LANCASTER UNIVERSITY

Modelling of Survival and Incidence for Colorectal Cancer in Malaysia

by

Anis Kausar Ghazali

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Health and Medicine

Lancaster Medical School

April 2018

Declaration of Authorship

I, ANIS KAUSAR GHAZALI, declare that this thesis titled, 'Modelling of Survival and Incidence for Colorectal Cancer in Malaysia' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

LANCASTER UNIVERSITY

Abstract

Faculty of Health and Medicine
Lancaster Medical School

Doctor of Philosophy

by Anis Kausar Ghazali

Colorectal cancer is the third most common cancer worldwide, with an estimated 1.36 million new cases and 694,000 deaths recorded in 2012. This number is expected to increase by 80% reaching approximately 2.4 million cases in 2035, and contributing to 1.3 million deaths globally per year.

In Malaysia, colorectal cancer is the second most common cancer, in both men and women, behind lung cancer in men and breast cancer in women respectively. The most recent estimate gives the incidence of colorectal cancer in Malaysia as 21.3 cases per 100,000 population and 9.8 deaths per 100,000 population for the six years period reported between 2008 and 2013. The incidence of colorectal cancer is known to vary by place and by time period. However, most research on the spatial and variation in cancer incidence has taken place outside South East Asia, and such research on this is still scarce in Malaysia. The research presented in this thesis investigates individual factors affecting survival, and the spatial variation in incidence and survival of colorectal cancer in Malaysia. This has not been done before in Malaysia.

There are three objectives addressed in this thesis: (1) to investigate the individual characteristics that affect survival of colorectal cancer in Malaysia, (2) to investigate and model the spatial variation of survival in colorectal cancer in Malaysia, (3) to investigate and model the spatial variation in the incidence of colorectal cancer in Malaysia.

The research was involved 4412 of colorectal cancer patients in Malaysia with histologically verified primary colorectal cancer who were diagnosed between 2008 and 2013 (ICD-10, C18-C20), recorded in the database of National Cancer Patient Registry- Colorectal Cancer (NCPR-CC) Malaysia. We investigated the effect of individual characteristics such as age, gender, education as well as clinical characteristics such as cancer staging, cancer site and treatment modalities on survival prognosis after a diagnosis of colorectal cancer using a Cox regression model. The analysis was then been extended to model the spatial variation in survival for colorectal cancer patients in Malaysia, accounting for individual and socioeconomic characteristics using a spatial survival model. We then applied a Generalized Linear Mixed Effects model, which is derived from the log-Gaussian Cox Process, in order to model the incidence of colorectal cancer in Peninsular Malaysia.

Our findings show that the severity of disease at diagnosis as measured by cancer staging, tumour grading and the presence of distant metastases, plays an important role in the prognosis of patients with colorectal cancer in Malaysia, and that this remains even after controlling for spatial correlation on space. Our research allows us to shows the geographical variation in survival of colorectal cancer in Malaysia, and what variation persists once individual and socioeconomic characteristics are taken into account. Our model that developed to predict the colorectal cancer incidence found that that some places had greater risk of an incidence exceeding the national average. The map for probability of exceedance relative risk of colorectal cancer incidence in the North West of Peninsular Malaysia shows variation in the risk for colorectal cancer cases across the region. We noted that town areas are highly likely to exceed the threshold of relative risk of increased number of colorectal cancer cases, and this effect is present even though we account for the additional population there.

Spatial variation in survival and incidence of colorectal cancer in Malaysia needs to be investigated further. To the best of our knowledge, this is the first research that uses spatial modelling to identify potential factors affecting the incidence and survival for colorectal cancer in Malaysia, as well as to map the risks in survival and incidence. Our findings can help public health authorities to plan better management of the resources used to prevent and treat this disease.

Acknowledgements

I would like to express my utmost gratitude to the following individuals who have supported me throughout my study:

My supervisors, Dr Benjamin Taylor and Dr Thomas Keegan, for their invaluable academic support throughout my study and for the knowledge they have imparted which I will bring back to use to serve my country.

My beloved husband and children for believing in me. Their love gave me the strength to face all the challenges. I know that no matter how tough things get, I can count on them.

My parents and siblings, even though we are miles apart, we are never far.

I would also like to thank my sponsors, the Ministry of Higher Education and Universiti Sains Malaysia, Malaysia.

Last but not least, I would like to thank my coursemates and Malaysian friends in Lancaster for sharing with me and my family times that have given us so many wonderful memories to cherish and for forging a bond that will endure.

Contents

List of Figures

List of Tables

Abbreviations

ASR	**A**ge **S**tandardised **R**ate
CRC	**C**olo**R**ectal **C**ancer
ICD-10	**I**nternational **C**lassification of **D**iseases, Tenth Edition
MCMC	**M**arkov **C**hain **M**onte **C**arlo
NCR	**N**ational **C**ancer **R**egistry
NCPR	**N**ational **C**ancer **P**atient **R**egistry
PH	**P**roportional **H**azards
SES	**S**ocio **E**conomic **S**tatus

Chapter 1

Study setting and population

This chapter describes the study setting and data source of this research. We begin by explaining the rationale of the study, then state our aims and objectives and outline the design of the study. Next we give an overview of the background of our research population and explain the source of our data; we lastly detail the ethical clearance of this research.

1.1 Rationale of the study

One of the major aims of population-level cancer research is to improve survival prognoses for patients. A patient's survival may depend on several known and unknown factors. Factors influencing the survival and incidence of colorectal cancer could be individual characteristics and/or environmental attributes of the places where patients live. While environmental and social factors affecting the variation in cancer incidence and survival rates have been broadly investigated, research specific to Malaysia is still scarce.

To address this issue, we will use advanced techniques in spatial survival analysis to model and interpret the spatial variation in colorectal cancer survival prognosis

across the whole of Malaysia, and evaluate the role of individual, environmental and socioeconomic risk factors within this context. We will use these spatial models to investigate the potential factors affecting incidence, and predict the incidence of colorectal cancer in Peninsular Malaysia. Our study will also derive risk maps from the models.

To the best of our knowledge, this is the first cancer research in Malaysia that utilises spatial modelling methods and maps incidence and survival risks, thus it would represent a novel contribution to the scientific study of colorectal cancer there.

The output in this part of the proposed research when disseminated through peer-reviewed journal articles (currently in preparation), conferences and presentations, will be of great interest to public health practitioners and epidemiologists in Malaysia and more widely, especially those involved in providing effective intervention and preventive health programmes.

1.2 Objectives of the study

1. To investigate the individual characteristics that affect survival in colorectal cancer patients in Malaysia,

2. To investigate and model the spatial variation in survival of colorectal cancer in Malaysia,

3. To investigate and model the spatial variation in incidence of colorectal cancer in Peninsular Malaysia.

1.3 Study Design

This study was a retrospective record review study where the data was sourced from the database maintained by the Malaysian National Cancer Patient Registry (NCPR). The inclusion criteria for our study were as follows :

- All patients with histologically verified primary colorectal cancer who were diagnosed between 2008 and 2013 (ICD-10, C18-C20) in the database of National Cancer Patient Registry - Colorectal Cancer (NCPR-CC),
- Malaysian citizenship.

1.4 Study population and data sources

1.4.1 Background of Malaysia

FIGURE 1.1: Malaysia by states (Department of Statistics Malaysia, 2016)

Malaysia is a country in South East Asia located between 2 and 7 degrees north of the Equator. It comprises 13 states and 3 Federal Territories, namely: Selangor,

Perak, Kedah, Kelantan, Perlis, Penang, Terengganu, Pahang, Johor, Malacca, Negeri Sembilan, and the Federal Teritory of Kuala Lumpur and Putrajaya in Peninsular Malaysia, while Sabah, Sarawak and the Federal Territory of Labuan are in East Malaysia (Figure 1.1). To the North of Peninsular Malaysia is Thailand, while its Southern neighbour is Singapore (Department of Statistics Malaysia, 2016).

Peninsular Malaysia is separated from the states of Sabah and Sarawak (East Malaysia or the Borneo island) by the South China Sea. Sarawak is located in the southwest of the island of Borneo and Sabah is located in the north. The Federal Territory of Labuan is an island located west of Sabah. Sabah and Sarawak bounded to the south by Indonesia. Sarawak also shares a border with Brunei. The states of Malaysia are divided into 144 administrative districts and 837 sub-districts, also known as 'mukim' for the whole of Malaysia.

The population of Malaysia in 2016 was estimated to be 31.6 million, comprising 16.3 million men and 15.3 million women. Among Malaysian citizens, ethnic Bumiputra (Malay and indigenous) made up the greatest percentage of the population with 68.6%, followed by Chinese (23.4%), Indian (7.0%) and Other Nationalities (1.0%). The percentage of the Malaysian population who are non-citizens was 10.3% (Department of Statistics Malaysia, 2016).

1.4.2 Healthcare system in Malaysia

Healthcare in Malaysia is provided by two sectors; the government sector and the private sector. There are two types of hospitals run by the government. Public hospitals, which are run by the Ministry of Health and university hospitals run by the Ministry of Higher Education. University hospitals are affiliated to universities. University hospitals provide healthcare service to the public as well as serve as teaching hospitals for the students from the universities they are affiliated with.

The Malaysian healthcare system mainly consists of tax-funded government-run universal services. The public sector health services are organized under a civil service structure and are centrally administered by the Ministry of Health. In Malaysia, the public sector provides about 82% of in-patient care and the private sector provides about 18% of in-patient care.

The private health sector provides health services, mainly in urban areas, through physician clinics and private hospitals with a focus on curative care. Private companies run diagnostic laboratories and ambulance service (Safurah et al., 2013).

The cost for a visit to the government or university outpatient clinic s at polyclinics, district hospitals and tertiary hospitals is just 1.00 RM (Ringgit Malaysia) for non-specialist treatment and RM 5.00 to RM 30.00 for specialist treatment. No charge for patients admitted in 3rd class ward, RM 5.00 charged for 2nd class and RM 10.00 for 1st class. University hospitals also have the same fees as they are also public hospitals but under Ministry of Higher Education . All these charges are applicable for Malaysians only. Other costs for special services and investigations can be found in the Ministry of Health website (Ministry of Health Malaysia, 2013a)

1.4.3 Primary care referral for symptoms of colorectal cancer carcinoma in Malaysia

At present, there is no national population-based screening for colorectal cancer in Malaysia. Patients who initiate a health check or present with sypmtoms are screened and this is mainly done in the primary care setting in any government or university outpatient clinic or polyclinic, district hospitals or tertiary hospitals. In an attempt to improve early detection of colorectal cancer, various efforts have been made to establish colorectal cancer screening programme nationally in primary care setting.

The first clinical practice guideline on colorectal cancer management has been produced by (Ministry of Health Malaysia, 2017) consists of a comprehensive strategy of screening, diagnosis, staging, appropriate treatment and follow-up. Based on clinical practice guidelines management of colorectal cancer, the screening of colorectal cancer should be offered at the age of 50 years and continue until 75 years old for average risk population. Patients that comes to the health centres will be assessed for any symptoms of colorectal cancer. Patients with symptoms will be examined further with certain criteria and will be referred for colonoscopy (urgent referral within two weeks or referred for elective colonoscopy). This depends on the clinical presentation during examination. For those without symptom, they will be stratified based on risk assessment based on the family history as average risk, moderate risk and high risk. Average risk group will undergo Immunofaecal Occult Blood Testing (IFOBT) and positive results will be referred for colonoscopy testing. If negative , they will be suggested to repeat IFOBT yearly. On the other hand, those with moderate and high risk family history of colorectal cancer will straight away be referred for colonoscopy. As for now, this is the guideline used by the Ministry of Health Malaysia for assessing colorectal cancer in Malaysia.

1.4.4 Collection of CRC Data in Malaysia

There are two national cancer registries in Malaysia: the National Cancer Registry (NCR) and the National Cancer Patient Registry (NCPR). Both are managed by the Ministry of Health; the NCR is administered by the Disease Control Division and the NCPR by the National Institute of Health (Ministry of Health Malaysia, 2013b, National Cancer Patient Registry, 2009).

The NCR captures data on diagnoses from all regions in Malaysia. Diagnoses are reported to a state registry and from thence to the National Registry. However, reporting of cases to the state registries from hospitals is voluntary and therefore

is not always completed. The Registry is not passive, though; it conducts active case finding and routine checks. Assessment of the completeness of registration in the NCR is difficult because it is not clear how many of the 165 main hospitals in Malaysia are sending records to the registry, or how accurately diagnoses have been recorded even when they were sent (Omar and Ibrahim Tamin, 2011).

The NCPR collects data on registrations of cancer from participating sites. These participating sites are 34 hospitals that diagnose and treat cancer patients in Malaysia. The objectives of the Registry are to describe the natural history of cancers and to determine the effectiveness of treatments, to monitor safety, and to evaluate access to treatments. The Registry collects data on four cancers: colorectal cancer, blood cancers, breast cancer and nasopharyngeal cancers. The Registry records diagnoses and collects clinical data on risk factors, treatments and patient outcomes. This makes the NCPR data useful for research into the effects of treatments and survival from cancers (Hassan et al., 2014).

This study using the data from the National Cancer Registry-Colorectal Cancer (NCPR-CRC). This registry was funded by the Ministry of Health Malaysia.Registration of cases were made through the online website (https://app.acrm.org.my/ccd). All data were stored in a secure database with restricted access to authorised persons only. For each source data provider, different persons are responsible for data entry. It may consist of surgeon/Gastroenterologist, research officers, medical officers, and also trained allied health personnel (nurses and medical assistants).

This registry was registered and approved in the National Medical Research Register and was ethically approved by the Medical Research Ethics Board, Malaysia in 2007. It conforms to the Declaration of Helsinki which protects and uphold the rights of all subjects in research.

For data quality control, NCPR-CRC continuously cross check their data with National Cancer Registry (NCPR?S mother registry) to capture any missing cases.

They also do cross checking with pathology department from respective Hospitals. Data cleaning was also conducted time to time depending on the necessity.

Appendix 1.4.4.1 stated the data fields collected by NCPR-CRC. Date of diagnosis of colorectal cancer refers to the first date of confirmed biopsy of the patients from pathology report. If pathology report is unavailable, then clinical basis will be used for confirmation. Meanwhile, treatments data refer to any kind of treatments that the patients received after he/she was confirmed diagnosis of colorectal cancer. Any death in colorectal cancer patients were cross checked and confirmed with Jabatan Pendaftaran Negara (National Registration Department, Ministry of Home Affairs) by using identification number of the patients.

NCPR-CRC estimates that the overall completeness of the data was around 70%. However, how does the completeness of data varies across areas, race, age group or other variables were not identified at the present.

1.4.4.1 Ethical Clearance

The colorectal cancer data analysed in this study is sourced from the database recorded by the Malaysian National Cancer Patient Registry - Colorectal Cancer (NCPR-CC). We obtained ethical clearance from the Ministry of Health Medical Research Ethical Committee (MREC), Malaysia, the Economic Planning Unit (EPU), Malaysia, and from the Faculty of Health and Medicine Research Ethics Committee (FHMREC), Lancaster University.

The main ethical consideration in this study is that of access to the patient's confidential medical data as recorded in the database. Data collection was anonymised and was kept confidential by the researcher. In order to get a good level of spatial resolution whilst preserving anonymity, only the name of patients 'mukim' and postcode (in Malaysia, the postcodes cover much larger areas than in the UK) was extracted from the address. We stored a copy of the original data in

a secure encrypted network research folder set up by Lancaster University's ISS team. The principal researcher's version of the data was stored in a personal file space electronically encrypted with a security password. The data was transferred electronically from the National Cancer Patient Registry to Lancaster University in encrypted form using the University's ZendTo secure file transfer service (http://zendto.lancs.ac.uk/).

This thesis presents research done to investigate the factors affecting, and spatial variation in, the survival and incidence of colorectal cancer in Malaysia. Here I give an overview of the thesis' structure.

We have divided the thesis into six chapters. The first chapter, Chapter 1, explains the setting of our study, the population and data chosen for this research, and states our investigative objectives. Then we present a review of the epidemiological literature concerning colorectal cancer and its incidence and survival in Chapter 2.

The three objectives in this study are presented individually in Chapters 3, 4 and 5. Each of these chapters includes introductory, methods, results and discussion sections.

Finally, in Chapter 6, we provide a general summary of the findings of all objectives investigated in this thesis, as well as a discussion and some concluding comments on these findings. We also suggest areas that have potential for further epidemiological and methodological research. References and appendices are organised on a chapter by chapter basis.

NATIONAL CANCER PATIENT REGISTRY - COLORECTAL
A. Patient Details and Demographic Details (1)

For Office Use only:
ID: /
Centre:

Instruction: Where check boxes ☐ are provided, check (√) one or more boxes. Where radio buttons ◉ are provided, check (√) one box only.

Centre Code: **Or Reporting centre name:** **Follow-up status:** ◉ New patient ◉ Existing patient on follow-up

Date of Notification (dd/mm/yyyy): / /

SECTION 1 : PATIENT DETAILS & DEMOGRAPHICS

1. Name : * *(Please print in capital letters)*

2. NRIC : *
MyKad/ MyKid: - - Old IC:
Other ID document No:
Specify document type (if others): ◉ Registration number ◉ Birth Certificate ◉ Hospital RN ◉ Date of Birth ◉ Police ID Card ◉ Pension card ◉ Passport ◉ Mother's I/C ◉ Armed Force ID ◉ Lab number ◉ Unregistered ◉ Drivers Licence ◉ Father's I/C ◉ Work Permit # ◉ Patient ID ◉ Others, please specify:

3. Address:
Postcode: Town / City:
State : ◉ Johor Darul Takzim ◉ Pahang Darul Makmur ◉ Sarawak ◉ Wilayah Persekutuan Labuan, Sabah ◉ Kedah Darul Aman ◉ Perak Darul Ridzuan ◉ Selangor Darul Ehsan ◉ Wilayah Persekutuan Putrajaya, Selangor ◉ Kelantan Darul Naim ◉ Perlis Indera Kayangan ◉ Terengganu Darul Iman ◉ Melaka ◉ Pulau Pinang ◉ Wilayah Persekutuan Kuala Lumpur ◉ Negeri Sembilan Darul Khusus ◉ Sabah ◉ Foreigner, please specify:

4. Contact number: Homephone: - H/P: -

5. Date of Birth * (dd/mm/yyyy) : / / ☐ Estimated / presumed year (autofill if MyKad is available)
If the exact date is not known, please enter 01/07/yyyy & check the estimated / presumed year box

6. Age: (autocalculated)

7. Gender * ◉ Male ◉ Female

8. Ethnic group: *
◉ Malay ◉ Orang Asli ◉ Murut ◉ Iban ◉ Kedayan ◉ Chinese ◉ Melanau ◉ Bajau ◉ Dusun ◉ Others, please specify: ◉ Indian ◉ Bidayuh ◉ Kadazan

SECTION 2 : EDUCATION, SOCIAL RISK FACTOR

1. Education level : *
◉ No Education ◉ Secondary ◉ Primary ◉ Tertiary

2. Social Risk Factor - Smoking Status: *
◉ Never ◉ Current ◉ Former (quit > 30 days)

SECTION 3 : SYMPTOM / MEDICAL HISTORY

1.Symptoms: *
☐ Blood in the stool / PR Bleeding ☐ Intestinal obstruction ☐ Alteration of bowel habit ☐ Anemia ☐ Abdominal pain ☐ Weight loss ☐ Others, please specify:

2. Medical History:
1) Diabetes Mellitus ◉ Yes ◉ No

3. Past History of other cancer(s):
* a) ◉ Yes ◉ No
(b) Please specify the type:-
☐ i. Colorectal ☐ ii. Endometrial ☐ iii. Gastric ☐ iv. Small Bowel ☐ v. Hepatobiliary ☐ vi. Urinary tract ☐ vii. Ovarian ☐ Others, please specify:

SECTION 4 : Family History (First Degree Relatives)

1) Colorectal Cancer
◉ Yes → **i. Number of relatives:**
◉ No

2) Other Cancer
◉ Yes → **i. Number of relatives:**
◉ No
ii. Type of cancer: ☐ Lung ☐ Ovary ☐ Endometrium ☐ Breast ☐ Prostate ☐ Others, please specify:

NATIONAL CANCER PATIENT REGISTRY - COLORECTAL
B1. Solid Tumours

For Office Use only:
ID: /
Centre:

Instruction: Where check boxes ☐ are provided, check (√) one or more boxes. Where radio buttons ◉ are provided, check (√) one box only.

I. Patient Name and NRIC Number: *(Patient identifier for paper CRF)* **II. Centre Code:**

SECTION 5 : PRIMARY DIAGNOSIS

(Autofill as Date First Diagnosis but editable)

1. Date of diagnosis: * *(dd/mm/yy)* / /
☐ check if it is Date First Diagnosis ☐ check if it is synchronous
☐ check if it is Date of Metachronous

2. Basis of diagnosis: *(Autofill as Histology of primary tumour but editable)*
☐ Death Certificate only ☐ Clinical Investigations ☐ Cytology ☐ Histology of primary tumour
☐ Clinical ☐ Specific tumour markers ☐ Histology of metastasis ☐ Others, please specify:

3. Primary Cancer Site: *

3.1 Malignant neoplasms based on ICD10:
(Malignant neoplasms based on ICD10 : C18 for Colon, C20 for Rectum, C19 for Rectosigmoid junction)

3.2 Please specify site:-
☐ Caecum ☐ Transverse colon ☐ Rectosigmoid
☐ Ascending colon ☐ Splenic flexure ☐ Rectum
☐ Hepatic flexure ☐ Descending colon ☐ Anorectal
☐ Sigmoid colon ☐ Colon, unspecified

SECTION 6: FINAL STAGING

1. Final Tumour Staging:
a) Stage T: ◉ T1 ◉ T2 ◉ T3 ◉ T4
b) Stage N: ◉ N0 ◉ N1 ◉ N2
c) Stage M: ◉ M0 ◉ M1
d) TNM Staging: ◉ 0 ◉ I ◉ IIA ◉ IIB ◉ IIIA ◉ IIIB ◉ IIIC ◉ IV ◉ Not staged *(autoclassify)*

SECTION 7 : TREATMENT MODALITIES

1. Treatment Modalities:
☐ No Therapy ☐ Radiotherapy ☐ Chemotherapy and Biological Therapy ☐ Supportive care / palliative care
☐ Surgery ☐ Endoscopy ☐ Complementary / Alternative Treatment

2. Surgery Method:
◉ Open ◉ Laparoscopic

3. Operation performed: *
◉ Right Hemicolectomy ◉ Proctocolectomy ◉ Hartmann's Procedure ◉ Ultra Low AR
◉ Extended Right Hemicolectomy ◉ Sub total colectomy ◉ Local excision ◉ Pelvic exenteration
◉ Transverse colectomy ◉ Laparotomy only ◉ Others, please specify:
◉ Left Hemicolectomy ◉ Loop colostomy ◉ High AR
◉ Sigmoid colectomy ◉ APR ◉ Low AR

4. Name of protocol / Regimen *
◉ MAYO's ◉ FOLFIRI & Bevacizumab
◉ De Grammont ◉ FOLFIRI & Cetuximab
◉ FOLFIRI ◉ FOLFOX & Bevacizumab
◉ FOLFOX ◉ FOLFOX & Cetuximab
◉ Capecitabine (Xeloda) ◉ XELOX & Bevacizumab
◉ Uracil and Tegafur (UFT) ◉ XELOX & Cetuximab
◉ XELOX (Xeloda & Oxaliplatin) ◉ Others, please specify:

SECTION 8: PATHOLOGY

1. Staging procedure:- ◉ Biopsy specimen (if checked, TNM staging not needed) ◉ p TNM ◉ yp TNM (for patients who had received preoperative neoadjuvant chemo/radiotherapy)

2. Pathological Staging: * TNM:
T: ◉ T1 ◉ T2 ◉ T3 ◉ T4
N: ◉ N0 ◉ N1 ◉ N2
M: ◉ M0 ◉ M1

3. Histologic type: * ◉ Adenocarcinoma (usual-type) ◉ Others, please specify:

4. Differentiation: ◉ Well ◉ Moderate ◉ Poor ◉ Not Applicable

5. Synchronous tumour present: ◉ Yes ◉ No

6. Regional lymph nodes * a) Number of nodes examined: b) Number of nodes positive:

SECTION 9: PATIENT STATUS

1. Patient status : *
◉ Alive → **Date of Follow-up (dd/mm/yy) :** / /
◉ Death → **Date of death (dd/mm/yy) :** / /
◉ Lost to Follow Up → **Date of Last Follow Up (dd/mm/yy) :** / /

References

Department of Statistics Malaysia (2016). Malaysia at a glance. https://www.dosm.gov.my/v1/index.php?r=columncone&menu_id=ZmVrN2FoYnBvZE05T1AzK0RLcEtiZz09.

Hassan, M. R. A., Khazim, W. K. W., Mustapha, N. R. N., Said, R. M., Tan, W. L., Suan, M. A. M., and Soelar, S. A. (2014). "The Second Report of the National Cancer Patient Registry-Colorectal Cancer, 2008-2013." Technical report, National Cancer Patient Registry-Colorectal Cancer and Clinical Research Centre (CRC).

Ministry of Health Malaysia (2013a). Caj pesakit luar. http:http://www.moh.gov.my/index.php/pages/view/162.

Ministry of Health Malaysia (2013b). Disease control division. http://www.moh.gov.my/english.php/pages/view/436.

Ministry of Health Malaysia (2017). "Clinical Practice Guideline: Management of Colorectal Carcinoma." Technical report, Malaysian Health Technology Assessment (MaHTAS).

National Cancer Patient Registry (2009). The national cancer patient registry. http://www.acrm.org.my/ncpr/.

Omar, Z. A. and Ibrahim Tamin, N. (2011). "National Cancer Registry Report Malaysia Cancer Statistics Data and Figure 2007." *Malaysia: National Cancer Registry*, pages 85–7.

Safurah, J., Mohd Noh, K., Othman, N. H., and Healy, J. (2013). "Malaysia Health System Review." *Asia Pacific Observatory on Health Systems and Policies*.

Chapter 2

Literature Review

2.1 Introduction

The chapter begins by providing a clinical overview of colorectal cancer: its definition, the signs and symptoms, how it is investigated and diagnosed, and clinical staging. Section 2.3 in this chapter considers the burden of colorectal cancer; its incidence and mortality rates worldwide, in all Asian countries, in South-East Asia, and then finally in Malaysia particularly. The latter section provides general information about Malaysia's geographical and population characteristics, and current published rates of colorectal cancer. Sections 2.4 to 2.6 provide a review of previous studies into survival and prognostic factors for patients with colorectal cancer. Section 2.7 examines previous survival studies for colorectal cancer in Malaysia.

2.2 Clinical Overview of Colorectal Cancer

This section describes the anatomy of the colon and rectum, and defines colorectal cancer together with its clinical features, diagnosis, symptoms and staging.

2.2.1 Anatomy of the Colon and Rectum

The system responsible for processing food and nutrients in our body is known as the digestive system or gastrointestinal tract. It has two main parts, the small intestine and the large intestine. The colon and rectum are part of the large intestine. The colon is shaped like a tube and measures about 5 feet in length. It is divided into the ascending colon, the transverse colon, the descending colon and the sigmoid colon. Fecal material matter in the colon moves gradually under peristalsis to the rectum, which is the last 6 inches of the large intestine and passes out from the body through the anus. (American Cancer Society, 2014)

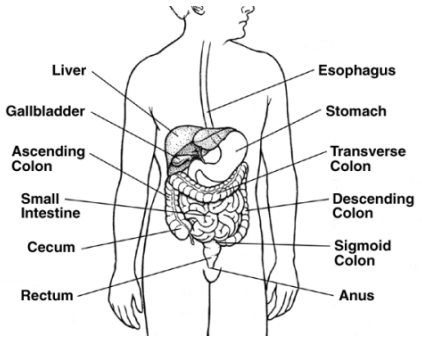


FIGURE 2.1: Human Anatomy: the Digestive System

2.2.2 Definition of Colorectal Cancer

Colorectal cancer refers to cancer in the colon or rectum. Cancer starts out as small growths occurring in the mucosal layer of the colon. These small growths are

known as polyps; they are benign (non-cancerous). Over a few years, the polyps may grow to become malignant (cancerous) tumours referred to collectively as colorectal cancer. (American Cancer Society, 2014).

2.2.3 Signs and symptoms

The presenting symptoms of colorectal cancer vary depending on the site of the cancer and stage of the disease. Patients may present with symptoms of alteration in bowel habit, intestinal obstruction, pain with an abdominal mass, unexplained reduction in weight, the presence of blood in the stool, or anaemia (Sack and Rothman, 2000). These symptoms may not be noticed by the patient for a long time.

Patients with colorectal cancer in the right side of the colon usually have symptoms of anaemia, loss of weight or abdominal pain. Patients with cancer in the left side of the colon often have an alteration in bowel habit or rectal bleeding (Simpson and Scholefield, 2008). Hamilton et al. (2005) determined that rectal bleeding, loss of weight, number of episodes of abdominal pain, constipation and number of episodes of diarrhoea were the five significant features most associated with colorectal cancer before diagnosis.

A systematic review conducted by Astin et al. (2011) on 23 works investigating symptoms of colorectal cancer reported that rectal bleeding is a major symptom of the disease. The review reported that an individual will be most at risk of having colorectal cancer if they have symptoms of rectal bleeding together with alteration in bowel habit and loss of weight.

2.2.4 Investigation and Staging for Colorectal Cancer

Colorectal cancer can be diagnosed by investigative procedures such as colonoscopy, flexible sigmoidoscopy, double contrast barium enema and CT scan. The primary investigative procedure for colorectal cancer is colonoscopy. A biopsy sample is taken from the colon and sent to the histopathology laboratory to confirm the diagnosis. CT scan can be used to assess cancer staging (Dorundi and Bannerjea, 2008).

Cancer staging shows the extent of the disease and how far it has spread. The cancer staging in this study is based on the TNM staging system. According to Greene et al. (2002), the T refers to the depth of the tumour, N refers to the number of lymph nodes that are involved and M refers to the occurrence of metastases.

2.2.5 Treatment for colorectal cancer

Treatment for colorectal cancer depends on the stage and location of the cancer, as well as the risks and benefits associated with the treatment themselves.. Surgery to remove the tumour tissue is the main treatment for colon cancer . Adjuvant therapy (additional ,which is the addition of treatment after surgery may also be involved depending on the severity of the disease , for example if the cancer is likely to come back of show growing into other tissues(American Cancer Society, 2014). For rectal cancer, surgery is also chosen as the main treatment. Additional treatments, such as chemotherapy and radiation, are often used before surgery (neo-adjuvant therapy) and/or after surgery (adjuvant therapy) to reduce the risk of recurrence and metastasis. If the cancer has spread through the wall of the rectum into nearby tissue and/or lymph node (regional stage), the radiation and

chemotherapy are often given together before surgery, with additional chemotherapy often given after surgery. The chemotherapy drugs used in the treatment of rectal cancer are the same as those used for colon cancer. 5-fluorouracil (5-FU) capecitabine, oxal- iplatin, and irinotecan are the chemotherapy drugs that most often used in the treatment of colorectal cancer (American Cancer Society, 2014).

In late stage colorectal cancer, where the cancer has spread to distant organs and tissues such as the liver, lung, peritoneum (lining of the abdomen), or ovaries, not all patients will be recommended for surgery. Chemotherapy, radiation, and biologically targeted therapies may be given alone or in combination to relieve, delay, or prevent symptoms and to prolong survival of the patients (American Cancer Society, 2014).

2.3 Incidence and Mortality Rates of Colorectal Cancer

2.3.1 Worldwide

The incidence and mortality data reported in this section were extracted from the GLOBOCAN 2012 report (Ferlay et al., 2014). This GLOBOCAN data was presented by type of region, classed as either developed or less developed regions. Countries in more developed regions include North America, Europe, Australia/New Zealand and Japan. The incidence and mortality rates given are Age Standardized Rates (ASR), defined in the GLOBOCAN (Ferlay et al., 2014) report as 'a summary measure of the rate that a population would have if it had a standard age structure using World standard population, expressed as per 100,000 persons per year.

Cancer is a major health burden across the world, with over 14.1 million new cancer cases worldwide in 2012. Of these, around 1.35 million cases (9.6%) are new cases of colorectal cancer. The number of colorectal cancer cases is expected to increase by 80% by the year 2035, climbing to approximately 2.4 million new colorectal cancer cases and contributing to 1.3 million deaths worldwide (Douaiher et al., 2017). Colorectal cancer also is the third most common cancer in men and the second most common cancer in women (Ferlay et al., 2014).

For the past three decades(1975 to 2010), there was an increasing trend of colorectal cancer incidence world wide ranging from 100 to almost 400 cases per 100,000 population per year and the trend was found to be higher in men (Ferlay et al., 2014). The increasing trend of colorectal cancer incidence worldwide was projected to continue for the next 25 years.

In 2012, the estimated number of deaths from colorectal cancer was 690,000 worldwide, of which 53.8% and 46.2% were in men and women respectively. In the same year, the worldwide mortality rate for colorectal cancer was 8.4 deaths per 100,000 (10.0 per 100,000 in men and 6.9 per 100,000 in women). Though incidence is higher in developed regions in 2012, mortality rates attributable to colorectal cancer were higher in less developed regions, accounting for 52% of total colorectal cancer deaths worldwide (Ferlay et al., 2014).

2.3.2 Asian countries

In the GLOBOCAN (GLOBOCAN, 2012) report, Asia is divided into four regions: Eastern Asia, South-Eastern Asia, South-Central Asia and Western Asia. Lung cancer is the most-diagnosed cancer in Asia with an estimated 1.04 million new cases in 2012 (15.5% of all cancer cases). This is followed by stomach cancer (10.4%), breast cancer (9.6%) and colorectal cancer (9%); the latter of which represents an estimated 607,200 new cases in Asia in 2012, 347,500 cases in men

and 259,700 cases in women. This makes it the fourth most common cancer in all of Asia. The incidence of colorectal cancer in Asia was 13.7 cases per 100,000 population per year, comprising male and female incidence of 16.5 and 11.1 per 100,000 population per year respectively. 69% (421,400) of these estimated new Asian cases of colorectal cancer were in Eastern Asia, with 18.4 cases per 100,000 population per year. This information is presented in Table 2.1.

There were an estimated 331,600 deaths from colorectal cancer in Asia in 2012, of which 185,000 deaths were men and 146,600 deaths were women. The ASR mortality was 7.2 per 100,000 population per year, with 8.6 and 6.0 per 100,000 population per year in men and women respectively. The highest estimated number of deaths was in Eastern Asia, whose 207,700 deaths represent 62.6% of all deaths from colorectal cancer in Asian countries (Ferlay et al., 2014), as presented in Table 2.2. Figure 2.2 and Figure 2.3 show the maps of incidence of and mortality due to colorectal cancer in Asia.

Overall, there was an upward trend in colorectal cancer incidence in Asian countries for the past three decades(1975-2010). Countries with the highest incidence trends are Japan, China and Singapore from EastAsia and South-East Asia while colorectal cancer incidence in Thailand(South-East Asia) and India(South-Central) Asia are at the bottom two. In many low-income and middle-income countries, CRC incidence and mortality rates correlates with the adoption of a western lifestyle linked to ongoing societal and economic development. In highly developed countries, rates are stabilising or decreasing (Arnold et al., 2016a).

TABLE 2.1: Estimated number of new colorectal cancer cases (thousands) in Asia in 2012

Region	**Male (%)**	**Female (%)**	**Total (%)**
Eastern Asia	243.2 (70.0)	178.2 (68.7)	421.4 (69.4)
South-Eastern Asia	38.5 (11.1)	30.5 (11.7)	69.0 (11.4)
South-Central Asia	50.7 (14.6)	39.0 (15.0)	89.7 (14.8)
Western Asia	15.1 (4.3)	12.0 (4.6)	27.1 (4.4)
Total	347.5 (100)	259.7 (100)	607.2 (100)

TABLE 2.2: Estimated number of colorectal cancer deaths (thousands) in Asia in 2012

Region	Male (%)	Female (%)	Total (%)
Eastern Asia	115.5 (62.4)	92.2 (62.9)	207.7 (62.6)
South-Eastern Asia	24.1 (13.0)	19.1 (13.0)	43.2 (13.0)
South-Central Asia	37.0 (20.0)	28.4 (19.4)	65.4 (19.7)
Western Asia	8.4 (4.6)	6.9 (4.7)	15.3 (4.7)
Total	185 (100)	146.6 (100)	331.6 (100)

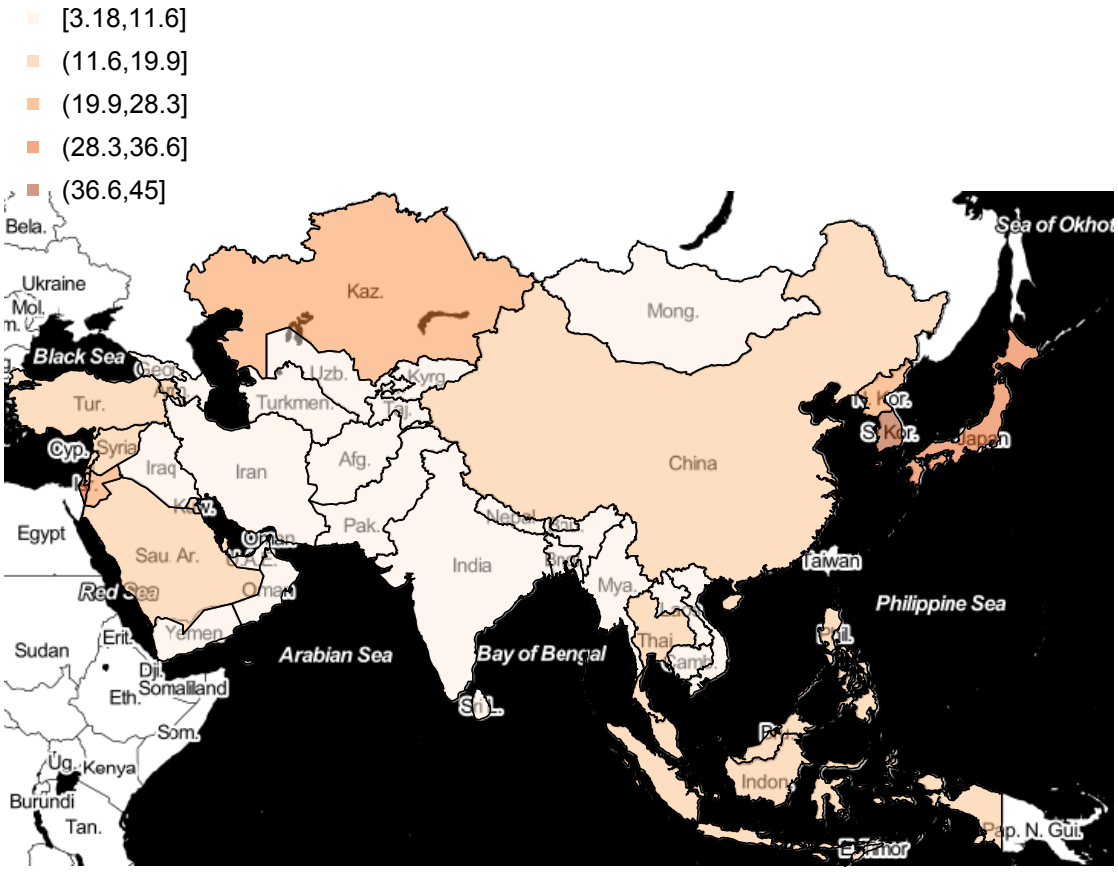


FIGURE 2.2: Map of Asia showing Age Standardized Rate of Incidence of Colorectal Cancer in Asia

*see text for definition of Age Standardized Rate (ASR)

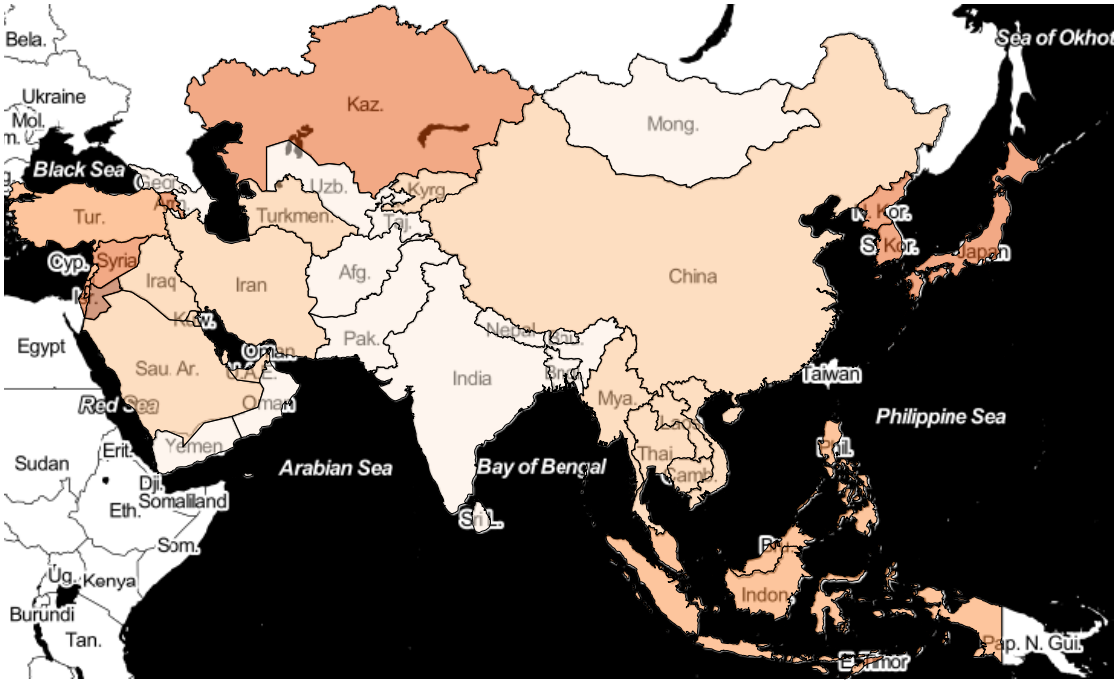


FIGURE 2.3: Map of Asia showing Age Standardized Rate of Mortality due to Colorectal Cancer in Asia

*see text for definition of Age Standardized Rate (ASR)

2.3.3 South-East Asian Countries

This area comprises Malaysia and its closest surrounding countries; namely Brunei, Cambodia, Indonesia, Laos, Myanmar, Philippines, Singapore, Thailand, Timor Leste and Vietnam. GLOBOCAN estimated that there were 68,900 new cases of colorectal cancer in South-East Asia in 2012. Of that number, an estimated 38,500 cases were in men and 30,400 in women. The age standardised rate of colorectal cancer incidence in South-East Asia was 12.5 cases per 100,000 population per year; 15.2 in men and 10.2 in women (Ferlay et al., 2014). According to the GLOBOCAN report, Singapore, which is adjacent to Malaysia, had the highest age standardized incidence rate of this group of countries, at 33.70 per 100,000

population per year. The Singapore Registry Report 2009 (Peng et al., 2013) stated that colorectal cancer was the most common cancer among males between 2009 to 2013 and the second most common cancer in females; representing 17.3% and 13.2% of all cancer cases in males and females respectively (Peng et al., 2013). Malaysia had the third highest estimated incidence in 2012 among South-East Asian countries with an incidence of 13.4 cases per 100,000 population; Figure 2.4 (Ferlay et al., 2014).

In South-East Asia, the estimated number of deaths attributed to colorectal cancer in 2012 was 43,200 cases, of which, 24,100 deaths were men and 19,100 women. The estimated age standardised mortality rate for colorectal cancer in South-East Asia was 7.9 deaths per 100,000 population overall with 9.7 and 6.4 ASR deaths per 100,000 population in men and women respectively. Brunei had the highest mortality rate attributed to colorectal cancer in 2012 with 12.0 deaths per 100,000 population while Malaysia had 9.4 deaths per 100,000 population (Ferlay et al., 2014). However, the incidence and mortality data from GLOBOCAN was based on data obtained from only two states (Penang and Sarawak) in Malaysia and not from the other 12 states. More detail is given in the next section.

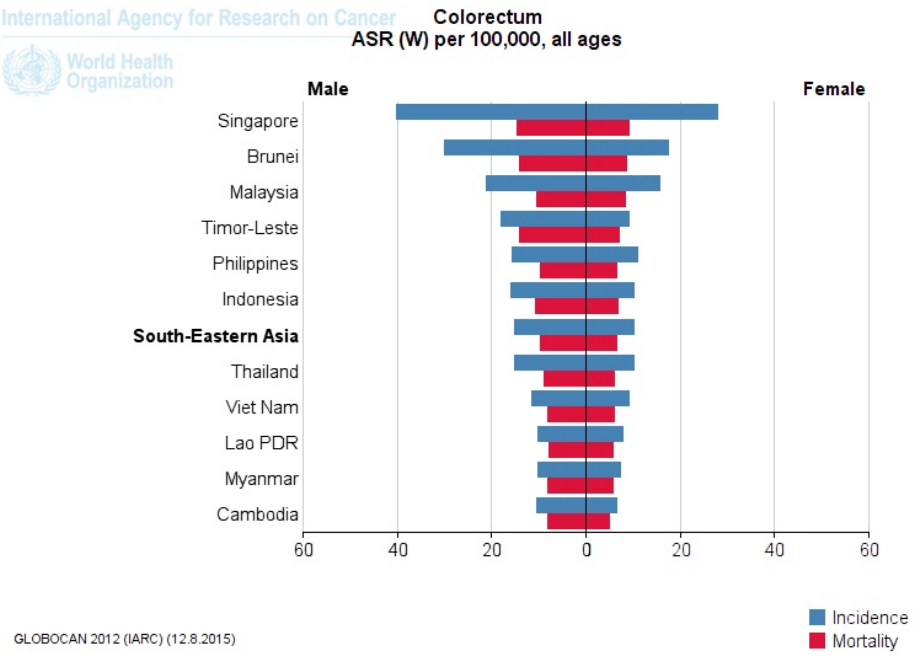


FIGURE 2.4: Plot of age standardized rates of incidence and mortality of colorectal cancer in South-East Asia, taken from the GLOBOCAN website (GLOBOCAN, 2012)

2.3.4 Malaysia

In Malaysia, cancer is one of the leading causes of mortality; the Ministry of Health in Malaysia reports cancer as the third leading cause of death after heart disease and disease of pulmonary circulation, and septicaemia (Omar and Ibrahim Tamin, 2011).

Colorectal cancer ranks as the second most common cancer in Malaysia for both sexes; exceeded only by lung cancer in men and breast cancer in women. The 2nd Annual Report by the National Cancer Patient Registry - Colorectal Cancer reported that the cumulative incidence rate was 21.3 cases per 100,000 population between 2008 and 2013 (Hassan et al., 2014). 55.9% of the overall cases were men. Grouping the population by ethnicity, Malay (42.7%) had the most colorectal cancer cases reported, followed by Chinese (40.3%), Indian (5.8%) and

Other Ethnicity (10.8%); after accounting for age, having Chinese ethnicity corresponded to the highest Age Standardized Rates of colorectal cancer incidence with 27.4 cases per 100,000 population. The Age Standardized Incidence Rates for Malay and Indian ethnicities was 19.4 and 17.6 cases per 100,000 population respectively (Hassan et al., 2014).

The mortality rate from 2008-2013 was 9.8 cases per 100,000 population. Among the three largest ethnic groups, the highest proportion of deaths was among Malay (49.2%), followed by Chinese (42.3%) and Indian (39.4%). Like the incidence rates above, the age adjusted mortality rate was highest among those of Chinese ethnicity with 11.9 deaths per 100,000 population, followed by Malay and Indian with 9.6 and 7.1 deaths per 100,000 population respectively (Hassan et al., 2014).

2.4 Survival Rates for Colorectal Cancer

Cancer is a chronic disease. How long individuals will survive depends on the severity of the disease at diagnosis, and on other factors affecting survival time that it is important to determine. Many studies have been carried out worldwide to predict cancer survival time and identify factors affecting this including socio-demographics, clinical factors and treatment modalities.

A recent study noted that incidence and mortality rates for colorectal cancer vary widely worldwide and are associated with human development levels; in particular where societal or cultural changes might incrementally reflect the adoption of a more western lifestyle (Arnold et al., 2016b).

Previous research has shown that there is a trend of increasing survival of colorectal cancer in Europe. A recent population based survival study in England and Wales by Quaresma et al. (2015) reported that survival for colorectal cancer has been steadily increasing over a 40 year period (1971 to 2011). The 5-year survival rate

in this period had increased from 24.6% to 58.2% for colon cancer and 24.2% to 59.7% for rectal cancer in England. The same trend was also found in Wales where the 5-year survival rates increased from 25.0% to 57.7% and 22.9% to 58.5% for colon and rectal cancers respectively (Quaresma et al., 2015). This increasing trend may be a result of the implementation of cancer screening in 2006, which aims to prevent invasive malignant tumours by diagnosing cancer at an early stage. Tang et al. (2015) performed a meta analysis of four randomized controlled trials: the study suggested that one death from colorectal cancer is prevented for every 1000 people that had been screened by flexible sigmoidoscopy.

Bujanda et al. (2010) compared the survival rates between groups of patients diagnosed with colorectal cancer in Spain during two separate periods of time; one group of patients diagnosed between 1980 and 1994 and another group diagnosed in 2001. The 5-year survival rates in these two groups were 35% and 57% respectively and the difference was statistically significant, showing an increasing trend in 5-year survival over 20 years. They concluded that the increased 5-year survival trend was parallel to both an increase in the administration of adjuvant chemotherapy and a decline in postoperative mortality (Bujanda et al., 2010). Similarly, a systematic review and meta-analysis noted that additional chemotherapy improved overall survival in colorectal cancer patients with liver metastases who underwent curative resection (Araujo et al., 2015). Another study, Oliphant et al. (2013), showed that longer term survival in the study was significantly influenced by surgical specialisation.

Colorectal cancer survival rates also improved over time in Asian countries. A population-based study carried out among colorectal cancer patients in China (Fang et al., 2013) found that the overall 5-year survival rate improved throughout the 50 year period between the 1960's and the 2000's, increasing from 49% in the 1960's to 77% in the 2000's. Fang et al. (2013) noted that it could have been

the result of advances in surgical techniques and chemotherapy, and changes to lifestyle.

In another study comparing the results from different pieces of research in a few Asian countries, they concluded that the 5-year survival rate remained at almost 60%, where China had the highest and India the lowest 5-year survival rate (31.2%) (Moghimi-Dehkordi and Safaee, 2012).

Overall, the survival rate for colorectal cancer showed an increasing trend in both Western and Eastern countries. Individual level factors that may affect colorectal cancer survival will be discussed in the next section.

2.5 Risk Factors for Colorectal Cancer

We searched PubMed database and Google Scholar to identify articles on colorectal cancer. The search strategy combined the Medical Subject Heading (MeSH) with search terms that include “neoplasm”, “malignancy”, “cancer”, “colon” and “rectum”. Text word search include variations of “colorectal”, “colon”, “rectal”, “cancer”, “tumour”, “malignant”, “prognostic”, “prognosis”, “incidence”, “risk factor”, “survival”. Other terms is changed based on search topic. For example, to search age as a prognostic factor, the term ¿‘Age” OR “Young” OR “Old” is added to the existing terms. Publications were limited to human studies published between January 2000 to December 2017 in English language. We also searched the available national cancer registries of individual Asian countries to make a comparison on cancer statistics

Factors that increase the risk of developing colorectal cancer can be divided into two categories: modifiable factors and non-modifiable factors.

Non-modifiable risk factors include having inflammatory bowel disease, some hereditary factors and old age (Haggar and Boushey, 2009). Wei et al. (2004) compared risk factors for colon cancer and rectal cancer; they reported that people who have first degree relatives with colon or rectal cancer are more at risk of getting either of these cancers. Specifically, these patients are almost twice as likely to develop colon cancer (RR 1.94, 95% CI (1.80 , 2.10)), and 27% more likely to develop rectal cancer (95% CI (1.08,1.50)) than patients without family history of these cancers.

Modifiable factors include obesity, lack of physical activity, a diet high in red and/or processed meat, smoking, and excessive drinking of alcohol and are all known to be associated with a diagnosis of colorectal cancer (Johnson et al., 2013).

2.6 Prognostic Factors for Colorectal Cancer

We previously mentioned that colorectal cancer is one of the primary causes of mortality worldwide. Therefore, understanding how different factors affect the prognosis and eventual survival of patients with this disease is of great importance. In this section, we discuss previous research drawn from a variety of populations studying factors that affect the survival outcomes of patients with colorectal cancer, and discuss several in particular that are relevant to the data we are using in this thesis. These factors are: age, sex, race, tumour location and socioeconomic status. We describe herein the effects of these factors on the survival of colorectal cancer patients in different populations. We acknowldge that this list is not exhaustive and there are other factors that may also affect the survival of colorectal cancer that we do not have data for yet, such as genetic influences and comorbidities; these are areas that might be considered in future research.

2.6.1 Age

More than 90% of the patients diagnosed with colorectal cancer were found to be more than 55 years old in the Li et al. (2015) study. Age at diagnosis also affects the survival from the disease. Despite previous findings that associate advanced age with poor survival, there are other studies that report that there is poor outcome in younger colorectal cancer patients. Generally, colorectal cancer in young patients usually presents with poor tumour morphology and late stage at diagnosis compared to older individuals (Li et al., 2015, Orsini et al., 2015) and with other factors that lead to poor survival (Chan et al., 2010, LIN et al., 2005). Fu et al. (2014) in their 30-year (1985 to 2011) retrospective review study in China found that age was not significant as an independent prognostic factor for colorectal cancer after controlling for other factors including sex and clinical and pathologic characteristics.

On the other hand, some studies found that young patients have a better survival prognosis despite presenting with more advanced disease at diagnosis. This is especially so after surgery, compared to the elderly. This phenomenon is reported in numerous studies worldwide, such as in the US (Li et al., 2015, 2014), Europe (McKay et al., 2014, Widdison et al., 2011) and in Asian countries (Yeole et al., 2001, Yuan et al., 2013). A large population-based study reported that the eldest group of patients ($>$ 80 years) have almost twice the risk of death after operation with hazard ratio 1.95, 95% CI (1.90, 2.02), after controlling for other factors (Li et al., 2015). The result is explained by a higher percentage of co-morbodities and complications after surgery experienced by older patients. This is agreed by Widdison et al. (2011) and Stornes et al. (2014), who claimed that age influences the selection of treatment modalities in older patients; for example, older patients may be treated less often with surgery (including major radical surgery)

or radio-chemotherapy, which results in more local recurrence and poorer eventual outcomes. This suggests a more complex causal pathway.

2.6.2 Sex

Incidence of colorectal cancer is considerably higher in men than women in many countries worldwide (Ferlay et al., 2014). Findings on the influence of sex differences in colorectal cancer prognosis have been less consistent. Some studies, for example those by Hendifar et al. (2009), Majek et al. (2013), McArdle et al. (2003) ,report a better survival in women, while other studies did not find any significant difference (Lydrup and Höglund, 2015). An analysis of a large representative cohort, comprising 24 clinical trials, that enrolled more than 30,000 colon cancer patients between 1978 and 2003 found that women with early stage colon cancer have significantly higher survival rates than men (Cheung et al., 2013). In their review, the researchers point out that the better survival in women may be due to their longer life expectancy on average compared to men, it is also possible that men may have a biologically more-aggressive disease or a poorer response to adjuvant therapy which leads to poorer survival. However, the subset analysis when stratified by age group showed that differences between sexes were more apparent in the older age group (aged > 65 years) than in patients aged ≤ 65 years, with a significant interaction ($\mathbb{P}_{\text{interaction}} < 0.01$).

A study by Kotake et al. (2016) focused on Asian countries. In their large study spanning a 20 year period (1985 to 2004) and involving more than 80,000 colorectal cancer patients in Japan, it was stated that sex is an independent prognostic factor for survival of colorectal cancer after controlling for other factors. Women have a lower risk of death compared to men with an adjusted hazard ratio of 0.92, 95% CI $(0.90, 0.95)$. Similarly, in Taiwan, Chou et al. (2013) reported that the 5-year survival rates for colon and rectal cancer were approximately 2-5% higher in women

than men. Similarly, in Tehran, a significant difference in 5-year survival rate was found between the sexes, where females had better survival (75%) compared to males (63%), and sex was also included as an independent prognostic factor in a Cox Model (Heidarnia et al., 2013). Another study by Lydrup and Höglund (2015) found that sex had no significant influence on colorectal cancer survival when controlling for age at diagnosis, neo-adjuvant treatment, tumour stage and year of having surgery.

2.6.3 Race

Different races or ethnic groups may have different cultures, lifestyles, environments and food preferences, and these factors may influence their health outcomes in general. Racial disparities have been reported to influence the survival of colorectal cancer in many countries (Le et al., 2009, Sabounchi et al., 2012, Veach et al., 2014).

Chien et al. (2005) evaluated the relationship between race/ethnicity and colorectal cancer outcome among 18 different racial/ethnic groups in the United States of America involving more than 150,000 colorectal cancer patients diagnosed between 1988 and 2000. The analysis of this large nationally representative population-based sample quantifies a significant variation in survival between ethnic groups. Similarly, a later US study of rectal cancer patients diagnosed between 2000 and 2009 demonstrated that race was a significant prognostic factor in multivariable analysis with a hazard ratio of 1.49 and 95% CI (1.10, 2.02) (Nitzkorski et al., 2013). This is also consistent with a further study by Ahmadi et al. (2014) in Asian countries which reported race as an independent prognostic factor for colorectal cancer patients. Conversely, Hassan et al. (2009) claims that ethnicity was associated with different age and stage at diagnosis but did not influence the patients' treatment or survival outcome.

2.6.4 Smoking

Smoking is a risk factors for many diseases including colorectal cancer. A 2014 report by the US Surgeon General (US Department of Health and Human Services et al., 2014), concluded that smoking is a risk factors for colorectal cancer and also that smoking is causally related to higher mortality from colon and rectal cancer. A systematic review and meta-analysis which included 16 studies on smoking and survival in colorectal cancer estimated a combine all cause mortality hazard of death in current smokers to be significantly higher, at 26% (HR 1.26 (95%CI 1.14- 137) compared to non-smoker patients (Walter et al., 2014). In a later study, Sharp et al. (2017b) investigated the association of between smoking at diagnosis and cancer specific survival in colon cancer. In this population based cohort study involving 18,166 colon cancer patients, it was found that smoking was an independent prognostic factor for colon cancer. When looking at the association between smoking status and colon cancer survival, they noted that the significant association was found in patients with surgical treatment only and suggested that this might be due to the underlying mechanism related to surgery. Later, they carried out a study on the relationship between smoking and survival of 10,794 rectal cancer cases in Ireland. Current smokers with have rectal cancer were found to have significantly poorer survival (HR=1.15, 95%CI 1.06-1.24) compared to non-smokers but the association was not significant in the former smokers at diagnosis of rectal cancer (Sharp et al., 2017a).

Recently, there is was a large review and meta-analysis studies of the impact of pre-diagnostic smoking on colorectal cancer prognosis, which involved 14 population-based cohort studies from 10 different countries in Europe and US. A significant association was found between current smoking (HR 1.29; 95% CI ; 1.04;1.60) and former smoking (HR 1.12; 95% CI 1.04;1.20) with poorer overall colorectal cancer

mortality but for the cause specific mortality, the association was not significant after adjusting for other covariates in the model (Ordóñez-Mena et al., 2017).

2.6.5 Diabetes

Previous studies reported the association between diabetes and survival in colorectal cancer. In the Netherland, a study was carried out that involved almost 10,000 colorectal cancer patients diagnosed in a 10-year period(1997 and 2007). These researchers found that patients with pre-existing diabetes at diagnosis had a higher risk of death from colorectal cancer compared to non-diabetic patients even after controlling for socio-demographic and clinical factors (Van de Poll-Franse et al., 2012). Lower overall survival may also be attributable to poor diabetes control, the result of a lack of attention to diabetes during cancer treatment. This latter idea was postulated by the American Diabetes Association and American Cancer Society in 2010 concerning diabetes and cancer (Zanders et al., 2014).

Opposite findings were found by a smaller study of 1039 colorectal cancer cases diagnosed between 2003 to 2005 from the Italian Cancer Registries database which indicated that diabetes in patients increased their survival. They claimed that this might be due to prevention and treatment stabilizes the condition and controls its complications and reduces mortality (Bella et al., 2013).

This is similar to a recent study which examined the association of four metabolic syndrome disease characteristics (diabetes, obesity, hypertension and dyslipidaemia) to progression-free survival in patients in with Stages I to Stages III colorectal cancer. Their results suggested that only diabetes mellitus had a significant positive association with progression-free survival and they found that the use of Metformin in the treatment of diabetic patients improved their outcome of progression free survival (Croft et al., 2018).On the other hand, a large systematic review and meta-analysis involving 21 eligible cohorts which involved more than one million

patients of colorectal cancer had different findings. This meta-analysis indicates that colorectal cancer patients with pre-existing diabetes mellitus was adversely associated with lower(worse) overall survival. However the association was not significant with cancer-specific survival (Li et al., 2017)

2.6.6 Site

The prognosis of colorectal cancer has been shown to vary depending on the anatomical site of the tumour. Lee et al. (2013)'s study in the USA and Jafarabadi et al. (2011)'s in Iran reported that patients with cancer at the rectum had better survival compared to those who had cancer at the colon. However, Fang et al. (2013) found a better 5-year survival was significantly associated with tumour location in the colon than in the rectum, among colorectal cancer cases diagnosed between 1960 to 2008 in China.

Patients with right-side tumours have been reported to have poor prognosis compared to left-side tumours in some studies (Golan et al., 2013, Meguid et al., 2008, Powell et al., 2012). Those with right-side tumours had a 4.2% increased risk of death (adjusted hazard ratio of 1.042 and 95% CI (1.02, 1.07)) compared with patients that had tumours on the left side of the colon, including the rectum (Meguid et al., 2008). These findings are in the line with a recent study by Price et al. (2015) which revealed that patients with tumours on the left side of the colon had a 25% greater survival rate than patients with tumours on the right side of the colon, after adjusting for other known factors that may influence the patient's survival.

2.6.7 Stage

The stage at diagnosis plays the biggest role in affecting the survival of colorectal cancer patients. Survival times in patients with colorectal cancer vary significantly across the different stages of the disease as has been indicated in findings by previous studies (Chien et al., 2013, Lemmens et al., 2005, Li et al., 2014, Quah et al., 2008). As the stage of disease at diagnosis advances, survival becomes considerably shorter.

Aravani et al. (2009) reported that 93% of patients who presented with earlier stage disease (Dukes A) survived longer than 5 years compared to only 6% of those with late stage disease (Dukes D). In another study, Ghabeljoo et al. (2011) confirmed that pathologic stage was an independent prognostic factor for both colon and rectal cancer, where the strongest hazard of death was observed in stage IV cancers, with hazard ratio 8.42 and 95% CI (2.09, 33.93). Conversely, Van der Pool et al. (2011) saw a positive improvement over time in survival in patients whose cancer was at stage IV at diagnosis, which could have been a result of advancement in chemotherapy or the proportion of patients that had also had liver surgery. This is in agreement with another study by Stelzner et al. (2005).

2.6.8 Socioeconomic status

Socioeconomic status has an association with colorectal cancer survival acknowledged in a number of studies. Findings from Dik et al. (2014), Harris et al. (2009), Le et al. (2008), Manser and Bauerfeind (2014) and Kelsall et al. (2009) show that declining socioeconomic status is associated with poorer survival. The group with the lowest socioeconomic status in Hines et al. (2014) had a 24% higher risk of death than that of the highest socioeconomic status (hazard ratio 1.24, 95% CI (1.16,1.32)). Hole and McArdle (2002) observed the impact of socioeconomic

deprivation on outcome after surgery of 2269 colorectal cancer patients and found that the lowest 5-year survival rate was in the most deprived group of patients (30.3%) compared to 32.5% in the intermediate and 35% in the most affluent group.

The inequalities of survival may be explained by less treatment being received by the lower socioeconomic status groups (Hines et al., 2014). Similarly, Aarts et al. (2010) in his review of 55 articles, discussed that unfavourable results have been seen in treatment, survival and mortality in the lower socioeconomic status groups where these patients tend to receive neo-adjuvant therapies less frequently than those in higher socioeconomic status groups.

Overall, evidence from literature suggests that research investigating factors affecting survival in colorectal cancer in certain populations is important and may help public health practitioners to plan and manage better healthcare.

2.7 Survival and Incidence Studies in Malaysia

In the previous section, we described various studies carried out both worldwide and in purely Asian countries concerning the factors influencing survival in patients with colorectal cancer. In this section we present published research on colorectal cancer that pertains specifically to patients in Malaysia.

Ghee (2014) reviewed 56 research articles concerning colorectal cancer that were conducted in Malaysia. They looked into genetic causes and biomarkers, KAP (knowledge attitude and practice) studies and quality of life studies among colorectal cancer patients. The literature on epidemiology and survival in colorectal cancer in Malaysia is not extensive. The most common site of colorectal cancer was found to be the left side of the colon. Azmi et al. (2007) studied 119 patients

who underwent surgery in one of the hospitals in Malaysia, and noted that colorectal cancer was more common in men and in patients of Malay ethnicity. More than fifty percent of the cases were found to be in the recto-sigmoid junction and rectum. In a later study, Magaji et al. (2014) investigated a larger sample involving 1212 colorectal cancer patients treated in University Malaya Medical Centre between 2000 and 2010. Similar to previous findings, they reported more male patients than female, but here more than 60% of the patients were Chinese.

Cancer survival in Malaysia has been studied at a regional level. A retrospective record review study considered colorectal cancer patients diagnosed over the ten year period from 1996 to 2005 in Hospital Universiti Sains Malaysia (HUSM) Kelantan, (Ghazali et al., 2010). The study reported that the overall 5-year survival rate was 34.3% and identified three significant independent prognostic factors: Dukes staging at diagnosis, presence of liver metastases and treatment modalities. Apparent delay in screening (Al-Naggar et al., 2015, Hilmi et al., 2010) and seeking treatment (Rashid et al., 2009) in high risk groups lead to poor overall survival in colorectal cancer patients.

Low socioeconomic groups are particularly at risk. Kong et al. (2010) investigated the impact of socioeconomic class on colorectal cancer patients in two major hospitals in Peninsular Malaysia (Kuala Lumpur) and West Malaysia (Kuching, Sarawak). Socioeconomic status was based on occupation as defined by the National Statistics Socioeconomic Classification. Patients from a lower socioeconomic class presented with later stage disease and had a smaller 5-year survival rate. Most of the studies only performed a descriptive analysis without further assessing the prognostic factors affecting survival in colorectal cancer.

Studies about colorectal cancer incidence in Malaysia are not extensive. One study has estimated the cumulative incidence rate for colorectal cancer patients diagnosed between 2008 to 2013 in Malaysia by sex and ethnicity for colorectal

cancer reported in the report by NCPR-CRC (Abu et al., 2016). The study reported that the cumulative incidence of colorectal cancer in the six year period was 21.32 cases per 100,000 population. Males had higher cumulative incidence in with 24.6 per 100,000 population compared to females with 18.4 cases per 100,000 population during the same period. Chinese ethnicity had the highest overall incidence followed by Malay and Indian with an incidence of 27.35, 18.95 and 17.55 cases per 100,000 population respectively (Abu et al., 2016) .

Later, another study investigated the modifiable risk factors for colorectal cancer in Malaysia. These included alcohol consumption, being overweight and physical inactivity amongst the Malaysian population. Their analysis aimed to estimate the population attributable fraction (PAF, the fraction of all cases (both exposed and unexposed in a population) that would not have occurred had the exposure not occurred. Their findings suggested that approximately 18% of the CRC cases in the year 2013 would be preventable, if proper interventions to limit physical inactivity, overweight and alcohol consumptions were implemented (Naing et al., 2017).

Other studies reported the distribution of colorectal cancer cases within selected countries. These studies, Samat et al. (2013), Samat and Shattar (2014) and Shah et al. (2014), demonstrated the spatial pattern in colorectal cancer incidence and mapped hot and cold spot regions. However, no spatial modelling was involved, and only one state was chosen in each study (namely, Penang, Kelantan and Kuala Lumpur). We reviewed these studies in the introduction section of chapter 5. We will see later that case-ascertainment is an issue for incidence studies in Malaysia.

Socio-demographic and clinical factors are factors that affect both survival from and incidence of the colorectal cancer. There are other factors that affect the survival and incidence such as environmental factors , for example the socioeconomic status of the population.. We were also interested to examine the effect of where

someone lives on survival and incidence, that is spatial factors, which we are interested in investigating in this research. To date, this is the first study looking into spatial variation in survival and the incidence of colorectal cancer for the whole Malaysian population, a research gap this thesis seeks to fill.

References

Aarts, M. J., Lemmens, V. E., Louwman, M. W., Kunst, A. E., and Coebergh, J. W. W. (2010). "Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome." *European Journal Of Cancer*, 46:2681–2695.

Abu, H. M., Ismail, I., Mohd, S. M., Ahmad, F., Wan, K. W., Othman, Z., Mat, S. R., Tan, W., Mohammed, S., Soelar, S., et al. (2016). "Incidence and mortality rates of colorectal cancer in Malaysia." *Epidemiology and health*.

Ahmadi, A., Nazari, S. S. H., and Mobasheri, M. (2014). "Does ethnicity affect survival following colorectal cancer? A prospective, cohort study using Iranian cancer registry." *Medical Journal of the Islamic Republic of Iran*, 28:83.

Al-Naggar, R. A., Al-Kubaisy, W., Yap, B. W., Bobryshev, Y. V., and Osman, M. T. (2015). "Attitudes towards colorectal cancer (CRC) and CRC screening tests among elderly Malay patients." *Asian Pacific Journal of Cancer Prevention*, 16:667–674.

American Cancer Society (2014). "American Cancer Society. Colorectal Cancer Facts and Figures 2014-2016." Technical report, American Cancer Society.

Araujo, R. L., Gönen, M., and Herman, P. (2015). "Chemotherapy for patients with colorectal liver metastases who underwent curative resection improves long-term outcomes: systematic review and meta-analysis." *Annals of surgical oncology*, 22:3070–3078.

Aravani, A., J.Thomas, M.Day, D.Forman, E.Morris, and R.TatarekGintowt (September2009). "Survival by Stage of Colorectal Cancer in England." *Northern and Yorkshire Cancer Registry and Information Service*, 1:1–2.

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2016a). "Global patterns and trends in colorectal cancer incidence and mortality." *Gut*, pages gutjnl–2015.

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2016b). "Global patterns and trends in colorectal cancer incidence and mortality." *Gut*, 1:gutjnl–2015.

Astin, M., Griffin, T., Neal, R. D., Rose, P., and Hamilton, W. (2011). "The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review." *Br J Gen Pract*, 61:e231–e243.

Azmi, M., MA, M. Z., Norashikin, M., AsmaAssadah, M., H.Norra, R.Kalavathy, R.M.K.Mohamad, and H.Hasmah (2007). "Five-Year Review of Histopathological Findings of Colorectal Cancer Patients Operated in Hospital Tengku Ampuan AfzanKuantan Pahang Malaysia." *Int. Med Journal*, 6:1–14.

Bella, F., Minicozzi, P., Giacomin, A., Crocetti, E., Federico, M., De Leon, M. P., Fusco, M., Tumino, R., Mangone, L., Giuliani, O., et al. (2013). "Impact of diabetes on overall and cancer-specific mortality in colorectal cancer patients." *Journal of cancer research and clinical oncology*, 139:1303–1310.

Bujanda, L., Sarasqueta, C., Hijona, E., Hijona, L., Cosme, A., Gil, I., Elorza, J. L., Asensio, J. I., Larburu, S., Enríquez-Navascués, J. M., et al. (2010). "Colorectal cancer prognosis twenty years later." *World Journal of Gastroenterology: WJG*, 16:862–867.

Chan, K., Dassanayake, B., Deen, R., Wickramarachchi, R., Kumarage, S., Samita, S., and Deen, K. (2010). "Young patients with colorectal cancer have poor survival in the first twenty months after operation and predictable survival in the medium and long-term: analysis of survival and prognostic markers." *World journal of surgical oncology*, 8:82.

Cheung, W. Y., Shi, Q., O'connell, M., Cassidy, J., Blanke, C. D., Kerr, D. J., Meyers, J., Van Cutsem, E., Alberts, S. R., Yothers, G., et al. (2013). "The predictive and prognostic value of sex in early-stage colon cancer: a pooled analysis of 33,345 patients from the ACCENT database." *Clinical colorectal cancer*, 12:179–187.

Chien, C., Morimoto, L. M., Tom, J., and Li, C. I. (2005). "Differences in colorectal carcinoma ctage and survival by race and ethnicity." *Cancer*, 104:629–639.

Chien, L.-C., Schootman, M., and Pruitt, S. L. (2013). "The modifying effect of patient location on stage-specific survival following colorectal cancer using geosurvival models." *Cancer Causes & Control*, 24:473–484.

Chou, C.-L., Weng, S.-F., Lin, J.-K., and Chang, S.-C. (2013). "Role for gender in colorectal cancer risk: a Taiwan population-based study." *International journal of colorectal disease*, 28:1001–1008.

Croft, B., Reed, M., Patrick, C., Kovacevich, N., and Voutsadakis, I. A. (2018). "Diabetes, obesity, and the metabolic syndrome as prognostic factors in stages I to III colorectal cancer patients." *Journal of gastrointestinal cancer*, pages 1–9.

Dik, V. K., Aarts, M. J., Van Grevenstein, W. M., Koopman, M., Van Oijen, M. G., Lemmens, V. E., and Siersema, P. D. (2014). "Association between socioeconomic status, surgical treatment and mortality in patients with colorectal cancer." *British Journal of Surgery*, 101:1173–1182.

Dorundi, S. and Bannerjea, A. (2008). "Colorectal cancer: early diagnosis and predisposing causes." *Surgery-Oxford International Edition*, 26:323–328.

Douaiher, J., Ravipati, A., Grams, B., Chowdhury, S., Alatise, O., and Are, C. (2017). "Colorectal cancer?global burden, trends, and geographical variations." *Journal of surgical oncology*, 115:619–630.

Fang, Y.-J., Wu, X.-J., Zhao, Q., Li, L.-R., Lu, Z.-H., Ding, P.-R., Zhang, R.-X., Kong, L.-H., Wang, F.-L., Lin, J.-Z., et al. (2013). "Hospital-based colorectal cancer survival trend of different tumor locations from 1960s to 2000s." *PloS one*, 8:e73528.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2014). "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012." *International journal of cancer*, 136:E359–E386.

Fu, J., Yang, J., Tan, Y., Jiang, M., Wen, F., Huang, Y., Chen, H., Yi, C., Zheng, S., and Yuan, Y. (2014). "Young patients (? 35years old) with colorectal cancer have worse outcomes due to more advanced disease: a 30-year retrospective review." *Medicine*, 93:1–7.

Ghabeljoo, M., Jafarabadi, M. A., Mohammadi, S. M., Hajizadeh, E., Kazemnejad, A., and Fatemi, S. R. (2011). "Patterns of survival for anatomical sites of colorectal cancer with shift to advanced lesions in Iran." *Asian Pac J Cancer Prev*, 12:1225–31.

Ghazali, A. K., Musa, K. I., Naing, N. N., and Mahmood, Z. (2010). "Prognostic factors in patients with colorectal cancer at Hospital Universiti Sains Malaysia." *Asian Journal Of Surgery*, 33:127–133.

Ghee, L. K. (2014). "A Review of Colorectal Cancer Research in Malaysia." *Med J Malaysia*, 69:22–32.

GLOBOCAN (2012). Estimated cancer incidence, mortality and prevalence worldwide. international agency for research on cancer; 2012. http://globocan.iarc.fr/Default.aspx.Accessed04.08.15.

Golan, T., Urban, D., Berger, R., and Lawrence, Y. R. (2013). "Changing prognosis of metastatic colorectal adenocarcinoma: Differential improvement by age and tumor location." *Cancer*, 119:3084–3091.

Greene, F., Page, D., Fleming, I., Fritz, A., and Balch, C. (2002). "AJCC Cancer Staging Manual Sixth Edition." Technical report, American Joint Committee On Cancer.

Haggar, F. A. and Boushey, R. P. (2009). "Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors." *Clinics in colon and rectal surgery*, 22:191.

Hamilton, W., Round, A., Sharp, D., and Peters, T. (2005). "Clinical features of colorectal cancer before diagnosis: a population-based case–control study." *British journal of cancer*, 93:399.

Harris, A., Bowley, D., Stannard, A., Kurrimboccus, S., Geh, J., and Karandikar, S. (2009). "Socioeconomic deprivation adversely affects survival of patients with rectal cancer." *British Journal of Surgery*, 96:763–768.

Hassan, M. O., Arthurs, Z., Sohn, V. Y., and Steele, S. R. (2009). "Race does not impact colorectal cancer treatment or outcomes with equal access." *The American Journal of Surgery*, 197:485–490.

Hassan, M. R. A., Khazim, W. K. W., Mustapha, N. R. N., Said, R. M., Tan, W. L., Suan, M. A. M., and Soelar, S. A. (2014). "The Second Report of the National Cancer Patient Registry-Colorectal Cancer, 2008-2013." Technical report, National Cancer Patient Registry-Colorectal Cancer and Clinical Research Centre (CRC).

Heidarnia, M. A., Monfared, E. D., Akbari, M. E., Yavari, P., Amanpour, F., and Mohseni, M. (2013). "Social determinants of health and 5-year survival of colorectal cancer." *Asian Pacific Journal of Cancer Prevention*, 14:5111–5116.

Hendifar, A., Yang, D., Lenz, F., Lurje, G., Pohl, A., Lenz, C., Ning, Y., Zhang, W., and Lenz, H.-J. (2009). "Gender disparities in metastatic colorectal cancer survival." *Clinical Cancer Research*, 15:6391–6397.

Hilmi, I., Hartono, J., and Goh, K. (2010). "Negative perception in those at highest risk–potential challenges in colorectal cancer screening in an urban asian population." *Asian Pac J Cancer Prev*, 11:815–822.

Hines, R., Markossian, T., Johnson, A., Dong, F., and Bayakly, R. (2014). "Geographic residency status and census tract socioeconomic status as determinants of colorectal cancer outcomes." *American journal of public health*, 104:e63–e71.

Hole, D. and McArdle, C. (2002). "Impact of socioeconomic deprivation on outcome after surgery for colorectal cancer." *British Journal Of Surgery*, 89:586–590.

Jafarabadi, M. A., Mohammadi, S. M., Hajizadeh, E., Kazemnejad, A., and Fatemi, S. R. (2011). "Does the prognosis of colorectal cancer vary with tumor site?" *Gastroenterology and Hepatology from bed to bench*, 4:199.

Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B., and Berry, D. A. (2013). "Meta-analyses of colorectal cancer risk factors." *Cancer causes & control*, 24:1207–1222.

Kelsall, H. L., Baglietto, L., Muller, D., Haydon, A. M., English, D. R., and Giles, G. G. (2009). "The effect of socioeconomic status on survival from colorectal cancer in the Melbourne Collaborative Cohort Study." *Social science & medicine*, 68:290–297.

Kong, C.-K., Roslani, A. C., Law, C.-W., Law, S., and Arumugam, K. (2010). "Impact of socio-economic class on colorectal cancer patient outcomes in Kuala Lumpur and Kuching, Malaysia." *Asian Pac J Cancer Prev*, 11:969–974.

Kotake, K., Asano, M., Ozawa, H., Kobayashi, H., and Sugihara, K. (2016). "Gender differences in colorectal cancer survival in Japan." *International journal of clinical oncology*, 21:194–203.

Le, H., Ziogas, A., Lipkin, S. M., and Zell, J. A. (2008). "Effects of socioeconomic status and treatment disparities in colorectal cancer survival." *Cancer Epidemiology and Prevention Biomarkers*, 17:1950–1962.

Le, H., Ziogas, A., Taylor, T. H., Lipkin, S. M., and Zell, J. A. (2009). "Survival of distinct Asian groups among colorectal cancer cases in California." *Cancer*, 15:259–270.

Lee, Y.-C., Lee, Y.-L., Chuang, J.-P., and Lee, J.-C. (2013). "Differences in survival between colon and rectal cancer from SEER data." *PloS one*, 8:e78709.

Lemmens, V., Janssen-Heijnen, M., Verheij, C., Houterman, S., Repelaer van Driel, O., and Coebergh, J. (2005). "Co-morbidity leads to altered treatment and worse survival of elderly patients with colorectal cancer." *British Journal of Surgery*, 92:615–623.

Li, J., Liu, J., Gao, C., Liu, F., and Zhao, H. (2017). "Increased mortality for colorectal cancer patients with preexisting diabetes mellitus: an updated meta-analysis." *Oncotarget*, 8:62478.

Li, J., Wang, Z., Yuan, X., Xu, L., and Tong, J. (2015). "The prognostic significance of age in operated and non-operated colorectal cancer." *BMC Cancer*, 15:83.

Li, Q., Cai, G., Li, D., Wang, Y., Zhuo, C., and Cai, S. (2014). "Better long-term survival in young patients with non-metastatic colorectal cancer after surgery, an analysis of 69,835 patients in SEER database." *PLoS One*, 9:e93756.

LIN, J.-T., WANG, W.-S., YEN, C.-C., LIU, J.-H., YANG, M.-H., CHAO, T.-C., CHEN, P.-M., and CHIOU, T.-J. (2005). "Outcome of colorectal carcinoma in

patients under 40 years of age." *Journal of Gastroenterology and Hepatology*, 20:900–905.

Lydrup, M.-L. and Höglund, P. (2015). "Gender aspects of survival after surgical treatment for rectal cancer." *Colorectal Disease*, 17:390–396.

Magaji, B. A., Moy, F. M., Roslani, A. C., and Law, C. W. (2014). "Descriptive epidemiology of colorectal cancer in University Malaya Medical Centre, 2001 to 2010." *Asian Pac J Cancer Prev*, 15:6059–6064.

Majek, O., Gondos, A., Jansen, L., Emrich, K., Holleczek, B., Katalinic, A., Nennecke, A., Eberle, A., Brenner, H., Group, G. C. S. W., et al. (2013). "Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany." *PLoS One*, 8:e68077.

Manser, C. N. and Bauerfeind, P. (2014). "Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: a systematic review." *GASTROINTESTINAL ENDOSCOPY*, 80:42–60.

McArdle, C., McMillan, D., and Hole, D. (2003). "Male gender adversely affects survival following surgery for colorectal cancer." *British Journal of Surgery*, 90:711–715.

McKay, A., Donaleshen, J., Helewa, R. M., Park, J., Wirtzfeld, D., Hochman, D., Singh, H., and Turner, D. (2014). "Does Young Age Influence The Prognosis Of Colorectal Cancer: A Population-based Analysis." *World journal of surgical oncology*, 12:370.

Meguid, R. A., Slidell, M. B., Wolfgang, C. L., Chang, D. C., and Ahuja, N. (2008). "Is there a difference in survival between right-versus left-sided colon cancers?" *Annals of Surgical Oncology*, 15:2388.

Moghimi-Dehkordi, B. and Safaee, A. (2012). "An overview of colorectal cancer survival rates and prognosis in Asia." *World journal of gastrointestinal oncology*, 4:71.

Naing, C., Lai, P. K., and Mak, J. W. (2017). "Immediately modifiable risk factors attributable to colorectal cancer in Malaysia." *BMC public health*, 17:637.

Nitzkorski, J. R., Willis, A. I., Nick, D., Zhu, F., Farma, J. M., and Sigurdson, E. R. (2013). "Association of race and socioeconomic status and outcomes of patients with rectal cancer." *Annals of surgical oncology*, 20:1142–1147.

Oliphant, R., Nicholson, G., Horgan, P., Molloy, R., McMillan, D., and Morrison, D. (2013). "Contribution of surgical specialization to improved colorectal cancer survival." *British Journal of Surgery*, 100:1388–1395.

Omar, Z. A. and Ibrahim Tamin, N. (2011). "National Cancer Registry Report Malaysia Cancer Statistics Data and Figure 2007." *Malaysia: National Cancer Registry*, pages 85–7.

Ordóñez-Mena, J., Walter, V., Schöttker, B., Jenab, M., O?Doherty, M. G., Kee, F., Bueno-de Mesquita, B., Peeters, P., Stricker, B. H., Ruiter, R., et al. (2017). "Impact of prediagnostic smoking and smoking cessation on colorectal cancer prognosis: a meta-analysis of individual patient data from cohorts within the CHANCES consortium." *Annals of Oncology*, 29:472–483.

Orsini, R. G., Verhoeven, R. H., Lemmens, V. E., van Steenbergen, L. N., de Hingh, I. H., Nieuwenhuijzen, G. A., and Rutten, H. J. (2015). "Comparable survival for young rectal cancer patients, despite unfavourable morphology and more advanced-stage disease." *European Journal of Cancer*, 51:1675–1682.

Peng, L., Ling, C., Yew, C., Aizen, J., Yun, L., and Ho, W. (2013). "Singapore Cancer Registry Interim Annual Registry Report: Trends in Cancer Incidence in Singapore 2009-2013." *National Registry of Diseases Office (NRDO)*.

Powell, A., Wallace, R., McKee, R., Anderson, J., Going, J., Edwards, J., and Horgan, P. (2012). "The relationship between tumour site, clinicopathological characteristics and cancer-specific survival in patients undergoing surgery for colorectal cancer." *Colorectal Disease*, 14:1493–1499.

Price, T. J., Beeke, C., Ullah, S., Padbury, R., Maddern, G., Roder, D., Townsend, A. R., Moore, J., Roy, A., Tomita, Y., et al. (2015). "Does the primary site of colorectal cancer impact outcomes for patients with metastatic disease?" *Cancer*, 121:830–835.

Quah, H., Chou, J., Gonen, M., Shia, J., Schrang, D., and Saltz, L. (2008). "Pathologic stage is most prognostic of disease-free survival in locally advanced rectal cancer patients after preoperative chemoradiation." *Cancer*, 113:57–64.

Quaresma, M., Coleman, M. P., and Rachet, B. (2015). "40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study." *The Lancet*, 385:1206–1218.

Rashid, M., Aziz, A. F. A., Ahmad, S., Shah, S. A., and Sagap, I. (2009). "Colorectal cancer patients in a tertiary referral centre in Malaysia: a five year follow-up review." *Asian Pac J Cancer Prev*, 10:1163–6.

Sabounchi, S., Keihanian, S., and Anand, B. S. (2012). "Impact of race on colorectal cancer." *Clinical Colorectal Cancer*, 11:66–70.

Sack, J. and Rothman, J. M. (2000). "Colorectal cancer: natural history and management." *Hospital Physician*, 36:64–73.

Samat, N. and Shattar, A. (2014). "Spatial accessibility of colorectal cancer patients to health care facilities in Kelantan State: a study on medical geography." *World Applied Science Journal*, 31:1772–1782.

Samat, N., Shattar, A. K. A., Sulaiman, Y., Manan, A. A., and Weng, C. N. (2013). "Investigating geographic distribution of colorectal cancer cases: an example from Penang state, Malaysia." *Asian Social Science*, 9:38.

Shah, S. A., Neoh, H. M., Rahim, S., Azhar, Z. I., Hassan, M. R., Safian, N., and Jamal, R. (2014). "Spatial analysis of colorectal cancer cases in Kuala Lumpur." *Asian Pac J Cancer Prev*, 15:1149–1154.

Sharp, L., McDevitt, J., Brown, C., Carsin, A.-E., and Comber, H. (2017a). "Association between smoking at diagnosis and cause-specific survival in patients with rectal cancer: Results from a population-based analysis of 10,794 cases." *Cancer*, 123:2543–2550.

Sharp, L., McDevitt, J., Brown, C., and Comber, H. (2017b). "Smoking at diagnosis significantly decreases 5-year cancer-specific survival in a population-based cohort of 18 166 colon cancer patients." *Alimentary pharmacology & therapeutics*, 45:788–800.

Simpson, J. and Scholefield, J. H. (2008). "Treatment of colorectal cancer: surgery, chemotherapy and radiotherapy." *Surgery-Oxford International Edition*, 26:329–333.

Stelzner, S., Hellmich, G., Koch, R., and Ludwig, K. (2005). "Factors predicting survival in stage IV colorectal carcinoma patients after palliative treatment: a multivariate analysis." *Journal of Surgical Oncology*, 89:211–217.

Stornes, T., Wibe, A., Romundstad, P., and Endreseth, B. (2014). "Outcomes of rectal cancer treatment - influence of age?" *International journal of colorectal disease*, 29:825–834.

Tang, V., Boscardin, W. J., Stijacic-Cenzer, I., and Lee, S. J. (2015). "Time to benefit for colorectal cancer screening: survival meta-analysis of flexible sigmoidoscopy trials." *bmj*, 350:h1662.

US Department of Health and Human Services et al. (2014). "The health consequences of smoking?50 years of progress: a report of the Surgeon General." *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*, 17:1081.

Van de Poll-Franse, L., Haak, H., Coebergh, J., Janssen-Heijnen, M., and Lemmens, V. (2012). "Disease-specific mortality among stage I–III colorectal cancer patients with diabetes: a large population-based analysis." *Diabetologia*, 55:2163–2172.

Van der Pool, A., Damhuis, R., Ijzermans, J., de Wilt, J., Eggermont, A., Kranse, R., and Verhoef, C. (2011). "Trends in incidence, treatment and survival of patients with stage IV colorectal cancer: a population-based series." *Colorectal Disease*, 14:56–61.

Veach, E., Xique, I., Johnson, J., Lyle, J., Almodovar, I., Sellers, K. F., Moore, C. T., and Jackson, M. C. (2014). "Race matters: analyzing the relationship between colorectal cancer mortality rates and various factors within respective racial groups." *Frontiers in public health*, 2:239.

Walter, V., Jansen, L., Hoffmeister, M., and Brenner, H. (2014). "Smoking and survival of colorectal cancer patients: systematic review and meta-analysis." *Annals of Oncology*, 25:1517–1525.

Wei, E. K., Giovannucci, E., Wu, K., Rosner, B., Fuchs, C. S., Willett, W. C., and Colditz, G. A. (2004). "Comparison of risk factors for colon and rectal cancer." *International journal of cancer*, 108:433–442.

Widdison, A., Barnett, S. W., and Betambeau, N. (2011). "The impact of age on outcome after surgery for colorectal adenocarcinoma." *Ann R Coll Surg Engl 2011; 93: 445?4502*, 93:445–450.

Yeole, B., Sunny, L., Swaminathan, R., Sankaranarayanan, R., and Parkin, D. (2001). "Population-based survival from colorectal cancer in Mumbai,(Bombay) India." *European Journal Of Cancer*, 37:1402–1408.

Yuan, Y., Li, M.-D., Hu, H.-G., Dong, C.-X., Chen, J.-Q., Li, X.-F., Li, J.-J., and Shen, H. (2013). "Prognostic and survival analysis of 837 Chinese colorectal cancer patients." *World Journal Of Gastroenterology*, 19:2650.

Zanders, M., Vissers, P., Haak, H., and van de Poll-Franse, L. (2014). "Colorectal cancer, diabetes and survival: epidemiological insights." *Diabetes & metabolism*, 40:120–127.

Chapter 3

The Effect of Individual-Level Factors in Survival Prognosis for Colorectal Cancer in Malaysia

3.1 Introduction

Cancer has been reported as the third leading cause of death in Malaysia by the Ministry of Health in Malaysia, after heart disease and pulmonary circulatory disease, and septicaemia. The five most common cancers in Malaysia in 2007, as reported by the National Cancer Registry of Malaysia, were breast cancer, colorectal cancer, lung cancer, nasopharyngeal cancer and cervical cancer at 18.1%, 12.3%, 10.2%, 5.2% and 4.6% of total cancers respectively (Omar and Ibrahim Tamin, 2011).Colorectal cancer is the second most common type of cancer for both sexes, exceeded only by lung cancer in men and breast cancer in women (Omar and Ibrahim Tamin, 2011).

Cancer survival in Malaysia has been studied at a regional level. A retrospective record review study analysed survival rates of patients diagnosed with colorectal cancer between 1996 and 2005 in Hospital Universiti Sains Malaysia (HUSM), Kelantan, (Ghazali et al., 2010). The study reported that the overall 5-year survival rate was 34.3% and three significant independent prognostic factors identified were Dukes staging at diagnosis, the presence of liver metastases and treatment modalities.

To date, there is no study looking into survival and prognostic factors for colorectal cancer for the whole Malaysian population. In this chapter, we aim to identify the effect of individual-level factors on survival prognosis for patients with colorectal cancer in Malaysia.

3.2 Data

The original data that we received from the database of the NCPR-CC consisted of 4501 patients with histologically verified primary colorectal cancer diagnosed between 2008 and 2013 (ICD-10, C18-C20). After excluding patients without Malaysian citizenship, patients with negative age and negative survival time, there were 4412 subjects' data available for analysis.

Variables chosen to be used to describe patients' social demographic were age, race, gender, education level and smoking status. The 'age' of the patients is defined as the age when the patients were diagnosed with colorectal cancer, calculated as the difference between the date of diagnosis and date of birth. 'Race' was recorded as Malay, Chinese, Indian, other Malaysian, foreigner, missing, or unknown race. 'Education level' was based on the patient's highest grade/education level completed or achieved; this is recorded as 'nil' if the patient did not receive any formal education. Primary school consists of six years of education,

referred to as Standard one to Standard six. Five years of Secondary school education follow, for children aged 13 to 17. After Secondary school, tertiary education can be undertaken. This includes pre-university, matriculation and other higher levels of education, and can be at public or private institutions of higher learning. 'Smoking status' indicates if the patient had a confirmed history of usage of any form of tobacco. This includes cigarette, cigar and pipe smoking, and tobacco chewing. They are categorized as 'Never' or 'Non-smoker', 'Former smoker' (quit > 30 days), 'current smoker', 'unknown', or 'missing'.

For clinical characteristics, the presence of diabetes mellitus, tumour site, stage at diagnosis, the presence of metastases, and tumour differentiation were chosen to be included in the study. Presence of diabetes mellitus was recorded as 'No', 'Yes', 'unknown', or 'missing'. 'Tumour site' describes the site of the primary tumour at diagnosis, and is categorised according to ICD-10 classifications as either at the colon, the rectosigmoid junction or the rectum. These, along with 'missing', comprise the variable levels used in this data set. Stage of cancer is based on final TNM Staging, recorded in the data as Stage I, II, IIA, IIB, IIIA, IIIB, IV, 'not staged', 'missing', 'not available' or as the missing value NA. The categorisation of 'tumour differentiation' is based on the proportion of the tumour that is gland forming. If the tumour is $> 95\%$ gland forming, it is recorded as well differentiated; 50-95% gland forming as moderately differentiated, and $< 50\%$ gland forming as poorly differentiated. In the data given to us, tumour differentiation was recorded as 'well differentiated', 'moderately differentiated', 'poorly differentiated', 'not applicable', 'missing' and 'NA'. Distant metastases were reported either by surgical or histopathological report and have been separated into each possible metastatic site. Data on each site of metastasis is recorded as 'No', 'Yes' and 'NA'.

The 'treatment modalities' variables record seven different types of treatment received by patients, namely: no therapy, radiotherapy, chemotherapy and biological therapy, supportive care/palliative care, endoscopy, complementary/alternative

treatment and unknown. Each of these was recorded as either 'TRUE' or 'FALSE' for the stated treatment.

Patient status in this data is given as either dead, alive, or lost to follow up. The date and cause of death were documented in our data where applicable. For each patients, survival time was computed from the date of diagnosis to date of death and follow up was done to the end of 2013 or the censoring date in this study was 31st of December 2013.

3.2.1 Modification of the Data Categories

There are many instances in this dataset of information being missing or incomplete, but these are recorded inconsistently with a variety of indicators such as 'missing', 'not applicable', 'not available', 'unknown' or 'NA'. These do not always match up with the categories defined for each variable in the data. We therefore decided to combine these various missing data as 'unknown' or 'not applicable' if these categories were stated as such in the patients' form, or just as 'missing' otherwise. We also obtained advice from our data provider about the 'uncertain' category recorded in the database to justify our decision in data categorization. For example, even though the variable 'tumour differentiation' has categories of 'not applicable', 'not available' and 'missing', most of the data in these categories did not tally with the data definition. Our data provider clarified this, and assured us that all data in those categories of this variable were actually just 'missing'.

In consideration of the social demographic variables, we firstly recategorised race as Malay and non-Malay because of the small number of patients in each race that was non-Malay. Smoking status was categorised to: non-smoker, former smoker, active smoker and missing status. Education level was classified as 'Nil', 'Primary', 'Secondary', 'Tertiary' and 'missing' group.

With regard to the clinical characteristics, diabetes status was simplified to the three categories: yes, no and missing. 'Cancer site' was unchanged, but 'stage of cancer' was recategorized to 'Stage I', 'StageII', 'StageIII', 'Not staged', and 'missing'. Since the 'site of distant metastases' contains many categories, each of which only contains a small number of data points, we decided to combine all the metastases regardless of their specific site. Irrespective of where and when it has been detected, the 'presence of distant metastases' was therefore reclassified into 'yes', 'no' and 'missing'. 'Tumour differentiation' was recorded as 'well', 'moderate', 'poor' and 'missing'.

The treatment modalities were categorized into four types of treatment received by the patients. They are patients who underwent surgery alone, patients who underwent surgery followed by chemotherapy and/or radiotherapy, patients who underwent chemotherapy and/or radiotherapy and patients who got other alternative treatments or palliative care. Patients without information of treatment received were recorded as an unknown group.

The specific cause of death provided in the data was not verified and could not therefore be deemed reliable, so we decided to perform the analysis on all-cause mortality. The data records whether each patient is dead or alive at the end of the study period; we relabel each patient's status as either dead or censored. The censored group are patients who were alive until the end of the study period as recorded in the database.

3.2.2 About Missing Data

There are a total of 4412 patients in this study and only 742 patients have complete data for all chosen covariates. The reason behind the degree of missing data is likely that data collection and data entry procedures are not standardised across all hospital sites. Patient information was gathered either from the physician directly

or by Registry staff reading through the patients' notes, it is written onto a patient form and then entered into the database by the cancer registry staff. There is a possibility of third party error in the database. Because of these irregularities, we feel the combination of the various missing data categories described in the previous section to be justified.

We compare the overall survival using Kaplan Meier survival curves for data with all observations (4412) and the data with complete cases only (742 observations). The overall survival probability in all data was significantly lower compared to those with the complete covariate data (Figure 3.3). We also investigated how the risk of death increased with more missing covariates, and we acknowledge that there is a possibility of introducing bias if we ignore or exclude the cases with missing data (Figure 3.2). Thus, we decided to keep the missing data in this study as a 'missing' group, choosing to analyze separately the full data set (allowing 'missing' as one of the categories) and the data with complete cases.

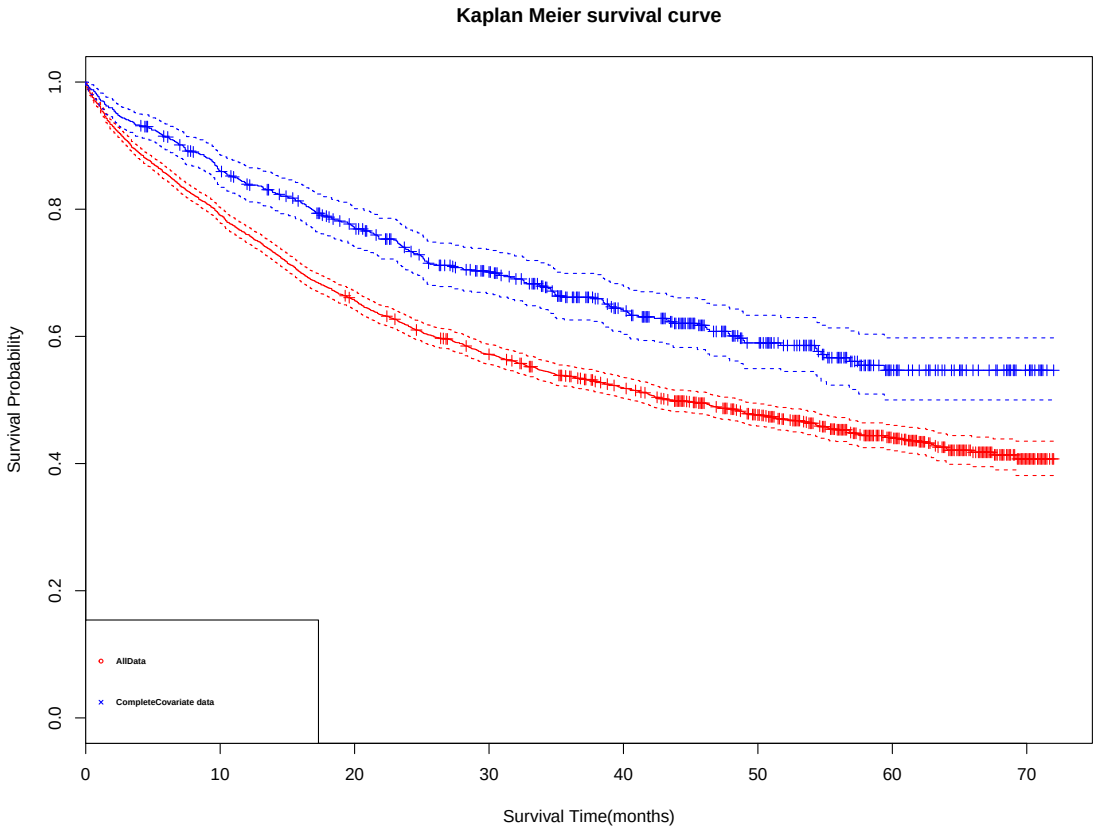


FIGURE 3.1: Kaplan Meier plot comparing survival probability for all data vs complete covariate data

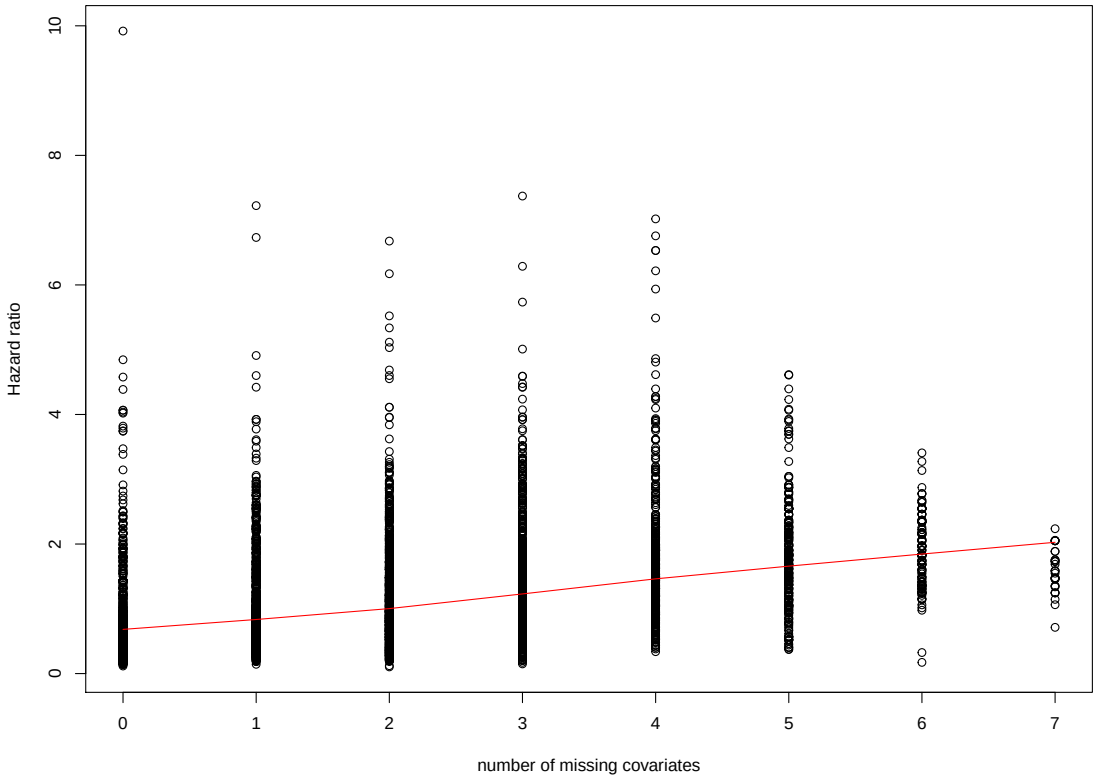


FIGURE 3.2: The hazard of death increases with increasing 'missing' covariates

3.3 Survival Probabilities

The probability of surviving until a particular time is presented in a plot of survival curves, calculated using the Kaplan Meier estimator (Kaplan and Meier, 1958), also known as the product limit estimator, which can be calculated by most common statistical software packages. The estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps determined by the observed survival and censored times (Hosmer and Lemeshow, 1999). The Kaplan Meier estimator is given by

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right)$$

where n_j is the number of individuals at risk, d_j is the number of deaths in the time interval from $t_{(k)}$ to $t_{(k+1)}$, with k= 1, 2, 3,... r events occurring at times $t_1 \leq t < t_2 \leq \ldots \leq t_r$. The Kaplan Meier estimate of the survivor function is a step function, in which the estimated survival probabilities are constant between adjacent death times (regardless of further censored events occurring) and decrease at each death time.

We also calculated the overall 5-year survival rate and the overall median survival rate for colorectal cancer patients in our study. The survival rate is the percentage of people in a study who are still alive for a certain period of time after they have been diagnosed with the disease. In this study, the 5-year survival rate for colorectal cancer is defined as the percentage of patients with colorectal cancer who are still alive five years after diagnosis. The median survival time is the first observed time at which cumulative survival was 50% or less; that is, the time when half of the patients have died.

3.4 Statistical Model

3.4.1 Fitting a Cox Proportional Hazard Regression Model

A Cox regression model was used to explore the relationship between patients' survival and the chosen explanatory variables (Cox, 1972) . The Cox model is written as

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

$$h_i(t) = h_0(t) \exp\{X_i\beta\}$$

$h_i(t)$ gives the value of the hazard function for the i^{th} individual at a particular time t. Here X_i is a vector of explanatory variables for individual i, β is a vector of regression coefficients to be estimated and h_0 is the baseline hazard function, which is left unspecified (though it can be estimated after the model has been fitted). The baseline hazard tells us how the rate of events changes over time: whether it increases, decreases or remains constant. Analysis was performed using the `survival` package in `R` (Therneau, 2015).

The reader will note that we have chosen not to undertake variable selection in the development of our models. The main reason for this is that we are not seeking to identify new predictors of survival in our study, rather we are seeking to evaluate the role of established predictors of survival and adjust those estimates to account for the other well-known confounding factors. While the present study is not sufficiently powered to be able to detect statistically significant effects of all variables at the 5% level, going forward in time, as cancer registries collect more data, and more complete data, it will be important to include each of these known predictors. At some point in the future, it would will be appropriate to undertake a study that seeks to identify *which are* the important variables in this context. Our opinion is that it is too early, at present, to undertake such a study.

3.4.2 Model Diagnostics

The validity of the proportional hazards (PH) assumption can be assessed using Scaled Schoenfeld residuals. We used the `cox.zph` function from the `survival` package in `R` (Therneau, 2015). We tested the PH assumptions on each of the covariates in the model and for the overall model, which is the global test (Fox

and Weisberg, 2011). The null hypothesis is that the (scaled) Schoenfeld residuals are independent of failure times or equivalently that the hazard is constant over time. This implies that the Schoenfeld residuals are uncorrelated with time. A p-value was computed for each covariate, along with a global test for the model as a whole.

We also fitted the Bayesian Cox time-varying coefficient model using the `R` package `dynsurv` to check if a model with time-varying coefficients is more appropriate. This package has a function `bayesCox` which is used to fit the time-varying coefficient model (Wang et al., 2017):

$$h_i(t) = h_o(t) \exp\{X_i \beta(t)\}$$

i.e the regression coefficients are allowed to vary with time. We looked at the result using a range of priors and produce the plot for each variable. We produced plots of $\beta(t)$ against t to check for time-varying effects.

We checked the fit of the model and identified influential observations by performing diagnostic checks on the standardised residuals. Martingale residuals were used to determine the correct functional form of continuous covariates; a straight line plot indicates that a linear term is needed. The df-beta residual plot was used to identify influential observations. Each of the potential extreme observations was assessed for its influence by considering the percentage change in the variable's coefficient ($\Delta\hat{\beta}\%$) when the model is refitted without this observation. The clinical plausibility of the observation was also considered in assessing the influential cases (Hosmer and Lemeshow, 1999).

Cox Snell residuals were used for assessing overall model fit and determining whether the Cox regression model is a suitable choice for this data. A straight line with a unit slope and zero intercept indicates that the fitted survival model is satisfactory. Finally, findings are presented with hazard ratio (HR) and 95% confidence interval.

3.5 Results

3.5.1 Descriptive Statistics

3.5.1.1 Summary Statistics - A Profile of Colorectal Cancer Patients in Malaysia

The number of colorectal cancer cases reported in the Registry increased by around 34%, from 639 in 2008 to 964 in 2009/10. The cases started to decrease in 2010 to only 363 cases in 2013. There were 1979 (45%) deaths from all cases in the 6 year period, while the remaining 2433 (55%) were alive and institutionally censored (Table 3.1).

TABLE 3.1: Colorectal Cancer Cases by Year

Year	Number of cases	Died, n (%)	Censored, n (%)
2008	639	377 (59)	262 (41)
2009	964	517 (54)	447 (46)
2010	964	473 (49)	491 (51)
2011	734	313 (43)	421 (57)
2012	748	248 (33)	500 (67)
2013	363	51 (14)	312 (86)
Total	**4412**	**1979 (45)**	**2433 (55)**

Patients' ages ranged from 9 to 98 years old with mean 61.2 years and standard deviation 12.7. 55% of the patients were male and there were more male deaths. Table 3.2 illustrates the socio-demographic profile of colorectal cancer patients in

this study. 43.1% patients were of the Malay race, 56.4% were non-Malays and 22% had missing ethnicity. In the six-year duration of the study, 10% were active smokers, 12% are former smokers and 35% were non-smokers. 59% of patients in the data had no information on their level of education (Table 3.2).

22% of the colorectal cancer patients were recorded in the registry as having diabetes mellitus. The commonest tumour site, according to ICD-10, was in the colon, followed by the rectum and the rectosigmoid junction. Among patients for whom cancer stage was recorded, the majority were diagnosed at Stage III, according to their final TNM staging. With regards to tumour differentiation, 57% of patients were classified as having a moderately differentiated tumour.

Surgery is the most common treatment for colorectal cancer. The majority of patients (70.4%) had surgery as their primary treatment. Some of them had chemotherapy or radiotherapy or both in addition to surgery. A small number of patients (3.0%) had other treatments which include alternative treatment or palliative care (Table 3.3).

There were a lot of missing data in our study. The highest proportion of missing data was seen in the 'Education Level' variable (59.2%), though smoking status and cancer stage also had 43.4% and 34.6% missing respectively.

3.5.1.2 Survival Probabilities and Kaplan Meier Survival Curves

The Kaplan Meier estimates for all data are shown in Figure 3.3. The overall survival rate at the end of the first year was almost 80%, and the overall survival rate for the entire 5-year period in this study was 44%, 95% CI (42%,46%).

Kaplan Meier survival curves show that female patients had better overall survival compared to males but the difference was not statistically significant. People of the Malay race show significantly lower survival compared to those of non-Malay

TABLE 3.2: Profile of colorectal cancer patients - socio-demographic characteristics

Variables	Number of patients, (%)	Died, n (%)	Censored, n (%)
Total(N)	**4412 (100)**	**1979 (49)**	**2433 (55)**
Sex			
Female	1894 (42.9)	824 (41.6)	1070 (44.0)
Male	2470 (56.0)	1142 (57.7)	1328 (54.6)
Missing	48 (1.1)	13 (0.7)	35 (1.4)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Race			
Non Malays	2489 (56.4)	1043 (52.7)	1446 (59.4)
Malays	1901 (43.1)	929 (46.9)	972 (40.0)
Missing	22 (0.5)	7 (0.4)	15 (0.6)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Smoking status			
Non smoking	1543 (35.0)	643 (32.5)	900 (37.0)
Former smoker	528 (12.0)	248 (12.5)	280 (11.5)
Active smoker	423 (9.6)	214 (10.8)	209 (8.6)
Missing	1918 (43.4)	874 (44.2)	1044 (42.9)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Education			
Nil	399 (9.0)	201 (10.2)	198 (8.1)
Primary	553 (12.6)	247 (12.5)	306 (12.6)
Secondary	651 (14.8)	271 (13.7)	380 (15.6)
Tertiary	195 (4.4)	60 (3.0)	135 (5.6)
Missing	2614 (59.2)	1200 (60.6)	1414 (58.1)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**

origin. The tertiary education group showed significantly better overall survival compared to other education groups and survival rates got progressively worse for lower education groups. Non-smokers showed higher overall survival than all other smoking status groups (Figure 3.4).

The Kaplan Meier plot illustrates that colon cancer patients had better overall survival compared to patients in the other groups. The overall survival of rectal cancer and rectosigmoid cancer were very similar for the first two years after diagnosis, though rectal cancer had a marginally better long term prognosis. The overall survival rate for rectosigmoid cancer was better than that for rectal cancer beyond two years after diagnosis.

The overall survival for those with moderately differentiated tumours was significantly better compared to poorly differentiated tumours and the 'not applicable' group. The overall survival was also significantly different for varying stages of

TABLE 3.3: Profile of colorectal cancer patients - clinical characteristics

Variables	Number of patients, (%)	Died, n (%)	Censored, n (%)
Total(N)	**4412 (100)**	**1979 (49)**	**2433 (55)**
Diabetes			
No	3021 (68.4)	1349 (68.2)	1672 (68.7)
Yes	982 (22.3)	449(22.7)	533 (21.9)
Missing	409 (9.3)	181 (9.1)	228 (9.4)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Cancer site			
Colon	2394 (54.3)	981 (49.6)	1413 (58.1)
Rectosigmoid	631 (14.3)	305 (15.4)	326 (13.4)
Rectum	1379 (31.2)	692 (34.9)	687 (28.2)
Missing	8 (0.2)	1 (0.1)	7 (0.3)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Staging			
Stage I	215 (4.9)	34 (1.7)	181 (7.4)
Stage II	600 (13.6)	141 (7.1)	459 (18.9)
Stage III	802 (18.2)	320 (16.2)	482 (19.8)
Stage IV	647 (14.7)	455 (23.0)	192 (7.9)
Not staged	620 (14.0)	213 (10.8)	407 (16.7)
Missing	1528 (34.6)	816 (41.2)	712 (29.3)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Tumour Grade			
Well	358 (8.1)	129 (6.5)	229 (9.4)
Moderate	2497 (56.7)	986 (49.8)	1511 (62.1)
Poor	161 (3.6)	96 (4.9)	65 (2.7)
Missing	1396 (31.6)	768 (38.8)	628 (25.8)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Distant Metastases			
No	1922 (43.6)	630 (31.8)	1292 (53.1)
Yes	1521 (34.4)	771 (39.0)	750 (30.8)
Missing	969(22.0)	578(29.2)	391(16.1)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**
Treatment modalities			
Surgery Alone	1658 (37.6)	659 (33.3)	999 (41.1)
Surgery Chemo Radio	1454 (32.9)	604 (30.5)	850 (34.9)
Chemo Radio	437 (9.9)	233 (11.8)	204 (8.4)
Other treatments	140 (3.2)	100 (5.0)	40 (1.6)
Unknown	723 (16.4)	383 (19.4)	340 (14.0)
Total	**4412 (100)**	**1979 (100)**	**2433 (100)**

cancer at diagnosis, with the poorest survival seen in Stage IV and the highest survival in Stage I. The overall survival for 'unknown' stage was inbetween the survival rates for Stage III and Stage IV. The overall survival in the group displaying no metastases was significantly better than the group known to have metastases present, but the 'unknown' group had the lowest overall survival of the three.

Survival was not significantly affected by the presence of diabetes. Considering treatment modalities, the overall survival rates among patients who received

surgery with adjuvant therapies (chemotherapy or/and radiotherapy) were higher than those of the 'surgery alone' group in the first two years after diagnosis; though the surgery alone group had better overall survival thereafter. The overall survival of those who had radio/and chemotherapy alone was not much different from the 'unknown' treatment group. The lowest survival was seen in the 'others' treatment group (Figure 3.5).

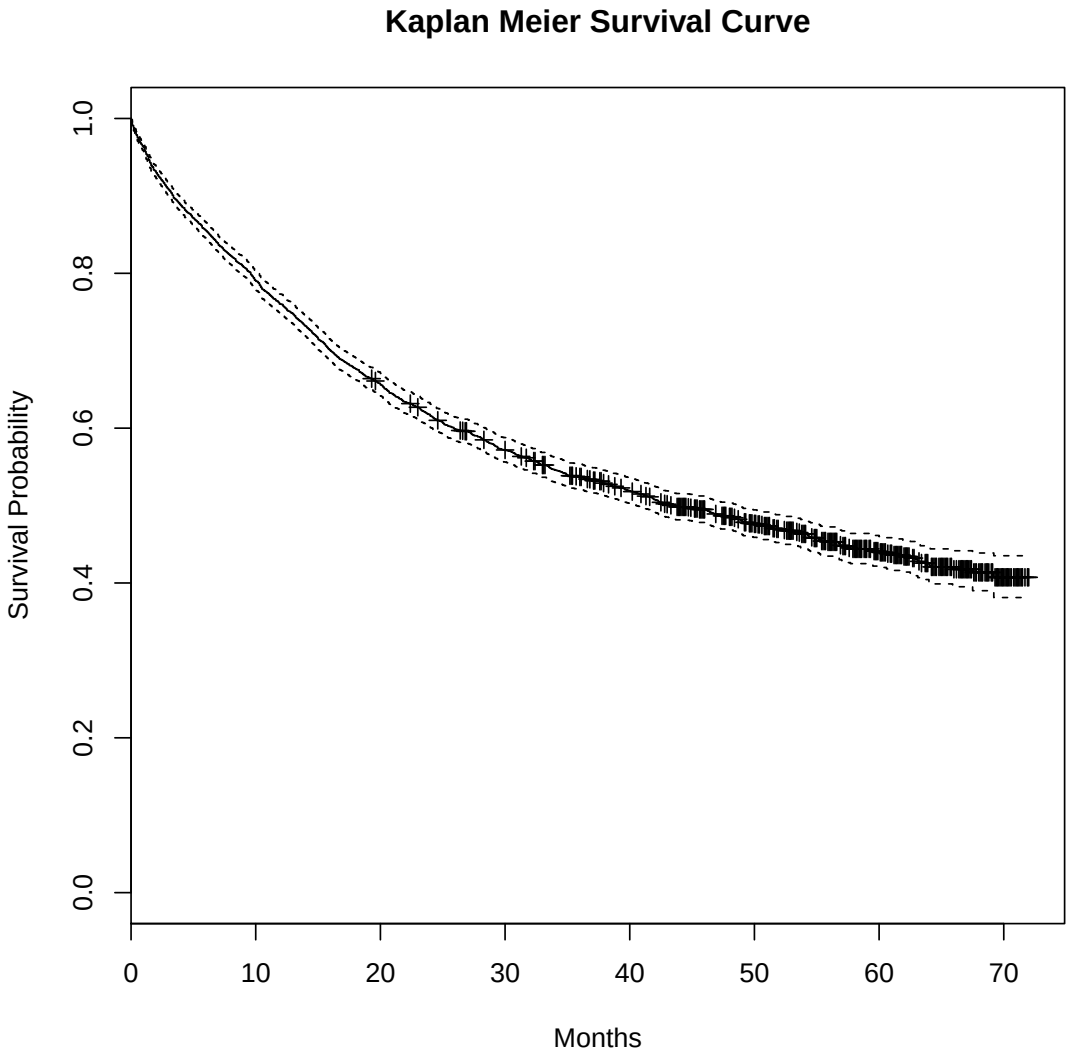


FIGURE 3.3: Kaplan Meier curve shows the overall survivor function for colorectal cancer cases in the study

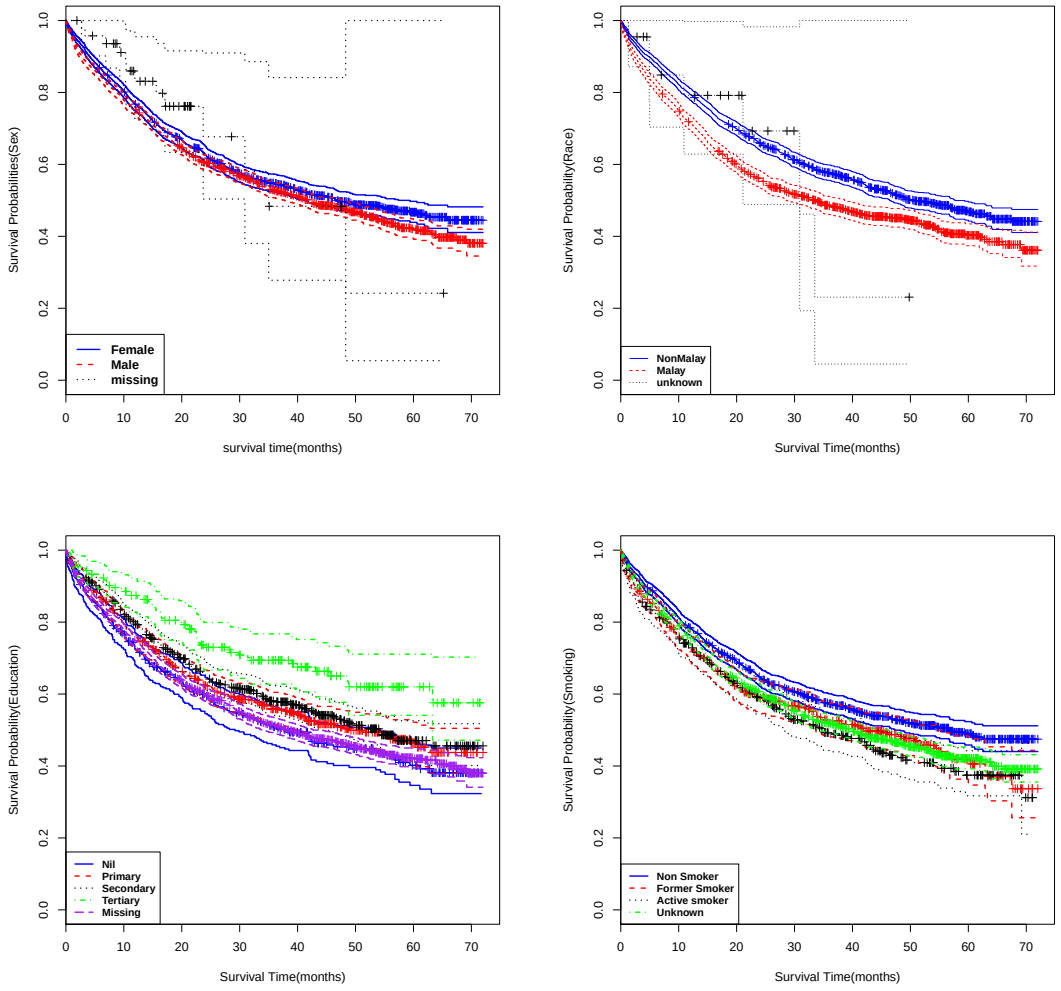


FIGURE 3.4: Kaplan Meier survival curves for sex, race, smoking and education

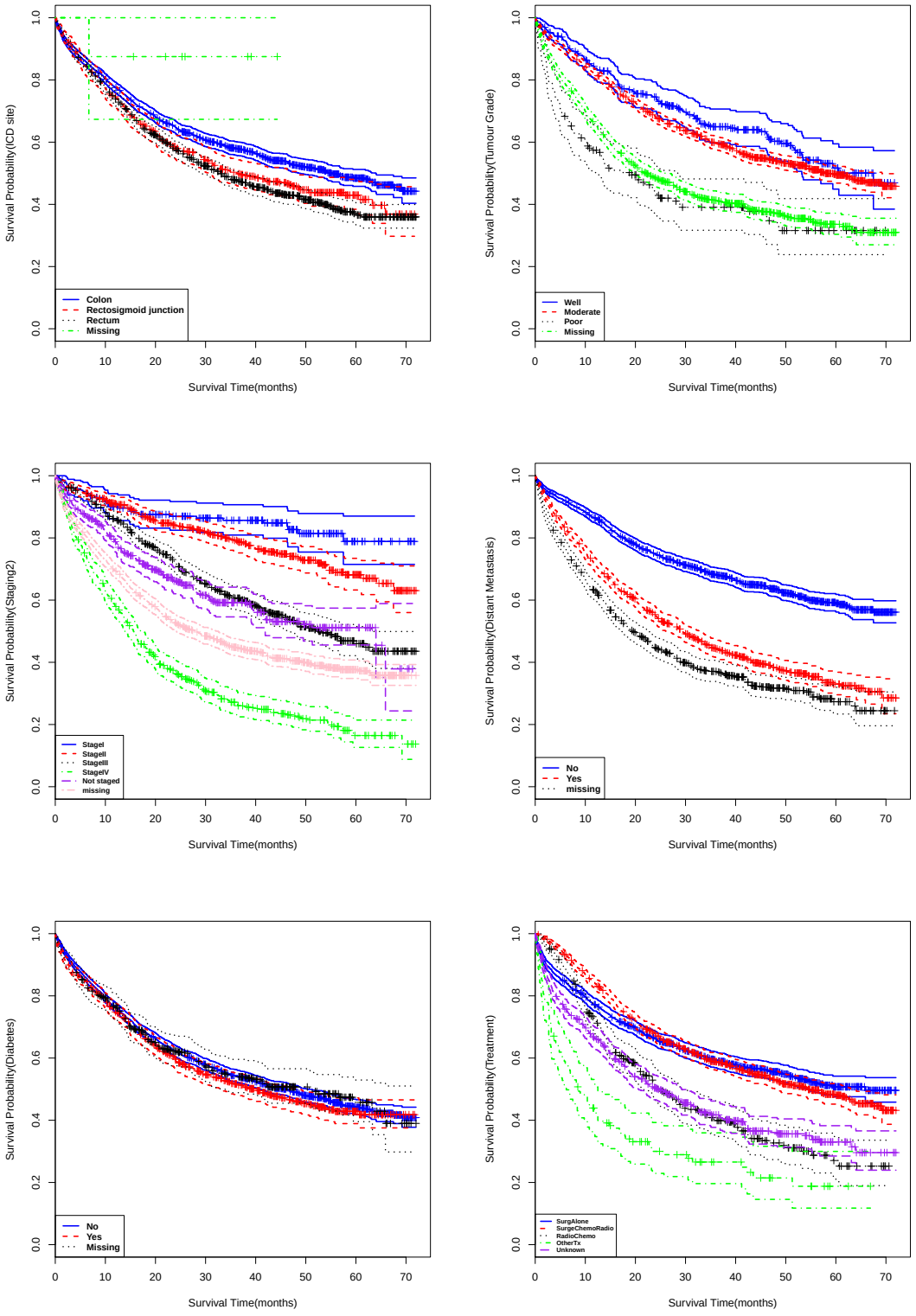


FIGURE 3.5: Kaplan Meier survival curves for cancer site, tumour grade, stage of cancer, presence of distant metastases, diabetes and treatment modalities

3.5.2 Cox Regression Model: Deducing The Effect of Individual-Level Factors on Colorectal Cancer Survival

We fitted a Cox regression model to each of two sets of data: one set containing all data (4412 cases) and one comprising only those with complete covariate information (742 cases). Age, third-degree education, cancer staging, poor tumour differentiation, the presence of distant metastases and receiving 'other' treatments were factors that increased the risk of death for colorectal cancer patients in both models.

Our model shows that the most important prognostic factor affecting survival in colorectal cancer patients is the stage at diagnosis. Patients diagnosed when their cancer was already at Stage IV had an almost 6-fold greater risk of dying from colorectal cancer than those who were diagnosed at Stage I.

Patients who were active smokers had a greater risk of dying compared to non-smokers in both models but the effect was not significant in the model with all data. Malay race and tumour at the recto-sigmoid junction increased significantly the risk of dying in the model with all patients. Diabetic patients had a significantly greater risk of dying compared to non-diabetic patients in the model with complete covariate data, but this effect was not significant in the all-case data set.

Overall, the factors that affect the survival of patients with colorectal cancer were not greatly different in each model. So, we decided to choose the model with all-case data as our final model. Table 3.4 presents the effects of covariates on colorectal cancer survival, given in terms of hazard ratios.

TABLE 3.4: Cox Regression Model for complete covariate data and all data

Variable	No. cases (N=4412)	All cases HR, (95% CI)	Complete cases HR, (95% CI)
Age		1.01 (1.01, 1.02)	1.01 (1.00, 1.02)
Sex			
Female	1894	1	1
Male	2470	1.01 (0.91, 1.11)	1.08 (0.79, 1.46)
Missing	48	0.67 (0.38, 1.18)	
Race			
Non Malay	2489	1	1
Malay	1901	1.32 (1.21, 1.45)	1.22 (0.95, 1.58)
Missing	22	1.38 (0.64, 2.97)	
Smoking			
Non-smoking	1543	1	1
Former Smoker	528	1.25 (1.07, 1.47)	1.02 (0.72, 1.46)
Active smoker	423	1.17 (0.99, 1.39)	1.70 (1.16, 2.50)
Missing	1918	1.04 (0.93, 1.18)	
Education			
No Education	399	1	1
Primary	553	0.92 (0.76, 1.12)	0.85 (0.58, 1.23)
Secondary	651	0.96 (0.79, 1.16)	0.97 (0.66, 1.43)
Tertiary	195	0.60 (0.45, 0.81)	0.56 (0.32, 0.98)
Missing	2614	1.01 (0.87, 1.19)	
Diabetes			
No	3021	1	1
Yes	982	1.07 (0.96, 1.20)	1.43 (1.06, 1.92)
Missing	409	0.85 (0.72, 1.00)	
Cancer site			
Colon	2394	1	1
Rectosigmoid junction	631	1.16 (1.02, 1.32)	0.87 (0.60, 1.28)
Rectum	1379	1.08 (0.97, 1.19)	0.88 (0.66, 1.17)
Missing	8	0.24 (0.03, 1.72)	
Staging			
Stage I	215	1	1
Stage II	600	1.53 (1.05, 2.24)	1.68 (0.81, 3.47)
Stage III	802	2.86 (2.00, 4.08)	3.72 (1.86, 7.45)
Stage IV	647	5.59 (3.90, 8.00)	6.02 (2.85, 12.70)
Not staged	620	2.66 (1.85, 3.84)	3.98 (1.85, 8.57)

3.5.2.1 Checking collinearity

In our analyses, if some of the independent variables were highly correlated, then the problem of multicollinearity could arise. These are a concern in data analyses because it means that at one predictor variable can be used to predict another - this creates redundant information in the model.

Since all of our variables are categorical (except for age), and computing correlation of such variables is not ideal, we looked at the correlation between the estimated parameters of the model instead. Our results shows no collinearity problem in our variables/predictors in the model. We additionally used the Variation inflation factor to check for multicollinearity. If the VIF for each variable in the model was less than 10, this indicates that there was no multicollinearity problem in the model. If the VIF was more than 10, this means that there is a multicollinearity problem and the variable should be excluded from the model. The variables demonstrated Variance Inflation Factor (VIF) of less than 10 which indicated no multicollinearity problem among those variables

3.5.2.2 Checking for Interactions

The possibleinteraction between variables in the model were checked by fitting cross product terms. The interactions were checked between the variables smoking,education, treatment, staging, site, tumor diffrentiation and metastasis.No interaction was found to significantly improve model fit

3.5.2.3 Diagnostic Test: Checking the Proportional Hazards Assumption

The test statistics based on the Schoenfeld residuals indicate that there is a statistically significant trend with time in the residuals for five of our variables (Staging,

Treatment, Race, Grading and ICD Site). A global test of residuals for the whole model was also found to be highly significant ($p < 0.001$); this indicates that the discrepancy between predicted and observed covariate value does vary with time and that the proportional hazards assumption may not hold.

The plot in `R` translates the residuals by adding $\hat{\beta}$, so that covariates without time-trend should form a smooth line close to the fixed estimate of β along the whole time range. The Schoenfeld residual plots (Figure 3.6) suggest some departure from the proportional hazard assumption, but the trend is not consistent. The apparent initial positive slope for some variables maybe overly influenced by one or two large negative residuals.

We therefore fitted a dynamic survival model, detailed on page 62, using the `dynsurv` package. We produced plots of $\beta(t)$ against t (Figure 3.7) and these showed no evidence that the coefficients of variables in the model do vary over time. The confidence intervals and the effect size remain roughly constant over time.

We experimented with a range of different priors for the time-varying coefficient model, created using `bayesCox` in the `dynsurv` package, and the results do not show any trends with time. Residual plots for these time-dependent models (Figure 3.7) show no evidence that the coefficients of variables in the model do vary over time. Therefore, even though the Schoenfeld residuals initially suggest some departure from the PH assumption, it was not considered necessary to introduce time-varying covariates to the model, as evidenced by the plots from the dynamic survival model. Therefore, we decided that use of the Cox proportional hazard model as our final model was justified.

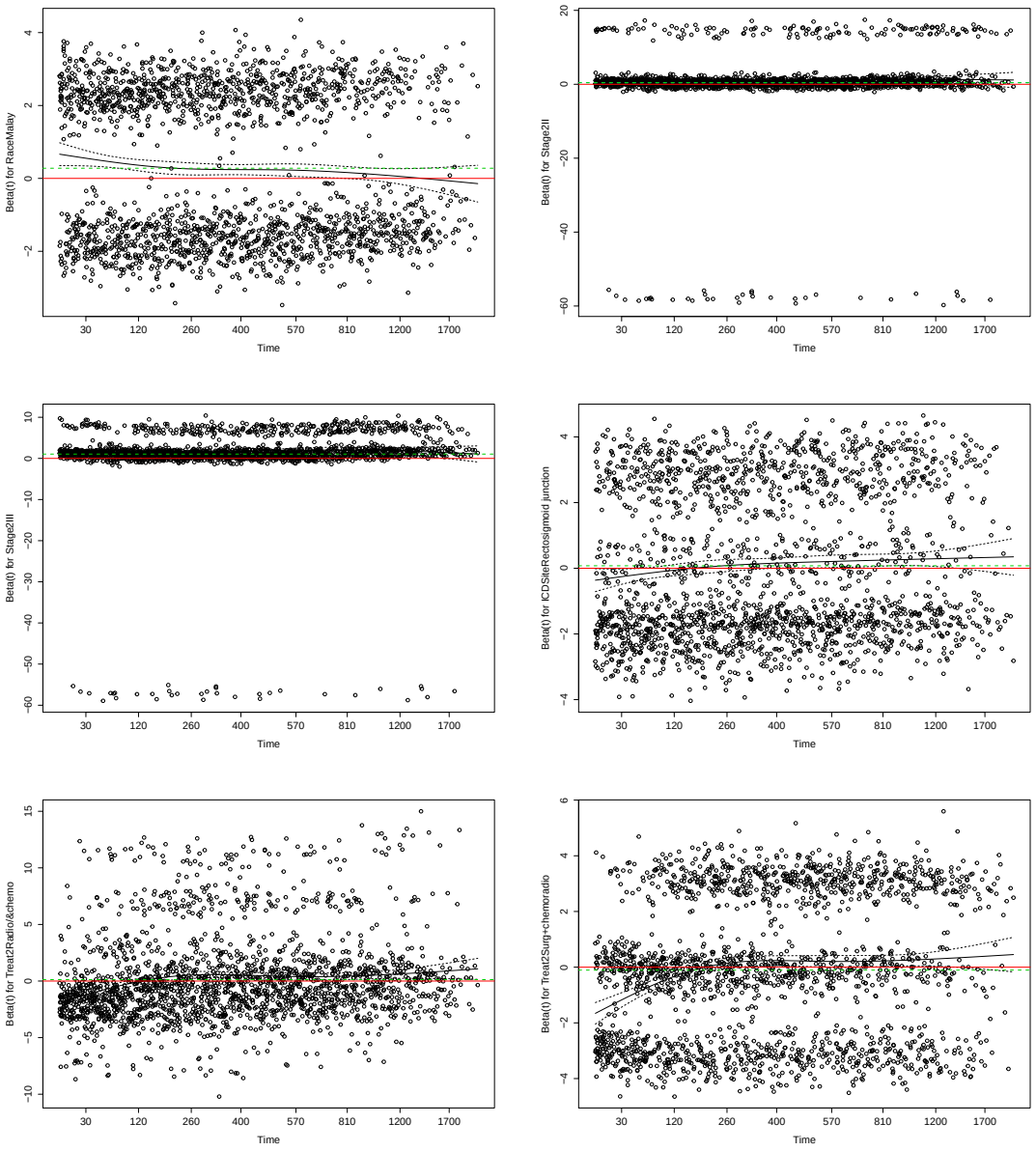


FIGURE 3.6: Scaled Schoenfeld residual plot

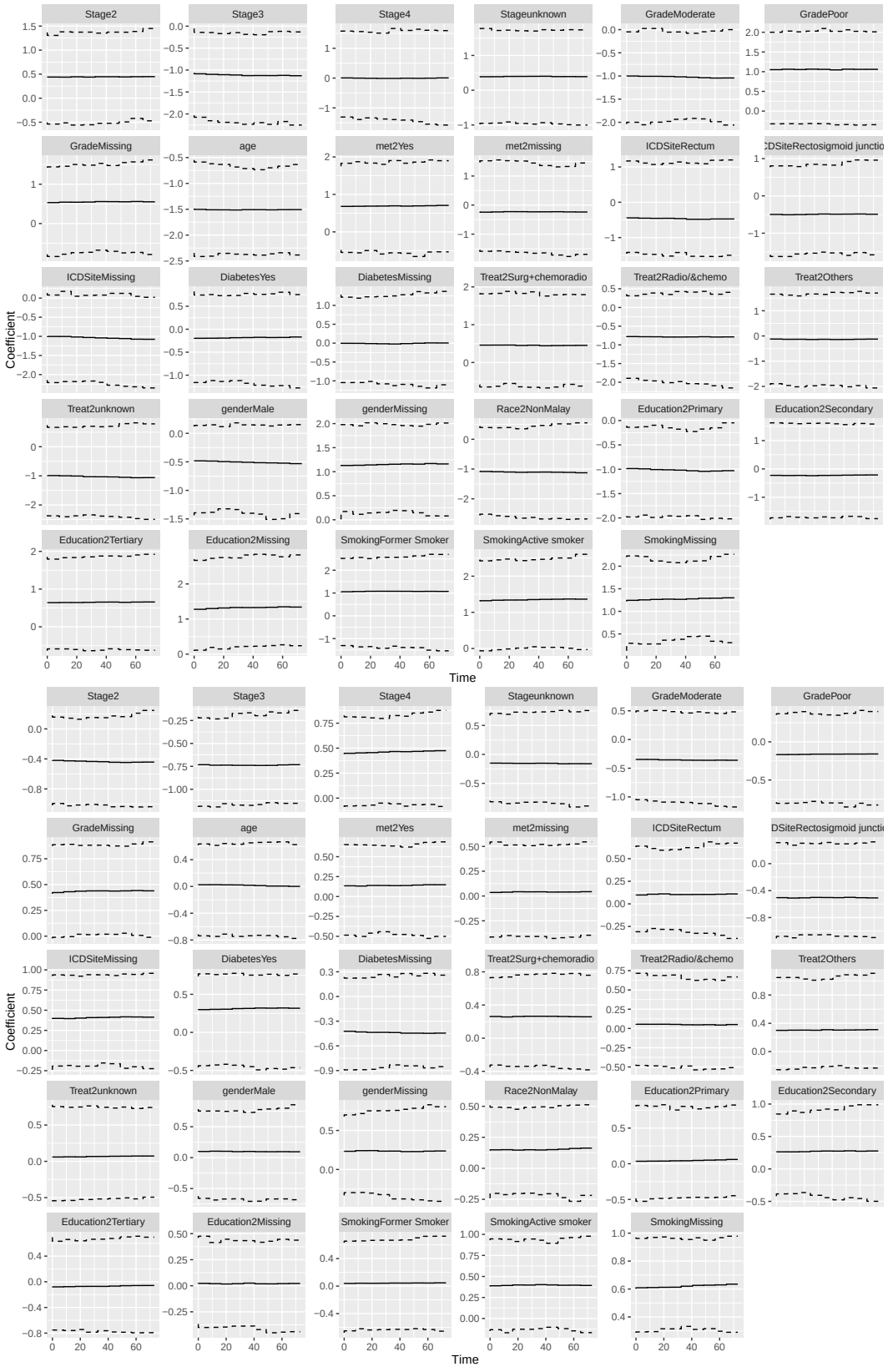


FIGURE 3.7: Plots of residuals derived from the time-varying coefficient model

The Cox-Snell residuals showed that our final Cox regression model fits the data

very well. This was shown by the plot of cumulative hazard versus Cox-Snell residual which produced an almost straight line with a unit slope and zero intercept as shown in Figure 3.8.

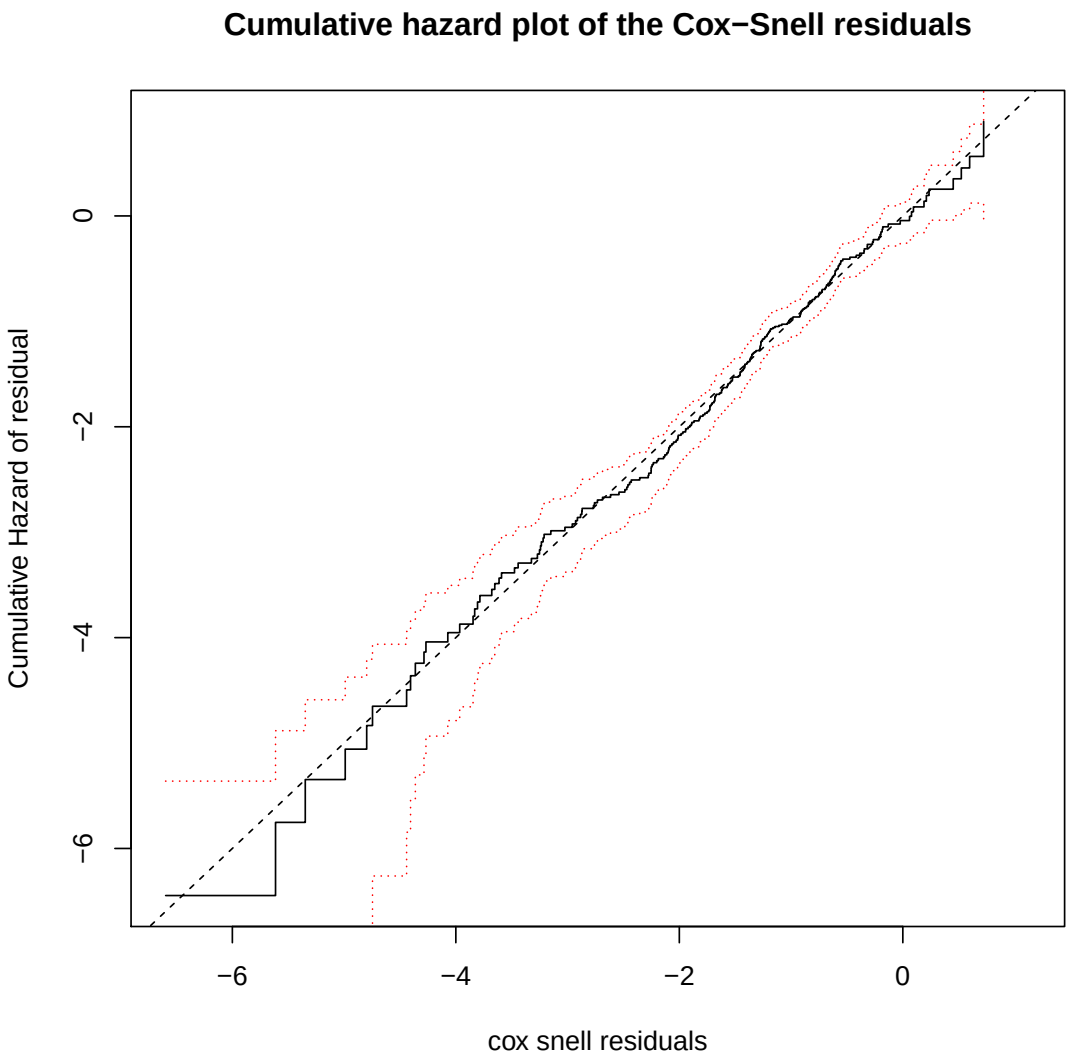


FIGURE 3.8: Cox-Snell Residual for assesing model fitness

3.6 Discussion

This study is a preliminary investigation of the factors that will influence the survival of patients who have been diagnosed with colorectal cancer. The data we have used come from the Malaysian National Cancer Patient Registry for Colorectal Cancer (NCPR-CC), which sources the data from 34 reference centres for colorectal cancer across Malaysia. Our analysis is based on data from 6 years of diagnoses and contains records on 4412 people, who are all Malaysian citizens.

An alternative approach that we could be applied to our data is the net survival method (Perme et al., 2012). In estimating net survival, there are approaches that can be applied: the cause specific approach and the all-cause mortality approach. The first requires knowing the cause of death for each individual and the second one requires all-cause mortality in the study group and an estimate of the "expected" mortality of a disease-free group having the same demographic characteristics as the study group. In a situation where the cause of death is not known, net survival estimators assume that the available expected cancer mortality rate correctly reflects mortality rates from other causes . This can be obtained from general population life tables. The mortality due to cancer can then be deduced from the all-cause and other-cause mortalities. This is also known as the excess hazard, if it refers to hazard.

Epidemiological studies usually use the second of the above approaches because cause of death is often unavailable or unreliable. Cancer epidemiologists use "Relative-survival methods" for net survival estimation when there is no information on the cause of death (Roche et al., 2013). The net survival method is applicable to measure cancer survival after excluding the influence of other causes of death and is very useful in large scale studies such as EUROCARE-4 (European cancer registry based study on survival and care of cancer patients) and iCONCORD (Global surveillance of cancer survival study) (Coleman et al., 2008,

Sant et al., 2009) where population survival differs substantially between countries (Perme et al., 2012).

We may consider using this method in future research with access to more data in order to get better picture of survival comparisons between countries, for instance extending our study to South East Asian countries. However, in this thesis we chose to use the more widely applied Cox model; one advantage of doing this is that there already exists software for handling our planned work in spatial survival data analysis, though the extension to spatial net survival models would be novel and useful to the research community.

Older age, having Malay ethnicity, higher cancer staging, the presence of distant metastases, having a poorly differentiated tumour, having 'other' treatment modalities and having ever smoked were all associated with statistically significant increases in risk of death in patients with colorectal cancer, while higher education level statistically decreased the risk of dying from colorectal cancer.

We found that cancer staging was the main factor affecting the risk of death in patients diagnosed with colorectal cancer. This data shows that there was an almost six-fold increase in the risk of death in patients with stage IV cancer compared to those with stage I cancer. This is consistent with previous findings (Kotake et al., 2016, Maringe et al., 2013, McKay et al., 2014) and is obviously supported by biological theory because the stage of cancer describes the extent and severity of the disease at the commencement of treatment.

Many studies have reported that age was an independent prognostic factor for colorectal cancer (McKay et al., 2014, Stornes et al., 2014, Widdison et al., 2011). Fu et al. (2014) noted that the youngest age group in their study ($\leq$35 years old) had significantly poorer overall survival rate compared to older patients due to the greater proportion of these younger patients presenting at a late stage of cancer. However, having accounted for other variables, age was not an independent

prognostic factor in their study. Previous research analyzed age as a categorical variable, while our study analyzed age as a continuous variable. The data supported a linear trend in log risk with age. Our findings show that there is a very slight increase in the risk of death for each extra year of age at diagnosis.

Malaysia is a multi-ethnic country; the three main ethnic groups in Malaysia are Malay, Chinese and Indian. The predominant ethnic group in Malaysia is Malay, constituting 63.1% of the total population in Peninsular Malaysia. Hassan et al. (2016) reported that the Chinese ethnicity has the highest age adjusted mortality rate (11.85 deaths per 100,000 population), compared to 9.56 and 7.08 deaths per 100,000 population in Malay and Indian races respectively. However, they did not assess the risk of death from colorectal cancer for each of these ethnic groups separately. In our study, we have re-categorized race into Malay and Non-Malay to compare their risk of death from colorectal cancer. There are fewer Malay patients than non-Malay in this dataset (43% vs 56.4%), and we find that Malays had shorter survival than non-Malays; they had a 30% greater risk of dying from colorectal cancer than the non-Malays in our study.

Different populations may have different types and levels of education and little is hitherto known about how patient education level may affect survival of colorectal cancer. Cavalli-Björkman et al. (2011) grouped their patients into three levels of education: low, middle and high education, and found that both middle and high education had significantly lower risk of death compared to the low education group. They assume that the low education group had received less treatment and different types of surgery than others in the two higher educational categories.

For our data, we found that patients with a tertiary-level education had a 40% lower risk of death than those without any formal education. Education level is often associated with socioeconomic status and those with higher socioeconomic status commonly demonstrate better survival than those who do not. One study

has suggested that higher survival rate can be achieved if the colorectal cancer patients adopted w ere to adopt a healthy lifestyle, and had better accessibility to medical care and higher level education (Rasouli et al., 2017).

In the meantime, public health information could be disseminated widely to help the population better what signs to look for and when to visit the doctor should early signs of cancer manifest themselves. Information on of available screening of screening programs should be made widely available to the population to encourage attendance.

Smoking has been reported to lead to poorer survival in colorectal cancer patients.Walter et al. (2014), Phipps et al. (2013) and (Boyle et al., 2013) all support the finding that smoking adversely affected all-cause mortality in colorectal cancer patients. A possible mechanism by which smoking impacts colon cancer survival is related in some way, to surgery. A recent study found that colon cancer patients who were current smokers who received surgery alone had a significantly higher risk of death (Adjusted HR = 1.14, 95% CI: 1.07;1.12) from colon cancer compared to those in the never smoker group (Sharp et al., 2017).

It is reasonable to expect an effect of cancer incidence from smoking, but there is also evidence that smoking affects survival from colorectal cancer (Phipps et al., 2011).The association between current smoker and survival of colorectal cancer in our study was slightly insignificant but we did see a significant association between smoking and reduced survival in the former smoker group. Former smokers had a statistically significant 25% increase in the risk of dying from colorectal cancer compared to non-smokers. The risk for former smokers is higher even than for active smokers, who have a 17% increased risk of death compared to non-smokers. This could be because the classification of former smoker in this study is quite short (quit $>$ 30 days). We did not take into account the length of time for which ex-smokers had not smoked, or the number of cigarettes smoked per day. The

effect of smoking may still be present if the former smokers had smoked for long periods of time, for example, more than 10 years. A previous study (Chao et al., 2000) on cigarette smoking and cancer mortality reported that there is a significant relationship between duration of smoking with mortality from colorectal cancer. Moreover, the risk of deaths among former smokers decreases with the number of years since they have stopped smoking.

Furthermore, in our study, cases's smoking status was derived by from medical records, and was not routinely recorded therein. Consequently assessement and assigning of exposure to smoking may be subject to misclassification, as we noted earlier that smoking also was found as the variable with the highest missing value in our data. Therefore, any association of between current smoker and survival from colorectal cancer in our study may not be picked up by our study. We suggest registries make sure that this variable is routinely documented in cancer registration form as it is an important variable to look for.

Cancer in the rectosigmoid junction and rectum are often grouped together as rectal cancer or as left-sided colorectal cancer, and the outcome from these alternative classifications may be different to that which we observe here. Previous studies have reported that cancer at the right-side (colon) has poorer outcomes than the cancer at the left-side (combining recto-sigmoid and rectal cancers) (Price et al., 2015). Our study classified the cancer sites separately as colon, rectosigmoid juction or rectum and we found that the cancer site was not a significant predictor of outcome.

With regards to treatment modalities, our findings showed that the combination of surgery and therapies (chemotherapy and/or radiotherapy) decreased the risk of death compared to those who had surgery alone. This finding was similar to a previous work on Malaysian colorectal cancer patients diagnosed in 2008-2009 by Hassan et al. (2016). They suggested that their results were an indicator of the

importance of adjuvant therapies (chemotherapy or radiotherapy or both) after surgery to prolong the patients' survival from colorectal cancer. On the other hand, patients with non-surgical treatments, such as chemotherapy, radiotherapy or other alternative treatments, are reported to have a higher risk of death (Ghazali et al., 2010, Yeole et al., 2001).

This study demonstrated a slightly increased risk of death in men, similar to the findings of Lydrup and Höglund (2015), but the result here was not statistically significant. Cheung et al. (2013) stated a possibility that men may have a biologically more aggressive disease or a poorer response to adjuvant therapies, which contributes to their poorer survival. On the other hand, Majek et al. (2013) concluded that a better prognosis among women with colorectal cancer may be explained by the use of postmenopausal hormonal replacement therapy (HRT). Another previous study, Mandelson et al. (2003), suggested that HRT extended survival in colon cancer.

Having diabetes mellitus, the known presence of distant metastases, and having a poorly differentiated tumour were all found to be strong prognostic factors in other studies (Aguero et al., 2012, Mills et al., 2013). In our study, the presence of distant metastases and having a poorly differentiated tumour do have a significant effect on the survival outcome for colorectal cancer patients, but having diabetes mellitus does not. This contradictory result may be explained by the low number of patients with this characteristic in our study compared to previous studies.

A particularly problematic aspect of our study was the large amount of missing data. Much covariate data was unavailable; for instance, 59% of data in the 'education level' variable was missing. We noted that possible reasons for the data being missing were non-standardised data collection and data entry procedures, and the possibility of third party error in the recording of the data. We explored the missing data cases and found that the risk of death increased with increasing

number of missing covariates in our cases. As this might introduce bias into our analysis, we decided not to ignore the cases with missing data; the missing values were likely to be connected with clinically relevant factors such as cancer staging. We assessed the likely effect on our results in two ways. First, we ran the analysis including 'missing' as a level of each predictor variable. Secondly, we analysed the subset of the study population for those who have complete data (n=742). We compared the point estimate of the hazard ratio for both models. It was reassuring to see that the results were very similar for both complete and all-case datasets. We decided to choose all-data analysis as our main model as it is the most appropriate dataset to use for the spatial analysis to be conducted in the next chapter; this next analysis will involve the whole population and where they live.

Despite this challenge, this remains the largest study of colorectal cancer survival ever carried out in the country. With increasing rates of chronic disease in developing economies such as Malaysia, the results from this study are important and will help in understanding the individual effects on colorectal cancer prognosis in Malaysia.

We found that more than half of the patients presented late in the disease at Stage III, Stage IV and missing stages. Our analysis revealed that the severity of the disease lead to poor prognosis in colorectal cancer in the population after adjusting for other individual characteristics such as age, gender and ethnicity. The five year survival for patients with stage III cancer around 20%. In United States(US), the equivalent figures are between 90 and 53% for Stage III and around 12% for Stage IV (American Cancer Society, 2018). Cancer survival is affected by a number of factors. Stage at diagnosis is one, and the success of the treatment regime another. Age, race, smoking, and unemployment are others (Shi et al., 2013). To improve survival in Malaysia one strategy would be to increase the percentage of cases who present for diagnosis at StageI or II. The five year survival rate in US for colon

cancer patients with stageI and II cancers are from 95 to 62%(American Cancer Society, 2018).

The success of such a strategy would depend on a number of factors. Public health campaigns can be helpful in persuading people with possible early signs of cancer, such as blood in stool, to go to the doctor. For example the National Bowel Cancer campaign in the UK has been successful in raising awareness, though they recognise that a subsequent change in survival will take time to achieve (Cancer Research UK, 2014).

In Malaysia, in contrast to UK, there are barriers to health care based on income and social and economic circumstance. These too must be lessened or removed so that money and physical availability of suitable healthcare does not prevent people getting diagnosed and cared for when they have cancer.

Additionally, we think that health education programs targeting high risk group and emphasizing the importance of early detection of cancer as well as knowledge on the importance of cancer treatment should be implemented. Formulation of a better screening program needs to be extended so that it is a genuinely national program. Currently, Malaysia has no national screening for colorectal cancer. More promotional activities with regards to cancer are recommended to increase population survival rate in the future.

From the Kaplan Meier survival curve, we noted that in the non-education group had the poorest survival. Their survival was similar to those in missing education group, and it is possible that a bias has been introduced here since those with education missing might be more likely to be of lower educational status. Even though we found no interaction between cancer staging and education in this study, it is likely that education plays an important role in survival from colorectal cancer.

Bringing awareness of public health issues to the population can be fitted by a general improvement in education level across populations, so benefits would accrue

were the general population better informed about risks for poor health in middle age. In the UK in 1999 general advice was given to the British population on good lifelong health by the chief medical officer. One of the ten points was "take up cancer screening opportunities". Michael Marmot recognises that improving knowledge does not improve behaviour, but actions to take up screening will help identify cases sooner, which usually benefits survival (Marmot, 2015). Emphasis needs to be given to the importance of early detection of colorectal cancer; both by increasing public awareness among the Malaysian population and by the formulation of a better screening program. This early detection would lead to diagnosis at the earlier stages of cancer, which we have herein shown to best increase survival rates in patients with colorectal cancer.

References

Aguero, F., Murta-Nascimento, C., Gallén, M., Andreu-García, M., Pera, M., Hernández, C., et al. (2012). "Colorectal cancer survival: results from a hospital-based cancer registry." *Rev Esp Enferm Dig*, 104:572–7.

American Cancer Society (2018). Survival rates for colorectal cancer, by stage. https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html#references.

Boyle, T., Fritschi, L., Platell, C., and Heyworth, J. (2013). "Lifestyle factors associated with survival after colorectal cancer diagnosis." *British journal of cancer*, 109:814.

Cancer Research UK (2014). Campaign evaluation. https://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/be-clear-on-cancer/bowel-cancer-campaign/evaluation#BCOC_bowel_evaluation1.

Cavalli-Björkman, N., Lambe, M., Eaker, S., Sandin, F., and Glimelius, B. (2011). "Differences according to educational level in the management and survival of colorectal cancer in Sweden." *European journal of cancer*, 47:1398–1406.

Chao, A., Thun, M. J., Jacobs, E. J., Henley, S. J., Rodriguez, C., and Calle, E. E. (2000). "Cigarette smoking and colorectal cancer mortality in the cancer prevention study II." *Journal of the National Cancer Institute*, 92:1888–1896.

Cheung, W. Y., Shi, Q., O'connell, M., Cassidy, J., Blanke, C. D., Kerr, D. J., Meyers, J., Van Cutsem, E., Alberts, S. R., Yothers, G., et al. (2013). "The predictive and prognostic value of sex in early-stage colon cancer: a pooled analysis of 33,345 patients from the ACCENT database." *Clinical colorectal cancer*, 12:179–187.

Coleman, M. P., Quaresma, M., Berrino, F., Lutz, J.-M., De Angelis, R., Capocaccia, R., Baili, P., Rachet, B., Gatta, G., Hakulinen, T., et al. (2008). "Cancer survival in five continents: a worldwide population-based study (CONCORD)." *The lancet oncology*, 9:730–756.

Cox, D. (1972). "Regression models and life-tables." *Journal of the Royal Statistical Society. Series B (Methodological)*, 34:187–220.

Fox, J. and Weisberg, S. (2011). "Cox proportional hazards regression for survival data in R." *An Appendix to An R Companion to Applied Regression*, 2011:19.

Fu, J., Yang, J., Tan, Y., Jiang, M., Wen, F., Huang, Y., Chen, H., Yi, C., Zheng, S., and Yuan, Y. (2014). "Young patients (? 35years old) with colorectal cancer have worse outcomes due to more advanced disease: a 30-year retrospective review." *Medicine*, 93:1–7.

Ghazali, A. K., Musa, K. I., Naing, N. N., and Mahmood, Z. (2010). "Prognostic factors in patients with colorectal cancer at Hospital Universiti Sains Malaysia." *Asian Journal Of Surgery*, 33:127–133.

Hassan, M. R. A., Ismail, I., Suan, M. A. M., Ahmad, F., Khazim, W. K. W., Othman, Z., Said, R. M., Tan, W. L., Rahmah, S., Mohammed, N. S., et al. (2016). "Incidence and mortality rates of colorectal cancer in Malaysia." *Epidemiology and health*, 38:1–5.

Hosmer, D. W. and Lemeshow, S. (1999). "Assesement of model adequacy." In *Applied Survival Analysis Regression Modeling of Time to Event Data*. JOHN WILEY and SONS,INC.

Kaplan, E. L. and Meier, P. (1958). "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association*, 53:457–481.

Kotake, K., Asano, M., Ozawa, H., Kobayashi, H., and Sugihara, K. (2016). "Gender differences in colorectal cancer survival in Japan." *International journal of clinical oncology*, 21:194–203.

Lydrup, M.-L. and Höglund, P. (2015). "Gender aspects of survival after surgical treatment for rectal cancer." *Colorectal Disease*, 17:390–396.

Majek, O., Gondos, A., Jansen, L., Emrich, K., Holleczek, B., Katalinic, A., Nennecke, A., Eberle, A., Brenner, H., Group, G. C. S. W., et al. (2013). "Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany." *PLoS One*, 8:e68077.

Mandelson, M. T., Miglioretti, D., Newcomb, P. A., Harrison, R., and Potter, J. D. (2003). "Hormone replacement therapy in relation to survival in women diagnosed with colon cancer." *Cancer Causes & Control*, 14:979–984.

Maringe, C., Walters, S., Rachet, B., Butler, J., Fields, T., Finan, P., Maxwell, R., Nedrebø, B., Påhlman, L., Sjövall, A., et al. (2013). "Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000–2007." *Acta Oncologica*, 52:919–932.

Marmot, M. (2015). "The health gap: the challenge of an unequal world." *The Lancet*, 386:2442–2444.

McKay, A., Donaleshen, J., Helewa, R. M., Park, J., Wirtzfeld, D., Hochman, D., Singh, H., and Turner, D. (2014). "Does young age influence the prognosis of colorectal cancer: a population-based analysis." *World journal of surgical oncology*, 12:370.

Mills, K. T., Bellows, C. F., Hoffman, A. E., Kelly, T. N., and Gagliardi, G. (2013). "Diabetes and colorectal cancer prognosis: a meta-analysis." *Diseases of the colon and rectum*, 56:1304.

Omar, Z. A. and Ibrahim Tamin, N. (2011). "National Cancer Registry Report Malaysia Cancer Statistics Data and Figure 2007." *Malaysia: National Cancer Registry*, pages 85–7.

Perme, M. P., Stare, J., and Estève, J. (2012). "On estimation in relative survival." *Biometrics*, 68:113–120.

Phipps, A. I., Baron, J., and Newcomb, P. A. (2011). "Prediagnostic smoking history, alcohol consumption, and colorectal cancer survival: the Seattle Colon Cancer Family Registry." *Cancer*, 117:4948–4957.

Phipps, A. I., Shi, Q., Newcomb, P. A., Nelson, G. D., Sargent, D. J., Alberts, S. R., and Limburg, P. J. (2013). "Associations between cigarette smoking status and colon cancer prognosis among participants in North Central Cancer Treatment Group Phase III Trial N0147." *Journal of Clinical Oncology*, 31:2016–2023.

Price, T. J., Beeke, C., Ullah, S., Padbury, R., Maddern, G., Roder, D., Townsend, A. R., Moore, J., Roy, A., Tomita, Y., et al. (2015). "Does the primary site of colorectal cancer impact outcomes for patients with metastatic disease?" *Cancer*, 121:830–835.

Rasouli, M. A., Moradi, G., Roshani, D., Nikkhoo, B., Ghaderi, E., and Ghaytasi, B. (2017). "Prognostic factors and survival of colorectal cancer in Kurdistan province, Iran: A population-based study (2009–2014)." *Medicine*, 96:1–7.

Roche, L., Danieli, C., Belot, A., Grosclaude, P., Bouvier, A.-M., Velten, M., Iwaz, J., Remontet, L., and Bossard, N. (2013). "Cancer net survival on registry data: use of the new unbiased Pohar-Perme estimator and magnitude of the bias with the classical methods." *International journal of cancer*, 132:2359–2369.

Sant, M., Allemani, C., Santaquilani, M., Knijn, A., Marchesi, F., Capocaccia, R., Group, E. W., et al. (2009). "EUROCARE-4. Survival of cancer patients

diagnosed in 1995–1999. Results and commentary." *European Journal of Cancer*, 45:931–991.

Sharp, L., McDevitt, J., Brown, C., and Comber, H. (2017). "Smoking at diagnosis significantly decreases 5-year cancer-specific survival in a population-based cohort of 18 166 colon cancer patients." *Alimentary pharmacology & therapeutics*, 45:788–800.

Shi, R., Kodali, M. D., Wang, S. C., Lavu, K. C., Liu, L., Shows, J. R., and Mills, G. M. (2013). "Mortality risk factors and survival of colon cancer patient." *Journal of Clinical Oncology*, 31:e14653–e14653.

Stornes, T., Wibe, A., Romundstad, P., and Endreseth, B. (2014). "Outcomes of rectal cancer treatment - influence of age?" *International journal of colorectal disease*, 29:825–834.

Therneau, T. M. (2015). *A Package for Survival Analysis in S.* version 2.38.

Walter, V., Jansen, L., Hoffmeister, M., and Brenner, H. (2014). "Smoking and survival of colorectal cancer patients: systematic review and meta-analysis." *Annals of Oncology*, 25:1517–1525.

Wang, X., Chen, M.-H., Wang, W., and Yan, J. (2017). *dynsurv: Dynamic Models for Survival Data.* R package version 0.3-6.

Widdison, A., Barnett, S. W., and Betambeau, N. (2011). "The impact of age on outcome after surgery for colorectal adenocarcinoma." *The Annals of The Royal College of Surgeons of England*, 93:445–450.

Yeole, B., Sunny, L., Swaminathan, R., Sankaranarayanan, R., and Parkin, D. (2001). "Population-based survival from colorectal cancer in Mumbai,(Bombay) India." *European Journal Of Cancer*, 37:1402–1408.

Chapter 4

Colorectal Cancer in Malaysia: A Spatial Survival Analysis

4.1 Introduction

A patient's survival may depend on several known and unknown factors and it may additionally vary spatially across a region. Socioeconomic status, accessibility to healthcare and other environmental factors are likely to contribute to survival rates.

Numerous studies have been carried out to investigate and model the spatial variation in survival for various types of cancer. For example, in north west England Henderson et al. (2002) model the spatial variation in survival for leukaemia; Fairley et al. (2008) assess the spatial variation in prostate cancer survival in England and Hsieh et al. (2016) evaluate the factors affecting the spatial variation in breast cancer survival in Queensland, Australia. Each of these studies report that spatial variation in cancer survival exists.

Spatial variation in the survival of colorectal cancer specifically has been observed in several studies conducted in developed countries. Henry et al. (2009) observed disparities in survival across New Jersey among more than 25,000 people diagnosed with colorectal cancer between 1996 and 2003. The lowest survival rates were found mostly in economically deprived areas while those in affluent areas had longer survival times; the lack of healthcare accessibility is assumed to be one of the key predictive factors here.

Baade et al. (2013) quantified the geographical variation in survival for colorectal cancer patients using a discrete-time multilevel logistic survival analysis. Their study, involving over 400 regions of Statistical Local Area (SLA) in Queensland, Australia, demonstrated that patients in rural and disadvantaged areas had significantly poorer survival rates than patients in urban and affluent areas, after controlling for individual characteristics and cancer stage.

In Spain, Etxeberria et al. (2014) mapped space-time patterns of colorectal cancer mortality risk by sex and by two age groups - the middle-aged (50 to 69 years) and the elderly ($\geq$70 years) - during the period 1975 to 2008. Their findings demonstrated spatial variation in mortality risk across the region by both sex and age group.

Material deprivation and geographical accessibility to healthcare was found to influence survival in colorectal cancer in a study involving the cases from three cancer registries in France and one cancer registry in England (Dejardin et al., 2014). This study showed that both of the above factors were relevant to patient survival, but that the effect differs between the countries. Material deprivation was significantly associated with cancer survival in England while lack of accessibility to healthcare lead to poorer survival in France. The findings from these studies suggest that it is important to investigate spatial variation in cancer survival in Malaysia.

Some previous studies have examined the spatial distribution of colorectal cancer incidence in a subset of states in Malaysia but none have investigated the spatial variation in survival of this disease (Samat and Shattar, 2014, Samat et al., 2013, Shah et al., 2014). To our knowledge, no studies have examined the epidemiology of this cancer using spatial modelling, and in particular none have extended research to include the whole of Malaysia or the whole Malaysian population.

Our aim in this chapter is to model the spatial variation in survival for colorectal cancer patients in Malaysia accounting for individual and socioeconomic risk factors. We also aim to investigate how individual and socioeconomic factors might affect survival from colorectal cancer, adjusting for spatial location.

Identifying the factors that influence the difference in survival across the region may help the public health authorities better plan healthcare delivery and thus eventually reduce disparities in colorectal cancer survival in Malaysia.

4.2 Methods

4.2.1 Data: Point Level Analysis

For three states in Peninsular Malaysia and one state in East Malaysia, we carried out a spatial survival analysis based on individual location from addresses given in the data. They were Kelantan, Kedah and Kuala Lumpur from Peninsular Malaysia and Sarawak from the East Malaysia. We chose Kelantan and Kedah because these states had a case ascertainment rate of greater than 95% as estimated by our data provider, and Kuala Lumpur because it represents the capital state of Malaysia for which it is estimated that over 90% of cases are recorded.

The population of the three states in Peninsular Malaysia (Kelantan, Kedah and Kuala Lumpur) is about 4.6 million people (census 2010) while Sarawak, which represents East Malaysia has 2,420,009 people in the same census year. Malays represents the majority of the population in the three states in Peninsular Malaysia which are Kelantan , Kedah and Kuala Lumpur with Malays being 94.6%, 77.6% and 45% of the total population in Kedah, Perlis, Kelantan and Kuala Lumpur respectively. Other races in the population include Chinese, Indian and other races. In Sarawak, the state that represents East Malaysia in our study, Malays represents 24.1% of the total population as Sarawak has their own indigenous ethnicity(43.9%) known as Iban, Bidayuh and Melanau which are not part of the population in Peninsular Malaysia (Department of Statistics Malaysia, 2016).

Addresses in Malaysia are assigned to a house, and will often, but not always, contain a house name, a street name and a postcode. Unlike in the UK, a postcode might contain all addresses in a small town. Often addresses are incomplete, with a name and a village name only. In these circumstances we have derived coordinates for the address based on the procedure outlined below.

We created a data variable called 'address' to contain the coordinates for each patients' home address. The point coordinates had been determined from the addresses available in the database, though point coordinates were available for only 87%, 65%, 46% and 12% of the addresses in Kuala Lumpur, Kedah, Sarawak and Kelantan respectively. For addresses where no address coordinate was available we generated an approximate address at as fine a level as was possible. We started with small street (a street of less than one kilometre length) and progressed to housing area to village to 'mukim' (sub-division of the district) as necessary. An average of 30% of the approximate coordinates were at the village level and 10% at the mukim level while the rest are at a finer level (small street and housing area) (Table). We used Google Maps and 1MalaysiaMap (MaCGDI, 2012) to search for the coordinates of the addresses. We ran our point-level analyses separately for each state.

TABLE 4.1: Table of percentage(%) addresses assigns in the spatial analysis model

States	Accurate address(%)	Small street	Village(%)	Small mukim(%)	Total %
Kuala Lumpur	87	8	5	0	100
Kedah	65	10	19	6	100
Sarawak	46	8	38	8	100
Kelantan	12	8	55	25	100

4.2.2 Data: District Level Analysis

The district level analysis includes all districts in Peninsular and East Malaysia and was based on polygonal data. However, the analysis for Peninsular Malaysia and East Malaysia was done separately as they are substantially physically separated. We therefore separated the shapefile of Malaysia into Peninsular Malaysia and East Malaysia.

In order to conduct analysis at the district level, each patient was assigned to their correct district based on their town variable as recorded in the data. Each of the

520 unique names of towns seen in the data belong to one of the 144 administrative districts in Malaysia, of which 87 districts are in Peninsular Malaysia and 57 districts in East Malaysia. Figure 4.1 shows the map of Malaysia by district.

We included a measure of socioeconomic index (Rahman and Zakaria, 2012) and the intensity of hospitals as covariates in our model. The latter is a proxy for access to healthcare facilities, the lack of which may delay patients seeking or obtaining medical attention early on and hence decrease survival.

To model the correlation in space, we used an exponential correlation function, using the distance between the centroids of regions to determine the correlation.

FIGURE 4.1: Malaysia by districts

4.2.2.1 Accessibility to Healthcare

There are 165 hospitals in Peninsular Malaysia which include 90 public hospitals and 75 private hospitals with in-patient care facilities. The hospitals in our data do not include the National Heart Institute, the National Eye Institute or psychiatric hospitals. We obtained the coordinates of all these hospitals from Google Maps. We create a proxy-measure for accessibility of healthcare using the estimated number of hospitals per unit area (the intensity).

Let $\delta(x)$ denote a small region containing the point x. The intensity of hospitals at location x is given by:

$$\lambda(x) = \lim_{|\delta x| \to 0} \left\{ \frac{\mathbb{E}[N(\delta x)]}{|\delta x|} \right\}$$

where

$|\delta x|$ is the area of the small region

$N(\delta x)$ is the number of points falling in δx

$\mathbb{E}[N(\delta x)]$ is the expected number of points falling in δx

The intensity of hospitals in each district was calculated using the `density.ppp` function from the `spatstat` package in `R`. Then, we included this intensity as one of the parameters in the spatial survival model. The intensity of hospitals in Peninsular and East Malaysian districts respectively is presented in Figure 4.2 and Figure 4.3.

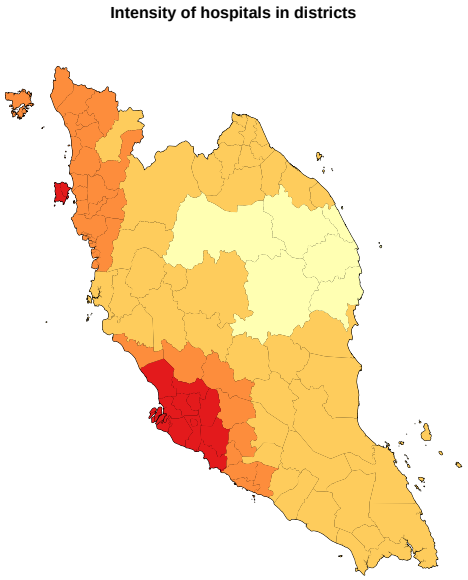


FIGURE 4.2: Intensity of hospitals in Peninsular Malaysia
([0.00,0.50], (0.50,1.50], (1.50,2.50]. (2.50,3.50])

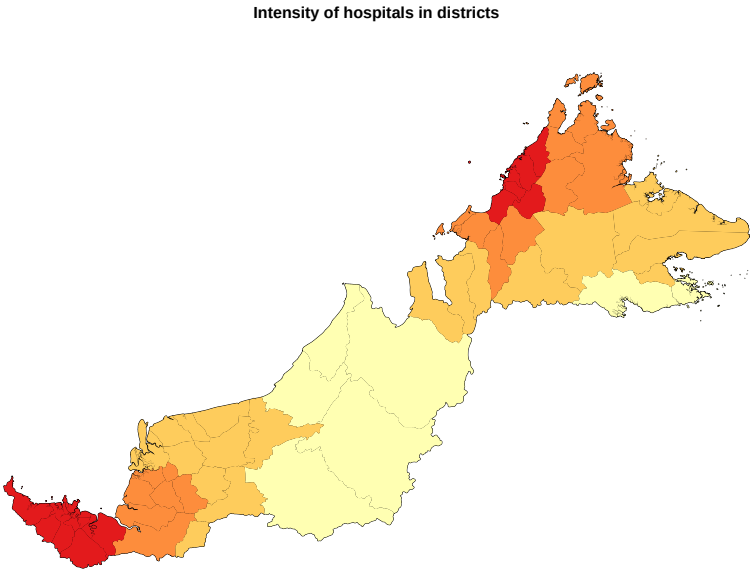


FIGURE 4.3: Intensity of hospitals in East Malaysia
([0.00,0.10], (0.10,0.20], (0.20,0.30]. (0.30,0.40])

4.2.2.2 Socioeconomic Status

We used an index based on information from a census in 2000 as a measure of socioeconomic status in Peninsular Malaysia (Rahman and Zakaria, 2012). This index refers to possession of certain items or attributes as follows:

- Percentage of households who own a car
- Percentage of households who own a motorcycle
- Percentage of households who have air-conditioning
- Percentage of households who own a video/dvd player
- Percentage of households with a tertiary education level (degree)

The index can have a positive or negative value. A more positive index for a particular area indicates that the facilities in that area go beyond basic needs, and vice versa for more negative indices (Rahman and Zakaria, 2012). This index was available for 82 of the 87 districts in Peninsular Malaysia, and we used the `autoKrige` function from the `gstat R` package to impute the value of the index for the remaining five districts (Pebesma and Heuvelink, 2016). The `autoKrige` function implements the technique of ordinary kriging, a method for smoothing spatial data and predicting values for new locations (in this case, the centroids of the districts with missing socioeconomic index). These are presented in Figure 4.4.

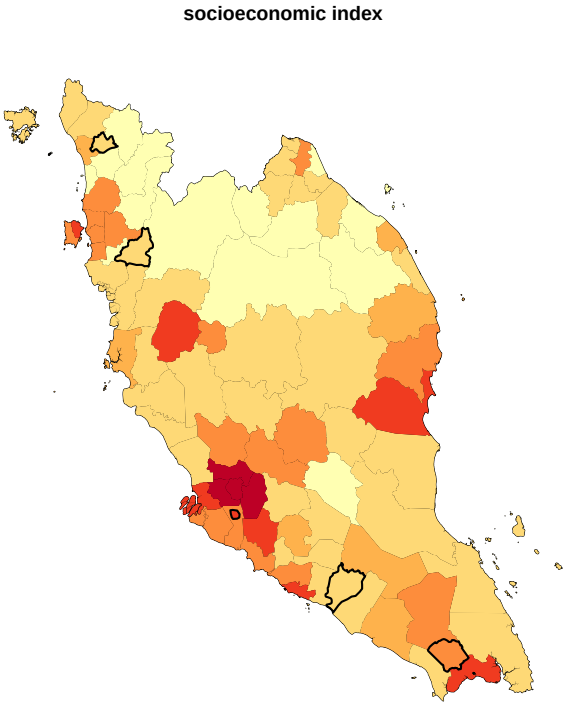


FIGURE 4.4: SES index of districts in Peninsular Malaysia (census 2000) ([-0.30,-0.15], (-0.15,0.00], (0.00,0.05]. (0.05,0.25], (0.25,0.5], (0.50,1.00])

*The districts outlined in bold are those with imputed socioeconomic index value

4.2.3 Statistical Analysis

We create a parametric proportional hazards model using the `spatsurv` package (Taylor and Rowlingson, 2017) to analyse our spatially referenced survival data. The hazard function takes the following form:

$$h(t_i;\psi,Y_i) = h_0(t_i;\omega)\exp\{X_i\beta + Y_i\},$$

where

h_0 is the baseline hazard function

t is the observed time for the i^{th} individual

X_i is a vector of covariate values for the i^{th} individual

$\psi = (\beta, \omega, \eta)$ are covariate effects, the parameter of the baseline hazard and the parameter of the covariance function of a spatially latent Gaussian field Y, respectively

Y_i is the value of the field at the location of individual i

Our original data contained the addresses of the patients. We conducted two different analysis on the data; point level analysis and district level analysis as described previously.

We used MCMC to estimate the parameters in our model. We looked for evidence of satisfactory convergence and mixing in the MCMC chain by considering the `mcmcplot` of β, ω, η and Y. We compared plots of prior and posterior to check that our data were sufficient to allow identifiability of the parameters in our model.

The model has been fitted using three different distributions for the baseline hazards: Weibull, Exponential and B-spline. The models were compared using the *Watanabe-Akaike information criterion* (`waic`) value.

We plot the posterior baseline and cumulative hazard for each model as well as the spatial covariance function and correlation against distance. We also mapped the probability that the covariate-adjusted relative risk exceeds certain thresholds. These plots represent the risk over space that is not accounted for by the covariates in our model.

4.3 Results

Table 4.2 below shows the number of colorectal cancer cases and deaths in four chosen states for the spatial survival model in this study.

TABLE 4.2: Colorectal cancer cases in four states chosen for spatial survival model

States	Cases(n)	Death(n(%))	Alive(n(%))
Kedah	556	225(45.9)	301(54.1)
Kelantan	268	152(56.7)	116(43.3)
Kuala Lumpur	290	151(52.1)	139(47.9)
Sarawak(East Malaysia)	583	244(41.9)	339(58.1)
Total	1697	772(45.5)	925(54.5)

4.3.1 Point Level Analysis

4.3.1.1 MCMC Convergence from Trace Plots

We began by checking MCMC convergence by considering the trace plots for each model for the four chosen states in the point analysis. The plots shown in Appendix 4.4 show that mixing is good and there is evidence for satisfactory convergence. We next look at autocorrelation in the latent field, and then check the information content in the data by plotting the priors against the posterior for each parameter. The idea of these plots is that the data should move the prior somewhat; if the prior and posterior are too similar, then there is little information in the data on this particular parameter.

4.3.1.2 Model Comparison

We compared three possible models for the baseline hazard using WAIC; Table 4.3 shows these values. The model with the smallest WAIC value is considered to have the best fit. In all four of the chosen states, the model using a Weibull

distribution to describe the baseline hazard function yielded the smallest WAIC value.

State	WAIC		
	Weibull	Exponential	B-spline
Kelantan	**2428.152**	2434.033	2430.226
Kedah	**4125.213**	4145.084	4130.128
Kuala Lumpur	**2472.562**	2475.856	2473.387
Sarawak	**4149.445**	4153.215	4149.778

TABLE 4.3: Table comparing the WAIC values of models with Weibull, Exponential and B-spline distributed baseline hazard functions, by state

Table 4.4 and Table 4.5 summarise the parameter effects described by the spatial survival model for the four states chosen for the point level analysis. Age slightly increased the risk of death for colorectal cancer and the association was significantly found in Kedah and Sarawak states. Sex was not significantly associated with the risk of death from colorectal cancer in any of the four states chosen in this study. Being of Malay race was found to increase the risk of death from colorectal cancer. We found that patients that had higher education level had a decrease in the risk of death from colorectal cancer, in particular those with tertiary education (degree) had the lowest risk of death from colorectal cancer. However, in Kedah, we noted that those with secondary level education had more than twice the risk of death than patients with no formal education.

Regarding smoking status, our evidence was mixed regarding the deviation of the effect. In Kelantan, being a former smoker, active smoker or having missing smoking status resulted in lower risk of death from colorectal cancer than the non smoking group, but the effect was not significant. In contrast, in Kuala Lumpur and Sarawak, former smokers and active smokers had more risk of death from colorectal cancer compared to those who do not smoke. However, it was found to

be significant in the smoker group for Sarawak only. Former smoker group and missing smoking status positively affect the risk of death in Kedah but the effect was not significant.

Colorectal cancer patients diagnosed with Stage IV cancer had the highest risk of death from colorectal cancer in all states except Kuala Lumpur, where those at Stage III at diagnosis had higher risk of death than Stage IV, but the effect was not significant. The group with missing stage also had high risk of death in all states but the effect was only significant in the state of Kedah.

The site of cancer did not significantly affect survival of colorectal cancer in any of the states in our study. Tumours graded moderately or poorly differentiated, or without a grading, were all at increased risk of death from colorectal cancer compared to those graded as well differentiated. The presence of distant metastases also had positive association with colorectal cancer survival even after controlling for spatial correlation. However, the effect was only significant in Kuala Lumpur and Kedah. Patients that underwent 'other' treatment had the lowest survival of all treatment groups after accounting for correlation in space, but this was found to be significant in Kuala Lumpur only. Diabetes mellitus was not significantly associated with survival in colorectal cancer once adjusted for correlation on space.

Overall, when the spatial location is taken into account, cancer staging and tumour differentiation play an important role in affecting survival in colorectal cancer where the risk of death increased with higher cancer staging and poorer differentiation of the tumour. On the other hand, high education level (tertiary) significantly reduced the risk of death from colorectal cancer in these four selected states. Other factors found to be significantly affecting the risk of death in certain states, after controlling for spatial location, were the patient's age, race, the presence of distance metastases and treatment modalities received by the patients. Patients' sex,

smoking status and diabetes status did not significantly affect the hazard of death in colorectal cancer patients in all selected states in the model.

TABLE 4.4: Table of parameter estimates from the fitted model for four states in Malaysia. Significant estimates (those whose credible interval does not contain 1) are given in bold font for ease of identification.

	Kelantan(N=268)	*KLumpur(N=290)*	*Kedah(N=556)*	*Sarawak(N=588)*
Variable	**Med (95% CRI)**	**Med (95% CRI)**	**Med (95% CRI)**	**Med (95% CRI)**
Age	0.99 (0.98, 1.01)	1.01 (0.98, 1.03)	**1.02 (1.01, 1.03)**	**1.01 (1.00, 1.03)**
Sex				
Female				
Male	1.14 (0.71, 1.85)	0.94 (0.53, 1.63)	1.17 (0.86, 1.61)	0.81 (0.61, 1.08)
Race				
Non Malay				
Malay	1.17 (0.61, 2.29)	1.31 (0.75, 2.36)	**1.47 (1.08, 2.06)**	**1.47 (1.02, 2.07)**
Education				
Nil				
Primary	**0.38 (0.16, 0.94)**	**0.37 (0.14, 0.93)**	1.36 (0.75, 2.45)	1.17 (0.69, 1.97)
Secondary	**0.18 (0.08, 0.43)**	**0.33 (0.11, 0.92)**	**2.38 (1.26, 4.63)**	0.94 (0.54, 1.59)
Tertiary	**0.13 (0.05, 0.36)**	**0.06 (0.01, 0.45)**	0.52 (0.18, 1.37)	**0.43 (0.17, 0.98)**
Missing	0.46 (0.23, 1.01)	0.76 (0.36, 1.52)	**2.65 (1.49, 4.73)**	0.92 (0.61, 1.44)
Smoking status				
Non Smoker				
Former Smoker	0.90 (0.48, 1.77)	1.25 (0.58, 2.96)	1.65 (0.96, 2.68)	**1.88 (1.15, 2.92)**
Active smoker	0.72 (0.30, 1.55)	1.89 (0.82, 4.49)	0.94 (0.60, 1.49)	1.07 (0.68,1.76)
Missing	0.77 (0.44, 1.36)	0.88 (0.47, 1.62)	1.06 (0.71, 1.57)	1.16 (0.81, 1.63)
Staging				
Stage I				
Stage II		0.47 (0.12, 2.09)	2.79 (0.97, 10.7)	0.85 (0.35, 2.35)
Stage III	**3.30 (1.35, 8.69)**	1.50(0.41, 6.19)	**3.20 (1.15, 11.8)**	1.34 (0.60, 3.65)
Stage IV	**6.59 (2.58, 19.14)**	1.31 (0.36, 5.51)	**7.11 (2.62, 27.1)**	**3.19 (1.38, 8.48)**
Not staged	1.22 (0.39, 4.05)	0.81 (0.16, 4.19)	2.62 (0.87, 9.79)	1.68 (0.71, 4.85)
Missing	6.53 (2.69, 17.66)	1.31 (0.39, 5.02)	**4.29 (1.56, 15.5)**	1.85 (0.83, 4.91)
Cancer Site				
Colon				
Rectum	0.88 (0.51, 1.45)	1.27 (0.57, 2.64)	0.85 (0.55, 1.38)	1.14 (0.73, 1.7)
Rectosigmoid	1.10 (0.67, 1.73)	0.82 (0.45, 1.39)	1.25 (0.95, 1.69)	1.20 (0.89, 1.63)

TABLE 4.5: Table of parameter estimates from the fitted model for four states in Malaysia. Significant estimates (those whose credible interval does not contain 1) are given in bold font for ease of identification.

	Kelantan(N=268)	*KLumpur(N=290)*	*Kedah(N=556)*	*Sarawak(N=588)*
Variable	**Med (95% CRI)**	**Med (95% CRI)**	**Med (95% CRI)**	**Med(95% CRI)**
Tumor grade				
Moderate	**2.18 (1.02, 5.88)**	**2.74 (1.08, 8.00)**	1.48 (0.83, 2.77)	**1.98 (1.11, 3.64)**
Poor	2.24 (0.54, 8.88)	**7.38 (1.49, 48.0)**	1.94 (0.69, 4.92)	**4.92 (2.03, 12.4)**
Missing	**3.10 (1.31, 8.26)**	**3.38 (1.31, 11.7)**	**1.88 (1.06, 3.70)**	**2.82 (1.48, 5.42)**
Metastasis				
No				
Yes	1.08 (0.64, 1.82)	**2.50 (1.30, 5.02)**	**1.58 (1.08, 2.29)**	1.44(0.98, 2.16)
Missing	0.86 (0.38, 2.12)	2.12 (0.85, 5.36)	1.41 (0.83, 2.47)	1.36 (0.85, 2.28)
Treatment				
Surgery Alone				
Surg,Chemo,radio	**0.59 (0.37, 0.95)**	1.15 (0.63, 2.21)	0.71 (0.481, 1.03)	0.90 (0.64, 1.26)
Radio,Chemo	1.50 (0.57, 3.42)	2.52 (0.99, 6.33)	0.62 (0.29, 1.27)	1.06 (0.67, 1.69)
Other	1.97 (0.496, 3.69)	**4.75 (1.54, 15.02)**	1.83(0.95, 3.63)	1.37 (0.19, 5.12)
Unknown	1.37 (0.62, 2.84)	1.76 (0.71, 4.52)	0.89 (0.56, 1.40)	**1.75 (1.01, 2.94)**
Diabetes				
No				
Diabetes	1.46 (0.87, 2.31)	1.21 (0.66, 2.24)	0.84 (0.60, 1.19)	1.01 (0.68, 1.48)
Missing	0.57 (0.30, 1.01)	0.91 (0.42, 1.95)	0.91 (0.54, 1.56)	0.98 (0.35, 2.22)
σ	0.55 (0.33, 0.88)	1.32 (0.92, 1.88)	0.54 (0.311, 0.83)	0.38 (0.20, 0.65)
ϕ	5010 (2285, 11241)	776 (495, 1309)	3839 (1617, 8458)	16384 (5828, 44570)

4.3.2 Maps for Probability of Exceedance Risk

We produced maps of the probability that hazard of death exceeds 1.1 and 1.25 for areas within the four states. The choice of the hazard values to exceed was arbitrary and solely for illustrative purposes.

The risk maps for the three states in Peninsular Malaysia are shown in Figure 4.5. Within Kelantan and Kedah, there were more regions displayed as at highest risk of their hazard exceeding 1.1 than there were when the threshold was increased to 1.25. However, the threshold increase did not cause much difference in the number of areas at highest risk within the Kuala Lumpur region, though we note that very few areas in Kuala Lumpur, the capital of Malaysia, had a probability higher than 0.25 of having hazard of death exceeding either threshold.

Figure 4.6 shows exceedance risk maps for Sarawak, the chosen state in East Malaysia. We see that very few areas of Sarawak are likely to have a hazard of death exceeding 1.25.

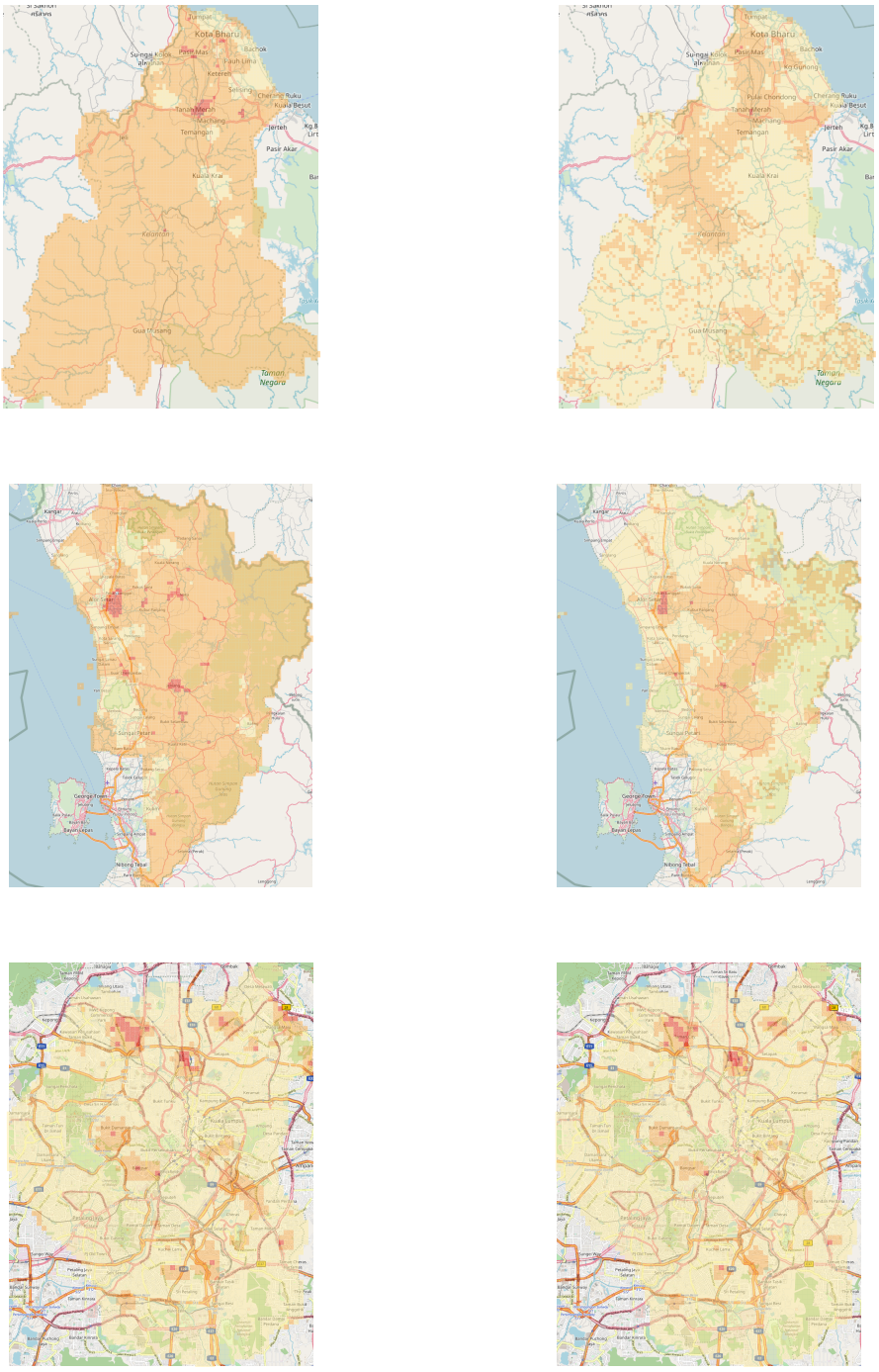


FIGURE 4.5: The leaflet plot of $\mathbb{P}[\exp(Y) > 1.1]$ (left) and $\mathbb{P}[\exp(Y) > 1.25]$ (right) for Kelantan (top), Kedah (middle row) and Kuala Lumpur (bottom) (□ [0.0,0.25], ■ (0.25,0.50], ■ (0.50,0.75])

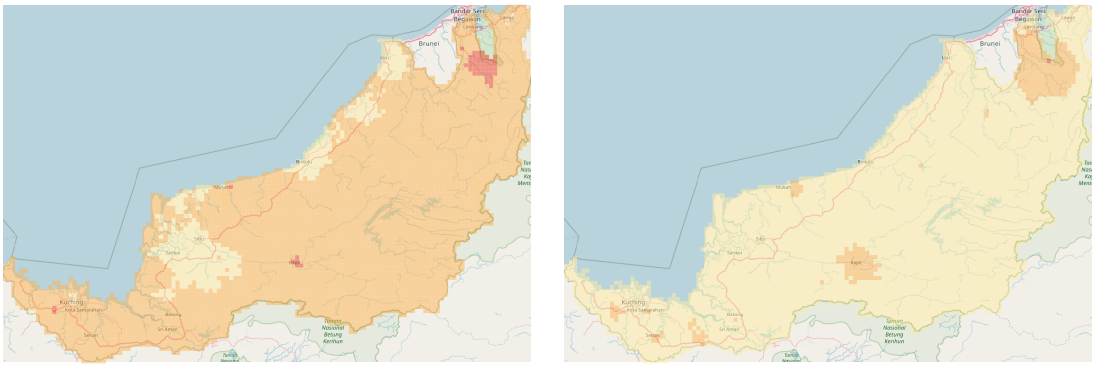


FIGURE 4.6: The leaflet plot of $\mathbb{P}[\exp(Y) > 1.1]$ (left) and $\mathbb{P}[\exp(Y) > 1.25]$ (right) for Sarawak, East Malaysia
([0.0,0.25], (0.25,0.50], (0.50,0.75])

4.3.3 Baseline Hazard and Spatial Correlation

Next, we provide plots of the baseline hazard function and the posterior spatial correlation function across space for each region, as shown in Figures 4.7, 4.8, 4.9 and 4.10. The baseline hazard function shows the shape of the hazard when the covariates are all equal to zero. The shape of the baseline hazard in the plot below describes a very slightly decreasing trend. The decrease in hazard was so slight that we might consider an exponential model instead, however the WAIC was lower for the Weibull model which indicates that this model is a better fit for the data.

The spatial correlation function shows how similar the hazard is across space, and how fast that similarity decays. A correlation plot with a fast drop (small ϕ) shows that there is only little dependence in hazard with distance. On the other hand, if the correlation plot has a slow drop, this shows that there is strong spatial dependance in the hazard. It means that even though there may be a large distance between places, the correlation in their hazard is high. The interpretation of ϕ is that for distances over ϕ apart, there is little dependence in space.

Kuala Lumpur had the smallest ϕ value and Sarawak the biggest ϕ value, as presented in Table 4.5. The plot shows that the posterior covariance function for Kuala Lumpur is fast to decay, indicating that spatial correlation exists within small distances while in Sarawak the correlation in space can extend to areas a long way away, even to 20km apart.

Kelantan

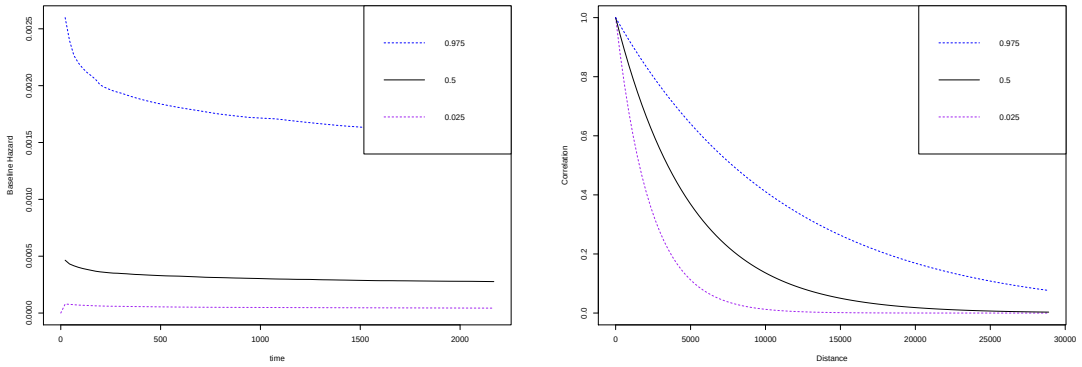


FIGURE 4.7: Plots of the baseline hazard function (left) and posterior spatial covariance function for Kelantan

Kedah

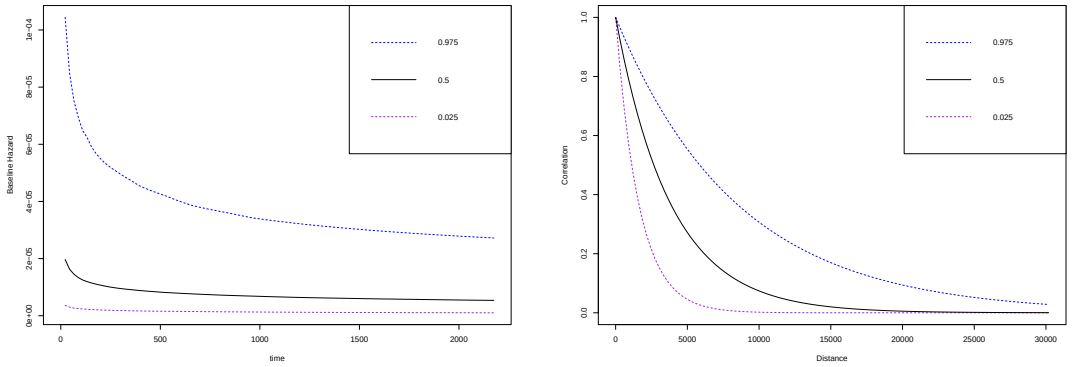


FIGURE 4.8: Plots of the baseline hazard function (left) and posterior spatial covariance function for Kedah

Kuala Lumpur

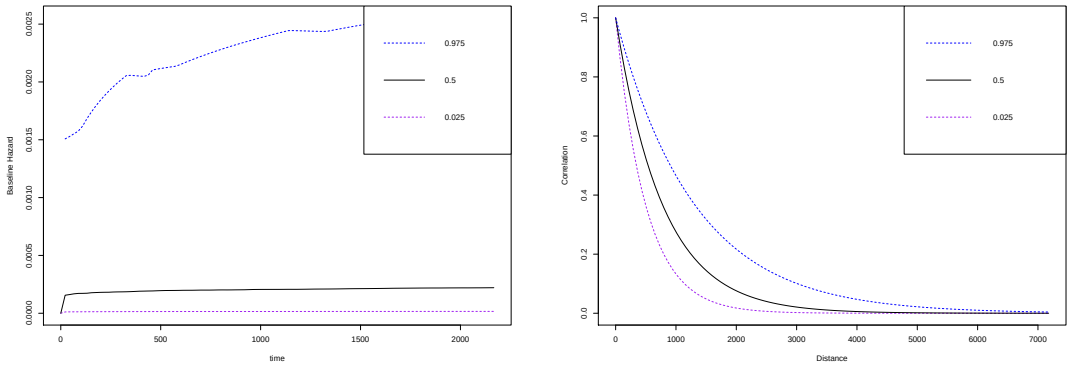


FIGURE 4.9: Plots of the baseline hazard function (left) and posterior spatial covariance function for Kuala Lumpur

Sarawak

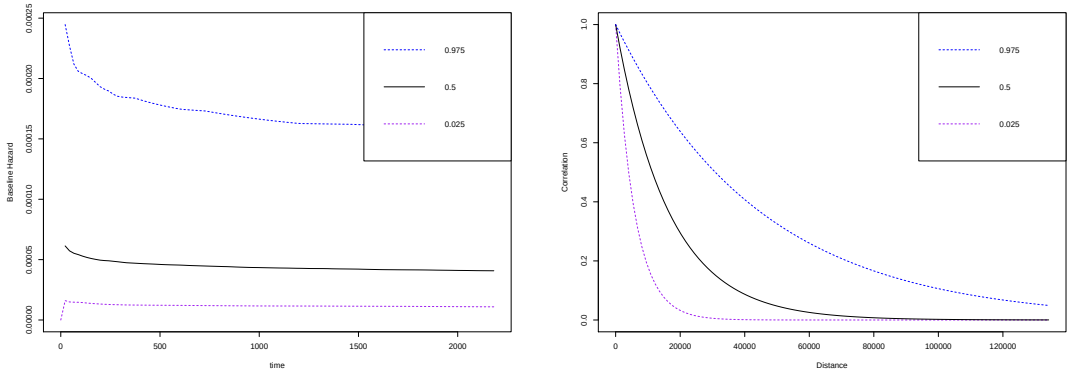


FIGURE 4.10: Plots of the baseline hazard function (left) and posterior spatial covariance function for Sarawak

4.3.4 District Level Analysis

Analysis at the district level involved the creation of two spatial survival models, one for Peninsular (West) Malaysia and one for East Malaysia. Table 4.6 shows the relative risk of covariates used in these models. After controlling for spatial location and socioeconomic factors, cancer staging still plays an important role in determining the risk of death from colorectal cancer in Malaysia. Patients

diagnosed at Stage IV had six times (median 6.37, 95% CRI (4.34, 9.75)) and seven times (median 7.33, 95% CRI (2.99, 24.00)) higher risk of death from colorectal cancer in West and East Malaysia respectively than patients diagnosed at Stage I. Each year increase in age led to a slight increase (median 1.01, 95% CRI (1.01, 1.02)) in relative risk of death. Other factors that significantly affect survival in colorectal cancer patients in Malaysia were race, tumour differentiation and the presence of distant metastases.

We included two parameters to represent the socioeconomic distinctions in the population; the intensity of the hospitals in each district and the middle class household item index (socioeconomic index). However the socioeconomic index was only available for Peninsular Malaysia. Both of these variables showed a decrease in the risk of death as the value of the parameters increases, however neither result was significant.

TABLE 4.6: Table of parameter estimates from the fitted model. Significant estimates (those whose credible interval does not contain 1) are given in bold font for ease of identification.

Covariates	Pen. Malaysia Med (95% CRI)%	East Malaysia Med (95% CRI)%
Age	**1.01 (1.01, 1.02)**	**1.01 (1.01, 1.02)**
Sex		
Female		
Male	1.05 (0.93, 1.19)	0.96 (0.76, 1.24)
Missing	0.81 (0.40, 1.42)	0.60 (0.02, 5.29)
Race		
Non Malay		
Malay	**1.26 (1.13, 1.43)**	**1.47 (1.05, 2.01)**
Missing	1.24 (0.20, 4.58)	1.53 (0.27, 5.87)
Education		
Nil		
Primary	0.92 (0.72, 1.17)	1.00 (0.68, 1.46)
Secondary	1.00 (0.78, 1.27)	1.01 (0.68, 1.47)
Tertiary	**0.63 (0.43, 0.88)**	0.60 (0.32, 1.10)
Missing	1.03 (0.85, 1.28,)	0.82 (0.61, 1.18)
Smoking		
Non Smoker		
Former Smoker	1.17 (0.97, 1.40)	1.45 (0.99, 2.00)
Active smoker	1.13 (0.93, 1.37)	1.12 (0.76, 1.63)
Missing	1.03 (0.90, 1.19)	1.09 (0.80, 1.50)
Staging		
Stage I		
Stage II	**1.65 (1.09, 2.57)**	2.04 (0.81 ; 6.85)
Stage III	**3.17 (2.13, 4.83)**	**3.63 (1.43,10.90)**
Stage IV	**6.37 (4., 9.75)**	**7.33 (2.99,24.00)**
Not staged	**2.79 (1.88, 4.46)**	**5.06 (2.03, 16.00)**
Missing	**4.14 (2.82, 6.45)**	**3.91 (1.56, 12.70)**
ICDSite		
Colon		
Rectum	1.05 (0.90 ; 1.23)	1.32 (0.95 ; 1.84)
Rectosigmoid	1.06 (0.94 ; 1.19)	1.16 (0.89 ; 1.49)
Missing	0.29 (0.01 ; 1.52)	
TumourDiff		
Well		
Moderate	1.17 (0.94, 1.46)	**1.77 (1.08, 3.01)**
Poor	**2.45 (1.81, 3.40)**	**2.99 (1.46, 5.88)**
Missing	**1.59 (1.28, 1.99)**	**2.84 (1.73, 4.97)**
Metastasis		
No		
Yes	**1.40 (1.21, 1.62)**	**1.59 (1.19 ; 2.13)**
Missing	**1.47 (1.21 ; 1.77)**	1.39 (0.93, 2.11)
Treatment		
Surgery Alone		
Surg,Chemo,radio	**0.87 (0.76, 0.99)**	0.98 (0.75, 1.27)
Radio,Chemo	1.11 (0.91, 1.35)	1.22 (0.83, 1.81)
Other	**1.84 (1.41, 2.38)**	2.36 (0.99, 4.96)
Unknown	1.13 (0.95, 1.33)	1.56 (0.99, 2.35)
Diabetes		
No		
Yes	1.12 (0.99, 1.28)	1.08 (0.77, 1.43)
Missing	0.88 (0.72, 1.06)	0.73 (0.28, 1.56)
Hosp. intensity	0.97 (0.81, 1.143)	0.69 (0.03, 13.70)
SE Index	0.85 (0.53, 1.39)	
σ	0.38 (0.20, 0.65)	0.51 (0.34, 0.74)
ϕ	16384 (5828, 44570)	44602 (24370, 78650)

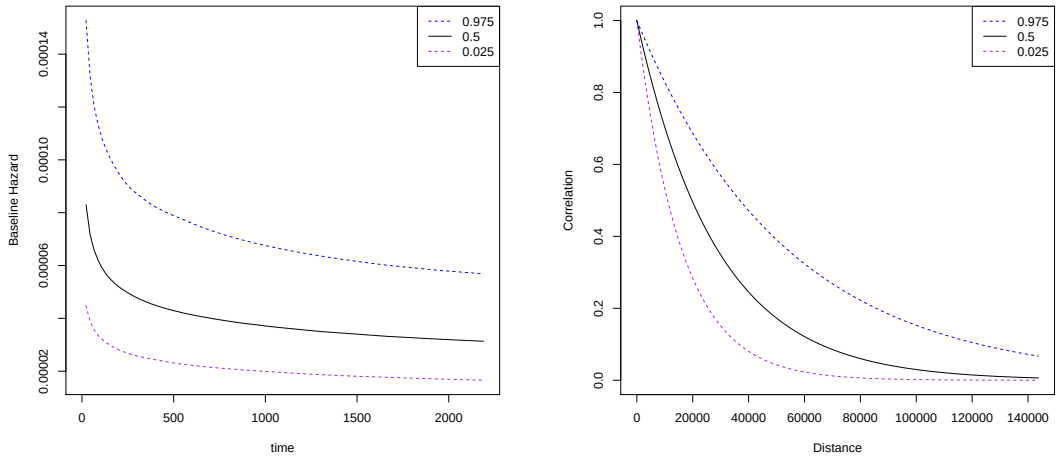


FIGURE 4.11: Plots of the baseline hazard function (left) and posterior spatial correlation function for Peninsular Malaysia. Plot shows that the correlation of hazard starts to decrease when the distance of cases was reaching 20km apart; supported by ϕ value in Table 4.6

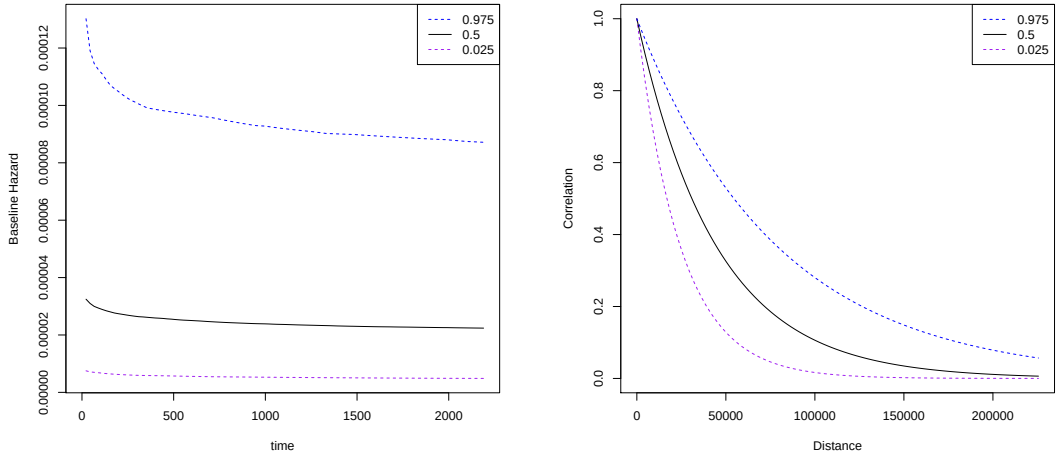


FIGURE 4.12: Plots of the baseline hazard function (left) and posterior spatial correlation function for East Malaysia. Plot shows that the correlation of hazard starts to decrease when the distance of cases was reaching 45km apart; supported by ϕ value in Table 4.6

4.3.5 Maps for Probability of Exceedance Risk - District Level

Figure 4.13 shows the risk map for the probability of exceedance risk of hazard of 1.1 and 1.25 in Peninsular Malaysia. These plots show $\mathbb{P}[\exp(Y) > 1.1]$ and $\mathbb{P}[\exp(Y) > 1.25]$. Three regions in Peninsular Malaysia were identified as having a higher probability that the hazard of death would exceed 1.1, that is, those with probability greater than 0.75 of exceeding the stated hazard. The probability starts to lessen when we increase the exceedance threshold to 1.25.

Figure 4.14 shows the risk map for the probability of exceedance risk of hazard of 1.25 and 1.5 in East Malaysia. East Malaysia has two states, Sabah and Sarawak. The areas of highest probability of exceeding the stated hazard of death in East Malaysia were all located in Sarawak. There was one district in Sarawak, Limbang, that was highly likely to have a hazard of death exceeding 1.25. Upon increasing the threshold to 1.5, only one part of the Sarawak region continues to have probability of exceedance between 0.50 to 0.75, and none with probability greater than 0.75.

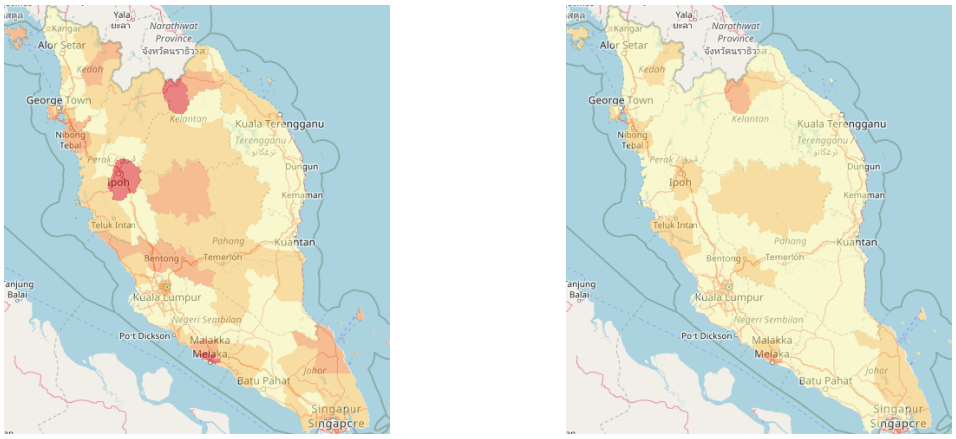


FIGURE 4.13: The leaflet plot of $\mathbb{P}[\exp(Y) > 1.1]$ (left) and $\mathbb{P}[\exp(Y) > 1.25]$, Peninsular Malaysia
([0.00,0.20], (0.20,0.40], (0.40,0.60]. (0.60,0.80])

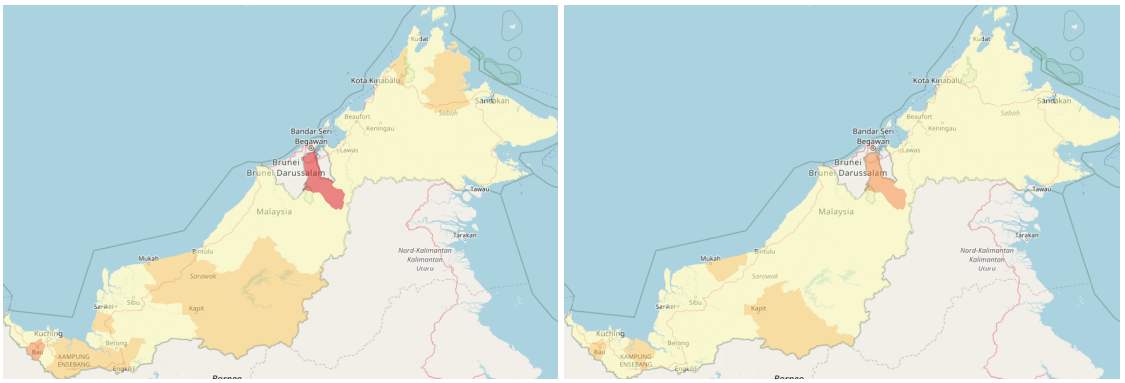


FIGURE 4.14: The leaflet plot of $\mathbb{P}[\exp(Y) > 1.25]$ (left) and $\mathbb{P}[\exp(Y) > 1.5]$, East Malaysia
([0.00,0.25], (0.25,0.50], (0.50,0.75]. (0.75,1.00])

4.4 Discussion

In this study, we extend the survival model from our previous objective to look at the geographical variation of colorectal cancer survival in Malaysia. Our findings show that there is a spatial variation in survival prognoses, or the hazard of death, for colorectal cancer in Malaysia even having adjusting for individual level and area level covariates.

Overall, from the point location analysis, we noted that the level of disease severity as measured by cancer staging, tumour grading and the presence of distant metastases, remain as important predictors of survival times for patients with colorectal cancer in Malaysia even after controlling for spatial correlation. Survival is also better in patients with higher levels of education.

Our results for the district level analysis for both Peninsular and East Malaysia were not much different from the point level analysis. Cancer staging, tumour differentiation and the presence of distant metastases continue to have a significant effect on survival for colorectal cancer patients. We also see here, though, that Malays had a significant 26% higher risk of dying from colorectal cancer than non-Malays. An increase in age slightly increased the risk of death from colorectal cancer.

We found that high socioeconomic index in an area did not significantly affect the risk of death from colorectal cancer. Our findings were similar to those of Barclay et al. (2015) who found that the socioeconomic status of the population did not significantly influence outcomes in patients with colorectal cancer. In comparison, a systematic review by Manser and Bauerfeind (2014) found that socioeconomic status had a significant impact on survival of colorectal cancer, where the risk of death was greater among patients with low socioeconomic status. Regarding socioeconomic status, recall that our measure, based on the 2000 census, was

aggregated to district level. Hence a potential explanation for not observing a significant effect may be due to the presence of ecological bias: that our measure of socioeconomic status did not pertain to individuals. In fact, considering that education is often directly related to socioeconomic status, our results do therefore show evidence that higher socioeconomic status is protective. However, further research in this area is required, and a more finely-resolved spatial map of deprivation could help us to better identify this effect.

It is possible that there is collinearity between our variable 'education', an individual level variable, and our socioeconomic index, which is an area level factor measure one of whose domains is education. We think that this two measures are not likely to be well correlated for the following reasons. Socioeconomic index includes a measure, at an area level, of the proportion of population with tertiary education in a district (area level). Education is one of five domains of which the socioeconomic measure is comprised. We realized the education might driven to the socioeconomic status, but in is not the sole driver of socioeconomic. The widely used UK Index of Multiple Deprivation (IMD) is comprised of 7 domains, one of which is based on education, and the index measures to the small areas across England, called Lower-Layer Super Output Areas (LSOAs) (Ministry of Housing, Communities Local Government UK, 2015).

The education variable in the model represents the individual education level of the patients in this study, classified into five categories, nil, primary, secondary, tertiary and missing status. Thus we think that the education and the SE index variable in the model represent different things. To see if this is likely, we checked if there was any correlation between the two variables but it was not significant with the correlation coefficient, ρ=0.01.

We noticed that there were patients that had changed their address to another place after diagnosis, we decided not to choose the distance of the patients address

to the hospital as our measure of accessibility to healthcare. Address at diagnosis is important as an 'exposure' or proxy for umeasured exposures, but we had no record of length of residence at the address at diagnosis, leaving open the likelihood of misclassifying cases by exposure to place. For example people move house or job for many reasons, but sometimes for health reasons. They may for example move nearer a hospital when ill, or away from an exposure when concerned and there is evidence that population movement can lead to misclassification in epidemiological studies, such as when birth address is used in studies of birth defects (Canfield et al., 2006). Instead, we decided to look at how the density of the hospital in the area affect the patients' survival in our study.

The relationship between survival and distance to treatment hospital is not clear cut: it is not necessarily the closest hospital to patients that they will choose to go to seeking treatment and for this reason, we use the smoothed intensity (number of hospitals per unit area) as a covariate in our model. We expect the coefficient of this covariate to be positive, since, in places where there is a greater concentration of hospitals, patients tend to get diagnosed quicker by being able to recognise and act on early symptoms. We found that greater accessibility to healthcare decreased the risk of death from colorectal cancer but the effect was not significant in our study, which contradicts a previous study reporting that lack of access to healthcare was significantly associated with being diagnosed at a more advanced stage of colorectal cancer (Wan et al., 2013), which we know adversely affects patient survival. Different ways of assessing accessibility to healthcare may influence the direction and significance of the effect. For example, the previous study measured the shortest time taken to travel to the nearest appropriate health facility, while our study measured the number of hospitals per unit area. Furthermore, other factors that could be mediators to the accessibility to healthcare, such as the transportation system, affordability of care and cultural barriers, were not considered in our study.

Regarding the exceedance probabilites, some areas of Kuala Lumpur remain more likely than others to exceed a relative risk of death of 1.1 and 1.25. This might warrant further investigation as to why certain places in the capital city had lower survival rates than others. However, note that our point level analysis only controlled for individual risk factors. Therefore, the environment or area level factors such as socioeconomic status (Lejeune et al., 2010) and accessibility to healthcare (Wan et al., 2013) might also affect spatial variation in survival.

Three areas were found to have high risk of death (where the HR $>$ 1.1 relative to other areas in Peninsular Malaysia). They were Jeli district in the North-East, Kinta district in the North-West and Melaka Tengah located in the West of Peninsular Malaysia. In East Malaysia, Sabah had better survival compared to Sarawak as shown by the probability of exceedance risk maps. The reason for these differences could be such things as delay in chemotherapy treatment (Xu et al., 2014) and comorbidities (Søgaard et al., 2013) which were not assessed in our study, and this would make an interesting direction for further work in this area.

There are some limitations to the work of this study. One of the limitation of this study is that we could not locate exactly the address for all cases study, the result of incomplete documentation of full address in the registry database. For records where there was an incomplete addresss recorded researcher had to locate the addresses manually as been explained in the methodology section and try and make the best effort possible to find the most accurate approximated address possible. The accuracy of the patients' location varies across the states and this may possibly affect our analysis in a state with very low accuracy, which in this case is Kelantan. Uncertainty in the spatial location affects our ability to estimate the spatial correlation parameter in our model, which is typically not well identified in spatial analyses in any case. It is possible to account for the

uncertainty in location in a formal manner - see the conclusions chapter, but this is beyond the scope of this thesis.

Malaysia does not have a formal socioeconomic status instrument consistent across the population (for example, the UK have an index of deprivation). Currently the Malaysian Statistics Department only produce data on income at the district level; data on income at finer levels does not exist. We hope that in future research, it will be possible to consider household income and employment status as additional indications of socioeconomic status.

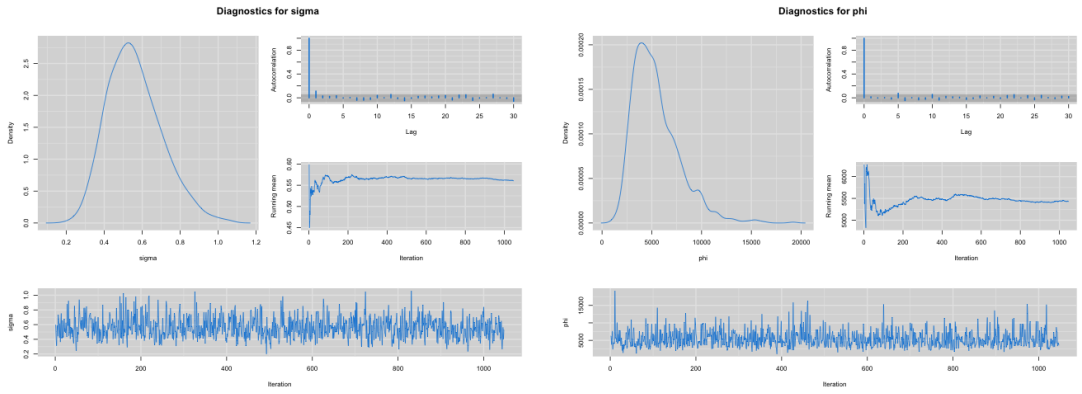
There are several strengths of spatial analysis method used here. First it allows us to combine data from the individual level with data from aggregated levels and map our findings. It also allows us to estimate and map the effect of unobserved environmental confounders through the use of a spatially correlated random effect term. We applied MCMC for our analysis as it delivers full joint inference for all model parameters. Some limitations to be considered is that this analysis assumes a particular model form and correlation structure for the spatial variation and that it can be difficult in practice to identify the parameters of the spatial process. Attaining good convergence and mixing of MCMC can be difficult without access to bespoke software ? but again this is not a issue for us.

Despite these limitations, this is the first study examining the variation in survival for colorectal cancer in Malaysia. There is variation in the survival or risk of death from colorectal cancer in the population, having control for the potential individual and area level factors. Areas close to town centers including the urban areas had lower risk of death compared to other areas in all four states chosen to be in our point analysis model. In the state level analysis, out of 144 districts in the whole of Malaysia, we identified 4 districts, three in Peninsular Malaysia and one in East Malaysia which had higher risk of death compared to the others. Our findings are a new input for the ministry of health to set a target for the population with

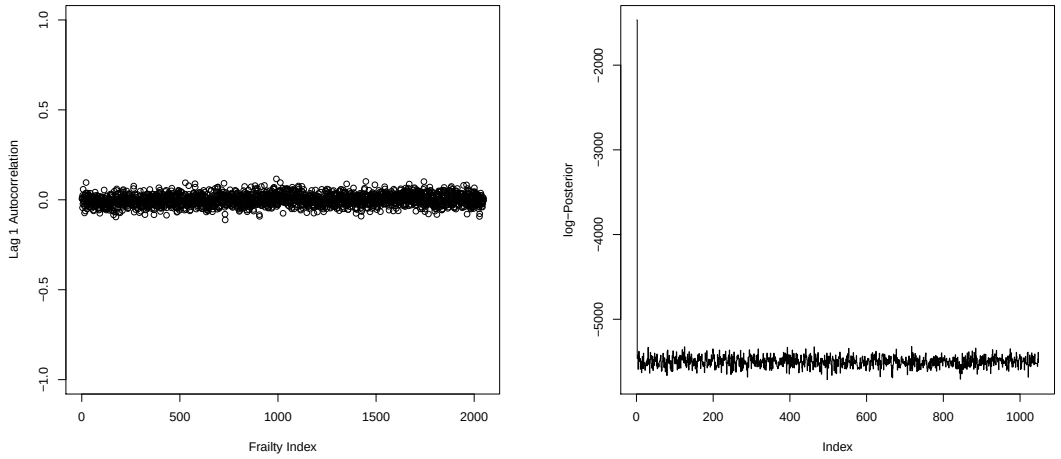
high risk of having poorer survival in providing cancer control services as well as enhance health activities that are cancer-reated in order to improve survival in the population.

Appendix I: Diagnostic Test

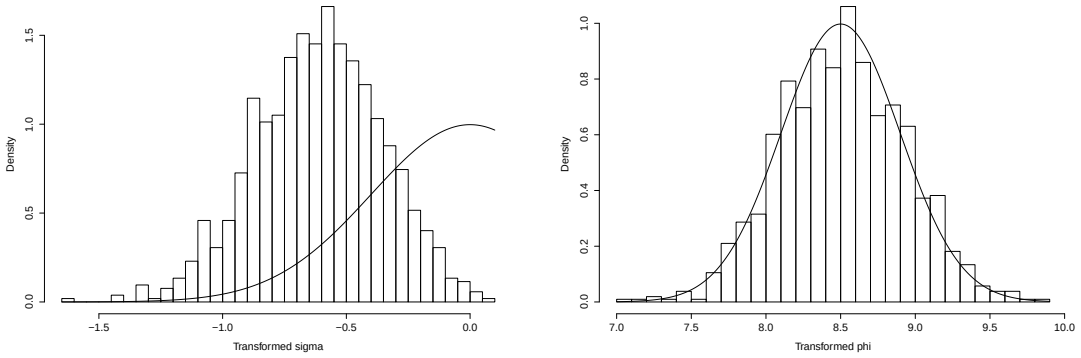
KELANTAN:



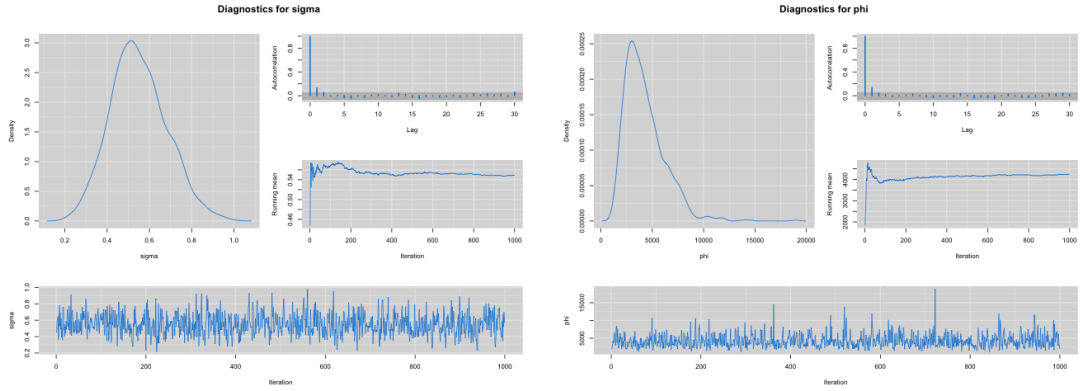
Lag 1 auto correlation log posterior



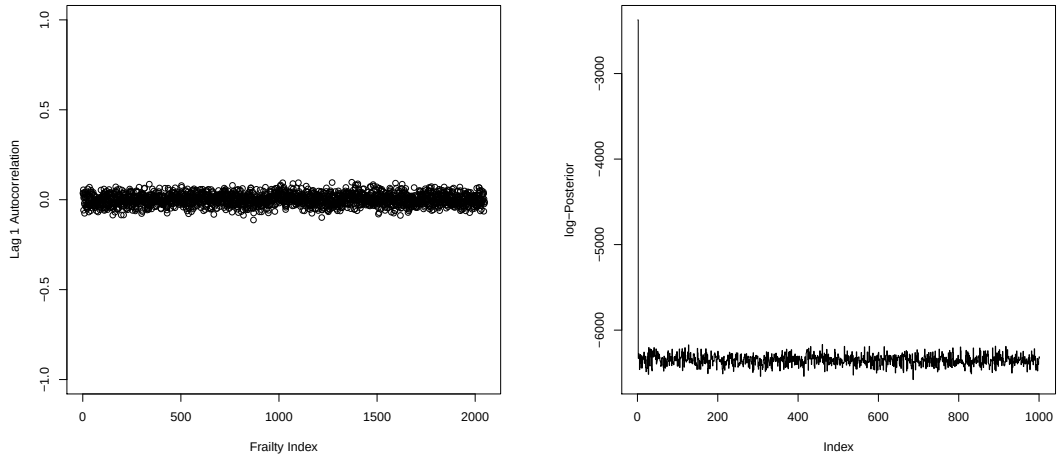
Prioposterior plot



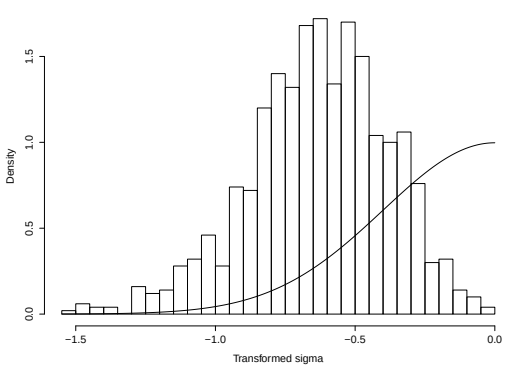
KEDAH:

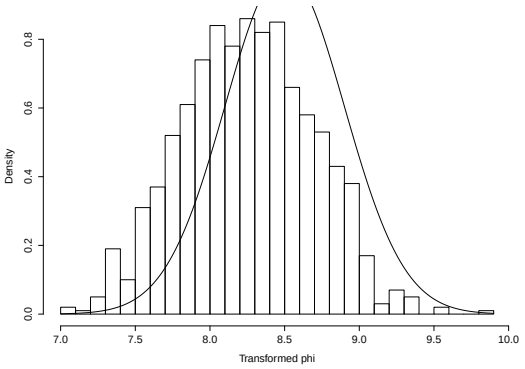


Lag 1 auto correlation log posterior

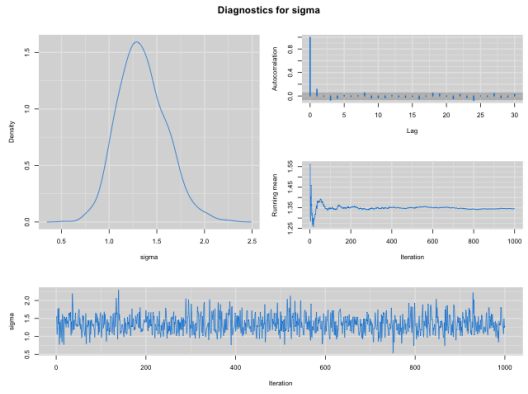


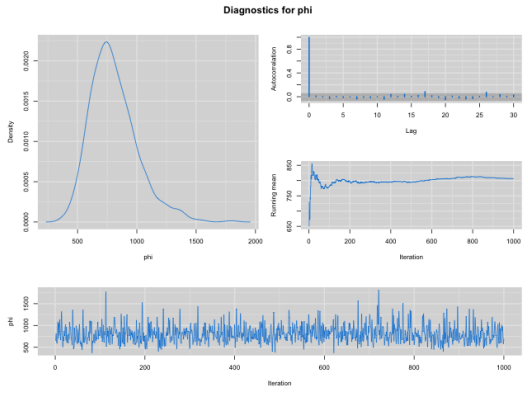
Prioposterior plots



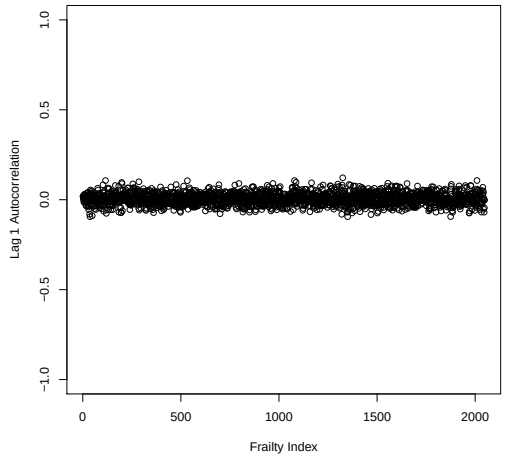


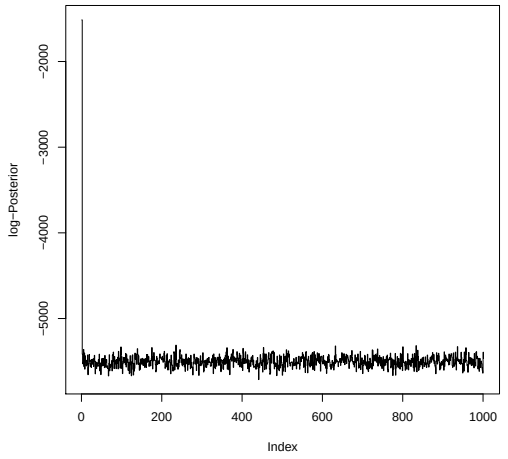
KUALA LUMPUR:



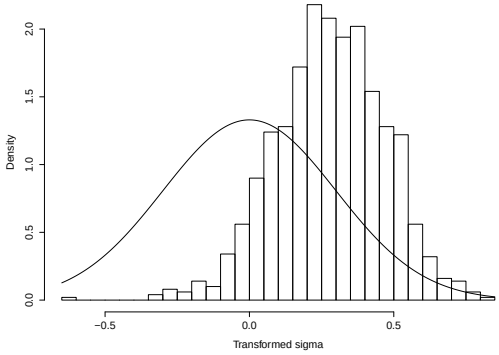


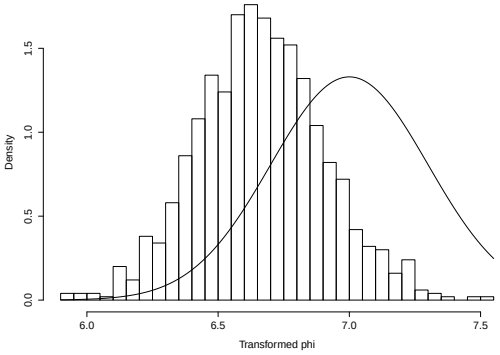
Lag 1 auto correlation log posterior



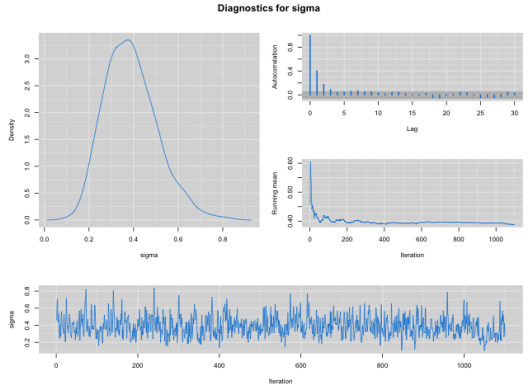


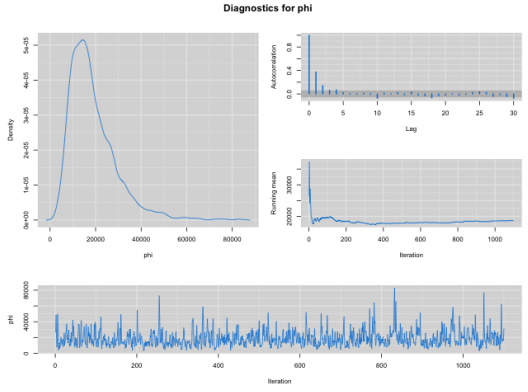
Prioposterior plots



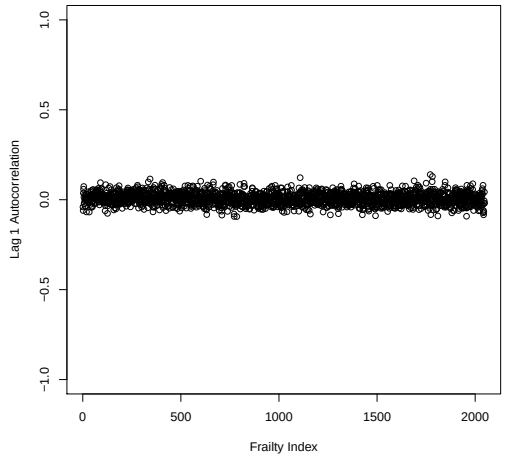


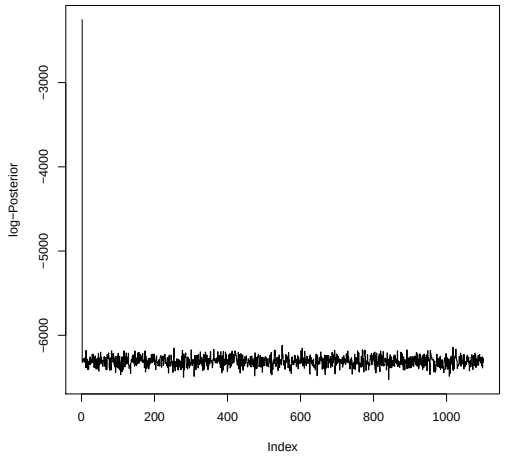
SARAWAK:



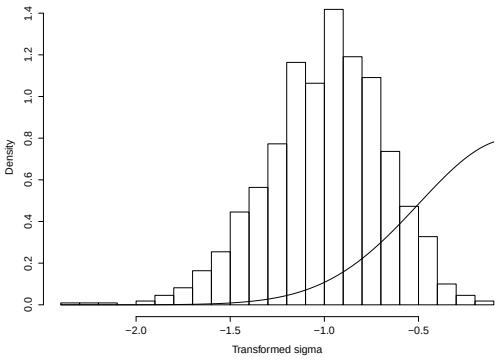


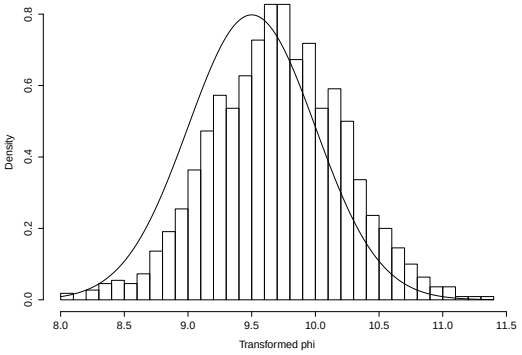
Lag 1 auto correlation log posterior





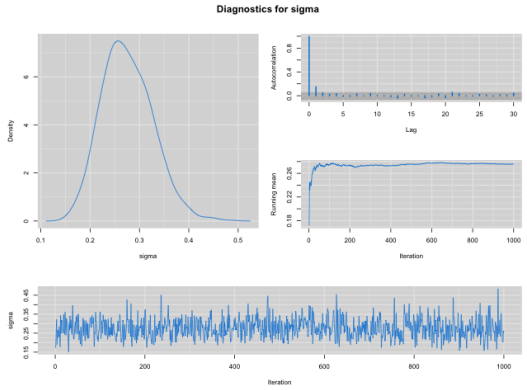
Prioposterior plots

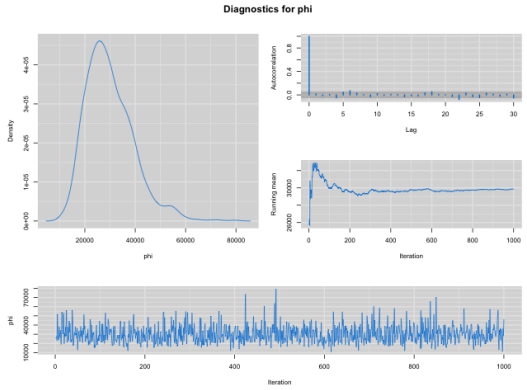




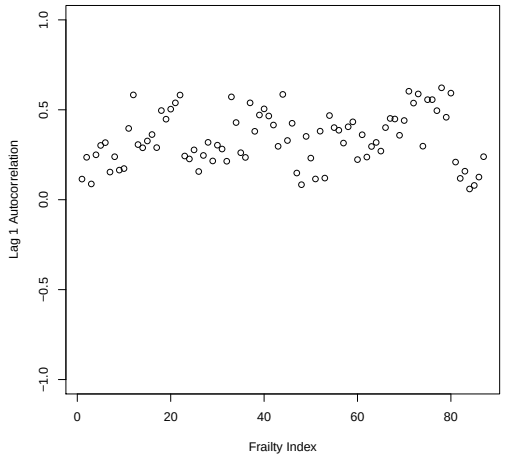
PENINSULAR MALAYSIA:

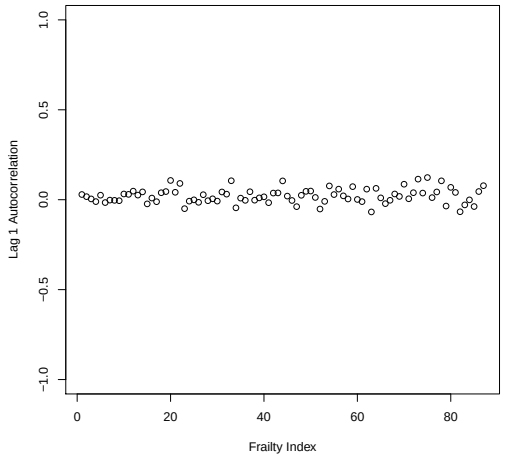
MCMC convergence from trace plots

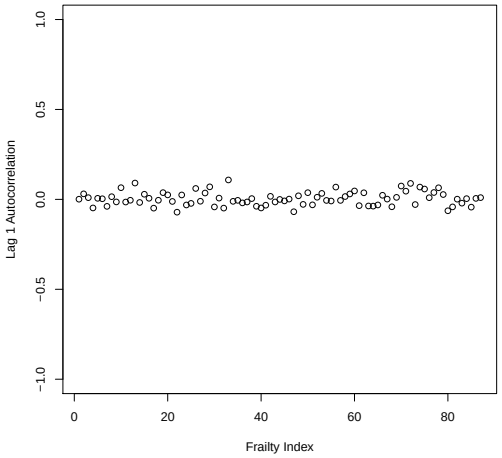


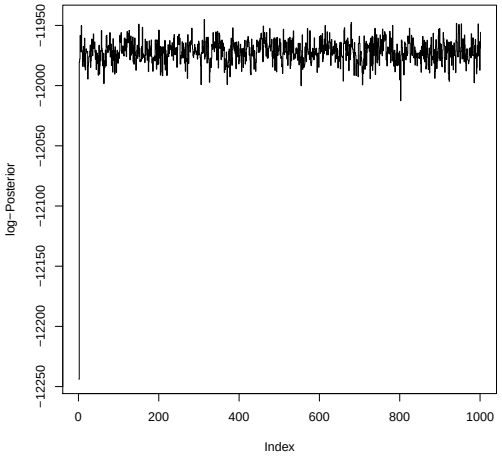


Lag 1 auto correlation log posterior

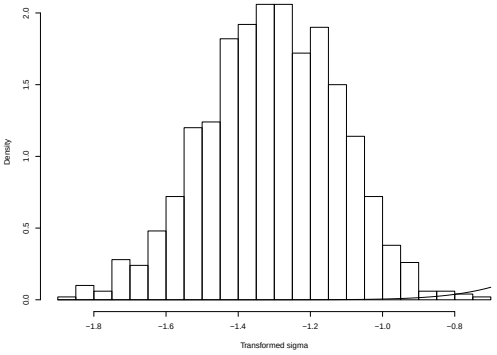


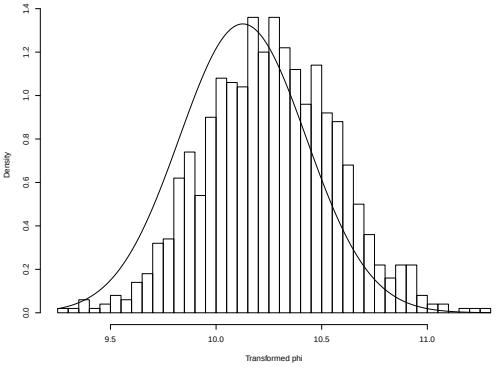




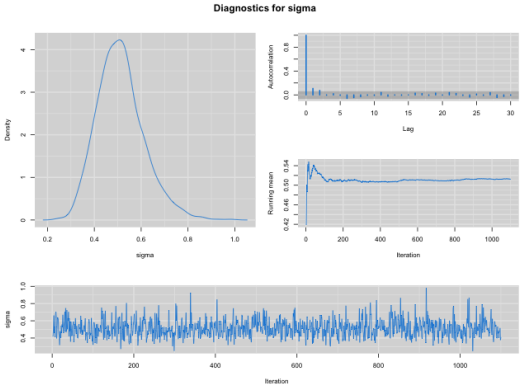


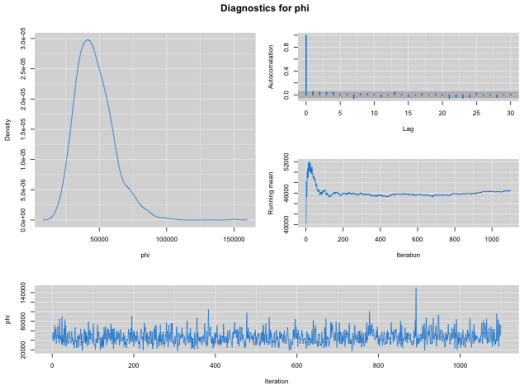
Prioposterior plots





EAST MALAYSIA





Lag 1 auto correlation log posterior

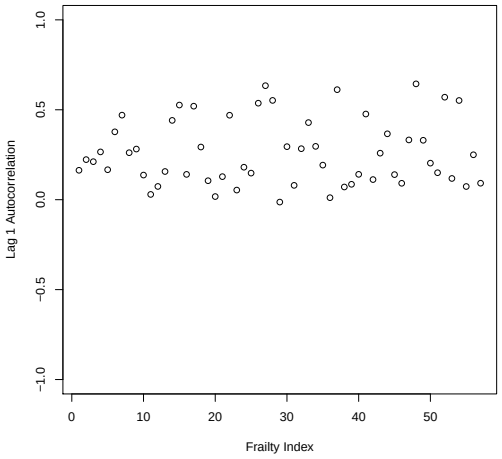


FIGURE 4.15: Autocorrelation plot, lag1

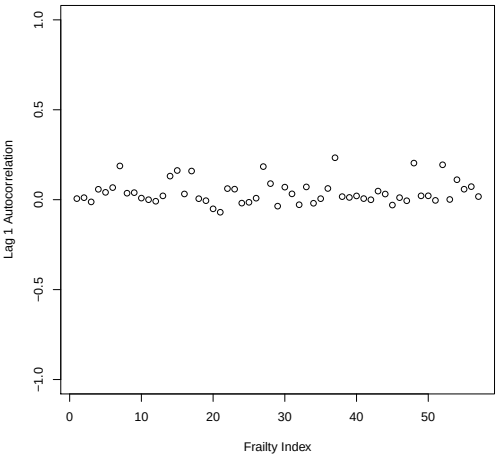


FIGURE 4.16: Autocorrelation plot, lag5

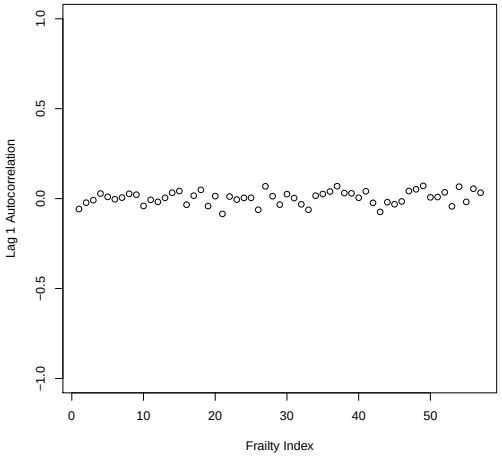


FIGURE 4.17: Autocorrelation plot, lag10

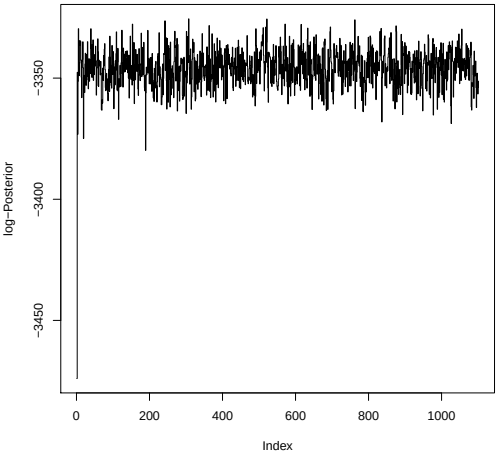


FIGURE 4.18: logposterior

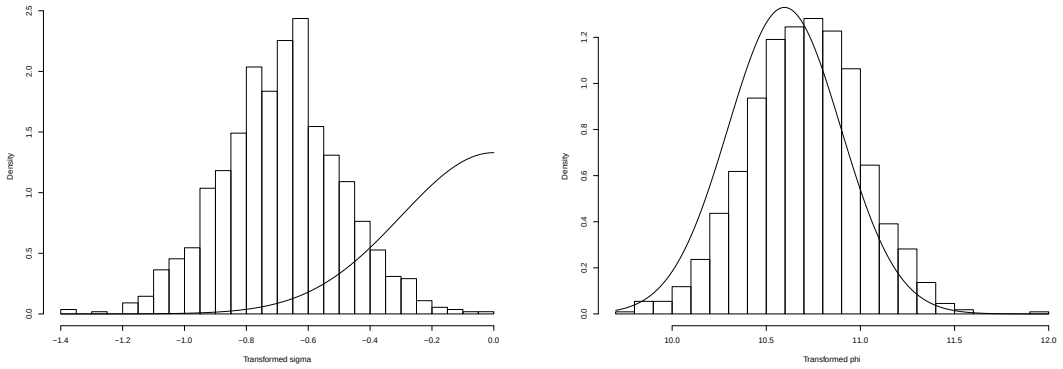


FIGURE 4.19: Priorposterior plots

References

Baade, P. D., Dasgupta, P., Aitken, J. F., and Turrell, G. (2013). "Geographic remoteness, area-level socioeconomic disadvantage and inequalities in colorectal cancer survival in Queensland: a multilevel analysis." *BMC cancer*, 13:493.

Barclay, K. L., Goh, P.-J., and Jackson, T. J. (2015). "Socio-economic disadvantage and demographics as factors in stage of colorectal cancer presentation and survival." *ANZ journal of surgery*, 85:135–139.

Canfield, M. A., Ramadhani, T. A., Langlois, P. H., and Waller, D. K. (2006). "Residential mobility patterns and exposure misclassification in epidemiologic studies of birth defects." *Journal of Exposure Science and Environmental Epidemiology*, 16:538.

Dejardin, O., Jones, A., Rachet, B., Morris, E., Bouvier, V., Jooste, V., Coombes, E., Forman, D., Bouvier, A., and Launoy, G. (2014). "The influence of geographical access to health care and material deprivation on colorectal cancer survival: evidence from France and England." *Health & Place*, 30:36–44.

Department of Statistics Malaysia (2016). Malaysia at a glance. https://www.dosm.gov.my/v1/index.php?r=columncone&menu_id=ZmVrN2FoYnBvZE05T1AzK0RLcEtiZz09.

Etxeberria, J., Ugarte, M. D., Goicoa, T., and Militino, A. F. (2014). "Age- and sex-specific spatio-temporal patterns of colorectal cancer mortality in Spain (1975-2008)." *Population health metrics*, 12:17.

Fairley, L., Forman, D., West, R., and Manda, S. (2008). "Spatial variation in prostate cancer survival in the Northern and Yorkshire region of England using Bayesian relative survival smoothing." *British journal of cancer*, 99:1786.

Henderson, R., Shimakura, S., and Gorst, D. (2002). "Modeling spatial variation in leukemia survival data." *Journal of the American Statistical Association*, 97:965–972.

Henry, K. A., Niu, X., and Boscoe, F. P. (2009). "Geographic disparities in colorectal survival." *International journal of health geographics*, 8:48.

Hsieh, J. C.-F., Cramb, S. M., McGree, J. M., Dunn, N. A., Baade, P. D., and Mengersen, K. L. (2016). "Spatially varying coefficient inequalities: evaluating how the impact of patient characteristics on breast cancer survival varies by location." *PloS one*, 11:e0155086.

Lejeune, C., Sassi, F., Ellis, L., Godward, S., Mak, V., Day, M., and Rachet, B. (2010). "Socio-economic disparities in access to treatment and their impact on colorectal cancer survival." *International journal of epidemiology*, 39:710–717.

MaCGDI (2012). 1malaysia map. https://www.mygeoportal.gov.my/1malaysia-map.

Manser, C. N. and Bauerfeind, P. (2014). "Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: a systematic review." *Gastrointestinal endoscopy*, 80:42–60.

Ministry of Housing, Communities Local Government UK (2015). English indices of deprivation 2015. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015.

Pebesma, E. and Heuvelink, G. (2016). "Spatio-temporal interpolation using gstat." *RFID Journal*, 8:204–218.

Rahman, N. A. and Zakaria, S. (2012). "The household-based socio-economic index for every district in Peninsular Malaysia." *World Academy of Science, Engineering and Technology, International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, 6:1389–1395.

Samat, N. and Shattar, A. (2014). "Spatial accessibility of colorectal." *World Applied Science Journal*, 31:1772–1782.

Samat, N., Shattar, A. K. A., Sulaiman, Y., Manan, A. A., and Weng, C. N. (2013). "Investigating geographic distribution of colorectal cancer cases: an example from Penang state, Malaysia." *Asian Social Science*, 9:38.

Shah, S. A., Neoh, H. M., Rahim, S., Azhar, Z. I., Hassan, M. R., Safian, N., and Jamal, R. (2014). "Spatial analysis of colorectal cancer cases in Kuala Lumpur." *Asian Pac J Cancer Prev*, 15:1149–1154.

Søgaard, M., Thomsen, R. W., Bossen, K. S., Sørensen, H. T., and Nørgaard, M. (2013). "The impact of comorbidity on cancer survival: a review." *Clinical epidemiology*, 5:3.

Taylor, B. and Rowlingson, B. (2017). "spatsurv: an R package for Bayesian inference with spatial survival models." *Journal of Statistical Software*, 77:1–32.

Wan, N., Zhan, F. B., Zou, B., and Wilson, J. G. (2013). "Spatial access to health care services and disparities in colorectal cancer stage at diagnosis in Texas." *The Professional Geographer*, 65:527–541.

Xu, F., Rimm, A. A., Fu, P., Krishnamurthi, S. S., and Cooper, G. S. (2014). "The impact of delayed chemotherapy on its completion and survival outcomes in stage II colon cancer patients." *PloS one*, 9:e107993.

Chapter 5

Modelling Incidence of Colorectal Cancer In Peninsular Malaysia

5.1 Introduction

Cancer is recognised as a major health burden across the world, as discussed in Section 2.3. Of the 14.1 million new cases of cancer diagnosed every year, around 1.36 million (9.7%) are new cases of colorectal cancer. This number is expected to increase by 80% by the year 2035 to approximately 2.4 million new colorectal cancer cases which will contribute to 1.3 million deaths worldwide (Douaiher et al., 2017).

Incidence of colorectal cancer varies widely geographically, with some countries exhibiting incidence rates 10 times larger than other countries. Australia/New Zealand have the highest rates of incidence with age standardized rates of 44.8 and 32.2 per 100,000 population in men and women respectively, while the lowest incidence was reported in Western Africa with comparative rates of 4.5 and 3.8 per 100,000 in men and women respectively. Similar patterns of colorectal cancer incidence were found in both sexes (Ferlay et al., 2015).

The incidence of colorectal cancer is higher in developed Asian countries such as Japan, Singapore and Korea than in developing Asian countries such as Thailand and India (GLOBOCAN, 2012). It is important to evaluate geographical (spatial) variation in incidence and risk factors affecting cancer incidence in order to (1) identify high risk populations, (2) help to prioritise the target population for screening and (3) facilitate health intervention.

There are several possible influences on the spatial variation in colorectal cancer incidence. Differences in risk factors such as eating habits, lifestyle (Li et al., 2017), socioeconomic status (Doubeni et al., 2012a, Jandova et al., 2016, Singh and Jemal, 2017) as well as race and genetic factors (Sung et al., 2005) may result in the geographical variation in colorectal cancer incidence that we observe. Disparity of screening policies, practices and rates among countries may lead to variation in colorectal cancer incidence; partly because increased screening may lead to increased detection of cancers (Li et al., 2017, M.M.Center et al., 2009), and also because screening rates have previously been linked with accessibility to healthcare. Areas with good accessibility to healthcare are reported to have higher cancer incidence than in areas living with limited accessibility to healthcare, including screening centres (Hao et al., 2009, Jackson et al., 2016).

Findings from previous studies in Malaysia show that the distribution of colorectal cancer varies within the country. Shah et al. (2014) observed the spatial variation of colorectal cancer cases by subdistrict or small counties, stage of cancer and ethnicity in Kuala Lumpur, Malaysia. The ethnicity of the cases was predominantly Chinese. The 'hot spots' were located in the North-West areas of the city and the 'cold spots' in the north east. The researchers discussed that this was probably due to the socioeconomic disparities between the areas, whereby the North-West had better socioeconomic status than the north east. However, no spatial modelling was employed in identifying the variation. Additionally, they suggested that the lack of mass screening to detect colorectal cancer among the population as

the possible cause for the low incidence in the cold spot areas, but they did not provide any evidence to support this.

The spatial distribution of colorectal cancer cases has also been described in the north (Penang) (Samat et al., 2013) and the east (Kelantan) of Peninsular Malaysia (Samat and Shattar, 2014). These researchers investigated the geographic distribution of colorectal cancer cases, sought to identify spatial clustering, and evaluated the spatial accessibility to healthcare facilities for these patients. They found that there were high numbers of cases concentrated in the main city centres, probably due to better accessibility of the population to screening facilities.

Our study aims to build on previous epidemiological research into colorectal cancer in Malaysia and the region. We aim to model the number of colorectal cancer cases in the four states : Kedah, Perlis, Penang and Perak, located in the North-West of Peninsular Malaysia. According to the data provider, these four states have near complete case ascertainment. We also aim to map the probability that the relative risk exceeds certain threshold, see section 5.3.3. Lastly, we will use the model to predict incidence in areas of Peninsular Malaysia where there is incomplete case ascertainment.

Our model, which is derived from the log-Gaussian Cox process (Diggle et al., 2013) can be conceptualized as a Generalized linear mixed effect model with spatially correlated random effects. This will be explained further in this Methods section. We will use point-located data in our analysis, aggregated to a fine grid, covering our study region. Using this 'complete data' model, we extrapolate our findings to the rest of Peninsular Malaysia and also discuss how to improve these predictions using model-based methods.

To the best of our knowledge, this is the first cancer research in Malaysia that uses spatial modelling to identify potential risk factors for colorectal cancer in Malaysia, as well as mapping those risks and predicting incidence in new areas.

The outcome of this study will be a valuable resource for healthcare providers, by helping them to anticipate demand for resources in identified high risk locations. It will also support the Cancer Registry in helping to understand how to improve case ascertainment for future cancer research.

5.2 Methods

5.2.1 Study area

The four Peninsular Malaysia states considered as having near complete case-ascertainment rate are Kedah, Penang, Perak and Perlis, and are shown in Figures 5.1 and 5.2. These states comprise the study area in which we will model the colorectal cancer incidence. The total population of these areas is almost 6 million people (census2010), and is principally Malay (58.1%), with 26.8%, 9.4% and 1.9% of the population identifying as Chinese, Indian or other race respectively. 3.8% of the population were non-citizens of Malaysia; these were excluded from our study.

FIGURE 5.1: Black border regions shows the 28 regions within the studied 4 states (North-West Peninsular Malaysia) from Malaysia map

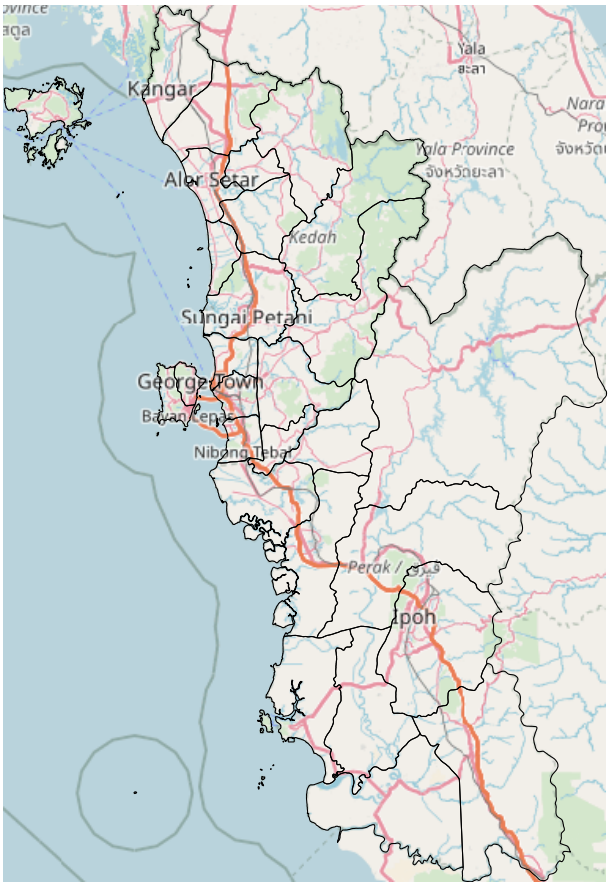


FIGURE 5.2: North-West Peninsular Malaysia

5.2.2 Data Description

The data used for this chapter comprise the colorectal cancer cases diagnosed between 2008 to 2013 in the North-West of Peninsular Malaysia. After excluding seven cases without address information, there remain 1248 colorectal cancer cases from these four states (556, 379, 247 and 69 cases in Kedah, Penang, Perak and Perlis respectively). These 1248 cases represent 39.6% of all colorectal cancer cases in Peninsular Malaysia.

For this chapter, we used each patient's geographical location for analysis. We applied the same procedure for acquiring the coordinate location of the cases as we did in Chapter 4, that is, we searched for the coordinates of the point locations given by the addresses available in the data. We used Google Maps, 1MalaysiaMap

(MaCGDI, 2012), and other resources from Google to search for the coordinates of the addresses. For those individuals where no address coordinate was available through the above databases, we generated an approximate address. We started with small street (of length less than 1km) to housing area to village to 'mukim' (sub-division of district) as also explained in the previous chapter (Table .

We used the 2010 data from Worldpop as the source of our population data for the analysis (Worldpop, 2017). These are raster data of 100m resolution. We projected the Worldpop data onto our region and aggregated it to a regular square grid of 4km sides. For the analysis, the number of people in each grid cell represent the population at risk.

TABLE 5.1: Table of percentage(%) addresses assigns in the North West of Peninsular Malaysia

States	Accurate address(%)	Small street	Village(%)	Small mukim(%)	Total %
Penang	84.9	6.9	8.2	0.0	100
Kedah	65.0	10.0	19.0	6.0	100
Perlis	37.8	4.3	57.9	0.0	100
Perak	61.6	12.1	24.7	1.6	100

We chose three explanatory variables to be evaluated as potential predictors for colorectal cancer incidence. They were:

- level of health service provision; determined using the number of hospitals per unit area as a proxy.
- a measure of socioeconomic status
- the proportion of Chinese people in the population

We have explained how we derived the value of hospital intensity and socio-economic index in Chapter 4. We chose the proportion of Chinese people as a predictor because this race had the highest age standardised incidence rate of colorectal cancer in Malaysia in a previous study (Hassan et al., 2016). This agrees with the analysis of colorectal incidence by race described in results section 5.3.1.

5.2.3 Statistical Analysis

5.2.3.1 Modelling Incidence of Colorectal Cancer

We modelled the number of colorectal cancer cases for the four states in North-West Peninsular Malaysia over the six year period 2008 to 2013.

As mentioned earlier, we used a point process model, fitted using the `R` software package `lgcp` (Taylor et al., 2015). It is a generalised linear mixed effects model in which the number of cases takes a Poisson distribution. The observed number of cases is explained by the chosen fixed effects (socioeconomic status, hospital intensity, proportion of Chinese people) and spatially correlated random effects. The model, also known as the spatial log-Gaussian Cox process is as follows:

$$X(s) \sim \text{Poisson}\{R(s)\}$$

$$R(s) = C_A \exp\{\alpha \log P(s) + Z(s)\beta + Y(s)\}$$

$$\log R(s) = \log(C_A) + \alpha \log P(s) + Z(s)\beta + Y(s)$$

$X(s)$ denotes the observed number of cases (counts) in the grid cell containing spatial location s. C_A is the cell area . We chose to use the covariate transformation $\log P(s)$ in the model to give a more flexible relationship between population and the number of cases (rather than just assuming proportionality as is common for Poisson models). If it transpires that $\alpha \approx 1$, then this, in essence similar to including $P(s)$ as an offset in the Poisson model. $Z(s)$ is a vector of area level covariates (fixed effect) with associated effect β. $Y(s)$ is a spatial random effect. The interpretation of $Y(s)$ is that having accounted for the fixed effects, $Y(s)$ represents variation in risk not accounted for by the effects. We assume that $Y(s)$ is similar at locations S_1 and S_2 that are close to each other, but nearly uncorrelated for S_1 and S_2 that are far apart. We use an exponential model to

describe the decay in correlation for two points. S_1 and S_2 that are distance d apart, we assume $\text{corr}(Y(S_1), Y(S_2)) = \exp\frac{-d}{\phi}$ for some parameters $\phi > 0$. We further assume the marginal variance of Y is $\sigma^2 > 0$.

We normalized the area level covariates chosen for the model as we want to see which factors had the biggest affect on colorectal cancer cases in our study.

5.2.3.2 Prediction

In order to predict the expected number of cases outside the region used to create the model, or, to be exact, for the whole of Peninsular Malaysia, we need estimates for all terms in the model, C_A, $\lambda(s)$, $Z(s)$, β, α, ϕ, σ^2 and $Y(s)$. Note that outside North-West Peninsular Malaysia, our predictions are extrapolations and should be treated with some caution.

We arrange the prediction grid so that it is an extension of the grid previously used for analysis. Therefore, the grid we used in North-West Peninsular Malaysia is a subset of the grid covering all of Peninsular Malaysia. The main complication with producing extrapolated predictions of the expected number of cases concerns the process Y. Since Y is correlated spatially, producing these predictions would require repeated inversion of a very large matrix which is beyond the scope of desktop computers.

Our model assumes that $\mathbb{E}(Y(s)) = \frac{-\sigma^2}{2}$ for any s, as this is how Y is parameterized in the package `lgcp` used to fit these data (Taylor et al., 2015). Therefore, rather than simulate a correlated Y on the area outside North-West Peninsular Malaysia, we instead simulated uncorrelated Y, so that Y has expectation $\frac{-\sigma^2}{2}$ and variance σ^2 for each retained sample of σ from the $MCMC$ chain used to fit the model.

We also sought to obtain an estimate of incidence of cases of colorectal cancer across the whole of Peninsular Malaysia. To do this, is in principle, straightforward from our model; it just involves aggregating the predicted number of cases in each cell over Peninsular Malaysia and dividing by the population at risk. However, this is complicated by the fact that the National Patient Cancer Registry - Colorectal Cancer database is known to be incomplete beyond North-West Peninsular Malaysia. Each sample i of our MCMC chain, $(\alpha^{(i)}, \beta^{(i)}, \sigma^{(i)}, \phi^{(i)})$ yields a different prediction of the total number of cases, T_i. Since we observed $T^* = 3155$ cases in our database, we know that the T must be $> T^*$. Therefore in our prediction for the total number of cases, we give inference for $\max[T_i, T^*]$, including a confidence interval.

We then mapped the predicted incidence for colorectal cancer for the whole of Peninsular Malaysia. We were aware of the uncertainty of the prediction presented in the map, hence we created another plot, a plot of the posterior probability that the incidence in each area exceeds the national average. All this is shown in results section 5.3.4. We also predict the mean case ascertainment rate across the whole of Peninsular Malaysia with a 95% credible interval, calculated as the number of observed cases divided by the number of expected (predicted) cases.

5.3 Results

5.3.1 Exploratory Analysis

The Malaysian colorectal cancer patients in our data identify as ethnically Malay, Chinese or Indian. Table 5.2 shows the proportion of the population (number and percentage) of each race in North-West Peninsular Malaysia. The Malay are the main race in this region, comprising between 56.5 and 93.2% of the total population in these four states. Indians are the least populous in all four states. In all states but Perak, Indian people make up less than 10% of the population. The proportion of the population with Chinese ethnicity is almost equal in Perak and Penang (30% and 32% respectively), but in Kedah and Perlis, people of Chinese origin make up just 16.5% and 5.5% of the total population respectively.

TABLE 5.2: Population by race in NorthWest Peninsular Malaysia, n (%)

	Kedah	Perlis	Perak	Penang
Malays	394975 (74.8)	55053 (93.2)	216281 (56.5)	100519 (59.7)
Chinese	87510 (16.6)	3241 (5.5)	117023 (30.6)	54114 (32.2)
Indians	45259 (8.6)	794 (1.3)	49431 (12.9)	13691 (8.1)
Total	527744 (100)	59088 (100)	382735 (100)	168324 (100)

Table 5.3 shows the ethnic proportions among patients with colorectal cancer in North-West Peninsular Malaysia between 2008 and 2013.

TABLE 5.3: Colorectal cancer cases by race in North-West Peninsular Malaysia

	Kedah	Perlis	Perak	Penang
Malays	337 (60.2)	49 (74.2)	115 (46.6)	112 (29.6)
Chinese	170 (30.4)	17 (25.8)	111 (44.9)	232 (61.2)
Indians	53 (9.5)	No cases	21 (8.5)	35 (9.2)
Total	560 (100)	66 (100)	247 (100)	379 (100)

Next, we calculate the age standardised incidence rates by race for colorectal cancer patients diagnosed between 2008 to 2013, see Table 5.4. Among these states, Kedah has the highest overall age standardised incidence rate with 5.6 (5.2, 6.1) per 100 000 population per year and Perak has the smallest age standardized incidence rate with 1.8 (1.5, 1.9) per 100 000 population per year. Those of Chinese ethnicity had the highest age standardised incidence of colorectal cancer for all four states in the North-West of Peninsular Malaysia. We note that there were no cases of colorectal cancer from patients of Indian ethnicity in Perlis, most likely because it is a small region with a small proportion of Indian people living there.

TABLE 5.4: Age standardized incidence rates by race of colorectal cancer between 2008 to 2013 (per 100 000 population)

Incidence rate (95%CI)				
	Kedah	**Perak**	**Perlis**	**Penang**
Overall	5.6 (5.2,6.1)	1.8 (1.5,1.9)	5.2 (4.0,6.5)	4.5 (4.0,4.9)
Malay	4.6 (4.1,5.1)	1.7 (1.4,2.1)	4.4 (2.9,5.9)	4.0 (3.3,4.8)
Chinese	9.5 (8.1,10.9)	1.9 (1.6,2.3)	11.7 (6.1,17.3)	4.8 (4.2,5.5)
Indian	6.2 (4.3,8.1)	1.4 (0.8,1.9)	No cases	4.3 (2.8,5.7)

5.3.2 Modelling the Risk of Colorectal Cancer in North-West Peninsular Malaysia

We used the `R` package `lgcp` (Taylor et al., 2015, 2013) to run MCMC until we obtain satisfactory convergence. This was achieved with 1,000,000 iterations, an initial burn-in of 10,000 iterations and thinning every 1000^{th} sample. We looked for evidence of good mixing and satisfactory convergence as follows.

To check for convergence and good mixing we examined trace plots and plots of autocorrelation (Figures 5.4 to 5.6). Since these are marginal summaries, we also produced a plot of the log-posterior(target) (Figure 5.3) over the duration of the chain to check for global convergence (see Taylor et al. (2013)). For clarity, Figure 5.4 shows the lag 1, 5 and 15 autocorrelation for Y in each cell of the grid: it is easier to visualize this way than examining thousands of traceplots by eye. For α and ϕ, the most challenging parameters with which to achieve good mixing, we also compared plots of the prior and posterior (Figure 5.7) to check how well these were identified by our data. That the prior and posterior in both cases look different is an indication that our data were able to identify both σ and ϕ. For all other parameters in the model mixing was quick and convergence was confirmed by visual inspection.

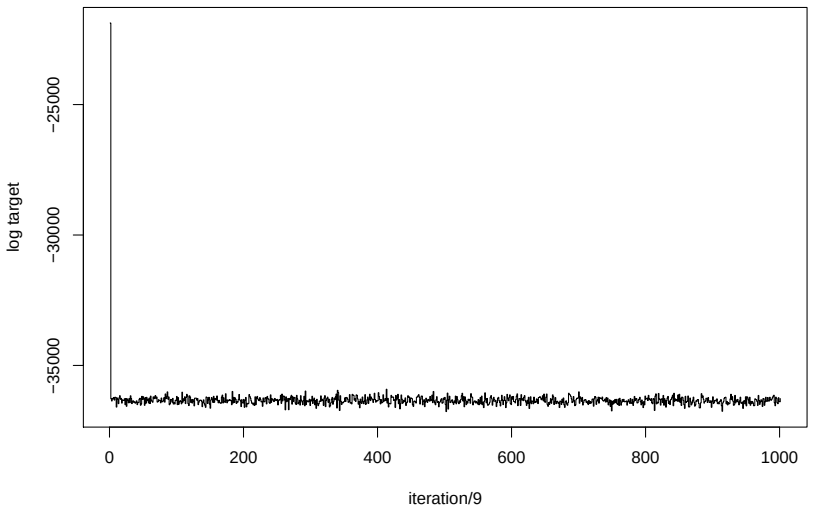


FIGURE 5.3: Diagnosing convergence to a posterior mode: a plot of the log(target)

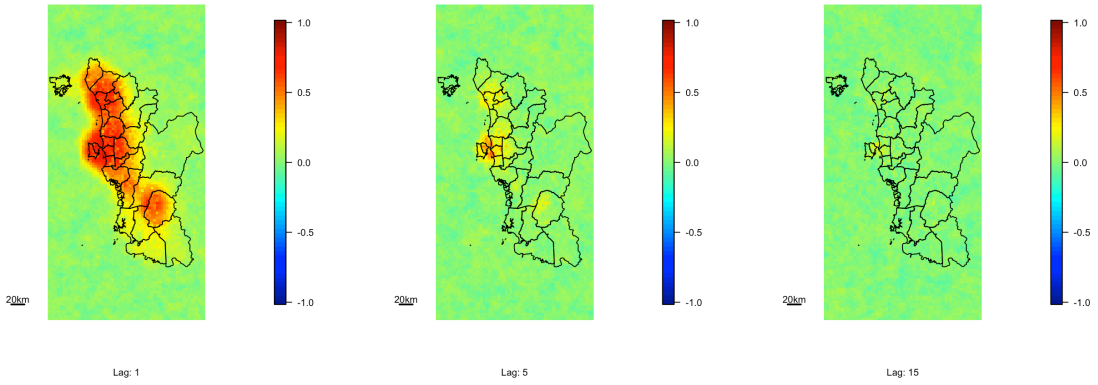


FIGURE 5.4: Lag 1, 5, 15 autocorrelation plots (from left to right)

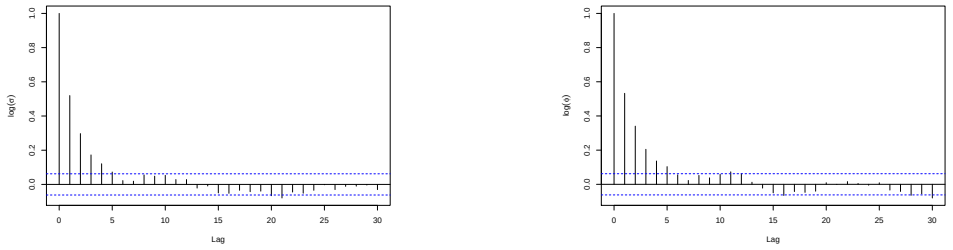


FIGURE 5.5: Autocorrelation plots for $\log \sigma$ and $\log \phi$

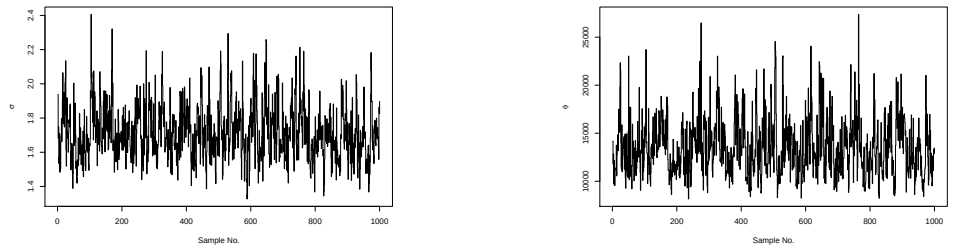


FIGURE 5.6: Trace plots of the parameters $\log \sigma$ and $\log \phi$

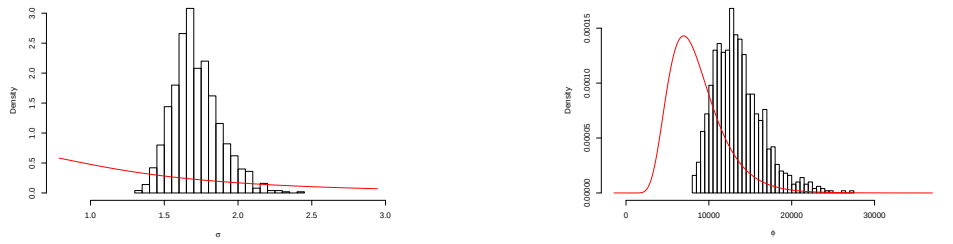


FIGURE 5.7: Plots of the prior and posterior values of $\log \sigma$ and $\log \phi$

Having ascertained satisfactory convergence, we proceed with making inferences from the model. Table 5.5 summarises the parameter estimates from our model. The variable that had the largest effect on incidence was hospital intensity: a unit increase in normalised hospital intensity led to a risk 1.58 (CRI (0.82, 2.91)) times that of existing colorectal cancer cases in a place/area. A unit increase in normalized value of socioeconomic index led to a 1.18 (CRI (0.58, 2.25)) increased risk of colorectal cancer cases, and a unit increase in the proportion of Chinese people in the population led to an increase in risk of 1.11 (CRI (0.64, 1.94)). However none of the variables in the model were statistically significant.

A matter that complicates inference is that the socioeconomic index and the proportion of Chinese people are correlated variables. We would like to keep both of these variables in the model because we want to know the joint effect of both on incidence of colorectal cancer. Therefore, interpretation of the coefficients of these variables needs to be undertaken with care. In Figure 5.8 we present the effect of these variables using a bivariate contour plot. For most of the areas, the effects of both variables were not significant, (i.e the credible interval contained 1), however the conditional effects were significant for low or high values of the other variable. We explain in more detail about this in Appendix 5.4.

We also fit models where we include one of the variables individually and we found that the effects were slightly larger and they were also only marginally insignificant, with the vast majority of the credible interval greater than 1, see Table 5.6 and Table 5.7.

Table 5.5: Parameter estimates for the LGCP point pattern model for the colorectal cancer data in North-West Peninsular Malaysia

	median	lower 95% CRI	upper 95% CRI
σ	1.69	1.44	2.07
ϕ	13137	9074	21073
$\exp(\beta_{Intercept})$	5.9×10^{-10}	2.59×10^{-10}	1.32×10^{-9}
$\exp(\alpha)$	2	1.8	2.24
$\exp(\beta_{HospIntensity})$	1.58	0.816	2.91
$\exp(\beta_{seIndex})$	1.18	0.583	2.25
$\exp(\beta_{ChineseProp})$	1.11	0.639	1.94

Table 5.6: LGCP Model 2: without the 'proportion of Chinese' variable

	median	lower 95% CRI	upper 95% CRI
σ	1.72	1.48	2.05
ϕ	13438	9484	21674
$\exp(\beta_{Intercept})$	5.95×10^{-10}	3.57×10^{-10}	1.56×10^{-9}
$\exp(\alpha)$	2	1.81	2.18
$\exp(\beta_{HospIntensity})$	1.50	0.925	2.46
$\exp(\beta_{SEIndex})$	**1.38**	0.966	2.04

Table 5.7: LGCP Model 3: without 'Socioeconomic index' variable

	median	lower 95% CRI	upper 95% CRI
σ	1.72	1.45	2.11
ϕ	14232	9796	21375
$\exp(\beta_{Intercept})$	6.24×10^{-10}	3.3×10^{-10}	1.51×10^{-9}
$\exp(\alpha)$	1.96	1.78	2.15
$\exp(\beta_{HospIntensity})$	1.37	0.779	2.33
$\exp(\beta_{ChineseProp})$	**1.29**	0.951	1.82

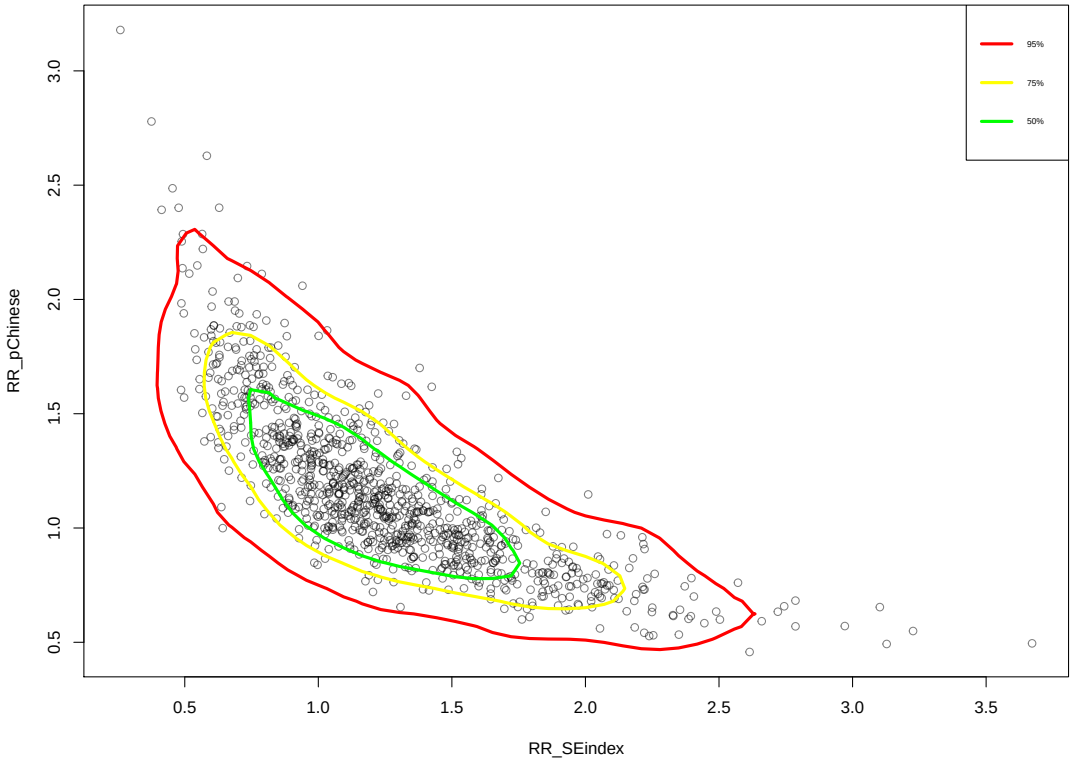


FIGURE 5.8: Contour plot showing the joint relative risk for SE index and the proportion of Chinese ethnicity

We checked the model fit by plotting the observed versus expected cases, shown in Figure 5.9. We see that our model fits the data well.

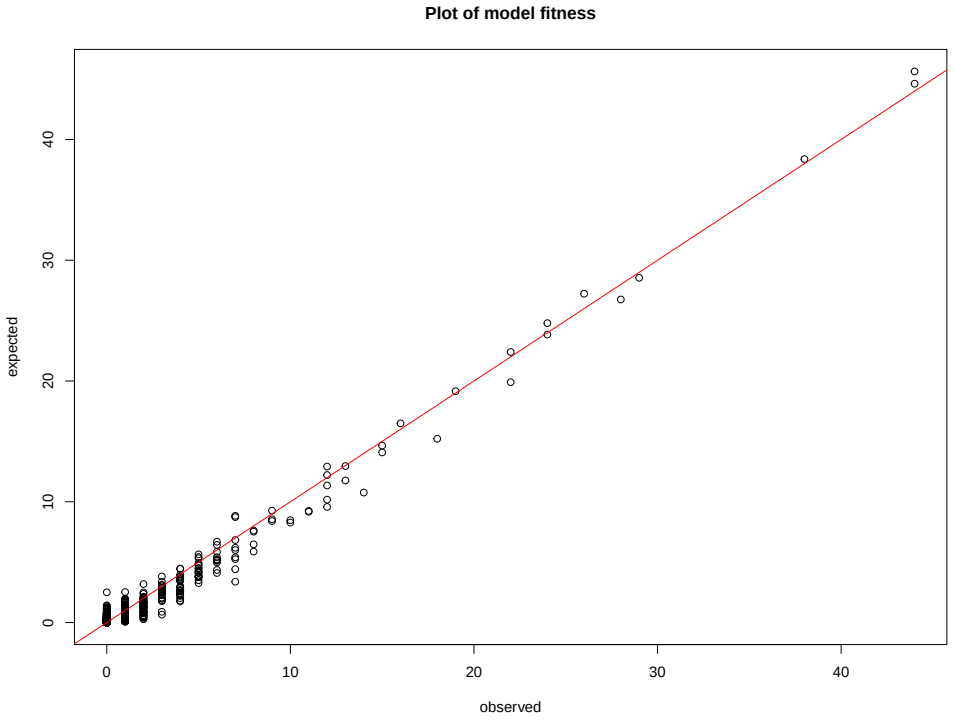


FIGURE 5.9: Plot for model fit

5.3.3 Maps of Exceedance Probabilities

It is of scientific interest to identify areas that have particularly high risk, even after adjusting for covariates. We can do this by plotting exceedance probabilities on a map. We consider $\mathbb{P}[\exp(Y) > t]$ for t =1.5, 2.5 and 5; that is, the probability that the relative risk is greater than 1.5, 2.5 and 5 over space.

Darker areas in the map highlight areas where it is likely that the relative risk thresholds were exceeded. The left map shows that in some areas the probability of the relative risk exceeding 1.5 is more than 0.75. There are a small number of areas where there was evidence that the relative risk was greater than 5.

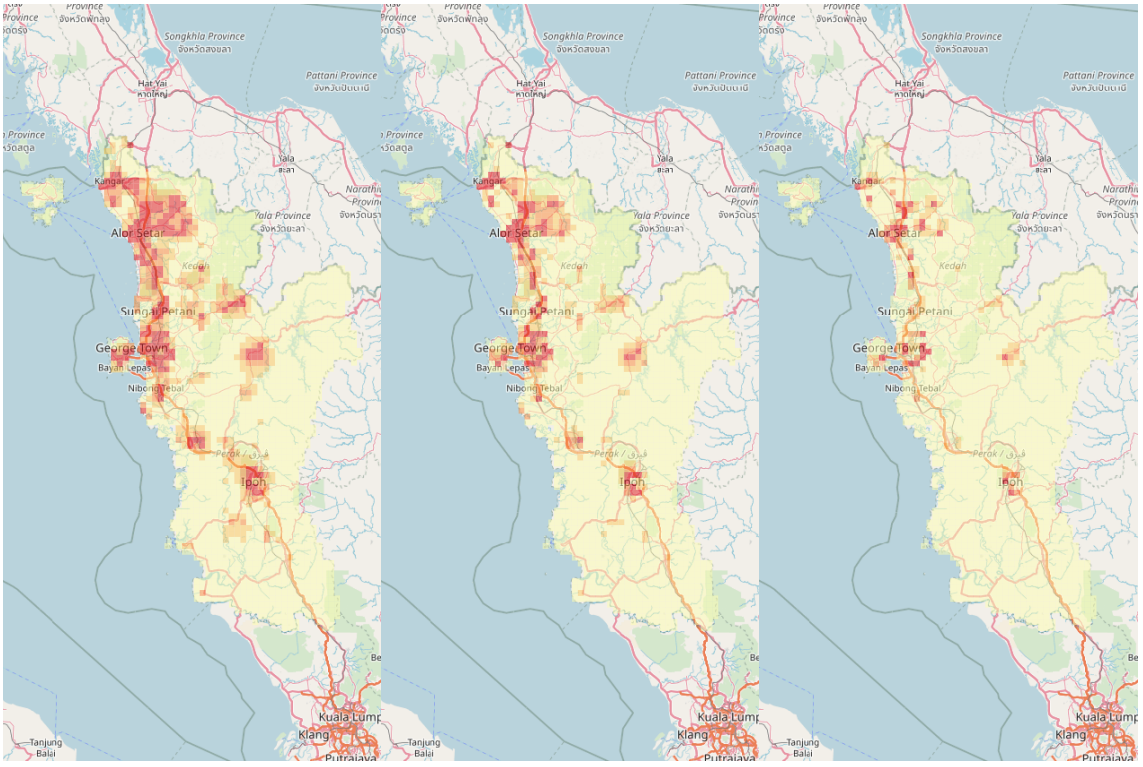


FIGURE 5.10: Maps illustrating $\mathbb{P}[\exp(Y) > t]$ for $t = 1.5, 2.5, 5$ from left to right
([0.0,0.25], (0.25,0.50], (0.50,0.75], (0.75,1.00])

5.3.4 Maps of Estimated Incidence of Colorectal Cancer for Peninsular Malaysia

Figure 5.11 shows the predicted annual incidence per 100,000 population. Our model predicts that the mean true incidence for colorectal cancer in Malaysia is 2.89 (95% CI (2.40, 5.25)) per 100,000 population per year. The plot shows that the predicted incidence in the western region of Peninsular Malaysia was higher than the eastern region. Taking into account the uncertainty in our predictions, we produced a map of the probability that incidence exceeds the national average,

shown in Figure 5.12. The map shows that the areas most likely to have an incidence above the national average are around Kuala Lumpur and Penang.

FIGURE 5.11: Predicted incidence of colorectal cancer in Peninsular Malaysia ([0,2], (2,10], (10,20], (20,40])

FIGURE 5.12: Probability that incidence of colorectal cancer is above the national average ([0.0,0.15], (0.15,0.30], (0.30,0.45], (0.45,0.60])

5.3.5 Estimating the Case Ascertainment Proportion

As mentioned earlier, for areas outside North-West Malaysia there is incomplete case ascertainment. By summing up the predicted numbers of cases per grid cell over Peninsular Malaysia, we are able to form a prediction of the total number of cases. Comparing this with the observed number of cases gives us the case ascertainment rate.

Case ascertainment, $\text{CA} = 100 \times \frac{\text{observed number of cases}}{\text{predicted number of cases}}$.

Each iteration of the MCMC algorithm gives us a different predicted number of cases, say $P^{(i)}$ for the i^{th} iteration. Therefore

$$\text{CA}^{(i)} = 100 \times \frac{\text{observed number of cases}}{\max P^{(i)}}$$

is our estimate of the case ascertainment based on the i^{th} iteration and we can form a credible interval in the usual way, which here gives an overall case ascertainment rate of 93.1%, CRI (47.5%, 100%).

5.4 Discussion

In this chapter, we used a log-Gaussian Cox process to model the observed number of cases of colorectal cancer in Malaysia. We then constructed an estimate of the annual disease incidence and also an estimate of the case ascertainment rate.

The intensity of hospitals per unit area, a socioeconomic index and the proportion of Chinese population were not statistically significant in our model. We found that the town centres and urban areas in North-West Peninsular Malaysia had greater risk of having colorectal cancer cases relative to other areas.

Our use of hospital intensity per unit area as a factor affecting the number of colorectal cancer cases in the area, is chosen for its relationship with accessibility of healthcare or screening centres, which may influence cancer reporting. It has been reported (Samat et al., 2013) that in Malaysia, the areas with better accessibility to healthcare or screening facilities are likely to document more cases relative to areas with lower accessibility. We found that the number of colorectal cancer cases increased with increasing hospital intensity, which corroborates the previous research, though we did not find this effect to be statistically significant. Further research should seek to investigate better measures of screening provision in order to confirm this, but these data were not available to us.

Our study found that socioeconomic index and the proportion of Chinese in the area were positively correlated. We produced a bivariate density plot for these covariates and found that neither of these variables were significantly associated with the number of colorectal cancer cases in this study. This is similar to Ladabaum et al. (2014) who reported that the neighbourhood socioeconomic status and ethnic enclave are highly correlated and are significantly associated with the incidence of colorectal cancer among Asian populations in California. However, their study did not involve spatial modelling to identify the associated factors and

arguably, our model would give a better indication of the relationship with these two variables if we had information at a finer spatial scale.

The map for probability of exceedance relative risk of colorectal cancer in North-West Peninsular Malaysia shows variation in the risk for colorectal cancer cases across the region. We noted that town areas are highly likely to exceed the threshold of relative risk of increased number of colorectal cancer cases, and this effect is present even though we account for the additional population there. Apart from the variables for which we have controlled, knowledge of and attitude towards screening for colorectal cancer also may affect the variation in the number of cases in the country.

The Malaysian population was reported to have very low participation in colorectal cancer screening (Yusoff et al., 2012). A study on the rural population in Malaysia, found that the level of awareness of the risks and symptoms of colorectal cancer was very low (Su et al., 2013). On the other hand, Naing et al. (2014) suggests that with Malaysia's suburban population, knowledge of the warning symptoms of colorectal cancer was adequate, though the awareness of colorectal cancer screening programmes was low, but also a high proportion expressed their willingness to undergo the screening test. This suggests that raising awareness should be a priority in Malaysia, through education and media campaigns.

In this study we were not able to consider the lifestyle factors that have been reported to affect colorectal cancer such as dietary habits (Azeem et al., 2015) and physical activity (Doubeni et al., 2012b, Tayyem et al., 2013) as we do not have the data for these variables.

To our knowledge, this is the only study to date that has modelled incidence and spatial variation of colorectal cancer in Malaysia. The results of this study can be used to inform those involved in public health; describing the incidence and

risk of colorectal cancer in the North-West of Peninsular Malaysia, and predict the incidence of colorectal cancer in the whole of Peninsular Malaysia.

Our study had several limitations. We only managed to obtain the exact coordinates of addresses for 60% of the cases, the rest were an approximate coordinate. Future research could address this shortcoming directly using aggregated point process models (Taylor et al., 2017). A further shortcoming of the study is that cancer registrations are incomplete, meaning that there is under-ascertainment of cases. An investigation in to the likely reasons for and levels of this would be useful. It is also possible to adapt our modelling approach to formally account for the non ascertainment, though this would not be completely straightforward. The idea would be to use a censoring-type approach using the ordinary Poisson likelihood for areas with complete ascertainment, like North-West Peninsular Malaysia, that is:

$$\mathbb{P}(\text{number of observed cases}|\text{parameters})$$

and otherwise,

$$\mathbb{P}(\text{number of cases} > \text{observed number of cases}|\text{parameters})$$

This is equal to

$$1 - \sum_{i=0}^{\text{observed number of cases}} P(\text{number of cases} = i|\text{parameters})$$

Additionally, our chosen measure of socioeconomic status could be improved, it could be more spatially resolved. That this measure pertains to large geographical areas means that care should be taken in interpreting these effects and readers should be aware of potential ecological bias here. Our measure was derived from the 2000 census, published by previous study Rahman and Zakaria (2012), whereas our data involves colorectal cancer cases between 2008 to 2013. Even though we use the best data available we could have reduced error in our exposure measure had we had a more up to date and spatially fine metric for socioeconomic status.

Future analysis could also extend our work into East Malaysia: we were forced to limit our study to Peninsular Malaysia because of better all-round data quality.

The findings from this chapter shows that urban areas near to the town centres and health facilities had higher probability of having more colorectal cancer cases compared to other areas in the states.

Our findings of the incidence prediction for colorectal cancer incidence in Peninsular Malaysia shows that the west of Peninsular Malaysia had higher probability of having incidence which exceeds the national average even though we account for the additional population there.

One possible reason for increased burden of cancer in urban areas is that people with a known health condition are more likely to move to be near health care facilities (Kibele and Janssen, 2013).Likewise, higher incidence in urban areas could be explained by selective emigration towards diagnosis in towns where healthcare facilities tend to be diagnosed.

There was a variation in the colorectal cancer incidence in the population. Besides socioeconomic status and health accessibility that we have discussed, other factors need to be investigated by the public health authorities regarding why certain areas had such a high incidence compared to the others. There might be underreported cases as we have mentioned earlier that the case ascertainment rate varies across the country. Furthermore, despite higher incidence in the urban areas, this population had a better survival according to the findings from our previous chapter on spatial survival model.

Therefore, the incidence mapping that we produced might help the policy makers to better plan management to target the places with high incidence as well as giving more education in the places with very low incidence because of underreported cases. Perhaps, more education on cancer prevention and detection is needed to reach out to the community to improve cancer detection in rural areas

and to reduce the incidence of colorectal cancer for the whole population. A further recommendation would be to extend facilities for diagnosis and routine care to clinics situated outside cities so diagnosis and routine care can be dealt with in the region of the origin of the patient.

The main findings from this study research relevant to policy makers are:

1. that more funding is required in order to acquire better data on colorectal cancer, with better geographical information.

2. that from the perspective of trying to understand socioeconomic factors, more detailed information is required, perhaps an index of multiple deprivation, similar to that used in the UK.

3. that publications involving estimates of colorectal cancer incidence at the national level should be treated with caution (due to the non-ascertainment issue)

Appendix II: Correlation of SE index and proportion of Chinese

Recall that Figure 5.8 the density of the $\exp(\beta_{SEin})$ and $\exp(\beta_{pCh})$. This plot was produced using the `kde2d` function from the `MASS` (Venables and Ripley, 2002) package in `R`. We show the area that covers 95%, 75% and 50% of the samples. This contour plot was done using the `HPDregionplot` from the `emdbook` (Bolker, 2008) package in `R`.

Different values of each parameter will give different credible intervals of relative risk and whether the effect of one factor is significant or not will depend on the relative risk for the others. We can give relative risks conditional on the other factor. These are represented in Table 5.8 and Table 5.9. In most cases, the effect of both variables was not statistically significant, we see that the credible interval for the conditional risk included 1. For the conditional risk, the effect was significant for extreme low or high values of the other variable.

TABLE 5.8: Table of credible interval for relative risk of colorectal cancer in SE index in conditional to the relative risk for proportion of Chinese

$\exp(\beta_{\text{pCH}})$	CRI $(\exp(\beta_{\text{SEindex}})\vert\exp(\beta_{\text{pCH}}))$
0.60	(**1.58, 2.38**)
0.65	(**1.08, 2.04**)
0.80	(**1.18, 1.93**)
0.95	(0.99, 1.61)
1.10	(0.92, 1.62)
1.25	(0.81, 1.46)
1.40	(0.73, 1.62)
1.55	(0.78, 1.31)
1.70	(0.66, 1.12)
1.85	(0.71, 1.04)
2.00	(0.66, 1.06)
2.15	**0.58, 0.95**)
2.30	(**0.62, 0.62**)

TABLE 5.9: Table of credible interval for relative risk of Chinese proportion conditional on the relative risk of SE index

$\exp(\beta_{\text{SEin}})$	CRI $(\exp(\beta_{\text{pCH}})\vert\exp(\beta_{\text{SEin}}))$
0.50	(**2.28, 3.64**)
0.65	**1.51, 2.91**)
0.80	**1.23, 2.26**)
0.95	(0.99, 2.13)
1.10	(0.85, 1.61)
1.25	(0.80, 1.41)
1.40	(0.60, 1.26)
1.55	(0.56, 1.16)
1.70	(0.60, 1.20)
1.85	(0.54, 1.03)
2.00	(**0.50, 0.69**)
2.15	(**0.49, 0.72**)
2.30	(**0.50, 0.56**)

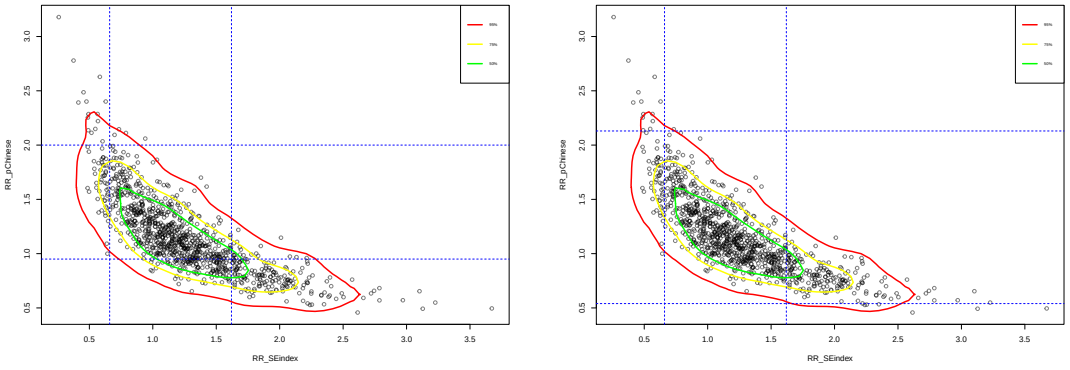


FIGURE 5.13: Contour plot shows the relative risk values for SE index and the proportion of Chinese. The blue lines show the thresholds at which the relative risk of each variable was not significant when calculated conditional on the other

References

Azeem, S., Gillani, S. W., Siddiqui, A., Jandrajupalli, S. B., Poh, V., and Syed Sulaiman, S. (2015). "Diet and colorectal cancer risk in Asia - A systematic review." *Asian Pac J Cancer Prev*, 16:5389–5396.

Bolker, B. M. (2008). *Ecological Models and Data in R.* Princeton University Press.

Diggle, P. J., Moraga, P., Rowlingson, B., and Taylor, B. M. (2013). "Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm." *Statistical Science*, pages 542–563.

Douaiher, J., Ravipati, A., Grams, B., Chowdhury, S., Alatise, O., and Are, C. (2017). "Colorectal cancer?global burden, trends, and geographical variations." *Journal of surgical oncology*, 115:619–630.

Doubeni, C. A., Laiyemo, A. O., Major, J. M., Schootman, M., Lian, M., Park, Y., Graubard, B. I., Hollenbeck, A. R., and Sinha, R. (2012a). "Socioeconomic status and the risk of colorectal cancer: an analysis of over one-half million adults in the NIH-AARP Diet and Health Study." *J Natl Cancer Inst 2012*, 15:3636.

Doubeni, C. A., Major, J. M., Laiyemo, A. O., Schootman, M., Zauber, A. G., Hollenbeck, A. R., Sinha, R., and Allison, J. (2012b). "Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence." *Journal of the National Cancer Institute*, 104:1353–1362.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015). "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012." *International journal of cancer*, 136:E359–E386.

GLOBOCAN (2012). Estimated cancer incidence, mortality and prevalence worldwide. international agency for research on cancer; 2012. http://globocan.iarc.fr/Default.aspx.Accessed04.08.15.

Hao, Y., Jemal, A., Zhang, X., and Ward, E. M. (2009). "Trends in colorectal cancer incidence rates by age, race/ethnicity, and indices of access to medical care, 1995–2004 (United States)." *Cancer Causes & Control*, 20:1855.

Hassan, M. R. A., Ismail, I., Suan, M. A. M., Ahmad, F., Khazim, W. K. W., Othman, Z., Said, R. M., Tan, W. L., Rahmah, S., Mohammed, N. S., et al. (2016). "Incidence and mortality rates of colorectal cancer in Malaysia." *Epidemiology and health*, 38:1–5.

Jackson, C. S., Oman, M., Patel, A. M., and Vega, K. J. (2016). "Health disparities in colorectal cancer among racial and ethnic minorities in the United States." *Journal of gastrointestinal oncology*, 7:S32.

Jandova, J., Ohlson, E., Torres, M. R., DiGiovanni, R., Pandit, V., Elquza, E., Nfonsam, V., et al. (2016). "Racial disparities and socioeconomic status in the incidence of colorectal cancer in Arizona." *The American Journal of Surgery*, 212:485–492.

Kibele, E. and Janssen, F. (2013). "Distortion of regional old-age mortality due to late-life migration in the Netherlands?" *Demographic Research*, 29:105–132.

Ladabaum, U., Clarke, C. A., Press, D. J., Mannalithara, A., Myer, P. A., Cheng, I., and Gomez, S. L. (2014). "Colorectal cancer incidence in Asian populations in California: effect of nativity and neighborhood-level factors." *The American Journal of Gastroenterology*, 109:579–588.

Li, K., Lin, G.-Z., Li, Y., Dong, H., Xu, H., Song, S.-F., Liang, Y.-R., and Liu, H.-Z. (2017). "Spatio-temporal analysis of the incidence of colorectal cancer in Guangzhou, 2010–2014." *Chinese journal of cancer*, 36:60.

MaCGDI (2012). 1malaysia map. https://www.mygeoportal.gov.my/1malaysia-map.

M.M.Center, A.Jemal, R.A.Smith, and E.Ward (2009). "Worldwide variations in colorectal Cancer." *CA: a cancer journal for clinicians*, 59:366–378.

Naing, C., Jun, Y. K., Yee, W. M., Waqiyuddin, S. B., Lui, L. C., Shaung, O. Y., and Haw, F. J. (2014). "Willingness to take a screening test for colorectal cancer: a community-based survey in Malaysia." *European Journal of Cancer Prevention*, 23:71–75.

Rahman, N. A. and Zakaria, S. (2012). "The household-based socio-economic index for every district in Peninsular Malaysia." *World Academy of Science, Engineering and Technology, International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, 6:1389–1395.

Samat, N. and Shattar, A. (2014). "Spatial accessibility of colorectal." *World Applied Science Journal*, 31:1772–1782.

Samat, N., Shattar, A. K. A., Sulaiman, Y., Manan, A. A., and Weng, C. N. (2013). "Investigating geographic distribution of colorectal cancer cases: an example from Penang state, Malaysia." *Asian Social Science*, 9:38.

Shah, S. A., Neoh, H. M., Rahim, S., Azhar, Z. I., Hassan, M. R., Safian, N., and Jamal, R. (2014). "Spatial analysis of colorectal cancer cases in Kuala Lumpur." *Asian Pac J Cancer Prev*, 15:1149–1154.

Singh, G. K. and Jemal, A. (2017). "Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: over six decades of changing patterns and widening inequalities." *Journal of environmental and public health*, 2017:19.

Su, T. T., Goh, J. Y., Tan, J., Muhaimah, A. R., Pigeneswaren, Y., Khairun, N. S., Normazidah, A. W., Tharisini, D. K., and Majid, H. A. (2013). "Level

of colorectal cancer awareness: a cross sectional exploratory study among multi-ethnic rural population in Malaysia." *BMC Cancer*, 2013:376.

Sung, J. J., Lau, J. Y., Goh, K., Leung, W., on Colorectal Cancer, A. P. W. G., et al. (2005). "Increasing incidence of colorectal cancer in Asia: implications for screening." *The lancet oncology*, 6:871–876.

Taylor, B., Davies, T., Rowlingson, B., and Diggle, P. (2015). "Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate log-Gaussian Cox processes in R." *Journal of Statistical Software*, 63:1–48.

Taylor, B. M., Andrade-Pacheco, R., and Sturrock, H. J. (2017). "Continuous inference for aggregated point process data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Taylor, B. M., Davies, T. M., Rowlingson, B. S., Diggle, P. J., et al. (2013). "lgcp: an R package for inference with spatial and spatio-temporal log-Gaussian Cox processes." *Journal of Statistical Software*, 52:1–40.

Tayyem, R. F., Shehadeh, I. N., AbuMweis, S. S., Bawadi, H. A., Hammad, S. S., Bani-Hani, K. E., Al-Jaberi, T. M., and Alnusair, M. M. (2013). "Physical inactivity, water intake and constipation as risk factors for colorectal cancer among adults in Jordan." *Asian Pac J Cancer Prev*, 14:5207–12.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Worldpop (2017). Malaysia 100m population. http://www.worldpop.org.uk/data/summary/?doi=10.5258/SOTON/WP00158.

Yusoff, H. M., Daud, N., Noor, N. M., and Rahim, A. A. (2012). "Participation and barriers to colorectal cancer screening in Malaysia." *Asian Pacific Journal of Cancer Prevention*, 13:3983–3987.

Chapter 6

General Discussion and Further Research

The research in this thesis focuses on two important parts of the epidemiology of colorectal cancer: measuring incidence and survival. Our research has addressed three main areas; identifying the individual prognostic factors for colorectal cancer, the spatial variation in survival of colorectal cancer and the incidence of colorectal cancer in Malaysia. This chapter summarises the main points from the findings of the three objectives of this study, as presented in chapters 3, 4 and 5 of this thesis.

For the first objective in this research, dicussed in Chapter 3, we investigated the effects of individual characteristics on the risk of death for colorectal cancer patients in Malaysia. We analysed data pertainng to over 4412 cases of colorectal cancer diagnosed between 2008 and 2013 in Malaysia. To our knowledge, this is the largest study of colorectal cancer survival ever carried out in the country. Our findings show that survival in colorectal cancer patients depends largely upon the stage at diagnosis. This is consistent with other studies (Kotake et al., 2016, Maringe et al., 2013, McKay et al., 2014).

One particularly concerning aspect was that there was a significant quantity of missing data in our study; data that involved important predictors such as cancer staging and tumour differentiation. The 'education level' variable had as much as 59% of its records missing.

In Chapter 4, we expanded the first analysis to take into account spatial correlation; we also added area level explanatory variables, namely socioeconomic index and accessibility to healthcare facilities. Better healthcare provision and higher socioeconomic index in the districts where patients live decrease the risk of death from colorectal cancer, but in this study the associations were not statistically significant. Our findings indicate wide spatial variation in colorectal cancer survival across Malaysia, after controlling for individual and area level characteristics. From the map of probability exceedance risk, we notice that a few areas in Kuala Lumpur, a known affluent city in Malaysia, have a greater probability of high hazard of death than we might expect. These findings can help policymakers and health practitioners to identify areas with poor survival outcome. The spatial survival model that we use can be applied to other diseases to investigate possible spatial variation and to interpret risk in space.

In Chapter 5, we studied another important aspect of epidemiology: the modelling of incidence. Accessibility to healthcare, here measured by the number of hospitals per unit area, positively influenced the estimated number of colorectal cancer cases in the population but the effect was not significant. The socioeconomic index and the proportion of Chinese people were positively correlated, but were not significantly associated with the risk of colorectal cancer. Town centres and urban areas had a greater risk of having more cases relative to other areas in North-West Peninsular Malaysia. In our model, we noted spatial variation in the predicted incidence in Peninsular Malaysia. The western regions of Peninsular Malaysia are likely to have greater incidence of colorectal cancer than the eastern regions of Peninsular Malaysia.

Comparing the incidence and survival of colorectal cancer in Malaysia, we noted that although the incidence of colorectal cancer in Malaysia is much lower compared to other high income Asian countries such as Singapore and Korea, these other countries had far greater overall 5-year survival rate (C.M.Oh et al., 2016, Peng et al., 2013). This is probably due to the late stage of diagnosis in Malaysian cases, as our study also found that the majority of colorectal cancer patients in our study were diagnosed at the later stages of disease (stages III and IV). This could be linked to low socioeconomic status, poor accessibility to healthcare and/or poor treatment adherence.

Our study was not able to identify a significant effect on either colorectal cancer incidence or survival from the socioeconomic index and accessibility to healthcare variables. This may be due to different measurements used in assessing those factors compared to those used in the literature. In a previous study, they assessed the individual socioeconomic status with markers such as income and occupation, while in our study we measured the socioeconomic index for the whole population of each area; and instead of measuring the spatial distance to healthcare facilities, we measured the number of hospitals per unit area. Furthermore, apart from the variables we can control, knowledge and attitude towards screening for colorectal cancer may also explain some of the variation in survival and incidence of colorectal cancer in the population. In a previous study, they reported that the Malaysian population had low rates of participation, knowledge and awareness of screening for colorectal cancer (Naing et al., 2014, Su et al., 2013, Yusoff et al., 2012). There could be other factors that might affect the incidence and survival of colorectal cancer that need to be considered in future research. For example, co-morbidities (Søgaard et al., 2013) and delay in treatments such as chemotherapy (Xu et al., 2014) may affect the survival outcome. On the other hand, non-clinical or cultural aspects such as lifestyle choices, diet and physical activity may be considered as risk factors for colorectal cancer in future research (Doubeni et al., 2012).

The main strength of our study is it can be considered as at the forefront of research in spatial modelling for colorectal cancer in Malaysia, and also that it is the largest study of its type to date. To our knowledge, this is the only study that uses spatial survival analysis to model and investigate geographical variation in the risk of, and predict incidence in, colorectal cancer in Malaysia.

The methods we have used have not been applied to data from Malaysia before. In particular, incidence of colorectal cancer has historically been difficult to estimate is South East Asia, but has become less complex and more reliable with the advent of national population-based cancer registries in the more economically developed countries in the region such as Japan and Singapore (Sung et al., 2005). Spatial studies of colorectal cancer, and cancer more generally, have had need of access to reliable cancer registry data, and these have not been widely available until now.

Estimates of population level incidence rely unlimitedly on the ability to count accurately the number of new cases of cancer in a time period (a calendar year commonly) and to have for the same period an accurate estimate of the population at risk. With reliable census data it is possible to accurately estimate a population size denominator, however counting the new cases of cancer is less easy. The development of the cancer registry system in Malaysia has improved matters and we have in this study taken advantage of the availability of cancer registry data. The accuracy of incidence or prevalence estimates based on registry data will be affected by the accuracy and completeness of the data held and provided by the registry. In previous analyses of cancer incidence in the region (Sung et al., 2005) incidence figures were available for nearby countries: Japan, Singapore and Thailand, but not for Malaysia. Two points are pertinent here. First, is that out paper fills a gap in knowledge of the country-level incidence of colorectal cancer. Our estimate of overall incidence in Malaysia is 2.89 cases/100,000 population/year. Based on this figure Malaysia's colorectal cancer incidence is the lowest in South and South East Asia (see Sung et al. (2005)).

Cancer incidence has been increasing in South East Asia. Incidence in affluent South East Asian countries is now similar to those in the West (IARC,WHO, 2010). This increase had been blamed on a shift in diet and exercise patterns from traditional to those more common in affluent societies, with higher meat and fat consumption, and lower levels of vegetable consumption and lower delay calorie expenditure (Godos et al., 2016). At face value our results would indicate that the increased in affluent country rates of colorectal cancer have not yet impacted Malaysia. If that is the case, then, the country should be aware that there is likelihood that as affluence increases so will colorectal cancer incidence and measures should be put in place now to identify and treat cases. Better perhaps, would be to educate the population to consume a healthy diet and to regularly exercise to prevent exposures linked to the risk of colorectal cancer (Huxley et al., 2009).

It is possible that our estimate of national rate of cancer incidence underestimates the true level. This would be so if the completeness of ascertainment of cases is low. We estimated that the completeness of the registry that provided us data to be 93% (see section 5.3.5). However, this estimate is based on the premise of the observed cases, and this is subject to error. Additionally, the accuracy and reliability of incidence estimate would be improved were we to have had this level of completeness and been able to use data for all states in the country rather than for the three we used. The impact on incidence estimates of incomplete registration can be profound. In England, it is estimated that the national cancer registry is 98-99% complete and that this level of"incompleteness" can lead to inaccuracies in estimates of survival of around 1% (Møller et al., 2011). If our estimate were to be nearer the lower 95% credible limit of our estimate, the Malaysian registry would be missing half of all cases of cancer, with a consequent impact on incidence estimates.

We have referred elsewhere to the need for a properly national cancer registry in Malaysia, To improve on this research this is vital. However, cancer registration

can only work if cases of cancer are diagnosed, and for this a way forward would be for ore people in Malaysia to come forward to their primary care doctors at an earlier stage than they do so presently, and for the diagnoses cases of cancer to be sent from the hospital system into the registry without loss of information. At that point work such as this can provide the means for the health authorities to identify how best to reduce risk and subsequent incidence of disease. In this endeavour, spatial methods such as those used here can be especially useful in identifying where the risk of exceedance of a set threshold is highest, and thus where resources can be targeted.

The quality of the data that we receive might affect the results of this study. For example, two important predictors in our study, education and smoking status had very high missing values, 59% and 43% of the total cases respectively. These groups also had higher risk of death compared to the reference group (no-education and non-smoking), but they were not significant. Based on existing epidemiological evidence it is likely that these two predictors will indeed affect survival, though through different causal routes. Had the full data for these variables been available, the Central Limit Theorem implies we would have been better able to identify the effect size. Under the additional assumption that higher risk patients (e.g. older, emergency cases) are less likely to provide these data, and there is some support for this assumption as detailed in section 3.2.2, we might also expect the effect size to be underestimated as our parameter inferences (e.g. the effect associated with being in the lowest education group) were based on a healthier population.

Moving to the spatial model, we did not find a significant effect for the socio-economic index nor for our measure of accessibility to healthcare on the survival and incidence of colorectal cancer in the population. It is difficult to predict the potential change in our estimates of these effects compared to if we had more reliable information on the associated covariates; this relates to the concept of ecological bias. Although education group is potentially linked to socioeconomic

status, it is only part of the picture and does not account for other factors that might affect SES. To obtain reliable SES data from individuals would require a range of questions to be answered as part of the initial data entry form. This might seem an unnatural thing to try to elicit routinely as part of a colorectal cancer diagnosis by medical practitioners. Similar comments apply to our chosen measure of healthcare accessibility: again this information would be better obtained from each patient.

In the future, the following suggestions may improve the quality of data and its resulting analysis. Reliable measurements of environmental factors are needed to provide good insight into the effects of potential risk factors for the disease. For example, a better metric is needed to measure the socioeconomic status and accessibility to healthcare in the country. Ensuring complete, accurate and consistently recorded data in the National Cancer Registry database is vital for reliable analysis and meaningful results. Perhaps active case ascertainment across the country and also better training of staff that interact with the database is required in order to minimise the amount of missing data.

To improve the methodology, two possible extensions can be considered. Firstly, we can perform the Poisson formal modelling of exactly and inexactly located cases (Taylor et al., 2017) and secondly, we can use a censoring-type approach to deal with the cell counts that are being underestimated as discussed at the end of Chapter 6.

The findings from this study may enlighten both health practitioners and patients on the subject of colorectal cancer in Malaysia and help policymakers, authorities and health professionals to develop better healthcare and adequately plan for disease management in Malaysia. Emphasis on increasing public knowledge of the risks, symptoms and prevalence of colorectal cancer, together with prominent campaigns to promote screening, should be a focus for the future.

References

C.M.Oh, Y.J.Won, K.W.Jung, H.J.Kong, H.Cho, J.K.Lee, D.H.Lee, and K.H.Lee (2016). "Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2013." *Cancer research and treatment: official journal of Korean Cancer Association*, 48:436.

Doubeni, C. A., Major, J. M., Laiyemo, A. O., Schootman, M., Zauber, A. G., Hollenbeck, A. R., Sinha, R., and Allison, J. (2012). "Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence." *Journal of the National Cancer Institute*, 104:1353–1362.

Godos, J., Bella, F., Torrisi, A., Sciacca, S., Galvano, F., and Grosso, G. (2016). "Dietary patterns and risk of colorectal adenoma: a systematic review and meta-analysis of observational studies." *Journal of Human Nutrition and Dietetics*, 29:757–767.

Huxley, R. R., Ansary-Moghaddam, A., Clifton, P., Czernichow, S., Parr, C. L., and Woodward, M. (2009). "The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence." *International journal of cancer*, 125:171–180.

IARC,WHO (2010). World health statistics annual. geneva: Who databank. http://www-dep.iarc.fr.

Kotake, K., Asano, M., Ozawa, H., Kobayashi, H., and Sugihara, K. (2016). "Gender differences in colorectal cancer survival in Japan." *International journal of clinical oncology*, 21:194–203.

Maringe, C., Walters, S., Rachet, B., Butler, J., Fields, T., Finan, P., Maxwell, R., Nedrebø, B., Påhlman, L., Sjövall, A., et al. (2013). "Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000–2007." *Acta Oncologica*, 52:919–932.

McKay, A., Donaleshen, J., Helewa, R. M., Park, J., Wirtzfeld, D., Hochman, D., Singh, H., and Turner, D. (2014). "Does young age influence the prognosis of colorectal cancer: A population-based analysis." *World journal of surgical oncology*, 12:370.

Møller, H., Richards, S., Hanchett, N., Riaz, S., Lüchtenborg, M., Holmberg, L., and Robinson, D. (2011). "Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates." *British journal of cancer*, 105:170.

Naing, C., Jun, Y. K., Yee, W. M., Waqiyuddin, S. B., Lui, L. C., Shaung, O. Y., and Haw, F. J. (2014). "Willingness to take a screening test for colorectal cancer: a community-based survey in Malaysia." *European Journal of Cancer Prevention*, 23:71–75.

Peng, L., Ling, C., Yew, C., Aizen, J., Yun, L., and Ho, W. (2013). "Singapore Cancer Registry Interim Annual Registry Report: Trends in Cancer Incidence in Singapore 2009-2013." *National Registry of Diseases Office (NRDO)*.

Søgaard, M., Thomsen, R. W., Bossen, K. S., Sørensen, H. T., and Nørgaard, M. (2013). "The impact of comorbidity on cancer survival: a review." *Clinical epidemiology*, 5:3.

Su, T. T., Goh, J. Y., Tan, J., Muhaimah, A. R., Pigeneswaren, Y., Khairun, N. S., Normazidah, A. W., Tharisini, D. K., and Majid, H. A. (2013). "Level of colorectal cancer awareness: a cross sectional exploratory study among multi-ethnic rural population in Malaysia." *BMC cancer*, 13:376.

Sung, J. J., Lau, J. Y., Goh, K., Leung, W., on Colorectal Cancer, A. P. W. G., et al. (2005). "Increasing incidence of colorectal cancer in Asia: implications for screening." *The lancet oncology*, 6:871–876.

Taylor, B. M., Andrade-Pacheco, R., and Sturrock, H. J. (2017). "Continuous inference for aggregated point process data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Xu, F., Rimm, A. A., Fu, P., Krishnamurthi, S. S., and Cooper, G. S. (2014). "The impact of delayed chemotherapy on its completion and survival outcomes in stage II colon cancer patients." *PloS one*, 9:e107993.

Yusoff, H. M., Daud, N., Noor, N. M., and Rahim, A. A. (2012). "Participation and barriers to colorectal cancer screening in Malaysia." *Asian Pacific Journal of Cancer Prevention*, 13:3983–3987.