the chain; at the same time, zigzag patterns demonstrate that the parameter proceeds more independently (Sorensen, et al., 2002).

The plot of MCMC iterations should resemble white noise, and the auto-correlational plot function must be minimized immediately to zero when the MCMC iterations are satisfactorily long. Therefore, in such situations, it can be claimed that convergence has been obtained. Moreover, different starting values can regularly be utilized to formulate numerous MCMC chains. It could be concluded that MCMC simulation obtained its stationary convergence; if overlap happens in the trace plot of chains, then a proper burn-in all result in the estimations for the accurate posterior distributions.

CHAPTER THREE

STATISTICAL RESULTS AND INTERPRETATIONS

CHAPTER THREE

3. STATISTICAL RESULTS AND INTERPRETATIONS

3.1. Introduction

In this research, 100 adult patients underwent coronary angiography at Erbil Cardiac Center, and their data was obtained. Patients were followed up from January 2016 until the end of December 2020. Data was also obtained from 60 healthy people who underwent the same coronary angiography.

This chapter contains the primary explanatory analysis of the Bypass Graft surgery data obtained from patient and control cases. In addition, it provides in-depth checking for Multicollinearity among predictor variables. Model building using logistic regression is also included in this chapter, in which the final model is selected.

This chapter also provides the main results of predictor variables affecting Bypass Graft surgery by applying classical and Bayesian logistic models. The data analysis is conducted using statistical software of **R** language and **STATA** for classical logistic regression, **WinBugs** for Bayesian logistic regression, and **Jamovi** for generating graphs.

3.2. Explanatory Analysis

Constructing a descriptive analysis for the data and examining how the variables connect to the outcome variable is preferable. Table 1 indicates that some predictor factors and the dependent binary variable may relate to the test value. We first demonstrate the continuous variables by to control and patient groups, and there is a noticeable difference in Age between the patient and the control group. The result shows that the mean Age for the control group is 44.87 years old, while the patient people tend to be 59.51, which tells us more into Age, more chance to be suffering, and the P-value of the T-test can confirm that with a value less than 0.05.

Table 3.1 T-test result and descriptive statistics for the study variables

Variables	Control	Patient	T-test[ƚ]	P-value
	Mean ± SD	Mean ± SD		
Age	44.87 ± 14.31	59.51 ± 8.51	-8.12	0.0000
Blood Sugar	105.87 ± 14.97	180.08 ± 99.57	-5.73	0.0000
HbA1c	5.07 ± 0.55	7.21 ± 2.25	-7.23	0.0000
Blood Urea	28.13 ± 5.44	36.88 ± 15.33	-4.26	0.0000
Creatinine	0.7 ± 0.27	0.899 ± 0.32	-4.05	0.0001
CHO	142.62 ± 27.19	160.61 ± 40.78	-3.04	0.0028
TG	140.55 ± 65.5	175.4 ± 82.32	-2.79	0.0059
HDL	39.02 ± 8.106	36.51 ± 10.03	1.64	0.1030
LDL	84.38 ± 21.6	100.57 ± 36.38	-3.13	0.0021
Weight	69.72 ± 12.56	77.87 ± 15.82	-3.4	0.0009
Length	1.67 ± 0.07	1.65 ± 0.1	1.59	0.1134
BMI	24.7 ± 3.09	28.61 ± 5.59	-4.98	0.0000
SBP	127.85 ± 10.61	128.99 ± 19.86	-0.41	0.6800
DBP	74.13 ± 6.09	78.1 ± 12.92	-2.23	0.0271

[ƚ] Independent sample T-test was performed for each variable with the 5% significant level.

Moreover, the high Blood Sugar value is more likely to put an individual at risk of heart problems, as from the initial output, the mean values are far from each other between the control and patient people, with 105.87 and 180.08, respectively. To support this, the driven T-test was found to be statistical significance since its P-value is less than 0.05

Another interesting result is BMI measurement, which is recorded at 24.7 for people in the control group, whereas a 28.61 mean value is found for the patient group. Based on this, the T-test reported that the mean values differed statistically significantly. On the other hand, HDL , SBP and Length were highlighted to be nonstatistically significant, with P-values (0.1030 and 0.6800) greater than 0.05, respectively. For Gender as well as smoking status.

Table 3.2 Cross-tabulation table of Chi-square for Gender and Smoking versus outcome variable

Variables		Control		Patient		Chi-Square	P-value
		N	%	N	%		
Gender	Male	34	56.67	64	64	0.84	0.0357
	Female	26	43.33	36	36		
Smoke	Yes	10	16.67	35	35	6.24	0.013
	No	50	83.33	65	65		

The percentage of males in the patient group is about double that of females with 64% and 36% constantly, according to Table 3.2, while in the control group, their differences are only 13.34%. This leads us to confirm that males are in the position to get an operation due to Bypass Graft concern, and the Chi-square test shows an association between these two variables. Similarly, smoking seemed to have a strong influence.

3.3. Check Multicollinearity

When building a model for the Bypass Graft variable using many covariates, we expect predictor covariates to be highly correlated to each other. When one or more covariates can be written as a linear combination of other variables in the model, then issues with the analysis of the data can happen. In such a case, we can say that there is an existence of Multicollinearity. The Multicollinearity issue can affect the parameter estimation variance, and hence inference about the relationship between the response variable and predictor variables can be incorrect (Midi, et al., 2010).

Usually, collinearity between predictor variables can be detected using a correlation matrix. However, it is challenging to detect collinearity between three or more variables. Ideally, there are two methods to deal with collinearity, which are principal component regression and ridge regression.

Collinearity between variable x_j and the other variables can be measured by the proportion of the variance in x_j explained by the other variables when x_j is regressed on the other covariates. This proportion can be obtained from the square of the multiple correlation coefficient, denoted by R_j. It can be shown that the variance of the regression coefficient β_j is proportional to $1/(1 - R_j^2)$, so large values of R_j^2 are associated with imprecise estimates of β_j. This is called the variance inflation factor (*VIF*).

So,

$$VIF = \frac{1}{(1 - R_j^2)} \quad \dots 3.1$$

Here the tolerance is the reciprocal of $1-R_j^2$. If covariate j is orthogonal to the other covariates, then its *VIF* will be 1. Hence, *VIF* provides a measure of the impact of

Collinearity on the variance of the estimated parameters. In this study, the value of *VIF* more significant than five is considered a collinearity issue (Belsley, 2014).

In this study, the collinearity issue occurred due to the methods used to collect the data, as some of the covariates belong to the same cluster of medical tests. In the first step, we calculated *VIF* including all variables in the model to detect Multicollinearity to build a logistic model for the Bypass Graft dataset. We have found that there are some covariates whose *VIF* values are higher than 5. The value of *VIF* above five is detected in covariates of Weight, Length, BMI, Cholesterol, LDL, HCT, HGB, TnT-Troponin, and CK-MB.

Table 3.3 VIF checking result before building models

Variable	VIF	Tolerance
Weight	143.72	0.0070
BMI	119.49	0.0083
Length	40.42	0.0247
CHO	17.29	0.0578
LDL	13.08	0.0764
HGB	11.92	0.0839
HCT	10.44	0.0958
TnT-Troponin	7.68	0.1303
CK-MB	7.46	0.1340
HbA1c	4.9	0.2039
Blood Sugar	4.37	0.2290
MCH	4.12	0.2430
TG	3.23	0.3092
MCV	3	0.3335
Gender	2.99	0.3347
Blood Urea	2.81	0.3558
Creatinine	2.8	0.3576
Eosinophil	2.47	0.4055
T3	2.4	0.4158
HDL	2.34	0.4270
SBP	2.32	0.4313
DBP	2.3	0.4340
GOT	2.09	0.4782

Variable	VIF	Tolerance
T4-2	2	0.4996
GPT	1.95	0.5138
Age	1.92	0.5209
TSH	1.91	0.5228
RDW	1.89	0.5283
Smoke	1.87	0.5350
T4-1	1.84	0.5424
ALP	1.74	0.5735
TSB	1.73	0.5772
WBC	1.63	0.6131
PDW	1.59	0.6284
MPV	1.39	0.7176

We have calculated a correlation matrix among covariates with high *VIF* to investigate the relationship among them. In Table 3.4 and correlation heatmap indicate whether there is a statistically significant relationship between variables. One remedy for dealing with Multicollinearity is removing variables that make the issue. Thus, we have removed covaries(s) related to other variables. Weight, Length, LDL, HCT, and CK-MB were deleted because they are correlated with other variables.

Table 3.4 Correlation matrix result

	Weight	BMI	Length	CHO	LDL	HGB	HCT	TnT-Troponin	CK-MB
Weight	1								
BMI	0.8378*	1							
Length	0.4222*	-0.131	1						
CHO	0.0536	0.0767	-0.0361	1					
LDL	0.0714	0.0596	0.02	0.9049*	1				
HGB	0.2848*	0.0688	0.3839*	0.0124	0.0132	1			
HCT	0.2313*	0.027	0.3600*	-0.086	-0.0637	0.9022*	1		
TnT-Troponin	-0.0134	0.0044	-0.0434	0.1333	0.114	0.1384	0.0779	1	
CK-MB	0.0242	0.0231	-0.0137	0.1093	0.0974	0.2013*	0.1612*	0.9048*	1

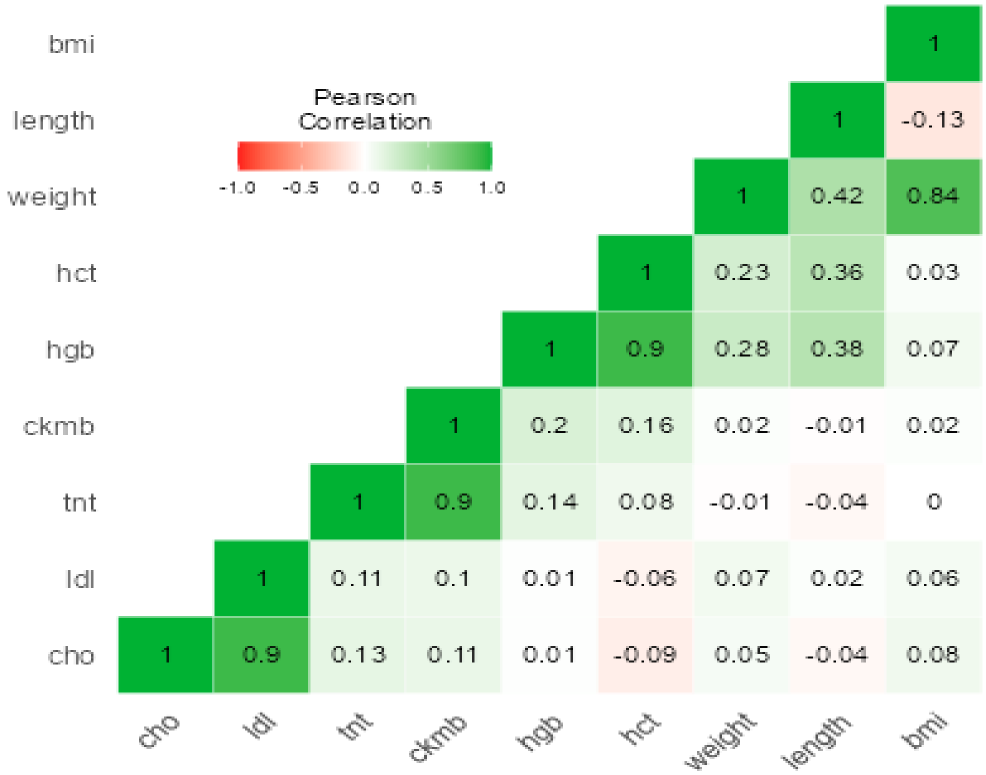

Fig. 3.1 Correlation Matrix using heatmap chart

Figure 3.1 displays the correlation values between the independent variables, and it can easily detect the highest correlation values, enabling us to remove them before fitting the model.

After removing variables that caused Multicollinearity among predictor variables, the *VIFs* for the remaining variable is calculated again. The Multicollinearity issue is solved, and the value of *VIFs* among all predictor variables is less than 5.

Table 3.5 VIF result of variables less than 5

Variable	VIF	Tolerance
HbA1c	4.73	0.2112
Blood Sugar	3.88	0.2580
MCH	2.96	0.3384
Blood Urea	2.74	0.3654
Creatinine	2.71	0.3690
MCV	2.5	0.4007
Eosinophil	2.39	0.4178
CHO	2.31	0.4331
T3	2.28	0.4380
DBP	2.24	0.4455
HGB	2.22	0.4501
SBP	2.16	0.4638
Gender	2.03	0.4923
GOT	2	0.4991
T4-2	1.9	0.5276
TG	1.87	0.5339
GPT	1.83	0.5466
TSH	1.79	0.5580
Smoke	1.79	0.5592
RDW	1.79	0.5594
TnT-Troponin	1.78	0.5611
HDL	1.74	0.5747
T4-1	1.69	0.5924
Age	1.65	0.6046
TSB	1.65	0.6060
WBC	1.58	0.6333
PDW	1.52	0.6571
ALP	1.51	0.6618
BMI	1.47	0.6811
MPV	1.38	0.7248

3.4. Model Building

Likewise, in all other modeling methodologies, selecting variables should be checked to be included in the final model. Akaike Information Criterion (AIC) is used to assess model quality. The forward approach can be used in our case. There are no variables in the initial model. All variables are then considered for inclusion, and the one with the lowest P-value but only if it is lower than a specified threshold can be added to the model. When combined with the first variable, the variable that has the lower P-value is then taken into account once more and included in the model. The procedure continues until no additional variable significantly improves the model. We have run logistic regression separately for each variable.

Table 3.6 Logistic regression for each variable

One Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	211.7002		1.3356	
Surgery ,Blood Sugar (Model-1)	148.6203	63.0799	0.9539	0*
Surgery ,HbA1c (Model-2)	136.8805	74.8199	0.8805	0*
Surgery, Blood Urea (Model-3)	186.2500	25.4503	1.1891	0*
Surgery, Creatinine (Model-4)	193.6281	18.0721	1.2352	0.0001*
Surgery, CHO (Model-5)	202.4482	9.2520	1.2903	0.0028*
Surgery, TG (Model-6)	202.9819	8.7184	1.2936	0.059
Surgery, HDL (Model-7)	209.0382	2.6620	1.3315	0.103
Surgery, GOT (Model-8)	209.8275	1.8728	1.3364	0.1682
Surgery, GPT (Model-9)	211.6522	0.0480	1.3478	0.8275
Surgery, ALP (Model-10)	211.4190	0.2813	1.3464	0.5973
Surgery, TSB (Model-11)	209.7091	1.9912	1.3357	0.1747
Surgery, T3 (Model-12)	205.0563	6.6440	1.3066	0.0104*
Surgery, T4-1 (Model-13)	210.8751	0.8251	1.3430	0.3654
Surgery, T4-2 (Model-14)	210.9380	0.7622	1.3434	0.4659
Surgery, TSH (Model-15)	206.9239	4.7764	1.3183	0.0564

One Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	211.7002		1.3356	
Surgery, TnT-Troponin (Model-16)	124.2905	87.4097	0.8018	0.1132
Surgery, WBC (Model-17)	177.6650	34.0353	1.1354	0*
Surgery, Eosinophil (Model-18)	183.6272	28.0730	1.1727	0*
Surgery, HGB (Model-19)	211.4326	0.2677	1.3465	0.6075
Surgery, MPV (Model-20)	210.3991	1.3012	1.3400	0.2577
Surgery, MCV (Model-21)	210.9138	0.7864	1.3432	0.2577
Surgery, MCH (Model-22)	206.5077	5.1925	1.3157	0.0229*
Surgery, RDW (Model-23)	211.6285	0.0717	1.3477	0.7901
Surgery, PDW (Model-24)	205.8333	5.8670	1.3115	0.0351*
Surgery, PLT (Model-25)	179.5797	32.1205	1.1474	0*
Surgery, BMI (Model-26)	186.3761	25.3242	1.1899	0*
Surgery, Age (Model-27)	159.0299	52.6704	1.0189	0*
Surgery, Gender (Model-28)	210.8544	0.8459	1.3428	0.3598
Surgery, Smoke (Model-29)	205.1353	6.5649	1.3071	0.0123*
Surgery, SBP (Model-30)	211.5291	0.1711	1.3471	0.6819
Surgery, DBP (Model-31)	206.5860	5.1143	1.3162	0.0271*

It can be seen from the Table 3.6 that a model with HbA1c has the lowest deviance and AIC. Thus, we can build a model by including HbA1c. We have excluded all variables that are not significant.

Table 3.7 AIC and deviance of two predictor variables

Two Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	136.8805		0.8805	
Surgery, HbA1c, Blood Sugar (Model-2 and 1)	127.4812	9.3993	0.8343	0
Surgery, HbA1c, Age (Model-2 and 27)	100.1682	36.7122	0.6636	0
Surgery, HbA1c, WBC (Model-2 and 17)	118.5153	18.3652	0.7782	0
Surgery, HbA1c, PLT (Model-2 and 25)	126.8466	10.0339	0.8303	0
Surgery, HbA1c, Eosinophil (Model-2 and 18)	120.7973	16.0832	0.7925	0
Surgery, HbA1c, Blood Urea (Model-2 and 3)	126.0175	10.8630	0.8251	0
Surgery, HbA1c, BMI (Model-2 and 26)	131.5211	5.3593	0.8595	0
Surgery, HbA1c, Creatinine (Model-2 and 4)	124.9469	11.9336	0.8184	0
Surgery, HbA1c, CHO (Model-2 and 5)	132.303	4.5775	0.8644	0
Surgery, HbA1c, T3 (Model-2 and 12)	136.8602	0.0202	0.8929	0
Surgery, HbA1c, Smoke (Model-2 and 29)	129.0385	7.8420	0.8440	0
Surgery, HbA1c, PDW (Model-2 and 24)	133.2963	3.5841	0.8706	0
Surgery, HbA1c, MCH (Model-2 and 22)	134.1061	2.774	0.8757	0
Surgery, HbA1c, DBP (Model-2 and 31)	136.3911	0.4894	0.8899	0

In Table 3.7, we have included two predictor variables to see which is significant with the first model with only HbA1c. It can be noted that the variable Age has the lowest AIC and deviance. Therefore, we have two variables for the next model and assess the third covariate. We assessed which variable to include or exclude based on the AIC in each step of adding another variable to the model (see Appendix A:Table 1 to Table 7).

Table 3.8 AIC and deviance of ten predictor variables

Ten Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	22.1517		0.2634	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea (Model-5 and 1)	23.0267	-0.8750	0.1375	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, T3 (Model-5 and 2)	17.9683	4.1834	0.2498	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Creatinine (Model-5 and 3)	20.5828	1.5690	0.2661	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, PDW (Model-5 and 7)	22.0952	0.0566	0.2756	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, CHO (Model-5 and 6)	21.9496	0.2022	0.2747	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Smoke (Model-5 and 4)	22.0068	0.1449	0.2750	0

In Table 3.8 it can be seen that the tenth variable to be added into the final model based on the AIC and residual deviances is HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, and T3. Thus, the final model will include all those ten variables to predict Bypass Graft data. In the next model, we have found no differences between models containing 11 predictor variables. This is because all AIC and residual deviances are the same (see Appendix A: Table 8).

3.5. Main results

3.5.1. Binary Logistic regression

The dataset was subjected to multiple logistic regression analyses in order to determine the likelihood of some biographic, health, and other parameters having an impact on being in surgery. RStudio version 1.4.1717 was used to conduct the analysis. We first implemented the simple logistic regression plot for each of the variables individually with the curve plot, as seen in the figures below, to get a clue about the nature association among the response and covariates.

It is vital to spot some potential points here: Blood Sugar, HbA1c, WBC, Age, and Creatinine variables showed a relatively exciting link with the dependent variable. For other risk factors that appear but have no correlation with surgery have be mentioned in (Appendix B).

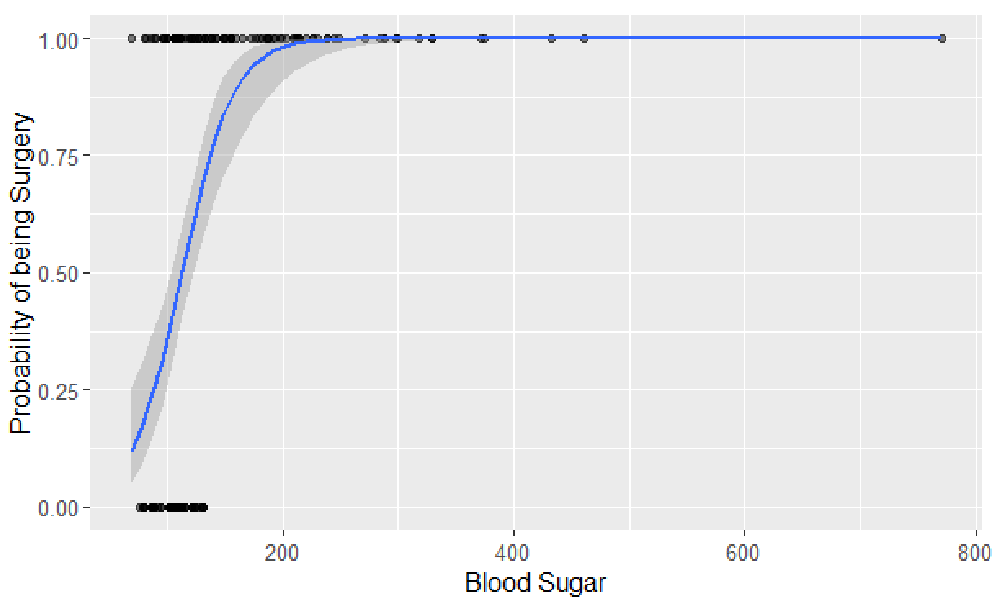

Fig. 3.2 Logit regression for variable of Blood Sugar

Figure 3.2 shows a clear positive relationship between the response variable and Blood Sugar.

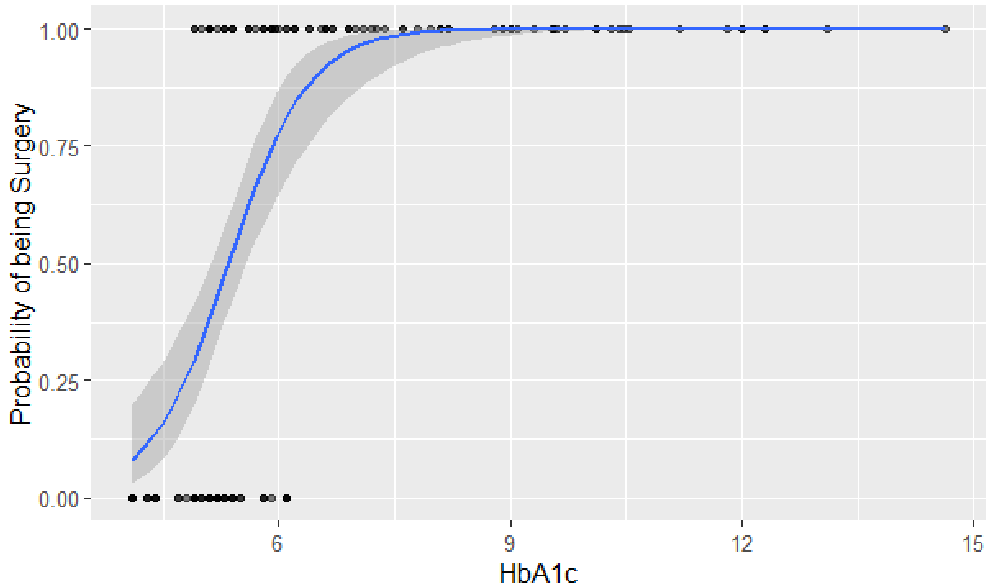

Fig 3.3 Logit model fit of surgery on HbA1c

Figure 3.3 shows a clear positive relationship between the response variable and HbA1c.

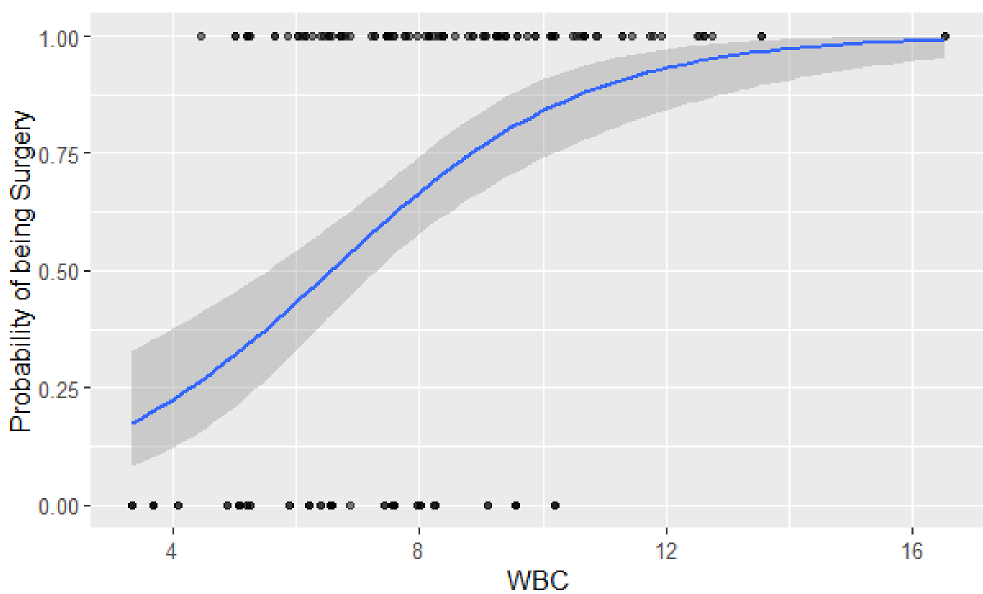

Fig. 3.4 Logit model fit of surgery on WBC

Figure 3.4 shows the positive relationships between WBC with the Bypass surgery variable.

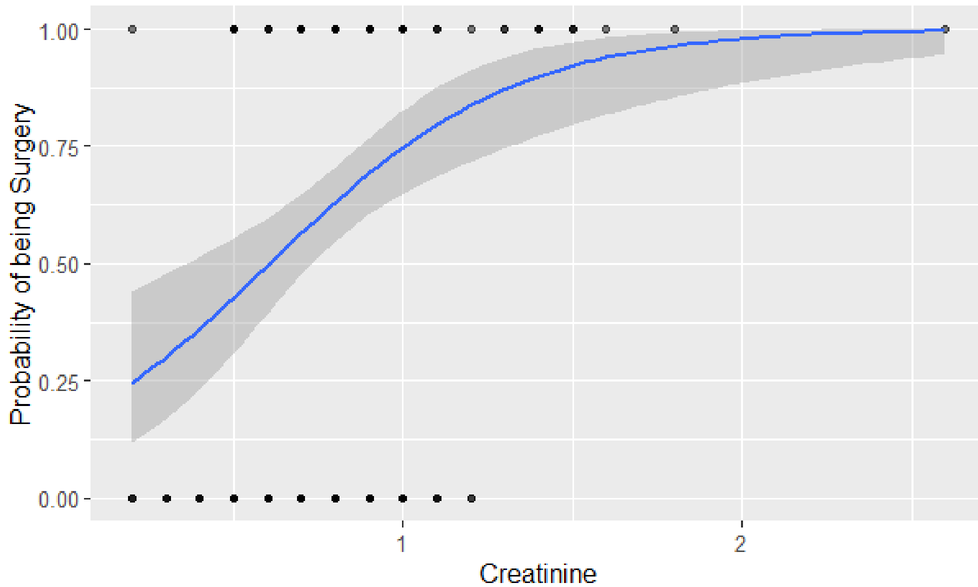

Fig. 3.5 Logit model fit of surgery on Creatinine

Figure 3.5 highlights the positive relationship between Bypass Graft surgery and the Creatinine variable.

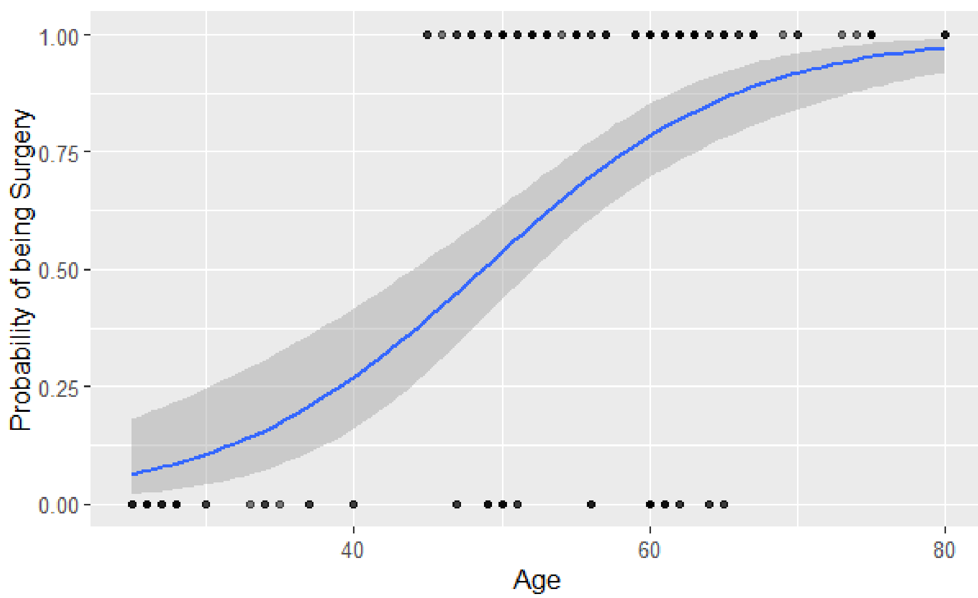

Fig. 3.6 Logit model fit of surgery on Age

Figure 3.6 highlights the positive relationship between Bypass Graft surgery and Age.

Considering the descriptive statistics and the Figures above, we can see that the final model has been taken into account with predictor variables DBP, Age, BMI, PLT, MCH, Eosinophil, WBC, HbA1c, T3, and Blood Sugar. Therefore, based on the visual inspection and model building, the final model equation can be as follow:

$$\rho = \frac{1}{1 + e^{-(0.513+0.302*DBP+0.959*\text{Age}+0.969*\text{MCH}+2.887\text{WBC}+0.265*Blood\ Sugar)}} \ldots\ 3.2$$

Table 3.9 Result of Estimated parameters, Odds Ratio in Logistic Regression

Variables	Odds Ratio	β_i	Std. Err.	Z-test	P-value	95% Conf. Interval	
						Lower	Upper
DBP	1.352	**0.302**	0.187	2.190	**0.029**	1.032	1.772
Age	2.608	**0.959**	1.182	2.110	**0.034**	1.072	6.341
BMI	2.722	**1.001**	1.703	1.600	0.109	0.799	9.276
PLT	0.952	**-0.049**	0.025	-1.900	0.057	0.905	1.002
MCH	2.636	**0.969**	1.155	2.210	**0.027**	1.116	6.223
Eosinophil	0.001	**-6.908**	0.001	-1.640	0.101	0.000	7.008
WBC	17.938	**2.887**	24.007	2.160	**0.031**	1.302	20.149
HbA1c	6.140	**1.815**	15.588	0.710	0.475	0.042	9.369
T3	0.001	**-6.908**	0.005	-1.600	0.109	0.000	4.348
Blood Sugar	1.303	**0.265**	0.155	2.220	**0.027**	1.031	1.646
Intercept	1.670	**0.513**	0.271	3.130	**0.002**	1.210	2.295

Table 3.9 indicates the parameter estimation of the final logistic model with its odds ratio for predictor variables. The significant variables contributing to the prediction of Bypass Graft operation are DBP, Age, MCH, WBC, and Blood Sugar.

The most effective attribute was WBC; as per the output, it shows a significant and positive impact on the response variable, holding other variables constant. The result provides us with (17.9) more chances to be selected for Bypass Graft operation with a one-unit increase in this variable in contrast to a low unit.

Age is also a significant predictor of the probability of Bypass Graft surgery. The odds ratio of Age is 2.6 with a 95% confidence interval of 1.07 to 6.3. This indicates that older patients have 2.6 times more likely to undergo Bypass Graft surgery.

The odds ratio for Blood Sugar is 1.3, with a 95% confidence interval of 1.03 to 1.65. This means that a case has a 1.30 times more likelihood of having surgery compared to not having surgery for every one-unit rise in Blood Sugar. This means patients with Blood Sugar had a 30% of increase in the probability of surgery.

MCH also played an essential role in increasing risks for the response variable as with one unit increase in MCH measurement, which would lead to odds (2.6) times more probability to face Bypass Graft surgery than patients with lower MCH.

When comparing low DBP units to high DBP units, DBP had a significant role in increasing risks for Bypass Graft surgery, as one unit increase in DBP measurement increased 35% of the likelihood of having Bypass Graft surgery.

The Chi-square test, which tests the null hypothesis that all coefficients are zero, revealed that the null hypothesis is rejected with a P-value = 0.000. This means that the overall model is statistically significant.

Table 3.10 Classification table of predicted data

Logistic model			
Classified	Patient D	Control ~D	Total
+	99	1	100
-	1	59	60
Total	100	60	160
Classified + if predicted Pr(D) >=0.5 True D is defined as surgY !=0			
Sensitivity		Pr(+|D)	99.00%
Specificity		Pr(-|~ D)	98.33%
Positive predictive value		Pr(D| +)	99.00%
Negative predictive value		Pr(~ D| -)	98.33%
False + Rate for true ~ D		Pr(+|~ D)	1.67%
False + Rate for true D		Pr(-| D)	1.00%
False + Rate for Classified +		Pr(~ D| +)	1.00%
False + Rate for Classified -		Pr(D| -)	1.67%
Correctly classified			98.75%

Another way to assess the model's goodness of fit is by calculating the classification table of a ROC curve. Instances with a probability less than 0.50 are anticipated in the categorization table, as shown in Table 3.5, to have Bypass surgery, whereas other cases are expected not to have the surgery. The two groups' calculated probability should ideally be considerably different. The model correctly predicted 99 patients who had undergone Bypass surgery out of 100 patients. We have calculated the model accuracy using Rstudio software. The percentage of correctly identified data determines the model's accuracy. In Table 3.10 it can be noted that the overall model accuracy is 98%

It is excellent that the classification prediction accuracy is more than 98%, and only 2% of categorization errors are misclassified.

Lastly, we draw the ROC (Receiver Operating Characteristic) Curve, which shows the proportion of true positives predicted by the model. The area under the curve, or AUC, measures how well our model can predict outcomes. The bigger the AUC, the better the model can predict the Bypass Graft data. For example, figure 3.7 shows that the area under the curve is 0.99, which means that our model is best at making predictions.

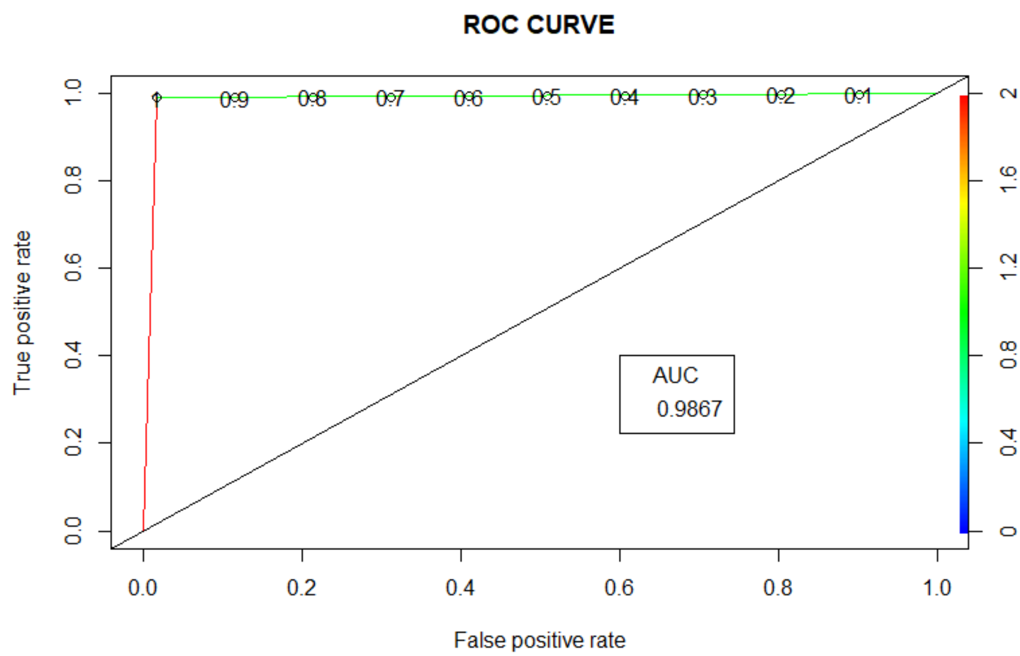

Fig. 3.7 ROC curve of predicted values with AUC

3.5.2. Bayesian logistic regression

A further model is used in Bayesian logistic regression. We have only used those variables which were significant predictors using classical logistic regression in section 3.5.1, which utilizes Bayesian inference.

The logistic equation has the following generic form:

$$\mathrm{g(x)} = \ logit\left(\frac{p_i}{1-p_i}\right) = \ \beta_0 + \ \beta^T \quad \dots 3.3$$

In Bayesian logistic regression, parameter estimates are obtained using Bayesian inference, a practical way of defining the probability distribution for explanatory variables, as opposed to maximum likelihood approaches used in classical logistic regression β. An indicator variable called Y_i is introduced in the chapter one (takes 1 if there is CABG surgery and 0 otherwise).

Then:

$$Y_i \sim Bernoulli(p_i) \qquad , \qquad i = 1, \dots, N$$

The linear predictor $\eta_i$is:

$$\eta_i = \ \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} \quad \dots 3.4$$

This is connected to the fitted probability through the logit function,

$$logit\left(\frac{p_i}{1-p_i}\right) = \ \eta_i$$

The Bayesian logistic model is formulated by specifying prior distributions on the logistic regression coefficients:

$$\boldsymbol{\beta_j} \sim p(\theta_j) \qquad\qquad j = 1, \dots, K$$

The variables included in the Bayesian model are DBP, Age, MCH, WBC, and Blood Sugar. In addition, a set of weak, "vague," normal priors with very low precision was utilized to parameterize estimates in the model. These priors were distributed using a normal distribution.

Table 3.11 Posterior odds ratios and their standard deviation obtained from the binary model using Bayesian logistic model

Variable	Odds Ratio	SD	95% Crl. Interval	
			Lower	Upper
Constant	1.671	0.281	1.218	2.315
Age	1.13	0.009	1.083	1.175
DBP	1.037	0.017	1.005	1.073
MCH	0.814	0.079	0.667	0.981
WBC	0.616	0.065	0.494	0.748
Blood Sugar	0.990	0.003	0.983	0.995

Multiple chains of samples were performed simultaneously for the parameters with no discernible change in parameter estimations between chains of samples with different starting values. Additional investigation was done with starting "burn-in" lengths of 5000, 10000, and 30000. There was no considerable change in the parameter estimations, and the samples appeared to settle around iteration number 5000 and with increasing burn-in time. Nevertheless, when sample quantity and "burn-in" increased, the kernel density charts seemed smoother. Table 3.11 indicates the contribution of each factor to the Bypass Graft surgery using the posterior distribution. It can be noted that Age, DBP, and Blood Sugar have little impact on

Bypass surgery. However, for each unit increase in WBC, the odds of having Bypass surgery reduced by 38% (OR=0.616, (95%Crl: 0.494 to 0.748)). Regarding the MCH, for each unit increase in MCH, the odds of having Bypass surgery reduced by 19% (OR=0.81, (95%CrL: 0.66 to 0.98)). The Bayesian odds ratio for variables of DBP and Blood Sugar has little effect on the Bypass Graft surgery since their values are close to 1.

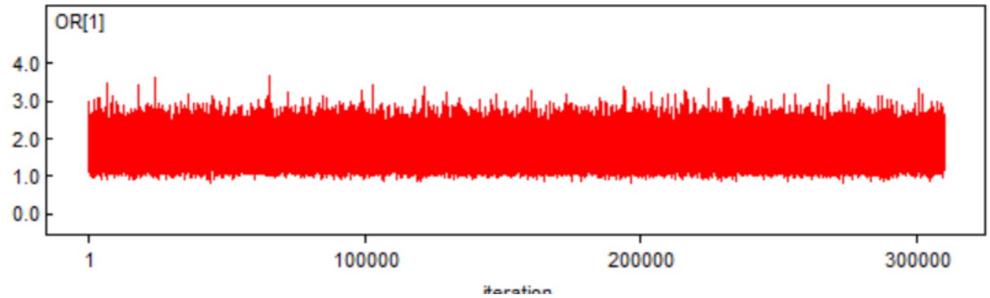

Fig. 3.8 Convergence for parameter estimates of constant

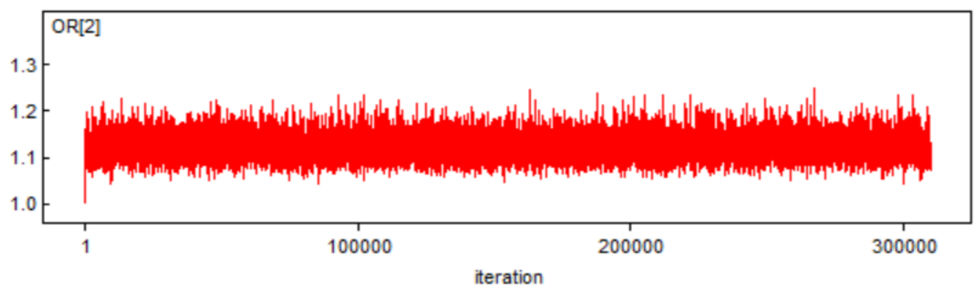

Fig. 3.9 Convergence for parameter estimates of Age

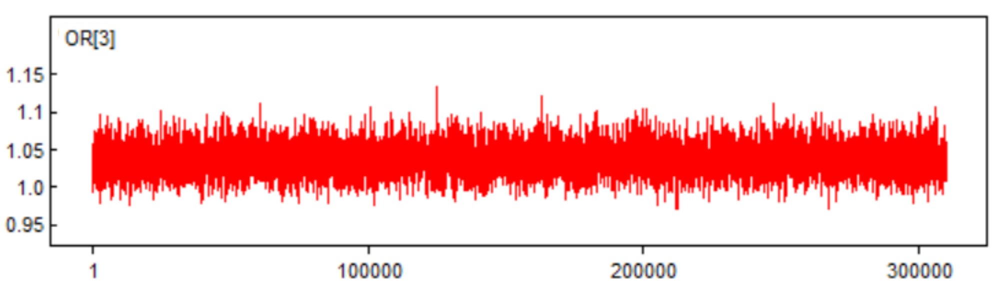

Fig 3.10 Convergence for parameter estimates of DBP

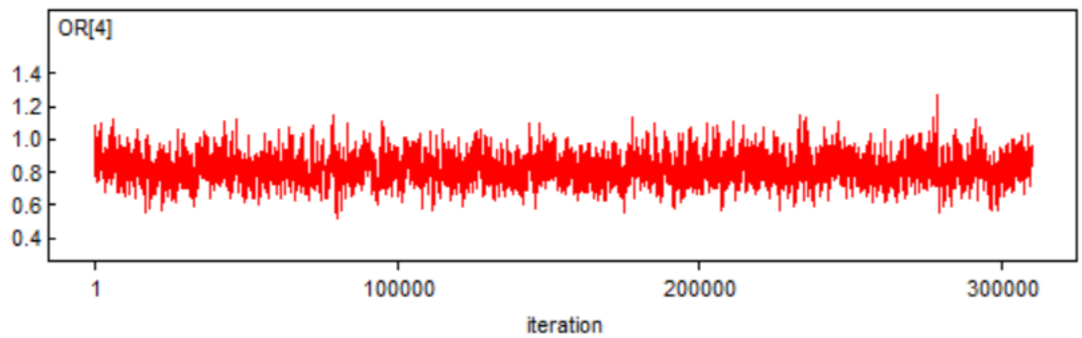

Fig. 3.11 Convergence for parameter estimates of MCH

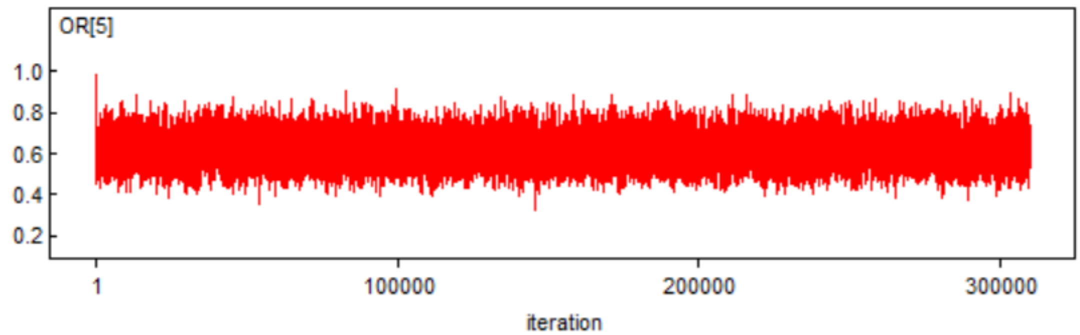

Fig. 3.12 Convergence for parameter estimates of WBC

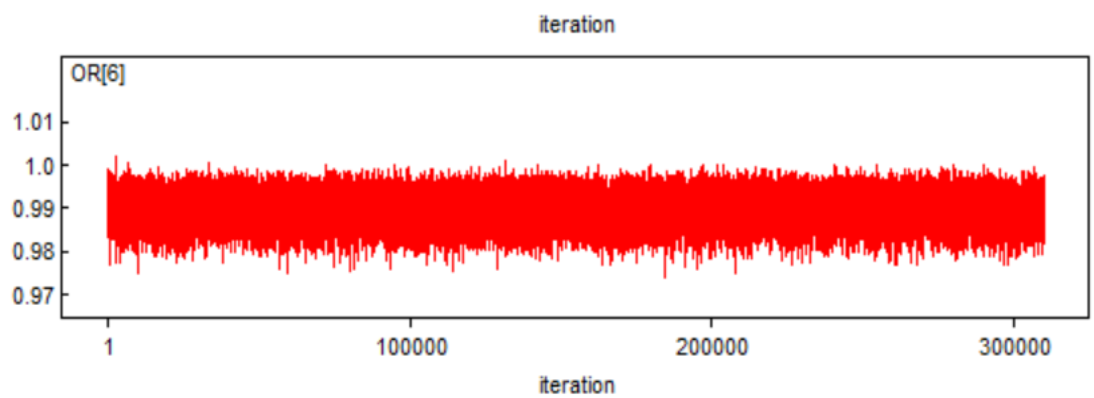

Fig. 3.13 Convergence for parameter estimates of Blood Sugar

The last half of the simulated sequences served as the basis for inference. The time series plot created by "history" in the graph can be visually examined to see that the Markov chains converged for parameter estimation of Age, MCH, WBC, DBP, and Blood sugar variables. The plots illustrate that parameter are reached equilibriums.

While auto-correlation results in sluggish mixing and likely individual non-convergence to the limited distribution since the chain will seek to locate less space in finite time, the high correlation between a chain's coefficients tends to perform slow convergence.

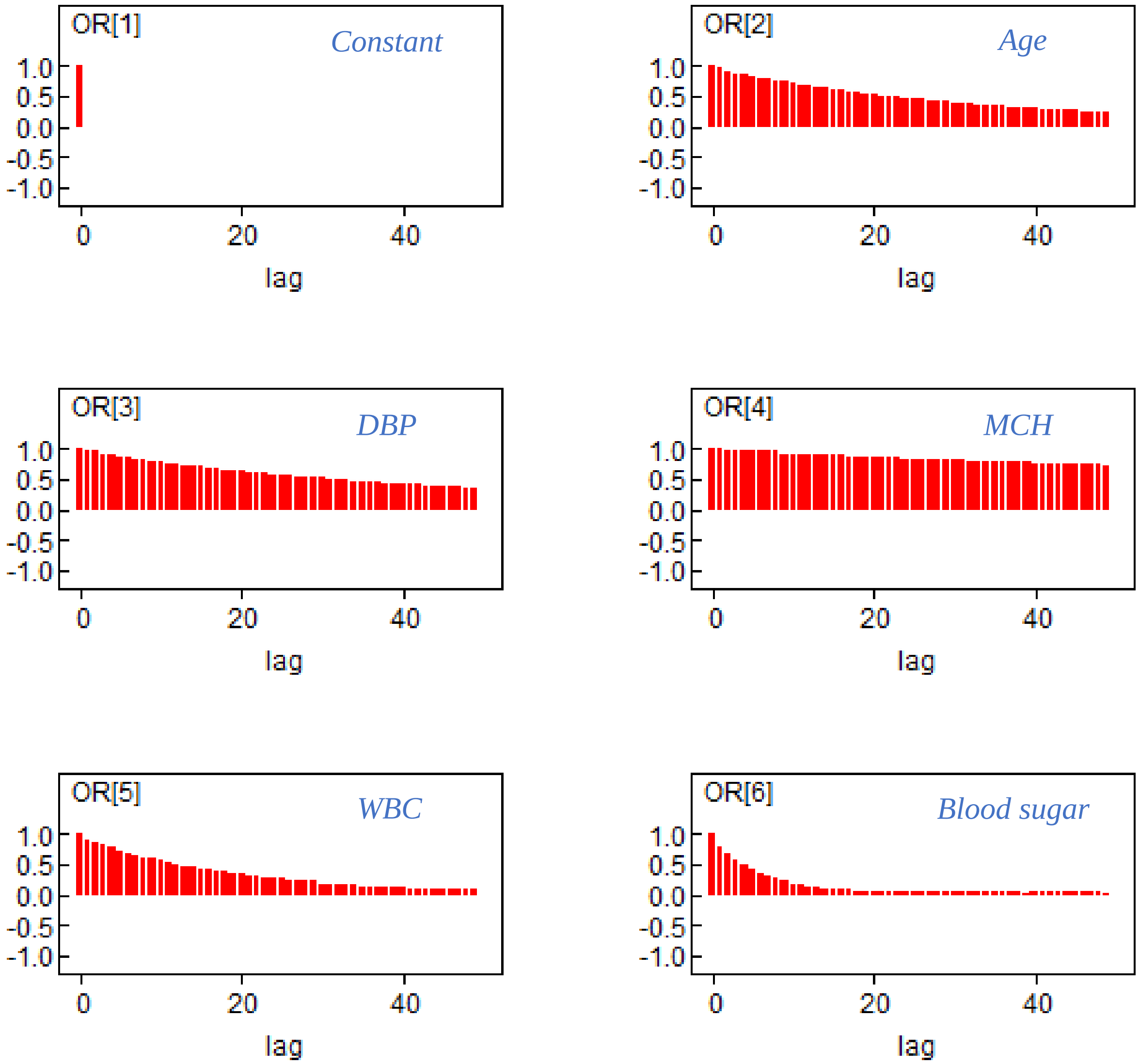

Fig. 3.14 Autocorrelation function for parameter estimation

The autocorrelation is not high for intercept, DBP, and MCH. Even though the auto-correlations for these variables are minor, it can be seen that the delayed auto-correlation clearly shows that there is a mixing with the intercept, Gender, and constant parameters since the plots show a persistent correlational link extending back in time. Re-parameterizing, the model is one of the key strategies for addressing excessive auto-correlation. There is a high correlation between Age, Blood Sugar, and WBC.

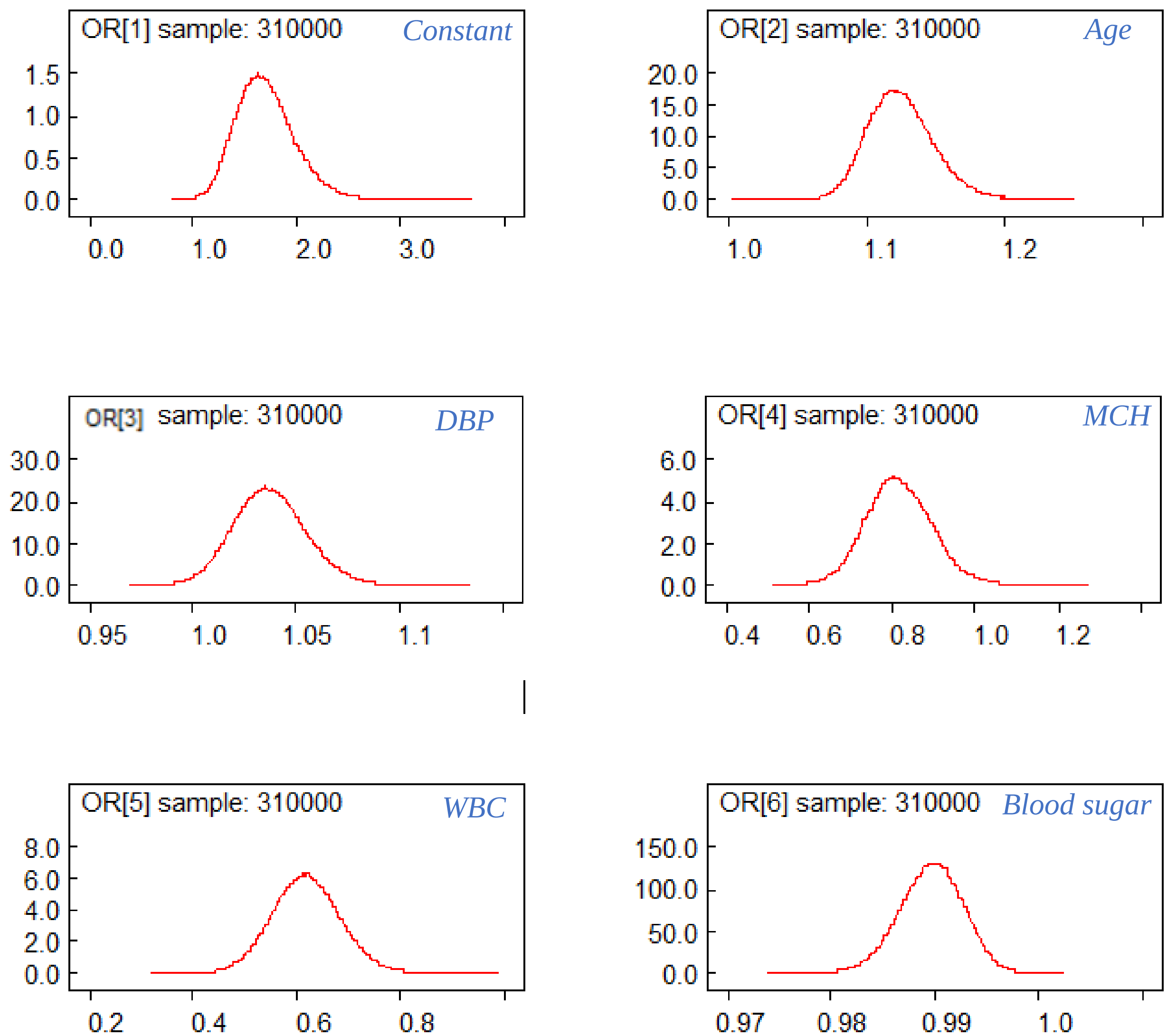

Fig. 3.15 Kernel Density for variable estimations

Figure 3.15 shows the Kernel Density of all estimated parameters of the variables. The Kernel Density posterior estimation shows that the parameters are away from zero, and the shapes are typically distributed.

3.6. Discussion

This thesis aimed to determine factors affecting Bypass Graft surgery in Erbil. We have found that HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, and T3 variables have a relatively exciting connection with the Bypass Graft using visual inspection. Multiple logistic regression was used to determine factors affecting Bypass Graft surgery. Variables were selected using the forward selection method.

We have found that WBC is one of the most significant variables affecting Bypass Graft surgery. The result showed that patients with higher WBC have 17.9 times more probability of Bypass surgery. This result is confirmed by the study of (Aizenshtein, et al., 2020) as they have shown that High WBC counts may make coronary artery Bypass Graft (CABG) surgery more likely to result in cardiovascular complications.

Our result was to highlight that the risk factor of HbA1c has a significantly increased risk, more likely 6.14 times in undergoing patients to CABG. (Narayan, et al., 2017), (Halkos, et al., 2008) and (Zheng, et al., 2017) revealed that the HbA1c level is significantly related to increased risks of all-cause Bypass surgery. For coronary artery Bypass Grafting patients, preoperative hemoglobin A1c testing is strongly associated with more precise risk categorization. Additionally, we have found that a higher average quantity of red blood cells named MCH level led to a higher risk of Bypass surgery.

The ability to provide superior results in high-risk individuals about Age significantly affects Bypass Graft surgery. This finding is supported by a study by (Lemaire, et al., 2020) and (Nicolini, et al., 2017) showed that patients under the Age of 60 who received CABG had a decreased risk of poor outcomes than elderly patients. In addition, compared to more seniors, people over 60 exhibit a distinct clinical pattern of CAD presentation.

According to our data, more significant DBP Variability was related to an increased risk of CABG surgery 1.352 times. (Roberts, et al., 1977) showed more than one-third of people with coronary artery Bypass Graft (CABG) surgery develop systemic hypertension.

Blood Sugar has become one of the other significant risk factors due to its association with excessive glycated hemoglobin levels, which probably cause Bypass surgery by 1.303 times. (Liu, et al., 2021) Furthermore, (Van Straten, et al., 2010) supported the result as they showed that Blood Sugar significantly increased the total composite adverse events in individuals with ischemic heart disease following CABG.

According to the results of the data that we collected on both groups, besides significant factors, which cause facing Bypass surgery, there were some nonsignificant factors such as; HbA1c, BMI, PLT, T3, and Eosinophil that do not affect Bypass surgery depending on our data. Whereas some studies confirmed our nonsignificant factors, for instance, (Jin, et al., 2005) showed that Body mass index does not appear to be a significant risk factor for CABG mortality. In addition, (Kremke, et al., 2015) revealed no statistically significant link between platelet transfusion and postoperative mortality. Furthermore, (Arter, 2021) confirmed our results; the study showed that decreased T3 hormone levels in CABG surgery did not influence the development of atrial fibrillation.

However, Because of our sample size, our findings showed that HbA1c and Eosinophil were nonsignificant but contradicted (Narayan, et al., 2017), and (Tanaka, et al., 2012) respectively explained those two factors were significant and one of the risk factors in undergoing CABG surgery. This disparity might be explained by the fact that (Columb & Atkinson, 2016) there was medical research seeking to establish generalizable findings from data acquired from randomized samples from such groups. Therefore, larger sample sizes should yield results that are more accurate. As a result, every clinical research's traditional design and interpretation should be included sample size and power concerns.

In this study, we have used Bayesian logistic regression for those significant factors affecting Bypass surgery. We have found that the results differed from the classical logistic regression, as the odds ratio results were different. This is because we have incorporated vague prior distribution into our posterior distribution. Even though the autocorrelations for these variables are minor, it can be seen that the delayed autocorrelation clearly shows that there is mixing with the intercept, gender, and constant parameters since the plots show a persistent correlational link extending back in time. Re-parameterizing, the model is one of the key strategies for addressing excessive autocorrelation. There is a high correlation between Age, Blood Sugar, and WBC. The time series plot created by "history" in the graph can be visually examined to see that the Markov chains converged for parameter estimation of Age, MCH, WBC, DBP, and Blood pressure variables. The plots illustrate that parameters are reached equilibriums.

CHAPTER FOUR

CONCLUSION AND RECOMMENDATION

CHAPTER FOUR

4. CONCLUSION AND RECOMMENDATION

4.1. Conclusion

1- Covaries(s) have been removed because of multicollinearity. Variables of Weight, Length, LDL, HCT, and CK-MB were deleted because they are correlated with other variables.

2- Using multivariate logistic regression, we have found that risk factors associated with CABG surgery are DBP, Age, MCH, WBC, and Blood Sugar.

3- ROC curve showed that the area under the curve is 0.99, which means that our model is the best at making predictions.

4- We have also used Bayesian logistic regression for those significant factors affecting Bypass surgery.

5- We have found that the results differed from the classical logistic regression as the odds ratio results differed. This is because we have incorporated vague prior distribution into our posterior distribution.

6- The risk factors associated with Bypass Graft surgery using Bayesian logistic regression were Age, WBC, and MCH. Visual inspection revealed that there is some autocorrelation among estimated parameters.

4.2. Recommendations

There are some recommendations as follows:

1. Increasing sample size by obtaining data from different hospitals to investigate the environmental effect of CABG surgery.
2. Using different statistical approaches to identify risk factors using random forest, Bayesian Neural Network (BNN), and Lasso regression.
3. Providing informative prior can be one of the options for getting better results from the Bayesian model.
4. The ordinal Regression model can be another suitable mode to investigate different level of the Bypass graph surgery.

References

- ABRAMOV, D. ET AL. (2000) Trends in coronary artery bypass surgery results: a recent, 9-year study. The Annals of thoracic surgery, 70(1), pp. 84-90.

- ADHIKARI, S., ROSE, S. & NORMAND, S. (2020) Nonparametric Bayesian Instrumental Variable Analysis: Evaluating Heterogeneous Effects of Coronary Arterial Access Site Strategies. Journal of the American Statistical Association, 115(532), pp. 1635-1644.

- AGRESTI, A.(2018) An introduction to categorical data analysis. s.l.:John Wiley & Sons.

- AIZENSHTEIN, A. ET AL. (2020) Effects of Preoperative WBC Count on Post-CABG Surgery Clinical Outcome. Southern Medical Journal, 113(6), pp. 305-310.

- ALEXANDER, J. & SMITH, P. (2016) Coronary-artery bypass Grafting. New England Journal of Medicine, 374(20), pp. 1954-1964.

- ARTER, D. (2021) Effect of Low Triiodothyronine (T3) Hormone Levels on The Development of Atrial Fibrillation After Coronary Artery Bypass Surgery.

- BELSLEY, D. (2014) Conditioning diagnostics. s.l.:Wiley StatsRef : Statistics Reference Online.

- CASELLA, G. & GEORGE, E. (1992) Explaining the Gibbs sampler. The American Statistician, 46(3), pp. 167-174.

- CEVENINI, G. ET AL. (2007) A comparative analysis of predictive models of morbidity in intensive care unit after cardiac surgery–Part II. BMC Medical Informatics and Decision Making, 7(1), pp. 1-13.

- CHIB, S. & GREENBERG, E. (1995) Understanding the metropolis-hastings algorithm. The american statistician, 49(4), pp. 327-335.

- COLUMB, M. & ATKINSON, M. (2016) Statistical analysis: sample size and power estimations. Bja Education, 16(5), pp. 159-161.

- CONGDON, P. (2005) Bayesian models for categorical data. s.l.:John Wiley & Sons.

- COX, D. & SNELL, E. (2018) Analysis of binary data. s.l.:Routledge.

- CZEPIEL, S. (2002) Maximum likelihood estimation of logistic regression models: theory and implementation. p. 83.

- DOBSON, A. & BARNETT, A. (2018) An introduction to generalized linear models. s.l.:Chapman and Hall/CRC.

- DUNSON, D. (2001) Commentary: practical advantages of Bayesian analysis of epidemiologic data. American journal of Epidemiology, 153(12), pp. 1222-1226.

- EDWARDS, F. ET AL. (1988) Use of a Bayesian statistical model for risk assessment in coronary artery surgery. The Annals of thoracic surgery, 45(4), pp. 437-440.

- FARAWAY, J. (2016) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. s.l.:Chapman and Hall/CRC.

- FIELD, A., MILES, J. & FIELD, Z. (2012) Discovering Statistics Using R. Great Britain: Sage Publications.

- GELFAND, A. & SMITH, A. (1990) Sampling-based approaches to calculating marginal densities. Journal of the American statistical association, 85(410), pp. 398-407.

- GELMAN, A. (2002) Prior distribution. Encyclopedia of environmetrics, 3(4), pp. 1634-1637.

- GELMAN, A., JAKULIN, A., PITTAU, M. & SU, Y. (2008) A weakly informative default prior distribution for logistic and other regression models. 2(4), pp. 1360-1383.

- GEMAN, S. & GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence, 6(6), pp. 721-741.

- GENKIN, A., LEWIS, D. & MADIGAN, D. (2007) Large-scale Bayesian logistic regression for text categorization. technometrics, 49(3), pp. 291-304.

- GILL, J. (2014) Bayesian methods: A social and behavioral sciences approach. Boca Raton: Fla: Chapman & Hall/CRC.

- HALKOS, M. ET AL. (2008) Elevated preoperative hemoglobin A1c level is predictive of adverse events after coronary artery bypass surgery. The Journal of thoracic and cardiovascular surgery, 136(3), pp. 631-640.

- HASTINGS, W. (1970) Monte Carlo sampling methods using Markov chains and their applications. pp. 97-109.

- HOSMER, D., LEMESHOW, S. & STURDIVANT, R. (2013) Applied logistic regression. s.l.:Hoboken.

- JIN, R., GRUNKEMEIER, G., FURNARY, A. & HANDY JR, J. (2005) Is obesity a risk factor for mortality in coronary artery bypass surgery?. Circulation, 111(25), pp. 3359-3365.

- JR, H., W., D., LEMESHOW, S. & STURDIVANT, R. X. (2013) Applied logistic regression (Vol. 398). s.l.:John Wiley & Sons.

- KHAN, S. ET AL. (1990) Increased mortality of women in coronary artery bypass surgery: evidence for referral bias. Annals of internal medicine, 112(8), pp. 561-567.

- KLEINBAUM, D. ET AL. (2002) Logistic regression. New York: Springer-Verlag.

- KREMKE, M. ET AL. (2015) The association between platelet transfusion and adverse outcomes after coronary artery bypass surgery. European Journal of Cardio-Thoracic Surgery, 48(5), pp. e102-e109.

- KURKI, T. & KATAJA, M. (1996) Preoperative prediction of postoperative morbidity in coronary artery bypass Grafting. The Annals of thoracic surgery, 61(6), pp. 1740-1745.

- LANCASTER, T. (2004) An introduction to modern Bayesian econometrics (p. 401). Oxford: Blackwell.

- LANG, C. D., HE, Y. & BITTL, J. A. (2015) Bayesian inference supports the use of bypass surgery over percutaneous coronary intervention to reduce mortality in diabetic patients with multivessel coronary disease. International Journal of Statistics in Medical Research, 4(1), 26-34., 4(1), pp. 26-34.

- LE, C. T. & EBERLY, L. E. (2016) Introductory biostatistics. s.l.:John Wiley & Sons.

- LEMAIRE, A. ET AL. (2020) The impact of age on outcomes of coronary artery bypass Grafting.. Journal of Cardiothoracic Surgery, 15(1), pp. 1-8.

- LIPPMANN, R. & SHAHIAN, D. (1997) Coronary artery bypass risk prediction using neural networks. The Annals of thoracic surgery, 63(6), pp. 1635-1643.

- LIU, M. ET AL. (2021) Effect of diabetes mellitus on long-term outcomes of surgical revascularization in patients with ischemic heart failure: a propensity score-matching study.. Chinese Medical Journal, 134(10), pp. 1146-1151.

- MACK, M. P. A. ET AL. (2004) Comparison of coronary bypass surgery with and without cardiopulmonary bypass in patients with multivessel disease. The Journal of thoracic and cardiovas, 127(1), pp. 167-173.

- MARSHALL, G., SHROYER, A., GROVER, F. & HAMMERMEISTER, K. (1994) Bayesian-logit model for risk assessment in coronary artery bypass Grafting. The Annals of thoracic surgery, 57(6), pp. 1492-1500.

- MCCULLAGH, P. & NELDER, J. (2019) Generalized linear models. s.l.:Routledge.

- MELIN, J. ET AL. (1985) Alternative diagnostic strategies for coronary artery disease in women: demonstration of the usefulness and efficiency of probability analysis. Circulation, 71(3), pp. 535-542.

- METROPOLIS, N. ET AL. (1953) Equation of state calculations by fast computing machines. The journal of chemical physics, 21(6), pp. 1087-1092.

- MIDI, H., SARKAR, S. K. & RANA, S. (2010) Collinearity diagnostics of binary logistic regression model. Journal of interdisciplinary mathematics, 13(3), pp. 253-267.

- NARAYAN, P. ET AL. (2017) Preoperative glycosylated hemoglobin: a risk factor for patients undergoing coronary artery bypass. The Annals of thoracic surgery, 104(2), pp. 606-612.

- NELDER, J. & WEDDERBURN, R. (1972) Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), pp. 370-384.

- NICOLINI, F. ET AL. (2017) The impact of age on clinical outcomes of coronary artery bypass Grafting: long-term results of a real-world registry. BioMed research international, Volume 2017.

- NOEL, C. ET AL. (2020) Predictors of surgical readmission, unplanned hospitalization and emergency department use in head and neck oncology: A systematic review. Oral Oncology, Volume 111, p. 105039.

- OTHMAN, G., SAEED, R., KADIR, D. & TAHER, H. (2019) Relation of angiography to hematological, hormonal and some biochemical variables in coronary artery bypass Graft patients. Journal of Physics: Conference Series, 1294(6), p. 062110.

- PAN, W. (2001) Akaike's information criterion in generalized estimating equations. Biometrics, 57(1), pp. 120-125.

- PERRIER, S. ET AL. (2017) Predictors of atrial fibrillation after coronary artery bypass Grafting: a Bayesian analysis. The Annals of thoracic surgery, 103(1), pp. 92-97.

- PIGEON, J. & HEYSE, J. (1999) An improved goodness of fit statistic for probability prediction models. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 41(1), pp. 71-82.

- QUEEN, J., QUINN, G. & KEOUGH, M. (2002) Experimental design and data analysis for biologists. s.l.:Cambridge university press.

- RACHARDI, A. ET AL. (2020) The prediction of postoperative morbidity in coronary artery bypass Grafting using Naïve Bayes Classification and Bayes Factor. In AIP Conference Proceedings, 2242(1), p. 030020.

- REES, M. & DINESCHANDRA, J. (2006) Risk stratification in assessing risk in coronary artery bypass surgery. 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06), pp. 303-308.

- REID, C. ET AL. (2009) An Australian risk prediction model for 30-day mortality after isolated coronary artery bypass: the AusSCORE. Thoracic and cardiovascular surgery, 138(4), pp. 904-910.

- RESNIC, F. ET AL. (2004) Exploration of a Bayesian updating methodology to monitor the safety of interventional cardiovascular procedures. Medical Decision Making, 24(4), pp. 399-407.

- ROBERT, C., CASELLA, G. & CASELLA, G. (2010) Introducing monte carlo methods with r. New York: Springer, Volume 18.

- ROBERTS, A. ET AL. (1977) Systemic hypertension associated with coronary artery bypass surgery: predisposing factors, hemodynamic characteristics, humoral profile, and treatment. The Journal of Thoracic and Cardiovascular Surgery, 74(6), pp. 846-859.

- SORENSEN, D., GIANOLA, D. & GIANOLA, D. (2002) Likelihood, Bayesian and MCMC methods in quantitative genetics.. s.l.:s.n.

- SPIEGELHALTER, D., BEST, N., CARLIN, B. & VAN DER LINDE, A. (2002) Bayesian measures of model complexity and fit. Journal of the royal statistical society, 64(4), pp. 583-639.

- TANAKA, M. ET AL. (2012) Eosinophil count is positively correlated with coronary artery calcification. Hypertension Research, 35(3), pp. 325-328.

- TU, J., SYKORA, K., NAYLOR, C. & 1, S. C. O. T. C. C. N. O. O. (1997) Assessing the outcomes of coronary artery bypass Graft surgery: how many risk factors are enough?. Journal of the American College of Cardiology, 30(5), pp. 1317-1323.

- UGOLINI, C. & NOBILIO, L. (2004) Risk adjustment for coronary artery bypass Graft surgery: an administrative approach versus EuroSCORE. International Journal for Quality in Health Care, 16(2), pp. 157-164.

- VAN STRATEN, A. ET AL. (2010) Diabetes and survival after coronary artery bypass Grafting: comparison with an age-and sex-matched population. European Journal of Cardio-Thoracic Surgery, 37(5), pp. 1068-1074.

- WILHELMSEN, M., DIMAKOS, X., HUSEBØ, T. & FISKAAEN, M. (2009) Bayesian modelling of credit risk using integrated nested laplace approximations. NR publication, pp. 1-25.

- ZHENG, J. ET AL. (2017) Does HbA1c level have clinical implications in diabetic patients undergoing coronary artery bypass Grafting? A systematic review and meta-analysis. International journal of endocrinology, Volume 2017.

Appendix

Appendix A: Developing a model for identifying characteristics that influence Bypass Graft based on forward selection model

Table 1: AIC and deviance of three predictor variables.

Three Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	100.1682		0.6636	
Surgery, HbA1c, Age, WBC (Model-2 and 3)	79.1898	20.9784	0.5449	0
Surgery, HbA1c, Age, Eosinophil (Model-2 and 5)	86.6300	13.5382	0.5914	0
Surgery, HbA1c, Age, Creatinine (Model-2 and 8)	92.8065	7.3617	0.6300	0
Surgery, HbA1c, Age, Blood Urea (Model-2 and 6)	92.3707	7.7975	0.6273	0
Surgery, HbA1c, Age, PLT (Model-2 and 4)	90.2547	9.9135	0.6141	0
Surgery, HbA1c, Age, Blood Sugar (Model-2 and 1)	85.6700	14.4982	0.5854	0
Surgery, HbA1c, Age, Smoke (Model-2 and 11)	93.9848	6.1835	0.6374	0
Surgery, HbA1c, Age, BMI (Model-2 and 7)	87.1510	13.0172	0.5947	0
Surgery, HbA1c, Age, CHO (Model-2 and 9)	99.2250	0.9432	0.6702	0
Surgery, HbA1c, Age, PDW (Model-2 and 12)	98.5023	1.6660	0.6656	0
Surgery, HbA1c, Age, MCH (Model-2 and 13)	94.0081	6.1601	0.6376	0
Surgery, HbA1c, Age, DBP (Model-2 and 14)	100.1682	3.3E-06	0.6761	0
Surgery, HbA1c, Age, T3 (Model-2 and 10)	97.9764	2.1919	0.6624	0

Table 2: AIC and deviance of four predictor variables.

Four Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	79.1898		0.5449	
Surgery, HbA1c, Age, WBC, Blood Sugar (Model1 and 6)	65.6650	13.5248	0.4729	0
Surgery, HbA1c, Age, WBC, Eosinophil (Model1 and 2)	66.4836	12.7062	0.4780	0
Surgery, HbA1c, Age, WBC, BMI (Model1 and 8)	64.3570	14.8329	0.4647	0
Surgery, HbA1c, Age, WBC, PLT (Model1 and 5)	73.0963	6.0936	0.5194	0
Surgery, HbA1c, Age, WBC, Blood Urea (Model1 and 4)	75.3487	3.8412	0.5334	0
Surgery, HbA1c, Age, WBC, Creatinine (Model1 and 3)	75.4077	3.7821	0.5338	0
Surgery, HbA1c, Age, WBC, Smoke (Model1 and 7)	74.3043	4.8855	0.5269	0
Surgery, HbA1c, Age, WBC, MCH (Model1 and 11)	75.4195	3.7703	0.5339	0
Surgery, HbA1c, Age, WBC, T3 (Model1 and 13)	75.5080	3.6819	0.5344	0
Surgery, HbA1c, Age, WBC, PDW (Model1 and 10)	79.0284	0.1615	0.5564	0
Surgery ,HbA1c, Age, WBC, CHO (Model-1 and 9)	78.3208	0.8690	0.5520	0
Surgery ,HbA1c, Age, WBC, DBP (Model1 and 12)	77.3860	1.8039	0.5461	0

Table 3: AIC and deviance of five predictor variables

Five Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	64.3570		0.4647	
Surgery, HbA1c, Age, WBC, BMI, Blood Sugar (Model-3 and 1)	56.8333	7.5236	0.4302	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil (Model-3 and 2)	49.0363	15.3207	0.3815	0
Surgery, HbA1c, Age, WBC, BMI, PLT(Model-3 and 4)	58.3402	6.0167	0.4396	0
Surgery ,HbA1c, Age, WBC, BMI, Smoke(Model-3 and 7)	60.4645	3.8925	0.4529	0
Surgery, HbA1c, Age, WBC, BMI, Blood Urea(Model-3 and 5)	63.4996	0.8573	0.4719	0
Surgery, HbA1c, Age, WBC, BMI, Creatinine(Model-3 and 6)	60.4049	3.9520	0.4525	0
Surgery, HbA1c, Age, WBC, BMI, MCH (Model-3 and 8)	62.7891	1.5679	0.4674	0
Surgery, HbA1c, Age, WBC, BMI, T3(Model-3 and 9)	62.0448	2.3121	0.4628	0
Surgery, HbA1c, Age, WBC, BMI, DBP(Model-3 and 12)	62.0645	2.2925	0.4630	0
Surgery, HbA1c, Age, WBC, BMI, CHO (Model-3 and 11)	64.2500	0.1069	0.4766	0
Surgery, HbA1c, Age, WBC, BMI, PDW(Model-3 and 10)	64.1245	0.2324	0.4758	0

Table 4: AIC and deviance of six predictor variables.

Six Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	49.0363		0.3815	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar (Model-2 and 1)	40.6584	8.3779	0.3416	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, PLT (Model-2 and 3)	48.1821	0.8542	0.3886	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Creatinine (Model-2 and 6)	44.5898	4.4465	0.3662	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Smoke (Model-2 and 4)	45.9692	3.0671	0.3748	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, T3 (Model-2 and 8)	47.6188	1.4175	0.3851	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, DBP (Model-2 and 9)	46.2177	2.8185	0.3764	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, MCH (Model-2 and 7)	48.9574	0.0788	0.3935	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Urea (Model-2 and 5)	47.5269	1.5093	0.3845	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, PDW (Model-2 and 11)	48.8768	0.1594	0.3930	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, CHO (Model-2 and 10)	48.4330	0.6033	0.3902	0

Table 5: AIC and deviance of seven predictor variables.

Seven Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	40.6584		0.3416	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, Creatinine (Model-1 and 3)	37.58835	3.0700	0.3349	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, Smoke (Model-1 and 4)	36.1942	4.4642	0.3262	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP (Model-1 and 6)	34.4601	6.1982	0.3154	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, Blood Urea (Model-1 and 8)	38.7370	1.9214	0.3421	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, T3 (Model-1 and 5)	35.8661	4.7923	0.3242	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, PLT (Model-1 and 2)	36.8207	3.8377	0.3301	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, CHO (Model-1 and 10)	40.5825	0.0758	0.3536	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, PDW (Model-1 and 9)	40.2484	0.4100	0.3516	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, MCH (Model-1 and 7)	40.5203	0.1381	0.3532	0

Table 6: AIC and deviance of eight predictor variables.

Eight Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	34.4601		0.3154	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, T3 (Model-3 and 5)	31.7925	2.6677	0.3112	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, Smoke (Model-3 and 2)	32.6144	1.8457	0.3163	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT (Model-3 and 6)	26.0560	8.4042	0.2754	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, Creatinine (Model-3 and 1)	32.1371	2.3231	0.3134	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, Blood Urea (Model-3 and 4)	29.6298	4.8304	0.2977	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PDW (Model-3 and 8)	34.4491	0.0111	0.3278	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, MCH (Model-3 and 9)	33.4080	1.0521	0.3213	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, CHO (Model-3 and 7)	33.9574	0.5028	0.3247	0

Table 7: AIC and deviance of nine predictor variables.

Nine Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	26.0560		0.2754	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, Blood Urea (Model-3 and 5)	24.6070	1.4490	0.2788	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, T3 (Model-3 and 1)	24.8830	1.1730	0.2805	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, Creatinine (Model-3 and 4)	25.2531	0.8029	0.2828	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, Smoke (Model-3 and 2)	26.0521	0.0039	0.2878	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH (Model-3 and 7)	22.1517	3.9042	0.2635	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, CHO (Model-3 and 8)	25.8425	0.2135	0.2865	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, PDW (Model-3 and 6)	25.7847	0.2713	0.2862	0

Table 8: AIC and deviance of eleven predictor variables.

Eleven Predicted Variable				
Cons_Model	Residual Deviance	Change	AIC	Sig
	17.9683		0.2498	
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea, T3 (Model-1 and 2)	0	0	0.1500	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea, Creatinine (Model-1 and 3)	0	0	0.1500	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea, CHO (Model-1 and 5)	0	0	0.1500	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea, Smoke (Model-1 and 6)	0	0	0.1500	0
Surgery, HbA1c, Age, WBC, BMI, Eosinophil, Blood Sugar, DBP, PLT, MCH, Blood Urea, PDW (Model-1 and 4)	0	0	0.1500	0

Regarding building model to stages after eleven predictors, they share the same Residual Deviance. Therefore, this step is regarded as the last need to complete the model.

Appendices B: Logit model fit of other predictor variables that appear but have no correlation with surgery

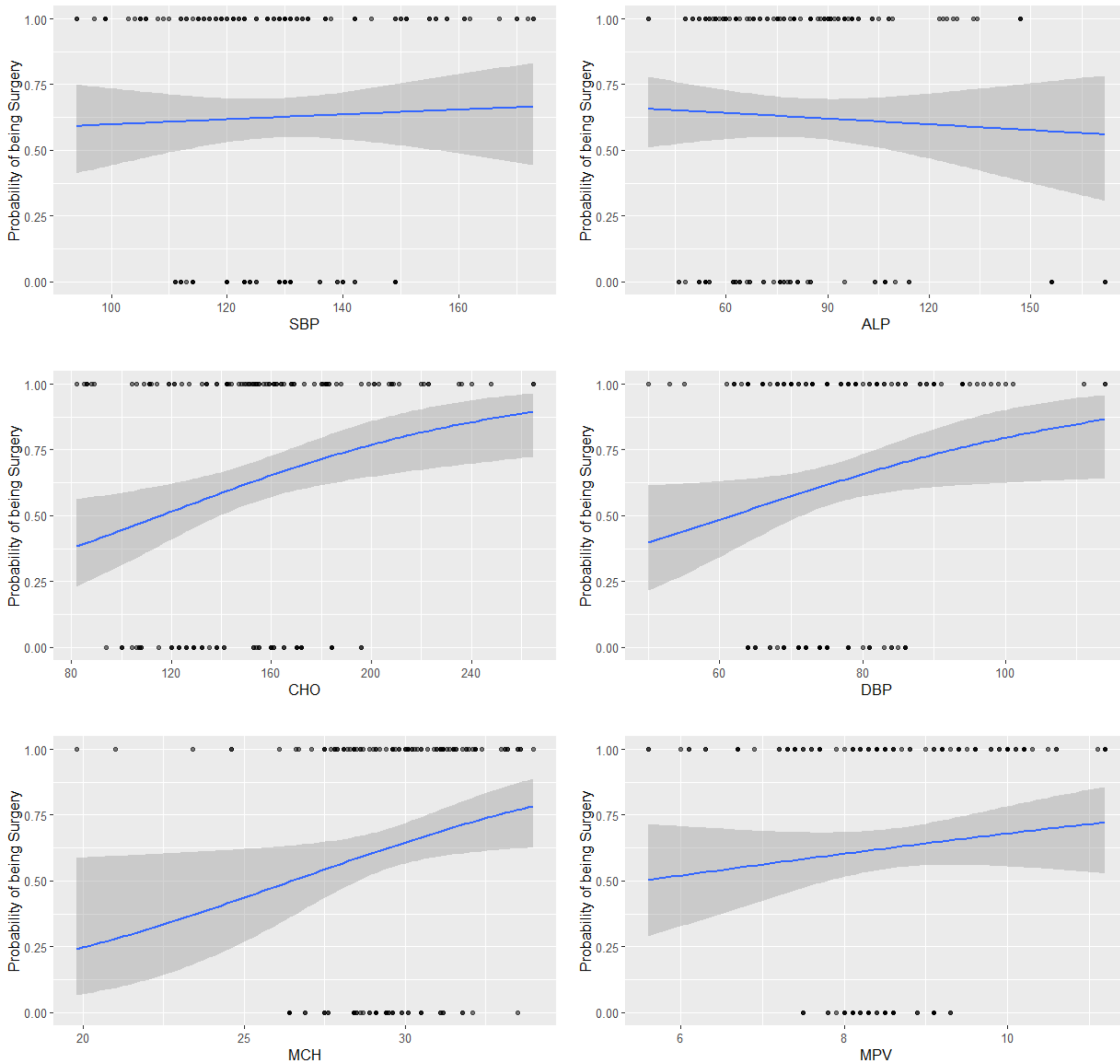

Fig. 1: Logit model fit of surgery on SBP, ALP, CHO, DBP, MCH, and MPV

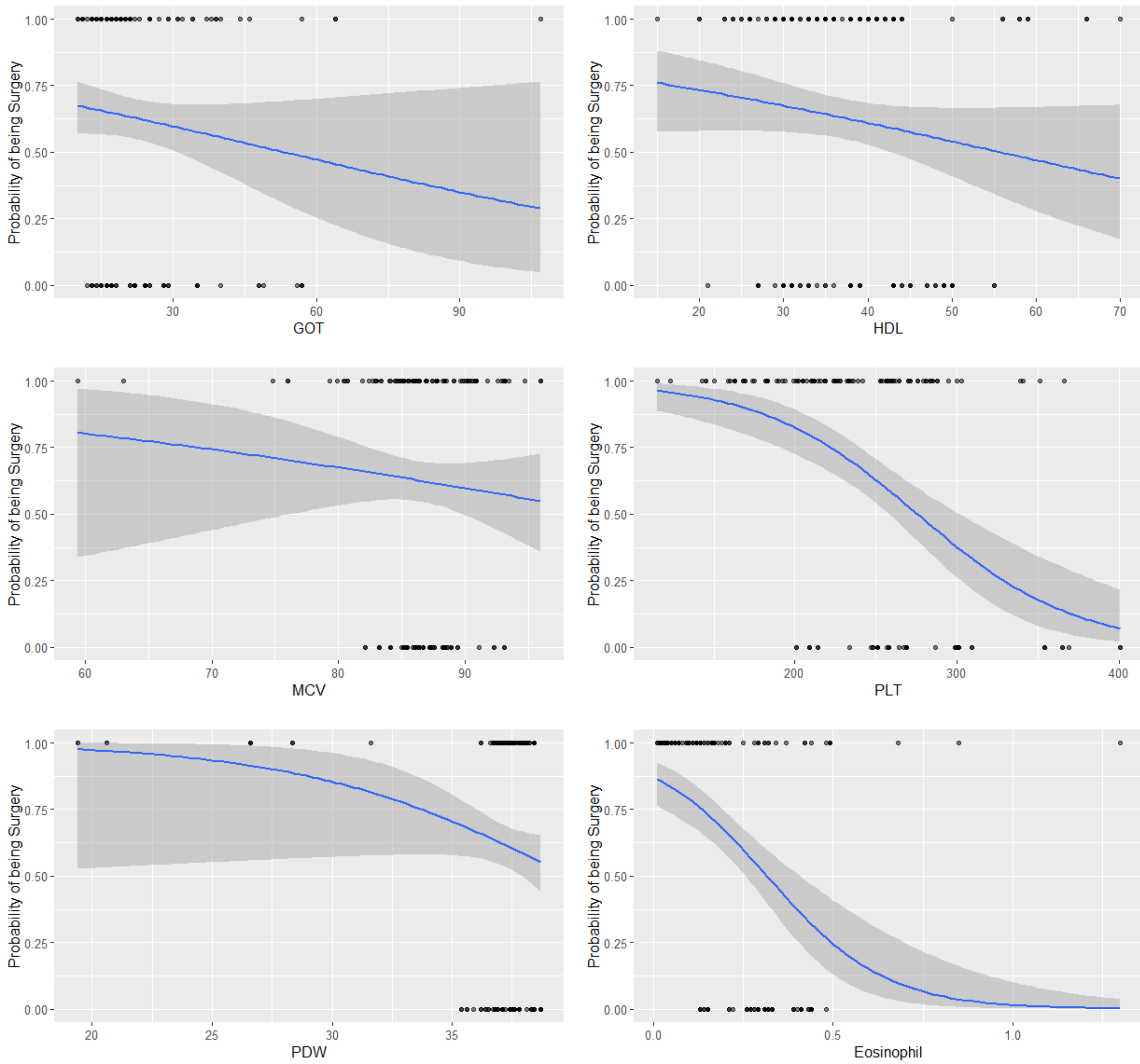

Figure 2: Logit model fit of surgery on GOT, HDL, MCV, PLT, PDW, and Eosinophil

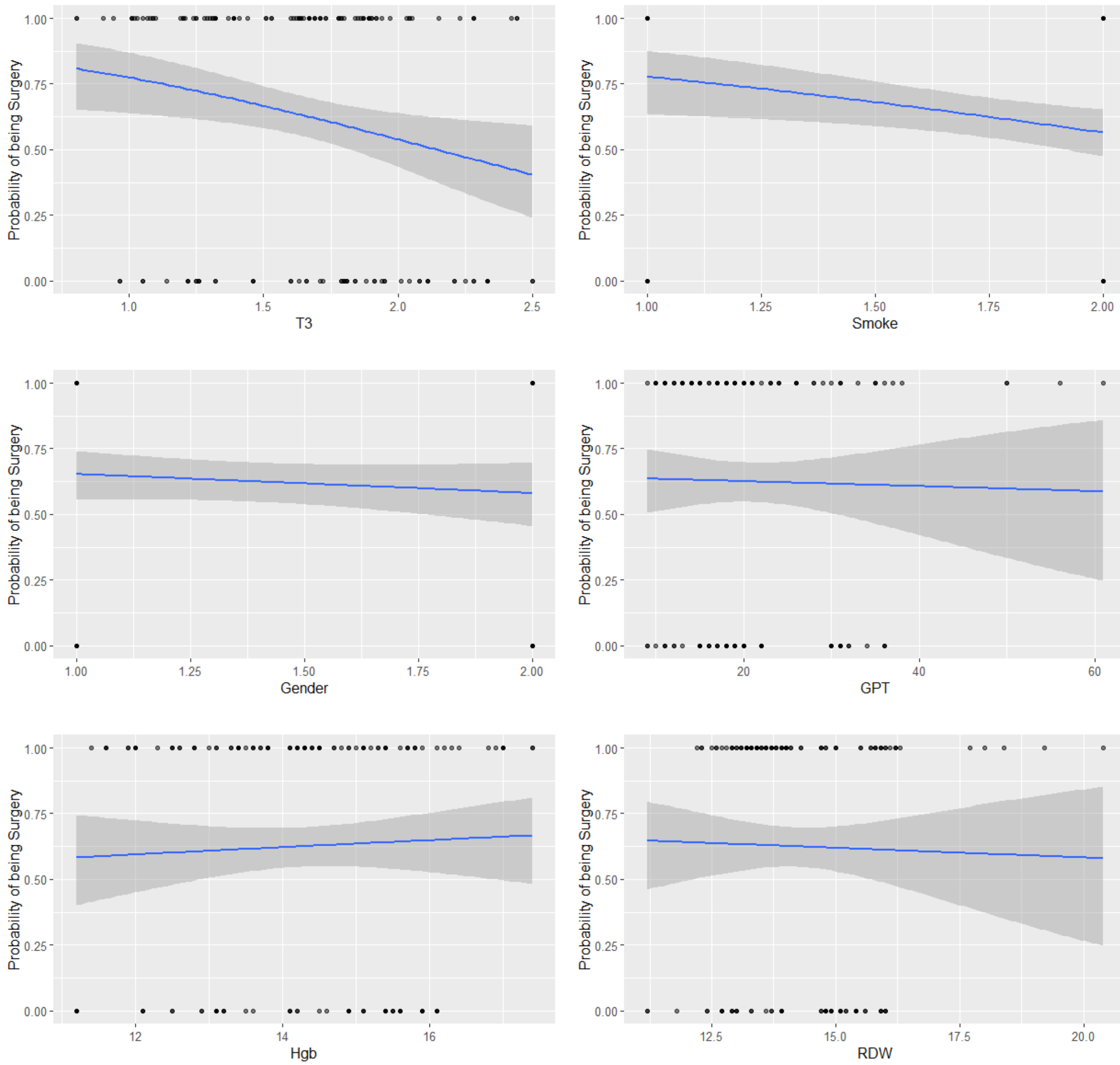

Fig. 3: Logit model fit of surgery on T3, Smoke, Gender, GPT, HGB, and RDW

پوختە

نەخۆشی شادەماری دڵ (CHD) دەکرێ وا پێناسە بکرێ کە نەخۆشییەکە، مادەی پلاک – ماددەیەکی مۆمییە –لەناو شادەمارەکانی دڵدا دروست دەبێت. ئامانجی ئەم توێژینەوەیە دیاریکردنی ئەو هۆکارە مەترسیدارانەیە کە کاریگەرییان لەسەر نەشتەرگەری Bypass Grafting هەیە. لەم توێژینەوەیەدا، سەد نەخۆشی پیگەیشتوو لە سەنتەری دڵ لە هەولێر کە هێڵکاری دەمارییان بۆکراوە و داتایان لێ وەرگیراوە. نەخۆشەکان تا کۆتایی دیسێمبەری 2020 لێکۆڵینەوە و بەدواداچونیان بۆکراوە. جگە لەمانە، داتاکانی ئەم توێژینەوەیە لە 60 کەسی ساغ و تەندروستەوە وەرگیراون کە هەمان پرۆسەی هێڵکاری دەمارییان بۆکراوە.

لەم توێژینەوەیدا Multiple logistic regression بەکارهێنرا بۆ دیاریکردنی ئەو هۆکارانەی کاریگەرییان لەسەر نەشتەرگەری Bypass Grafting هەیە. گۆڕاوەکان بە بەکارهێنانی میتۆدی هەڵبژاردنی پلە بە پلە (stepwise forward selection) نیشانکران. توێژینەوەکەمان ئەوەی بۆ دەرکەوتووە کە گۆڕاوەکانی HbA1c، تەمەن، BMI،WBC ، Eosinophil ، Blood Sugar ، MCH، PLT،DBP و T3 پەیوەندییەکی تاڕادەیەک سەرنجڕاکێشیان لەگەڵ Bypass Grafting هەیە بە بەکارهێنانی پشکنینی بینراو.

ئەو گۆڕاوە گرنگانەی کە بەشدارن لە پێشبینیکردنی Bypass Grafting بە بەکارهێنانی Multiple logistic regression بریتین لە DBP، Age , WBC, MCH وBlood Sugar. ئەو هۆکارە مەترسیدارانەی کە پەیوەندییان بە نەشتەرگەری Bypass Grafting هەیە بە بەکارهێنانی Bayesian logistic regression بریتیبوون لە Age ، WBC، و MCH.

لە ئەنجامی ئەم توێژینەوەیە کۆمەڵێک دەرەنجامی جیاواز بەدەستهاتووە لەنێوان Classical and Bayesian logistic models. پێوستە توێژینەوەی زیاتر لەوبارەیەوە بکرێ، تێیدا سامپڵی گەورەتر بەکاربێ و بەجۆرێک کە بەیسیەن مۆدێل Informative prior distribution لەخۆبگرێ.

له خوارەوە گرنگترین دەرەنجامەکانی توێژینەوەکە دەخەینەڕو:

نەخۆشی شادەماری دڵ (CHD) دەکرێ وا پێناسە بکرێ کە نەخۆشییەکە، مادەی پلاک - ماددەیەکی مۆمییە -لەناو شادەمارەکانی دڵدا دروست دەبێت. ئامانجی ئەم توێژینەوەیە دیاریکردنی ئەو هۆکارە مەترسیدارانەیە کە کاریگەرییان لەسەر نەشتەرگەری"Bypass Grafting" هەیە. لەم توێژینەوەیەدا، سەد نەخۆشی پیگەیشتوو لە سەنتەری دڵ لە هەولێر هیلکاری دەمارییان بۆکراوە و ٠داتایان لێ وەرگیراوە. نەخۆشەکان تا کۆتایی دیسێمبەری 2020 لێکۆڵینەوە و بەدواداچوونیان بۆکراوە. لەم توێژینەوەیە بۆمان دەرکەوتووە کە ئەو هۆکارە مەترسیدارانەی وابەستەن بە نەشتەرگەری CABG بریتین لە DBP، تەمەن، WBC،MCH و Blood Sugar بە بەکارهێنانی Multiple logistic regression.

ئێمە لەم توێژینەوەیەدا Bayesian logistic regression مان بەکارهێناوە بۆ ئەو هۆکارە گرنگانەی کە کاریگەرییان لەسەر نەشتەرگەری بای پاس هەیە. ئەم توێژینەوەیە ئەوەمان بۆ دەرکەوت کە دەرەنجامەکان جیاوازبوون لەClassical logistic regression ، بەهۆی ئەوەی کە دەرەنجامەکانی Odds Ratio جیاوازبون. ئەمەش بەهۆی ئەوەوەیە کە ئێمە Prior distribution نادیارمان لە Posterior distribution ئاخنیوە. ئەو فاکتەرە مەترسیدارانەی کە پەیوەستبوون بە نەشتەرگەری Bypass Grafting بەبەکارهێنانی Bayesian logistic regression بریتیبوون لە تەمەن، WBC وMCH پشکنینی گرافیکی ئەوەی دەریخست کە هەندێک Autocorrelation لە نێوان پارامێتەرە خەملێنراوەکاندا هەیە. بەمجۆرە، دووبارە پارامێتەرکردنی مۆدێلەکە یەکێکە لە ستراتیژییە سەرەکییەکان بۆ مامەڵەکردن و چارەسەرکردنی لەڕادەبەدەری Autocorrelation.

زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

پێشبینیکردنی هۆکارە مەترسیدارەکانی داتاکانی قەستەرەی دڵ لە هەولێر بە بەکارهێنانی شێوازی جێبەجێکاری بەیز و نابەیزی

نامەیەکە

پێشکەشی ئەنجومەنی کۆلێژی بەڕێوەبردن و ئابووری کراوە لە زانکۆی سەڵاحەددین- هەولێر وەکو بەشێک لە پێداویستیەکانی بە دەستهێنانی پلەی ماستەر لە زانستی ئامار

لەلایەن

ئەژین محمد خضر

٢٠١٥ – بەکالۆریۆس لە ئامار – زانکۆی سەڵاحەددین – هەولێر

بە سەرپەرشتیاری

پ.ی.د. دلێر حسین قادر

هەولێر, کوردستان

ئۆکتۆبەر 2022