STATISTICAL APPRAISAL IN SOLVING SOME MEDICAL PROBLEMS

LIM FONG PENG

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

STATISTICAL APPRAISAL IN SOLVING SOME MEDICAL PROBLEMS

LIM FONG PENG

THESIS SUBMITTED IN FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **LIM FONG PENG**

Registration / Matric No: **SHB110003**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Project Paper / Research Report/ Dissertation / Thesis ("this Work"):

"**STATISTICAL APPRAISAL IN SOLVING SOME MEDICAL PROBLEMS**"

Field of Study : **BAYESIAN APPROACH IN SURVIVAL ANALYSIS**

I do solemnly and sincerely declare that:

(1) I am the sole author / writer of this Work,
(2) This Work is original,
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work,
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work,
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained,
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: PROFESSOR DR IBRAHIM BIN MOHAMED

Designation:

ABSTRACT

Interest in some medical problems has raised the need for the development of appropriate statistical techniques in order to provide reliable solutions. We look at two local medical scenarios which are of current interest; firstly, identifying the optimal number of lymph nodes removed for maximizing the survival and adequate nodal staging of local breast cancer patients, and secondly, studying the outlier detection in cross-over design for kinesiology study. In this thesis, we will discuss alternative and new methods to provide the solution to the scenarios above. For the breast cancer study, we investigate the influence of the number of lymph nodes removed (LNR) on survival of breast cancer patients using Chi-square test of independence and Wilcoxon test. We proceed to find the best-fitted logistic and Cox's regression models using forward selection and Bayesian model averaging procedures. The models are then used to assess the prognostic values of independent factors for survival at all thresholds of the number of LNR. For both types of regression models, we use not only the Wald statistic but also present the use of the Akaike Information Criterion to determine the optimal number of LNR which results in maximum differential in survival of the breast cancer patients. Similar procedure will be extended to the case of finding the dependence of number of LNR to the adequate nodal staging of the patients. For the kinesiology study, we employ both non-Bayesian and Bayesian framework to detect outliers in a 2×2 cross-over design. We consider the mixed model with different factors representing subject, period, treatment and carry-over effects. In non-Bayesian framework, we consider the classical studentized residual and provide a studentized residual using median absolute deviation to identify possible outlying subjects. The performances of both procedures in detecting subject outliers are compared via simulation. On the other hand, in Bayesian framework, we assume that the random subject effect and the errors are normal distributed. However, the outlying subjects come from normal distribution with different variance.

Due to the complexity of the resulting joint posterior distribution, we obtain the information on the posterior distribution from samples by using Markov Chain Monte Carlo method. We use two real data sets, the Malaysian Breast Cancer data and kinesiology data, obtained from the University of Malaya Medical Centre (UMMC). This study is able to provide solutions to the problems which are very beneficial to the local medical practitioners. The findings are very important as guidelines in the surgical management of breast cancer patients and in the usage of kinesiotapes in sports.

ABSTRAK

Kepentingan dalam beberapa masalah perubatan telah meningkatkan keperluan bagi pembangunan teknik statistik yang sesuai bagi menyediakan penyelesaian yang boleh dipercayai. Kami melihat dua jenis senario perubatan yang merupakan kegemaran semasa; pertama, mengenal pasti bilangan optimum nodus limfa yang dikeluarkan dalam peningkatan survival dan kecukupan pementasan nodus pesakit kanser payudara tempatan, dan kedua, mengkaji pengesanan titik tersisih dalam kajian cross-over bagi bidang kinesiologi. Dalam tesis ini, kami akan membincangkan kaedah-kaedah alternatif dan baru untuk menyediakan penyelesaian bagi senario di atas. Bagi kajian kanser payudara, kami menyiasat pengaruh bilangan nodus limfa yang dikeluarkan (LNR) terhadap survival pesakit kanser payudara dengan menggunakan ujian kemerdekaan Chi-square dan ujian Wilcoxon. Kami meneruskan untuk mencari model-model logistik dan Cox yang dilengkapi terbaik dengan menggunakan prosedur-prosedur pemilihan ke hadapan dan kaedah model Bayesian purata. Model-model tersebut kemudian digunakan untuk memperolehi nilai ramalan setiap peramal bebas untuk survival pada setiap ambang bilangan LNR. Bagi kedua-dua jenis model regresi, kami menggunakan bukan sahaja statistik Wald tetapi juga memperkenalkan penggunaan Akaike Information Criterion untuk menentukan bilangan optimum LNR supaya memberikan perbezaan maksimum dalam survival pesakit kanser payudara. Prosedur yang sama akan diperluaskan kepada kes yang mencari pergantungan bilangan LNR untuk kecukupan pementasan nodus pesakit. Bagi kajian kinesiologi, kami menggunakan kedua-dua rangka kerja bukan Bayesian dan Bayesian untuk mengesan titik tersisih dalam 2×2 kajian cross-over. Kami mengambil kira model bercampur dengan faktor-faktor berbeza yang mewakili subjek, tempoh, rawatan dan kesan sampingan. Dalam rangka kerja bukan Bayesian, kami membincangkan pengiraan studentized reja klasik dan mencadangkan satu studentized reja baharu yang

menggunakan sisihan mutlak median untuk mengesan subjek-subjek tersisih yang mungkin wujud. Prestasi kedua-dua prosedur dalam pengesanan subjek-subjek tersisih dibandingkan melalui simulasi. Di sebaliknya, dalam rangka kerja Bayesian, kami menganggap bahawa kesan subjek rawak dan ralat akan dijana daripada taburan normal. Walau bagaimanapun, subjek tersisih berasal daripada taburan normal dengan varians yang berbeza. Disebabkan taburan posterior tercantum yang kompleks, kami mendapatkan maklumat mengenai taburan posterior dari sampel yang dijana melalui pensampelan Markov Chain Monte Carlo (MCMC). Kami menggunakan dua set data sebenar; data Kanser Payudara Malaysia dan data kinesiologi, yang diperolehi daripada *University of Malaya Medical Centre* (UMMC). Kajian ini dapat memberikan penyelesaian kepada masalah yang amat memberi manfaat kepada pengamal perubatan tempatan. Hasil kajian adalah amat penting sebagai garis panduan dalam pengurusan pembedahan pesakit kanser payudara dan dalam penggunaan kinesiotapes dalam sukan.

ACKNOWLEDGEMENT

First of all, I sincerely express my infinite gratitude to my supervisor, Professor Dr. Ibrahim bin Mohamed for his encouragement, patience and critical reviews towards the completion of this project. Without his guidance and help, I could never have accomplished this difficult task. I would also like to extend my gratitude to Professor Dr. Nur Aishah binti Mohd Taib, Associate Professor Dr. Noorizam Daud and Dr. Adriana Irawati Nur binti Ibrahim for their guidance, providing related information, and advice. I would like to thank Dr. Goh Siew Li for providing the kinesiology data. I feel thankful to all lecturers who had taught me throughout my basic degree up to the completion of PhD study.

My appreciation goes to my dear parents, Mr. Lim Tin Lee and Madam Tee May Lee, and my family for their encouragement and continuous support. I would like to share my joy with them and they will always be in my heart.

I am especially indebted to those who had given me advice and providing suggestions throughout my study. I would like to thank all my fellow friends who had given me support and encouragement when I encountered problems and on the brink of giving up.

Thank you Universiti Putra Malaysia (UPM) for giving the opportunity for me to continue my degree by providing the scholarship for my study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background of the study

Statistics is the process of collection, analysis, interpretation, presentation, and organization of data. Statistics has been shown to be useful to describe different scenario found in scientific, industrial, or societal problems. The choice of relevant statistical techniques for a data analysis can be made based on the needs of the existing problems. In other words, the statistical techniques should not be applied rigidly in all situations according to their general guidelines, but need to be tailored according to the situation. Here, we are interested to further show the important role of statistics in medical studies by suggesting alternative and new statistical methods as the solution for some medical problems, in particular breast cancer and the sport related studies.

In this century, cancer arises as a great hazard against human health. Cancer is one of the top ten leading causes of death. In the report of World Health Statistics 2008, it is estimated that 7.4 million people died of cancer in 2004 and, if current trends continue, 80 million more will have died after 2015. Breast cancer has attracted a lot of interest among the researchers for the past few decades since it is the most common cancer among women in many countries. Breast cancer incidence rates are increasing steadily in all low- and medium-resource countries. The cause of breast cancer is not known, although risk factors such as family history of breast or ovarian cancer (particularly first-degree relatives on either the mother's or father's side); early age at menarche and late age at first childbirth, menopausal hormone use, obesity, and alcohol intake have been identified to increase the risk (Sankaranarayanan et al., 2011).

Prevention is better than cure. Effective breast cancer awareness program is needed to reduce the incidence of breast cancer. Early detection and appropriate treatment are very important in preventing breast cancer deaths. As highlighted by

Sankaranarayanan et al. (2011), different factors will influence the choice of treatments. In recent years, there has been an explosion of life-saving treatment advances against breast cancer, such as radiation, surgery, hormonal therapy and chemotherapy. However, a multimodal approach is generally adopted as treatment for patients with breast cancer. Surgery is fundamental to management on the treatment choices of breast cancer patients. The surgical options include breast-conserving surgery with radiation therapy, or mastectomy with reconstruction, or mastectomy alone. The axillary lymph nodes should be explored and histologically studied in order to decide on the relevant treatment and prognosis.

Axillary lymph node dissection is a surgical procedure to remove the lymph nodes from under the arm (Bembenek and Schlag, 2000). It is the standard way to diagnose cancer by investigating the axillary lymph nodes in most breast cancer centers. Lymph node-negative means the lymph nodes is free from cancer while, lymph node-positive refer to cancer spreading to the lymph nodes. The results are then used to diagnose and stage breast cancer according to the number of positive lymph nodes and eventually to make informed decisions on the treatment. Also, axillary lymph node dissection is one of the way to remove cancer on cell that may have spread to the lymph nodes and reduce the risk of recurrence of cancer in the axilla. Previous studies provided substantial evidence that axillary lymph node dissection provides excellent local control of disease in the axilla, which may lead to improved overall survival (Diab et al., 1998; Haffty et al., 1997). As the number of lymph nodes removed is an important prognostic predictor for survival in breast cancer, question arises regarding the minimum number of nodes that should be removed in order to maximize the survival of breast cancer patients (Krag et al., 2003). Here, in the study, we investigate the influence of the number of lymph nodes removed on the survival for both node-positive and node-negative patients. The study is then proceeded to determine the

optimal number of lymph nodes to be removed for maximizing the survival in Malaysia context. Breast cancer patients in Malaysia have different level of exposure and awareness on the subject compared to patients from other developed country. Hence, there is a different in the characteristics of breast cancer patients for the studies conducted by different countries (Yip et al., 2006; Taib et al., 2008; Taib et al., 2011).

On the other hand, axillary lymph node dissection is a standard surgical technique used in the staging and treatment of the node positive breast cancer patient. Since the number of lymph nodes removed is significantly associated with the survival of breast cancer patients, it is imperative that accurate staging for the axilla is performed i.e. removing a minimum number of lymph nodes that accurately gives information on whether or not cancer has spread to the lymph node. Although axillary lymph node staging surgery has evolved towards sentinel node biopsy in early breast cancer, axillary dissection remains the gold standard in staging the axilla in low and middle income countries where the majority of patients are symptomatic and in advanced stages of the disease (Saxena et al., 2012). The minimum thresholds of lymph nodes removed for staging have been recommended (Fisher et al., 1981; Singletary et al., 2002; National Comprehensive Cancer Network, 2014; Erbes et al., 2014). However, these numbers vary depending on the surgeon's own experience, the thoroughness of pathologist reporting, and patient anatomy. As stated by Sakorafas et al. (2000), traditionally, clinical staging of axillary lymph nodes is performed by physical examination. If lymph nodes are felt, they are further categorized as being enlarged owing either to benign causes or to malignant involvement, depending on the consistency of the node (AJJC, 1993). However, clinical evaluation of axillary lymph nodes can be highly inaccurate. The simplest imaging technique for lymph nodes in patients with breast cancer is axillary mammography, but it is far from satisfactory. Ultrasonography or color Doppler studies of the axillary node basin may be useful in the identification of involved axillary

lymph nodes. Computed tomography (CT) of the axilla is more informative than physical examination or mammography. On the other hand, patient with different levels of axillary lymph node dissection remove different number of lymph nodes. For example, partial axillary lymph node dissection removes levels I and II, without transecting the muscle, and an average of 15–20 nodes are retrieved. Low axillary lymph node dissection removes level I and an average of 4–6 nodes are retrieved. Notably, the methodologies used to obtain these optimal numbers were not statistically sound. Most of them use tabulated data only to reach the conclusion (Fisher et al., 1981; Carter et al., 1989; Somner et al., 2004). Therefore, in this study, we are interested to further investigate the optimal number of lymph nodes to be removed for a reliable staging of the axilla and increase the survival of the breast cancer patients in Malaysia context.

Exercising is very important to maintain one's good health especially for those who are involved seriously in sport. The safety and performance aspect are of concern. In recent times, there is a great interest in the use of kinesiotape (KT) in sport, noticeably by well-known sportsman including tennis player. KT is an elastic tape that had been introduced in the 1970's to mimic human skin in elasticity and thickness for correcting of muscle function (Kase et al., 2003). Combination of its unique design and its peculiar taping methods has attracted many to consider the tape as an alternative therapy to alleviate musculoskeletal symptoms and to improve the performance of the sports.

In sport, KT is used widely among practitioners and sportsmen although there is little scientific evidence to support KT use (Williams et al., 2012). This leads one to wonder if there is a novel mechanism through which the tape is able to impart a sense of improved athletic performance, barring the placebo effect. In a review of KT study, many musculoskeletal outcome measures, such as muscle strength, proprioception,

power, range of movement, endurance and various functional performance tests had been studied, but none of these parameters demonstrated convincing association with KT use (Baltaci et al., 2011; Huang et al., 2011; De Hoyo et al., 2013; Ujino et al., 2013). At present, the effect of KT on the other outcome measure, including the peak oxygen consumption (VO_2 peak), has not been investigated yet. VO_2 peak is usually used to gauge cardiorespiratory fitness of an individual. In addition, we will also look at the problem of outlier in the data set from a study in kinesiology. In the presence of outliers, the resulting statistical inferences of the kinesiology studies may be inaccurate. This is because the presence of outliers will affect the investigation of the effectiveness of certain treatments we study (Chow and Tse, 1990; Liu and Weng, 1991). Hence, before conducting any relevant statistical analyses, the detection and removal of possible outliers from data set is an important step to ensure the accuracy of its corresponding outcomes. Here, we present two methods of outlier detection in crossover design for kinesiology study, using Bayesian and non-Bayesian approaches.

1.2 Statement of problem

In Malaysia scenario, no work can be found on determining the optimal threshold of lymph nodes removed for maximal survival and adequate nodal staging of local breast cancer patients. On the other hand, the detection of outliers in crossover study is also an important issue since the outcome of the study may be affected by including the outliers in the analysis. With the availability of data set provided by the UMMC, this study will present appropriate statistical techniques in order to provide reliable solutions to the problems. We provide alternative methods of identifying the optimal threshold of lymph nodes to be removed for maximal survival and adequate nodal staging of local breast cancer patients. We also provide both non-Bayesian and Bayesian frameworks to detect outliers in a cross-over design with application in

kinesiology study. The findings of the study are very important as guidelines in the surgical management of breast cancer patients and in the usage of kinesiotape in sports.

1.3 Objectives of study

Based on the statement of problem above, the researcher has outlined the following objectives for this study:

1 To provide alternative methods of identifying the optimal threshold of lymph nodes removed for maximizing the survival of local breast cancer patients.

2 To provide alternative methods of identifying the optimal threshold of lymph nodes removed for adequate nodal staging of local breast cancer patients.

3 To provide a method of detecting outliers in 2×2 crossover design with application in kinesiology study using non-Bayesian framework.

4 To provide a method of detecting outliers in 2×2 crossover design with application in kinesiology study using Bayesian framework.

5 To apply the proposed methods on real data set.

1.4 Thesis outline

This research attempts to provide new statistical methods of identifying the optimal threshold of lymph nodes to be removed for breast cancer patients and; investigating the outlier detection of 2×2 crossover design using Bayesian and non-Bayesian frameworks. This thesis is divided into four parts: introduction, breast cancer study, kinesiology study and conclusion.

Chapter 2 provides a literature review on the fundamental of axillary lymph node dissection for breast cancer patients and on the problem of optimal number of lymph nodes to be removed for maximal survival. A brief discussion on the 2×2 crossover

design with application in kinesiology study using Bayesian and non-Bayesian framework also included.

Chapter 3 presents the description of Malaysia breast cancer data obtained from University of Malaya Medical Centre (UMMC) breast cancer registry. We conduct the survival analysis of Malaysia breast cancer data and discuss the output in detail.

Chapter 4 presents the study of optimal number of lymph nodes removed for maximal survival in breast cancer patients. The influence of the number of lymph nodes removed on survival of breast cancer patients is investigated using chi-square test of independence and Wilcoxon test. We proceed to find the best-fitted logistic and Cox regression models using forward selection and the Bayesian model averaging (BMA) procedures. The models are then used to assess the prognostic values of independent prognostic factors of survival at all thresholds of the number of lymph nodes removed. For both types of regression models, we use not only the Wald statistic but also present the use of the Akaike Information Criterion to determine the optimal number of lymph nodes to be removed that give maximum differential in survival of local breast cancer patients. In this study, we apply the logistic and Cox regression models for analyses as both models are the simplest and well-understood by the medical practitioners. Besides, the medical data collected contain information on the survival status and time, which are suitable for Cox and logistic modeling. Furthermore, the BMA procedure is considered as it accounts for model uncertainty in linear regression models, but not in the forward selection procedure. According to Raftery et al. (1997), ignoring of model certainty may leads to the underestimation of uncertainty when making inferences about quantities of interest. As a Bayesian solution to this problem, the BMA procedure involves averaging over all possible models when making inferences about quantities of interest.

Chapter 5 presents the study of optimal number of lymph nodes removed for adequate nodal staging in breast cancer patients. Similar methodology used in Chapter 4, by excluding the Cox regression model, is considered here specifically for reliable staging of the axilla. In this chapter, we only consider the best-fitted logistic model as the response of the logistic model is the probability of lymph nodes positive/involved which does not involve the information on survival status.

Chapter 6 provides a literature review on the fundamental of the 2×2 crossover design with application in kinesiology study using Bayesian and non-Bayesian framework also included.

Chapter 7 presents the description of a kinesiology data obtained from University of Malaya Medical Centre (UMMC).

Chapter 8 presents two methods for outlier detection in standard 2×2 crossover design in non-Bayesian framework. We discuss the classical calculation of studentized residual and propose a new studentized residual using median absolute deviation to identify possible outlying subjects. The performances of both procedures in detecting subject outliers are compared via simulation. We then illustrate their applications on a real data set from a study in kinesiology.

Chapter 9 introduces a method for outlier detection in a standard 2×2 crossover design in Bayesian framework. We assume that the random subject effect and the errors to be generated from normal distributions. However, the outlying subjects are assumed to come from normal distribution with different variance. Due to the complexity of the

resulting joint posterior distribution, we obtain the information on the posterior distribution from samples by using Markov Chain Monte Carlo method.

Chapter 10 presents the summary in this research work and the suggestions for further research works.

CHAPTER 2: LITERATURE REVIEW

2.1 Breast cancer

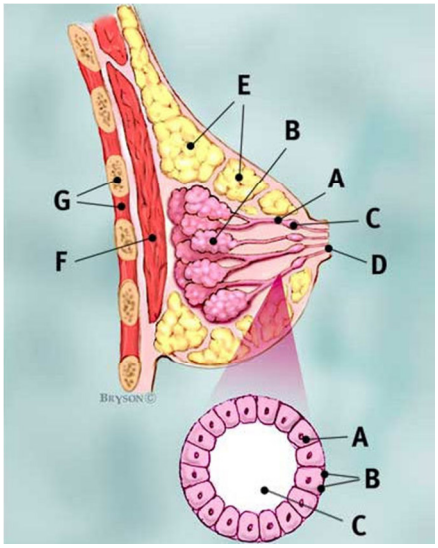
Cancer is a result of mutated cells. There are countless cells in a human body and cells are the body's basic unit of life. Each cell is pre-programmed to live, reproduce and die according to a precise schedule. Deoxyribonucleic acid (DNA) is the substance that directs a cell's activities. Normally, cells stop reproducing when the body has no need for them. However, carcinogens, which are any substances or radionuclide that are agent directly involved in causing cancer, may cause DNA in cells to change or mutate. Instead of dying, these abnormal cells continue to grow and multiply in an uncontrolled manner to form tumors.

There are two main types of tumors, based on shapes and effects, namely benign and malignant. Benign tumors are not cancerous. They refer to cells that reproduce at a slower rate and grow at their original site. They can be surgically removed and generally do not pose a serious health risk. Benign tumors usually do not recur once removed. On the other hand, malignant tumors are cancerous. They consist of cells that grow rapidly and invade normal tissues, and even spread to other parts of the body through the circulatory system or lymphatic system. Malignant tumors may be surgically removed. However, if their cells are not removed completely, they may metastasize or spread to other parts of the body to cause cancer recurrence. Cancer is a term used for diseases in which abnormal cells divide without control and are able to invade other tissues.

In our body, human cells metabolize naturally. However, many cancer risk factors such as chemical carcinogens, ionizing radiation, free radicals, microorganisms (bacteria, fungi, viruses), metabolic toxins, family history of cancer, endocrine malfunctions and immune disorders can cause the DNA in cells to change or mutate and

these abnormal cells begin to grow uncontrollably. In reality, there are abnormal cells or abnormal cell growth in every human body at every life stage. A healthy immune system is like an army in the body that provides around-the-clock protection by destroying these abnormal cells. However, the body cannot destroy these abnormal cells under some circumstances. For example, if the immune system is too weak, the body cannot identify, fight and overcome these abnormal cells effectively. These abnormal cells then reproduce uncontrollably, growing exponentially from a single cancer cell to form a malignant tumor and suppress surrounding tissues in the body, finally disrupting the natural biological order in the body to destroy health and lead to death.

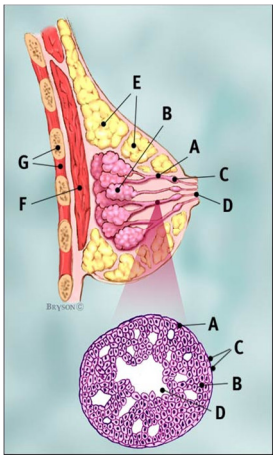
According to Breastcancer.org (2015), the term "breast cancer" refers to a malignant tumor that developed from cells in the breast known as *carcinoma in situ* (CIS). Usually breast cancer either begins in the cells of the lobules (L), which are the milk-producing glands, or the ducts (D), the passages that drain milk from the lobules to the nipple as shown in Figure 2.1(a). Less commonly, breast cancer can begin in the stromal tissues, which include the fatty and fibrous connective tissues of the breast. The malignant tumor can be categorized into non-invasive or invasive tumor. Non-invasive tumor (*ductal carcinoma in situ* (DCIS) or *lobular carcinoma in situ* (LCIS)) has not spread beyond the milk duct into any normal surrounding breast tissue, while invasive tumor is a tumor that has spread outside the milk duct or milk-masking glands and has grown into normal tissue inside the breast (*invasive ductal carcinoma* (IDC) or *invasive lobular carcinoma* (ILC)). The DCIS, LCIS, IDC and ILC images are presented in Figure 2.1 (b) – (d). Over time, cancer cells can invade nearby healthy breast tissue and make their way into the underarm lymph nodes, small organs that filter out foreign substances in the body as shown in Figure 2.2. If cancer cells get into the lymph nodes, they then have a pathway into other parts of the body. In fact, the stage of breast cancer



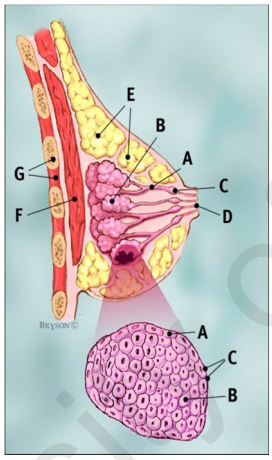
(a) Normal cell in Lobules

Breast profile: (a) to (e)
A Ducts
B Lobules
C Dilated section of duct to hold milk
D Nipple
E Fat
F Pectoralis major muscle
G Chest wall/rib cage

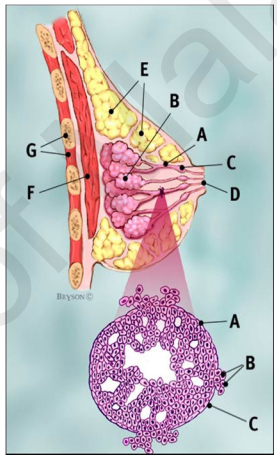
Enlargement (a)
A Normal duct cells
B Basement membrane
C Lumen (center of duct)



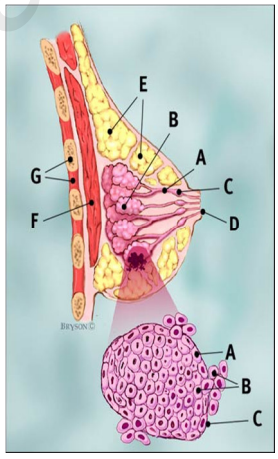
(b) DCIS



(c) LCIS



(d) IDC



(e) ILC

Enlargement: (b)
A Normal duct cells
B Ductal cancer cells
C Basement membrane
D Lumen (center of duct)

Enlargement: (c)
A Normal lobular cells
B Lobular cancer cells
C Basement membrane

Enlargement: (d)
A Normal duct cells
B Ductal cancer cells breaking through the basement membrane
C Basement membrane

Enlargement: (e)
A Normal lobular cells
B Lobular cancer cells breaking through the basement membrane
C Basement membrane

Figure 2.1: Breast structure with cancer (Breastcancer.org, 2015)

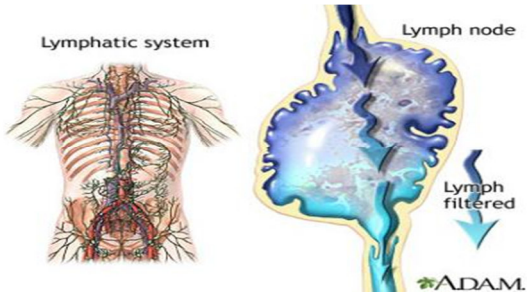


Figure 2.2: Lymphatic system and lymph node (Whitlock, 2014)

is defined based on how far the cancer cells have spread beyond the original tumor and to be discussed in Chapter 3.

Malaysia is a developing country in the Asia Pacific region with a population of 28.1 million. According to Taib et al. (2011), breast cancer is the commonest cancer in Malaysia with the age standardized incidence rate for females of 47.4 per 100,000 women. Malaysia has a multiracial composition with Malays being the majority followed by the Chinese and Indians. Lim et al. (2008) pointed out that the incidence was higher in Chinese women (ASR 59.9 per 100,000 women) compared to Indian (ASR 54.2) and Malay (ASR 34.9). Breast cancer specific information such as observed survival can be obtained with the mandatory reporting of deaths in Malaysia (Yip et al., 2006; Taib et al., 2008).

To reduce the incidence of advanced stage breast cancer, breast cancer awareness among women is needed. There are many factors that can influence a woman's risk of getting breast cancer, such as a person's age or race, family history, genetic risk factors, pregnancy, breastfeeding, food intake, stress level, alcohol, obesity, diet, physical activity (Herrinton and Brinton, 1993; Lam et al., 2000; Amir et al., 2010; Stephenson et al., 2010; Martin et al., 2010). The key to cancer prevention lies in a healthy, balanced diet and a healthy lifestyle to prevent cells from mutating into cancer cells and to nourish the immune system so that it can perform optimally when fighting cancer cells.

2.2 Axillary lymph node dissection

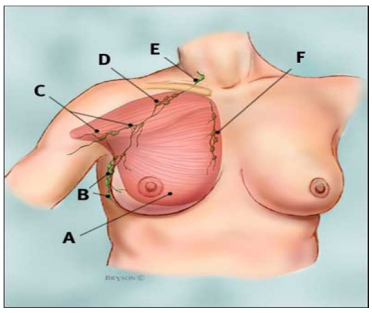
Lymph nodes are small clumps of immune cells that act as filters for the lymphatic system. The lymphatic system runs throughout the body and carries fluid and cells. The lymph nodes in the underarm, which are called the axillary lymph nodes, are the first place breast cancer is likely to spread. Hence, lymph node status is highly related to prognosis of breast cancer. Lymph node-negative means the lymph nodes do not contain cancer while lymph node-positive means the lymph nodes contain cancer. As mentioned by Chang and Hilsenbeck (2010), cancer found only in the breast (lymph node-negative) has the best prognosis of breast cancer. Prognosis is poorer when cancer has spread to the lymph nodes (lymph node-positive). The more lymph nodes that contain cancer, the poorer prognosis of breast cancer tend to be.

Axillary lymph node dissection is known as a surgical procedure that incises the axilla (also called armpit) to identify, examine, or remove lymph nodes. It has been the standard technique used in the staging and the treatment of node positive breast cancer. Although axillary lymph nodes staging has evolved towards sentinel node biopsy in early breast cancer, axillary dissection remains the gold standard in staging the axilla in low and middle income countries. Also, axillary lymph node dissection is done to remove any cancer that may have spread to the lymph nodes and reduce the risk of recurrence of cancer in the axilla.

Normally, axillary lymph node dissection is carried out during a modified radical mastectomy or radical mastectomy. It may also be performed with breast-conserving surgery. It is always a better choice when a sentinel lymph node biopsy is not suitable or if the sentinel lymph node is positive. There are three levels of an axillary lymph node dissection, from least aggressive to most aggressive, as showed in Table 2.1 and Figure 2.3. In fact, a traditional axillary lymph node dissection usually includes removal of the nodes in levels I and II from the “fat pad” under the arm. In the

Table 2.1: Three levels of an axillary lymph node dissection
(Magee-Womens Hospital of University of Pittsburgh Medical Center (UPMC), 2015)

Level	Description
I	The surgical removal of all tissue below the lower edge of the pectoralis minor muscle.
II	The surgical removal of the tissue lying underneath the pectoralis minor muscle.
III	The most aggressive dissection and is the surgical removal of tissue lying above the pectoralis minor muscle.



A pectoralis major muscle
B axillary lymph nodes: levels I
C axillary lymph nodes: levels II
D axillary lymph nodes: levels III
E supraclavicular lymph nodes
F internal mammary lymph nodes

Figure 2.3: Lymph node areas adjacent to breast area (Breastcancer.org, 2012)

output of this surgery procedure, a negative lymph node means that no cancer is found in the lymph node, while a positive lymph node means that cancer is found in the lymph node.

Side effects can happen any time during axillary lymph node dissection. Some may happen during, immediately after, or a few days or weeks after the procedure. These side effects include infection, bruising, seroma, lymphedema, chronic pain and so on. Van Bemmel et al. (2011) reviewed several methods to minimize seroma formation and associated morbidity. They found that the number of lymph nodes and extent of lymph nodes involved have shown to be significant influencing factors for seroma formation according to the results of some studies. In addition, Hashemi et al. (2004) and Woodworth et al. (2000) reported that removal of a larger number of lymph nodes results in greater injury of the lymph vessels. This may increase the painful of patients and might cause the formation of seroma.

2.3 Optimal number of lymph nodes removed for maximizing the survival of breast cancer patients

Axillary lymph node dissection is a standard surgical technique used in the treatment of node positive breast cancer. Previous studies provided substantial evidence that axillary lymph node dissection gave excellent local control of disease in the axilla, which may lead to improved overall survival. (Atkins et al., 1972; Langlands et al., 1980; Cabanes et al., 1992; Sosa et al., 1998; Orr et al., 1999; Bembenek and Schlag, 2000; Krag and Single, 2003; Sanghani, 2009).

Studies have found not only the importance of number of lymph nodes involved but also those removed. Lymph node ratio which takes into account number of nodes involved divided by number of nodes removed has been found to be an important prognostic factor (Taib et al. ;2008, 2011), and hence, highlighting the importance of the number of nodes removed in the management of breast cancer (Sosa et al., 1998; van der Wal, 2002; Weir et al., 2002; Krag and Single, 2003). Consequently, the optimal number of nodes to be removed for maximizing the survival has attracted a lot of attention among medical researchers.

For node-positive breast cancer patients, Sosa et al. (1998) investigated whether extent of axillary dissection is associated with survival based on the remaining cohort of 464 patients with stage I breast cancer, which had undergone axillary lymph node dissection. They estimated the overall survival, disease-free survival and recurrence for breast cancer patients according to two groups of the number of lymph nodes removed (< 10 or ≥ 10; < 15 or ≥ 15). The data suggested that the survival was improved with the removal of ≥ 10 and ≥ 15 lymph nodes, respectively. The criteria used for these lymph node groups were based on similar and previously-reported criteria. It was shown that there was no statistical difference between these lymph node groups.

Meanwhile, Bembenek and Schlag (2000) provided an overview of the current knowledge of the breast cancer and the surgical therapy. They suggested a removal of a minimum of 10 lymph nodes for the survival benefit based on previous studies (Cady and Sears, 1986; Graversen et al., 1988; Axelsson et al., 1992; Sosa et al., 1998).

Sakorafas et al. (2000) noted that Fisher et al. (1985) began a randomized trial to compare alternative local and regional treatments of breast cancer in 1971, all of which employ breast removal. There were 1765 patients at 34 United State and Canadian institutions participating in the National Surgical Adjuvant Breast Project enrolled in the trial and randomly assigned to treatment. However, only 1665 patients were judged to be eligible for a mean of 126 months. Their findings recommended a minimum threshold of 6 lymph nodes so that no axillary recurrence was observed.

Furthermore, Krag and Single (2003) analyzed the data from the Surveillance, Epidemiology, and End Results (SEER) database, from which 72102 patients were selected whose breast cancer had been diagnosed in 1988 or later and who were aged 40-79 years at diagnosis, had a single primary lesion, and had 0 to 3 positive lymph nodes. Their study only reported that the hazard rate of death was between 8% and 9% less for each additional 5 nodes removed, but lack of the discussion regarding the survival on the particular threshold of lymph nodes removed.

For node-negative breast cancer patients, Axelsson et al. (1992) examined the effect of extent of axillary node dissection in group of 7,145 patients enrolled onto their "low-risk" protocols. All patients had negative nodes and received no adjuvant systematic therapy. They found a highly significant correlation between the number of nodes removed and axillary recurrence-free survival, overall recurrence-free survival, and overall survival at a median observation time of 76 months. They found that a significantly better prognosis in all survival end points when a minimum threshold of 10 axillary lymph nodes are dissected.

Besides, Camp et al. (2000) carried out a study to determine whether the considered pathologic parameter can predict the outcomes of lymph node negative in breast cancer patients. The cohort for this study consisted of 290 patients who underwent breast resection at Yale – New Haven Hospital for invasive ductal breast carcinoma from 1 July 1983 to 1 July 1993, and followed through 8 September 1998. One of their findings conversely showed that a maximum threshold of 20 is considered for the benefit in survival of the patients. They reported that patients with 20 or more lymph nodes in their axillary resection had an overall survival of 82% compared to 87.3% for those with fewer lymph nodes removed. The 5-year survival rates were 84.7% and 96.3%, respectively.

In addition, Blancas et al. (2006) investigated whether the number of lymph nodes removed at axillary dissection is associated with recurrence and survival in node-negative breast cancer patients. The eligible patients of the study had to have survived at least 30 days from the time of diagnosis, to have undergone an axillary dissection with at least one lymph node recovered (not eligible if a sentinel lymph node biopsy was performed), and to have received no neoadjuvant systemic treatment. These eligible patients were diagnosed between 1 January 1982 and 31 December 2000 and treated at Clinic Hospital, University of Valencia, Spain. Their study found that node-negative breast cancer patients who have fewer than 6 nodes removed during axillary node dissection have worse outcome with a greater risk of relapse and a shorter breast cancer-specific survival.

On the other hand, Weir et al. (2002) evaluated the association between the number of lymph node removed at axillary dissection and recurrence and survival for patients with node-negative invasive breast cancer. Subjects of the study were 2278 women with pathologically node-negative invasive breast cancer, diagnosed from 1989 to 1993 in British Columbia, Canada. Women aged > 90 years, with pure in-situ,

bilateral invasive breast cancer or T4, N1, N2, or M1 stage, or who had axillary radiation were excluded. Two groups were defined for analysis: node-negative with no systemic therapy (n = 1468) and node-negative with systemic therapy (n = 810). Median follow-up was 7.5 years. The results showed that node-negative patients who received systemic therapy did not have a higher regional relapse rate or shorter overall survival when fewer nodes were removed.

Gao et al. (2014) performed a systematic and retrospective evaluation to examine the association between the number of tumor-free axillary lymph nodes removed and the outcomes for patients with pathologically negative nodes in a cohort of 603 patients with lymph node negative breast cancer. Their findings suggested that the number of tumor-free lymph nodes removed is an independent predictor in cases of lymph node-negative breast cancer. They mentioned that patients who had more than 10 tumor-free lymph nodes removed had a higher risk of death from breast cancer compared to patients who had 10 or fewer tumor-free lymph nodes removed. They chose 10 as a critical value because the median number of lymph nodes removed in the study was 10. They concluded that simple pursuit of a higher number of tumor-free lymph nodes removed may be of little use for improving survival rate.

2.4 Optimal number of lymph nodes removed for adequate nodal staging

The number of lymph nodes removed is one of the most important prognostic factors for survival in breast cancer. Hence, it is imperative that accurate staging for the axilla is performed. Although axillary lymph node staging surgery has evolved towards sentinel node biopsy in early breast cancer, in low and middle income countries, where the majority of patients are symptomatic and in advanced stages of the disease, axillary dissection remains the standard way used in staging the axilla.

The minimum threshold of lymph nodes removed for nodal staging varies across studies. Fisher et al. (1981) reported that the majority of histologic positive nodes breast cancer patients considered in their study have 1-3 involved nodes. The more nodes were removed, the more involved nodes were identified. Patients included in their study came from 34 National Surgical Adjuvant Breast Project member institutions in the United State and Canada. There were 641 women having a conventional radical mastectomy (node negative patients = 355 and node positive patients = 286) and 296 women having a total mastectomy with removal of 1 to 10 lymph nodes (node negative patients = 197 and node positive patients = 99).

On the other hand, Singletary et al. (2002) revised the American Joint Committee on Cancer (AJCC) staging system for breast cancer. This revised staging system was officially adopted for use in tumor registries in January 2003. They reported that the AJCC Cancer Staging Manual, Sixth Edition (2002) required 6 axillary lymph nodes to be removed and examined.

Under the auspices of the College of American Pathologists, Fitzgibbons et al. (2000) considered prognostic and predictive factors in breast cancer and stratified them into categories reflecting the strength of published evidence. They recommended a minimum threshold of 10 lymph nodes for staging. Similar suggestion was reported by the National Comprehensive Cancer Network (2014) in the United State.

Furthermore, Erbes et al. (2014) also mentioned that current guidelines recommend the removal of at least 10 lymph nodes (Carter et al., 1989; Somner et al., 2004), based on a mathematical model which determined the cut off at 10 lymph nodes to allow a 90% certainty of a true negative axillary status (Kiricuta and Tausch, 1992; Somner et al., 2004). Erbes et al. (2014) analyzed retrospectively the lymph node removed in 182 patients with axillary lymph node dissection after neoadjuvant chemotherapy and 351 patients with primary axillary lymph node dissection. They

found that the lower lymph nodes had no impact on clinical outcome. They showed that the lymph node count of less than 10 by axillary lymph nodes dissection after neoadjuvant chemotherapy might not be indicative for an insufficient axillary staging.

In addition, Somner et al. (2004) investigated the minimum number of lymph nodes needed in an axillary lymph node dissection specimen to be confident that the axilla is free from metastases. Data of the study were collected from the Edinburgh Breast Unit. Patients with large and high grade tumor were selected for axillary lymph node dissection. About 609 consecutive axillary lymph node dissections performed between October 1999 and December 2002 were reviewed. Full data about the underlying invasive breast cancer were available for 520 patients. Somner et al. (2004) suggested that 16 nodes should be regarded as a target to ensure a high level of confidence that the nodes are negative.

In literature, these optimal numbers vary based on surgeon philosophy and technique, the thoroughness of pathologist reporting, and patient anatomy. Notably, there is scarce information on the scientific basis of how these optimal numbers were obtained.

2.5 Chi-square test of independence

Chi-square test is one of the basic tests for statistical significance that is particularly appropriate for testing hypotheses about frequencies arranged in a frequency or contingency table (William G. Zikmund, 2010). There are several types of chi-square tests depending on the way the data are collected and the hypothesis being tested. It may be used both as a test of goodness-of-fit (comparing a collection of categorical data with some theoretical expected distribution) and as a test of independence (determine whether an association exists between two categorical variables of a population). For both tests, the underlying arithmetic is the same, but the

way expected values are calculated is different (McDonald, 2009).

In this study, chi-square test of independence is used to determine whether an association exists between two categorical variables of a population. For a test of independence, categorical data may be displayed in a contingency table as given in Table 2.2 (Diener-West, 2008).

Table 2.2: Layout of a contingency table

Criterion 2	Criterion 1					
	1	2	3	⋯	c	Total
1	n_{11}	n_{12}	n_{13}	⋯	n_{1c}	r_1
2	n_{21}	n_{22}	n_{23}	⋯	n_{2c}	r_2
3	n_{31}				⋮	⋮
⋮	⋮				⋮	⋮
r	n_{r1}	⋯	⋯	⋯	n_{rc}	r_r
Total	c_1	c_2	⋯	⋯	c_c	n

Under the null and alternative hypotheses below:

H_0: The two categorical variables are independent

(i.e., there is no relationship between them)

H_1: The two categorical variables are dependent

(i.e, there is a relationship between them),

the test statistic is

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

and the degrees of freedom are

$$(r-1)(c-1),$$

where r = number of rows,

c = number of columns,

O_i = the observed frequency in the i^{th} cell of the table,

E_i = the expected frequency in the i^{th} cell of the table.

Here, the chi-square statistic compares the observed count in each table cell to the count which would be expected under the assumption of no association between the row and column classifications. That is, observed counts are compared to expected counts. If the computed value of the χ^2 test statistic is greater than χ^2_α, the upper-tail critical value from a chi-square distribution with $(r-1)(c-1)$ degree of freedom, then the null hypothesis is rejected. Finally, the existence of the association between the two categorical variables can be determined.

Sosa et al. (1998) compared the breast cancer patient and tumor characteristics among nodal groups using the chi-square statistics. They show the distribution of patients according to the number of lymph nodes removed at surgery. The results showed that the lymph node groups were similar in terms of age, menopausal status, type of surgery, tumor size, hormonal receptor status, and adjuvant therapy. They found that there was no statistical difference in the number of lymph nodes removed between the two decades considered in their study.

Kuru (2006) investigated whether total number of nodes removed, negative nodes removed, and ratio of positive nodes to total nodes removed are predictors of survival in node positive patients. The comparisons between the total number of nodes removed and the probable potential prognostic factors and number of positive nodes based on chi-square analysis. The results indicate that the comparisons of the categories of total number of nodes removed with the categories of age, size, lymphovascular invasion (LVI), ER status, histological type, grade, menopausal status and adjuvant systematic therapy showed no differences. The proportion of patients with 4 or more positive nodes was significantly higher among patients with more than 15 axillary nodes removed compared to 1 – 15 nodes removed.

Gao et al. (2014) investigated the relationship between the number of lymph nodes removed and survival of patients diagnosed with lymph node-negative breast

cancer. Demographic data by the number of lymph nodes removed then were analyzed by chi-square test. They found that there was no statistical difference in the survival of patients between the categories of patient age, marital status, histologic grade, tumor size and adjuvant therapy.

2.6 Wilcoxon test

The Wilcoxon test is a nonparametric test for comparing survival curves and is an extension of the Wilcoxon rank sum test in the presence of censoring. It also requires that the censoring patterns for the two treatment groups be the same, but it does not assume proportional hazards. Hence, Wilcoxon test is preferred to the log-rank test when the hazard functions are thought to vary in ways other than proportionally.

Refer to Collett (2003), the Wilcoxon test, sometimes known as the Breslow test, is used to test the null hypothesis that there is no difference in the hazard functions for two groups of survival data. The Wilcoxon test is based on the statistic

$$U_W = \sum_{j=1}^{r} n_i\left(d_{1j} - e_{1j}\right),$$

where $d_{1j} =$ the number of deaths at time $t_{(j)}$ in the first group,

$e_{1j} = \frac{n_{1j}d_j}{n_j}$,

$n_j = n_{1j} + n_{2j}$,

n_{1j} = individuals at risk of death in the first group just before time $t_{(j)}$,

n_{2j} = individuals at risk of death in the second group just before time $t_{(j)}$.

The variance of the Wilcoxon statistic U_W is given by

$$V_W = \sum_{j=1}^{r} n_j^2\, v_{1j},$$

where $v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$, and so the Wilcoxon test statistic is

$$W_W = \frac{U_W^2}{V_W},$$

which has a chi-square distribution on one degree of freedom when the null hypothesis is true.

Sosa et al. (1998) made the survival curve comparisons between lymph node groups using the Wilcoxon test. From their results, the variables that were significantly associated with improved overall survival were younger age, premenopausal status, lack of breast cancer recurrence, and tumor size ≤ 1 cm, as well as increased number of nodes removed in the LN10 (<10 or $\geq$10 nodes) or LN15 (<15 or $\geq$15 nodes) groups. They also found that variables associated with the improved disease-free survival were tumor size ≤ 1 cm, no adjuvant therapy, later decade of diagnosis, and increased number of nodes removed in the LN10 or LN15 groups.

Peyre et al. (2008) investigated the relationship between number of lymph nodes removed and survival in esophageal cancer. Univariate analysis was performed using Wilcoxon test as appropriate to determine factors associated with survival at 5 years. Those factors with a *p*-value less than 0.2 then were used to construct a Cox regression model for all-cause mortality to determine independent predictors of survival. The choice of 0.2 had been used by Goodman et al. (1996), Milleron et al. (2004) and Peyre et al. (2008).

2.7 Logistic regression analysis

The goal of a logistic regression model is to understand a binary or proportional response (dependent variable) on the basis of one or more predictors. The response variable, y, is binary and parameterized in term of 1 or 0, in which 1 indicates a success and 0 as a failure. Success is to be thought of in a very wide sense, such as in term of yes or no, present or not present, dead or alive. There are two major uses to which statisticians employ a logistic model. It is widely used in the interpretation of parameter

estimates as odds ratio, and related to the calculation of the fitted values, which can be understood as the probability of success. Both of these usages play important roles in fields such as social science, health and medical research.

Logistic regression enables us to use regression models to predict the probability of a particular categorical response for a given set of independent variables (Hilbe, 2009). Logistic regression analysis is based on a generalized linear model for binary response using a logit link function (Stauffer, 2008). Let z be a binary random variable, taking on the values 1 and 0 with probability p and $1-p$ respectively, where $0 \leq p \leq 1$. If there are n independent random variables $z_1, z_2, \dots, z_n$ with probability p_i, then their joint probability distribution is given by

$$f(z_1, z_2, \dots, z_n; \theta_1, \theta_2, \dots, \theta_n) = \prod_{i=1}^{n} p_i^{z_i}(1-p_i)^{1-z_i}$$
$$= exp\left[\sum_{i=1}^{n} z_i \cdot log\left(\frac{p_i}{1-p_i}\right) + \sum_{i=1}^{n} log(1-p_i)\right]$$

which is a member of the exponential family.

The link function for logistic regression is given by the logit function

$$logit = g(p) = log\left(\frac{p}{1-p}\right),$$

the log of the odds ratio. Here, the odds ratio represents the probability of an event of interest compared with the probability of not having an event of interest. For the logistic regression model with one explanatory variable x, the link function, or logit, is given by

$$logit = g(p) = log\left(\frac{p}{1-p}\right) = \beta_1 + \beta_2 x \quad .$$

For the more general logistic regression model with one or more explanatory variable $\boldsymbol{x} = [x_1, x_2, \cdots, x_k]$, the link function is given by

$$logit = g(p) = log\left(\frac{p}{1-p}\right) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \ ,$$

where k is the number of parameters, $\beta_1, \beta_2, \cdots, \beta_k$, and $x_1 = 1$. Hence, the inverse of the logit link function g is then given by

$$p = \frac{exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)} .$$

Peyre et al. (2008) investigated the relationship between number of lymph nodes removed and survival in esophageal cancer. Logistic regression was performed to determine independent predictors of survival at 5 years. They showed that the number of nodes removed was one of the important predictor of survival.

Ghosh et al. (2010) examined the involution and density association in a large benign breast disease (BBD) cohort. They investigated associations between involution and parenchymal pattern using logistic regression analysis with parenchymal pattern as the outcome and involution as the predictor variable. Among 317 women with BBD in the Mayo Mammography Health Study, there was an inverse association between involution and quantitative percent density, and a strong positive association of involution with nondense area. No association was seen between involution and dense area.

2.8 Cox regression model

Cox proportional hazards model is the basic model for survival data and proposed by Cox (1972). It has also come to be known as the proportional hazards model. Although the model is based on the assumption of proportional hazard, no particular form of probability distribution is assumed for the survival times. The model is therefore referred to as a semi-parametric model. Cox proportional hazards model is commonly used to determine which combination of potential explanatory variables affect the form of the hazard function. It may also be used to obtain an estimate of the hazard function itself for an individual.

As explained by Collett (2003), suppose that patients are randomized to receive either a standard treatment or a new treatment, and let $h_S(t)$ and $h_N(t)$ be the hazards of death at time t for patients on the standard treatment and new treatment, respectively.

According to a simple model for the survival times of the two groups of patients, the hazard at time t for a patient on the new treatment is proportional to the hazard at that same time for a patient on the standard treatment. The Cox proportional hazards model then can be expressed in the form

$$h_N(t) = \psi h_S(t),$$

for any non-negative value of t, where ψ is a constant.

In this model, ψ is known as the relative hazard or hazard ratio. The value of ψ is the ratio of the hazards of death at any time for an individual on the new treatment relative to an individual on the standard treatment. If $\psi < 1$, the hazard of death at t is smaller for an individual on the new drug, relative to the standard. However, if $\psi > 1$, the hazard of death at t is greater for an individual on the new drug, and the standard treatment is superior.

The Cox regression model above can be generalized to the situation where the hazard death at a particular time depends on the values $x_1, x_2, \dots, x_p$ of p explanatory variables, $X_1, X_2, \dots, X_p$. Let $x = (x_1, x_2, \dots, x_p)'$ and $h_0(t)$ be the hazard function for an individual for whom the values of all the explanatory variables that make up the vector $\boldsymbol{x}$ are zero. The function $h_0(t)$ is called the baseline hazard function. The hazard function for ith individual can then be written as

$$h_i(t) = \psi(\boldsymbol{x}_i) h_0(t),$$

where $\psi(\boldsymbol{x}_i)$ is a function of the values of the vector of explanatory variables for the ith individual. Finally, let $\psi(\boldsymbol{x}_i) = exp(\eta_i)$, where $\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$, the general Cox regression model then becomes

$$h_i(t) = exp\big(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}\big) h_0(t)\,.$$

To determine whether extent of axillary dissection in patients with stage I breast cancer is associated with survival, Sosa et al. (1998) performed a multivariate analysis using Cox proportional hazards model in order to adjust for confounding prognostic

variables. For each lymph node group, they created separate models for overall survival, disease-free survival and recurrence-free survival. The hazard ratios and their corresponding 95% confidence intervals for the models were then calculated.

Van der Wal et al. (2002) conducted a study to determine the prognostic importance of the lymph node ratio (for node-positive patients) and the total number of removed lymph nodes (for node-negative patients) in addition to known prognostic factors in relation to disease recurrence or survival. The Cox proportional hazards model was used to calculate hazard ratios and 95% confidence intervals. Survival curves were also obtained using the Cox proportional hazards model, depicted as proportional survival.

Weir et al. (2002) evaluated the association between the number of lymph nodes removed at axillary dissection and recurrence and survival for patients with node-negative invasive breast cancer. They performed multivariate tests of the effect of number of nodes removed using Cox proportional hazards analysis.

In addition, Kuru (2006) investigated whether total number of nodes removed, negative nodes removed, and ratio of positive nodes to total nodes removed are predictors of survival in node positive patients. He carried out a multivariate analysis using Cox proportional hazards model. He then calculated the hazard ratios and 95% confidence intervals for the risk of dying from breast carcinoma.

Duraker et al. (2011) evaluated the prognostic significance of lymph node ratio, number of metastatic lymph nodes divided by number of removed nodes in 924 breast carcinoma patients with 1 – 3 metastatic axillary lymph nodes. The relative importance of the features was investigated using the Cox proportional hazards model.

Wu et al. (2013) examined the prognostic value of axillary lymph node ratio as compared to the number of involved nodes in patients with axillary lymph node-positive

breast cancer with mastectomy without radiation. Cox stepwise regression analysis was used for multivariate analysis, and significant variables in univariate analysis then were included in the Cox proportional hazards model.

2.9 Selection procedures

2.9.1 Forward selection

In the multiple regression model building, an efficient model selection strategy is needed to ensure a parsimonious model that fits the data well. The selected model should be simple to interpret and smooth the data rather than over fitting it.

Forward selection is a variable selection procedure that successively adds a single explanatory variable to a regression model a step at a time. At each step, each variable that is not already in the model is tested for inclusion in the model. Among the significant variables we obtain, the variable that having the lowest *p*-value will be chosen as the most significant one. Forward selection begin with a model including the explanatory variable that is most significant in the initial analysis, and continue adding variables until none of remaining explanatory variables are significant when added into the model. It is worth to point out that this multiple use of hypothesis testing means that the real probability of Type I error for an explanatory variable (i.e. the chance of including it into the model given it is not really necessary) does not equal the critical level we choose, that is 0.05. In fact, because of the complexity that arises from the complex nature of the procedure, it is essentially impossible to control the probability of error and it must be viewed as exploratory (Brant, 2004).

Consider a data set with k possible explanatory variable. A complete procedure of the forward selection can be illustrated as follow:

Step 1. Fit k single-variable regression models, each having one of the predictor variables in it. Calculate the overall model χ^2 value for each of the k fits. Select the fit with the highest value of χ^2.

Step 2. Using a significant level of $\alpha = 0.05$, the significance of predictor variable is tested. If it is not significant, terminate the procedure and conclude that none of the predictors are useful in predicting the response. If it is significant, retain the predictor variable, set $r = 1$, and proceed to step3.

Step 3. Fit $k - r$ reduced models, each having the r predictor variables from the previous stage of the selection process and one of the remaining candidate predictors. Calculate the overall model χ^2 value for each of the fits. Select the fit with the highest value of χ^2.

Step 4. Using a significant level of $\alpha = 0.05$, test the significance of the additional predictor variable using the χ^2 value with the model M_1containing the $r + 1$ predictors and the reduced model M_2 containing the r predictors from the previous stage of the selection process.

Step 5. If the χ^2 value is not significant, terminate the procedure and retain the r predictors from the previous stage. If it is significant, add the additional predictor to the r previously selected, increment r by 1, and return to step 3.

Weir et al. (2002) evaluated the association between the number of lymph nodes removed at axillary dissection and recurrence and survival for patients with node-negative invasive breast cancer. Multivariate tests of the effect of number of nodes removed were performed with Cox proportional hazards analysis. Using forward selection, the best-fitted prognostic model for each index of outcome then was constructed.

Taib et al (2011) investigated survival trends and factors that affect survival in Malaysia. As for identifying important prognostic factors, they employed the forward selection approach in the multivariate Cox regression model. The selection was based on the log-likelihood ratio statistic.

2.9.2 Bayesian model averaging

Referring to Raftery et al. (1997), a typical approach to data analysis is to carry out a model selection leading to a single "best" model and then to make inferences as if the selected model was the true model. However, this ignores a major component of uncertainty, namely uncertainty about model itself (Draper, 1995; Hodges, 1987; Leamer, 1978; Moulton, 1991; Raftery, 1988, 1996). As a consequence, uncertainty about quantities of interest can be underestimated. A complete Bayesian solution to this problem involves averaging over all possible combinations of predictors when making inferences about quantities of interest. This approach is called Bayesian model averaging (BMA). The idea of BMA was developed by Leamer (1978), and recently received a lot of attention in the literature. Leamer (1978) presented the basic paradigm for BMA and the fundamental idea that BMA accounts for the uncertainty involved in selecting the model. George (1999) discussed the use of BMA in the context of decision theory. Meanwhile, Draper (1995), Chatfield (1995), and Kass and Raftery (1995) reviewed BMA and the costs of ignoring model uncertainty. Hoeting et al. (1999) explained in detail the implementation of BMA. This approach provides optimal prediction ability (Madigan and Raftery, 1994).

According to Montgomery and Nyhan (2010), BMA is a more comprehensive approach to addressing model uncertainty, which allows us to assess the robustness of results to alternative specifications by calculating posterior distributions over

coefficients and models. Let X denote the $n \times p$ matrix of all the independent variables theorized to be predictors of outcome Y. Standard analyses would assume that

$$Y = X\beta + \epsilon\,,$$

where $\epsilon \sim N(0, \sigma^2 I)$. However, we might have uncertainty about which of the $q = 2^p$ model configurations from the model space $M = [M_1, M_2, \ldots, M_q]$ is the "correct" model.

The purpose of BMA is to explicitly incorporate this uncertainty into the model and therefore its inferences. The standard BMA approach represents the data as coming from a hierarchical mixture model. Firstly, a prior probability distribution is assign to the model parameters β and σ^2 and the model M_k. The model, M_k, is assumed to come from the prior probability distribution

$$M_k \sim \pi(M_k)$$

and the vector of model parameters is generated from the conditional distribution

$$\sigma^2 | M_k \sim \pi(\sigma^2 | M_k)$$

and $$\beta_\omega | \sigma^2, M_k \sim \pi(\beta_\omega | M_k, \sigma^2),$$

where $\Omega = \omega_1, \ldots, \omega_p$ represents a vector of zeroes and ones indicating the inclusion (or exclusion) of variables in model M_k.

The data generating process then is parameterized using the following conditional model:

$$Y | \beta_\omega, \sigma^2, M_k \sim N(X_\omega \beta_\omega, \sigma^2 I)\,.$$

The marginal distribution of the data under model M_k can therefore be written as

$$p(Y|M_k) = \iint p(Y|\beta_\omega, \sigma^2, M_k)\pi\,(\beta_\omega|\sigma^2, M_k)\pi(\sigma^2|M_k)d\beta_\omega d\sigma^2\,.$$

The posterior probability of model M_k is

$$p(M_k|Y) = \frac{p(Y|M_k)\pi(M_k)}{\sum_{k=0}^{q} p(Y|M_k)\pi(M_k)}.$$

It provides a coherent way of summarizing model uncertainty after observing the data. For instance, the expected value for the coefficient β_k can be easily derived after averaging across the model space M:

$$E(\beta_k|Y) = \sum_{k=0}^{q} p(M_k|Y)E(\beta_k|M_k, Y) .$$

$E(\beta_k|Y)$ represents the weighted expected value of β_k across every possible model configuration (with the weights determined by our priors and the performance of the models).

The difficulties associated with implementing the BMA approach are primarily computational. Calculating any quantities of interest involves solving or approximating $p(M_k|Y)$, which is often an intractable high-dimensional integral, for all $q = 2^p$ models under consideration. In practice, the calculations of these quantities use Bayes factors (Jeffreys, 1935,1961), a method for assessing the evidence in favor of two competing models, to compare each model with either the null model or the full specification (Kass and Raftery, 1995).

Given modest numbers of plausible covariates, even standard Markov chain Monte Carlo (MCMC) approaches become increasingly impractical as the model space expands. Hence, BMA computation has been radically improved. With the combination of increased computing power, the development of more analytically tractable prior specifications and the distribution of the BMA and Bayesian Adaptive Sampling (BAS) packages for R have made these techniques far more accessible.

For this study, we utilize the *bic.glm* software available for R in model selection. We use the Bayesian information criterion (BIC), as an approximation to a Bayes factor, in calculating the quantities of interest. The best-fitted model then is selected as that which corresponds to the lowest value of BIC.

2.10 Wald statistic and Akaike's information criterion

In statistical analysis approaches, a rigorous and theoretically justifiable approach to model fitting, selection and inference is required. Traditionally, with multiple linear regression modeling, analysts have used statistics such as parameter coefficient estimates and their significance, the coefficient of determination R^2, the residual standard error, the ANOVA F test, the adjusted R^2, and Mallows' C_p to evaluate the relative fit of models (Stauffer, 2008). In this study, we use the Wald statistic (χ^2) for this purpose.

According to Kyngas and Rissanen (2001), the Wald test is a way of testing the significance of particular explanatory variables in a statistical model. The Wald statistic, W, can be used to test under the hypotheses set as below,

$$H_0: \theta = \theta_0$$

$$H_A: \theta \neq \theta_0$$

and is calculated to be:

$$W = \frac{\left(\hat{\theta} - \theta_0\right)^2}{\frac{1}{I_n\left(\hat{\theta}\right)}}$$

$$= I_n\left(\hat{\theta}\right)\left[\hat{\theta} - \theta_0\right]^2$$

where $\hat{\theta}$ is the maximum likelihood estimate (MLE), and $I_n\left(\hat{\theta}\right)$ is the expected Fisher information evaluated at the MLE. When the null hypothesis is true, W tend to a chi-square distribution. The Wald test is easy to calculate but their reliability is questionable, particularly for small samples.

The Wald test, described by Polit (1996) and Agresti (1990), is one of a number of ways of testing whether the parameters associated with a group of explanatory variables are zero. If for a particular explanatory variable, or group of explanatory variables, the Wald test is significant, then we conclude that the parameters associated with these variables are not zero, so that the variables should be included in the model.

If the Wald test is not significant then these explanatory variables can be omitted from the model. When considering a single explanatory variable, Altman (1991) uses a *t*-test to check whether the parameter is significant. For a single parameter, the Wald statistic is the square of the *t*-statistic and so will give exactly equivalent results.

However, refer to Stauffer (2008), the statistic mentioned above test various assumptions of the model fit, such as whether the model is statistically equivalent to the null model. They do not directly assess the issue of whether the model is the best-fitting to the sample dataset. Unfortunately, they also sometimes tend to overfit the model to the sample dataset, with compounding of error. Therefore, one of a more appropriate information-theoretic approach to model fitting, using Akaike's information criteria (AIC), is recommended. This criterion provides a more rigorous and theoretically justified approach to model fitting that avoids the overfitting of models to the sample dataset and the compounding of error. The AIC was developed by the Japanese mathematician Hirotugu Akaike (1973, 1974). In general, for any probabilistic statistical model for a sample dataset with a likelihood function $\mathcal{L}$, AIC is defined using the deviance $= D = -2 \cdot log(\mathcal{L})$ as below:

$$\begin{aligned} \text{AIC} &= D + 2 \cdot k \\ &= -2 \cdot log(\mathcal{L}) + 2 \cdot k \end{aligned}$$

where $k =$ the number of parameters in the model. It is a first-order Taylor series approximation of the relative Kullback-Leibler distance (KL distance) between a model and the dataset. The best-fitted model in a collection of models which has the lowest AIC value. In our study, we propose the AIC instead of χ^2 to evaluate the relative fit of models. Their performance in model fitting is compared.

2.11 Summary

It is of interest to look at the problems of finding alternative statistical methods of identifying the optimal threshold of lymph nodes to be removed for survival and adequate nodal staging of breast cancer patients. We have presented the literature review of different methods as solutions for the problems we considered. This will be explored further in subsequent chapters.

CHAPTER 3: UMMC BREAST CANCER DATA – DESCRIPTION AND SURVIVAL ANALYSIS

3.1 Introduction

University Malaya Medical Centre (UMMC) is a government-funded medical institution located in Petaling Jaya, southwest corner of Kuala Lumpur. It was established in 1962. Breast cancer data for this study are obtained from the UMMC breast cancer registry, which contains information on 3280 patients diagnosed between January 1998 and December 2008. The patients are followed-up until 19 February 2013. Patients with missing information are excluded. These 1890 patients with complete details then are categorized into two groups: 1019 patients who do not have lymph nodes involvement (N_{neg}) and 871 patients with at least one of axillary nodes involved (N_{pos}). Both groups of patients are considered separately for the analysis in Chapter 4 but group as a whole in Chapter 5.

3.2 Description of data

The information recorded for each patient consists of age at diagnosis, race (Chinese, Malay, India or others), date of diagnosis, date of death or date of last contact, and pathological characteristics of tumor. The pathological characteristics considered include tumor size, tumor stage, lymphovascular invasion (LVI), total number of lymph nodes removed (NRN) and number of lymph nodes positive or involved (N_{pos}). In addition, the survival times of patients and status of patients are recorded at the end of the study. The mortality information is confirmed by referring to the records in the National Registration Department Malaysia. In our statistical modeling, we treated tumor size and NRN as continuous variables and the rest as categorical variables. We

note that N_{pos} was treated as continuous variable in Chapter 4 but as categorical variable in Chapter 5.

The first pathological characteristic considered in this study is tumor size. It indicates how large across the tumor is at its widest point and usually measured in millimeters (mm) or centimeters (cm). As we can see from Figure 3.1, tumor size 2 cm in diameter is as big as a peanut while 5 cm tumor is about the size of a lemon. Tumor size is used to determine the stage of cancer.

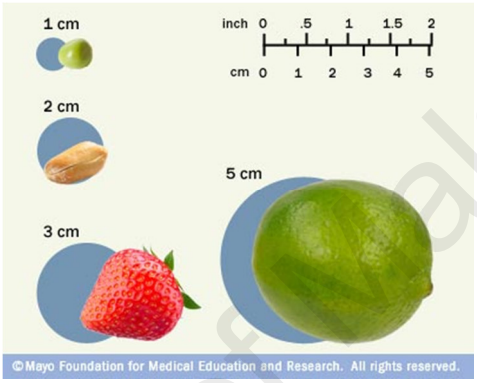


Figure 3.1: Four types of food representing the tumor size
(Mayo Foundation for Medical Education and Research)

The second pathological characteristic considered in this study is tumor stage. Staging is a way of describing where the cancer is located, how much the cancer has grown, and if or where it has spread. The staging of the breast cancer may provide some guidance for appropriate treatment regimen for patients. The most commonly used tool to describe the stage of cancer is the TNM system. TNM is an abbreviation for tumor (T), node (N), and metastasis (M). To determine the stage of cancer for each person, these three factors are considered as follows:

1. How large is the primary tumor and where is it located?
2. Has the tumor spread to the lymph nodes, and if so, how many nodes are involved?
3. Has the cancer metastasized to other parts of the body?

There are five stages: stage 0 (zero), which is noninvasive ductal carcinoma in situ (DCIS), and stages I through IV (one through four), which are used for invasive breast cancer. The details of these five stages are discussed as follows:

1. Stage 0 describes disease that is only in the ducts and lobules of the breast tissue and has not spread to the surrounding tissue of the breast. It is also called noninvasive cancer (see Figure 3.2).

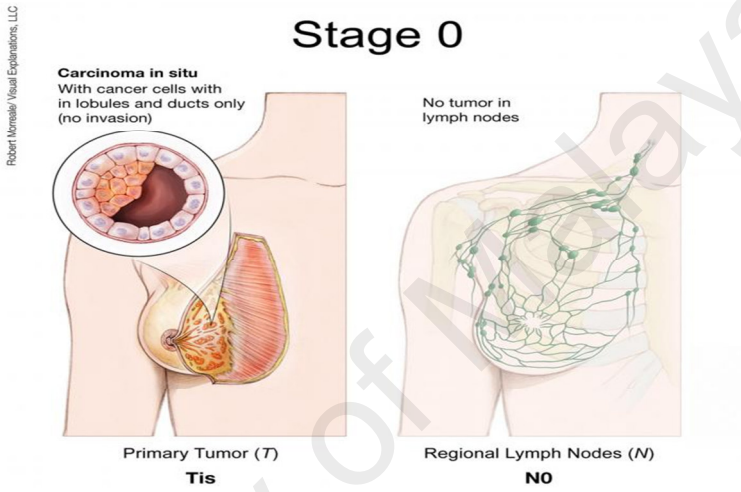


Figure 3.2: Abnormal cells are found in the lining of a breast duct and in the lobules of the breast (Cancer.Net, 2015)

2. Stage I describes invasive breast cancer and it is divided into stages IA and IB (see Figure 3.3).
    a. In stage IA, the tumor is 2 centimeters or smaller and has not spread outside the breast.
    b. In stage IB, no tumor is found in the breast or the tumor is 2 centimeters or smaller. Small clusters of cancer cells (larger than 0.2 millimeters but not larger than 2 millimeters) are found in the lymph nodes.

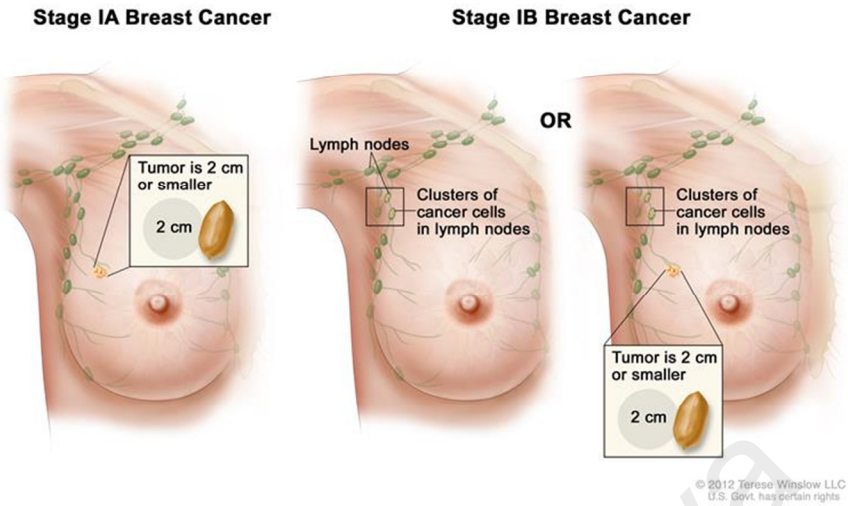


Figure 3.3: Stages IA and IB breast cancer (National Cancer Institute, 2015)

3. Stage II describes invasive breast cancer and it is divided into stages IIA and IIB (see Figures 3.4 and 3.5).
    a. In stage IIA, no tumor is found in the breast and cancer is found in 1 to 3 axillary lymph nodes or lymph nodes near the breastbone, or the tumor is 2 centimeters or smaller and cancer is found in 1 to 3 axillary lymph nodes or lymph nodes near the breastbone, or the tumor is larger than 2 centimeters but not larger than 5 centimeters and has not spread to the lymph nodes.
    b. In stage IIB, the tumor is larger than 2 centimeters but not larger than 5 centimeters and small clusters of cancer cells (larger than 0.2 millimeter but not larger than 2 millimeters) are found in the lymph nodes, or the tumor is larger than 2 centimeters but not larger than 5 centimeters and cancer is found in 1 to 3 axillary lymph nodes or lymph nodes near the breastbone, or the tumor is larger than 5 centimeters and has not spread to the lymph nodes.

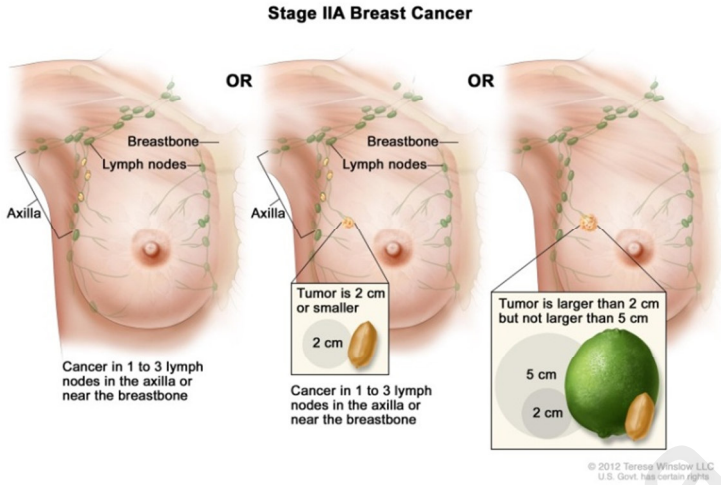


Figure 3.4: Stages IIA breast cancer (National Cancer Institute, 2015)

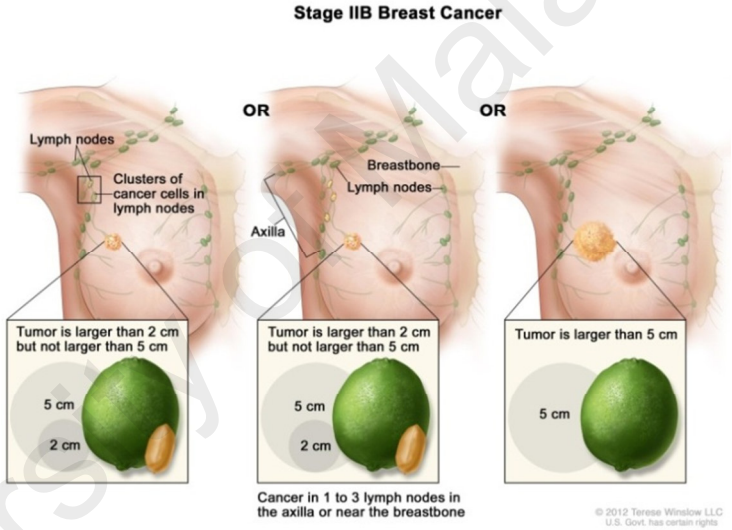


Figure 3.5: Stages IIB breast cancer (National Cancer Institute, 2015)

4. Stage III describes invasive breast cancer and it is divided into stages IIIA, IIIB and IIIC (see Figures 3.6 to 3.8).

    a. In stage IIIA, no tumor is found in the breast or the tumor may be any size and cancer is found in 4 to 9 axillary lymph nodes or lymph nodes near the breastbone, or the tumor is larger than 5 centimeters and small clusters of cancer cells (larger than 0.2 millimeter but not larger than 2 millimeters) are found in the lymph nodes, or the tumor is larger than 5 centimeters and cancer is found in 1 to 3 axillary lymph nodes or lymph nodes near the breastbone.

b. In stage IIIB, the tumor may be any size and cancer has spread to the chest wall and/or to the skin of the breast and caused swelling or an ulcer. Cancer may have spread to up to 9 axillary lymph nodes or the lymph nodes near the breastbone. Cancer that has spread to the skin of the breast may be inflammatory breast cancer.

c. In stage IIIC, no tumor is found in the breast or the tumor may be any size and may have spread to the chest wall and/or to the skin of the breast and caused swelling or an ulcer. Also, cancer has spread to 10 or more axillary lymph nodes, or to lymph nodes above or below the collarbone, or to axillary lymph nodes and lymph nodes near the breastbone. Cancer that has spread to the skin of the breast may be inflammatory breast cancer.

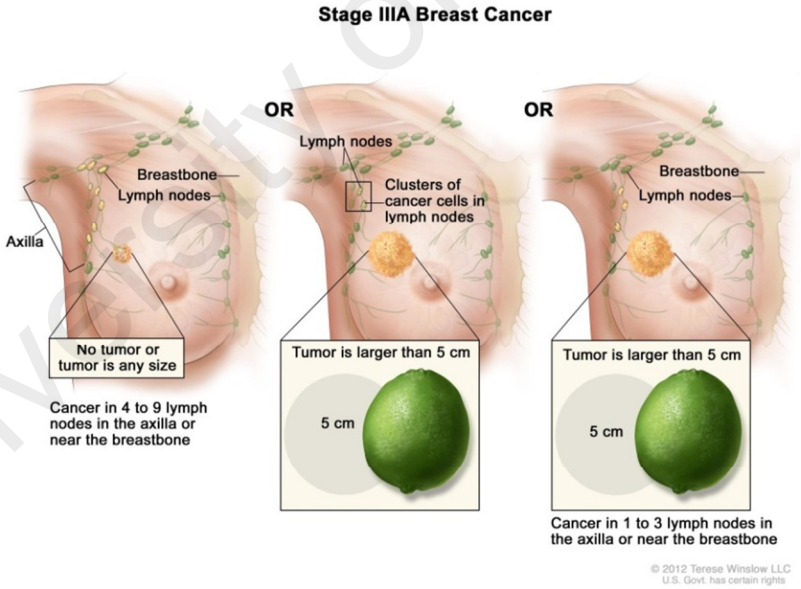


Figure 3.6: Stages IIIA breast cancer (National Cancer Institute, 2015)

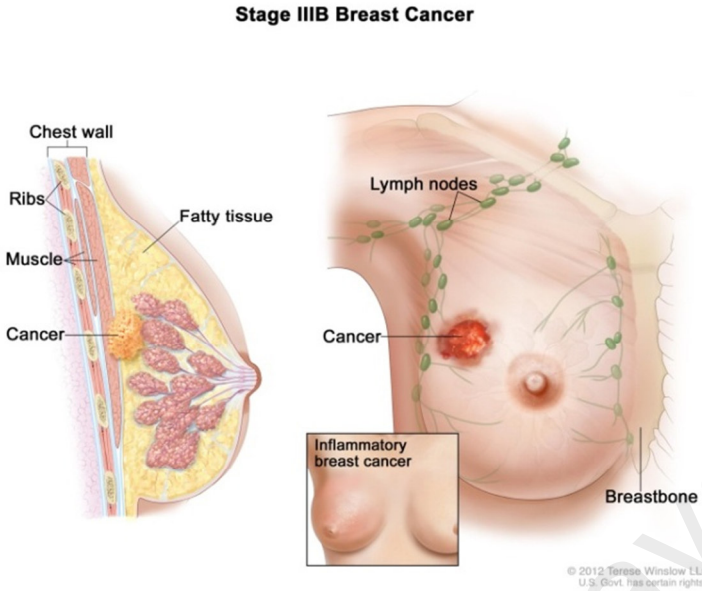


Figure 3.7: Stages IIIB breast cancer (National Cancer Institute, 2015)

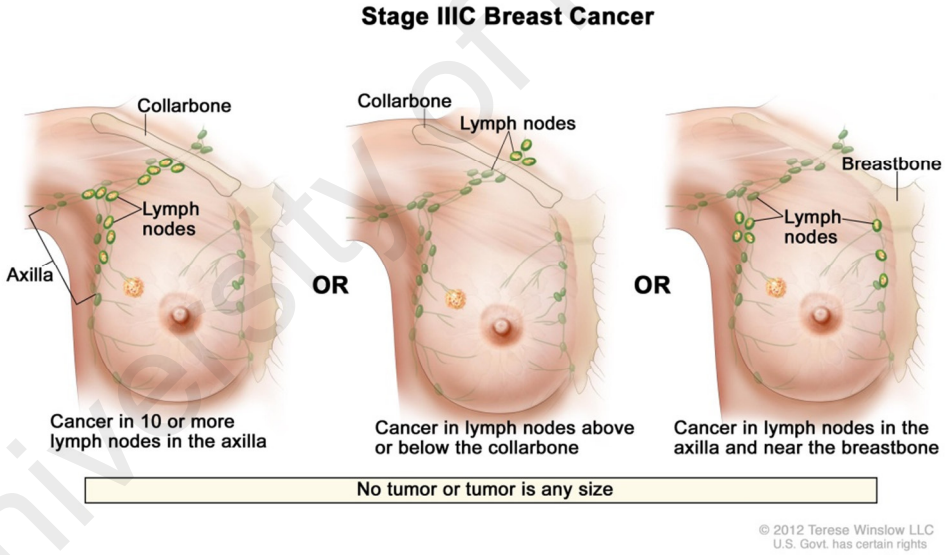


Figure 3.8: Stages IIIC breast cancer (National Cancer Institute, 2015)

5. Stage IV can be described as the cancer has spread to other parts of the body, most often the bones, lungs, liver, or brain (see Figure 3.9).

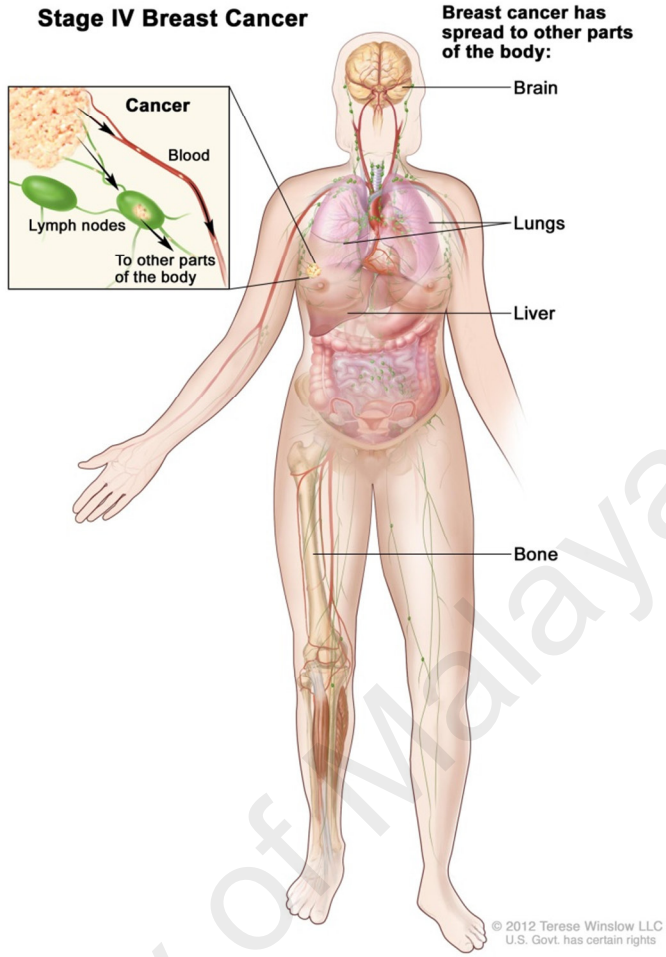


Figure 3.9: Stages IV breast cancer (National Cancer Institute, 2015)

The third pathological characteristic considered in this study is lymphovascular invasion denoted as LVI. The LVI is defined as tumor emboli present within a definite endothelial-lined space in the breast surrounding invasive carcinoma. The existence of LVI may help identify who is at increased risk for axillary lymph node and distant metastasis (Song et al., 2011).

The fourth and fifth pathological characteristics considered in this study are the total number of lymph nodes removed (NRN) and number of lymph nodes involved (N_{pos}), respectively. As a part of the lymphatic system, lymph nodes are small structures that work as filters for harmful substances. They contain immune cells that can help fight infection by attacking and destroying germs that are carried in through the lymph fluid. Lymph nodes removed (NRN) are the nodes that have been removed during

cancer surgery. Lymph node-negative (N_{neg}) means the lymph nodes do not contain cancer while lymph node-positive/involved (N_{pos}) means the lymph nodes contain cancer. The N_{pos} is rated N1, N2, or N3 depending on the number affected and the location. The details are as below:

1. N1: Cancer is found in 1-3 lymph nodes under the arm or lymph nodes within the breast.
2. N2: Cancer is found in 4-9 lymph nodes under the arm or lymph nodes within the breast.
3. N3: Cancer is found in 10 or more lymph nodes under the arm, or has spread under or over the collarbone. It may have been found in the underarm nodes as well as lymph nodes within the breast.

3.2.1 Background of data

Summary of the prognostic factors are given in Table 3.1. The age at diagnosis of a patient is divided into three levels; level one is for ≤ 40 years old, level two is for $41-59$ years old and level three is for ≥ 60 years old. The race is categorized into four levels; Chinese as level one, Malay as level two, India as level three and other races as level four.

The tumor stage is stratified into five levels where level zero is for stage 0, level one is for stage I, level two is for stage II, level three is for stage III, while level four represents stage IV. Further, the LVI are divided into two levels; "absent" status as level zero and "present" status as level. The tumor size (mm), NRN and N_{pos} are treated as continuous variables.

Two other important information recorded are the survival times measured in days and the status of patients. The survival times of patients take the number of days in which the individual enters the study until the date on which the individual die or last

known to be alive. The mortality information is confirmed with the records obtained from the National Registration Department Malaysia. Patients who are still alive at the end of the study or die because of non-breast cancer death are given status zero, while patients who die because of breast cancer are given status one.

Table 3.1: Description of breast cancer data

Prognostic Factors		**Level**
Age	≤ 40 41 – 59 ≥ 60	age1 age2 age3
Race	Chinese Malay India Others	race1 race2 race3 race4
Tumor stage	Stage 0 Stage I Stage II Stage III Stage IV	stage0 stage1 stage2 stage3 stage4
LVI	Absent Present	LVI0 LVI1
Tumor size		Continuous
NRN		Continuous
N_{pos}		Continuous

Table 3.2 gives the number of patients for every prognostic factor of breast cancer patients in the group as a whole and for each group of patients (N_{pos} and N_{neg} patients). Generally, it can be seen that most breast cancer patients were in the age range 41 – 59 years old, followed by age more than 60 and less than 40 years old. That is, patients between 41 to 59 years old are more prone to cancer compared to patients in the other two age groups. It is observed that Chinese has the highest number of patients who are diagnosed with breast cancer, followed by the Malay, the India and then the other race. The majority of breast cancer patients are Chinese because they seek treatment early at the hospital compare to other races (Taib et al., 2008; Taib et al., 2011).

Besides, the majority of breast cancer patients are diagnosed as having stage II and stage III, whereas stage II is the commonest stage. Next, it is found that around

Table 3.2: Breast cancer patient characteristics

Prognostic Factors			All patients		Node Positive Patients		Node Negative Patients	
			Frequency	%	Frequency	%	Frequency	%
Age	age1	≤ 40	264	13.97	116	43.94	148	56.06
	age2	41 – 59	1156	61.16	555	48.01	601	51.99
	age3	≥ 60	470	24.87	200	42.55	270	57.45
Race	race1	Chinese	1266	66.98	539	42.58	727	57.42
	race2	Malay	353	18.68	196	55.52	157	44.48
	race3	India	255	13.49	132	51.76	123	48.24
	race4	Others	16	0.85	4	25	12	75
Tumor stage	stage0	Stage 0	25	1.32	0	0	25	100
	stage1	Stage I	436	23.07	0	0	436	100
	stage2	Stage II	817	43.23	327	40.02	490	59.98
	stage3	Stage III	540	28.57	478	88.52	62	11.48
	stage4	Stage IV	72	3.81	66	91.67	6	8.33
LVI	LVI0	Absent	1024	54.18	300	29.3	724	70.7
	LVI1	Present	866	45.82	571	65.94	295	34.06
Tumor size (cm)		≤ 2	649	34.34	185	28.51	464	71.49
		2.1 – 5	946	50.05	472	49.89	474	50.11
		> 5	295	15.61	214	72.54	81	27.46
NRN		0	1	0.05	0	0	1	100
		1 – 5	107	5.66	27	25.23	80	74.77
		6 – 10	443	23.44	145	32.73	298	67.27
		11 – 15	610	32.28	277	45.41	333	54.59
		16 – 20	423	22.38	224	52.96	199	47.04
		21 – 25	179	9.47	97	54.19	82	45.81
		> 25	127	6.72	101	79.53	26	20.47
Status		Alive	1401	74.13	521	36.99	880	63.01
		Dead	489	25.87	350	72.12	139	27.88

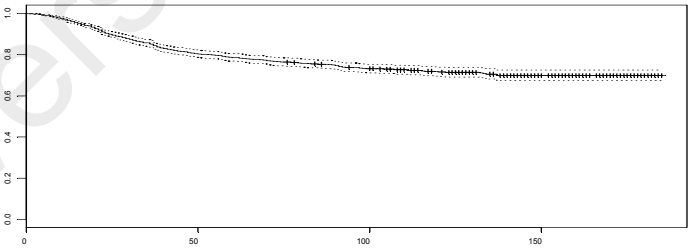
45.82% of patients had a positive lymphovascular invasion while 54.18% of patients had a negative lymphovascular invasion. More patients with N_{pos} had a positive lymphovascular invasion compared to patients with N_{neg}.

In addition, 50.05% of patients were found to have 2 cm to 5 cm size of tumor. It is obviously showed that patients with N_{pos} were diagnosed with large sized tumor compared to patients with N_{neg}. This might be due to lack of awareness among patients to go for early check-up. Further, most of the patients had between 6 and 20 lymph nodes removed during surgery. In this study, the survival was high for patients with N_{neg} (63%) but low for patients with N_{pos} (37%).

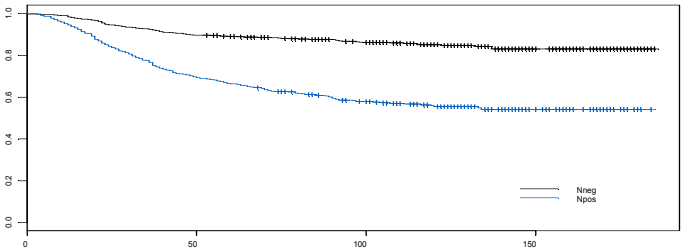
3.3 Survival analysis

3.3.1 Survival probability of breast cancer patients

Figures 3.10(a) and 3.10(b) give the plot of overall survival probability for the group as a whole and for each group of patients (N_{pos} and N_{neg} patients). Figure 3.10(b) shows that the Kaplan-Meier curves of N_{pos} and N_{neg} patients do not cross each other. The log-rank tests confirm that the survival of both groups are significantly different (p-value = 0). The five-year survival probability $S(60)$ of total patients is high, that is 0.786. The five-year survival probability is the estimated value in the confidence interval at time t = 60 months. It should be noted that the five-year rate has conventionally been used as an index for comparing survival across groups of patients by stage, race etc. and is often taken as a measure of cure rate. By comparing the five-year survival probabilities of both groups, we can conclude that N_{neg} patients had a better chance of survival compared to the N_{pos} patients as given in Table 3.3. It is noted that the 95% confidence intervals (C.I.) of the survival probabilities for both groups of patients do not overlap.



a) All patients



b) Node positive and node negative patients (p-value = 0)

Figure 3.10: Kaplan-Meier plot of overall survival for breast cancer patients
y-axis is survival probability, x-axis is observation

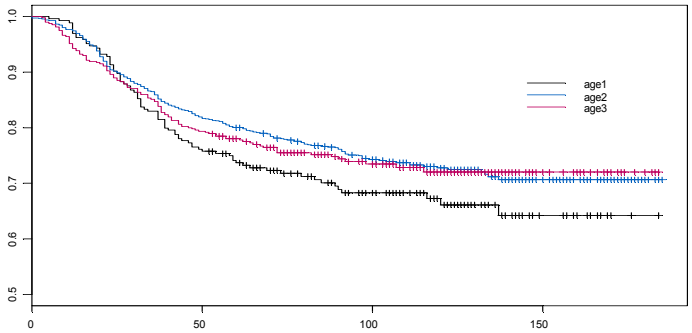
Table 3.3: Five-year probability of overall survival $S(60)$ for breast cancer patients

Patients	$S(60)$	95% C.I.
All	0.786	(0.768, 0.805)
Node Positive	0.664	(0.634, 0.697)
Node Negative	0.890	(0.871, 0.909)

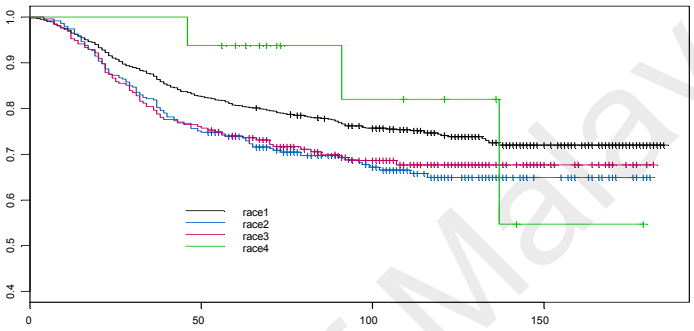
We then proceed to look at the survival probabilities for every prognostic factor. Figure 3.11 gives the Kaplan-Meier plots of each prognostic factor. The survival of patients in different levels of stage and LVI were significantly different. Their respective Kaplan-Meier curves do not cross each other and the *p*-values of the log-rank tests were less than 0.05. Meanwhile, the prognostic factor of age gave insignificant result and the Kaplan-Meier curve does cross.

Besides, Figure 3.12 gives the Kaplan-Meier plots of each prognostic factor for N_{pos} patients. The survival of patients in different levels of stage and LVI were significantly different. Their respective Kaplan-Meier curves do not cross each other and the *p*-values of the log-rank tests were less than 0.05. While, the other two prognostic factor; age and race gave insignificant result and the Kaplan-Meier curve does cross.

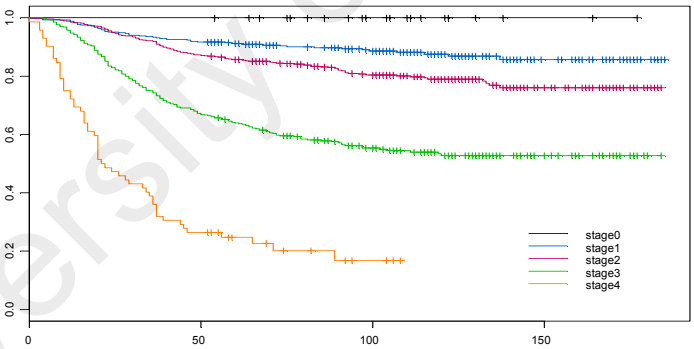
Furthermore, the Kaplan-Meier plots of each prognostic factor for N_{neg} patients were shown in Figure 3.13. The survival of patients in different levels of age and stage were significantly different, where their respective Kaplan-Meier curves do not cross each other and the *p*-values of the log-rank tests were less than 0.05. On the other hand, the survival of patients in different levels of race and LVI were not significantly different and the Kaplan-Meier curve does cross.



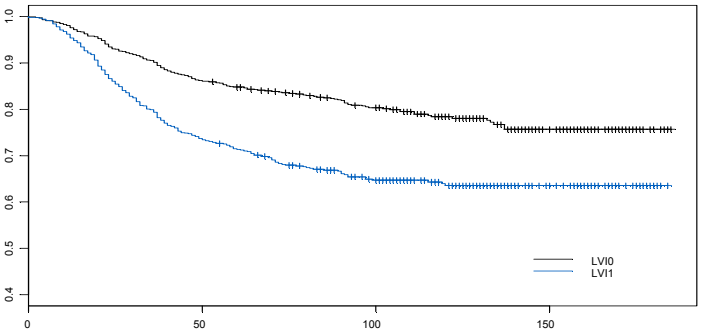
a) Age (*p*-value = 0.131)



b) Race (*p*-value = 0.004)

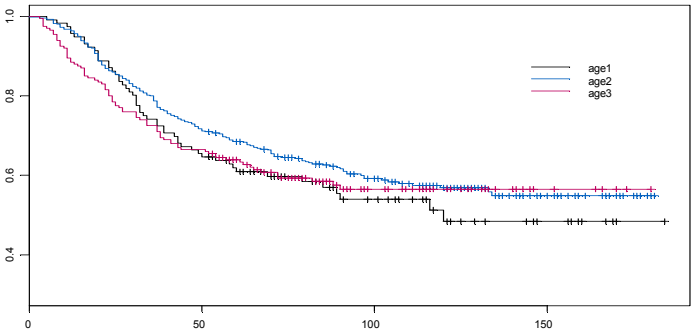


c) Stage (*p*-value = 0)

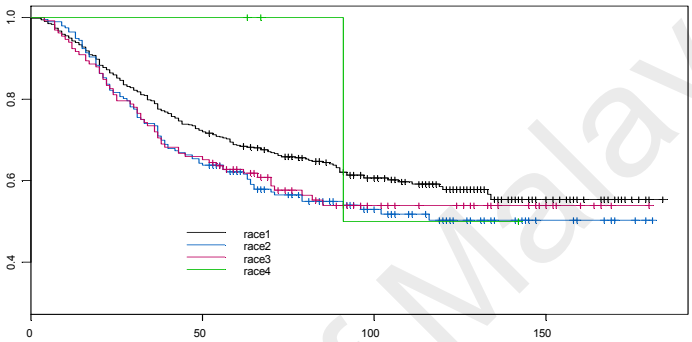


d) LVI (*p*-value = 0)

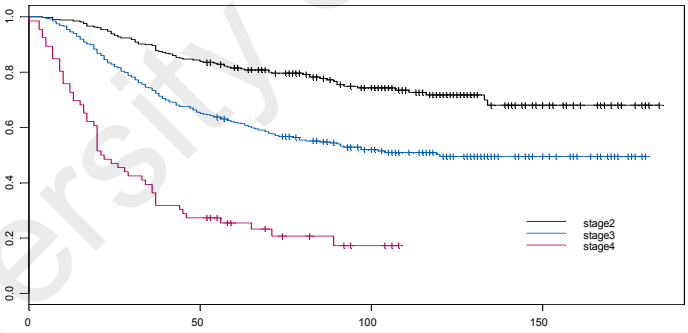
Figure 3.11: Kaplan-Meier plot of variables for all patients
y-axis is survival probability, *x*-axis is observation



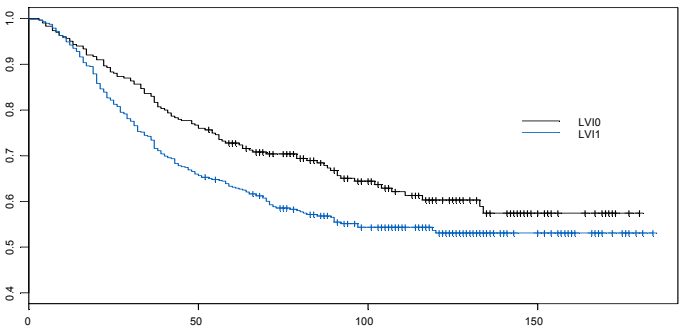
a) Age (p-value = 0.313)



b) Race (p-value = 0.148)

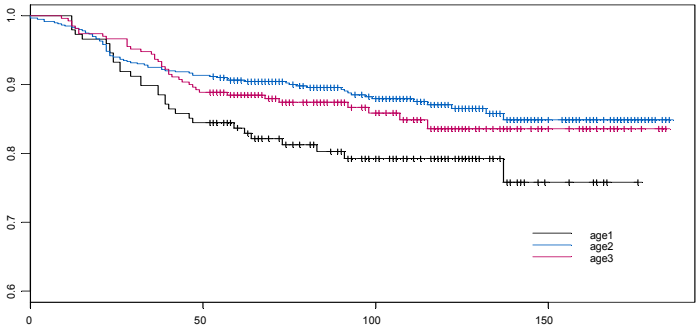


c) Stage (p-value = 0)

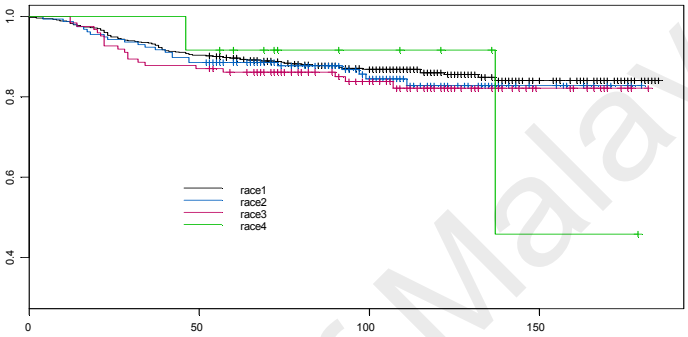


d) LVI (p-value = 0.007)

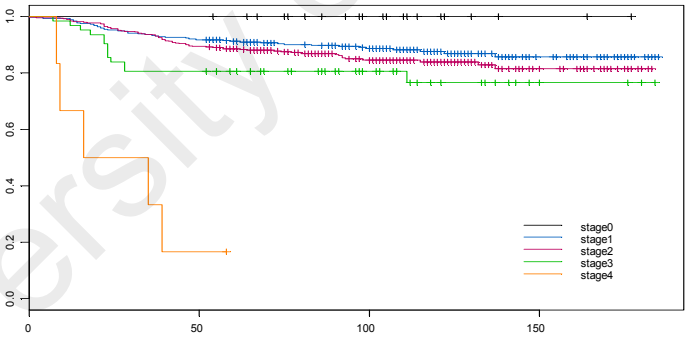
Figure 3.12: Kaplan-Meier plot of variables for node positive patients
y-axis is survival probability, x-axis is observation



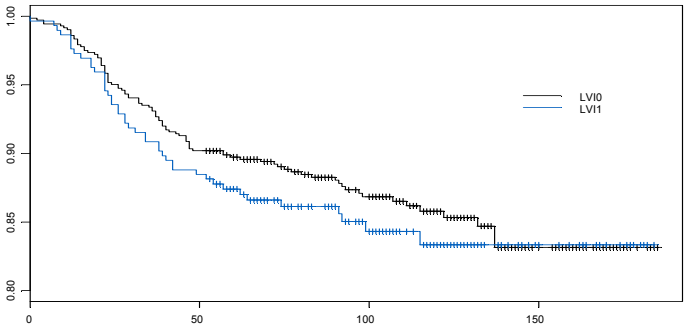
a) Age (p-value = 0.034)



b) Race (p-value = 0.804)



c) Stage (p-value = 0)



d) LVI (p-value = 0.396)

Figure 3.13: Kaplan-Meier plot of variables for node negative patients
y-axis is survival probability, x-axis is observation

Table 3.4 gives the five-year probabilities of breast cancer patients for each prognostic factor in the group of patients we considered. The five-year probabilities of survival showed that patients in age group between 41 to 59 years old had a better chance of survival compared to other age groups. In general, Chinese patients have best chance of survival compared to other races. We observed that Malay patients had significantly highest probability of mortality among both groups of N_{pos} and N_{neg} patients.

Table 3.4: Five-year probability survival $S(60)$ for breast cancer patients

Prognostic Factors		**All patients**		**Node Positive Patients**		**Node Negative Patients**	
		***S*(60)**	**95% C.I.**	***S*(60)**	**95% C.I.**	***S*(60)**	**95% C.I.**
Age	age1 (≤ 40)	0.737	(0.685, 0.792)	0.609	(0.526, 0.705)	0.834	(0.776, 0.897)
	age2 (41 – 59)	0.799	(0.776, 0.822)	0.684	(0.646, 0.724)	0.906	(0.882, 0.929)
	age3 (≥ 60)	0.779	(0.742, 0.817)	0.635	(0.571, 0.706)	0.884	(0.846, 0.923)
Race	race1 (Chinese)	0.807	(0.785, 0.829)	0.686	(0.648, 0.727)	0.895	(0.873, 0.918)
	race2 (Malay)	0.737	(0.692, 0.784)	0.618	(0.553, 0.690)	0.881	(0.832, 0.934)
	race3 (India)	0.738	(0.686, 0.794)	0.621	(0.543, 0.711)	0.860	(0.801, 0.924)
	race4 (Others)	0.901	(0.760, 1.000)	0.670	(0.423, 1.000)	0.846	(0.672, 1.000)
Tumor stage	stage0 (Stage 0)	1.000	(1.000, 1.000)	-	-	1.000	(1.000, 1.000)
	stage1 (Stage I)	0.911	(0.885, 0.938)	-	-	0.911	(0.885, 0.938)
	stage2 (Stage II)	0.856	(0.832, 0.881)	0.814	(0.773, 0.858)	0.884	(0.856, 0.913)
	stage3 (Stage III)	0.639	(0.600, 0.681)	0.618	(0.575, 0.663)	0.791	(0.691, 0.906)
	stage4 (Stage IV)	0.238	(0.157, 0.363)	0.245	(0.160, 0.377)	0.167	(0.028, 0.997)
LVI	Absent (LVI0)	0.847	(0.825, 0.869)	0.726	(0.677, 0.779)	0.897	(0.875, 0.919)
	Present (LVI1)	0.714	(0.684, 0.744)	0.631	(0.592, 0.672)	0.872	(0.834, 0.911)

In addition, the five-year survival probability for stage 0 breast cancer patients is 1. This means that all patients diagnosed with stage 0 breast cancer will be alive for at least five years after being diagnosed. In fact, patients diagnosed with this precancerous condition usually live long and healthy lives. Besides, patients in stage 1 had better chance of survival compared to other stages, while patients diagnosed with stage IV had significantly worse probability of survival compared to other stages. It was obviously shown that the early stage of cancer had a better survival compared to the advanced stage of cancer. Therefore, it is important that breast cancer patients come early for check-up upon noticing any symptom of breast cancer.

Further, we found that patients with negative lymphovascular invasion had better survival probability compared to those with positive lymphovascular invasion.

3.4 Summary

For the breast cancer data in this study, most of the patients were in the age range 41 to 59 years old and most were Chinese. The majority of them were stage 2 patients and their tumor size was less than 5 cm. Most of the patients had between 6 and 20 lymph nodes removed during surgery. The survival of patients was high for patients with N_{neg}, but low for patients with N_{pos}.

CHAPTER 4: BREAST CANCER – OPTIMAL NUMBER OF LYMPH NODES TO BE REMOVED FOR MAXIMAL SURVIVAL

4.1 Introduction

The number of lymph nodes removed is one of the most important prognostic factors for survival in breast cancer. In this chapter, we aim to investigate the influence of the number of lymph nodes removed on survival and then to determine the optimal number of lymph nodes to be removed for maximizing the survival. Data were obtained from University Malaya Medical Centre (UMMC) breast cancer registry. The analysis consists of three stages. Firstly, the chi-square test of independence is performed to determine the significant association between prognostic factors and survival status, while the Wilcoxon test is used to compare the hazard functions of the two or more levels at each observed time event. Secondly, we find the best-fitted logistic and Cox regression models using forward selection and Bayesian model averaging (BMA) procedures when applied on binary and time-to-event responses, respectively. Thirdly, the models are used to assess the prognostic values of independent prognostic factors of survival at all thresholds of number of lymph nodes removed. For both types of regression models, we use not only the Wald statistic (χ^2) but also introduce the use of the Akaike Information Criterion (AIC) to determine the optimal number of lymph nodes to be removed which results in maximum differential in survival of breast cancer patients. In this study, we apply the logistic and Cox regression models for analyses as both models are the simplest and well-understood by the medical practitioners. Besides, the medical data collected contain information on the survival status and time, which are suitable for Cox and logistic modeling. Furthermore, the BMA procedure is considered as it accounts for model uncertainty in linear regression models, but not in the forward selection procedure. According to Raftery et al. (1997), ignoring of model certainty may

leads to the underestimation of uncertainty when making inferences about quantities of interest. As a Bayesian solution to this problem, the BMA procedure involves averaging over all possible models when making inferences about quantities of interest.

4.2 Data description

Data for this study were obtained from the University of Malaya Medical Center (UMMC) breast cancer registry, which contains information on 3280 patients diagnosed between Jan 1998 and Dec 2008. The patients were followed-up until 19 February 2013. Patients with missing information were excluded. These 1890 patients with complete details were then categorized into two groups: 1019 patients who did not have lymph nodes involvement (N_{neg}) and 871 patients with at least one or more axillary nodes involved (N_{pos}). Both groups of patients were considered separately for the analysis. The information recorded for each patient can be found in Chapter 3.

4.3 Statistical analysis

4.3.1 Hypothesis testing

We first use the chi-square test of independence to assess individually whether the prognostic factors and survival status are independent of each other. This enables us to identify important prognostic factors associated with the survival of the breast cancer patients. Similarly, the statistical significance of the difference between the hazard functions for each level in individual prognostic factor is determined using Wilcoxon test. This is constructed by computing the observed and expected number of death in each of the level at each observed event time and then adding these to obtain an overall summary across all event time points. The null hypothesis tested here is that the risk of death is the same in all levels.

4.3.2 Regression modeling

It is also our intention to identify important prognostic factors taking into account the contribution of other factors. This is achieved using the logistic and Cox regression modeling. In logistic regression analysis, the model predicts the probability of death for a given set of independent prognostic factors based on a generalized linear model for binary response using a logit link function. Additionally, Cox regression model is used to estimate the risk or hazard of death at any time after the time origin of the study. The best-fitted models are obtained using forward selection and BMA procedures, respectively.

To obtain the best-fitted logistic and Cox regression models, the procedures are presented in detail as below:

Step 1. The data set are categorized according to all thresholds of NRN involved.

Step 2. For each threshold of NRN, a multiple logistic regression model is fitted using all predictor we consider in this study, using a significance level of $\alpha = 0.05$.

Step 3. At each threshold of NRN, the procedure of forward selection is conducted to find the best-fitted logistic regression model. The procedure starts with a simple model and adds terms sequentially until further additions do not significantly improve the fit.

Step 4. The best-fitted logistic regression model then is obtained for each thresholds of NRN.

Step 5. The Step 1 – 4 are repeated using the Bayesian model averaging (BMA) procedure.

Step 6. The complete procedure (Step 1 – 5) is repeated using Cox regression analysis in order to obtain the best-fitted Cox regression model for each thresholds of NRN.

4.3.3 Determination of optimal number of lymph nodes to be removed

The focus of the study is to determine the optimal number of lymph nodes to be removed (*NRN_T*), for two groups of patients (N_{neg} and N_{pos}). This is carried out in several steps. Firstly, we treat NRN as a categorical variable such that $NRN(k) = 0$ or 1, according to whether $NRN > k$ or $NRN \leq k$, with k taking all possible values of NRN in the data. Secondly, for each k, we find the best-fitted logistic regression model using forward selection and BMA. We then record the values of χ^2, and also those of AIC. Finally, the value of *NRN_T* is identified as that which corresponds to the highest value of χ^2, or the lowest value of AIC. A similar procedure is applied to the case using Cox regression model.

To determine the optimal number of lymph nodes to be removed (*NRN_T*) for maximal survival using logistic and Cox regression analyses, the procedure are summarized as follow:

Step 1. The data set are categorized according to all thresholds of NRN involved.

Step 2. Each NRN is treated as a categorical variable such that $NRN(k) = 0$ or 1, according to whether $NRN > k$ or $NRN \leq k$, with k taking all possible values of NRN in the data.

Step 3. Use the best-fitted logistic regression model (under forward selection procedure), which is obtained in section 4.3.2.

Step 4. The values of χ^2 and AIC are recorded from logistic regression models for all thresholds of nodes to be removed.

Step 5. Choose the *NRN_T*, which has the highest value of χ^2 or the lowest value of AIC.

Step 6. The Step 1 – 5 are repeated using the best-fitted logistic regression model (under BMA procedure) we obtain in section 4.3.2.

Step 7. The complete procedure (Step 1 – 6) is repeated using the best-fitted Cox regression models we obtain in the previous section.

4.4 Results

Figure 4.1 presents the distributions of NRN for the group as a whole, patients with N_{neg} only and patients with N_{pos} only. We notice that the same patterns are observed for both groups such that more than 30% of patients have at least 11 to 15 nodes removed during axillary dissection. This number seems to be larger than those found in the literature. This study hopes to provide a better guideline for deciding the sufficient number of lymph node to be removed that give maximum differential in survival of local breast cancer patients.

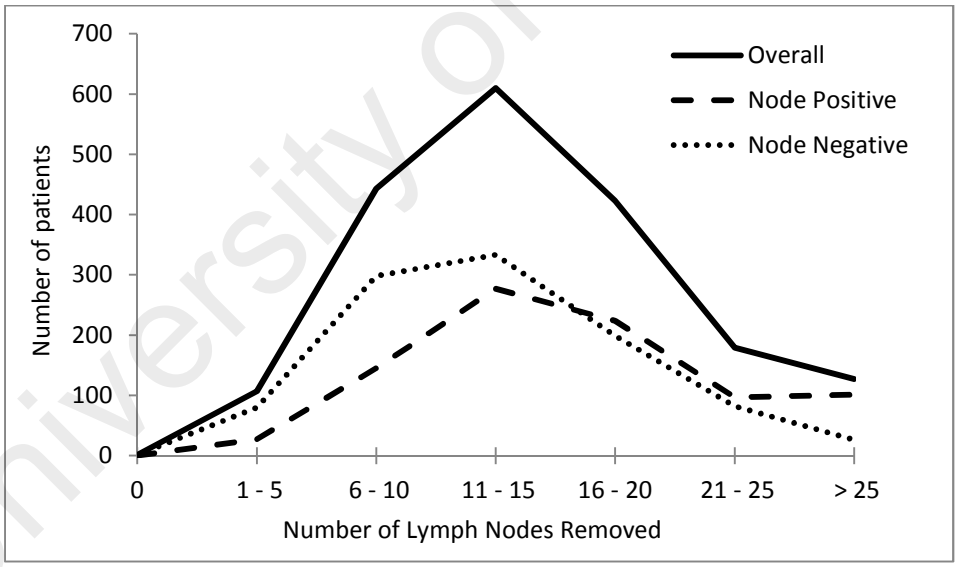


Figure 4.1: Number of lymph nodes removed for the group as a whole and for each group of patients (node positive and node negative patients)

As illustrated in Table 4.1, the chi-square test of independence reveals that the survival status of patients with N_{neg} were related to age at diagnosis and tumor stage, while the survival status of patients with N_{pos} were associated with LVI, tumor size, tumor stage and NRN.

Table 4.1: Significant prognostic factors of survival for node positive and node negative patients using chi-square test

Prognostic factors	Node negative patients		Node positive patients	
	χ^2	*p*-value	χ^2	*p*-value
Age	6.9297	0.0313*	1.4656	0.4806
Race	1.2572	0.7393	4.5286	0.2098
LVI	0.5725	0.4493	5.1158	0.0237*
Tumor Size	70.8077	0.1395	101.5575	0.0001*
Tumor Stage	33.5602	<.0001*	80.3832	<.0001*
NRN	42.4062	0.2492	66.9182	0.0453*

* Significant

In addition, the results of the Wilcoxon test in Table 4.2 indicate that there was a significant difference in the survivor functions for each level of age at diagnosis, tumor size, tumor stage and NRN for patients with N_{neg}. However, for patients with N_{pos}, the significant differences not only occur for each level of tumor size, tumor stage and NRN, as presented in the case N_{neg}, but also for each level of LVI.

Table 4.2: Significant difference between the survival functions for node positive and node negative patients in all prognostic factors of survival using Wilcoxon test

Prognostic factors	Node negative patients		Node positive patients	
	χ^2	*p*-value	χ^2	*p*-value
Age	6.5	0.038*	3.2	0.2
Race	1	0.797	5.8	0.124
LVI	0.8	0.373	8.3	0.00388*
Tumor Size	192	0*	357	0*
Tumor Stage	77.7	0*	152	0*
NRN	58.2	0.0147*	274	0*

* Significant

The results of chi square and Wilcoxon tests are not consistent as they are different in criteria of testing. The former investigates the significant association between prognostic factors and survival status, while the latter is used to compare the hazard functions of the two or more levels at each observed time event.

Using forward selection procedure, for patients with N_{neg}, the best-fitted logistic and Cox regression models indicate that tumor size was the important prognosis factor of survival, except for the case which *k* was 30 in the best-fitted Cox regression model.

In this particular case, the best-fitted Cox regression model show that not only tumor size but also tumor stage as the important prognosis factors of survival. As tumor size was the important prognosis factor of survival, for each increase in one unit of tumor size, the estimated odds of survival increase by 1.14 based on the best-fitted logistic model, while the hazard increases by 13% based on the best-fitted Cox model. On the other hand, for patients with N_{pos}, both best-fitted logistic and Cox regression models point out that the important prognosis factor were tumor size and tumor stage. For each increase in one unit of tumor size, the estimated odds of survival increase by 1.06 based on the best-fitted logistic model, while the hazard increases by 3% based on the best-fitted Cox model. Meanwhile, for each increase in one level of stage, the estimated odds of survival increase by 2.58 based on the best-fitted logistic model, while the hazard increases by 134% based on the best-fitted Cox model.

Figure 4.2 gives the plots for patients with N_{neg}, which use the logistic and Cox regressions with AIC and Wald's statistics as the decision rule. Using logistic regression, the plots recommend that the *NRN_T* for patients with N_{neg} was 19 nodes, where its value of χ^2 was the highest, or its value of AIC was the lowest. The corresponding values of χ^2 and AIC were 23.75 and 794.33 respectively. Using Cox regression, the plots suggest that the *NRN_T* for patients with N_{neg} was 19 nodes based on the value of χ^2 while 30 nodes based on the value of AIC. The corresponding values of χ^2 and AIC were 35.40 and 1846.78.

Meanwhile, Figure 4.3 gives the plots for patients with N_{pos}. From these plots, it is seen that the recommended number of nodes to be removed was 10 as it gave the highest value of χ^2, or the lowest value of AIC. The corresponding values of χ^2 and AIC were 81.82 and 1084.08 for the logistic regression model respectively, while 130.21 and 4398.73 for the Cox regression model. That is, *NRN_T* for patients with N_{pos} was 10.

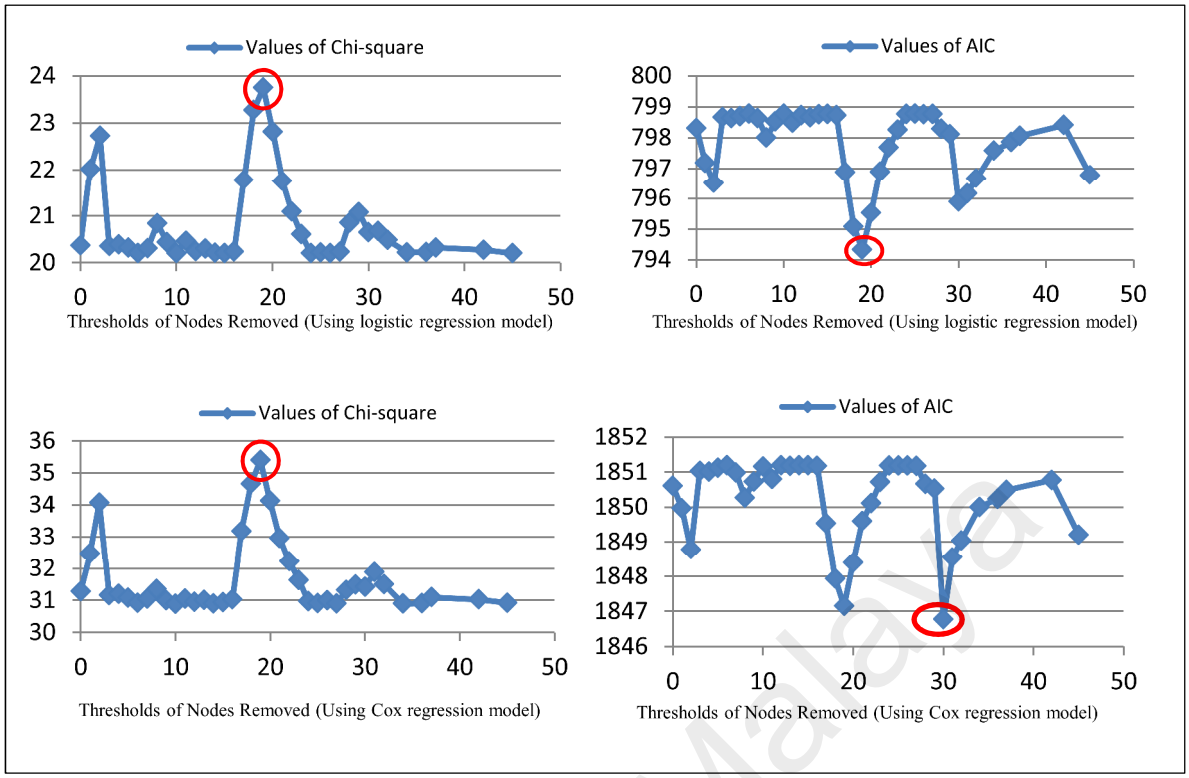


Figure 4.2: The optimal threshold for patients with N_{neg} using forward selection procedure

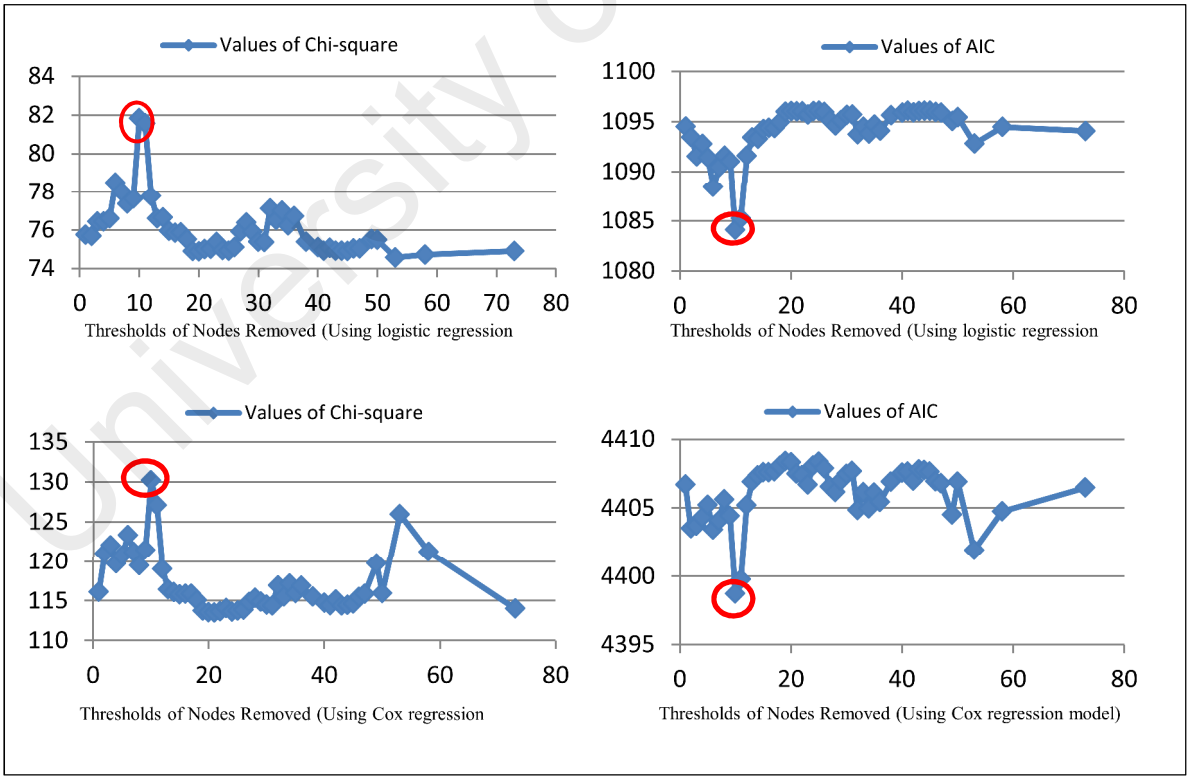


Figure 4.3: The optimal threshold for patients with N_{pos} using forward selection procedure

We notice that most of the results using AIC are consistent with those using χ^2 values for both logistic and Cox regression models using forward selection procedure, except cases for patients with N_{neg} using Cox regression model.

On the other hand, using the BMA procedure, the best-fitted logistic and Cox regression models show that the important prognosis factor was tumor size for patients with N_{neg} while tumor stage for patients with N_{pos}. Similar to those cases using the forward selection procedure, Figure 4.4 gives the plots for patients with N_{neg}, which use the logistic and Cox regressions with AIC and Wald's statistics as the decision rule. These plots recommend that the *NRN_T* for patients with N_{neg} was 19 nodes in where its value of χ^2 was the highest, or its value of AIC was the lowest. The corresponding values of χ^2 and AIC were 23.75 and 794.33 for the logistic regression model respectively, while 35.40 and 1847.15 for the Cox regression model.

Meanwhile, Figure 4.5 gives the plots for patients with N_{pos}. From these plots, it is seen that the recommended number of nodes to be removed was 10 as it gave the highest value of χ^2, or the lowest value of AIC. The corresponding values of χ^2 and AIC were 79.41 and 1086.22 for the logistic regression model respectively, while 121.03 and 4400.62 for the Cox regression model. That is, *NRN_T* for patients with N_{pos} was 10.

We also observe that the results using AIC are consistent with those using χ^2 values for both logistic and Cox regression models using the BMA procedure. It is recommended that 19 nodes should be removed for patient with N_{neg} and 10 nodes for patient with N_{pos}.

Furthermore, we also look at the survival characteristics of patients in groups categorized by the optimal threshold NRN. The analysis is conducted based on the results of the BMA procedure due to the consistency of results using both measures. For patient with N_{neg}, only size of tumor was identified as significant prognosis factor. The

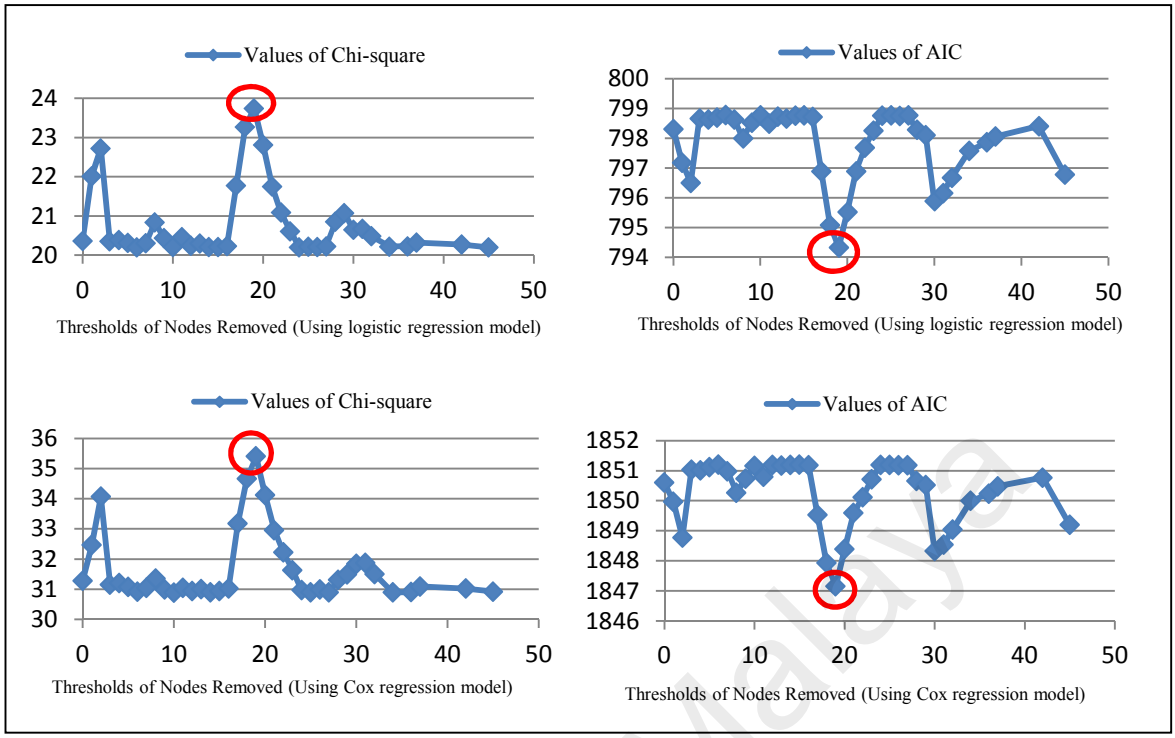


Figure 4.4: The optimal threshold for patients with N_{neg} using the BMA procedure

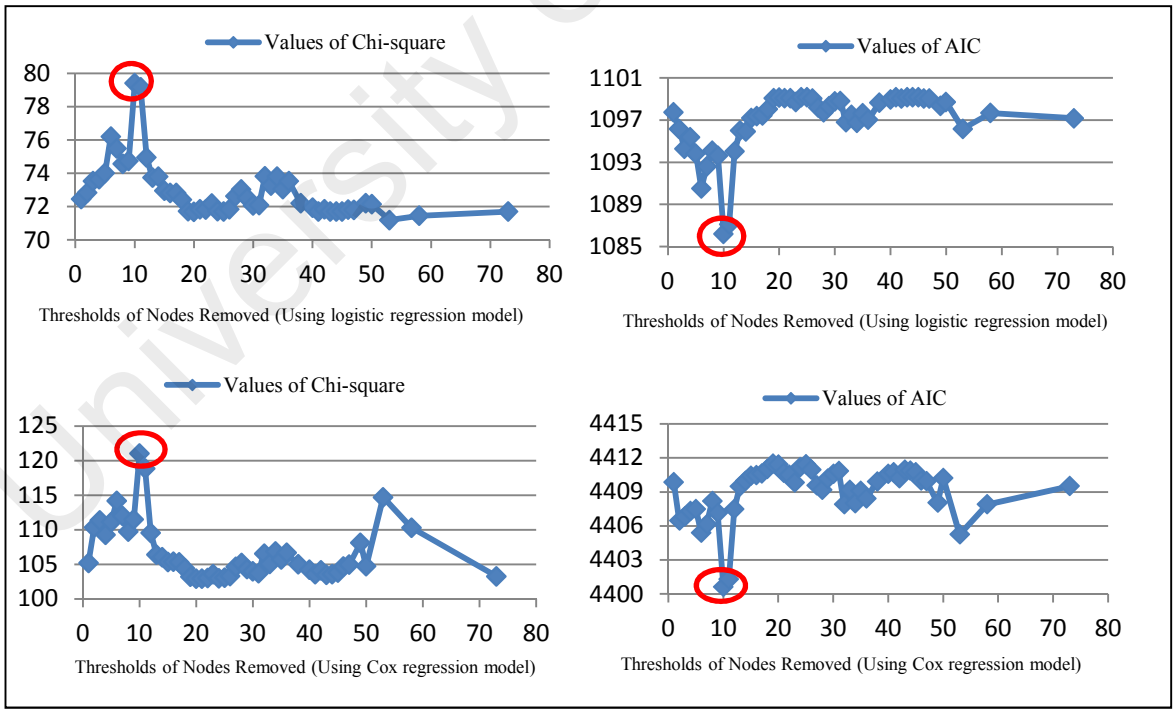


Figure 4.5: The optimal threshold for patients with N_{pos} using the BMA procedure

Kaplan-Meier plots of the survival in patients who had 19 or more nodes removed to those with less than 19 nodes removed were presented for the whole data (Figure 4.6) and for each level of tumor size as defined in Table 1 (Figure 4.7 – 4.9). In each situation, survival was significantly better when 19 or more nodes were removed for the case of tumor size less than 5 mm.

As for patient with N_{pos}, stage was identified as the only important prognosis factor. From the Kaplan-Meier plots given for the group as a whole (Figure 4.10) and for each stage of disease (Figure 4.11 – 4.13), we find that the survival in patients who had 10 or more nodes removed was significantly better as compared to those with less than 10 nodes removed for those with stage 3 cancer.

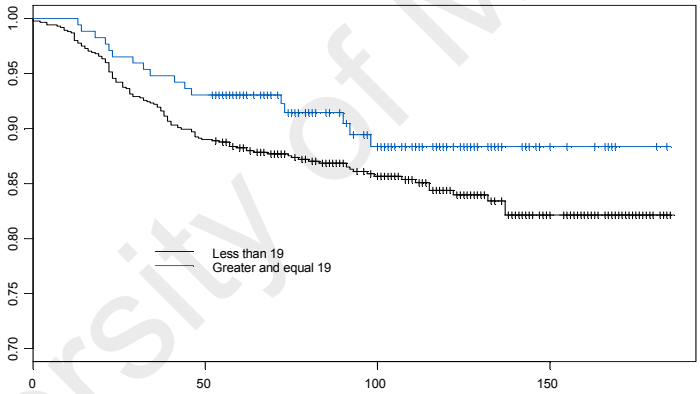


Figure 4.6: Kaplan Meier estimates of survival for all patients with N_{neg}
y-axis is survival probability, *x*-axis is observation
(Log-rank test: *p*-value = 0.12)

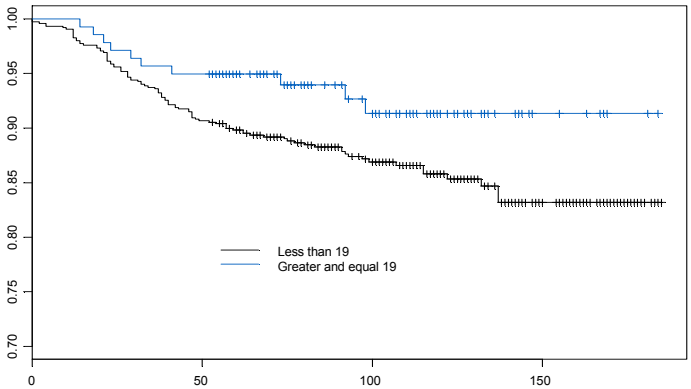


Figure 4.7: Kaplan Meier estimates of survival for N_{neg} patients with tumor size < 5mm
y-axis is survival probability, *x*-axis is observation
(Log-rank test: *p*= 0.0618)

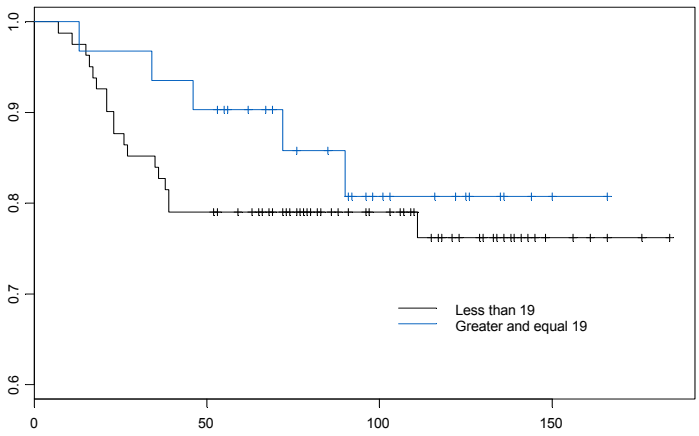


Figure 4.8: Kaplan Meier estimates of survival for N_{neg} patients with 5mm $\leq$ tumor size $< 10mm$
y-axis is survival probability, x-axis is observation
(Log-rank test: p-value = 0.443)

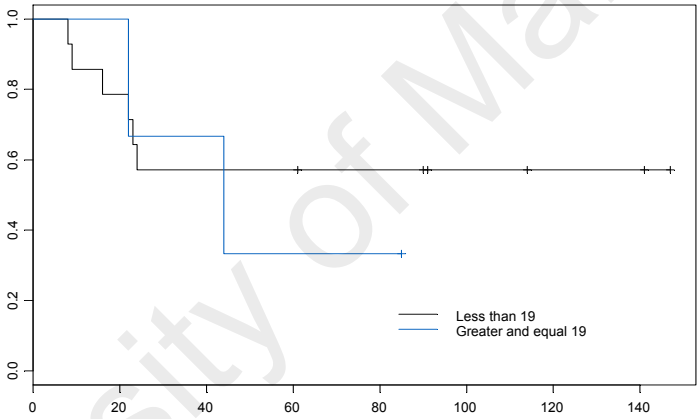


Figure 4.9: Kaplan Meier estimates of survival for N_{neg} patients of with tumor size $> 10mm$
y-axis is survival probability, x-axis is observation
(Log-rank test: p-value = 0.677)

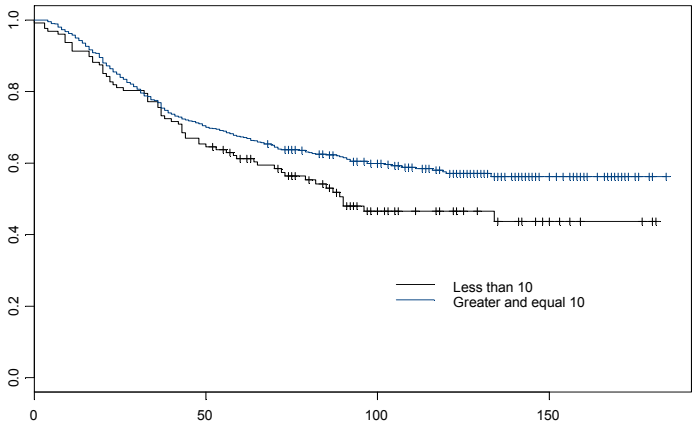


Figure 4.10: Kaplan Meier estimates of survival for all patients with N_{pos}
y-axis is survival probability, x-axis is observation
(Log-rank test: p-value = 0.0412)

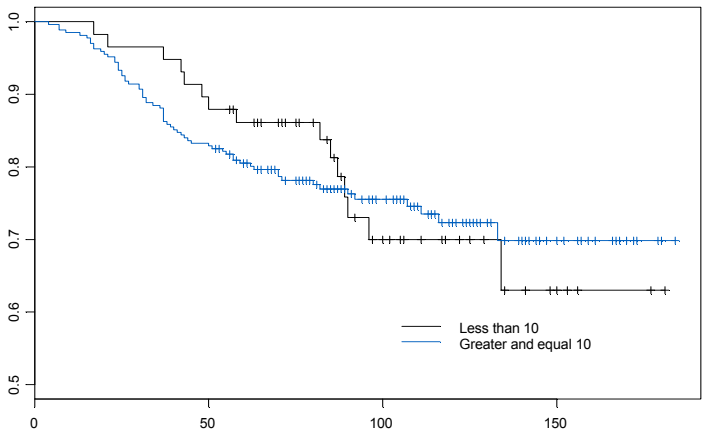


Figure 4.11: Kaplan Meier estimates of survival for patients with N_{pos} of stage 2
y-axis is survival probability, *x*-axis is observation
(Log-rank test: *p*-value = 0.908)

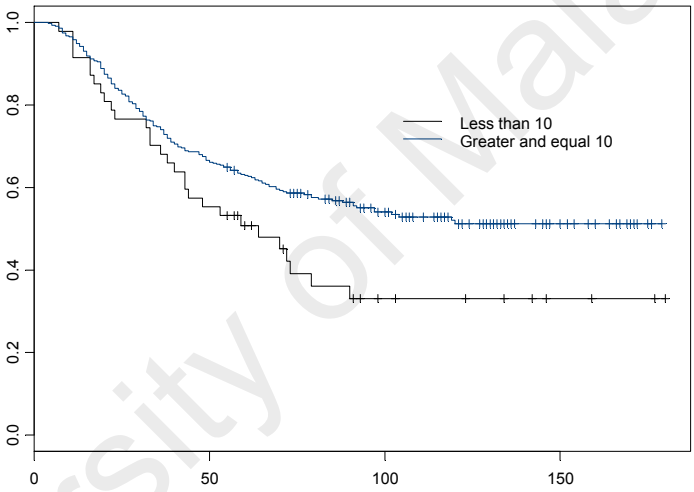


Figure 4.12: Kaplan Meier estimates of survival for patients with N_{pos} of stage 3
y-axis is survival probability, *x*-axis is observation
(Log-rank test: *p*-value= 0.0249)

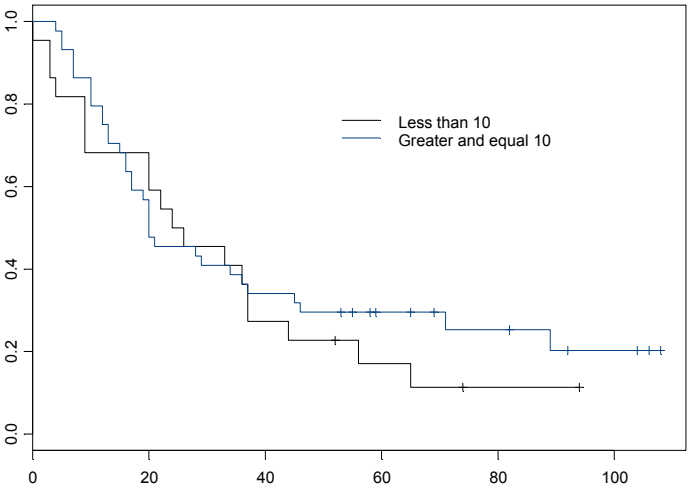


Figure 4.13: Kaplan Meier estimates of survival for patients with N_{pos} of stage 4
y-axis is survival probability, *x*-axis is observation
(Log-rank test: *p*-value = 0.486)

4.5 Discussion

The significant influence of the number of lymph nodes removed in the survival of breast cancer patients has been proven in the literature. Sosa et al. (1998) carried out a bivariate analysis using the Wilcoxon-Gehan statistic and they found that increasing the number of lymph nodes removed had improved the overall survival, disease-free survival and recurrence in breast cancer patients. Using the Cox regression model, van der Wal et al. (2002) showed that total number of lymph nodes removed was one of the significant prognostic factors of survival for node-negative and node-positive patients. Krag and Single (2003) investigated the survival according to age and a cutoff of 10 nodes removed in node-negative and 1-3 node positive women with breast cancer. They concluded that a larger number of nodes removed were associated with a better survival experience for each of the analysis groups. In our study, the number of lymph nodes removed is identified as an independent prognostic factor of survival, similarly treated in most of the previous studies.

Due to the importance of the number of nodes removed in the management of breast cancer, the optimal number of nodes to be removed for maximizing the survival then becomes the main concern for most of medical researchers. In a review of literature, some recommendations regarding the suitable threshold as the optimal cutoff point for survival benefit have been given as in the previous section. These numbers vary due to the difference in the surgeon's philosophy and technique, the thoroughness of pathologist reporting, and patient anatomy.

To obtain the optimal cutoff point that give maximum differential in survival of esophageal cancer patients, Peyre et al. (2008) calculated the χ^2 scores for the Cox regression model and the R^2 values for the logistic model for thresholds ranging from 1 to 60 nodes. Both regression models indicated that the best threshold ranges from 23 to 29 nodes removed. Meanwhile, Rizk et al. (2010) determined the optimum number of

nodes that should be resected to maximize 5-year survival of esophageal cancer patients by random forest multivariable regression. Maximum 5-year survival was modulated by tumor (T) classification. They recommended the resecting of 10 nodes for pT1, 20 for pT2, and ≥30 for pT3/T4. Based on the mortality rates using Cox regression model, Groth et al. (2010) reported that esophageal cancer patients should have at least 30 lymph nodes removed pathologically as part of esophageal resection in order to maximize all-cause and cancer-specific survival. On the other hand, based on the Cox regression model, Kuru (2006) stated that axillary lymph node dissection with a minimum threshold of 15 was associated with increased survival for node positive breast cancer patients.

In our study, we use Akaike Information Criterion (AIC) as a measure in determining the optimal number of lymph nodes that give maximum differential in survival of local breast cancer patients. While sentinel node biopsy is a standard of care for the management of node negative patients, this study suggests that a minimum of 10 nodes for patients with N_{pos} may be sufficient for maximizing the survival in a setting where sentinel node is not available. On the other hand, a minimum of either 19 or 30 nodes is recommended for patients with N_{neg} when two different model selection methods are applied. These numbers are obtained from both logistic and Cox regression modeling under procedures of forward selection and BMA, and based on both χ^2 and AIC values. For instance, when NRN=10, the value of χ^2 is the highest or the value of AIC is the lowest. These results are quite close to the findings in the literature, where they depend on their corresponding population of patients used in the study. It is worth pointing out that the results are consistent for both measures when procedure of BMA is applied. This strongly suggests that AIC can be considered as an alternative choice to χ^2 using Bayesian framework in future related works.

4.6 Summary

The survival of breast cancer patients is associated with the number of lymph nodes removed. To maximize the outcome of surgical resection for breast cancer, lymphadenectomy that removes a minimum of 10 nodes need to be included for patients with N_{pos}, while at least 19 nodes are recommended for patients with N_{neg}. These finding are very important as a guideline in the surgical management of breast cancer patients.

CHAPTER 5: BREAST CANCER – OPTIMAL NUMBER OF LYMPH NODES REMOVED FOR ADEQUATE NODAL STAGING

5.1 Introduction

The number of lymph nodes removed is one of the most important prognostic factors for survival in breast cancer. Hence, it is imperative that accurate staging for the axilla is performed. Although axillary lymph node staging surgery has evolved towards sentinel node biopsy in early breast cancer, in low and middle income countries, where the majority of patients are symptomatic and in advanced stages of the disease, axillary dissection remains the gold standard in staging the axilla. In this chapter, we intend to investigate the association of the number of lymph nodes removed on nodal involvement and then to determine the optimal number of lymph nodes to be removed for a reliable staging of the axilla. Similar data set of breast cancer in Chapter 4 was used. The analysis consists of three stages. Firstly, the chi-square test of independence is performed to determine the significant association between factors and nodal involvement. Secondly, we find the best-fitted logistic model using forward selection and Bayesian model averaging (BMA) procedures applied to binary response. We only consider the best-fitted logistic model as the response of the logistic model is the probability of lymph nodes positive/involved which does not involve the information on survival status. Thirdly, the model is used to assess the prediction values of independent factors of nodal involvement at all thresholds of number of lymph nodes removed. In this study, we use not only the Wald statistic (χ^2) but also present the use of the Akaike Information Criterion (AIC) to determine the optimal number of lymph nodes to be removed for adequate nodal staging in breast cancer patients.

The study population consists of 1890 patients who underwent at least a level II axillary dissection: 1019 patients who did not have lymph nodes involvement (N_{neg}) and 871 patients with at least one or more axillary node involved (N_{pos}).

5.2 Data description

Similar data set of breast cancer from Chapter 4 was used for the analysis in this chapter. The background of breast cancer data can be found in Chapter 3 and Chapter 4. For this study, 1890 patients were grouped as a whole for the analysis. Variable N_{pos} was excluded from the analysis since it is dependent on lymph node status.

5.3 Statistical analysis

The methodology of this study is mostly similar to the analysis in Chapter 4. We carry out the chi-square test of independence, logistic regression modeling, forward selection and Bayesian model averaging (BMA) procedures for the analysis. Refer to Chapter 4 for their details.

We first use the chi-square test of independence to assess individually whether the factors and nodal involvement are independent of each other. This enables us to identify important factors associated with the lymph node involvement of the breast cancer patients. It is also our intention to identify important factors taking into account the contribution of other factors. This is achieved using the logistic regression modeling. In logistic regression analysis, the model predicts the probability of lymph nodes positive/involved for a given set of independent factors based on a generalized linear model for binary response using a logit link function. The best-fitted model then is obtained using forward selection and BMA procedures, where the details of these procedures had been discussed in section 4.3.

The focus of the study is to determine the optimal number of lymph nodes to be removed (*NRN_T**), for adequate nodal staging in breast cancer patients. This is carried out in several steps as similar as the procedures of determining the optimal number of lymph nodes to be removed for maximizing the survival of breast cancer patients. Firstly, we treat NRN as a categorical variable such that $NRN(k) = 0$ or 1, according to whether $NRN > k$ or $NRN \leq k$, with k taking all possible values of NRN in the data. Secondly, for each k, we find the best-fitted logistic regression model using forward selection and BMA. We then record the values of χ^2, and also those of AIC. Finally, the value of *NRN_T** is identified as that which corresponds to the highest value of χ^2, or the lowest value of AIC.

5.4 Results

As presented in Table 5.1, the results of the chi-square test of independence show that the nodal involvement of patients was associated with race, LVI, tumor size and NRN. Using forward selection and the BMA procedures, the best-fitted logistic models indicated that race, LVI and tumor size were the important predictors for nodal involvement, except for cases which *k* were 7 and 8 using the BMA procedure. In these particular cases, the best-fitted logistic models point out that only LVI and tumor size were the important predictors for nodal involvement. As race, LkVI and tumor size were the important predictors for nodal involvement, the estimated odds of survival increase by 1.20 for each changes in race. The estimated odds that a patient with positive lymphovascular invasion survives is 4.32 greater than the corresponding odds for a patient with negative lymphovascular invasion. Meanwhile, for each increase in one unit of tumor size, the estimated odds of survival increase by 1.21.

Table 5.1: Significant factors of nodal involvement for breast cancer patients using chi square test

Prognostic factors	χ^2	*p*-value
Age	4.5734	0.1016
Race	25.1088	<.0001*
LVI	253.4905	<.0001*
Tumor Size	276.0854	<.0001*
NRN	158.1551	<.0001*

Under forward selection and the BMA procedures, Figure 5.1 and 5.2 give the plots for the group as a whole, which use the logistic regression with AIC and Wald's statistics as the decision rule. These plots recommend that *NRN_T** for patients was 10 nodes in where its value of χ^2 was the highest, or its value of AIC was the lowest. For both procedures, the corresponding values of χ^2 and AIC were 324.57 and 2188.94, respectively.

We also notice that the results using AIC are consistent with those using χ^2 values for logistic regression model using both procedures.

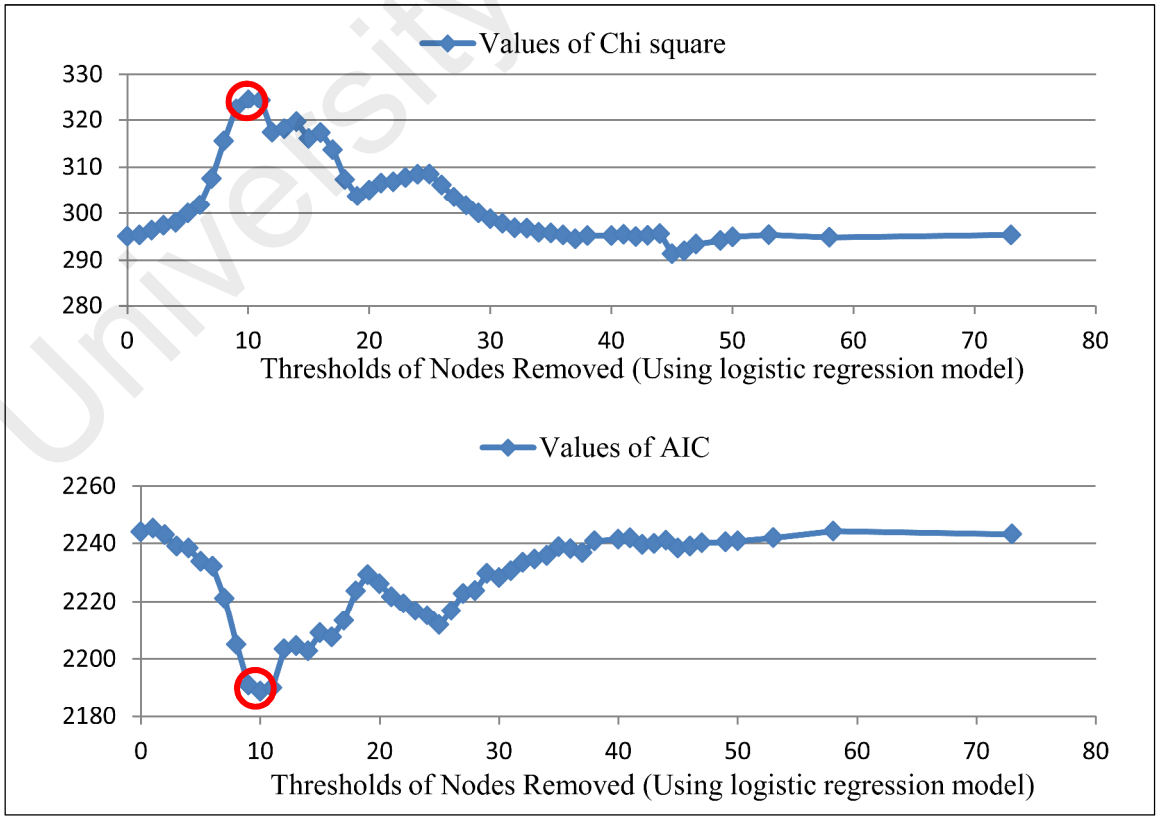


Figure 5.1: The optimal threshold for adequate nodal staging in breast cancer patients using forward selection procedure

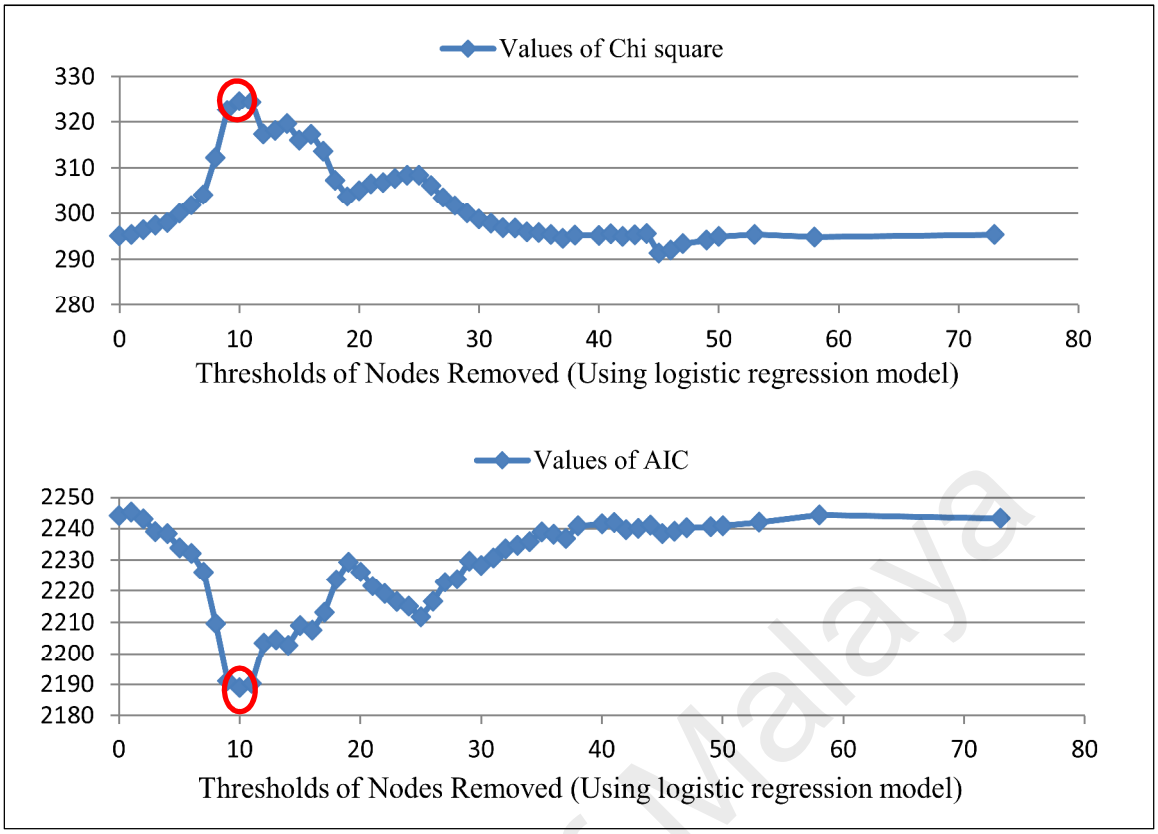


Figure 5.2: The optimal threshold for adequate nodal staging in breast cancer patients using the BMA procedure

5.5 Discussion

Axillary lymph node dissection is a standard surgical technique used in the staging and treatment of the axilla in node positive breast cancer. Previous studies provided substantial evidence that axillary lymph node dissection gave excellent local control of disease in the axilla, which may lead to improved overall survival (Atkins et al., 1972; Langlands et al., 1980; Cabanes et al., 1992; Sosa et al., 1998; Orr et al., 1999; Bembenek and Schlag, 2000; Krag and Single, 2003; Sanghani, 2009). Studies have found not only the importance of number of lymph nodes involved but also those removed. Lymph node ratio which takes into account number of nodes involved divided by number of nodes removed has been found to be an important prognostic factor (Taib et al. ;2008, 2011), and hence, highlighting the importance of the number of nodes removed in the management of breast cancer (Sosa et al., 1998; van der Wal, 2002;

Weir et al., 2002; Krag and Single, 2003; Saxena et al., 2012). Therefore, the optimal number of nodes to be removed for adequate staging has attracted a lot of attention.

Fisher et al. (1981) reported that the majority of histologic positive nodes breast cancer patients considered in their study have 1-3 involved nodes. The more nodes were removed, the more involved nodes were identified. On the other hand, the National Comprehensive Cancer Network (2014) in the US recommended a minimum threshold of 10 lymph nodes for staging, whereas Singletary et al. (2002) noted that the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, Sixth Edition (2002) required 6 axillary lymph nodes to be removed and examined. These numbers vary based on surgeon's philosophy and technique, the thoroughness of pathologist reporting, and patient anatomy. Notably, there is scarce information on the scientific basis of how these optimal numbers were obtained. Here, we aim to determine the optimal cut-off point for staging the axilla in breast cancer patients in Malaysia.

To investigate the optimal number of nodes to be removed to obtain the maximum differential in survival of esophageal cancer patients, Peyre et al. (2008) calculated the χ^2 values for the Cox regression model and the R^2 values for the logistic model for thresholds ranging from 1 to 60 nodes. Both regression models indicated that the best threshold ranged from 23 to 29 nodes removed. However, in this study, we propose the use of Akaike Information Criterion (AIC) as an alternative measure instead of χ^2 in determining the optimal number of lymph nodes that need to be removed for adequate nodal staging in local breast cancer patients.

The aim of this study then was to find the optimal number of lymph nodes to be removed for adequate nodal staging in breast cancer patients. This threshold value may provide guidance in the management of these patients. While sentinel node biopsy is a standard of care for the management of node negative patients, this study suggests that a minimum of 10 nodes for patients with N_{pos} may be sufficient for reliable staging of

lymph nodes in a setting where sentinel node is not available. This result is quite close to the findings in the literature, where it is all depend on the corresponding population of patients used in each study. It is worth pointing out the consistency of results using logistic regression model under forward selection and the BMA procedures. This strongly suggests that AIC can be considered as a good measure in future related works.

5.6 Summary

The nodal involvement of patients is related to race, LVI, tumor size and NRN. To reduce the risk of understaging for breast cancer, axillary lymph node dissection that removed a minimum of 10 nodes should be performed for breast cancer patients. This finding is very important as a guideline in the surgical management of breast cancer patients.

CHAPTER 6: LITERATURE REVIEW

6.1 Kinesiology studies

Kinesiology comes from the Greek word *kinesis*, which means motion. In the medical sciences, it is the name given to the study of muscles and the movement of the body, the mechanics of body movements. Dr. George Goodheart, an American chiropractor, the acknowledged founder of Applied Kinesiology in 1964, used the model of muscle testing to evaluate what he was doing chiropractically. The model of muscle testing he used was developed in the 1930s by the husband and wife team of Kendal and Kendal.

In neurology contacts, muscle testing is defined as "a means of testing the motor function of limbs". Muscle testing therefore was accepted as a valid technique and used extensively in orthopaedic medicine by physiotherapists, chiropractors and osteopaths. As part of the foundation of expanding the application of muscle testing, Dr Goodheart and his colleagues took on board work done by Drs Bennet and Chapman with regard to the lymphatic and vascular systems. They also looked at the subtle energy system as used within acupuncture. This then was the basis of muscle testing that was to develop and become known as Applied Kinesiology (AK). Applied Kinesiology was the name given by Dr George Goodheart, to the system of applying muscle testing diagnostically and therapeutically to different aspects of health care.

In the early 1970's, Dr. John Thie recognized the need to educate the public in many of the self-help techniques within AK and this educational program then became known as Touch For Health (TFH), a program for the layman and is taught and used throughout the world. Brian Butler then expanded further Systematic Kinesiology from the concepts of TFH in 1982 (Source: Association of Systematic Kinesiology in Ireland).

6.2 Kinesiotape

Kinesiotape tape (KT) is an elastic tape that had been introduced in the 1970 to mimic human skin in elasticity and thickness for correcting of muscle function (Kase et. al., 2003). Combination of its unique design and its peculiar taping methods has attracted many to consider the tape as an alternative therapy to alleviate musculoskeletal symptoms, and to enhance sports performance. It has expanded over the years and multiple disciplines as well as the public population have begun seeing its advantages over other taping methods. Physical therapists, physical therapists assistants, occupational therapists, certified occupational therapy assistants, chiropractors, massage therapists, athletic trainers, and both professional and amateur athletes are among those who use KT. This useful tape has been featured by many athletes in Olympic Games; including Phil Dalhausser, Kerri Walsh, and Patty Schnyder (Tabott, 2008). Dancers and dance trainers are another group that have seen and promoted the benefits KT can provide (Wozny, 2009).

6.3 Crossover design

According to Jones and Kenward (1989), in a crossover trial, each experiment subject receives two or more different treatments. The order in which each subject receives the treatments depends on the particular design chosen for the trial. The simplest design is the 2×2 design. In this design, each subject receives two different treatments, which are labeled as A and B. Half the subjects receive treatment A first and then, after a suitably chosen period of time, cross over to treatment B. The remaining subjects receive treatment B first and then cross over to treatment A. Jones and Kenward (1989) discuss the theory, methods and practice for crossover trials in details.

Crossover trials are widely used in medical and pharmaceutical industry. These trials are applicable in most situations such as veterinary research, animal feeding trials,

sport science and psychological experiments. For example, Parkes (1982) has described the use of a crossover trial to investigate occupational stress, and Raghavarao (1989) has described a potential application of crossover trials in non-pharmaceutical industry. In the literature, the use of crossover design for clinical trials also has been discussed extensively by Brown (1980), Huitson et al. (1982), Senn (1993), and Jones and Kenward (2003).

6.4 Outlier detection in 2 × 2 crossover design

In the standard 2 × 2 crossover design, we assume that there are two different groups of subjects. Each group receives the two treatments in a different order and the trial is to last for two treatment periods, with the order of treatments reversed in the second period. A common problem in crossover trials is the occurrence of extremely large or small observations. These extraordinary observations are called outliers and they may influence the conclusion drawn from the data set. An outlier is a data point which is significantly different from the remaining data (Aggarwal, 2013). Hawkins (1980) formally defined the concept of outlier as an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism.

Chow and Tse (1990) proposed two procedures based on Cook's likelihood distance and the estimated distance for the detection of outliers in crossover studies. Liu and Weng (1991) carried out procedures based on Hotelling T^2 statistics and residuals for the same purpose. Wang and Chow (2003) presented a general test procedure based on a mean-shift model. Furthermore, Ramsay and Elkum (2005) compared different outlier detection methods proposed by Chow and Tse (1990) and Liu and Weng (1991) via simulation study. They concluded that the estimated distance test performs better than other tests. Most recently, Karasoy and Daghan (2012) applied these existing

methods to a real data set in order to investigate outliers. In crossover studies, Enachescu and Enachescu (2009) initially used principal components for the identification of outliers. Meanwhile, Singh et al. (2014) provided details regarding a studentized residual test and the Lund test for identification of outlier subjects. It is therefore important that methods of identifying outliers in 2×2 crossover design are developed for proper handling of the data in studies. These methods are usually graphical or numerical.

6.5 Summary

It is of interest to look at the problems of investigating the outlier detection of 2×2 crossover design using Bayesian and non-Bayesian frameworks. We have presented the literature review on different methods as solutions for the problems we considered. This will be explored further in subsequent chapters.

CHAPTER 7: UMMC KINESIOLOGY DATA – DESCRIPTION

7.1 Introduction

This chapter presents a real data set from a study of kinesiology which is obtained from University of Malaya Medical Centre (UMMC) Sport Medicine Clinic. A two period crossover and randomized placebo-controlled trial of AB (treatment followed by sham taping)/BA (placebo followed by treatment taping) design is conducted. There are 77 subjects, from eighty one subjects volunteered, completed the study (AB = 37, BA = 40) which observed a minimum washout period of one week. Pre and post measurements of peak oxygen consumption or VO_2 peak (in ml/kg/min) recorded from a six-minute Astrand submaximal cycling exercise test conducted at least one week apart. Since there is none of the musculoskeletal outcome measures demonstrated convincing association with kinesiotape (KT) use, we therefore propose to investigate the effect of KT on the VO_2 peak. VO_2 peak is mainly used to gauge cardiorespiratory fitness of an individual. Theoretically, KT is expected to improve VO_2 peak. Even though it is a feasible theory, no study or investigation has explored the effect of kinesiotape on measurements of VO_2 peak thus far.

7.2 Subject recruitment and procedure

At the beginning of the study, a total of 81 subjects are deemed feasible by researchers. The study has a power of 90% to detect a standard deviation of 0.4 by assuming 10% attrition rate with 72 subjects completing the study. The subjects included in the study are volunteers who are between 18 – 25 years of age, any gender, free from cardiorespiratory and musculoskeletal disorders such as knee, hip or ankle pain that may preclude cycling testing. On the other hand, those with known allergies to

taping material unfit to perform exercise, or reluctant to follow the testing protocol, are excluded from the study.

The subjects involved are advised to refrain from strenuous activity the day before and avoid consumption of caffeine at least six hours prior to the test. Their demographic data and information of anthropometry are recorded before starting the testing. It is important that subjects are weighed in sport attire without their shoes and at least one hour after light meals in order to ensure the accuracy of the information recorded. Heart rate monitoring device, that is Polar Wearlink Coded, then is attached to the subjects throughout the testing session. The subjects have to take a rest for 5 – 10 minutes to make sure their heart rate reaches steady state. After familiar with the cycle ergometer (Monark 939E), the subjects proceed to start the cycling protocol and maintain a 55 – 65 cadences per minute pacing. The peak oxygen consumption (VO_2 peak) measurements finally are obtained after 6 minutes submaximal Astrand cycling protocol.

The study then is proceed by the taping on the subjects' bilateral quadriceps muscles either according to Dr. Kase's recommended Kinesio Tex®method(6); or a sham taping method following random assignment of sequence. The researcher in blocks of 10 conduct the randomization of the study using 1:1 sequence assignment in accordance to the order of subject enrollment into the studies. The subjects are requested to repeat the 6 minutes submaximal Astrand cycling after a lapse of at least one hour from the time of taping.

At the early stage of the analysis, graphical plots of the data are constructed for description and comparison purposes. A linear model analysis of variance (ANOVA) for the 2 x 2 crossover design is then used to determine if there are any carryover, direct treatments or periods effects in the data. Analysis of residuals is also conducted to check the assumptions underlaying the ANOVA model used in this study.

7.3 Data summary

As presented in Figure 7.1, a total of 81 university students (Age: 21.0 ± 1.0; Male = 24, Female = 57) volunteered to be recruited as the subjects of the study. However, only 77 subjects completed the tests (Age: 21.1 ± 1.0; Male = 24, Female= 53) with washout period = 17.8 ± 17.7 days, VO_2 peak of baseline 1 = 44.04 ± 10.40 and VO_2 peak of baseline 2 = 45.93 ± 8.37, while another 4 subjects are excluded from the study.

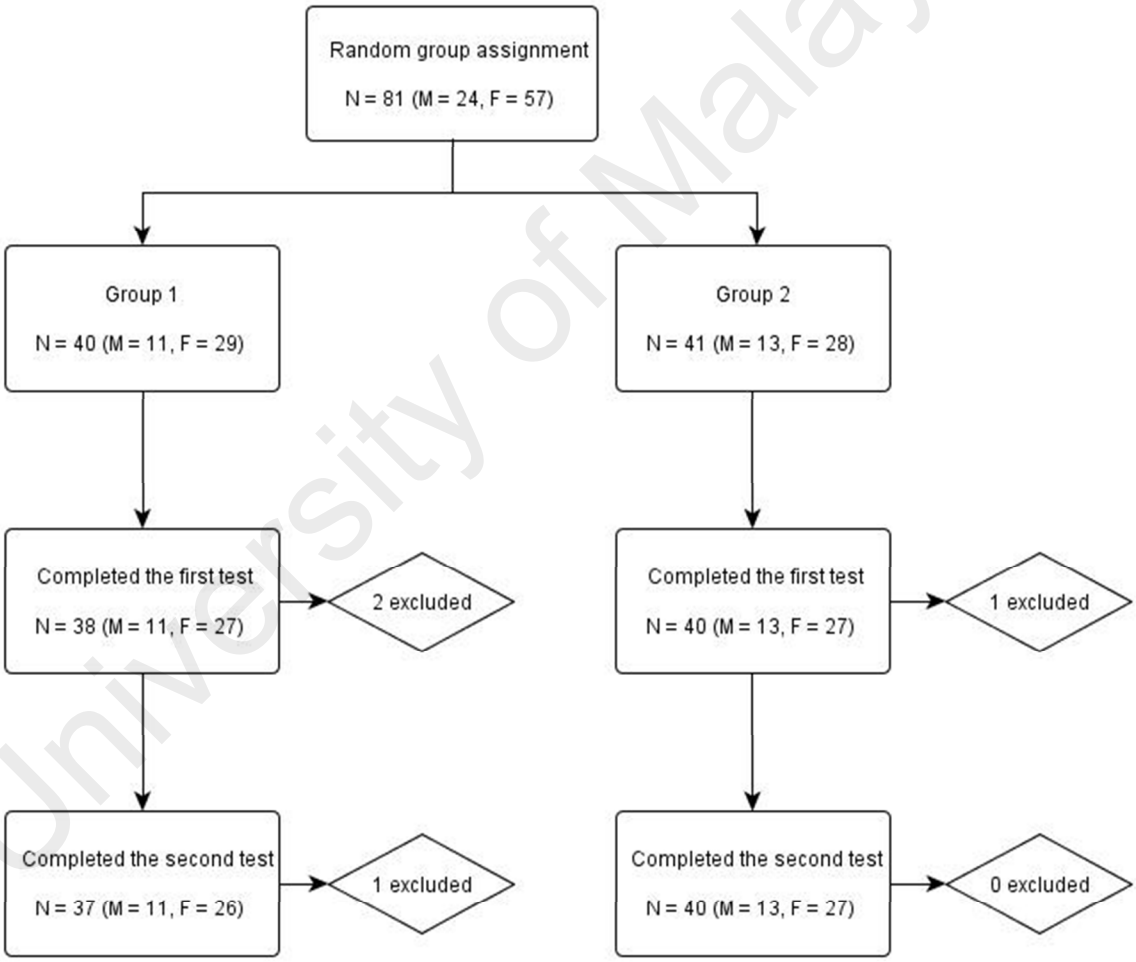


Figure 7.1: Flow chart containing details of number of subjects in each stage

Furthermore, Figure 7.2 illustrates the pre and post plot for Group 1 and 2, where 1A, 1B, 2A and 2B are the means of the VO_2 peak measurement in terms of the group and treatment they currently received at that period. For instance, 1A implies that Group 1 was receiving treatment A in period 1. For both groups, there is no much difference between the pre and post measurement of VO_2 peak in the period 1, but there is a distinct drop between them in the period 2.

Besides, Table 7.1 shows the group-by-period means of the VO_2 peak measurement while Figure 7.3a and Figure 7.3b present its corresponding plots. We notice that the pattern of Figure 7.3a and Figure 7.3b are similar. Despite of the periods, Figure 7.3b suggests that treatment B is superior to treatment A since the means of VO_2 peak for treatment B for both groups are higher than the ones for treatment A.

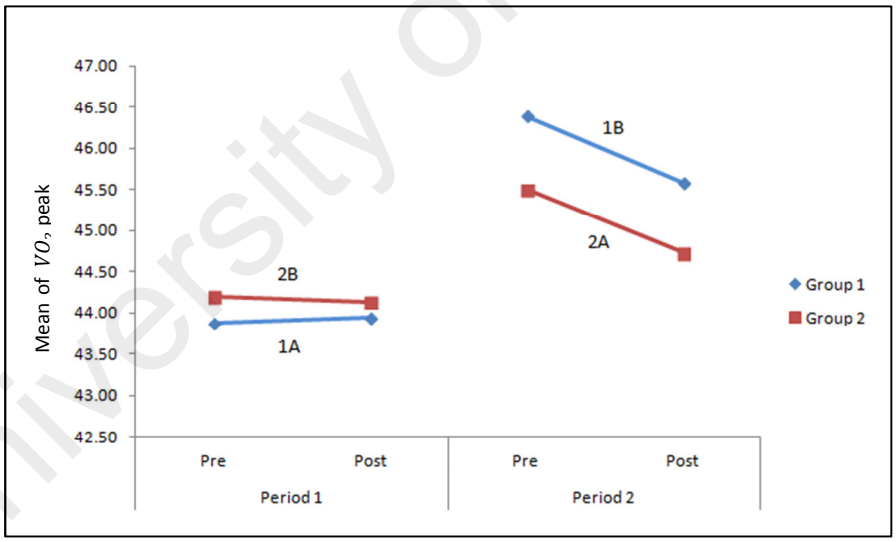
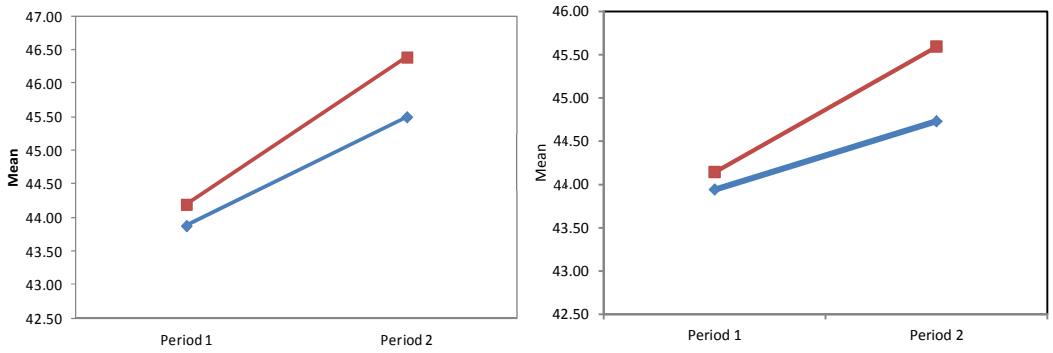


Figure 7.2: Pre and post plot

Table 7.1: Group-by-period means

	Period 1	**Period 2**	**Mean**
Group 1	43.94	45.59	44.76
Group 2	44.14	44.73	44.44
Mean	44.04	45.14	44.59



a) using only the pre measurements b) using only the post measurements

Figure 7.3: Groups-by-periods plots

For comparing the groups, we join the outermost points of each group and obtain the plot of differences against totals for the data and its convex hull, as illustrated in Figure 7.4. When comparing the groups in the horizontal direction (total axis), it suggests that there is no carryover effect in the data as there is no clear separation of the groups. On the other hand, as comparing the groups in the vertical direction (difference axis), the overlap suggests that there is no direct treatment effect in the data. From Figure 7.4, one might suspect the appearance of outliers in data since there are 2 subjects of Group 1, which stand out, have either an unusually large difference or an unusually large total.

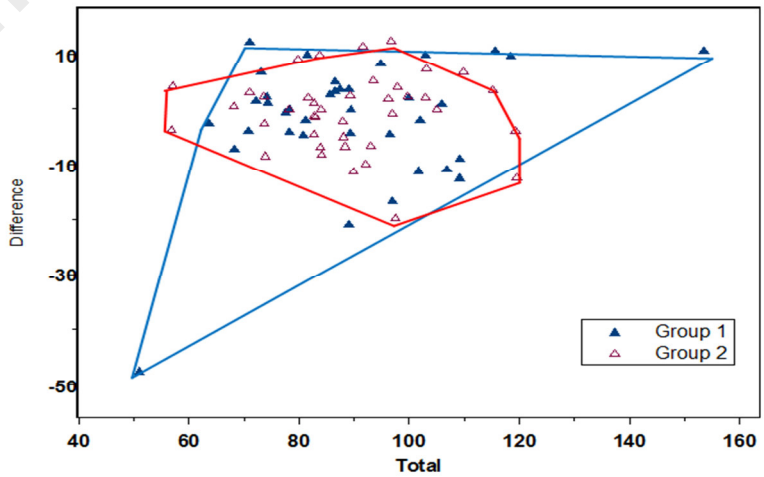


Figure 7.4: Plot of differences vs. total of the VO_2 peak for both periods

Table 7.2 presents the ANOVA result for the data using the full crossover model, which is discussed in details in Chapter 8. It is observed that there is significantly no carryover, direct treatment and period effects for the data, where their corresponding *p*-values are 0.78, 0.07 and 0.20. The tests are arbitrarily separated by a minimum duration of 1 week to minimize the inherent variation of VO_2 peak with time and training. Meanwhile, this will also allow enough time for the subjects to recover from the previous testing and time to negate the effects of the previous taping. In the analysis of residuals, Figure 7.5 – 7.7 suggest that the assumptions used for the model also violated which may due to outlier. Investigation of the occurrence of outlier will be investigated in the next two chapters.

Table 7.2: ANOVA result for the data

Source	df	SS	MS	F	P(F)
Btw subjects:					
Carryover	1	10.94	10.9372	0.08005148	0.7780071
B-S residual	75	10247.00	136.6267		
Within subjects:					
Periods	1	132.563	132.5630	3.289442	0.0737268
Treatments	1	65.878	65.8777	1.634702	0.2049962
W-S residual	75	3022.464	40.2995		
Total	153	13478.845			

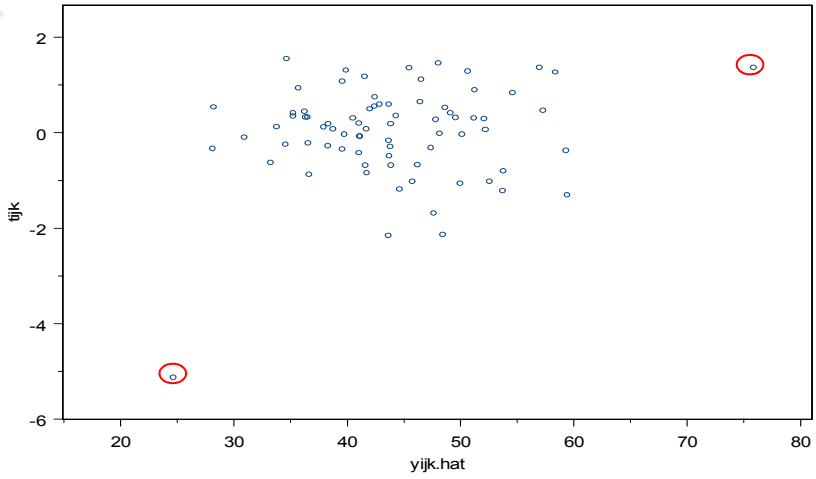


Figure 7.5: Plot of studentized within-subject residuals vs. fitted values of the data

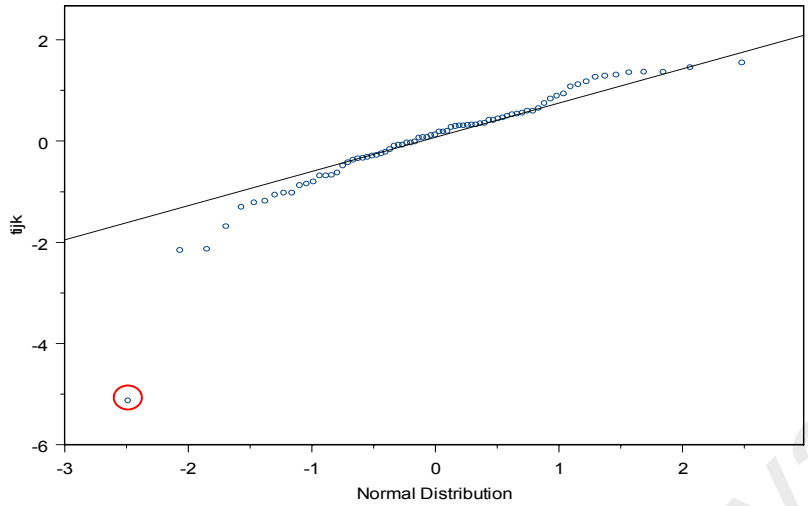


Figure 7.6: Normal probability plot of studentized within-subject residuals for the data (Shapiro-Wilk normality test, *p*-value = 0.00)

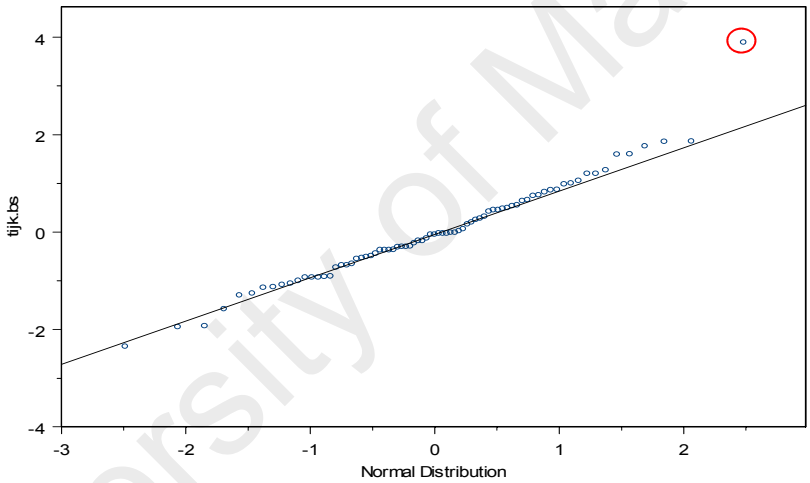


Figure 7.7: Normal probability plot of studentized between-subject residuals for the data (Shapiro-Wilk normality test, *p*-value = 0.06)

7.4 Summary

For the kinesiology data in this study, we consider 77 measurements of peak oxygen consumption or VO_2 peak (in ml/kg/min) recorded from a six-minute Astrand submaximal cycling exercise test conducted at least one week apart. Regardless of the taping method, it is observed that kinesiotape can possibly improve VO_2 peak, both on immediate application and the effect can be augmented with repeated use even after a week.

CHAPTER 8: KINESIOLOGY STUDY – OUTLIER DETECTION IN 2×2 CROSSOVER DESIGN USING NON-BAYESIAN FRAMEWORK

8.1 Introduction

In this chapter, we discuss methods for outlier detection in standard 2×2 crossover studies. Two outlier detection procedures are carried out based on residual analysis. Under a simplified model of 2×2 crossover design, we present the classical calculation of studentized residual ($SR1$) and provide a new studentized residual using median absolute deviation ($SR2$) to identify possible outlying subjects. We shall denote the procedure using $SR1$ by $P(SR1)$ and that using $SR2$ by $P(SR2)$. We investigate the performances of both procedures via simulation. As an illustration, these procedures are applied to two real data sets from studies of bioavailability and kinesiology, the later is discussed in Chapter 7.

8.2 Outlier detection for the standard 2×2 crossover design

In the standard 2×2 crossover design, we assume that there are two different groups of subjects. Each group receives the two treatments in a different order and the trial is to last for two treatment periods, with the order of treatments reversed in the second period. A common problem in crossover trials is the occurrence of extremely large or small observations. These extraordinary observations are called outliers and they may influence the conclusion drawn from data set. It is therefore important that methods of identifying outliers in standard 2×2 crossover design are developed for proper handling of the outlier in the data. These methods are either in graphical or numerical form. In this study, we carry out the outlier detection based on residuals. Under a simplified model of 2×2 crossover design, we present a classical calculation

of studentized residual ($SR1$) and provide a new studentized residual using median absolute deviation ($SR2$). Suitable outlier tests can then be applied to the resulting sets of studentized residuals in order to detect the possible outliers in the study.

Let Y_{ijk} be the response of the kth subject in sequence i during period j under the $d[i,j]$th treatment, where $i,j = 1,2$; m_i is the size of group with treatment $d[i,j]$ and $k = 1,2,\dots,m_i$. Refer to Jones and Kenward (1989), the full model is

$$Y_{ijk} = \mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]} + S_{ik} + e_{ijk} \tag{8.1}$$

where μ is the overall mean, p_j the fixed effect of the jth period, $\tau_{d[i,j]}$ the fixed effect of the treatment administered in period j of sequence i, $\lambda_{d[i,j-1]}$ the fixed effect of the carryover of the treatment administered in period $j-1$ of sequence i where $\lambda_{[i,0]} = 0$, S_{ik} the random effect of the kth subject, and e_{ijk} the random error. The variance components $\{S_{ik}\}$ and $\{e_{ijk}\}$ are assumed to be independent and normally distributed with mean 0 and variances σ_s^2 and σ_e^2, respectively. We also consider here the crossover model used by Chow and Tse (1990):

$$Y_{ijk} = \mu + p_j + \tau_{d[i,j]} + S_{ik} + e_{ijk} \tag{8.2}$$

with $\sum_i \sum_j \tau_{d[i,j]} = 0$, $\sum_j p_j = 0$ and no carryover effect is assumed. According to Liu and Weng (1991), when no period effect is assumed, model (8.2) can be reduced as follows:

$$\begin{aligned} Y_{ijk} &= \mu + \tau_{d[i,j]} + S_{ik} + e_{ijk} \\ &= \alpha_{d[i,j]} + S_{ik} + e_{kij}, \end{aligned} \tag{8.3}$$

where $\alpha_{d[i,j]} = \mu + \tau_{d[i,j]}$ and $k = 1,2,\dots,m_i$ for $i = 1,2$.

Regulatory authorities generally do not allow the exclusion of outliers from the statistical analysis of 2×2 crossover design based solely on statistical criteria. However, if such a data set does contain outliers, then it might be of interest to present the results of analysis with and without outliers.

8.2.1 Outlier detection using $SR1$

For calculation of $SR1$, we refer to model (8.3) in the previous section. The repeated measurements on each subject are assumed to be independent, normally distributed random variables with equal variances. The residual r_{ijk} is then given by

$$r_{ijk} = \left(1 - \frac{1}{m_i}\right) Y_{ijk} - \left(\frac{1}{m_i}\right) \left[\left(\sum_{t=1}^{m_i} Y_{ijt}\right) - Y_{ijk}\right] \tag{8.4}$$

for each i, respectively.

The r_{ijk} are estimators of the random error e_{ijk} in model (8.3). These r_{ijk} are normally distributed with mean zero and variance

$$V(r_{ijk}) = \left(1 - \frac{1}{m_i}\right) \sigma_e^2 . \tag{8.5}$$

Refer to Liu and Weng (1991) for the details of the derivation of residuals and their variance. Thus the studentized residuals are

$$SR1 = \frac{r_{ijk}}{\left(\hat{V}(r_{ijk})\right)^{1/2}} \tag{8.6}$$

where $\hat{V}(r_{ijk})$ is the estimated value of $V(r_{ijk})$ obtained by replacing σ_e^2 by the mean square of the within-subject residual. As stated by Liu and Weng (1991), $SR1$ are treated as standard normal variables.

According to Jones and Kenward (1989), the response values corresponding to unusually large $SR1$ are called outliers or discordant values. The larger the residual, the more discordant is the corresponding response. To identify the outlier, the largest $|SR1|$ is significantly large at the 10%, 5%, 2.5% and 1% levels if it is greater than 3.

8.2.2 Outlier detection using $SR2$

As stated by Rousseeuw and Croux (1993), although many robust estimators of location exist, the sample median is still the most widely used in the literature. If $\{x_1, \ldots, x_n\}$ is a batch of numbers with sample size of n, its sample median is denoted by

$$median(x_i) \tag{8.7}$$

which is simply the middle order statistic when n is odd. When n is even, the average of the order statistics with ranks $(n/2)$ and $(n/2)+1$. The median has a breakdown point of 50% (which is the highest possible), because the estimate remains bounded when fewer than 50% of the data points are replaced by arbitrary numbers. Its influence function is also bounded, with the sharpest bound for any location estimator (Hampel et al., 1986).

A very robust scale estimator is the median absolute deviation about the median (*MADN*), is given by

$$MADN = b\, median\{|\, x_i - median(x_i)|\}\,. \tag{8.8}$$

where $b = 1.4826$, as suggested by Rousseeuw and Croux (1993), Ruppert (2010), and Leyset al. (2013). The *MADN* has the best possible breakdown point and its influence function is bounded, with the sharpest possible bound among all scale estimators. The *MADN* was first promoted by Hampel (1974), who attributed it to Gauss. The constant b in (8.8) is needed to make the estimator consistent for the parameter of interest.

The sample median and the *MADN* are simple and easy to calculate, but nevertheless very useful. Their extreme sturdiness makes them ideal for detecting the outliers in the data, by computing the absolute value of

$$\frac{x_i - median(x_i)}{MADN} \tag{8.9}$$

for each x_i and flagging those x_i as spurious for which this statistic exceeds a certain cutoff, D.

For this study, x_i in equations (8.8) and (8.9) is replaced by r_{ijk}. Here r_{ijk} can be any observations since it is a non-parametric method and it does not require any assumptions. The value of b remains the same (Mustafi, 1991). Therefore the median absolute deviation about the median of r_{ijk} ($MADN_{red}$) and scaled residual ($SR2$) are given by

$$MADN_{red} = b\ median\{|\ r_{ijk} - median(r_{ijk})|\}. \quad (8.10)$$

and

$$SR2 = \frac{r_{ijk} - median(r_{ijk})}{MADN_{red}}, \quad (8.11)$$

respectively. Responses will be labeled outliers when $|SR2| > D$ in Table 8.1 – 8.4, where D is the critical value of the largest $|SR2|$ at the significance level of 10%, 5%, 2.5% and 1%.

Simulation studies are carried out to construct a table of critical values for the largest $|SR2|$ under model (8.3). The mean of treatment 2 (μ_2) is set to 60, 80, 90, 100, 110, 125. The values of the constant γ are chosen to be 0.5, 5, 10 and 15, so that the coefficient of the intra-subject variation for the treatment 1 are 0.5%, 5%, 10% and 15%, respectively. For each total sample size considered $(N = 20,\ 40\ ,60\ ,80\ ,100)$, we calculate the $SR2$, and determine the largest $|SR2|$. The procedure of finding the largest $|SR2|$ is repeated 1000 times. They are then sorted in ascending order and the 90%, 95%, 97.5% and 99% percentiles are obtained. These percentiles approximate the critical values for significance levels of 0.10, 0.05, 0.025 and 0.01, respectively, and tabulated in Table 8.1 – 8.4. The values generally increase as γ increases, for μ_2 different from 100.

8.3 Simulation study

In this section, we compare the power of the $P(SR2)$ in testing for subject outliers with that of the $P(SR1)$ in a standard 2×2 crossover design. Random samples are generated under a two-sequence, two-period crossover model based on the procedure used in Luzar-Stiffler and Stiffler (2005). Random samples Y_{ijk} are first generated based on the following formula:

$$Y_{ijk} = \gamma(z_{0jk} + z_{ijk}) + \mu_i \quad (8.11)$$

Table 8.1: Critical values of the largest $|SR2|$ at significance level of 10%

N	μ_2	γ			
		0.5	**5**	**10**	**15**
20	60	0.701	0.943	1.212	1.530
	80	0.728	1.212	1.688	1.591
	90	0.782	1.688	1.551	1.427
	100	1.220	1.220	1.220	1.220
	110	0.782	1.688	1.551	1.427
	125	0.718	1.105	1.582	1.645
40	60	0.716	1.087	1.526	1.917
	80	0.757	1.526	2.018	2.026
	90	0.839	2.018	2.043	2.013
	100	1.789	1.789	1.789	1.789
	110	0.839	2.018	2.043	2.013
	125	0.740	1.334	1.971	2.026
60	60	0.715	1.076	1.506	1.927
	80	0.755	1.506	1.964	1.989
	90	0.835	1.964	2.001	1.931
	100	1.714	1.714	1.714	1.714
	110	0.835	1.964	2.001	1.931
	125	0.739	1.317	1.946	2.001
80	60	0.724	1.170	1.708	2.129
	80	0.774	1.708	2.305	2.393
	90	0.873	2.305	2.436	2.396
	100	2.177	2.177	2.177	2.177
	110	0.873	2.305	2.436	2.396
	125	0.754	1.503	2.192	2.408
100	60	0.720	1.132	1.610	2.017
	80	0.766	1.610	2.128	2.178
	90	0.858	2.128	2.167	2.115
	100	1.900	1.900	1.900	1.900
	110	0.858	2.128	2.167	2.115
	125	0.748	1.413	2.061	2.171

Table 8.2: Critical values of the largest $|SR2|$ at significance level of 5%

N	μ_2	γ			
		0.5	**5**	**10**	**15**
20	60	0.704	0.973	1.271	1.640
	80	0.734	1.271	1.781	1.724
	90	0.794	1.781	1.693	1.641
	100	1.408	1.408	1.408	1.408
	110	0.794	1.781	1.693	1.641
	125	0.722	1.152	1.705	1.767
40	60	0.714	1.073	1.501	1.896
	80	0.754	1.501	1.964	1.959
	90	0.834	1.964	2.023	1.978
	100	1.728	1.728	1.728	1.728
	110	0.834	1.964	2.023	1.978
	125	0.738	1.311	1.916	2.003
60	60	0.718	1.111	1.585	1.983
	80	0.762	1.585	2.100	2.139
	90	0.849	2.100	2.146	2.121
	100	1.947	1.947	1.947	1.947
	110	0.849	2.100	2.146	2.121
	125	0.744	1.380	2.014	2.149
80	60	0.723	1.164	1.682	2.056
	80	0.772	1.682	2.241	2.353
	90	0.870	2.241	2.295	2.220
	100	1.999	1.999	1.999	1.999
	110	0.870	2.241	2.295	2.220
	125	0.753	1.473	2.106	2.282
100	60	0.724	1.165	1.702	2.104
	80	0.773	1.702	2.266	2.351
	90	0.871	2.266	2.356	2.300
	100	2.167	2.167	2.167	2.167
	110	0.871	2.266	2.356	2.300
	125	0.753	1.482	2.137	2.318

Table 8.3: Critical values of the largest $|SR2|$ at significance level of 2.5%

N	μ_2	γ			
		0.5	**5**	**10**	**15**
20	60	0.703	0.960	1.245	1.564
	80	0.732	1.245	1.749	1.660
	90	0.789	1.749	1.605	1.533
	100	1.306	1.306	1.306	1.306
	110	0.789	1.749	1.605	1.533
	125	0.720	1.131	1.638	1.728
40	60	0.711	1.041	1.446	1.890
	80	0.748	1.446	1.957	1.921
	90	0.821	1.957	1.991	1.824
	100	1.648	1.648	1.648	1.648
	110	0.821	1.957	1.991	1.824
	125	0.733	1.261	1.886	1.965
60	60	0.716	1.091	1.546	1.943
	80	0.758	1.546	2.034	2.061
	90	0.841	2.034	2.066	2.046
	100	1.833	1.833	1.833	1.833
	110	0.841	2.034	2.066	2.046
	125	0.741	1.341	1.976	2.047
80	60	0.721	1.139	1.663	2.046
	80	0.767	1.663	2.225	2.284
	90	0.860	2.225	2.255	2.180
	100	1.964	1.964	1.964	1.964
	110	0.860	2.225	2.255	2.180
	125	0.749	1.446	2.086	2.209
100	60	0.722	1.145	1.666	2.060
	80	0.769	1.666	2.192	2.278
	90	0.863	2.192	2.278	2.216
	100	2.015	2.015	2.015	2.015
	110	0.863	2.192	2.278	2.216
	125	0.750	1.453	2.087	2.245

Table 8.4: Critical values of the largest $|SR2|$ at significance level of 1%

N	μ_2	γ			
		0.5	**5**	**10**	**15**
20	60	0.707	1.001	1.327	1.734
	80	0.740	1.327	1.842	1.818
	90	0.805	1.842	1.815	1.769
	100	1.546	1.546	1.546	1.546
	110	0.805	1.842	1.815	1.769
	125	0.727	1.196	1.775	1.841
40	60	0.711	1.039	1.438	1.854
	80	0.747	1.438	1.950	1.901
	90	0.820	1.950	1.909	1.747
	100	1.527	1.527	1.527	1.527
	110	0.820	1.950	1.909	1.747
	125	0.733	1.257	1.871	1.905
60	60	0.720	1.130	1.649	2.024
	80	0.766	1.649	2.194	2.248
	90	0.857	2.194	2.281	2.260
	100	2.050	2.050	2.050	2.050
	110	0.857	2.194	2.281	2.260
	125	0.747	1.426	2.072	2.247
80	60	0.720	1.125	1.649	2.036
	80	0.765	1.649	2.209	2.209
	90	0.855	2.209	2.189	2.145
	100	1.851	1.851	1.851	1.851
	110	0.855	2.209	2.189	2.145
	125	0.747	1.432	2.057	2.187
100	60	0.725	1.180	1.745	2.158
	80	0.776	1.745	2.353	2.447
	90	0.877	2.353	2.466	2.438
	100	2.249	2.249	2.249	2.249
	110	0.877	2.353	2.466	2.438
	125	0.755	1.519	2.209	2.419

where z_{0jk} and z_{ijk} are i.i.d. standard normal ($i, j = 1,2;\ \ k = 1,2, \dots, m_i$). Note that z_{0jk} and z_{ijk} are used to account for the between- and within-subject variations, respectively. Without loss of generality, the mean of treatment 1 (μ_1) is set to 100, while the mean of treatment 2 (μ_2) is set to 60, 80, 90, 100, 110, 125, as suggested by Wang and Chow (2003). The ability of $P(SR1)$ and $P(SR2)$ in detecting outlier then can be observed when the difference between μ_1 and μ_2 increases. The values of the constant γ are chosen to be 0.5, 5, 10 and 15, so that the coefficient of the intra-subject variation for the treatment 1 are 0.5%, 5%, 10% and 15%, respectively. For our study, the effectiveness of kinesiotape in improving the VO_2 peak increases with the coefficient of the intra-subject variation of less than 20%.

For simplicity, in generating the random samples Y_{ijk}, we assume the values of m_i are equal. Three values of m_i, 10, 30 and 50, are considered in the simulation. Let the total sample size, $N = \sum m_i$. The corresponding values of N therefore take the values 20, 60 and 100, respectively. The first subject is made into an outlier by multiplying the responses Y_{111} and Y_{112} by a constant p which varies from 10% to 200%. As an example, let the responses $Y_{111} = 34$ and $Y_{112} = 52$, p = 50%, after multiplying the responses Y_{111} and Y_{112} by a constant p, then we have new responses $Y_{111} = 17$ and $Y_{112} = 26$ as the designated outlier. The process is repeated 200 times and the power of performance is assessed by computing the percentage of times that the outlier is identified correctly.

To conduct the power studies for $P(SR1)$, the procedure are summarized as below,

Step 1. Using S-plus, random samples of z_{0jk} and z_{ijk} are generated from the standard normal distribution, where $i, j = 1,2$ and $k = 1,2, \dots, n_i$.

Step 2. Calculate the sum of each pair z_{0jk} and z_{ijk}. Each value of sum then multiplied with the chosen constant γ.

Step 3. Set the mean of treatment 1 (μ_1) to 100. Repeat the Step 1 – 2 and calculate Y_{ijk} for different mean of treatment 2 (μ_2).

Step 4. The first subject is made into an outlier by multiplying the responses Y_{111} and Y_{112} by a constant p. The contaminated sample of Y_{ijk} then is obtained.

Step 5. Calculate the residual r_{ijk} in model (8.3) using the contaminated sample of Y_{ijk}.

Step 6. Conduct the analysis of variance for the contaminated sample of Y_{ijk}. Then, estimate the variance of r_{ijk}, $V(r_{ijk})$ in model (8.4), by replacing σ_e^2 with the within-subjects residual mean squares.

Step 7. Calculate the studentized residuals $SR1$ in model (8.5) using the r_{ijk} and $V(r_{ijk})$ in Step 4 and 5.

Step 8. The Steps 1 – 7 are repeated (200) times and the times that the outlier is identified correctly are recorded.

Step 9. Calculate the power, which is the percentage of times (out of 200 simulated samples) that the outlier is identified correctly.

Step 10. The Steps 1 – 9 are repeated for different constant γ ($\gamma = 0.5, 5, 10, 15$), mean of treatment 2 ($\mu_2 = 60, 80, 90, 100, 110, 125$) and constant p (p = 10, 30, 50, 130, 150, 200).

Step 11. The complete procedure (Steps 1 – 10) is repeated for different sizes of group with treatment i ($n_i = 10, 30, 50$). The corresponding total sample sizes N will be 20, 60 and 100, respectively.

To conduct the power studies for $P(SR2)$, we use the same contaminated sample of Y_{ijk} in the power studies for $P(SR1)$, which is obtained using Steps 1 – 4 in the procedure above. We then calculate the scaled residuals $SR2$, as given by model (8.8).

We repeat this calculation 200 times and record the times that the outlier is identified correctly. With the same combination of constant γ, mean of treatment 2 (μ_2) and constant p, the power of $P(SR2)$ can then be obtained. The complete procedure is repeated for the same sizes of group with treatment i in the power studies for $P(SR1)$.

Tables 8.5 – 8.7 show the percentages of correctly detecting the designated outlier for sample sizes of 20, 60 and 100, respectively. For example, refer to Table 8.7 for both methods of $P(SR1)$ and $P(SR2)$, when $\mu_2 = 60$ and p = 50%, increasing the values of γ decreases the percentage of correctly identifying the designated outlier. Note that γ is the coefficient of variation within the subjects. For all sample sizes considered, the percentages of detection for both $P(SR1)$ and $P(SR2)$ are almost 100% when $\gamma =$ 0.5 or 5. However, when $\gamma = 10$ or 15, the performance of $P(SR2)$ is always better than that of $P(SR1)$ since it has higher percentages of detection. These results show that $P(SR2)$ is obviously more powerful than $P(SR1)$ for detecting outliers in a standard 2×2 crossover design.

8.4 Numerical Examples

As an illustration, both procedures above are applied to Clayton and Leslie's data (1981) and kinesiology data.

8.4.1 Clayton and Leslie's data

Clayton and Leslie (1981) considered the blood concentration-time curve (AUC) data from two erythromycin formulations in a bioavailability study. In their study, a standard 2×2 crossover experiment was conducted with 18 subjects to compare a new erythromycin formulation (erythromycin stearate) with a reference formulation (erythromycin base). As no sequence identification of each subject was provided in Clayton and Leslie (1981), we adapt the order of periods given in Weiner (1989) and

Table 8.5: Percentage of correctly identifying the designated outlier for sample size of 20

μ_2	p (%)	$SR1$				$SR2$			
		γ				γ			
		0.5	**5**	**10**	**15**	**0.5**	**5**	**10**	**15**
60	10	100	100	100	74	100	100	100	99.5
	30	100	100	96.5	64.5	100	100	100	95
	50	100	100	96.5	74	100	100	98	88.5
	130	100	100	99	90	100	100	98	92.5
	150	100	100	99.5	92	100	100	100	94
	200	100	100	100	95.5	100	100	100	99
80	10	100	100	100	86	100	100	100	100
	30	100	100	98.5	53	100	100	100	100
	50	100	100	75.5	29	100	100	99	96
	130	100	99	74.5	43	100	100	93	95
	150	100	100	87.5	67.5	100	100	98.5	98.5
	200	100	100	100	94	100	100	100	99.5
90	10	100	100	100	92	100	100	100	100
	30	100	100	99	57	100	100	100	100
	50	100	100	77	24.5	100	100	100	98.5
	130	100	96.5	53	28.5	100	99.5	97.5	96.5
	150	100	100	85.5	60.5	100	100	99	98.5
	200	100	100	100	93.5	100	100	100	100
100	10	100	100	100	96	100	100	100	100
	30	100	100	99.5	67.5	100	100	100	100
	50	100	100	82.5	24.5	100	100	100	98.5
	130	100	95.5	45.5	22	100	100	98.5	96
	150	100	100	85.5	58.5	100	100	100	98.5
	200	100	100	100	96	100	100	100	100
110	10	100	100	100	97.5	100	100	100	100
	30	100	100	100	76.5	100	100	100	100
	50	100	100	85.5	34	100	100	100	99
	130	100	98.5	56	34.5	100	100	98	96
	150	100	100	91.5	63.5	100	100	100	98
	200	100	100	100	98	100	100	100	100
125	10	100	100	100	98.5	100	100	100	100
	30	100	100	100	85.5	100	100	100	100
	50	100	100	92.5	54.5	100	100	100	94.5
	130	100	100	86	61	100	100	95	93.5
	150	100	100	97.5	76.5	100	100	100	98.5
	200	100	100	100	99	100	100	100	100

Table 8.6: Percentage of correctly identifying the designated outlier for sample size of 60

μ_2	p (%)	$SR1$				$SR2$			
		γ				γ			
		0.5	**5**	**10**	**15**	**0.5**	**5**	**10**	**15**
60	10	100	100	100	76	100	100	100	99
	30	100	100	100	68	100	100	100	91.5
	50	100	100	98	74.5	100	100	96	88
	130	100	100	100	83.5	100	100	98	95
	150	100	100	99.5	88.5	100	100	99.5	95.5
	200	100	100	100	97.5	100	100	100	99
80	10	100	100	100	87	100	100	100	100
	30	100	100	99.5	33.5	100	100	100	99
	50	100	100	68.5	20	100	100	100	94
	130	100	100	71	37	100	100	96	94.5
	150	100	100	86	57.5	100	100	98.5	97
	200	100	100	100	92.5	100	100	100	100
90	10	100	100	100	94.5	100	100	100	100
	30	100	100	100	39.5	100	100	100	100
	50	100	100	64.5	11.5	100	100	100	97.5
	130	100	95	42.5	23.5	100	100	96.5	94
	150	100	100	83	50	100	100	99	98
	200	100	100	100	92.5	100	100	100	100
100	10	100	100	100	98.5	100	100	100	100
	30	100	100	100	47.5	100	100	100	100
	50	100	100	71.5	7.5	100	100	100	98
	130	100	96	33.5	16	100	100	98.5	96.5
	150	100	100	82	46.5	100	100	99	98.5
	200	100	100	100	96	100	100	100	99.5
110	10	100	100	100	100	100	100	100	100
	30	100	100	100	65.5	100	100	100	100
	50	100	100	87	15	100	100	100	98.5
	130	100	99	49.5	25	100	100	98	94
	150	100	100	89.5	54	100	100	100	99
	200	100	100	100	98	100	100	100	100
125	10	100	100	100	100	100	100	100	100
	30	100	100	100	86	100	100	100	100
	50	100	100	97.5	41.5	100	100	100	98
	130	100	100	88	53	100	100	96.5	95.5
	150	100	100	98	77.5	100	100	100	96.5
	200	100	100	100	99	100	100	100	100

Table 8.7: Percentage of correctly identifying the designated outlier for sample size of 100

μ_2	p (%)	$SR1$				$SR2$			
		γ				γ			
		0.5	**5**	**10**	**15**	**0.5**	**5**	**10**	**15**
60	10	100	100	100	76	100	100	100	100
	30	100	100	99	73.5	100	100	100	95
	50	100	100	99	81	100	100	96.5	94.5
	130	100	100	100	88	100	100	97	96
	150	100	100	100	90	100	100	99.5	97
	200	100	100	100	96	100	100	100	100
80	10	100	100	100	92.5	100	100	100	100
	30	100	100	99.5	35.5	100	100	100	99.5
	50	100	100	70.5	25.5	100	100	100	95
	130	100	100	66.5	38	100	100	95.5	95.5
	150	100	100	84	54	100	100	98	97
	200	100	100	99.5	93	100	100	100	99
90	10	100	100	100	97.5	100	100	100	100
	30	100	100	99.5	32.5	100	100	100	100
	50	100	100	66	11	100	100	100	97
	130	100	95	40.5	26.5	100	99.5	96	96
	150	100	100	77	42	100	100	99	97.5
	200	100	100	100	92	100	100	100	99
100	10	100	100	100	98	100	100	100	100
	30	100	100	100	43	100	100	100	100
	50	100	100	71	9.5	100	100	100	98.5
	130	100	91.5	31.5	16.5	100	100	96.5	96
	150	100	100	74	44.5	100	100	99.5	97
	200	100	100	100	92.5	100	100	100	100
110	10	100	100	100	99.5	100	100	100	100
	30	100	100	100	63.5	100	100	100	100
	50	100	100	86	15	100	100	100	99
	130	100	98.5	50	25	100	100	98.5	97
	150	100	100	83	56	100	100	100	97.5
	200	100	100	100	96.5	100	100	100	100
125	10	100	100	100	100	100	100	100	100
	30	100	100	100	90.5	100	100	100	100
	50	100	100	98	44.5	100	100	100	98
	130	100	100	87.5	53.5	100	100	98	96
	150	100	100	97.5	77	100	100	99.5	97.5
	200	100	100	100	99.5	100	100	100	100

assign subject 1 through 9 to sequence 1 and the remaining subjects to sequence 2. Figures 8.1 and 8.2 indicate the scatter plots of AUC data for period 1 and 2, respectively. A point (7^{th} measurement from group 1 with erythromycin base) seems tobe far from others at the top of Figure 8.1. Using both procedures ($P(SR1)$ and $P(SR2)$), subject 7 in group 1 is identified as an outlying subject as shown in Figures 8.3 and 8.4 respectively. Similar result is showed by either the two-sample Hotelling T^2 or likelihood distance as stated by Chow and Liu (2009). Subject 7 has exceptionally high value of blood concentration in period 1 (12.39 $\mu g \cdot h/mL$ against mean value 4.95 $\mu g \cdot h/mL$ for group 1) but low value in period 2 (0.99 $\mu g \cdot h/mL$ against mean value 3.51 $\mu g \cdot h/mL$ for group 1), suggesting the identification of the subject as an outlier is acceptable. Upon repeating the procedure after removing subject 7, $P(SR1)$ does not identify any more outliers but $P(SR2)$ detect two new outliers; subject 11 (blood concentration are 7.14 $\mu g \cdot h/mL$ and 9.83 $\mu g \cdot h/mL$ respectively in period 1 and period 2) and subject 2 (blood concentration are 4.84 $\mu g \cdot h/mL$ and 8.87 $\mu g \cdot h/mL$ respectively in period 1 and period 2). As can be seen, both subjects have high values of blood concentration in period 2 compared to their means (3.51 $\mu g \cdot h/mL$ in group 1 and 4.72 $\mu g \cdot h/mL$ in group 2) indicating the subjects are candidates to be outliers.

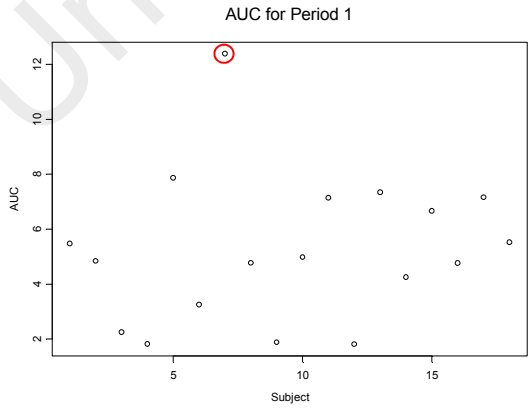


Figure 8.1: Scatter plot of AUC data for period 1

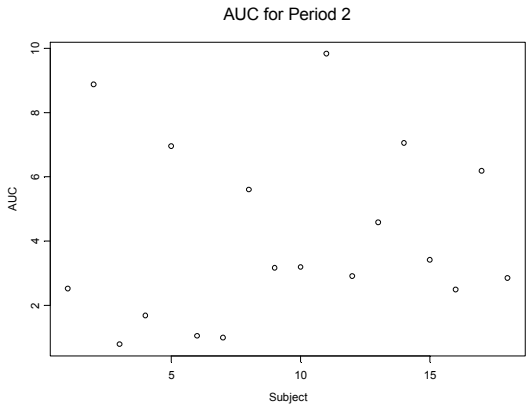


Figure 8.2: Scatter plot of AUC data for period 2

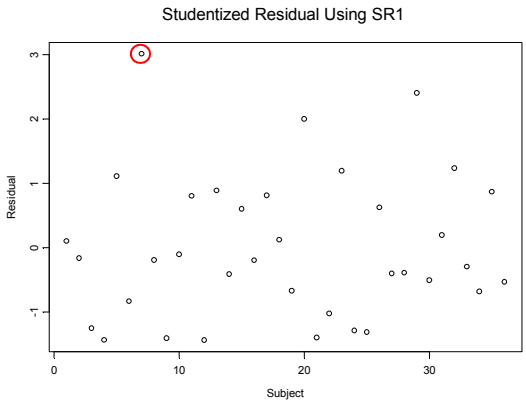


Figure 8.3: Studentized residual using $SR1$ for AUC data

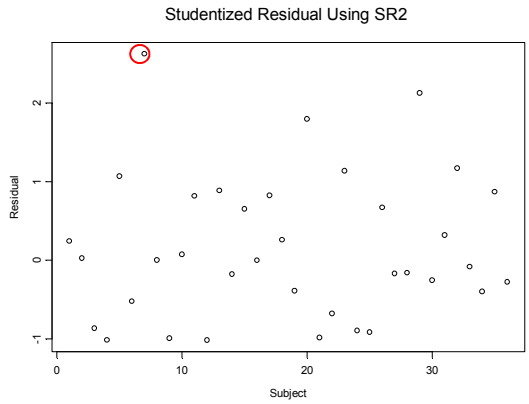


Figure 8.4: Studentized residual using $SR2$ for AUC data

8.4.2 Kinesiology data

Using the real data set of peak oxygen consumption or VO_2 peak, as described in Chapter 7, we consider 77 measurements of peak oxygen consumption or VO_2 peak (in ml/kg/min) recorded from a six-minute Astrand submaximal cycling exercise test conducted at least one week apart. Figures 8.5 and 8.6 indicate the scatter plots of VO_2 peak for period 1 and 2, respectively. A point (34^{th} measurement from group 1 with treatment A) seems to be far from others at the bottom of Figure 8.5. We fit the full data to model (8.3) and proceed to both procedures ($P(SR1)$ and $P(SR2)$) for detecting the possible outliers in the data.

From Figures 8.7 and 8.8, we see that at each plot, the 34^{th} measurement (from group 1 with treatment A) gives the largest values of $|SR1|$ and $|SR2|$. The largest values of $|SR1|$ is greater than 3 while the largest values of $|SR2|$ is greater than its corresponding tabulated critical value. Figures 8.7 and 8.8, therefore, reveal a dramatic outlier for subject 34.

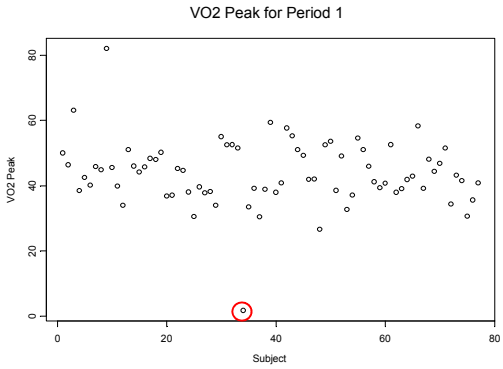


Figure 8.5: Scatter plot of VO_2 peak for period 1

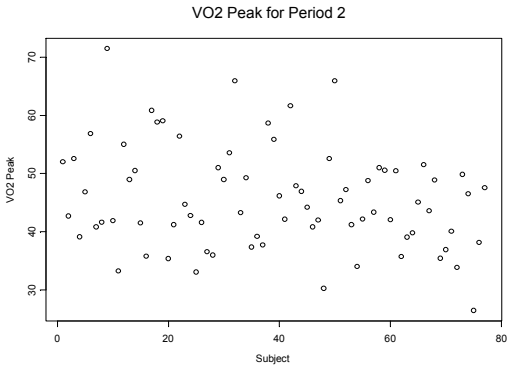


Figure 8.6: Scatter plot of VO_2 peak for period 2

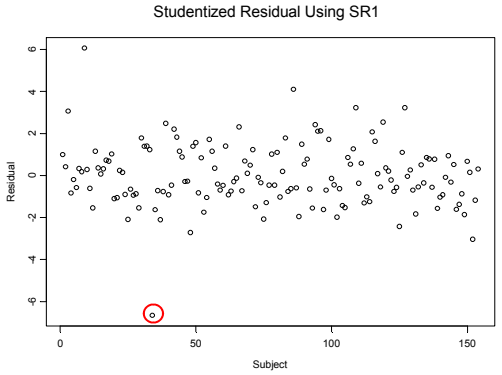


Figure 8.7: Studentized residual using $SR1$

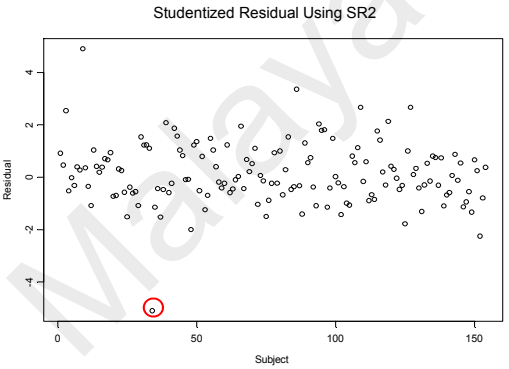


Figure 8.8: Studentized residual using $SR2$

To compare the ability of both procedures in detecting the possible outliers in the data, we proceed to test the subsequent largest values of $|SR1|$ and $|SR2|$ in the same manner. The sequential testing procedure stops when only the subsequent largest values of $|SR1|$ are less than 3 for $P(SR1)$, or the subsequent largest values of $|SR2|$ less than its corresponding tabulated critical values for $P(SR2)$. Both $P(SR1)$ and $P(SR2)$ detect six possible outliers from the data as listed in Table 8.8. As can be seen, at least one of the values of VO_2 in either period of the six subjects are distinctly different from the means given in the last two rows.

Table 8.8: List of outliers

Sequence	Subject	Group	Reading	
			Period 1	Period 2
1	34	1	1.75	49.26
2	9	1	81.98	71.47
3	32	1	52.55	65.92
4	50	2	53.56	65.92
5	3	1	63.06	52.55
6	75	2	30.62	26.45
Mean		1	43.36	46.58
Mean		2	44.14	44.73

8.5 Summary

In this study, we investigated the detection of outliers based on residuals in a standard 2×2 crossover design. We calculated two types of studentized residual: $SR1$ using a classical procedure and $SR2$ using a new procedure based on median absolute deviation. Their performances in testing for within-subject outliers are compared. Based on a simulation study, we concluded that $P(SR2)$ is more powerful than $P(SR1)$. As an illustration, both procedures were applied to Clayton and Leslie's data (1981) and kinesiology data. The results showed that $P(SR2)$ is comparable for $P(SR1)$ in detecting outliers.

CHAPTER 9: KINESIOLOGY STUDY – OUTLIER DETECTION IN 2×2 CROSSOVER DESIGN USING BAYESIAN FRAMEWORK

9.1 Introduction

This chapter considers the problem of outlier detection method in 2×2 crossover design via Bayesian framework. The usual Bayesian approach for outlier detection uses the maximum posterior probability as the criteria to identify outlier, and not use cut-off point (Chaloner and Brant, 1988; Shotwell and Slate, 2011). We study the problem of outlier detection in bivariate data fitted using generalized linear model (GLM) in Bayesian framework presented by Nawama et al. (2015) and the work outlined by Unnikrishnan (2010). We follow closely their works but adapt them into 2×2 crossover design. In Bayesian framework, we assume that the random subject effect and the errors to be generated from normal distributions. However, the outlying subjects come from normal distribution with different variance. Due to the complexity of the resulting joint posterior distribution, we obtain the information on the posterior distribution from samples by using Markov Chain Monte Carlo method.

9.2 Modified generalized linear model with outlier

Unnikrishnan (2010) has developed an outlier model using Bayesian approach and it is modeled as a generalized linear model (GLM). As mentioned by Unnikrishnan (2010), consider $N = \{1, \dots, N\}$ be a finite population with known N. For each unit $i \in N$, we have the real valued response variable y_i and known $p \times 1$ vector of explanatory variables $\boldsymbol{x_i}$ where $\boldsymbol{x_i'} = \begin{pmatrix} x_{i1} & \cdots & x_{ip} \end{pmatrix}$

Assume that a random sample of size n is obtained with a number of suspected outliers. Let v^h be the set of all outlying observations where h denotes the number of outliers. We consider the models with/without outliers based on GLM such that

$$f(y_i|\theta_i,\phi_i,\delta)=\begin{cases}exp\{\phi_i(y_i\theta_i)+c_i(\theta_i,\phi_i)+d_i(y_i)\} & i\in N-v^h\\ exp\left\{\frac{\phi_i}{\delta}(y_i\theta_i)+c_i(\delta,\theta_i,\phi_i)+d_i(y_i)\right\} & i\in v^h\end{cases} \quad (9.1)$$

where θ_i is a location parameter, ϕ_i and δ are scale parameters, and $c(\cdot)$, $d(\cdot)$ are known functions. This model was used by Pettit (1988) for the case of exponential distribution with known scale parameter δ. The parameters θ_i are modeled through a link function $h(\cdot)$ and is given by

$$h(\theta_i)=\boldsymbol{x_i'\beta}+\varepsilon_i\ , \qquad i=1,\dots,N \quad (9.2)$$

where $\boldsymbol{\beta}^T=(\beta_1\ \cdots\ \beta_p)$ is an unknown regression coefficients of $p\times 1$ vector parameters and the error components ε_i's are independently and normally distributed. In other words,

$$h(\theta_i)|\sigma^2\ \sim\ N(\boldsymbol{x_i'\beta},\sigma^2)\ .$$

For conventional outlier models, it is usually assumed that they have the same mean for all observations. However, we expect to see higher variance for outlying observations, that is, when $\delta>1$.

By considering the exponential regression model, Unnikrishnan (2010) suggested that model (9.1) can be reparameterized to fit their general framework as follows:

$$f(y_i|\theta_i,\phi_i,\delta)=\begin{cases}\phi\theta_i exp(-\phi\theta_i y_i) & i\notin v^h\\ \frac{\phi\theta_i}{\delta}exp\left(-\frac{\phi\theta_i}{\delta}y_i\right) & i\in v^h\end{cases} \quad (9.3)$$

with the link function

$$log\theta_i=\beta(x_i-\bar{x})+\varepsilon_i$$

where $\bar{x}$ is the mean of the sample.

Nawama et al. (2015) consider the problem of outlier detection in bivariate exponential samples using GLM via Bayesian approach. They follow closely the work outlined by Unnikrishnan (2010) but present every step of the detection procedure by considering the model (9.3). Due to the complexity of the resulting joint posterior

distribution, they obtain the information on the posterior distribution from samples generated by Markov chain Monte Carlo (MCMC) sampling, in particular, using the Gibbs sampler with Metropolis-Hastings (MH) algorithm. They use these samples to identify the observation which has the highest probability of being an outlier.

9.3 The model

We follow closely the works of Nawama et al. (2015) and Unnikrishnan (2010) but adapt them into 2×2 crossover design. Under the full model (8.1), for case without outliers, the expected value of y_{ijk} is

$$E(y_{ijk}) = E(\mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]} + S_{ik} + e_{ijk})$$
$$= \mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]}$$

while the variance of y_{ijk} is

$$Var(y_{ijk}) = Var(\mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]} + S_{ik} + e_{ijk})$$
$$= Var(S_{ik}) + Var(e_{ijk}) = \sigma_s^2 + \sigma_e^2,$$

Hence,

$$y_{ijk} \sim N(\mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]},\ \sigma_s^2 + \sigma_e^2).$$

On the other hand, for the case with outliers, the variance component $\{S_{ik}\}$ is assumed to be independent and normally distributed with mean 0 and variances $\delta^2\sigma_s^2$. Therefore,

$$y_{ijk} \sim N\left(\mu + p_j + \tau_{d[i,j]} + \lambda_{d[i,j-1]}, \delta^2\sigma_s^2 + \sigma_e^2\right).$$

Assume that a random sample size of $n = \sum m_i$ is obtained with a number of suspected outliers. Define $y_{ijk} = (y_{111}, y_{112}, \cdots, y_{11n})$. Let v^h be the set of all outlying observations, where h denotes the number of outliers. We consider the model with/without outliers such that

$$f\left(y_{ijk}\middle|\mu, p_j, \tau_{d[i,j]}, \lambda_{d[i,j-1]}, \delta\right)$$

$$= \begin{cases} [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\sigma_s^2 + \sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right] & \text{for } k \notin v^h \\ [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta^2\sigma_s^2 + \sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right] & \text{for } k \in v^h \end{cases}$$

(9.4)

Using the Bayesian approach, we consider normal prior distributions for the overall mean, μ, the period effect, p_j, the treatment effect, $\tau_{d[i,j]}$, and the carryover effect, $\lambda_{d[i,j-1]}$, as suggested by Chen and Huang (2015). For the parameter δ, Unnikrishnan (2010) assume that this extra variance component of the outlying observations is bounded above by a known constant δ_{max}, so that $1 < \delta < \delta_{max} < \infty$ and therefore Uniform(1, δ_{max}) prior is assigned to it. According to the suggestions in Unnikrishnan (2010), we shall assume that any distinct k-tuples are equally likely to be outliers and prior for v^h assigns equal probability of $\binom{N}{h}^{-1}$. In other words, we assume that

$$\mu \sim N\left(\mu_0, \sigma_\mu^2\right)$$

$$p_j \sim N\left(\mu_p, \sigma_p^2\right), \quad j = 1,2$$

$$\tau_{d[i,j]} \sim N(\mu_\tau, \sigma_\tau^2), \quad i = 1,2; j = 1,2$$

$$\lambda_{d[i,j-1]} \sim N\left(\mu_\lambda, \sigma_\lambda^2\right)$$

$$\delta \sim Uniform(1, \delta_{max})$$

$$p(v^h|h) = \binom{N}{h}^{-1}$$

(9.5)

where the hyperparameters $\mu_0, \sigma_\mu^2, \mu_p, \sigma_p^2, \mu_\tau, \sigma_\tau^2, \mu_\lambda, \sigma_\lambda^2, \delta_{max}, N, h$ are all pre-specified. Then, the joint likelihood function is given by

$$L(\boldsymbol{y}|\mu,\boldsymbol{p},\boldsymbol{\tau},\boldsymbol{\lambda},\delta,v^h)$$

$$=\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h} f\left(y_{ijk}\middle|\mu,p_j,\tau_{d[i,j]},\lambda_{d[i,j-1]}\right)\times\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h} f\left(y_{ijk}\middle|\mu,p_j,\tau_{d[i,j]},\lambda_{d[i,j-1]},\delta\right)$$

$$=\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right] \tag{9.6}$$

Consequently, from the result obtain in equations (9.5) and (9.6), the full joint posterior distribution for the parameters $\mu,\boldsymbol{p},\boldsymbol{\tau},\boldsymbol{\lambda},\delta,v^h$ is given by

$$f(\mu,\boldsymbol{p},\boldsymbol{\tau},\boldsymbol{\lambda},\delta,v^h|\boldsymbol{y})$$

$$\propto L\left(y_{ijk}\middle|\mu,p_j,\tau_{d[i,j]},\lambda_{d[i,j-1]},\delta,v^h\right)\times f\left(\mu,p_j,\tau_{d[i,j]},\lambda_{d[i,j-1]},\delta,v^h\right)$$

$$\propto\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n} f\left(y_{ijk}\middle|\mu,p_j,\tau_{d[i,j]},\lambda_{d[i,j-1]},\delta,v^h\right)\times f(\mu)\times f\left(p_j\right)\times f\left(\tau_{d[i,j]}\right)\times f\left(\lambda_{d[i,j-1]}\right)$$

$$\times f(\delta)\times f(v^h)$$

$$\propto\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times\left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}}exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_\mu^2}\right]\times\left(2\pi\sigma_p^2\right)^{-\frac{1}{2}}exp\left[-\frac{\left(p_j-\mu_p\right)^2}{2\sigma_p^2}\right]\times(2\pi\sigma_\tau^2)^{-\frac{1}{2}}exp\left[-\frac{\left(\tau_{d[i,j]}-\mu_\tau\right)^2}{2\sigma_\tau^2}\right]$$

$$\times(2\pi\sigma_\lambda^2)^{-\frac{1}{2}}exp\left[-\frac{\left(\lambda_{d[i,j-1]}-\mu_\lambda\right)^2}{2\sigma_\lambda^2}\right]\times\frac{1}{\delta_{max}-1}\times\left(\frac{N!}{(N-h)!h!}\right) \tag{9.7}$$

Since this posterior distribution is intractable, sampling is carried out using the MCMC sampling method, in particular using Metropolis-Hastings (MH) algorithm.

9.4 Sampling methods of the parameters

Note that model (9.4) involves multiple parameters that are structured hierarchically such that the dependency of the parameters is reflected in the joint probability distribution. The conditional posterior distributions of the parameters are intractable and therefore we use the MH algorithm for sampling purposes. The sampling methods for each of the parameters $\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h$ are given in detail as below.

a) Parameter δ

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter δ given $\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, v^h$ is given by

$$f(\delta|\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, v^h)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h} \frac{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}}{\delta_{max}-1} exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right].$$

Here we propose to use a proposal density for δ_{prop} as

$$g(\delta) = \frac{1}{\delta_{max}-1}$$

so that δ_{prop} has a uniform (1, δ_{max}) distribution. Using the MH algorithm, candidate point δ_{prop} is accepted with probability

$$\alpha(\delta, \delta_{prop}) = \min\left(1, \frac{f(\delta_{prop}|\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, v^h)g(\delta)}{f(\delta|\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, v^h)g(\delta_{prop})}\right)$$

$$= \min\left(1, \frac{\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}\left\{[2\pi(\delta_{prop}^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta_{prop}^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}}{\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}}\right)$$

The explanations of MH algorithm for calculating $\alpha(\delta, \delta_{prop})$ can be found in Appendix.

b) Parameter μ

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter μ given $\boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h$ is given by

$$f(\mu|\boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}}exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_\mu^2}\right].$$

Here, we introduce a function ω_k where

$$\omega_k = \begin{cases} 1 & \text{for } k \in v^h \\ 0 & \text{for } k \notin v^h \end{cases} \tag{9.8}$$

so that

$$f(\mu|\boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}^{1-\omega_k}$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}^{\omega_k}$$

$$\times \left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}}exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_\mu^2}\right]$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}exp\left[-\frac{(1-\omega_k)}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{\omega_k}{2}}exp\left[-\frac{\omega_k}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \left(2\pi\sigma_\mu^2\right)^{-\frac{1}{2}}exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_\mu^2}\right].$$

Here we choose the proposal density for μ_{prop} as

$$g(\mu) = \left(2\pi\sigma_{\mu}^{2}\right)^{-\frac{1}{2}} exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_{\mu}^{2}}\right],$$

so that μ_{prop} has a $N\left(\mu_0, \sigma_{\mu}^{2}\right)$ distribution. Using the MH algorithm, candidate point μ_{prop} is accepted with probability

$$\alpha\left(\mu, \mu_{prop}\right) = \min\left(1, \frac{f\left(\mu_{prop}|\boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h\right) g(\mu)}{f(\mu|\boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h)\mathrm{g}\left(\mu_{prop}\right)}\right)$$

$$= \min\left(1, \frac{A_1}{B_1}\right)$$

where

$$A_1 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times q_1]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times q_1]\}$$

$$B_1 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times r_1]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times r_1]\}$$

$$a = [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}$$

$$b = -\frac{(1-\omega_k)}{2(\sigma_s^2 + \sigma_e^2)}$$

$$c = [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{\omega_k}{2}}$$

$$d = -\frac{\omega_k}{2(\delta^2\sigma_s^2 + \sigma_e^2)}$$

$$q_1 = \left(y_{ijk} - \mu_{prop} - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

$$r_1 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

c) Parameter $\boldsymbol{p}$

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter $p_j, j = 1, 2$ given $\mu, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h$ is given by

$$f_j(p_j|\mu, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h, \boldsymbol{p})$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \left(2\pi\sigma_p^2\right)^{-\frac{1}{2}} exp\left[-\frac{(p_j-\mu_p)^2}{2\sigma_p^2}\right].$$

We use the function ω_k as defined in equation (9.8) so that

$$f_j(p_j|\mu, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h, \boldsymbol{p})$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{(1-\omega_k)}{2}} exp\left[-\frac{(1-\omega_k)}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{\omega_k}{2}} exp\left[-\frac{\omega_k}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \left(2\pi\sigma_p^2\right)^{-\frac{1}{2}} exp\left[-\frac{(p_j-\mu_p)^2}{2\sigma_p^2}\right].$$

Here we propose to use a proposal density for p_j as

$$g_j(p_j) = \left(2\pi\sigma_p^2\right)^{-\frac{1}{2}} exp\left[-\frac{\left(p_j-\mu_p\right)^2}{2\sigma_p^2}\right].$$

Therefore, by using MH algorithm, for parameter $\boldsymbol{p}$, we update p_1 and p_2 one by one, where for each p_j, the candidate point $p_{j_{prop}}$ is accepted with probability

$$\alpha\left(p_j, p_{j_{prop}}\right) = min\left(1, \frac{f_{j_{prop}}\left(p_{j_{prop}}\middle|\mu, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h, \boldsymbol{p}\right) g_j(p_j)}{f_j(p_j|\mu, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta, v^h, \boldsymbol{p}) g_{j_{prop}}\left(p_{j_{prop}}\right)}\right)$$

$$= min\left(1, \frac{A_2}{B_2}\right)$$

where

$$A_2 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times q_2]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times q_2]\}$$

$$B_2 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times r_2]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times r_2]\}$$

$$a = [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}$$

$$b = -\frac{(1-\omega_k)}{2(\sigma_s^2 + \sigma_e^2)}$$

$$c = [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{\omega_k}{2}}$$

$$d = -\frac{\omega_k}{2(\delta^2\sigma_s^2 + \sigma_e^2)}$$

$$q_2 = \left(y_{ijk} - \mu - p_{j_{prop}} - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

$$r_2 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

d) Parameter $\boldsymbol{\tau}$

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter $\tau_{d[i,j]}$, $i, j = 1, 2$, given $\mu, \boldsymbol{p}, \boldsymbol{\lambda}, \delta, v^h$ is given by

$$f_{d[i,j]}\left(\tau_{d[i,j]}\middle|\mu,\boldsymbol{p},\boldsymbol{\lambda},\delta,v^h,\boldsymbol{\tau}\right)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{1}{2}}exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]$$

$$\times (2\pi\sigma_\tau^2)^{-\frac{1}{2}}exp\left[-\frac{(\tau_{d[i,j]}-\mu_\tau)^2}{2\sigma_\tau^2}\right].$$

We use the function ω_t as defined in equation (9.8) so that

$$f_{d[i,j]}\left(\tau_{d[i,j]}\middle|\mu,\boldsymbol{p},\boldsymbol{\lambda},\delta,v^h,\boldsymbol{\tau}\right)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}exp\left[-\frac{(1-\omega_k)}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{\omega_k}{2}}exp\left[-\frac{\omega_k}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \left(2\pi\sigma_p^2\right)^{-\frac{1}{2}}exp\left[-\frac{(\tau_{d[i,j]}-\mu_\tau)^2}{2\sigma_\tau^2}\right].$$

Here we propose to use a proposal density for $\tau_{d[i,j]}$ as

$$g_{d[i,j]}\left(\tau_{d[i,j]}\right)=(2\pi\sigma_\tau^2)^{-\frac{1}{2}}exp\left[-\frac{\left(\tau_{d[i,j]}-\mu_\tau\right)^2}{2\sigma_\tau^2}\right].$$

Therefore, by using MH algorithm, for parameter $\boldsymbol{\tau}$, we update $\tau_{11},\tau_{12},\tau_{21}$ and τ_{22} one by one, where for each $\tau_{d[i,j]}$, the candidate point $\tau_{d[i,j]_{prop}}$ is accepted with probability

$$\alpha\left(\tau_{d[i,j]},\tau_{d[i,j]_{prop}}\right)=min\left(1,\frac{f_{d[i,j]_{prop}}\left(\tau_{d[i,j]_{prop}}\middle|\mu,\boldsymbol{p},\boldsymbol{\lambda},\delta,v^h,\boldsymbol{\tau}\right)g_{d[i,j]}\left(\tau_{d[i,j]}\right)}{f_{d[i,j]}\left(\tau_{d[i,j]}\middle|\mu,\boldsymbol{p},\boldsymbol{\lambda},\delta,v^h,\boldsymbol{\tau}\right)g_{d[i,j]_{prop}}\left(\tau_{d[i,j]_{prop}}\right)}\right)$$

$$=min\left(1,\frac{A_3}{B_3}\right)$$

where

$$A_3 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times q_3]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times q_3]\}$$

$$B_3 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a \times exp[b \times r_3]\} \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c \times exp[d \times r_3]\}$$

$$a = [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}$$

$$b = -\frac{(1-\omega_k)}{2(\sigma_s^2 + \sigma_e^2)}$$

$$c = [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{\omega_k}{2}}$$

$$d = -\frac{\omega_k}{2(\delta^2\sigma_s^2 + \sigma_e^2)}$$

$$q_3 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]_{prop}} - \lambda_{d[i,j-1]}\right)^2$$

$$r_3 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

e) Parameter $\boldsymbol{\lambda}$

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter $\lambda_{d[i,j-1]}$, $i,j = 1,2$, given $\mu, \boldsymbol{p}, \boldsymbol{\tau}, \delta, v^h$ is given by

$$f_{d[i,j-1]}\left(\lambda_{d[i,j-1]} \middle| \mu, \boldsymbol{p}, \boldsymbol{\tau}, \delta, v^h, \boldsymbol{\lambda}\right)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k \notin v^h}[2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k \in v^h}[2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right]$$

$$\times (2\pi\sigma_\lambda^2)^{-\frac{1}{2}} exp\left[-\frac{\left(\lambda_{d[i,j-1]} - \mu_\lambda\right)^2}{2\sigma_\lambda^2}\right].$$

We use the function ω_t as defined in equation (9.8) so that

$$f_{d[i,j-1]}\left(\lambda_{d[i,j-1]}\middle|\mu,\boldsymbol{p},\boldsymbol{\tau},\delta,v^h,\boldsymbol{\lambda}\right)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\sigma_s^2+\sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}exp\left[-\frac{(1-\omega_k)}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\left\{[2\pi(\delta^2\sigma_s^2+\sigma_e^2)]^{-\frac{\omega_k}{2}}exp\left[-\frac{\omega_k}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk}-\mu-p_j-\tau_{d[i,j]}-\lambda_{d[i,j-1]}\right)^2\right]\right\}$$

$$\times \left(2\pi\sigma_p^2\right)^{-\frac{1}{2}}exp\left[-\frac{\left(\lambda_{d[i,j-1]}-\mu_\lambda\right)^2}{2\sigma_\lambda^2}\right].$$

Here we propose to use a proposal density for $\lambda_{d[i,j-1]}$ as

$$g_{d[i,j-1]}\left(\lambda_{d[i,j-1]}\right)=\left(2\pi\sigma_p^2\right)^{-\frac{1}{2}}exp\left[-\frac{\left(\lambda_{d[i,j-1]}-\mu_\lambda\right)^2}{2\sigma_\lambda^2}\right].$$

Therefore, by using MH algorithm, for parameter $\boldsymbol{\lambda}$, we have $\lambda_{10},=\lambda_{20}=0$ and update λ_{11} and λ_{21} one by one, where for each $\lambda_{d[i,j-1]}$, the candidate point $\lambda_{d[i,j-1]_{prop}}$ is accepted with probability

$$\alpha\left(\lambda_{d[i,j-1]},\lambda_{d[i,j-1]_{prop}}\right)$$

$$=min\left(1,\frac{f_{d[i,j-1]_{prop}}\left(\lambda_{d[i,j-1]_{prop}}\middle|\mu,\boldsymbol{p},\boldsymbol{\tau},\delta,v^h,\boldsymbol{\lambda}\right)g_{d[i,j-1]}\left(\lambda_{d[i,j-1]}\right)}{f_{d[i,j-1]}\left(\lambda_{d[i,j-1]}\middle|\mu,\boldsymbol{p},\boldsymbol{\tau},\delta,v^h,\boldsymbol{\lambda}\right)g_{d[i,j-1]_{prop}}\left(\lambda_{d[i,j-1]_{prop}}\right)}\right)$$

$$=min\left(1,\frac{A_4}{B_4}\right)$$

where

$$A_4=\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a\times exp[b\times q_4]\}\times\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c\times exp[d\times q_4]\}$$

$$B_4=\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{a\times exp[b\times r_4]\}\times\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k=1}^{n}\{c\times exp[d\times r_4]\}$$

$$a = [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{(1-\omega_k)}{2}}$$

$$b = -\frac{(1-\omega_k)}{2(\sigma_s^2 + \sigma_e^2)}$$

$$c = [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{\omega_k}{2}}$$

$$d = -\frac{\omega_k}{2(\delta^2\sigma_s^2 + \sigma_e^2)}$$

$$q_4 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]_{prop}}\right)^2$$

$$r_4 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

f) Parameter v^h

Based on the full joint posterior distribution (9.7), the conditional posterior distribution for parameter v^h given $\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta$ is given by

$$f(v^h|\mu, \boldsymbol{p}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \delta)$$

$$\propto \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h}[2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\sigma_s^2+\sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right]$$

$$\times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h}[2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}} exp\left[-\frac{1}{2(\delta^2\sigma_s^2+\sigma_e^2)}\left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2\right].$$

For the case of $h = 1$, we let $v^1 = v = \{v_1\}$. To find a new value of v_1, we select a unit at random from v^c, say v_{prop}. If the proposal is accepted, then v_1 goes out and v_{prop} replace the value v_1 as the current outlier. Then, using MH algorithm, this state is accepted with probability

$$\alpha(v_1, v_{prop}) = min\left(1, \frac{\begin{matrix}\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{v_{prop}\notin v} f\left(y_{ijk}\middle|\mu, p_j, \tau_{d[i,j]}, \lambda_{d[i,j-1]}\right) \\ \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{v_{prop}\in v} f\left(y_{ijk}\middle|\mu, p_j, \tau_{d[i,j]}, \lambda_{d[i,j-1]}, \delta\right)\end{matrix}}{\begin{matrix}\prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v} f\left(y_{ijk}\middle|\mu, p_j, \tau_{d[i,j]}, \lambda_{d[i,j-1]}\right) \\ \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v} f\left(y_{ijk}\middle|\mu, p_j, \tau_{d[i,j]}, \lambda_{d[i,j-1]}, \delta\right)\end{matrix}}\right)$$

$$= min\left(1, \frac{A_5}{B_5}\right)$$

where

$$A_5 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h} s \times exp[t \times q_5] \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h} u \times exp[w \times q_5]$$

$$B_5 = \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\notin v^h} s \times exp[t \times r_5] \times \prod_{i=1}^{2}\prod_{j=1}^{2}\prod_{k\in v^h} u \times exp[w \times r_5]$$

$$s = [2\pi(\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}}$$

$$t = -\frac{1}{2(\sigma_s^2 + \sigma_e^2)}$$

$$u = [2\pi(\delta^2\sigma_s^2 + \sigma_e^2)]^{-\frac{1}{2}}$$

$$w = -\frac{1}{2(\delta^2\sigma_s^2 + \sigma_e^2)}$$

$$q_5 = \left(y_{prop} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

$$r_5 = \left(y_{ijk} - \mu - p_j - \tau_{d[i,j]} - \lambda_{d[i,j-1]}\right)^2$$

9.5 Numerical examples

The method as described in the previous section is now applied to Clayton and Leslie's data (1981) and kinesiology data. The values of μ_0, μ_π, μ_τ and μ_λ equal to 0 while

the values of $\sigma_\mu^2, \sigma_p^2, \sigma_\tau^2, \sigma_\lambda^2, \sigma_s^2$ and σ_e^2 equal to 1000; these values are suggested by Chen (2015), and δ_{max} equals to 10.

9.5.1 Clayton and Leslie's data

Using this real data set, as described in section 8.4.1, we run the method for 1000 iterations, with a burn-in of 500. We are especially interested in estimating the probability of a subject being an outlier. The probability of subject i being an outlier in this model can be estimated using the proportion of iterations that $v = \{i\}$. Figure 9.1 shows the estimated probability of being an outlier for subjects 1 to 18. Given that there is one outlier; subject 11 has the highest probability of being an outlier with the probability of approximately 0.30. This is to be expected since for subject 11 (that is, subject 2 from group 2), the blood concentration is unusually large in the data set for period 2. Therefore, we may conclude that subject 11 is likely an outlier. Note that subject 11 is one of the outliers identified using $P(SR2)$ as described in Section 8.4.1.

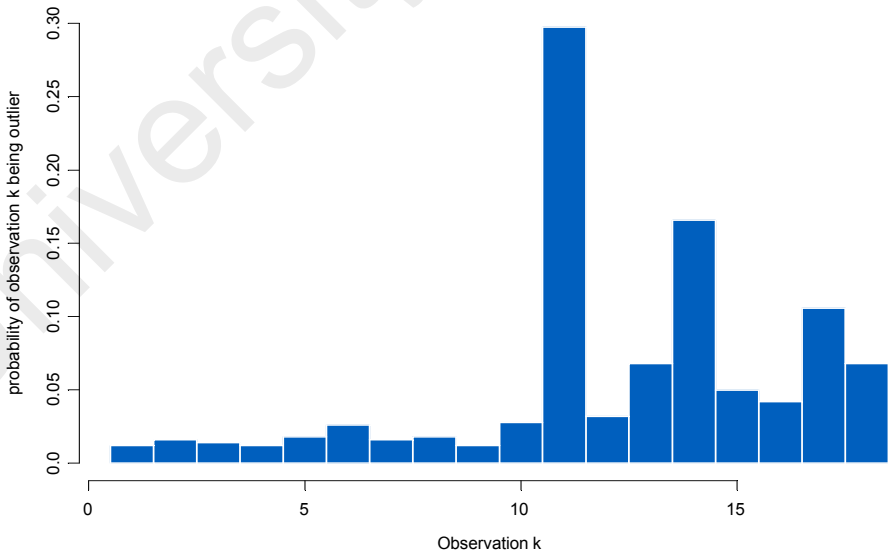


Figure 9.1: Clayton and Leslie's data: probability for an observation being outlier

9.5.2 Kinesiology data

The real data set of peak oxygen consumption or VO_2 peak, as described in Chapter 7, is used as illustration in this section. However, only 74 subjects who completed the study (AB = 37, BA = 37) are included for analyses. We run the method for 10000 iterations, with a burn-in of 5000. With the same interest in previous section, Figure 9.2 shows the estimated probability of being an outlier for subjects 1 to 74 using the proportion of iterations that $v = \{i\}$. Given that there is one outlier, subject 50 has the highest probability of being an outlier with the probability of approximately 0.32. This is likely because for subject 50 (that is, subject 13 from group 2), the VO_2 is unusually large in the data set for period 1. Therefore, we may conclude that subject 13 from group 2 is likely an outlier. Note that subject 50 is one of the outliers identified using $P(SR1)$ and $P(SR2)$ as described in Section 8.4.2.

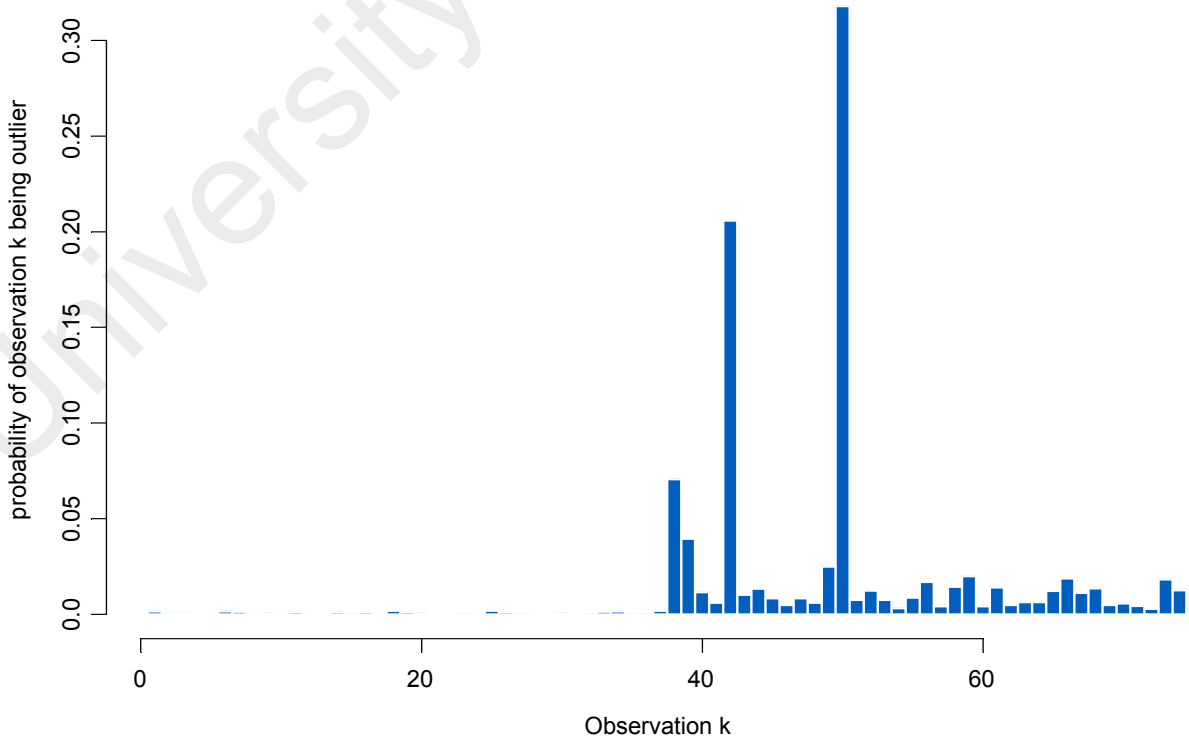


Figure 9.2: Kinesiology data: probability for an observation being outlier

9.6 Summary

In this chapter, we have considered the problem of detecting outlier using Bayesian approach in 2×2 crossover design. We have shown that with the chosen prior distributions for the parameter, we can obtain the information from samples generated by MCMC sampling, in particular using the MH algorithm. When applied to both Clayton and Leslie's data (1981) and kinesiology data, this method is able to detect an unusual large observation as being an outlier with the highest probability as compared to the other observations. These subjects are also identified as outliers using the non-Bayesian approach. We expect to identify other outliers found in Chapter 8 when we generalized the Bayesian approach presented in this chapter to the case of more than one outlier.

CHAPTER 10: CONCLUDING REMARKS

10.1 Summary of study

This study looks at three local medical problems of current interest. Firstly, this study considers the optimal number of lymph nodes to be removed (LNR) for maximizing the survival of breast cancer patients. We investigate the influence of the number of lymph nodes to be removed on survival using chi-square test of independence and Wilcoxon test. We then find the best-fitted logistic and Cox regression models using forward selection and BMA procedures. The models are used to assess the prognostic values of independent factors of survival at all thresholds of the number of LNR. For both types of regression models, we present the use of the Akaike Information Criterion (AIC) instead of Wald statistic (χ^2) in determining the optimal number of LNR that give maximum differential in survival of the breast cancer patients.

Secondly, this study considers the optimal LNR for adequate nodal staging of breast cancer patients. We explore the association of LNR on nodal involvement using chi-square test of independence. With logistic regression analysis, we use similar procedure above to find the best-fitted logistic regression model and determine the optimal LNR for adequate nodal staging of breast cancer patients.

Finally, this study considers the problem of detection outliers in 2×2 crossover design using non-Bayesian and Bayesian frameworks. In the non-Bayesian framework, we consider the classical studentized residual and provide a new studentized residual using median absolute deviation to identify possible outliers. The performances of both procedures in detecting outliers are compared via simulation. In Bayesian framework, since the joint posterior distribution is intractable, we obtain information on the posterior distribution from samples generated by Markov chain Monte Carlo (MCMC)

sampling, in particular with Metropolis-Hastings algorithm to obtain the estimated values of the parameters and to identify the outliers in the data sets.

The methods mentioned above have been applied to the Malaysian Breast Cancer data and kinesiology data, obtained from the University of Malaya Medical Centre (UMMC). This study is able to provide solutions to the problems which are very beneficial to the local medical practitioners. Hence, the findings are very important as guidelines in the surgical management of breast cancer patients and in the usage of kinesiotapes in sports.

10.2 Contributions

The study has contributed in the following ways:

1. We have provided an alternative method of identifying the optimal number of lymph nodes to be removed for maximizing the survival of local breast cancer patients.
2. We have provided an alternative method of identifying the optimal threshold of lymph nodes removed for adequate nodal staging of local breast cancer patients.
3. We have presented the use of studentized residual using median absolute deviation to identify outliers in 2×2 crossover design with application in kinesiology study.
4. We have shown the procedure to identify the outliers in 2×2 crossover design by generating sample using MCMC method with application in kinesiology study.

10.3 Further research

There are various possibilities for further research in these areas. Some suggestions are given as below:

1. To propose Bayesian information criterion (BIC), instead of AIC, as an alternative measure for the χ^2 in determination optimal cut-offs for the issues considered.
2. To propose more advanced group deleted Studentized residuals such as Generalized Studentized Residuals proposed by Imon (2005) and Generalized Pearson Residuals proposed by Imon and Hadi (2008), as the alternative choice for the studentized residual in the work of Chapter 8.
3. To extend the work in Chapter 9 to cases with two or more outliers.
4. To extend the work in Chapter 9 to unknown number of outliers.

LIST OF PUBLICATIONS

ISI Indexed

1. F. P. Lim, I. Mohamed, N. Daud and S. L. Goh. (2016). Comparison of Outlier Detection Methods in Standard 2×2 Crossover Design. *Sains Malaysiana*, 45(3), 499-506.

2. F. P. Lim, N. A. Taib, I. Mohamed and N. Daud. (2014). The Optimal Number of Lymph Nodes Removed in Maximizing the Survival of Breast Cancer Patients, *AIP Conference Proceedings 1605*, 40.

Oral Presentations

1. Presenter, "Determination of The Optimal Threshold of Lymph Nodes Removed for the Adequacy of Nodal Staging for Local Breast Cancer Patients using Bayesian Framework", Simposium Kebangsaan Sains Matematik ke-22, 24 – 26 November 2014, Hotel Grand BlueWave, Shah Alam.

2. Presenter, "The Optimal Number of Lymph Nodes Removed in Maximizing the Survival of Breast Cancer Patients", Simposium Kebangsaan Sains Matematik ke 21 (SKSM-21), 6-8 November 2013, Universiti Sains Malaysia.

3. Presenter, "Outlier Detection in Standard 2×2 Crossover Design Using Bayesian Framework", The 3rd ISM International Statistical Conference, 8-11 August 2016, Institute of Mathematical Sciences, University of Malaya.

Papers Submitted

1. F. P. Lim, I. Mohamed, A. I. N. Ibrahim, N. Daud and S. L. Goh. (2016). Outlier Detection in Standard 2×2 Crossover Design Using Bayesian Framework. *Sains Malaysiana*. Submitted on 23 April 2016.

REFERENCES

Aggarwal, C. C. (2013). *Outlier Analysis*. Springer New York Heidelberg Dordrecht London.

Agresti, A. (1990). *Categorical data analysis*. John Wiley and Sons, New York.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. *In B. N. Petrov and F. Csaki (eds.), Proceedings of 2nd International Symposium on Information Theory, Akademiani Kiado, Budapest*, 267 – 281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC*,*19*, 716-723.

Altman, D.G. (1991). *Practical statistics for Medical Research*. Chapman & Hall, London.

American Joint Committee on Cancer. (1993). *Manual for staging of cancer* (4th ed.). Philadelphia, DC: JB Lippincott.

American Joint Committee on Cancer. (2002). *Cancer staging manual* (6th ed.). DC: Springer-Verlag New York.

Amir, E., Freedman, O.C., Seruga, B., & Evans, G. (2010). Assessing women at high risk of breast cancer: a review of risk assessment models. *Journal of the National Cancer Institute*, *102*, 680-691.

Association of Systematic Kinesiology in Ireland. Retrieved from www.kinesiology.ie/ what-is-kinesiology/

Atkins, H., Hayward, J. L, Klugman, D. J., & Wayte A. B. (1972). Treatment of early breast cancer: a report after ten years of a clinical trial. *British Medical Journal*, *2*, 423-429.

Axelsson, C. K., Mouridsen, H. T. & Zedeler, K. (1992). Axillary dissection of level-I and -II lymph nodes is important in breast cancer classification. *European Journal of Cancer*, *28*, 1415-1418.

Baltaci, G., Irmak, R., & Soylu, A. R. (2011). Acute effects of kinesiotaping on muscular endurance and fatigue by using surface electromyography signals of masseter muscle. *Medicina Sportiva*, *15*, 13-16.

Bembenek, A., & Schlag, P. M. (2000). Lymph-node dissection in breast cancer. *Langenbeck's Archives of Surgery, 385*, 236-245.

Blancas, I., Garcia-Puche, J. L., Bermejo, B., Hanrahan, E. O., Monteagudo, C., Martínez-Agulló, A., Rouzier, R., Hennessy, B. T., Valero, V., & Lluch, A. (2006). Low number of examined lymph nodes in node-negative breast cancer patients is an adverse prognostic factor. *Annals of Oncology*, *17*, 1644-1649.

Brant, R. (2004). *Forward selection*. Retrieved from http://www.stat.ubc.ca/~rollin/teach/643w04/lec/node41.html

Breastcancer.org. (2012). Retrieved from http://www.breastcancer.org/pictures/breast_anatomy/axillary_lymph_nodes

Breastcancer.org. (2015). Retrieved from http://www.breastcancer.org/symptoms/understand _bc/what_is_bc

Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics*, *36*, 69 - 79.

Cabanes, P. A., Salmon, R. J., Vilcoq, J. R., Durand, J. C., Fourquet, A., Gautier, C., & Asselain, B. (1992). Value of axillary dissection in addition to lumpectomy and radiotherapy in early breast cancer. *The Breast Carcinoma Collaborative Group of the Institut Curie*, *The Lancet*, *339*(8804), 1245-1248.

Cady, B., & Sears, H. F. (1986). Usefulness and technique of axillary dissection in primary breast cancer. *Journal of Clinical Oncology*, *4*, 623-624.

Camp, R. L., Rimm, E. B., & Rimm, D. L. (2000). A high number of tumor free axillary lymph nodes from patients with lymph node negative breast carcinoma is associated with poor outcome. *Cancer, 88*(1), 108-113.

Cancer.Net. (2015). Retrieved from http://www.cancer.net/cancer-types/breast-cancer/stages

Carter, C. L., Allen, C., & Henson, D. E. (1989). Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*, *63*(1), 181–187.

Chang, J. C., & Hilsenbeck, S. G. (2010). Chapter 31: Prognostic and predictive markers. In J. R. Harris, M. E. Lippman, M. Morrow, C. K. Osborne (Eds.), *Diseases of the Breast* (4th ed.). Lippincott Williams & Wilkins.

Chaloner K. & Brant R. (1988). A Bayesian approach to outlier detection and residual analysis. *Biometrika*, *75*(4), 651-659.

Chatfield, C. (1995). Model uncertainty, data mining, and statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A*, 158, 419-466.

Chen, Y. I., & Huang, C. S. (2015). Bayesian bioequivalence test based on skew normal model for bioavailability measure. *Journal of the Chinese Statistical Association*, 53, 38-54.

Chow, S. C., & Liu, J. P. (2009). *Design and analysis of bioavailability and bioequivalence studies*. London: Chapman and Hall.

Chow, S. C., & Tse, S. K. (1990). Outliers detection in bioavailability/bioequivalence studies. *Statistics in Medicine*, *9*, 549-558.

Clayton, D., & Leslie, A. (1981). The bioavailability of erythromycin stearate versus enteric-coated erythromycin based when taken immediately before and after food. *Journal of International Medical Research*, *9*, 470-477.

Collett, D. (2003). *Modelling survival data in medical research.* USA: Chapman and Hall/CRC CRC Press LLC.

Cox, D.R. (1972). Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society - Series B*, 34, 187-220.

De Hoyo, M., Álvarez-Mesa, A., Sañudo, B., Carrasco, L., & Domínguez, S. (2013). immediate effect of kinesio taping on muscle response in young elite soccer players. *Journal of Sport Rehabilitation*, *22*, 53-58.

Diab, S. G., Hilsenbeck, S. G., de Moor, C., Clark, G. M., Osborne, C. K., & Ravdin, P. M. (1998). Radiation therapy and survival in breast cancer patients with 10 or more positive axillary lymph nodes treated with mastectomy. *Journal of Clinical Oncology*, *16*(5), 1655-1660.

Diener-West, M. (2008). *Use of the chi-square statistic*. John Hopkins Bloomberg School of Public Health.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B*, *57*, 45-97.

Duraker, N., Bati, B., Demir, D., & Caynak, Z. C. (2011). Prognostic significance of the number of removed and metastatic lymph nodes and lymph node ratio in breast carcinoma patients with 1 – 3 axillary lymph node(s) metastasis. *ISRN Oncology*, 645450. doi:10.5402/2011/645450

Enachescu, D. & Enachescu, C. (2009). *A new approach for outlying records in bioequivalence trials*. Paper presented at the XIII International Conference "Applied Stochastic Models and Data Analysis".

Erbes, T., Orlowska-Volk, M., Zur Hausen, A., Rucker, G., Mayer, S., Voigt, M., Farthmann, J., Iborra, S., Hirschfeld, M., Meyer, P. T., Gitsch, G., & Stickeler, E. (2014). Neoadjuvant chemotherapy in breast cancer significantly reduces number of yielded lymph nodes by axillary dissection. *BMC Cancer*, *14*, 4. doi:10.1186/1471-2407-14-4

Fisher, B., Wolmark, N., Bauer, M., Redmond, C., & Gebhardt, M. (1981). The accuracy of clinical nodal staging and of limited axillary dissection as a determinant of histologic nodal status in carcinoma of the breast. *Surgery, Gynecology & Obstetrics*, *152*, 765-772.

Fisher, B., Redmond, C., Fisher, E. R., Bauer, M., Wolmark, N., Wickerham, D. L., Deutsch, M., Montague, E., Margolese, R., & Foster, R. (1985). Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *The New England Journal of Medicine*, *312*, 674-81.

Fitzgibbons, P. L., Page, D. L., Weaver, D., Thor, A. D., Allred, D. C., Clark, G. M., Ruby, S. G., O'Malley, F., Simpson, J. F., Connolly, J. L., Hayes, D. F., Edge, S. B., Lichter, A., & Schnitt, S. J. (2000). Prognostic factors in breast cancer, College of American Pathologists consensus statement 1999. *Archives of Pathology & Laboratory Medicine*, *124*, 966–78.

Gao, F., He, N., & Wu, P. H. (2014). The number of tumor-free axillary lymph nodes removed as a prognostic parameter for node-negative breast cancer. *Chinese Journal of Cancer*, *33*(11), 569–573.

George EI. (1999). Bayesian model selection. In *Encyclopedia of Statistical Sciences Update 3*. New York, NY: Wiley.

Ghosh, K., Hartmann, L. C., Reynolds, C., Visscher, D. W., Brandt, K. R., Vierkant, R. A., Scott, C. G., Radisky, D. C., Sellers, T. A., Pankratz, V. S., & Vachon, C. M. (2010). Association between mammographic density and age-related lobular involution of the breast. *Journal of Clinical Oncology*, *28*(13), 2207-2212.

Goodman K. J., Correa, P., Tengana Aux, H. J., Ramirez, H., DeLany, J. P., Guerrero Pepinosa, O., Lopez Quinones, M., & Collazos Parra, T. (1996). *Helicobacter pylori* infection in the Colombian Andes: A population-based study of transmission pathways. *American Journal of Epidemiology*, *144*(3), 290-299.

Graversen, H. P., Blichert-Toft, M., Andersen, J. A., & Zedeler, K. (1988). Breast cancer: risk of axillary recurrence in nodal-negative patients following partial dissection of the axilla. *European Journal of Surgical Oncology*, *14*, 407-412.

Groth, S. S., Virnig, B. A., Whitson, B. A., DeFor, T. E., Li, Z. Z., Tuttle, T. M., & Maddaus, M. A. (2010). Determination of the minimum number of lymph nodes to examine to maximize survival in patients with esophageal carcinoma: data from the surveillance epidemiology and end results database. *The Journal of Thoracic and Cardiovascular Surgery*, *139*, 612-620.

Haffty, B. G., & Ward, B. A. (1997). Is breast-conserving surgery with radiation superior to mastectomy in selected patients? *Cancer Journal From Scientific American, 3*(1), 2-3.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*, 383-393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: the approach based on influence functions*. New York: John Wiley.

Hashemi, E., Kaviani, A., Najafi, M., Ebrahimi, M., Hooshmand, H., & Montazeri, A. (2004). Seroma formation after surgery for breast cancer. *World Journal of Surgical Oncology*, *2*, 44. doi: 10.1186/1477-7819-2-44

Hawkins, D. M. (1980). *Identification of outliers*. Chapman and Hall.

Herrinton, L. J., Saftlas, A. F., Stanford, J. L., Brinton, L. A., & Wolfe, J. N. (1993). Do alcohol intake and mammographic densities interact in regard to the risk of breast cancer? *Cancer*, *71*(10), 3029-3035.

Hilbe, J. M. (2009). *Logistic regression models.* United States of America: Taylor & Francis Group, LLC.

Hodges, J. S. (1987). Uncertainty, policy analysis, and statistics. *Statistical Science*, *2*, 259-291.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382-417.

Huang, C. Y., Hsieh, T. H., Lu, S. C., & Su, F. C. (2011). Effect of the Kinesio tape to muscle activity and vertical jump performance in healthy inactive people. *BioMedical Engineering OnLine*, *10*, 70. doi: 10.1186/1475-925X-10-70.

Huitson, A., Poloniecki, J., Hews, R., & Barker, N. (1982). A review of cross-over trials. *Statistician*, *31*, 71-80.

Imon, A. H. R. M. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied Statistics*, *32*(9), 929-946.

Imon, A. H. R. M. & Hadi, A. S. (2008). Identification of multiple outliers in logistic regression. *Communications in Statistics – Theory and Methods*, *37*(11), 1697-1709.

Jones, B. & Kenward, M. G. (1989). *Design and analysis of cross-over trials* (1st ed.). London, UK: Chapman and Hall.

Jones, B. & Kenward, M. G. (2003). *Design and analysis of cross-over trials* (2nd ed.). London, UK: Chapman and Hall.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambrige Philosophical Society 31*, 203-222.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Karasoy, D. & Daghan, G. (2012). Examination of outliers in bioequivalence studies. *Bulletin of Clinical Psychopharmacology*, *22*(4), 307-312.

Kase, K., Wallis, J., & Kase, T. (2003). *Clinical therapeutic applications of the kinesio taping method*. Tokyo, Japan: Ken Ikai Co. Ltd.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Kiricuta, C. I., & Tausch, J. (1992). A mathematical model of axillary lymph node involvement based on 1446 complete axillary dissections in patients with breast carcinoma. *Cancer*, *69*(10), 2496-2501.

Krag, D. N., & Single, R. M. (2003). Breast cancer survival according to number of nodes removed. *Annals of Surgical Oncology, 10*, 1152-1159.

Kuru, B. (2006). Prognostic significance of total number of nodes removed, negative nodes removed, and ratio of positive nodes to removed nodes in node positive breast carcinoma. *European Journal of Surgical Oncology*, *32*, 1082-1088.

Kyngas, H., & Rissanen, M. (2001). Support as a crucial predictor of good compliance of adolescents with a chronic disease. *Journal of Clinical Nursing*, *10*(6), 767-774.

Lam, P. B., Vacek, P. M., Geller, B. M., & Muss, H. B. (2000). The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer*, *89*(2), 369-375.

Langlands, A. O., Prescott, R. J., & Hamilton, T. (1980). A clinical trial in the management of operable cancer of the breast. *British Journal of Surgery*, *67*, 170-174.

Leamer, E. E. (1978). *Specification Searches*. New York : Wiley.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766.

Lim, G. C. C., Rampal, S. & Yahya, H. (2008). *In cancer incidence in Peninsular Malaysia, 2003-2005*. Kuala Lumpur: National Cancer Registry.

Liu, J. P., & Weng, C. S. (1991). Detection of outlying data in bioavailability/ bioequivalence studies. *Statistics in Medicine*, *10*, 1375-1389.

Luzar-Stiffler, V. & Stiffler, C. (2005). Higher-order crossover design outlier detection. *Proceedings of the 27th International Conference on Information Technology Interfaces. Zagreb: University of Zagreb*, 650-655.

Madigan, D. & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, *89*, 1535-1546.

Magee-Womens Hospital of University of Pittsburgh Medical Center, UPMC. (2015). Retrieved from http://www.upmc.com/locations/hospitals/magee/services/magee-womens-cancers/breast-cancer-program/treatment/breast-cancer-surgery/pages/axillary-lymph-node.aspx

Martin, L. J., Melnichouk, O., Guo, H., Chiarelli, A. M., Hislop, T. G., Yaffe, M. J., Minkin, S., Hopper, J. L., & Boyd, N. F. (2010). Family history, mammographic density, and risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *19*(2), 456-463. doi: 10.1158/1055-9965.EPI-09-0881

Mayo Foundation for Medical Education and Research. (2015). *Disease and conditions: breast cancer*. Retrieved from http://www.mayoclinic.org/diseases-conditions/breast-cancer/multimedia/tumor-size/img-20006260

McDonald, J. H. (2009). *Handbook of biological statistics.* United States of America: Sparky House Publishing.

Milleron, O., Pilliere, R., Foucher, A., de Roquefeuil, F., Aeqerter, P., Jondeau, G., Raaffestin, B. G., & Dubourg, O. (2004). Benefits of obstructive sleep apnoea treatment in coronary artery disease: a long-term follow-up study. *European Heart Journal*, *25*, 728-734.

Montgomery, J. & Nyhan, B. (2010). Bayesian model averaging: theoretical developments and practical applications. *Political Analysis*, *18*(2), 245-270.

Moulton, B. R. (1991). A Bayesian approach to regression selection and estimation with application to a price index for radio services. *Journal of Econometrics*, *49*, 169-193.

Mustafi, C. K. (1991). The median absolute deviation of residual life data and weak convergence of a related stochastic process. *Journal of Statistical Planning and Inference*, *27*, 231-236.

National Cancer Institute. (2015). *Breast cancer treatment: general information about breast cancer*. Retrieved from http://www.cancer.gov/types/breast/patient/breast-treatment-pdq#link/stoc_h2_1

National Comprehensive Cancer Network (NCCN) Guidelines for Patients Stage I and II Breast Cancer, Version 1. (2014). Retrieved from http://www.nccn.org/ patients/ guidelines/stage_i_ii_breast/index.html#30

National Comprehensive Cancer Network (NCCN) Guidelines for Patients Stage III Breast Cancer, Version 1. (2014). Retrieved from http://www.nccn.org/ patients/guidelines/ stage_iii_breast/index.html#37/z

Nawama, M., Ibrahim, A. I. N., Mohamed, I., Yahya, M. S., & Taib, N. A. (2015). Outlier detection using generalized linear model in Malaysian breast cancer data. *Sains Malaysiana, 44*(10), 1417-1422.

Orr, R. K. (1999). The impact of prophylactic axillary node dissection on breast cancer survival - a Bayesian meta-analysis. *Annals of Surgical Oncology*, *6*, 109-116.

Parkes, K. R. (1982). Occupational stress among student nurses: a natural experiment. *Journal of Applied Psychology*, *67*, 784 -796.

Pettit, L.I. (1988). Bayesian methods for outliers in exponential samples. *Journal of the Royal Statistical Society, Series B*, 50, 371-380.

Peyre, C. G., Hagen, J. A., DeMeester, S. R., Altorki, N. K., Ancona, E., Griffin, S. M., Hölscher, A., Lerut, T., Law, S., Rice, T. W., Ruol, A., van Lanschot, J. J., Wong, J., & DeMeester, T. R. (2008). The number of lymph nodes removed predicts survival in esophageal cancer: An international study on the impact of extent of surgical resection. *Annals of Surgery*, *248*, 549-556.

Polit, D. (1996). *Data analysis and statistics for nursing research*. Stamford, Connecticut: Appleton & Lange.

Raftery, A. E. (1988). Approximate Bayes factors for generalized linear models. *Technical Report 121*, University of Washington, Department of Statistics.

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika*, *83*, 251-266.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179-191.

Raghavarao, D. (1989). Crossover designs in industry, in design and analysis of experiments, with applications to Engineering and Physical Sciences (ed. S. Gosh). New York: Marcell Dekker.

Ramsay, T., & Elkum, N. (2005). A comparison of four different methods for outlier detection in bioequivalence studies. *Journal of Biopharmaceutical Statistics*, *15*(1), 43-52.

Rizk, N. P., Ishwaran, H., Rice, T. W., Chen, L. Q., Schipper, P. H., Kesler, K. A., Law, S., Lerut, T. E., Reed, C. E., Salo, J. A., Scott, W. J., Hofstetter, W. L., Watson, T. J., Allen, M. S., Rusch, V. W., & Blackstone, E. H. (2010). Optimum lymphadenectomy for esophageal cancer. *Annals of Surgery*, *251*, 46-50.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association: Theory and Methods*, *88*(424), 1273-1283.

Ruppert, D. (2010). *Statistics and Data Analysis for Financial Engineering*. Springer.

Sakorafas, G. H., Tsiotou, A. G., Balsiger, B. M. (2000). Axillary lymph node dissection in breast cancer-current status and controversies, alternative strategies and future perspectives. *Acta Oncologica*, *39*, 455-466.

Sanghani, M., Balk, E. M., & Cady, B. (2009). Impact of axillary lymph node dissection on breast cancer outcome in clinically node negative patients. *Cancer*, *115*, 1613-1620.

Sankaranarayanan, R., Swaminathan, R., Jayant, K., & Brenner, H. (2011). An overview of cancer survival in Africa, Asia, the Caribbean and Central America: the case for investment in cancer health services. *IARC Scientific Publications*, *162*, 257-291.

Saxena, N., Hartman, M., Yip, C. H., Bhoo-Pathy, N., Khin, L. W., Taib, N.A., Looi, L. M., Lim, S. E., Lee, S. C., & Verkooijen, H. M. (2012). Does the axillary lymph node ratio have any added prognostic value over pN staging for South East Asian breast cancer patients? *PLoS One, 7*(9), e45809.

Senn, S. (1993). *Cross-over trials in clinical research*. New York: John Wiley & Sons.

Shotwell, M. S., & Slate, E. H. (2011). Bayesian outlier detection with dirichlet process mixtures. *Bayesian Analysis*, *6*(4), 665-690.

Singh, R., Namdev, K. K. & Chilkoti, D. (2014). An assessment on handling of outlier subjects in bioequivalence study - a review. *Journal of Pharmaceutical and Bioanalytical Science*, *3*(1), 14-20.

Singletary, S. E., Allred, C., Ashley, P., Bassett, L. W., Berry, D., Bland, K. I., Borgen, P. I., Clark, G., Edge, S. B., Hayes, D. F., Hughes, L. L., Hutter, R. V., Morrow, M., Page, D. L., Recht, A., Theriault, R. L., Thor, A., Weaver, D. L., Wieand, H. S., & Greene, F. L. (2002). Revision of the American Joint Committee on Cancer staging system for breast cancer. *Journal of Clinical Oncology*, *20*(17), 3628-3636.

Somner, J. E., Dixon, J. M., Thomas, J, S. (2004). Node retrieval in axillary lymph node dissections: recommendations for minimum numbers to be confident about node negative status. *Journal of Clinical Pathology*, *57*(8), 845–848.

Song, Y. J., Shin, S. H., Cho, J. S., Park, M. H., Yoon, J. H., & Jegal, Y. J. (2011). The Role of Lymphovascular Invasion as a Prognostic Factor in Patients with Lymph Node-Positive Operable Invasive Breast Cancer. *Journal of Breast Cancer*, *14*(3), 198-203.

Sosa, J. A., Diener-West, M., Gusev, Y., Choti, M. A., Lange, J. R., Dooley, W. C., & Zeiger, M. A. (1998). Association between extent of axillary lymph node dissection and survival in patients with stage I breast cancer. *Annals of Surgical Oncology*, *5*(2), 140-149.

Stauffer, H. B. (2008). *Contemporary Bayesian and Frequentist Statistical Research Methods for Natural Resource Scientists.* New Jersey: John Wiley & Sons, Inc.

Stephenson, N., Beckmann, L., Chang-Claude, J. (2010). Carcinogen metabolism, cigarette smoking, and breast cancer risk: a Bayes model averaging approach. *Epidemiologic Perspectives & Innovations*, *7*, 10. doi:10.1186/1742-5573-7-10

Taib, N. A., Yip, C. H., & Mohamed, I. (2008). Survival analysis of Malaysian women with breast cancer: results from the University of Malaya Medical Centre. *Asian Pacific Journal of Cancer Prevention*, *9*, 197-202.

Taib, N. A., Akmal, M. N., Mohamed, I., & Yip, C. H. (2011). Improvement in survival of breast cancer patients - trends in survival over two time periods in a single institution in an Asia Pacific country, Malaysia. *Asian Pacific Journal of Cancer Prevention*, *12*, 345-349.

Talbott, C. (2008). After Olympic exposure, kinesio tape sticks around. *The Associated Press*. Retrieved from http://nbcsports.msnbc.com/id/26781777

Ujino, A., Eberman, L. E., Kahanov, L., Renner, C., & Demchak, T. (2013). The Effects of Kinesio Tape ad Stretching on Shoulder ROM. *International Journal of Athletic Therapy & Training*, 18, 24-28.

Unnikrishnan, N. K. (2010). Bayesian analysis for outliers in survey sampling. *Computational Statistics and Data Analysis*, 54, 1962-1974.

Van Bemmel, A. J. M., van de Velde, C. J. H., Schmitz, R. F., & Liefers, G. J. (2011). Prevention of seroma formation after axillary dissection in breast cancer: a systematic review. *European Journal of Surgical Oncology*, *37*, 829-835.

van der Wal, B. C. H., Butzelaar, R. M. J. M., Van der Meij, S., & Boermeester, M. A. (2002). Axillary lymph node ratio and total number of removed lymph nodes: predictors of survival in stage I and II breast cancer. *European Journal of Surgical Oncology*, *28*, 481-489.

Wang, W., & Chow, S. C. (2003). Examining outlying subjects and outlying records in bioequivalence trials. *Journal of Biopharmaceutical Statistics*, *13*(1), 43-56.

Weiner, D. (1989). *Bioavailability. Notes on the training course for new clinical statisticians* sponsored by the Biostatistics Subsections of Pharmaceutical Manufacture Association, March, 1989, Washington, DC.

Weir, L., Speers, C., Dyachkova, Y., & Olivotto, I. A. (2002). Prognostic significance of the number of axillary lymph nodes removed in patients with node-negative breast cancer. *Journal of Clinical Oncology*, *20*(7), 1793-1799.

Whitlock, J. (2014). *Lymph nodes: understanding the function of lymph nodes*. Retrieved from http://surgery.about.com/od/glossaryofsurgicalterms/g/Lymph Node.htm

William, G., & Zikmund, B. J. (2010). *Business Research Methods*. Canada: Cengage Learning.

Williams, I., Whatman, C., Hume, P. A., & Sheerin, K. (2012). Kinesio taping in treatment and prevention of sports injuries. *Sports Medicine*, *42*, 153-164.

Woodworth, P. A., McBoyle, M. F., Helmer, S. D., & Beamer, R. L. (2000). Seroma formation after breast cancer surgery: incidence and predicting factors. *The American Surgeon*, 66, 444-457.

Wozny, N. (2009). Stuck on it: Kinesio tape has won a following among dancers. *Your Body*, *Dance Magazine*, 26.

Wu, S. G., Chen, Y., Sun, J. Y., Li, F. Y., Lin, Q., Lin, H. X., & He, Z. Y. (2013). Using the lymph nodal ratio to predict the risk of locoregional recurrence in lymph node-positive breast cancer patients treated with mastectomy without radiation therapy. R*adiation Oncology*, *8*, 119. doi: 10.1186/1748-717X-8-119

Yip, C. H., Nur Aishah Mohd Taib., & Ibrahim Mohamed. (2006). Epidemiology of breast cancer in Malaysia. *Asian Pacific Journal of Cancer Prevention*, *7*, 369-374.