STATISTICAL MODELLING OF CARDIOVASCULAR DISEASE PATIENTS USING BAYESIAN APPROACHES

NURLIYANA BINTI JUHAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2020

DEDICATION

This thesis is dedicated to my beloved parents and husband, who pray all day and night for my success, instilled in me the virtues of perseverance and commitment and relentlessly encouraged me to strive for excellence.

ACKNOWLEDGEMENT

First and foremost, all praises to Allah the Most Merciful and Most Compassionate for giving me the strength and opportunity to complete this doctoral thesis. I would like to express my deepest gratitude to my main supervisor Dr. Zarina Hj. Mohd Khalid for her guidance, great support, kind advice and motivation. She always gives valuable comments and insightful feedbacks.

I would also like to express my special thanks to my co-supervisor, Prof. Dr. Yong Zulina Zubairi for providing me with all the necessary facilities and continuous guidance throughout the period of this study. The discussions with her always provide me sufficient knowledge and encourage me to complete this study. Thank you for the endless amount of reading and editing my thesis drafts.

I also extend my thanks to my co-supervisor Assoc. Prof. Dr. Ahmad Syadi Mahmood Zuhdi; without his data, this study would not be possible.

Once again, I would like to thank all my supervisors, without their continued support and interest, this thesis would not have been the same as presented here.

A sincere gratitude goes to the University Malaysia Sabah and Ministry of Higher Education for providing the financial assistance to enable me to undertake this study.

A million of love and thanks to my dear parents and husband. I am very grateful for all the sacrifices, support and prayers. For my little sisters, thanks for caring and cheering me up.

Last but not least, I would like to express my appreciation to my best friends and labmates in the University of Malaya, dear friends in Universiti Teknologi Malaysia and lecturers for their support and assistance during my work in this study. May all the efforts be rewarded by Allah S.W.T.

ABSTRACT

This study focuses on statistical modelling on cardiovascular disease (CVD) patients in Malaysia. A secondary dataset from the National Cardiovascular Disease Database-Acute Coronary Syndrome (NCVD-ACS) registry for the years 2006 to 2013 is utilised. Studies have shown that CVD affects males and females differently. Thus, a gender-specific analysis with regard to the risk factors and mortality among ST-Elevation Myocardial Infarction (STEMI) patients is needed. Initially, this study performed the standard multivariate logistic analysis where the aims are to identify risk factors associated with mortality for each gender and to compare differences, if any, among STEMI patients. The results showed that gender differences existed among STEMI patients. Even though females share the same risk factors as males, there are risk factors that relate only to females which may have increased their tendency to develop and increase the risk of mortality of CVD patients. An important contribution of this analysis is that it gives an understanding of possible gender-based differences in baseline characteristics, risk factors, treatments and outcomes which will help cardiac care specialists in improving current management of patients with CVD. Next, Bayesian analysis is proposed to develop a prognostic model of the STEMI patients. Bayesian Markov Chain Monte Carlo (MCMC) simulation approach is applied. Beside that, comparisons of the parameter estimates from the proposed Bayesian and frequentist models are made. The results showed that the proposed Bayesian modelling can deal correctly with the probabilities and provides parameter estimates of the posterior distribution which have natural clinical interpretations. In doing so, several programming codes for the Bayesian model development and convergence diagnostics in the Just Another Gibbs Sampler (JAGS) software in R interface are developed. In the final part of this study, a graphical probabilistic model framework defined using a Bayesian Network (BN) is proposed to identify and interpret the dependence structure between the predictors and health outcomes of STEMI patients. In doing so, the two learning processes are involved in obtaining the BN model from the data namely the structural learning and parameter learning. From the structural learning, 25 and 20 arcs were considered significant for males' and females' BN respectively. A few variables namely, Killip class, renal disease and age group were classified as key predictors as they were the most influential variables directly associated with the outcome of patients' status. Moreover, conditional probabilities for each feature were obtained. The novelty of this study is that it provides an indication on the strength of each arc in the network by exploiting the bootstrap resampling method in the structural learning. A graphical model is developed where the relationships in a diagrammatical form is capable to be displayed and the cause-effect relationships can be illustrated. An important implication of this model is that it identifies dependencies based on the different features of variables. It can also include expert knowledge to improve predictability for data driven research when information or resources regarding the variables are limited.

ABSTRAK

Kajian ini memberi tumpuan kepada pemodelan statistik pesakit kardiovaskular (CVD) di Malaysia. Satu set data sekunder daripada Pangkalan Data Kebangsaan Penyakit Kardiovaskular-Sindrom Koronari Akut (NCVD-ACS) bagi tahun 2006 hingga 2013 telah digunakan. Kajian menunjukkan bahawa CVD memberi kesan kepada lelaki dan wanita secara berbeza. Dengan itu, analisis khusus-jantina berkaitan dengan faktor risiko dan kematian di kalangan pesakit Infarksi Miokardium ST-aras tinggi (STEMI) diperlukan. Pada mulanya, kajian ini menjalankan analisis logistik multivariat yang bertujuan untuk mengenal pasti faktor risiko yang berkaitan dengan kematian bagi setiap jantina dan membandingkan perbezaan jantina, jika ada, di kalangan pesakit STEMI. Hasilnya menunjukkan perbezaan jantina wujud di kalangan pesakit STEMI. Walaupun wanita mempunyai faktor risiko yang sama dengan lelaki, terdapat faktor risiko yang berkaitan hanya dengan wanita yang mungkin meningkatkan kecenderungan mereka untuk mendapat dan meningkatkan risiko kematian pesakit CVD. Sumbangan penting dalam analisis ini adalah ia memberikan kefahaman tentang kemungkinan perbezaan jantina dalam ciri asas, faktor risiko, rawatan dan hasil yang akan membantu pakar penjagaan jantung dalam meningkatkan pengurusan pesakit CVD masa kini. Seterusnya, analisis Bayesian dicadangkan untuk membangunkan model prognostik pesakit STEMI. Pendekatan simulasi Rantai Markov Monte Carlo (MCMC) Bayesian digunakan. Di samping itu, perbandingan anggaran parameter daripada model Bayesian yang dicadangkan dan model frekuentis telah dibuat. Hasil kajian menunjukkan bahawa pemodelan Bayesian yang dicadangkan boleh mengganggarkan kebarangkalian dengan betul dan menyediakan anggaran parameter bagi taburan posterior yang mempunyai tafsiran klinikal semulajadi. Oleh yang demikian, beberapa kod pengaturcaraan untuk pembangunan model Bayesian dan diagnostik penumpuan dalam perisian *Just Another Gibbs Sampler* (JAGS) dengan antaramuka R telah dibangunkan. Di bahagian akhir kajian ini, satu rangka kerja model kebarangkalian grafik yang ditakrifkan menggunakan Rangkaian Bayesian (BN) telah dicadangkan untuk mengenal pasti dan mentafsir struktur kebersandaran antara pesakit dan hasil kesihatan pesakit STEMI. Dengan berbuat demikian, dua proses pembelajaran telah terlibat dalam mendapatkan model BN dari data iaitu pembelajaran berstruktur dan pembelajaran parameter. Dari pembelajaran berstruktur, masing-masing 25 dan 20 lengkok dianggap penting bagi BN lelaki dan wanita. Beberapa pembolehubah iaitu kelas Killip, penyakit buah pinggang dan kumpulan umur diklasifikasikan sebagai peramal utama kerana pembolehubah-pembolehubah ini adalah yang paling berpengaruh secara langsung dengan hasil status pesakit. Selain itu, kebarangkalian bersyarat untuk setiap ciri telah diperolehi. Penemuan baru kajian ini adalah ia memberi petunjuk kepada kekuatan setiap lengkok dalam rangkaian dengan mengeksploitasi kaedah persampelan cangkuk but dalam pembelajaran berstruktur. Model grafik telah dibangunkan di mana hubungan dalam bentuk rajah mampu dipaparkan dan hubungan sebab-akibat boleh digambarkan. Implikasi penting dalam model ini ialah ia mengenalpasti kebersandaran berdasarkan ciri-ciri pembolehubah yang berbeza. Ia juga boleh memasukkan pengetahuan pakar untuk meningkatkan kebolehramalan untuk penyelidikan yang berasaskan data apabila maklumat atau sumber berkaitan pemboleubah adalah terhad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	-	Acute Coronary Syndrome
AIC	-	Akaike Information Criterion
ANCOVA	-	Analysis of Covariance
ANN	-	Artificial Neural Network
aSAH	-	Acute Aneurysmal Subarachnoid Haemorrhage
AUC	-	Area Under Receiver Operating Characteristic Curve
BDe	-	Bayesian Dirichlet equivalent
BDeu	-	Bayesian Dirichlet equivalent uniform
BMI	-	Body-Mass Index
BN	-	Bayesian Network
CI	-	Confidence Interval
CVD	-	Cardiovascular Disease
DAG	-	Directed Acyclic Graph
DIC	-	Deviance Information Criterion
ETS	-	Environmental Tobacco Smoke
FDA	-	Food and Drug Administration
FPR	-	False Positive Rate
GS	-	Grow-Shrink
GVIF	-	Generalized Variance-Inflation Factors
HL	-	Hosmer and Lemeshow
ICU	-	Intensive Care Unit
IgAN	-	Immunoglobulin A Nephropathy
JAGS	-	Just Another Gibbs Sampling
MC	-	Monte Carlo
MCMC	-	Markov Chain Monte Carlo
MetS	-	Metabolic Syndrome
MI	-	Myocardial Infarction
MLE	-	Maximum Likelihood Estimation
MMHC	-	Max-Min Hill-Climbing
MMPC	-	Max-Min Parents and Children

MOH	-	Ministry of Health
MREC	-	Medical Review & Ethics Committee
NCVD	-	National Cardiovascular Disease Database
NHMS	-	National Health and Morbidity Survey
NSAIDs	-	Non-Steroidal Anti-Inflammatory Drugs
NSTEMI	-	Non-ST Elevation Myocardial Infarction
OR	-	Odds Ratio
PCAD	-	Premature Coronary Artery Disease
PCI	-	Primary Percutaneous Coronary Intervention
pdf	-	Probability Density Function
PK-BUGS	-	Pharmacokinetic BUGS
PSRF	-	Potential Scale Reduction Factors
ROC	-	Receiver Operating Characteristic
SE	-	Standard Error
SHS	-	Second-Hand Smoke
STEMI	-	ST-Elevation Myocardial Infarction
TPR	-	True Positive Rate
UA	-	Unstable Angina
VIF	-	Variance Inflation Factor
WHO	-	World Health Organization
WHO/ISH	-	World Health Organization/ International Society of Hypertension

LIST OF SYMBOLS

χ^2	-	Chi-square
O_i	-	Number of observations of type *i*
n	-	Number of cells in the table for the Chi-square test
E_i	-	Expected frequency of type *i*
R_i^2	-	Coefficient of determination for a model where the *i*th variable is fit against all other predictor variables in the model
df	-	Degree of freedom
$t \in \mathbb{R}$	-	Real input in logistic function
$\pi(t)$	-	Predicted probability of mortality
$F(x)$	-	Probability of the dependent variable equalling a case
β_0	-	Intercept value
β_i	-	Regression coefficient for the *i*th variable
$x' = (x_0, x_1, x_2, ..., x_p)$	-	CVD risk factors associated to mortality of patients in the NCVD-ACS data
ℓ	-	Log-likelihood
y_i	-	Vector of observed class
a	-	Number of exposed cases
b	-	Number of exposed non-cases
c	-	Number of unexposed cases
d	-	Number of unexposed non-cases
G	-	Likelihood ratio test which follows a Chi-squared distribution with *k* degrees of freedom
k	-	Difference in the number of parameters between the two models
j	-	Iteration count
$p(\theta)$	-	Prior distribution

$p(y\|\theta)$	-	Likelihood function
$p(y)$	-	Distribution of the data
μ	-	Parameter of Bernoulli distribution
$\propto$	-	Directly proportional to
$\|$	-	Conditional event or given
W	-	Within chain variance
B	-	Between chain variance
$\hat{R}$	-	Potential scale reduction factors
m	-	Number of chains
n	-	Number of iterations
s_j^2	-	Variance of the *j*th chain
θ_{ij}	-	Model parameter *i* in *j*th chain
$\bar{\theta}_j$	-	Sample posterior mean
$V\hat{a}r(\theta)$	-	Estimated variance
$S(0)$	-	Spectral density evaluated at frequency zero
T_n	-	Empirical distribution function constructed from the sample x_j
α	-	Significance level in the hypothesis tests
G	-	Directed acyclic graph
V	-	A set of nodes in the directed acyclic graph
A	-	A set of directed arcs in the directed acyclic graph
X	-	A set of random variables in the directed acyclic graph
$\perp\perp_G$	-	d-separate
$\rightarrow$	-	Direction arrow (to)
v	-	Intermediate node
B	-	Bayesian network
D	-	Dataset used in the Bayesian network
Θ	-	Parameters of the global distribution
$P(B\|D) = P(G,\Theta\|D)$	-	Bayesian network learning
$P(G\|D)$	-	Structural learning

$P(\Theta \mid G, D)$	-	Parameter learning
α	-	Imaginary sample size
r_i	-	Number of categories for the node X_i
q_i	-	Number of configurations of the categories of the parents of X_i
n_{ijk}	-	Number of samples which have the jth category for node X_i and the kth configuration for its parents
G^*	-	Modified network
M	-	Number of bootstrap resamples
P_B	-	Joint probability of a CVD event in relation with a set of risk factors using BN
$\theta_{X_i \mid \Pi_{X_i}}$	-	Joint probability distribution over risk factor X_i
s	-	Score
$f_k(s)$	-	Probability density function of the scores for class k
$F_k(s)$	-	Cumulative density function of the scores for class k
$F_0(t)$	-	Sensitivity
$F_1(t)$	-	Specificity
G_{HL}	-	Hosmer and Lemeshow test statistic
$\hat{\pi}_{ij}$	-	Estimated probability for observation j in group i
D^*	-	Deviance
$\log[L_m(\hat{B})]$	-	Maximized log-likelihood of the fitted model
$\log[L_s(\hat{B})]$	-	Maximized log-likelihood of the saturated model
$\forall$	-	For all

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Cardiovascular disease (CVD) is the number one cause of death in Malaysia (Mohammad *et al.*, 2018) and globally (WHO, 2017; Gutierrez *et al.*, 2018). CVD is defined as a group of disorders of the heart and blood vessels which include coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease and congenital heart disease (WHO, 2017). More people die annually from this non-communicable disease than from any other cause. Of these deaths, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke (WHO, 2017). It was estimated that 23.6 million people will die by the year 2030 due to CVD (Benjamin *et al.*, 2017). Amongst the more developed countries, the highest death rates from CVD are in Ukraine and Russian Federation with 718 and 654 deaths per 100,000 population respectively, while the lowest are in South Korea and Japan with 36.5 and 47.0 deaths per 100,000 respectively (Mozaffarian *et al.*, 2015).

Although various research has been done on CVD worldwide, it is important to better understand the disease pattern in Malaysia specifically and analyse the impacts of the study on clinical practice locally. It is worthwhile to note that, CVD accounted for 98.9 deaths per 100,000 population in Malaysia in 2012, or 29,400 deaths which is 20.1% of all deaths (WHO, 2017). Even worse, CVD remain as a principal cause of death in Malaysia for the last ten years, from 2005 to 2014 (Department of Statistics Malaysia Official Portal, 2016). Ample information on the burden of disease has also been obtained from death certifications and hospital admission records from the Malaysia Ministry of Health (MOH) hospitals where circulatory disease accounted for 6.99% of the total hospital admissions and 23.34% of all hospital deaths in 2014 (Ang and Chan, 2016; Ministry of Health Malaysia, 2016).

In Malaysia, the National Cardiovascular Disease Database (NCVD), a service supported by the MOH, plays an important role to collect information about CVD across Malaysia. NCVD enables us to obtain the incidence of CVD, and to evaluate its risk factors and treatment in the country. This information is useful in supporting the MOH, non-governmental organizations, private healthcare providers and industry in programme planning and evaluation, leading to CVD prevention and control. The NCVD is established to integrate various existing databases for individual hospitals either in MOH hospitals or private and other data sources to achieve a nationwide cardiovascular database.

The Acute Coronary Syndrome (ACS) registry was officially launched on 31st March 2006. The Malaysian National Cardiovascular Disease-Acute Coronary Syndrome (NCVD-ACS) registry was the first stage in realising the rationale of a nationwide cardiovascular database. Previously, the risk prediction of ACS is unclear and might be different from cardiovascular disease patients with chronic stable angina. These are the reasons why NCVD-ACS registry is needed, as it provides the real-life data that would represent the population of Malaysia (Wan Ahmad and Sim, 2006).

Prevention of CVD requires timely identification of patients at increased risk to target effective dietary, lifestyle, drug interventions or treatments (WHO, 2016). These can be done by studying the risk factors, analysing and interpreting them using various statistical methods. Therefore, the existence of high-quality data as well as suitable statistical methods to analyse data is of significant importance.

Understanding a disease often requires research that examines multiple variables and their relationships. These include clinical and laboratory data and other attributes such as risk factors, socioeconomic factors among others. In order to integrate these multiple variables, statistical model is often used where it can be expressed in mathematical expression that portrays the relationships among the variables. The model is able to provide prediction and explanation on the nature of the illness. One commonly used model is the regression model where we understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed (Suárez *et al.*, 2017).

Over the past few years, numerous regression models have been developed for predicting risk factors such as models based on logistic regression model (Awad and Al-Nafisi, 2014; Umamahesh *et al.*, 2014; Jousilahti *et al.*, 2016; Keto *et al.*, 2016; Yuan *et al.*, 2017) and Cox regression model (McClelland *et al.*, 2015; Wulsin *et al.*, 2015). This study commences by considering the logistic regression model with independent variables that include the risk factors associated with mortality among Malaysian CVD patients. For initial analysis, univariate analysis is performed to identify significant variables. This is followed by the multivariate analysis using the purposeful selection method to obtain the best model.

Using various statistical methods, this study attempts to model the data by validating, making head-to-head comparisons, and thus providing a parsimonious model that describes the data. Bayesian model is also considered to describe the data where Bayesian Markov Chain Monte Carlo (MCMC) simulation approach is applied in the analysis. The dataset used for this study is the NCVD-ACS dataset from the year of its establishment, 2006 until 2013.

Model performance is assessed through convergence diagnostics, overall model fit, model calibration and discrimination. Additionally, comparisons of the parameter estimates are made between Bayesian model and frequentist model. To our knowledge, Bayesian model using the MCMC method has not been used extensively in the analysis of CVD data in Malaysia.

Also, another aim of the study is to identify the dependence structures between variables and the outcome of CVD patients, graphical model based on the Bayesian network (BN) approach has been considered. The BN approach incorporated a few learning techniques namely structural learning and parameter learning. In addition, this study applied a bootstrap resampling approach to the structure learning, in the interest of estimate the strength of each identified dependence. Validation and the performance of the BN model are assessed using the area under receiver operating characteristic curve (AUC) and the accuracy test.

1.2 Problem Statement

A few studies have been carried out for the CVD datasets worldwide such as the treatment attitude and clinical outcome of elderly patients with ACS (Zuhdi *et al.*, 2016), prediction on CVD mortality among working men and women (Jousilahti *et al.*, 2016), determination of the population trends of CVD risk factors in the national capital region (Prabhakaran *et al.*, 2017) and among others. However, successful implementation of statistical methods remains a challenge among researchers as they are many common mistakes applied especially in biomedical research such as improper research design, inadequate sample size, unsuitable statistical test and overfitting regression model (Nuzzo, 2014; Moyé, 2016). A review of the literature reveals general awareness of this issue (George, 1985; Ioannidis, 2005; Nuzzo, 2014a; Moyé, 2016). In medical research, the purpose of statistical model is often used to describe relationships between variables and identify risk factors associated with the outcome of some illness; for example, Shah *et al.* (2015), examined if the likelihood of having coronary artery disease is influenced by factors involving age and smoking. Measures of association provide an initial impression of the extent of statistical dependence between variables. Besides, it is important to eliminate and evaluate the best set of variables to be used for building predictive models and identify the risk factors that influence the outcome.

In Malaysia, there is a lack of comprehensive studies on the risk factors of CVD. Previous studies of CVD only focused on treatment attitude (Zuhdi *et al*., 2016; Venkatason *et al*., 2016) and the samples are based only on Kuala Lumpur's population (Amiri *et al.*, 2014) and limited to certain states such as Kelantan (Mohd Noor *et al.*, 2013) and Pahang (Mohammad *et al.*, 2018). Moreover, there were also studies which focused only on certain age group such as adolescents (Thangiah *et al.*, 2017) and elderly (Azahar *et al.*, 2016). Most of these studies used descriptive and regression analyses. Besides, the prevalence of CVD in males is higher than in females and mortality rates are also different between the two genders (Department of Statistics Malaysia Official Portal, 2016). This suggest that risk factors associated with mortality between males and females are different. Hence, this study will demonstrate the development of logistic regressions using the NCVD-ACS dataset particularly on the ST-Elevation Myocardial Infarction (STEMI) patients in Malaysia. Based on the regression model, we would like to find out what are the risk factors and what is the prognosis of the patients. Is there any difference between male and female patients' prognosis?

An alternative to traditional frequentist statistical analysis is the Bayesian approach. The flexibility of the Bayesian approach is that we are able to update prior information on the underlying parameters with information from cumulative or past experience (Ntzoufras, 2009; Torman & Camey, 2015) and this has become the motivation in this study. A search through the literature revealed that implementation of Bayesian approach in CVD dataset are somewhat limited in Malaysia. Difficulties in performing Bayesian analyses for multivariate data and the availability of other easier methods are some of the reasons why the Bayesian approach is less popular compared to other traditional methods. Bayesian approach rests on the assumption that all model parameters are random quantities and hence can incorporate prior knowledge. Its use in predicting risk of mortality in CVD has been rather underutilized (Hannan *et al.*, 2005) even though this method is widely used for predictive analysis in other medical applications such as modelling risk of death in an intensive care unit (Wong and Ismail, 2016), identifying risk genes for schizophrenia and neurodevelopmental disorders (Nguyen *et al.*, 2017) and multiple treatment comparisons in female urinary incontinence (Carlin *et al.*, 2013). In this study, we will incorporate prior knowledge and investigate how does this affect the risk factors, the prognosis of the patients and how does this improve the prediction of the model. To our knowledge, there is no published study that discussed the performance of Bayesian approach using Malaysian medical data, particularly CVD patients in the NCVD-ACS registry. Therefore, this study considers MCMC approach where development of Bayesian model using Malaysian STEMI patients from the NCVD-ACS dataset is shown.

Modelling the risk factors associated with mortality of CVD taking all the related variables into account is rather challenging from a theoretical point of view, as the risk factors-mortality nexus is expected to be characterised by a rather complex dependence structure. Obvious computational challenges can be seen, especially when it involves quite a large dataset in which it requires long processing times and can only be performed with an adequate computer infrastructure in place (Torman and Camey, 2015; Liu, 2018). When research is driven mostly by available data and resources or information regarding the variables are limited, a model that can essentially identify dependencies based on the different features of variables and can include expert knowledge to improve predictability is needed. In this setting, graphical model such as Bayesian network seem to be a perfect fit. This model was increasingly being used in computer science (Lacave *et al.*, 2018; Scutari *et al.*, 2018), business analytics problems (van Wagenberg *et al.*, 2018) as well as medicine (Foroushani *et al.*, 2016; Zador *et al.*, 2017), and genetics (Scutari *et al.*, 2014). Both categorical and numerical data were involved in these previous studies. However, all of these studies have been based on international samples such as Spain, United State of America, United Kingdom and Iran and this model is relatively less explored in Malaysian CVD dataset. Also, most of these Bayesian network studies do not contain indication on the degree of confidence or strength of each arc in the network. In order to address this limitation, a graphical model based on Bayesian network will be applied to discover the dependence structure between variables in Malaysian CVD dataset and the conditional independence of a variable or groups of variables from a given variable or variables. Bootstrap resampling method will be considered in this study in which the degree of confidence of each arc can be obtained. In short, we would like to know what are the relationship defined by the Bayesian network where the prognosis is different.

In machine learning, fitting a model to our training dataset is one thing (Witten *et al.*, 2016), but how do we know that it generalizes well to unobserved data? How to make sure that it doesn't simply memorize the data we feed it and able to make good predictions on future samples? In order to address this constraint, this study will assess the performance of both proposed Bayesian regression and Bayesian network models and carry out the validation studies to identify whether the predictions are sufficiently accurate across different settings and populations.

1.3 Research Questions

In this study, we aim to address the following research questions:

(a) What are the CVD risk factors associated with mortality among the STEMI patients in Malaysia? Is there any difference for each gender?

(b) What is the alternative technique in identifying prognostic factor of STEMI patients?

(c) How well can the proposed statistical model fit the Malaysian STEMI data?

(d) What is the suitable model to describe the dependence structure between the predictors and health outcomes of STEMI patients?

(e) How accurate the proposed model in identifying the dependencies between variables?

1.4 Objective of the Study

The objectives of the study are as follow:

(a) To identify the CVD risk factors associated with mortality for each gender and compare differences, if any, among STEMI patients in Malaysia using frequentist approach.

(b) To propose prognostic model using Bayesian MCMC method for STEMI patients.

(c) To assess Markov chains convergence, model performance and result validation of the proposed Bayesian model.

(d) To establish a framework of dependence structure for STEMI patients based on graphical probabilistic model defined using a Bayesian network.

(e) To assess the accuracy of the proposed Bayesian network model.

1.5 Significance of the Study

In general terms, the analyses will give an accurate information of the illness nature for patients with CVD in Malaysia. The findings from this study will be beneficial in the following ways:

(a) This study gives an understanding of CVD-STEMI cases in Malaysia and provides accurate information to clinicians.

(b) This study provides predictive models using frequentist approach and Bayesian MCMC where risk factors, outcome of illness and prognosis of the disease can be identified.

(c) A graphical model is developed where it is capable of displaying relationships in a diagrammatical form and the cause-effect relationships can be illustrated to the clinicians.

(d) It is hoped that the results of the statistical analyses will be of use to clinicians treating CVD-STEMI patients by providing robust decision support system.

1.6 Scope of Study

The scope of the study is on patients with acute coronary syndromes (ACS) who are registered with the National Cardiovascular Disease Database (NCVD) with the main focus on ST-Elevation Myocardial Infarction (STEMI). The analyses were done on the information obtained from 2006 to 2013.

1.7 Organization of the Thesis

This thesis is divided into seven chapters:

(a) **Chapter 1** provides the background of the study, problem statement, objectives, significance of the study, scope and outline of chapters in the thesis.

(b) **Chapter 2** presents the literature reviews related to the study.

(c) **Chapter 3** elaborates the background of the source of data and the methodology for the construction of the proposed models in this study.

(d) **Chapter 4** presents the results and findings of modelling the mortality of STEMI male and female patients using multivariate logistic regression.

(e) **Chapter 5** discusses the performance of the proposed Bayesian modelling of STEMI patients using MCMC approach. This is followed by convergence diagnostics for the Markov chains and comparison between Bayesian and frequentist estimates in the final model.

(f) **Chapter 6** focuses on the development of BN models and parameter estimation for the local distributions. The performances of the BN models are also discussed.

(g) **Chapter 7** summarises and concludes the overall findings of this research. This chapter also includes the research contribution to knowledge and recommendations for future work.

REFERENCES

Abdollahi, A., Qorbani, M., Salehi, A. and Mansourian, M. (2009) 'ABO Blood Groups Distribution and Cardiovascular Major Risk Factors in Healthy Population', *Iranian J Publ Health*, 38(3), pp. 123–126.

Achcar, J. A., Martinez, E. Z. and Louzada-Neto, F. (2004) 'Binary data in the presence of misclassification', *Proceedings of 16th Symp. Int. Assoc. Statist. Comput.* Prague, Czech Republic: Physica-Verlag, Springer, pp. 581–588.

Agarwal, A. (2016) *Bayesian Variable Selection With Spike-And-Slab Priors*. The Ohio State University.

Agresti, A. (2019) *An introduction to categorical data analysis*. Third Edit. New Jersey: John Wiley & Sons, Inc.

Ahmadi, A., Soori, H. and Sajjadi, H. (2015) 'Modeling of in hospital mortality determinants in myocardial infarction patients, with and without type 2 diabetes, undergoing pharmaco-invasive strategy: the first national report using two approaches in Iran.', *Diabetes research and clinical practice*, 108(2), pp. 216–22.

Akobeng, A. K. (2007) 'Understanding diagnostic tests 3: receiver operating characteristic curves', *Acta Paediatrica*, 96(5), pp. 644–647.

Albers, C. J., Kiers, H. A. L., Van Ravenzwaaij, D., Vazire, S. and Savalei, V. (2018) 'Credible Confidence: A Pragmatic View on the Frequentist vs Bayesian Debate', *Collabra: Psychology*, 4(1).

Alfers, D. and Dinges, H. (1984) 'for Beta and Gamma Tail Probabilities', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65(3), pp. 399–419.

Alissa, E. M., Alnahdi, W. A., Alama, N. and Ferns, G. A. (2015) 'Bone mineral density and cardiovascular risk factors in postmenopausal women with coronary artery disease', *BoneKEy Reports*, 4, pp. 1–7.

Allahyari, E., Jafari, P. and Bagheri, Z. (2016) 'A Simulation Study to Assess the Effect of the Number of Response Categories on the Power of Ordinal Logistic Regression for Differential Item Functioning Analysis in Rating Scales', *Computational and Mathematical Methods in Medicine*, 2016, pp. 1–8.

Altman, D. G. and Bland, J. M. (1994) 'Diagnostic tests. 1: Sensitivity and specificity', *BMJ: British Medical Journal*, 308(6943), p. 1552.

Álvaro-Meca, A., Gil-Prieto, R. and Gil de Miguel, A. (2011) 'In-hospital death caused by pancreatic cancer in Spain: Application with a bayesian network', *International Journal of Biomedical Science*, 7(2), pp. 125–130.

Amiri, M., Majid, H. A., Hairi, F., Thangiah, N., Bulgiba, A. and Su, T. T. (2014) 'Prevalence and determinants of cardiovascular disease risk factors among the residents of urban community housing projects in Malaysia', *BMC Public Health*, 14(Suppl 3), pp. 1–9.

Ang, C. and Chan, K. (2016) 'A Review of Coronary Artery Disease Research in Malaysia', *The Medical Journal of Malaysia Hospitals*, 2013(June), pp. 42–57.

Angula, H. N., Stuart-Hill, G., Ward, D., Matongo, G., Diggle, R. W. and Naidoo, R. (2018) 'Local perceptions of trophy hunting on communal lands in Namibia', *Biological Conservation*, 218, pp. 26–31.

Arnold, R., Hayakawa, Y. and Yip, P. (2010) 'Capture-recapture estimation using finite mixtures of arbitrary dimension', *Biometrics*, 66(2), pp. 644–655.

Awad, A. and Al-Nafisi, H. (2014) 'Public knowledge of cardiovascular disease and its risk factors in Kuwait: A cross-sectional survey', *BMC Public Health*, 14(1), pp. 1–11.

Azahar, N., Ismail, Z., Hassan, R. and Yusoff, K. (2016) 'Mismatch between actual risk and perceived risk for cardiovascular disease among elderly in a rural population of Malaysia', *Journal of Hypertension*, 34, p. e459.

Bae, S., Kim, N. H., Park, C. and Kim, Z. (2017) 'Confidence Interval of Bayesian Network and Global Sensitivity Analysis', *AIAA Journal*, 55(11), pp. 3916–3924.

Baldwin, S. A. and Larson, M. J. (2017) 'An introduction to using Bayesian linear regression with clinical data', *Behaviour Research and Therapy*, 98, pp. 58–75.

Bally, M., Dendukuri, N., Rich, B., Nadeau, L., Helin-Salmivaara, A., Garbe, E. and Brophy, J. M. (2017) 'Risk of acute myocardial infarction with NSAIDs in real world use: bayesian meta-analysis of individual patient data.', *BMJ (Clinical research ed.)*, 357, pp. 1909.

Banks, A. D. (2008) 'Women and Heart Disease: Missed Opportunities', *Journal of Midwifery & Women's Health*, 53(5), pp. 430–439.

Barber, J. J., Gelfand, A. E. and Silander, J. A. (2006) ‘Modelling map positional error to infer true feature location’, *Canadian Journal of Statistics*, 34(4), pp. 659–676.

Barnett, A. G., Van Der Pols, J. C. and Dobson, A. J. (2005) ‘Regression to the mean: what it is and how to deal with it’, *International Journal of Epidemiology*, 34, pp. 215–220.

Bayes, T. (1763) ‘An essay towards solving a problem in the doctrine of chances’, *Philosophical transactions of the Royal Society of London*, 53, pp. 370–418.

Bédard, M., Fraser, D. A. S. and Wong, A. (2007) ‘Higher Accuracy for Bayesian and Frequentist Inference: Large Sample Theory for Small Sample Likelihood reprint differs from the original in pagination and typographic detail. 1 2’, *Statistical Science*, 22(3), pp. 301–321.

Bedrick, E. J., Christensen, R., Johnson, W., Bedrick, E. J., Christensen, R. and Johnson, W. (1997) ‘Bayesian Binomial Regression: Predicting Survival at a Trauma Center’, *The American Statistician*, 51(3), pp. 211–218.

Bellman, R. E. (2015) *Adaptive Control Processes: A Guided Tour (Vol.2045)*. Princeton university press.

Bendel, R. B. and Afifi, A. A. (1977) ‘Comparison of Stopping Rules in Forward Stepwise Regression’, *Journal of the American Statistical Association*. Taylor & Francis, Ltd.American Statistical Association, 72(357), p. 46.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E. and Cushman, M. (2017) ‘Heart Disease and Stroke Statistics—2017 Update’, *Circulation*, 135(10), pp. e146–e603.

Berg, B. A. (2004) *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*. World Scientific Publishing Company.

Bertuzzi, M., Negri, E., Tavani, A. and La Vecchia, C. (2003) ‘Family history of ischemic heart disease and risk of acute myocardial infarction.’, *Preventive medicine*, 37(3), pp. 183–7.

Besag, J., York, J. and Molli, A. (1991) ‘Bayesian image restoration, with two applications in spatial statistics’, *Annals of the Institute of Statistical Mathematics*, 43(1), pp. 1–20.

Bhatt, D. L., Ohman, E. M., Hirsch, A. T., Richard, A. J. and Wilson, P. W. F. (2006) ‘International Prevalence, Recognition, and Treatment of Cardiovascular Risk Factors in Outpatients With Atherothrombosis’, *JAMA : the journal of the American Medical Association*, 295(2).

Blalock, H. M. (1961) ‘Correlation and Causality: The Multivariate Case’, *Social Forces*, 39(3), pp. 246–251.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) ‘Latent Dirichlet Allocation’, *Journal of Machine Learning Research*, 3, pp. 993–1022.

Blomster, J. I., Woodward, M., Zoungas, S., Hillis, G. S., Harrap, S., Neal, B., Poulter, N., Mancia, G., Chalmers, J. and Huxley, R. (2016) ‘The harms of smoking and benefits of smoking cessation in women compared with men with type 2 diabetes: An observational analysis of the ADVANCE (Action in Diabetes and Vascular Disease: Preterax and Diamicron modified release Controlled Evaluation) trial’, *BMJ Open*, 6(1), pp. 1–7.

Bodnar, O., Link, A., Arendacká, B., Possolo, A. and Elster, C. (2017) ‘Bayesian estimation in random effects meta-analysis using a non-informative prior’, *Statistics in Medicine*, 36(2), pp. 378–399.

Bouckaert, R. R. (1995) *Bayesian Belief Networks: From Construction to Inference*. PhD Thesis, Universiteit Utrecht.

Brier, G. W. (1950) ‘Verification of forecasts expresed in therms of probability: Brier Score’, *1950*, 78(1), pp. 1–3.

Briggs, A. and Sculpher, M. (1998) ‘An introduction to Markov modelling for economic evaluation’, *PharmacoEconomics*, 13(4), pp. 397–409.

Brooks, S. P. and Gelman, A. (1998) ‘General Methods for Monitoring Convergence of Iterative Simulations’, *Journal of Computational and Graphical Statistics*, 7(4).

Broom, B. M., Do, K.-A. and Subramanian, D. (2012) ‘Model averaging strategies for structure learning in Bayesian networks with limited data’, *BMC Bioinformatics*, 13.

Bursac, Z., Gauss, C. H., Williams, D. K. and Hosmer, D. W. (2008) ‘Purposeful selection of variables in logistic regression’, *Source Code for Biology and Medicine*, 3, pp. 1–8.

Carlin, B. P., Hong, H., Tatyana Shamliyan, candidate A., François Sainfort, M. and Kane, R. L. (2013) ‘Case Study Comparing Bayesian and Frequentist Approaches for Multiple Treatment Comparisons Methods Research Report Case Study Comparing Bayesian and Frequentist Approaches for Multiple Treatment Comparisons’, Rockville, MD: Agency for Healthcare Research and Quality.

Chang, Y., Sharp, G. C., Li, Q., Shih, H. A., El Fakhri, G., Ra, J. B. and Woo, J. (2018) ‘Subject-specific brain tumor growth modelling via an efficient Bayesian inference framework’, in Angelini, E. D. and Landman, B. A. (eds) *Medical Imaging 2018: Image Processing*, pp. 89.

Chao, Y.-S., Wu, H., Scutari, M., Chen, T.-S., Wu, C.-J., Durand, M. and Boivin, A. (2017) ‘A network perspective on patient experiences and health status: the Medical Expenditure Panel Survey 2004 to 2011’, *BMC Health Services Research*. BMC Health Services Research, 17(1), pp. 579.

Chen, L. (2009) ‘Curse of Dimensionality’, in *Encyclopedia of Database Systems*. Boston, MA: Springer US, pp. 545–546.

Cheung, S. H. and Bansal, S. (2017) ‘A new Gibbs sampling based algorithm for Bayesian model updating with incomplete complex modal data’, *Mechanical Systems and Signal Processing*, 92, pp. 156–172.

Chiaka, E. S., Adam, M. B., Krishnarajah, Isthrinayagy Shohaimi, S. and Guure, C. B. (2015) ‘Bayesian Logistic Regression Model on Risk Factors of Type 2 Diabetes Mellitus’, *Mathematical Theory and Modeling*, 5(1), pp. 2224–5804.

Chin, S. P., Jeyaindran, S., Azhari, R., Wan Azman, W. A., Omar, I., Robaayah, Z. and Sim, K. H. (2008) ‘Acute coronary syndrome (ACS) registry-leading the charge for National Cardiovascular Disease (NCVD) Database.’, *Med J Malaysia*, 63(Suppl C, pp. 29–36.

Chomistek, A. K., Chiuve, S. E., Eliassen, A. H., Mukamal, K. J., Willett, W. C. and Rimm, E. B. (2015) ‘Healthy Lifestyle in the Primordial Prevention of Cardiovascular Disease Among Young Women’, *Journal of the American College of Cardiology*, 65(1), pp. 43–51.

Chu, H. and Cole, S. R. (2010) ‘Estimation of risk ratios in cohort studies with common outcomes: A bayesian approach’, *Epidemiology*, 21(6), pp. 855–862.

Cífková, R. and Krajčoviechová, A. (2015) ‘Dyslipidemia and Cardiovascular Disease in Women’, *Current Cardiology Reports*, 17(7).

Claeskens, G. and Hjort, N. L. (2008) *Model Selection and Model Averaging*, *Cambridge Books*. Cambridge University Press.

Cleophas, T. J. and Zwinderman, A. H. (2018) *Modern Bayesian Statistics in Clinical Research*. Switzerland: Springer.

Cole, S. R., Chu, H. and Greenland, S. (2014) ‘Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer’, *American Journal of*

Epidemiology. Oxford University Press, 179(2), pp. 252–260.

Cooper, G. F. and Herskovits, E. (1992) 'A Bayesian method for the induction of probabilistic networks from data', *Machine Learning*, 9(4), pp. 309–347.

Costner, H. L. (1969) 'Theory, Deduction, and Rules of Correspondence', *American Journal of Sociology*, 75(2), pp. 245–263.

Cox, D. R. (2018) *Analysis of Binary Data, Second Edition*. Routledge.

Damen, J. A. A. G., Hooft, L., Schuit, E. and Debray, T. P. A., (2016) 'Prediction models for cardiovascular disease risk in the general population: Systematic review', *BMJ (Online)*, 353.

Department of Statistics Malaysia Official Portal (2016) *Statistics on Causes of Death, Malaysia, 2014*. Retrieved December 5, 2016, from https://www.dosm.gov.my/v1/index.php

DeVon, H. A., Pettey, C. M., Vuckovic, K. M., Koenig, M. D. and McSweeney, J. C. (2016) 'A Review of the Literature on Cardiac Symptoms in Older and Younger Women.', *Journal of obstetric, gynecologic, and neonatal nursing*, 45(3), pp. 426–437.

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P. and Sutton, A. J. (2018) *Network meta-analysis for decision making*. John Wiley & Sons.

Digiacomo, S. I., Jazayeri, M. A., Barua, R. S. and Ambrose, J. A. (2019) 'Environmental tobacco smoke and cardiovascular disease', *International Journal of Environmental Research and Public Health*, 16(1), p. 96.

Ding, T., Hardiman, P. J., Petersen, I. and Baio, G. (2018) 'Incidence and prevalence of diabetes and cost of illness analysis of polycystic ovary syndrome: a Bayesian modelling study', *Human Reproduction*, 33(7), pp. 1299–1306.

Döbler, C. (2015) 'Stein's method of exchangeable pairs for the Beta distribution and generalizations', *Electronic Journal of Probability*, 20(109), pp. 1–34.

Ducher, M., Kalbacher, E., Combarnous, F., Vilaine, J. F., McGregor, B., Fouque, D. and Fauvel, J. P. (2013) 'Comparison of a Bayesian network with a logistic regression model to forecast IgA nephropathy', *Biomed Res Int*, 2013, pp.60.

Fach, A., Bünger, S., Zabrocki, R. and Schmucker, J. (2015) 'Comparison of Outcomes of Patients With ST-Segment Elevation Myocardial Infarction Treated by Primary Percutaneous Coronary Intervention Analyzed by Age Groups (85 Years); (Results from the Bremen STEMI Registry)', *The American Journal of Cardiology*, 116(12), pp. 1802–1809.

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*. 3rd edn. New York: Wiley.

Fisher, L. D. (1996) ‘Comments on Bayesian and Frequentist Analysis and Interpretation of Clinical Trials’, *Controlled clinical trials*, 17(5), pp. 423–434.

Food and Drug Administration (2010) *Guidance for Industry and FDA Staff Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. Center for Biologics Evaluation and Research, Rockville. Retrieved January 21, 2018, from http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationan Guidance/GuidanceDocuments/ucm071121.pdf

Foroushani, A. R., Mousavi, S. S., Mohammad, K. and Abedinia, N. (2016) ‘Using Bayesian networks model predicting pregnancy after psychiatric interventions in infertile couple’, *Journal of Biostatistics and Epidemiology,* 2(4), pp. 180–187.

Fosu, M., Jackson, O. and Twum, S. (2016) ‘Bayesian and Frequentist Comparison: An Application to Low Birth Weight Babies in Ghana’, *British Journal of Applied Science & Technology*, 16(2), pp. 1–15.

Fox, J. and Monette, G. (1992) ‘Generalized Collinearity Diagnostics’, *Journal of the American Statistical Association*, 87(417), pp. 178–183.

Fragoso, T. M., Bertoli, W. and Louzada, F. (2018) ‘Bayesian Model Averaging: A Systematic Review and Conceptual Classification’, *International Statistical Review*, 86(1), pp. 1–28.

Franzin, A., Sambo, F. and Camillo, B. Di (2017) ‘Systems biology bnstruct: an R package for Bayesian Network structure learning in the presence of missing data’, *Bioinformatics*, 33(8), pp. 1250–1252.

Franzini, L., Ardigò, D., Cavalot, F. and Vaccaro, O. (2013) ‘Women show worse control of type 2 diabetes and cardiovascular disease risk factors than men: Results from the MIND.IT Study Group of the Italian Society of Diabetology’, *Nutrition, Metabolism and Cardiovascular Diseases*, 23, pp. 235–241.

Friedman, N., Goldszmidt, M. and Wyner, A. (1999) ‘Data Analysis with Bayesian Networks: A Bootstrap Approach’, in *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence*, pp. 196–201.

Frigyik, B. A., Kapila, A. and Gupta, M. R. (2010) *Introduction to the Dirichlet Distribution*. Seattle, Washington.

Fuster-Parra, P., Tauler, P., Bennasar-Veny, M.and Aguiló, A. (2016) 'Bayesian network modeling: A case study of an epidemiologic system analysis of cardiovascular risk', *Computer Methods and Programs in Biomedicine*, 126, pp. 128–142.

Gan, Q., Smith, K. R., Hammond, S. K. and Hu, T. -w. (2007) 'Disease burden of adult lung cancer and ischaemic heart disease from passive tobacco smoking in China', *Tobacco Control*, 16(6), pp. 417–422.

Gao, F., Lam, C. S. P., Yeo, K. K. and Machin, D. (2016) 'Influence of Ethnicity, Age, and Time on Sex Disparities in Long-Term Cause-Specific Mortality After Acute Myocardial Infarction', *Journal of the American Heart Association*, 5(10), pp. e760.

Gearhart, C. and Kasturiratna, D. (2018) 'Implementation of Gibbs Sampling within Bayesian Inference and its Applications in Actuarial Science', *SIAM Journal*, pp. 219–231.

Geiger, D., Verma, T. and Pearl, J. (1990) 'Identifying independence in bayesian networks', *Networks*, 20(5), pp. 507–534.

Gelfand, A. E. and Smith, A. F. M. (1990) 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, 85(410), pp. 398–409.

Gelman, A. (1996) 'Inference and monitoring convergence', in *Markov chain Monte Carlo in practice*. Florida, United States of America: Chapman & Hall/CRC.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014) *Bayesian Data Analysis*. Third Edit. Boca Raton: Chapman and Hall/CRC.

Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, A. and Rubin, D. B. (1992) 'Inference from Iterative Simulation Using Multiple Sequences', *Statistical Science*, 7(4), pp. 457–511.

Geman, S. and Geman, D. (1984) 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), pp. 721–741.

Gentry, J., Gentleman, R. and Huber, W. (2018) *How To Plot A Graph Using Rgraphviz*.

George, S. L. (1985) 'Statistics in medical journals: A survey of current policies and proposals for editors’, *Medical and Pediatric Oncology*, 13(2), pp. 109–112.

Gerbarg, P. L. and Brown, R. P. (2016) ‘Pause menopause with Rhodiola rosea, a natural selective estrogen receptor modulator’, *Phytomedicine*, 23(7), pp. 763–769.

Geweke, J. (1992) ‘Evaluating the Accuracy of Sampling-Based Approaches to the Calculations of Posterior Moments’, *Bayesian Statistics*, 4, pp. 641–649.

Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. (1987) *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. San Diego, CA. : Academic Press.

Gonsalves, C. A., McGannon, K. R., Schinke, R. J. and Pegoraro, A. (2017) ‘Mass media narratives of women’s cardiovascular disease: a qualitative meta-synthesis’, *Health Psychology Review*, 11(2), pp. 164–178.

Goudie, R. J. B. and Mukherjee, S. (2016) ‘A Gibbs Sampler for Learning DAGs’, *J Mach Learn Res*, 17(30), pp. 1–39.

Graham, G. (2016) ‘Acute coronary syndromes in women: recent treatment trends and outcomes’, *Clinical Medicine Insights: Cardiology*, 10.

Greene, W. (2004) ‘The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects’, *The Econometrics Journal*, 7(1), pp. 98–119.

Greenland, S. and Mansournia, M. A. (2015) ‘Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions’, *Statistics in Medicine*, 34(23), pp. 3133–3143.

Griffin, A. T., Wiemken, T. L. and Arnold, F. W. (2013) ‘Risk factors for cardiovascular events in hospitalized patients with community-acquired pneumonia’, *International Journal of Infectious Diseases*, 17(12), pp. 1125–1129.

Gross, T. J., Araújo, R. B., Vale, F. A. C., Bessani, M. and Maciel, C. D. (2018) ‘Dependence between cognitive impairment and metabolic syndrome applied to a Brazilian elderly dataset’, *Artificial Intelligence in Medicine*, 90, pp. 53–60.

Gupta, T., Kolte, D., Khera, S. and Harikrishnan, P. (2016) ‘Smoker’s paradox in patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention’, *Journal of the American Heart Association*, 5(4), pp. 1–11.

Gutierrez, J., Alloubani, A., Mari, M. and Alzaatreh, M. (2018) ‘Cardiovascular Disease Risk Factors: Hypertension, Diabetes Mellitus and Obesity among Tabuk Citizens in Saudi Arabia.’, *The open cardiovascular medicine journal*. Bentham Science Publishers, 12, pp. 41–49.

Guure, C. B., Ibrahim, N. A., Dwomoh, D. and Bosomprah, S. (2015) ‘Bayesian statistical inference of the loglogistic model with interval-censored lifetime data’, *Journal of Statistical Computation and Simulation*, 85(8), pp. 1567–1583.

Hajian-Tilaki, K. (2013) ‘Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation’, *Caspian journal of internal medicine,* 4(2), pp. 627–635.

Hand, D. J. (2009) ‘Measuring classifier performance: a coherent alternative to the area under the ROC curve’, *Mach Learn*, 77, pp. 103–123.

Hanley, A. J. and McNeil, J. B. (1982) ‘The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve’, *Radiology,* 143, pp. 29–36.

Harrell, F. E. (2015) *Binary Logistic Regression*, in Regression modeling strategies, Springer, Cham, pp. 219–274.

Hassan, I. I., Muhammad, I. N., Ismail, S. B. and Kadir, A. A. (2017) ‘Effect of Health Education Program on Cardiovascular Risk Factors among Menopausal Women in Malaysia’, *International Medical Journal*, 24(1), pp. 51–55.

Hastings, W. K. (1970) ‘Monte Carlo sampling methods using Markov chains and their applications’, *Biometrika,* 57(1), pp. 97–109.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995) ‘Learning Bayesian Networks: The Combination of Knowledge and Statistical Data’, *Machine Learning,* 20(3), pp. 197–243.

Heidelberger, P. and Welch, P. D. (1983) ‘Simulation Run Length Control in the Presence of an Initial Transient’, *Operations Research*, 31(6), pp. 1109–1144.

Hilbe, J. . (2016) *Practical guide to logistic regression*. Chapman and Hall/CRC.

Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O’Callaghan, C. A., Lasserson, D. S. and Hobbs, F. D. R. (2016) ‘Global Prevalence of Chronic Kidney Disease – A Systematic Review and Meta-Analysis’, *PLOS ONE,* 11(7), pp. e0158765.

Højsgaard, S., Edwards, D. and Lauritzen, S. (2012) *Graphical Models with R, Use R!* New York: Springer.

Hosmer, D. W., Hosmer, T., Le Cessie, S. and Lemeshow, S. (1997) ‘A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model', *Statistics in Medicine*, 16(9), pp. 965–980.

Hosmer, D. W. and Lemesbow, S. (1980) 'Goodness of fit tests for the multiple logistic regression model', *Communications in Statistics - Theory and Methods*, 9(10), pp. 1043–1069.

Hosmer, D. W. and Lemeshow, S. (2000) *Applied Logistic Regression*. New York: John Wiley & Sons, Inc.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013) *Applied Logistic Regression., Third Edition*. New Jersey: John Wiley & Sons, Inc.

Howson, C. and Urbach, P. (1991) 'Bayesian reasoning in science', *Nature*, 350(6317), pp. 371–374.

Hubert, H. B. (1983) 'Obesity as an Independent Risk Factor for Cardiovascular Disease : A 26-year Follow-up of Participants in the Framingham Heart Study', *American Heart Association Journal*, 67(5), pp. 968–976.

Idris, N., Aznal, S., Chin, S. P. and Azman Wan Ahmad (2011) 'Acute coronary syndrome in women of reproductive age', *International Journal of Women's Health*, 3, pp. 375–380.

Imes, C. C. and Lewis, F. M. (2014) 'Family history of cardiovascular disease, perceived cardiovascular disease risk, and health-related behavior: a review of the literature.', *The Journal of cardiovascular nursing*, 29(2), pp. 108–29.

Ioannidis, J. P. (2005) 'Why most published research findings are false', *PLoS medicine*, 2(8), pp. e124.

Irony, T. and Singpurwalla, N. (1996) 'Noninformative priors do not exist: A discussion with Jose M. Bernardo', *Journal of Statistical Inference and Planning*, 65(1), pp. 159–189.

Jackman, S. (2009) *Bayesian Analysis for the Social Sciences*. Chichester, UK: John Wiley & Sons, Ltd.

Jager, J., Putnick, D. L. and Bornstein, M. H. (2017) 'More Than Just Convenient: The Scientific Merits Of Homogeneous Convenience Samples', *Monographs of the Society for Research in Child Development*, 82(2), pp. 13–30.

Jain, R. K. and Laiteerapong, N. (2015) 'Strategies for Improving Cardiovascular Health in Women With Diabetes Mellitus: a Review of the Evidence', *Current Diabetes Reports*, 15(11).

Jiang, F., Wang, J., Zhang, R. and Chen, M. (2015) 'Effects of active and passive smoking on the development of cardiovascular disease as assessed by a carotid intima-media thickness examination in patients with type 2 diabetes mellitus', *Clinical and Experimental Pharmacology and Physiology*, 42(5), pp. 444–450.

Jiménez-Valverde, A. (2012) 'Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling', *Global Ecology and Biogeograph*, 21(4), pp. 498–507.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) 'Chapter 21: Beta distributions', in *Continuous Univariate Distributions, Vol. 2*.

Jones, M., Hockey, R., Mishra, G. D. and Dobson, A. (2014) 'Visualising and modelling changes in categorical variables in longitudinal studies', *BMC Medical Research Methodology*, 14(1), pp. 1–8.

Jousilahti, P., Laatikainen, T., Peltonen, M. and Borodulin, K. (2016) 'Primary prevention and risk factor reduction in coronary heart disease mortality among working aged men and women in eastern Finland over 40 years: Population based observational study', *BMJ (Online)*, 352, pp. 1–8.

Kalantar-Zadeh, K., Block, G., Horwich, T. and Fonarow, G. C. (2004) 'Reverse epidemiology of conventional cardiovascular risk factors in patients with chronic heart failure', *Journal of the American College of Cardiology*, 43(8), pp. 1439–1444.

Kandpal, V., Sachdeva, M. P. and Saraswathy, K. N. (2016) 'An assessment study of CVD related risk factors in a tribal population of India', *BMC public health*, 16(1), 434.

Karreth, J. (2015) 'Using JAGS via R', pp. 1–24.

Kazembe, L. N., Chirwa, T. F., Simbeye, J. S. and Namangale, J. J. (2008) 'Applications of Bayesian approach in modelling risk of malaria-related hospital mortality', *BMC Medical Research Methodology*, 8, pp. 1–14.

Kempaller, V. J., Kiran, K. G., Patil, S., Kiran, U., Shetty, S. and Dinesh, S. (2015) 'A Cross Sectional Study on Pattern of Tobacco Consumption and Prevalence of Passive Smoking and Its Health Hazardawareness Among Elderly Patients.-A Coastal Karnataka Study', *Journal Of Advance Researches In Medical Sciences*, 7(1), pp. 5–10.

Keto, J., Ventola, H., Jokelainen, J. and Linden, K. (2016) 'Cardiovascular disease risk factors in relation to smoking behaviour and history: a population-based cohort study', *Open Heart*, 3(2), pp. e000358.

Killip, T. and Kimball, J. T. (1967) 'Treatment of myocardial infarction in a coronary care unit. A two year experience with 250 patients.', *The American journal of cardiology*, 20(4), pp. 457–64.

King, E. N. and Ryan, T. P. (2002) 'A Preliminary Investigation of Maximum Likelihood Logistic Regression versus Exact Logistic Regression', *The American Statistician*, 56(3), pp. 163–170.

Kleinbaum, D. G. and Klein, M. (2010) *Logistic Regression: A Self-learning Text, Third Edition*. Springer Science & Business Media.

Koehler, E., Brown, E. and Haneuse, S. J.-P. A. (2009) 'On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses.', *The American statistician*, 63(2), pp. 155–162.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2015) 'Machine learning applications in cancer prognosis and prediction', *Computational and Structural Biotechnology Journal*, 13, pp. 8–17.

Kramer, A. A. and Zimmerman, J. E. (2007) 'Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test', *Critical Care Medicine*, 35(9), pp. 2052–2056.

Kuipers, J. and Moffa, G. (2017) 'Partition MCMC for Inference on Acyclic Digraphs Partition MCMC for Inference on Acyclic Digraphs *', *Journal of the American Statistical Association*, 112(517), pp. 282–299.

Kuji, S., Kosuge, M., Kimura, K. and Nakao, K. (2017) 'The identi-Impact of Acute Kidney Injury on In-Hospital Outcomes of Patients With Acute Myocardial Infarction-Results From the Japanese Registry of Acute Myocardial Infarction Diagnosed by Universal Definition (J-MINUET) Substudy', *Circulation Journal*, 81, pp. 733–739.

Kumar, P. and Herath, H. S. B. (2015) 'Bayesian Analysis in Industrial Applications using Markov Chain Monte Carlo Simulations', *International Journal of Computational and Theoretical Statistics*, 2(1), pp. 49–66.

Kulesa, A., Krzywinski, M., Blainey, P. and Altman, N. (2015) 'Sampling distributions and the bootstrap', *Nature Methods*, 12(6), pp. 477–478.

Lacave, C., Molina, A. I. and Cruz-Lemus, J. A. (2018) 'Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks', *Behaviour and Information Technology*, pp. 1–15.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005)

‘How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS’, *Statistics in Medicine*, 24(15), pp. 2401–2428.

Larrañaga, P., Kuijpers, C. M., Poza, M. and Murga, R. H. (1997) ‘Decomposing Bayesian networks: triangulation of the moral graph with genetic algorithms’, *Statistics and Computing*, 7(1), pp. 19–34.

Lau, C. L. and Smith, C. S. (2016) ‘Bayesian networks in infectious disease eco-epidemiology’, *Reviews on Environmental Health*, 31(1).

Lawless, M. H., Harrison, K. A., Grandits, G. A. and Eberly, L. E. (2015) ‘Perceived stress and smoking-related behaviors and symptomatology in male and female smokers’, *Addictive Behaviors*, 51, pp. 80–83.

Leander, K., Hallqvist, J., Reuterwall, C. and de Faire, U. (2001) ‘Family history of coronary heart disease, a strong risk factor for myocardial infarction interacting with other cardiovascular risk factors: results from the Stockholm Heart Epidemiology Program (SHEEP).’, *Epidemiology (Cambridge, Mass.)*, 12(2), pp. 215–21.

Lee, P. N., Thornton, A. J., Forey, B. A. and Hamling, J. S. (2017) ‘Environmental Tobacco Smoke Exposure and Risk of Stroke in Never Smokers: An Updated Review with Meta-Analysis’, *Journal of Stroke and Cerebrovascular Diseases*, 26(1), pp. 204–216.

Legato, M. J., Johnson, P. A. and Manson, J. E. (2016) ‘Consideration of Sex Differences in Medicine to Improve Health Care and Patient Outcomes’, *JAMA*, 316(18), p. 1865.

LePage, R. and Billard, L. (1992) Exploring the Limits of Bootstrap. Vol. 270. John Wiley & Sons.

Leucht, S., Leucht, C., Huhn, M. and Chaimani, A. (2017) ‘Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia: Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors’, *American Journal of Psychiatry*, 174(10), pp. 927–942.

Lewis, F., Butler, A. and Gilbert, L. (2011) ‘A unified approach to model selection using the likelihood ratio test’, *Methods in Ecology and Evolution*, 2(2), pp. 155–162.

Ley, S. H., Li, Y., Tobias, D. K. and Manson, J. E. (2017) ‘Duration of Reproductive Life Span, Age at Menarche, and Age at Menopause Are Associated With Risk of Cardiovascular Disease in Women.', *Journal of the American Heart Association*, 6(11), pp. e006713.

Li, W., Zhang, P., Han, Q., Leung, H. and Song, W. (2018) 'A Novel QoS Prediction Approach for Cloud Service Based on Bayesian Networks Model', *IEEE International Conference on Mobile Services (MS)*, 6, pp. 111–118.

Li, X., Yu, J., Li, M. and Zhao, D. (2017) 'Discover high-risk factor combinations using Bayesian network from national screening data in China', *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1047–1051.

Liao, Y., Xu, B., Wang, J. and Liu, X. (2017) 'A new method for assessing the risk of infectious disease outbreak', *Scientific Reports*, 7(December 2016), pp. 1–12.

Lim, H. K., Ghazali, S. M., Kee, C. C., and Lim, K. K. (2013) 'Epidemiology of smoking among Malaysian adult males: prevalence and associated factors', *BMC Public Health*, 13(1), pp. 8.

Liu, J., Fan, M., Zhao, J., Zhao, B. and Zhang, C. (2016) 'Efficacy and safety of antithrombotic regimens after coronary intervention in patients on oral anticoagulation: Traditional and Bayesian meta-analysis of clinical trials', *International Journal of Cardiology*, 205, pp. 89–96.

Liu, Y. (2018) *Statistical Science for the Life and Behavioural Sciences*. Master Thesis, Universiteit Leiden.

Lloyd-Jones, D. M., Nam, B. H., D'Agostino Sr, R. B. and Levy, D. (2004) 'Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring', *JAMA*, 291(18), pp. 2204–2211.

Lu, H. T., Nordin, R., Wan Ahmad, W. A. and Lee, C. Y. (2014) 'Sex Differences in Acute Coronary Syndrome in a Multiethnic Asian Population', *Global Heart*, 9(4), pp. 381–390.

Ma, T.-Y. and Klein, S. (2017) 'Bayesian networks for constrained location choice modeling using structural restrictions and model averaging HANC, Healthy Aging and Networks in Cities: The combined role of built environments and socio-spatial networks View project The Impact of ATIS on', *European Journal of Transport and Infrastructure Research*, 18(1), pp. 91–111.

Madadizadeh, F., Asar, M. E. and Hosseini, M. (2016) 'Bayesian Statistical Methods and Their Valuable Applications in the Pharmaceutical Sciences', *Iranian* *journal of public health*, 45(8), pp. 1098–1099.

Mahmood, K. (2016, August 2) Heart disease No . 1 killer for women in Malaysia. *The Sun Daily*. pp. 2–3.

Margaritis, D. (2003) *Learning Bayesian Network Model Structure from Data*. PhD Thesis, Carnegie Mellon University.

McCarthy, M. A. (2007) *Bayesian methods for ecology*. Cambridge University Press.

McClelland, R. L., Jorgensen, N. W., Budoff, M. and Blaha, M. J. (2015) '10-Year Coronary Heart Disease Risk Prediction Using Coronary Artery Calcium and Traditional Risk Factors', *Journal of the American College of Cardiology*, 66(15), pp. 1643–1653.

McCullagh, P. and Nelder, J. A. (1989) 'Generalized Linear Models.', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154(2), p. 361.

McHugh, M. L. (2013) 'The Chi-square test of independence', *Biochemia Medica*, 23(2), pp. 143–9.

Meer, M. G. van der (2015) *Gender differences in diagnosis and prognosis of coronary artery disease*. PhD Thesis, Utrecht University.

Mehta, L. S., Watson, K. E., Barac, A. and Beckie, T. M. (2018) 'Cardiovascular Disease and Breast Cancer: Where These Entities Intersect: A Scientific Statement From the American Heart Association.', *Circulation*, 137(8), pp. e30–e66.

Mello, B. H. G. D., Oliveira, G. B. F., Ramos, R. F. and Lopes, B. B. C. (2014) 'Validation of the Killip-Kimball Classification and Late Mortality after Acute Myocardial Infarction', *Arquivos brasileiros de cardiologia*, 103(2), pp. 107–117.

Memon, A. G. and Khan, M. (2017) 'Echocardiographic Correlation of Clips Classification to Asses Left Ventricular Function in Patients with Acute Myocardial Infarction', *Journal of American Science*, 13(7), pp. 106–112.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H (1953) 'Equation of State Calculations by Fast Computing Machines', *The Journal of Chemical Physics*, 21(6), pp. 1087–1092.

Mickey, R. M. and Greenland, S. (1989) 'The impact of confounder selection criteria on effect estimation.', *American journal of epidemiology*, 129(1), pp. 125–137.

Midi, H., Sarkar, S. K. and Rana, S. (2010) 'Collinearity diagnostics of binary logistic regression model', *Journal of Interdisciplinary Mathematics*, 13(3), pp. 253–

267.

Miller, P. (2016) ‘Women and Cardiovascular Disease: What Can Health Care Providers Do to Reduce the Risks?’, *North Carolina Medical Journal*, 77(6), pp. 406–409.

Mills, K. T., Xu, Y., Zhang, W. and Bundy, J. D. (2015) ‘A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010’, *Kidney International*, 88(5), pp. 950–957.

Ministry of Health Malaysia (2016) ‘Health Facts 2016’, *Planning Division Health Informatics Centre*, pp. 1–19. Retrieved February 25, 2017, from http://www.moh.gov.my/images/gallery/publications/KKMHEALTH FACTS 2016.pdf

Mohammad, N. B., Rahman, N. A. A. and Haque, M. (2018) ‘Knowledge, Attitude, and Practice Regarding the Risk of Cardiovascular Diseases in Patients Attending Outpatient Clinic in Kuantan, Malaysia.’, *Journal of pharmacy & bioallied sciences*, 10(1), pp. 7–14.

Mohammed, M. J., Rakhimov, I. S., Shitan, M. and Ibrahim, R. W. (2016) ‘A new mathematical evaluation of smoking problem based of algebraic statistical method’, *Saudi Journal of Biological Sciences*, 23(1), pp. S11–S15.

Mohd Ghazali, S., Seman, Z., Cheong, K. C. and Hock, L. K. (2015) ‘Sociodemographic factors associated with multiple cardiovascular risk factors among Malaysian adults’, *BMC Public Health*, 15(1), pp. 1–8.

Mohd Noor, N., Sharifah Amirah, S., Lili Husniati, Y. and Mohd Noor, N. (2013) ‘Cardiovascular risk: associated factors, assessment and agreement between WHO/ISH risk prediction chart and Framingham Scoring System among Primary Care Patients in Kelantan, Malaysia’, *International Journal of Collaborative Research on Internal Medicine & Public Health*, 5(12), pp. 652–663.

Monroe, W. (2017) ‘Logistic Regression’, *Recall*, 1, pp. 1.

Montesinos-López, O. A., Montesinos-López, A., Crossa, J. and Eskridge, K. (2015) ‘Genomic-Enabled Prediction of Ordinal Data with Bayesian Logistic Ordinal Regression’, *G3 Genes|Genomes|Genetics*, 5(10), pp. 2113–2126.

Morey, R. D., Hoekstra, · Rink, Rouder J. N. and Lee, M. D. (2016) ‘The fallacy of placing confidence in confidence intervals’, *Psychon Bull Rev*, 23, pp. 103–123.

Moriyama, N., Noguchi, T., Nakanishi, M. and Ogawa, H. (2017) ‘Early development of acute kidney injury is an independent predictor of in-hospital mortality in patients with acute myocardial infarction consecutive patients who were admitted to the National Cerebral and Cardiovascular Center in Japan within 48 h after the onset of symptoms were’, *Journal of Cardiology*, 69, pp. 79–83.

Morris, W. K., Vesk, P. A., McCarthy, M. A. and Baker, P. J. (2015) ‘The neglected tool in the Bayesian ecologist’s shed: a case study testing informative priors’ effect on model accuracy’, *Ecology and Evolution*, 5(1), pp. 102–108.

Moyé, L. (2016) ‘Statistical methods for cardiovascular researchers’, *Circulation Research*, 118(3), pp. 439–453.

Mozaffarian, D., Benjamin, E. J., Go, A. S. and Turner, M. B. (2015) ‘Executive Summary: Heart Disease and Stroke Statistics—2015 Update’, *Circulation*, 131(4), pp. 434–441.

Mulder, J. and Wagenmakers, E. J. (2016) ‘Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments’, *Journal of Mathematical Psychology*, 72, pp. 1–5.

Mullachery, V., Khera, A. and Husain, A. (2018) ‘Bayesian Neural Networks’, *arXiv*, pp. 1–16.

Müller, P., Maier, J., Dietrich, J. W. and Deneke, T. (2015) ‘Association between left atrial low-voltage area, serum apoptosis, and fibrosis biomarkers and incidence of silent cerebral events after catheter ablation of atrial fibrillation’, *Journal of Interventional Cardiac Electrophysiology*, 44(1), pp. 55–62.

Murray, C. J. L. and Lopez, A. D. (2013) ‘Measuring the Global Burden of Disease’, *New England Journal of Medicine*, 369(5), pp. 448–457.

Murty, O. P., Seng, L. K., Nuraeiniza, I., Chee, A. S. and Syahir, B. M. (2007) ‘Females deaths due to cardiac causes in Malaysia : 10 years autopsy review’, *Malaysian J Forensic Pathol Sci*, 2(1), pp. 9–19.

Musen, M. A., Middleton, B. and Greenes, R. A. (2014) ‘Clinical Decision-Support Systems’, *Biomedical informatics*, pp. 643–647.

Nagarajan, R., Scutari, M. and Lèbre, S. (2013) *Bayesian Networks in R*.

Neapolitan, R. E. (2004) *Learning Bayesian Networks*. Vol.38. Upper Saddle River, NJ: Pearson Prentice Hall.

Neapolitan, R. and Jiang, X. (2018) *The Bayesian Network Story - The Oxford Handbook of Probability and Philosophy*. Oxford University Press, pp. 1–22.

Nemes, S., Jonasson, J. M., Genell, A. and Steineck, G. (2009) 'Bias in odds ratios by logistic regression modelling and sample size', *BMC Medical Research Methodology*, 9(1), p. 56.

Nguyen, H. T., Bryois, J., Kim, A. and Stahl, E. A. (2017) 'Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders', *Genome Medicine*, 9(1), pp. 1–22.

Nilsson, A. G., Sundh, D., Johansson, L. and Lorentzon, M. (2017) 'Type 2 Diabetes Mellitus Is Associated With Better Bone Microarchitecture But Lower Bone Material Strength and Poorer Physical Function in Elderly Women: A Population-Based Study', *Journal of Bone and Mineral Research*, 32(5), pp. 1062–1071.

Ntzoufras, I. (2009) *Bayesian Modeling Using WinBUGS, Wiley Series in Computational Statistics*. New Jersey: John Wiley & Sons, Inc.

Nuzzo, R. (2014) 'Scientific method: statistical errors', *Nature*, 506, pp. 150–152.

Oates, C. J., Smith, J. Q., Mukherjee, S. and Cussens, J. (2016) 'Exact Estimation of Multiple Directed Acyclic Graphs', *Statistics and Computing*, 26(4), pp. 797–811.

Ormerod, J. T., Stewart, M., Yu, W. and Romanes, S. E. (2017) 'Bayesian hypothesis tests with diffuse priors: Can we have our cake and eat it too?', *arXiv*, pp. 1710.

Otgontuya, D., Oum, S., Buckley, B. S. and Bonita, R. (2013) 'Assessment of total cardiovascular risk using WHO/ISH risk prediction charts in three low and middle income countries in Asia', *BMC Public Health*, 13(1), p. 539.

Park, S. H., Goo, J. M. and Jo, C.-H. (2004) 'Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists', *Korean Journal of Radiology*, 5(1), pp. 11–18.

Pascale, A., Nicoli, M. and Spagnolini, U. (2014) 'Cooperative Bayesian Estimation of Vehicular Traffic in Large-Scale Networks', *IEEE Transactions on Intelligent Transportation Systems*, 15(5), pp. 2074–2088.

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P. and Stamatakis, A. (2010) 'How many bootstrap replicates are necessary?', *Journal of Computational Biology*, 17(3), pp. 337–354.

Pearl, J. (1993) 'Bayesian Analysis in Expert Systems : Comment: Graphical Models, Causality and Intervention', *Statistical Science*, 8(3), pp. 266–269.

Pearl, J. (1996) 'Causation, action, and counterfactuals', in *6th conference on Theoretical aspects of rationality and knowledge*. The Netherlands: Morgan Kaufmann Publishers Inc., pp. 51–73.

Pearl, J. (2014) *Probabilistic reasoning in intelligent systems : networks of plausible inference*. San Francisco, California: Morgan Kaufmann Publishers.

Pearl, J., Geiger, D. and Verma, T. (1989) 'Conditional independence and its representations', *Kybernetika*, 25(7), pp. 33–44.

Peizer, D. B. and Pratt, J. W. (1968) 'A Normal Approximation for Binomial, F, Beta, and other Common, Related Tail Probabilities, II', *Journal of the American Statistical Association*, 63(324), pp. 1416–1456.

Peng, C.-Y. J., Lee, K. L. and Ingersoll, G. M. (2002) 'An Introduction to Logistic Regression Analysis and Reporting', *The Journal of Educational Research*, 96(1), pp. 3–14.

Plummer, M. (2013) 'JAGS Version 3.4.0 user manual', pp. 1–41.

Prabhakaran, D., Roy, A., Praveen, P. A. and Krishnan, A. (2017) '20-Year Trend of CVD Risk Factors', *Global Heart*, 12(3), pp. 209–217.

Preis, S. R., Pencina, M. J., Mann, D. M. and Fox, C. S. (2013) 'Early-Adulthood Cardiovascular Disease Risk Factor Profiles Among Individuals With and Without Diabetes in the Framingham Heart Study', *Diabetes Care*, 36(6), pp. 1590–1596.

Raftery, A. . and Lewis, S. . (1992) 'One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo.', *Statistical Science*, 7, pp. 493–497.

Raftery, A. E. and Lewis, S. (1991) *How many iterations in the Gibbs sampler?*. Washington: Univ Seattle Dept Of Statistics.

Raftery, A. E. and Lewis, S. M. (1996) 'Implementing MCMC.', *Markov Chain Monte Carlo in Practice* , pp. 115–130.

Rao, C. R. (1973) *Linear statistical inference and its applications*. New York: Wiley.

Redelmeier, D. A., Bloch, D. A. and Hickam, D. H. (1991) 'Assessing predictive accuracy: how to compare Brier scores.', *Journal of clinical epidemiology*, 44(11), pp. 1141–6.

Richardson, T. (1994) *Properties of cyclic graphical models*. Master Thesis, Carnegie Mellon Univ.

Robinson, R. W. (1977) *Counting unlabeled acyclic digraph*, in Combinatorial mathematics. Springer, Berlin, Heidelberg, pp. 28–43.

Rosenthal, E. L., Moore, L. S., Tipirneni, K. and Doerr, J. (2017) ‘Sensitivity and Specificity of Cetuximab-IRDye800CW to Identify Regional Metastatic Disease in Head and Neck Cancer’, *Clinical Canser Research*, pp. 2968.

Rufibach, K. (2010) ‘Use of Brier score to assess binary predictions’, *Journal of Clinical Epidemiology*, pp. 938–939.

Russell, S. J. and Norvig, P. (2016) *Artificial Intelligence: A Modern Approach*. Third Edit. Essex: Pearson Education Limited.

Sabroe, J. E., Antonsen, L., Hougaard, M. and Jensen, L. O. (2014) ‘Impact of renal insufficiency on mortality in patients with ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention’, *BMC cardiovascular disorders*, 14(1), pp. 15.

Sambo, F., Camillo, B. Di, Franzin, A. and Cobelli, C. (2015) ‘A Bayesian Network analysis of the probabilistic relations between risk factors in the predisposition to type 2 diabetes’, *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 2119–2122.

Sarnak, M. J., Levey, A. S., Schoolwerth, A. C. and Wilson, P. W. (2003) ‘Kidney Disease as a Risk Factor for Development of Cardiovascular Disease: A Statement From the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention’, *Circulation*, 108(17), pp. 2154–2169.

Scheideler, J. K., Taber, J. M., Ferrer, R. A. and Klein, W. M. P. (2017) ‘Heart disease versus cancer: understanding perceptions of population prevalence and personal risk’, *Journal of Behavioral Medicine*, 40(5), pp. 839–845.

Schruben, L., Singh, H. and Tierney, L. (1983) ‘Optimal Tests for Initialization Bias in Simulation Output’, *Operations Research*, 31(6), pp. 1167–1178.

Schruben, L. W. (1982) ‘Detecting Initialization Bias in Simulation Output’, *Operations Research*, 30(3), pp. 569–590.

Scott Long, J. (1997) ‘Regression models for categorical and limited dependent variables.’, *Advanced Quantitative Techniques in the Social Sciences* , 7.

Scutari, M. (2010) ‘Learning Bayesian Networks with the bnlearn R Package’, *Journal of Statistical Software*, 35(3), pp. 1–22.

Scutari, M. (2013) ‘On the Prior and Posterior Distributions Used in Graphical Modelling’, *Bayesian Analysis*, 8(3), pp. 505–532.

Scutari, M. and Denis, J. B. (2015) *Bayesian Networks With Examples in R*. Boca Raton: CRC Press Taylor & Francis Group.

Scutari, M., Graafland, C. E. and Gutiérrez, J. M. (2018) 'Who Learns Better Bayesian Network Structures: Constraint-Based, Score-based or Hybrid Algorithms?', in *Proceedings of Machine Learning Research 72, 2018 International Conference on Probabilistic Graphical Models*. Prague, Czech Republic, pp. 1–12.

Scutari, M., Howell, P., Balding, D. J. and Mackay, I. (2014) 'Multiple Quantitative Trait Analysis Using Bayesian Networks', *Genetics*, 198(1), pp. 129–137.

Scutari, M. and Ness, R. (2018) *Package " bnlearn "*.

Shah, S. A., Jing, L., Khalid, M. S. and Wan Ngah, W. Z. (2015) 'Prevalence and risk factors of premature coronary artery disease: A comparative cross-sectional study between two time frames in Malaysia', *Malaysian Journal of Public Health Medicine*, 15(3), pp. 157–166.

Shomad, M. B. A., Rahman, N. A. A., Rahman, N. I. A. and Haque, M. (2016) 'The prevalence of cardiovascular disease risk factors among students of international islamic university Malaysia, Kuantan Campus', *Journal of Applied Pharmaceutical Science*, 6(3), pp. 51–57.

Silvestro, D., Zizka, G. and Schulte, K. (2014) 'Disentangling the Effects of Key Innovations on the Diversification of Bromelioideae', *Evolution*, 68(1), pp. 163–175.

Simon, H. A. (1954) 'Spurious correlation: A causal interpretation', *Journal of the American Statistical Association*, 49, pp. 467–479.

Singh, K., Bartolucci, A. and Bae, S. (2010) 'The Bayesian multiple logistic random effects model for analysis of clinical trial data', *Journal of Data Science*, 8, pp. 495–504.

Singh, K. and Xie, M. (2010) 'Bootstrap method', *International Encyclopedia of Education*. Elsevier Ltd.

Smith, A. F. M. and Roberts, G. O. (1993) 'Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(1), pp. 3–23.

Smith, B. J. (2005) '(boa) Version 1 . 1 User ' S Manual'.

Smith, B. J. (2007) 'boa : An R Package for MCMC Output Convergence', *Journal of Statistical Software*, 21(11), pp. 1–37.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) ‘Bayesian Measures of Model Complexity anf Fit’, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 64(4), pp. 583–639.

Spirtes, P. (1994) *Conditional Independence in Directed Cyclic Graphical Models for Feedback*. Carnegie Mellon Department of Philosophy.

Spirtes, P., Glymour, C. N. and Scheines, R. (1989) *Causality from probability*. Vol. 112. Carnegie-Mellon University.

Spirtes, P., Richardson, T., Meek, C., Scheines, R. and Glymour, C. (1996) *Using D-separation to Calculate Zero Partial Correlations in Linear Models with Correlated Errors Philosophy Methodology Logic*. Carnegie-Mellon University.

Srinivasan, C. R., Saravanababu, S., Swaminathan, N. and Venkatesan, S. (2018) ‘Correlation of Killip Class and Stroke Volume In Acute Myocardial Infarction’, *International Journal of Scientific Research*, 7(7).

Steck, H. and Jaakkola, T. S. (2004) ‘Bias-Corrected Bootstrap and Model Uncertainty’, *Advances in Neural Information Processing Systems*, pp. 521–528.

Stephenson, T. A. (2000) *An introduction to Bayesian network theory and usage*. 9–1104.

Stoltzfus, J. C. (2011) ‘Logistic Regression: A Brief Primer’, *Academic Emergency Medicine*, 18(10), pp. 1099–1104.

Suárez, E., Pérez, C. M., Martínez, M. N. and Rivera, R. (2017) *Applications of Regression Models in Epidemiology*. John Wiley & Sons.

Symons, R., Masci, P. G., Francone, M. and Bogaert, J. (2016) ‘Impact of active smoking on myocardial infarction severity in reperfused ST-segment elevation myocardial infarction patients: the smoker’s paradox revisited’, *European Heart Journal*, 37(36), pp. 2756–2764.

Szumilas, M. (2010) ‘Explaining Odds Ratios’, *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), pp. 227–229.

Taguchi, E., Konami, Y., Inoue, M. and Sakamoto, T. (2017) ‘Impact of Killip classification on acute myocardial infarction: data from the SAIKUMA registry’, *Heart and Vessels*, 32(12), pp. 1439–1447.

Tan, Y. C., Sinclair, H., Ghoorah, K. and Kunadian, V. (2016) ‘Gender differences in outcomes in patients with acute coronary syndrome in the current era: A review', *European heart journal. Acute cardiovascular care*, 5(7), pp. 51–60.

Tang, Z. H., Liu, J., Zeng, F. and Zhou, L. (2013) 'Comparison of Prediction Model for Cardiovascular Autonomic Dysfunction Using Artificial Neural Network and Logistic Regression Analysis', *PLoS ONE*, 8(8).

Thangiah, N., Abdul Majid, H., Chinna, K. and Su, T. T. (2017) 'Clustering and Stability of Biological CVD Risk Factors Among Adolescents in Malaysia', *International Journal of Cardiology*, 249, p. S12.

Thomson, J. R., Mac Nally, R., Fleishman, E. and Horrocks, G. (2007) 'Predicting Bird Species Distributions in Reconstructed Landscapes', *Conservation Biology*, 21(3), pp. 752–766.

Torman, V. B. L. and Camey, S. A. (2015) 'Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes', *Emerging Themes in Epidemiology*, 12(1).

U.S. Department of Health and Human Services (2006) *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General, Publications and Reports of the Surgeon General*. Atlanta, GA.

Ueda, T., Kawakami, R., Sugawara, Y. and Saito, Y. (2014) 'Worsening of renal function during 1 year after hospital discharge is a strong and independent predictor of all-cause mortality in acute decompensated heart failure', *Journal of the American Heart Association*, 3(6), pp. 1–8.

Umamahesh, K., Vigneswari, A., Thejaswi, G. S. and Viswanathan, V. (2014) 'Incidence of cardiovascular diseases and associated risk factors among subjects with type 2 diabetes e An 11-year follow up study', *Indian Heart Journal*, 66, pp. 5–10.

Van Ravenzwaaij, D., Cassey, P. and Brown, S. D. (2018) 'A simple introduction to Markov Chain Monte Carlo sampling', *Psychonomic Bulletin & Review*, 25(1), pp. 143–154.

Van Wagenberg, C. P. A., Aramyan, L. H., de Lauwere, C. C. and Willems, D. J. M. (2018) 'A Bayesian Network as a tool to measure Supply Chain Resilience', in *International Conference of Agricultural Economists (ICAE 2018)*. Vancouver.

Venkatason, P., Zubairi, Y. Z., Ahmad, W. A. W. and Zuhdi, A. S. (2016a) 'Trends in evidence-based treatment and mortality for ST elevation myocardial infarction in Malaysia from 2006 to 2013: Time for real change', *Annals of Saudi*

Medicine, 36(3), pp. 184–189.

Venkatason, P., Salleh, N. M., Zubairi, and Zuhdi, A. S. M. (2016b) ‘The bizzare phenomenon of smokers’ paradox in the immediate outcome post acute myocardial infarction: an insight into the MalaysianNational Cardiovascular Database-Acute Coronary Syndrome (NCVD-ACS) registry year 2006–2013’, *SpringerPlus*, 5(1).

Vitolo, C., Scutari, M., Ghalaieny, M., Tucker, A. and Russell, A. (2018) ‘Modelling air pollution, climate and health data using Bayesian Networks: a case study of the English regions’, *Earth and Space Science*, 5(4), pp. 76-88.

Walley, R. J., Smith, C. L., Gale, J. D. and Woodward, P. (2015) ‘Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study’, *Pharmaceutical Statistics*, 14(3), pp. 205–215.

Wan Ahmad, W. A. and Sim, K.-H. (2006) *National Cardiovascular Disease Database (NCVD) Report of the Acute Coronary Syndrome (ACS) Registry 2006, National Cardiovascular Disease Database*. Kuala Lumpur.

Wan Ahmad, W. A. and Sim, K.-H. (2015) *Annual Report of the NCVD-ACS Registry, 2011-2013 Kuala Lumpur, Malaysia., National Cardiovascular Disease Database*. Kuala Lumpur.

Wang, T., Baron, K., Zhong and Elmquist, W. (2014) ‘Bayesian Approach to Estimate AUC, Partition Coefficient and Drug Targeting Index for Studies with Serial Sacrifice Design’, *Pharm Res.*, 31(3), pp. 649–659.

Wang, X., Niu, Y. and Li, X. (2017) ‘Bayesian ROC curve estimation under b inormality using an ordinal category likelihood’, *Communications in Statistics - Theory and Methods*, pp. 1–13.

Weiner, D. E., Tighiouart, H., Amin, M. G. and Sarnak, M. J. (2004) ‘Chronic Kidney Disease as a Risk Factor for Cardiovascular Disease and All-Cause Mortality: A Pooled Analysis of Community-Based Studies’, *Journal of the American Society of Nephrology*, 15(5), pp. 1307–1315.

Wen, Y. H., Wu, S. S., Lin, C. H. R. and Tseng, K. H. (2016) ‘A Bayesian Approach to Identifying New Risk Factors for Dementia’, *Medicine*, 95(21), pp. e3658.

WHO (2016) *Fact sheet No.317: Cardiovascular diseases (CVDs)*. Retrieved January 24, 2017, from http://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

WHO (2017) *WHO | Cardiovascular diseases (CVDs), WHO*. World Health

Organization. Retrieved April 2, 2018, from http://www.who.int/mediacentre/factsheets/fs317/en/

WHO (2018) ‘WHO - 10 leading causes of death in females’. Retrieved May 13, 2018, from https://www/who/int/gho/women_and_health/mortality/situatio n_trends_ causes_death/en/

Witten, I. H., Frank, E., Hall, M. A. and Pal, C. J. (2016) *Data mining : practical machine learning tools and techniques*. Morgan Kaufmann.

Witteveen, A., Nane, G. F., Vliegen, I. M. H., Siesling, S. and Ijzerman, M. J. (2018) ‘Comparison of Logistic Regression and Bayesian Networks for Risk Prediction of Breast Cancer Recurrence’, *Medical Decision Making*, 38(7), pp. 822–833.

Wong, R. S. Y. and Ismail, N. A. (2016) ‘An application of Bayesian approach in modeling risk of death in an intensive care unit’, *PLoS ONE*, 11(3), pp. 1–17.

Wright, S. (1921) ‘Correlation and causation’, *Journal of agricultural research*, 20(7), pp. 557–585.

Wulsin, L. R., Horn, P. S., Perry, J. L., Massaro, J. M. and D’agostino, R. B. (2015) ‘Autonomic Imbalance as a Predictor of Metabolic Risks, Cardiovascular Disease, Diabetes, and Mortality’, *J Clin Endocrinol Metab*, 100, pp. 2443–2448.

Xie, J., Liu, Y., Zeng, X., Zhang, W. and Mei, Z. (2017) ‘A Bayesian network model for predicting type 2 diabetes risk based on electronic health records’, *Modern Physics Letters B*, 31(19–21), p. 1740055.

Yang, Y., Liu, F., Wang, L. and Ma, X. (2017) ‘Association of Husband Smoking With Wife’s Hypertension Status in Over 5 Million Chinese Females Aged 20 to 49 Years’, *Journal of the American Heart Association*, 6(3).

Yap, C., Lin, X. and Cheung, K. (2015) ‘Comparison of a frequentist and Bayesian response-adaptive randomisation approach in multi-stage phase II selection trials with multiple experimental arms’, *Trials*, 16(2), pp. 58.

Yuan, J., Zou, X. R., Han, S. P. and Wang, X. Q. (2017) ‘Prevalence and risk factors for cardiovascular disease among chronic kidney disease patients: Results from the Chinese cohort study of chronic kidney disease (C-STRIDE)’, *BMC Nephrology*, 18(1), pp. 1–12.

Yunus, M., Ms, S., Mz, N. A., L, R. and Kh, T. (2003) ‘the Prevalence of Hypertension and Smoking in the Subdistrict of Dengkil, Selangor', *Malaysian Journal of Public Health Medicine*, pp. 5–9.

Zador, Z., Huang, W., Sperrin, M. and Lawton, M. T. (2017) 'Multivariable and Bayesian Network Analysis of Outcome Predictors in Acute Aneurysmal Subarachnoid Hemorrhage: Review of a Pure Surgical Series in the Postinternational Subarachnoid Aneurysm Trial Era', *Operative Neurosurgery*, 14(6), 603–610.

Zandecki, L., Sadowski, M., Janion, M. and Gasior, M. (2018) 'P579Sex effect on mortality in STEMI - a retrospective analysis of a large nationwide database', *European Heart Journal*, 39(suppl_1), pp. 564–579.

Zhang, Z. (2016) 'Model building strategy for logistic regression: purposeful selection', *Annals of Translational Medicine*, 4(6), pp. 111–111.

Zhou, L., Pan, S., Wang, J. and Vasilakos, A. V. (2017) 'Machine learning on big data: Opportunities and challenges', *Neurocomputing*, 237, pp. 350–361.

Zuhdi, A. S. M., Ahmad, W. A. W. and Hian, S. K. (2016) 'Acute coronary syndrome in the elderly: The Malaysian National Cardiovascular Disease Database-Acute Coronary Syndrome registry', *Singapore Medical Journal*, 57(4), pp. 191–197.

LIST OF PUBLICATIONS

1. **Juhan, N.**, Zubairi, Y. Z., Zuhdi, A. S. M., Khalid, Z. M. and Ahmad, W.A.W. (2018), Gender Differences in Mortality Among ST Elevation Myocardial Infarction Patients in Malaysia from 2006 To 2013. *Annals of Saudi Medicine*, 38(1), 481-487. ***(ISI-Indexed, IF: 0.633)***

2. **Juhan, N.**, Zubairi, Y. Z., Khalid, Z. M. and Zuhdi, A. S. M. (2019), Risk Factors of Mortality Among Male Patients with Cardiovascular Disease in Malaysia Using Bayesian Analysis. *Iranian Journal of Public Health*. (In Press) ***(ISI-Indexed, IF: 1.053)***

3. **Juhan, N.**, Khalid, Z. M., Zubairi, Y. Z., Zuhdi, A. S. M. and Ahmad, W.A.W. (2019), Risk Factors of Cardiovascular Disease Among ST Elevation Myocardial Infarction Male Patients in Malaysia. *Jurnal Teknologi (Sciences and Engineering)*, 81(3), 1-10. ***(Scopus-Indexed, IF: 0.180)***

4. **Juhan, N.**, Zubairi, Y. Z., Khalid, Z. M. and Zuhdi, A. S. M. (2019). A Comparison B etween Bayesian and Frequentist Approach in the Analysis of Risk Factors for Female Cardiovascular Disease Patients in Malaysia. *ASM Science Journal*. (In Press) **(*Scopus-Indexed, IF: 0.3*)**

5. **Juhan, N.**, Zubairi, Y. Z., Khalid, Z. M. and Zuhdi, A. S. M. (2018), Identifying Risk Factors for Female Cardiovascular Disease Patients in Malaysia: A Bayesian Approach. *Matematika*, 34(3), 15-23. ***(ESCI-Indexed)***