RELATIVE RISK ESTIMATION OF TUBERCULOSIS DISEASE MAPPING WITH STOCHASTIC SLIR MODELS

IJLAL BINTI MOHD DIAH

MASTER OF SCIENCE (APPLIED STATISTICS)
UNIVERSITI UTARA MALAYSIA
2017

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

IJLAL MOHD DIAH

calon untuk Ijazah   **MASTER OF SCIENCE (STATISTICS)**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"RELATIVE RISK ESTIMATION OF TUBERCULOSIS DISEASE MAPPING WITH STOCHASTIC EQUATIONS"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : **21 Jun, 2017.**
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
June 21, 2017.

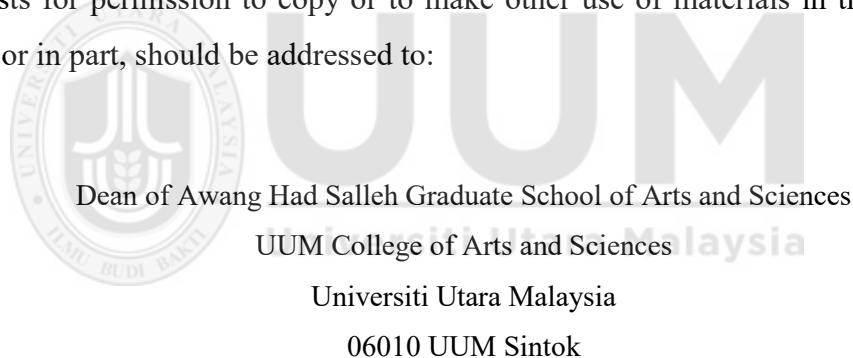|		
Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Nor Idayu Mahat**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Prof. Dr. Amran Ahmed**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Nor Aishah Ahad**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Nazrina Aziz**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Assoc. Prof. Dr. Maznah Mat Kasim**	Tandatangan *(Signature)*

Tarikh:
(Date) **June 21, 2017**

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Tuberkulosis (TB) adalah salah satu sebab utama kematian di negara membangun. Pemantauan penyakit ini pada masa kini hanya berdasarkan kepada jumlah kes yang dilaporkan. Sebagai alternatif, terdapat satu pendekatan yang lebih baik iaitu pemetaan penyakit yang menawarkan risiko relatif taburan geografi penyakit. Kajian terdahulu menggunakan Nisbah Kematian Piawai (SMR) dan model Poisson-gamma untuk menganggarkan risiko relatif tetapi model ini mempunyai beberapa kelemahan. Model SMR tidak dapat mengesan risiko relatif pada kawasan kecil manakala model Poisson-gamma tidak membenarkan penyesuaian kovariat. Oleh itu, matlamat kajian ini adalah untuk membangunkan model statistik alternatif bagi menganggar risiko relatif yang dinamakan model stokastik *Susceptible-Latently infected-Infectious-Recovered* (SLIR). Terdapat empat fasa dalam kajian ini. Pertama, model deterministik SLIR untuk penyebaran penyakit TB dibangunkan. Kemudian, model stokastik SLIR dibentuk. Seterusnya, model stokastik SLIR digunakan untuk menganggarkan risiko relatif bagi penyakit tersebut. Kemudian, prestasi model stokastik SLIR dibandingkan dengan model sedia ada berdasarkan nilai risiko relatif. Akhir sekali, peta risiko TB dibina. Untuk analisis berangka, kajian ini menggunakan satu set data kes TB yang dilaporkan di Malaysia dari 2008 hingga 2015. Penemuan menunjukkan bahawa terdapat perbezaan yang besar pada nilai anggaran risiko relatif apabila menggunakan model stokastik SLIR berbanding dengan model sedia ada. Ini dapat digambarkan dengan jelas melalui pemetaan penyakit di mana beberapa lokasi berubah warna daripada tona rendah (risiko rendah) kepada tona gelap (risiko lebih tinggi). Ini berlaku kerana mengambilkira komponen laten dalam model stokastik SLIR. Sebagai kesimpulan, kajian ini menawarkan model yang lebih baik dalam menganggarkan risiko relatif bagi penyakit TB. Penemuan ini juga dapat membantu kerajaan dalam mengutamakan lokasi yang memerlukan perhatian lanjut terutamanya dari aspek polisi kesihatan dan sokongan kewangan.
.

Kata kunci: Pemetaan penyakit, Risiko relatif, Model SLIR, Model stokastik, Tuberkulosis

Abstract

Tuberculosis (TB) is one of the death leading causes in developing countries. The current monitoring of the disease is based only on the total cases reported. Alternatively, a better approach called disease mapping offers geographic distribution of the disease relative risk. Previous studies used Standard Mortality Ratio (SMR) and Poisson-gamma models to estimate relative risk but these models have several drawbacks. SMR model cannot detect relative risk for small areas while Poisson-gamma model cannot allow for covariate adjustments. Hence, the objective of this study is to develop an alternative statistical model in estimating the relative risk called stochastic Susceptible-Latently infected-Infectious-Recovered (SLIR). There are four phases in this study. Firstly, the deterministic SLIR model for TB disease transmission is developed. Then, the stochastic SLIR model is constructed. Next, the stochastic SLIR model is used to estimate the relative risk for the disease. Later, the performance of the stochastic SLIR model is compared with other existing models based on relative risk values. Finally, the TB risk maps are constructed. For numerical analysis, this study used a data set of Malaysia TB cases reported from 2008 to 2015. Findings show that there is a large difference of relative risk estimation values when using stochastic SLIR model compared to existing models. This is clearly visible through disease mapping as some locations change colour from low tone (low risk) to darker tone (higher risk). This is due to the inclusion of latent component in the stochastic SLIR model. As a conclusion, this study offers a better model in estimating relative risk for TB disease. The findings may assist the government in prioritizing locations which need further attention especially in terms of health policy and financial support.

Keywords: Disease mapping, Relative risk, SLIR model, Stochastic model, Tuberculosis

Acknowledgement

In the name of Allah, The Most Gracious, Most Merciful, all praise The Lord of the worlds, peace and blessings be upon the Prophet Muhammad, the entire family and all his companions. Thank God, be grateful to the Almighty on His grace, I complete this thesis. I also would like to express my gratitude to all those who have made it possible for me to complete this thesis.

My foremost thanks go to my both supervisors as respect, Dr. Nazrina Aziz and Associate Professor Dr. Maznah Mat Kasim for the guidance and insight in the realization of this research, and also on proofread numerous drafts, suggestions and opinions given. I thank them for all advices; for the time they have given to this research; and most of all, for making me confident with my work. Without them, this thesis would not have been possible.

Heartfelt appreciation also goes to my beloved parents, Mohd Diah Hamdan and Fairuz Hassan and also my family for their encouragement, understanding, devoted and loving support. My special gratitude to my eldest brother, Muhammad Fadhli for making time to proofread this thesis and special thank also to my friends, Aznida Che Awang, Sufi Hafawati and Nuraimi Ruslan for the help and support given since the first day in my master study journey.

I also thank the Ministry of Health, Malaysia, for its cooperation in providing data for this research that enabled me to do the analysis and subsequently write this thesis. Finally, financial support from the Ministry of Higher Education is gratefully acknowledged.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations and Mathematical Symbols

TB   Tuberculosis

EPT   Extrapulmonary tuberculosis

MOH   Ministry of Health

WHO   World Health Organization

RR   Relative Risk

SIR   Susceptible-Infected-Recovered

SLIR   Susceptible-Latently infected- Infected- Recovered

SMR   Standard Mortality/Morbidity Ratio

CI   Confidence Interval

$S_{i,j}$ Total number of susceptible persons for area i, at time j

$L_{i,j}$ Total number of latently infected persons for area i, at time j

$I_{i,j}$ Total number of infectious persons for area i, at time j

$R_{i,j}$ Total number of recovered persons for area i, at time j

$\bar{I}_{i,j}$ The number of new infectious persons for area i, at time j

$\Re_{i,j}$ The number of newly recovered persons for area i, at time j

π Recruitment rate people per year

r Per year human birth rate

μ Natural mortality rate (per year)

μ_T TB caused mortality rate (per person per year)

λ Force of infection (per year)

v Progression rate from latent to active TB (per person per year)

p Probability of new infections that develop progressive primary active TB

c TB cure rate (per person per year)

β_0 Overall rate of the process

b_i Random effect that absorbs residual spatial variation

t Time
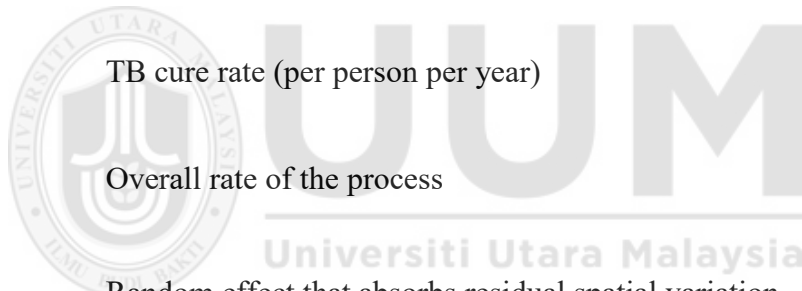
CHAPTER ONE
INTRODUCTION

1.1 What is Tuberculosis?

Tuberculosis or TB (abbreviation for tubercle bacillus) is a bacterial disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) organism which these slow-growing bacteria grow well in the area of the body that has a lot of blood and oxygen (Bhowmik, Chandira, & Pradesh, 2009). In the past, according to Kumar, Abbas, Fausto, and Mitchell (2007) TB was also called as consumption, phthisis or phthisis pulmonalis. Tuberculosis usually affects the lung (pulmonary TB or PTB), but also can affect any other part of the body, for example bones, kidneys, lymph nodes and brain as the infection can spread via blood from the lung which is called as extrapulmonary tuberculosis (EPT) (New York State Department of Health Tuberculosis (TB), 2007).

Konstantinos (2010) and Kethireddy (2010) stated that TB can be transmitted from a person to another through air. Tiny droplets released into the air when people with active TB infection sneeze, cough or spit. Even though the droplets dry out quickly, the bacteria can still remain airborne in the air for hours especially in small area with no fresh air.

The infection of TB can either be latent or active TB. When someone inhales air that containing *M. tuberculosis* that are expelled into the air by other person with infectious TB, that person will become infected. However, someone who had been infected with the bacteria, he or she does not necessarily become sick. This is

because that person can have either latent TB or TB disease (Hauck, Neese, Panchal, & El-Amin, 2009).TB disease occurs when a person who has latent TB becomes sick. Sometimes this is known as that person is having active TB or TB disease. Most of people with latent TB never become sick. According to Hauck et al. (2009), about 5% to 10% of the people who have latent TB, who do not receive treatment for it, will become sick at some time in their lives. People with active TB infection are caused by active duplicate tubercle bacilli. They easily spread the disease to others and show symptoms. While those who with latent infection have previously been infected but do not show any symptoms and are not contagious (Konstantinos, 2010). This means that people with latent TB cannot spread the disease to others. Latent TB is less serious as people with latent TB have fewer germs in their body and the germs are latent (sleeping) (Ai, Ruan, Liu & Zhang, 2016). However, during their lifetime, they remain at risk for active TB (Konstantinos, 2010).

General symptoms for those who have active infection include fever, cough with thick, cloudy and bloody mucus from the lungs for more than two weeks, chest pain, fatigue, weight loss, shortness breath and night sweats (Bhowmik et al., 2009). According to Konstantinos (2010), there are many factors related with an increased risk for being infected and further developing the disease such as migrants from high TB prevalence countries, healthcare workers, patients with HIV infection, diabetes mellitus, chronic renal failure, infancy and also people of older age. Bhowmik et al. (2009) stated that even though it takes a long process, treatment is often successful. TB treatment usually takes about six to nine months. However some TB infection cases should be treated for up to two years. A single antibiotic is usually used for latent TB treatment, while combinations of several antibiotics are

used to treat active TB disease. This can help to decrease the threat of the bacteria developing antibiotic resistance.

In summary, TB is one of the infectious diseases caused by *M. tuberculosis* bacterium which can only be transferred to others through air. Figure 1.1 shows the symptoms for someone being infected with TB. According to WHO (2014), human immunodeficiency syndrome (HIV) patients are high contributor to TB cases. It helps to speed up the spread of TB. In Africa, HIV becomes the single most important factor in determining the increase incidence of TB in the last 20 years (Mukhtar, Abdullah & Juliana, 2014).
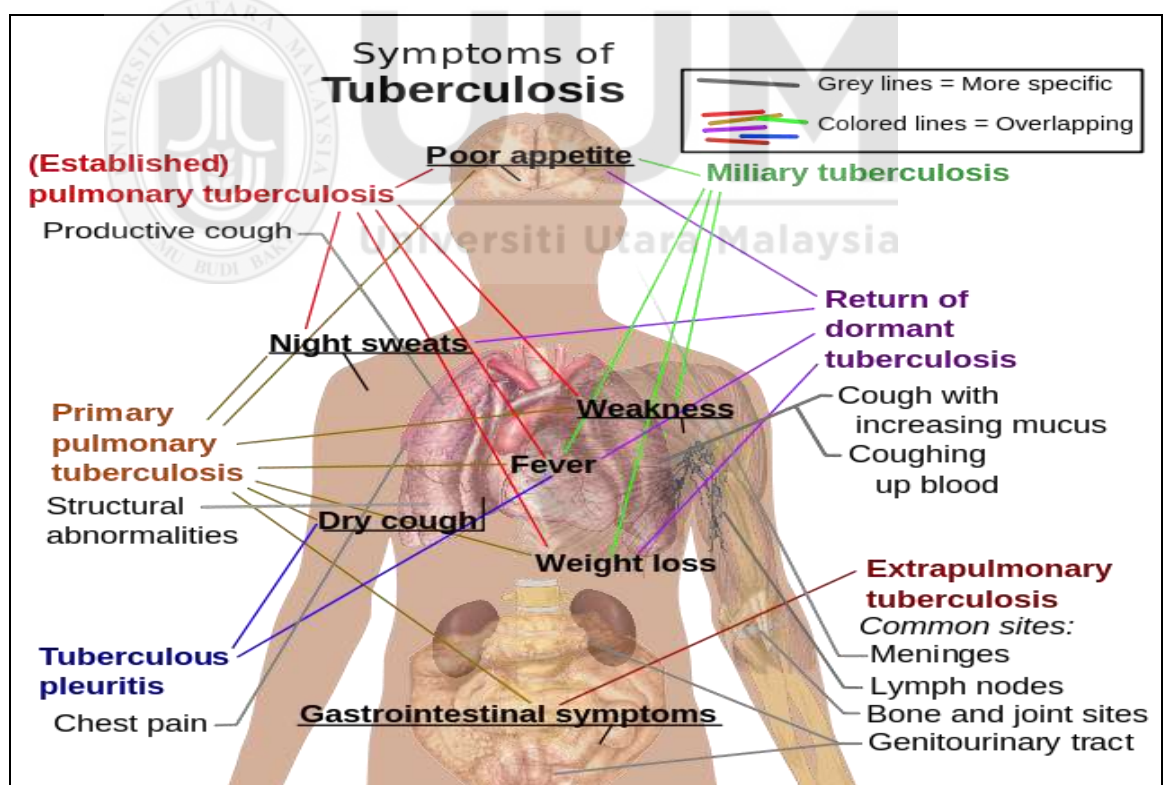
Figure 1.1. Symptoms of Tuberculosis (TB) (source: Tuberculosis [TB, Consumption], 2016)

1.1.1 Relationship between HIV, AIDS and Tuberculosis

Weakened immune system is taken advantage by TB which is called as an opportunistic infection. People who are HIV-positive will be at high risk to get TB. Globally, TB is the main cause of death for HIV patients (WHO, 2016). It is one of the top 10 causes of death worldwide and the leading cause of death from infectious disease after HIV/AIDS (WHO, 2016).

As there are two types of TB infection, latent and active, for those who are suffering HIV-positive, will be susceptible to the type of active TB, especially for those CD4 cells (a type of immune system cell) count is under 200. In comparison, a person with both HIV and TB has at least 10 times risk to gain active TB compares to the person with only TB (Narasimhan, Wood, Intyre, & Mathai, 2013). Basically, for those who are infected with both of these diseases will get acquired immunodeficiency syndrome (AIDS), a more advanced stage of HIV (U.S. Department of Health & Human Resources, 2016).

1.1.2 World TB Scenario
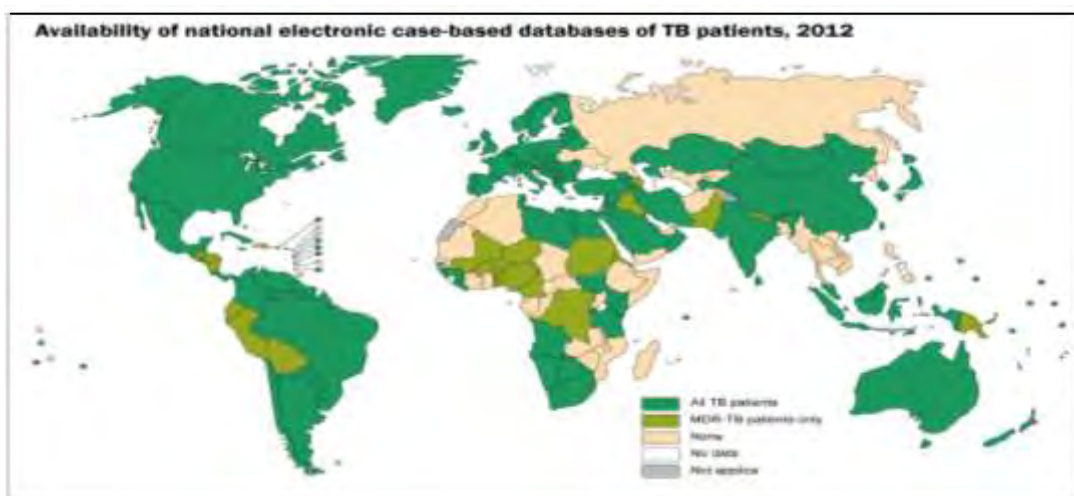
Figure 1.2. Tuberculosis world distribution map (source: WHO, 2013)

According to Global Tuberculosis Report 2013 by WHO, in 2012, about 8.6 million people have been estimated to develop TB and lead to the death of 1.3 million people whereby 320,000 of them are HIV positive. Almost 1.1 million people which is about 13% people who had TB in 2012 were HIV-positive. About 75% of the cases take place in the African region (refer to Figure 1.2). About 9 million people were infected with TB in 2013, including 1.1 million cases among people living with HIV. In 2013 the number of cases increased as 1.5 million fatalities caused by TB, including 360,000 among people who were HIV-positive. Globally in 2014, the number of TB cases reported have been increased to 9.6 million people and it is estimated that 12% of them were HIV positive. About 58% of the TB cases took place in South-East Asia and Western Pacific regions. There were an estimated 1.5 million deaths from TB, where 390,000 deaths among people who are HIV positive. In 2015, according to Global Tuberculosis Report 2016 by WHO, about 10.4 million new TB cases were estimated worldwide including 1.2 million (11%) with HIV positive. There were an estimated 1.8 million people died from TB, where 400,000 deaths among people with HIV positive. This shows that over 4,900 people died because of TB every day. African region shows the highest number of countries with the proportion of TB cases co-infected with HIV and exceed 50% in parts of southern Africa (refer to Figure 1.3). TB ranks along with HIV as a major cause of death worldwide (WHO, 2015).
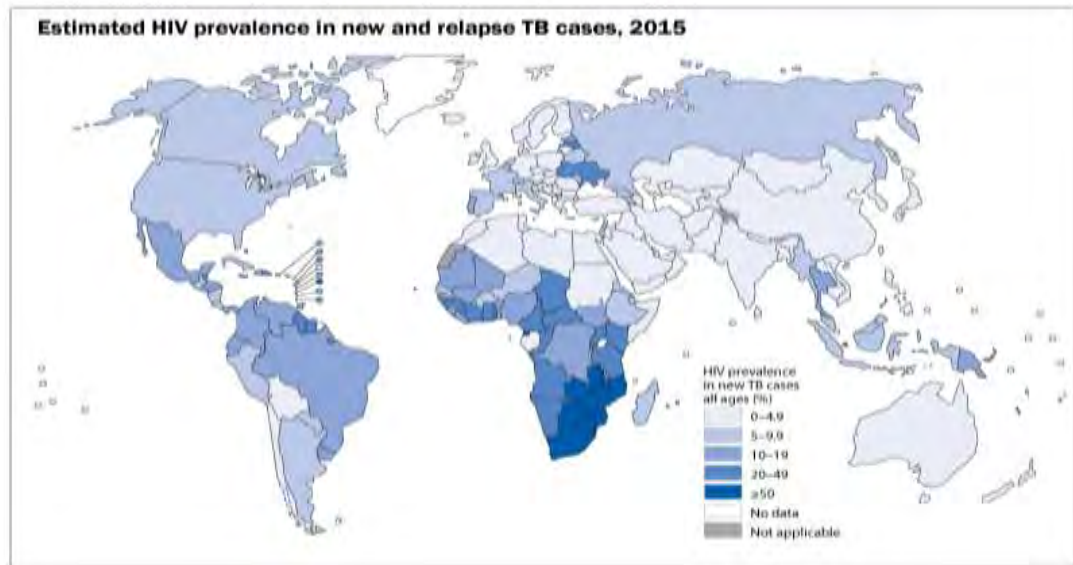
Figure 1.3. Estimated HIV prevalence in new and relapse TB cases (source: WHO, 2016)

Although men are shown to be the most affected by this disease, TB also ranked the third commonest cause of death among women worldwide. In 2012, a total of 3.5 million women had died of TB by which about 40% of them were HIV positive. Among the HIV-positive people who died in 2012, half of them are women. In the year 2013, TB has become one of the most important causes of death among women in reproductive age. Women also constitute 50% of the total TB death among HIV-positive people. It was reported that from 510,000 women who died due to TB, 180,000 of them were HIV-positive. Among children, TB is responsible for the death of 80,000 people who were HIV-negative and 550,000 others becoming ill with TB. However, it was observed that since the past 15 years, the mortality rates had decreased by 45% (WHO, 2013). Based on Global Tuberculosis Report 2016 by WHO (2016), in 2015, men were the highest group that contact with TB with 5.9 million (56%). This is followed by women with 3.5 million (34%) of them and 10% were children which are about 1.0 million. There were an estimated 0.14 million deaths from TB among HIV-positive women.
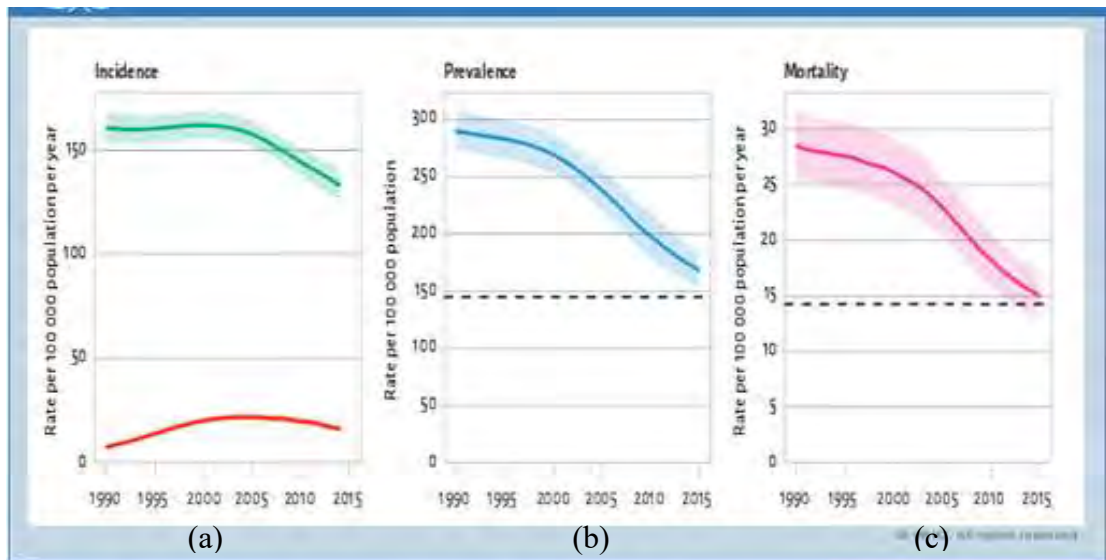
Figure 1.4. Global trends in estimated rates of TB incidence, prevalence and mortality. (source: WHO, 2015)

Figure 1.4 shows the global trends in estimated rates of TB incidence, prevalence and mortality. Graph (a) is the graph of estimated incidence rate including HIV-positive TB (green) and estimated incidence rate of HIV-positive TB (red). Meanwhile, graph (b) illustrates the trends in estimated prevalence of TB and graph (c) presents the mortality rate from 1990 to 2012 and forecast TB prevalence and mortality rates for year 2013 until 2015 respectively. The horizontal dashed lines represent the Stop TB Partnership targets of 50% reduction in prevalence and mortality rates by 2015 compared with 1990. The shaded areas represent the uncertainty bands. The mortality chart does not include TB deaths among HIV-positive people.

Stop TB Partnership is a unique international body with the power to align actors all over the world in the fight against TB. It is established in 2001. The Stop TB Partnership unite the international and technical organizations, governmental and non-governmental organizations, research and funding agencies, foundations, civil

society, community groups and also the private sector that are a collective force that transforming the fight against TB in more than 100 countries. This Stop TB Partnership has been replaced with the End TB Strategy that covers the period of 2016 until 2035.

According to Global Tuberculosis Report 2016 by WHO (2016), in 2015, it has been estimated about 10.4 million of new (incidence) TB cases worldwide. This shows that about 28, 500 people fell ill due to TB every day. TB disease remained one of the top 10 causes of death worldwide in 2015 despite a drop of number of TB deaths between year 2000 and 2015 (WHO, 2016). Based on report by WHO (2016), there are six countries that have been identified to have high TB new cases reported which is accounted for 60% of the new cases: China, India, Indonesia, Pakistan, South Africa and Nigeria.

1.1.3 Tuberculosis Scenario in Malaysia

In Malaysia, TB has been emerged as the number one killer in Malaysia in early 1940s and 1950s (Kuppusamy, 2004). According to Kuppusamy (2004), at that time, TB patients were admitted to many sanatoriums and were often managed by surgery. In the late 1950s, TB chemotherapy became available and at that time, TB already became a major cause of death (Kuppusamy, 2004). In 2001, it was the second highest communicable disease in Malaysia (Jiloris, Jamaliah & Yap, 2004). Until now, in Malaysia, TB still becomes major public health significance as stated by WHO (2014) with yearly, 80 new cases per 100,000 population in Malaysia were affected by TB.

Table1.1 Number of TB Cases in Malaysia

Year	No. of Cases	% of Change
2003	15,853	
2004	15,429	-2.67
2005	15,991	3.64
2006	16,665	4.21
2007	16,918	1.52
2008	17,496	3.42
2009	18,102	3.46
2010	19,337	6.82
2011	20,666	6.87
2012	22,710	9.89
2013	24,070	5.99
2014	24,711	2.66
2015	24,220	-1.99

(source: Ministry of Health Malaysia, 2012-2016)
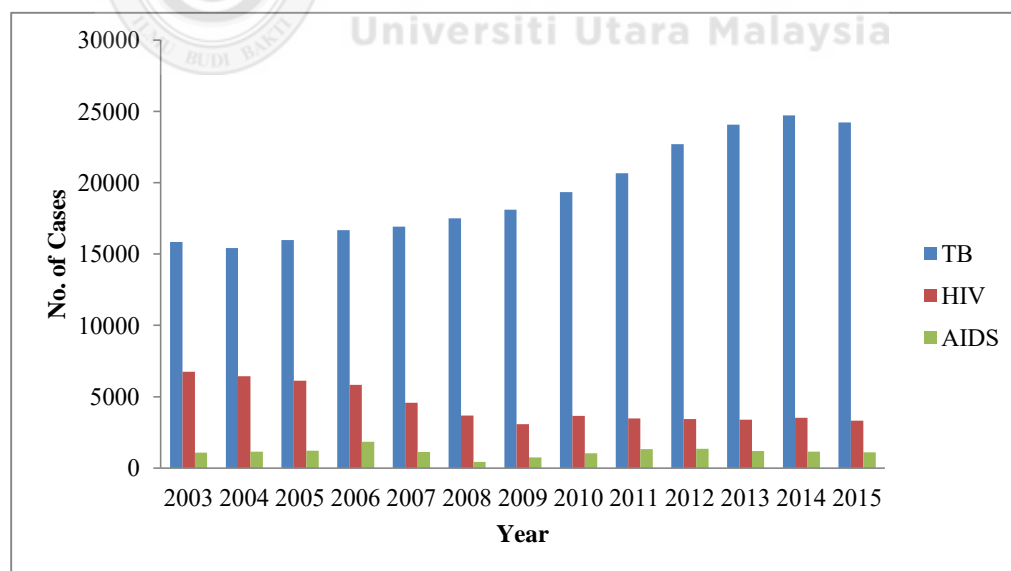
Figure 1.5. Number of TB, HIV and AIDS cases in Malaysia (source: Ministry of Health Malaysia, 2012-2016)

According to Ministry of Health Malaysia (2002), the numbers of TB and HIV co-infection reported have increased from six cases in the year 1990 to 933 cases in 2002. Based on Figure 1.5, the Ministry of Health Malaysia (2012-2016) stated that the number of TB cases had increased for a total of 8,367 cases from the year 2003 until year 2015 which is about 52.77 %. In 2011 to 2012, it is recorded the highest increase in the number of cases about 9%. In 2013, the number of cases notified is 24,070 and in 2014, it increases to 24,711. A total of 640 new cases from the previous year which is an increase of about 2.66 % (refer to Table 1.1). However, the number of TB cases drop to 24,220 in year 2015 with -1.99% drop from the previous year. This number tallies with the number of HIV cases which also dropped in year 2015 (refer to Figure 1.5). Figure 1.5 shows that the numbers of TB cases in Malaysia keep on increasing from year to year. Table 1.2 shows that in 2015, the number of TB patients with known HIV status is 23,094 which is 98% from the total number of TB patients.

Table 1.2

Number of HIV-positive TB Cases Reported in Malaysia 2015

TB cases notification, 2015	Number	(%)
Total cases notified	24,220	
Total new and relapse	23,565	
-% tested with rapid diagnostics at time of diagnosis		2
-% with known HIV status		98
-% pulmonary		87
-% bacteriologically confirmed among pulmonary		75

(source: WHO, 2017)

Table 1.3

The Number of Deaths due to TB Cases in Malaysia

Year	No. of Cases	% of Change
2003	1,029	
2004	1,318	28.09
2005	1,437	9.03
2006	1,431	-0.42
2007	1,504	5.10
2008	1,523	1.26
2009	1,582	3.87
2010	1,557	-1.58
2011	1,644	5.59
2012	1,414	-13.99
2013	1,597	12.94
2014	1,603	0.38
2015	1,696	5.80

(source: Ministry of Health Malaysia, 2012-2016)
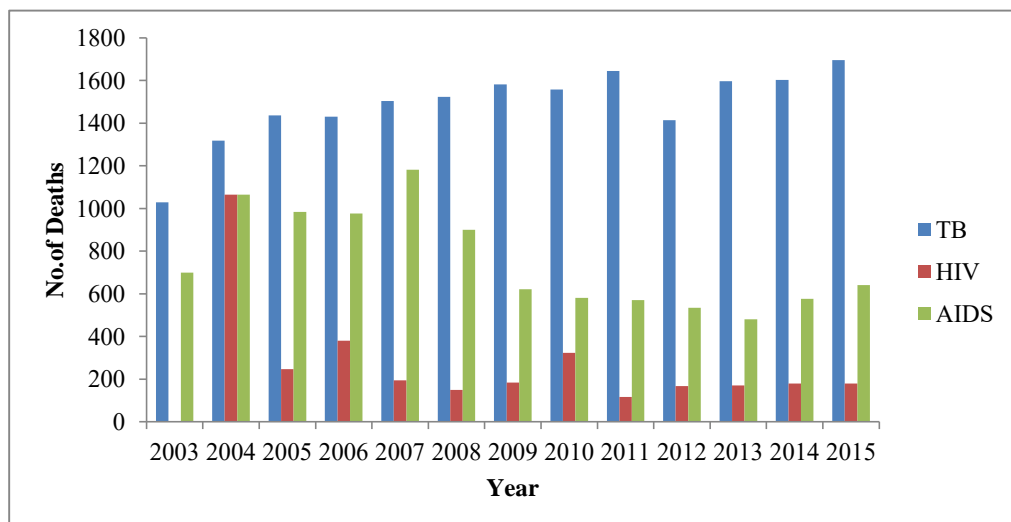
Figure 1.6. The number of deaths due to TB, HIV and AIDS cases in Malaysia
(source: Ministry of Health Malaysia, 2012-2016)

It is necessary to look into this problem as TB is a public health challenge in Malaysia with infection constantly increasing reported annually (Ministry of Health Malaysia, 2015) and has been identified as the most cause of death among all infectious diseases in Malaysia (Loh et al., 2017). Based on Figure 1.6, the number of deaths due to TB disease in Malaysia is still high despite a drop from 1,644 deaths in 2011 to 1,414 deaths in 2012. The reduction in the number of death is tallies with the HIV cases which also dropped in 2012 (refer to Figure 1.6). However, the number of deaths due to TB infectious increased to 1,597 in 2013 and 1,603 in 2014. This represents an increase of 0.38 % (refer to Table 1.4) or six cases. In year 2015, the number of deaths caused by TB increased to 5.8% from the previous year with 1,696 cases reported.

The rate of TB disease was reported high among adults rather than children (WHO, 2013). This is due to the vaccine called BCG (Bacille Calmette-Guerin) that is only effective in protecting children from this disease. For adults, the vaccine is not effective as it could not protect them well against lung TB. (National Institute of Allergy and Infectious Disease [NIAID], 2012).

Even though there is a treatment for TB, we need a strategy to control the problems of TB. This can be done through disease mapping because it can identify the hazard areas for disease. Once we identify the areas, the government can focus on these areas and provide better facilities and services. Prevention programs can also be implemented efficiently.

Disease mapping has been identified as a significant public health instrument, which may be used to identify high risk areas and generate etiologic hypotheses in the prevention and control strategies for disease (Nor Azah & Syafiqah Husna, 2013). Good disease mapping plays very important role in monitoring the health of community. It can provide detailed knowledge of local communities' health and this would allow government to provide the quality and quantity of resources appropriate to the local needs (Lawson & Williams, 2001). Good disease maps can also help public health specialists to identify the cause of the infection by monitoring the spread of the infectious disease. In order to get a good disease mapping, it is very important to get good estimation of relative risk. Normally, good relative risk can be obtained through a right statistical model (Nor Azah & Percy, 2012).

1.2 Research Background

Disease mapping is one of the methods used to understand the geographic distribution of a disease within a population (Lawson & Williams, 2001). Lawson and Williams also stated that in disease mapping, the mapping refers to the geographical distribution shown in visual form. Disease mapping has various names, some of which are: environmental epidemiology, small health studies and spatial epidemiology (Lawson, 2013).

According to Lawson (2013), there are two important characteristics of disease mapping; spatial or geographical distribution and disease distribution. In mapping the disease, the relative location of events is crucial, hence geographical information system is needed. Another focus is the disease distribution. The main issue here is how to analyze the disease incidence or prevalence when we have

13

geographic information. This is sometimes called as geo-referenced disease data, by labeling the outcome with spatial tags. Disease mapping can show up geographic areas with high and low incidence or mortality rates for specific diseases (Jin, Banerjee, and Carlin, 2007). According to Koch (2005), disease mapping for incidence and prevalence had long been used in public health, epidemiology and the study of disease in human populations.

To place epidemiological studies in context, it is crucial to estimate the background variability in underlying risk, simple explanation, hypothesis generation and allotment of health care resources. There are four main aims for using disease mapping (Shaddick, 2008). Maps are usually drawn for one of two reasons either to help to analyze the data in order to know which areas have a high or low incidence for the disease or for illustration purpose so that main conclusions of analysis can be drawn for others benefits. The purpose of the analysis in disease mapping is to estimate the true relative risk of a disease of interest throughout the geographical area of study and also to reduce the noise in a disease map. Hence, in this study, disease mapping is the focus of attention.

Among the first studies of disease mapping in epidemiology is that by John Snow in 1854 (Lawson &Williams, 2001). It was the map of the cholera victims that addresses links to the water supply locations in London. In this case, every street address of the victim of cholera has been recorded. Snow (1854) was the first person to clearly show that cholera could spread through contaminated water supply. He used dot maps to show the residences of the victims. In Figure 1.7, his ‗dot maps‘ of

cholera patients residence in the Golden Square, London that show different group of cases around the water pump in Broad Street.
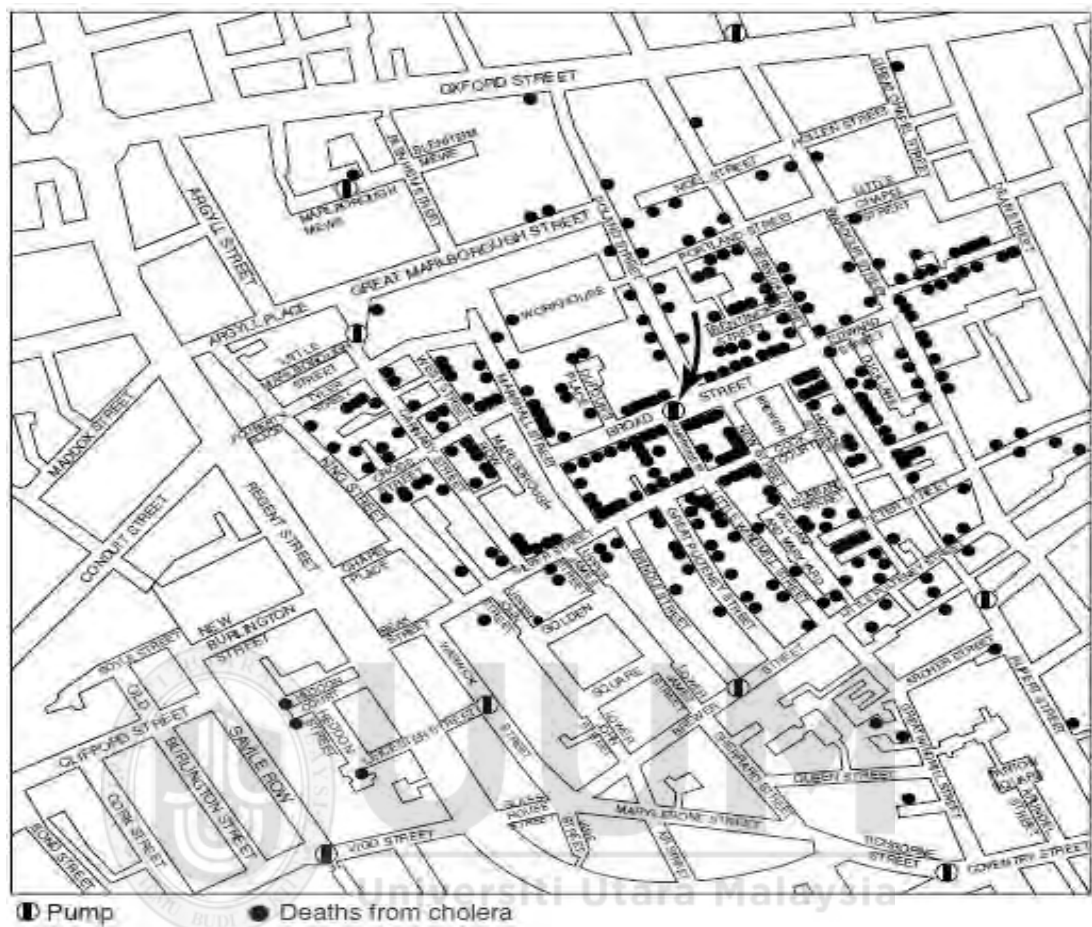
Figure 1.7. Dot map of deaths from cholera in London (the arrow points to the Broad Street Pump). Redrawn from Snow (1936) by permission of Oxford University Press. (source: Lawson & Williams, 2001)

The study had helped preventing the cholera epidemics in the United Kingdom since it can identify pipes which had become contaminated by faecal material from cases of cholera. This type of map had been used by others researchers such as Deneke (1895) in studying the cholera cases in the city of Hamburg and in the adjoining suburb of Altona. Holden (1880, as cited in Lawson & Williams, 2001) also used ‗dot maps' to demonstrate the value of disease mapping. His study recognized that the absence of sewage systems is more significant compared to the presence of unavoidable topographical features such as watercourses and elevation

15

above the sea level in typhoid disease. According to Lawson and Williams (2001), Lilienfeld and Lilienfeld (1981) used this type of map in a study of typhus distribution in Montgomery, Alabama for the cases reported in 1922-1950. Nowadays, this type of map is not preferred by researchers since the visual impression is influenced by the dot size. The dots may not be seen if the sizes are too small and they will overlap if the sizes are too large. The visual impression of dot maps is often misleading because it does not take account of differences in the size of the population at risk (Lawson & Williams, 2001).

According to Lawson and Williams (2001), in twentieth century, disease mapping became more common. The first map of infectious and non-infectious disease in England and Wales was produced by Stock (1936, 1937, 1939, as cited in Lawson & Williams, 2001). He standardized the map for differences in sex, age and urbanization. Stock (1939, as cited in Lawson & Williams, 2001) also introduced maps using the Standardized Mortality Ratio (SMR) for cancer. These maps are drawn using ratio data for areas. This type of map is called choropleth maps. The construction method is quite simple. Each shaded area is in accordance with the data value. Areas with the highest values usually will be filled with the darkest tones. Figure 1.8 shows the example of early use of this kind of map that was constructed for mortality data. The data set was given in the Medical Tables and Text Volumes of the Annual Statistical reviews of the Registrar General.
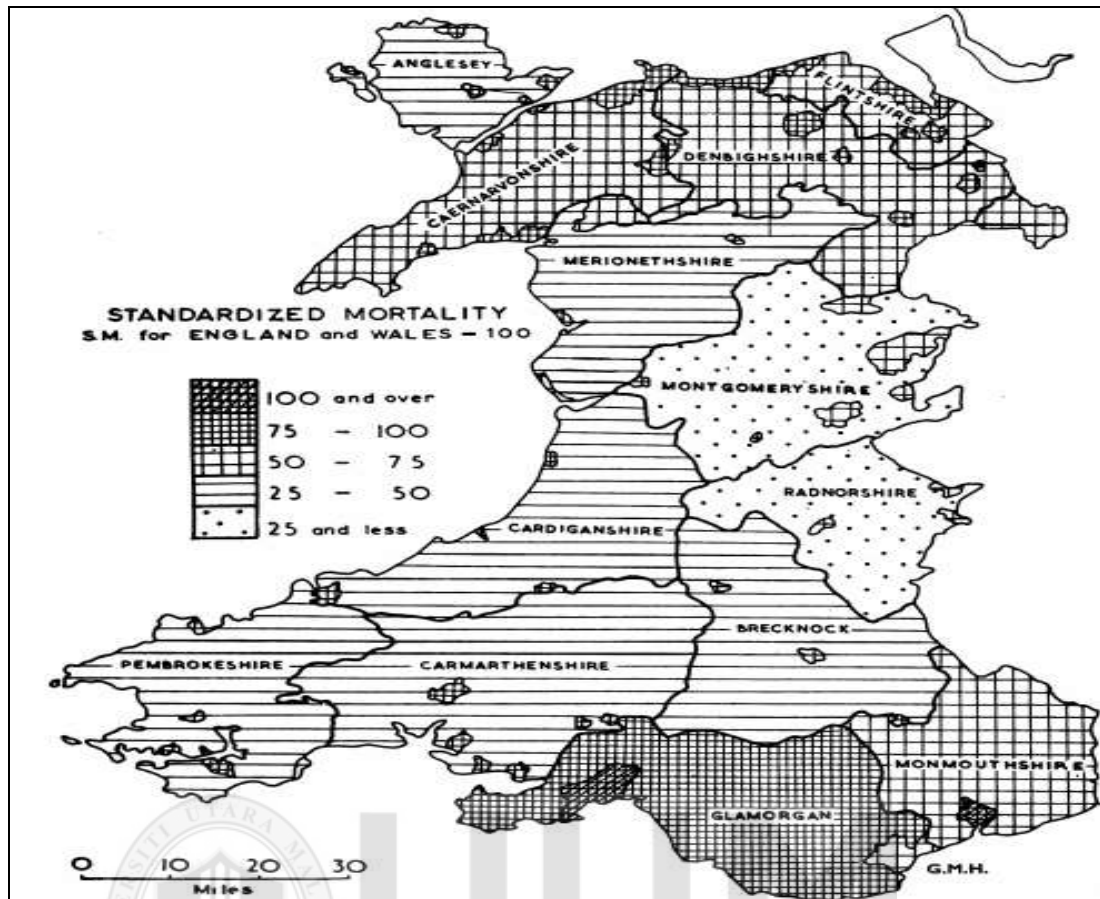
Figure 1.8. Cancer of lung and bronchus, Standardized Mortality for males, 1947 –
1953 (source: Howe, 1959)

This study focuses on this type of map since this study wants to know the
relative risk of TB. The map appearance for choropleth map is affected by various
factors such as number of classes, tones and class intervals. Robinson and
Sale (1969) stated that for the number of classes, it is best to have five or six shading
categories. If the number of categories is fewer, the information may be lost. This
can cause the map to be over-generalized. But if the number of categories is too
large, it will be difficult to interpret the data as there are many information.

The tones for choropleth maps used to represent the different values
(Robinson & Sale, 1969). This obviously can give an impact to the appearance of the
map. It is advisable to use clear distinguishable tones and be spaced evenly from

light to dark. The colour used should form an obvious sequence and rather than using multiple colours it is better to use different shades of one or two colours. Figure 1.9 shows the example of map using shades of the same colour. This example of map shows the high-low risk areas based on SMR method for dengue disease. The class intervals as mentioned earlier also influence the appearance of maps. The class interval refers to the cut-off points. According to Robinson and Sale (1969), the choice of class interval will give different visual impressions even on the same data values. The method used to determine the cut-off points between the classes determines whether the top category contains one area or more. This can lead to identify which area is of high mortality or low mortality based on the light or dark shade region.
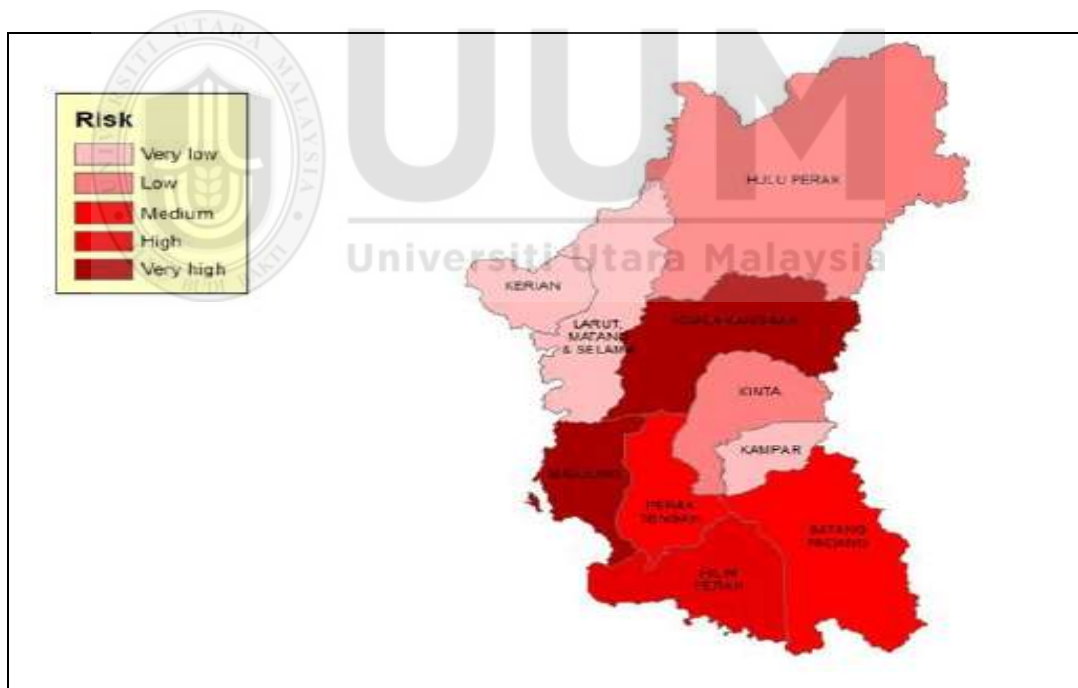
Figure 1.9. Disease map of estimated relative risks based on SMR method
(source:Nor Azah & Syafiqah Husna, 2013)

There are several ways in choosing the class interval. Robinson and Sale (1969) stated that the simplest is to use equal divisions of the range. This

method helps to give good impression of the underlying statistical distribution. Some may use quantiles to put an equal number of areas into each shading category.

In order to estimate the status of an area with respect to the occurrence of the disease, it is easier to evaluate what disease occurrence should be locally 'expected' in the area and then compare the observed incidence to the 'expected' events. This method has been conventionally practiced to analyze the tract-count data and also can be applied to case-event data. Case-event data are typically found as the street address of a case of the disease recorded because it usually took place within a fixed time period and also known as a continuous data while tract-count data correspond to observations which consist of aggregated count cases within the tracts or small study areas over specified periods of time (Lawson, 2006). Tract-count data are also known as discrete data. The specific disease has been retrieved within the small areas. According to Lawson et al. (2003), the ratio observed to expected counts within tracts is known as Standardized Mortality or Morbidity Ratio (SMR). This ratio is relative risk estimation within each tract. It is the simple ratio of the observed count within a tract to the expected count based on the _at risk' background or population. Poisson-gamma model is another method that used by many researchers in finding the relative risk estimation (Lawson et al., 2003). In this Poisson-gamma model, Poisson distribution is used because this is the basic model for count data.

Nor Azah and Percy (2012) stated that the main issue when studying geographical distributions of disease occurrence is the relative risk estimation. The disease map's production depends on the modeling to predict and estimate the risks. Therefore, better predictions and risk estimations would produce more precise

disease risk maps. A relative risk is the risk of an event (or of developing a disease) relative to exposure. A relative risk was also defined as a ratio of the probability of the event occurring in the exposed group over a non-exposed group (Larssen, Desai, Anahita & Shortell, 2013). It measures the strength of association between an exposure and a disease.

$$Relative\ risk = \frac{probability\ of\ event\ when\ exposed}{probability\ of\ event\ when\ non\text{-}exposed}$$

The current statistical method used by Malaysian government to identify hot spots or high-low risk areas are still based on the total numbers of TB occurrences across the regions. For example, large numbers of TB cases in certain areas represent high risks of TB occurrences without considering other factors like the population size or the land size of individual areas. District Hospital and District Health Clinic are responsible to provide this data for analysis. Preferably, the patient morbidity data should be obtained for a period of at least a year. The number of cases reported also can be used to estimate disease risk for TB disease. When the number of cases reported in a certain area is high, then the area is spotted as high risk for TB. Ministry of Health Malaysia used this method to know the health status of the Malaysian's community. However, this approach has disadvantage because it does not take into account other factors.

When marking the area as low or high area based on the number of cases occurred only without considering other factors such as the population size and the land size, the risk estimation of the disease may not be accurate. For example, if the number of cases recorded in Perlis is 15 people while in Kedah is 10 people, it is not

correct to mark Perlis as high area since the population size in Perlis is less than the number of population in Kedah, and the region area of Perlis is smaller compared to Kedah.

Therefore, this study is interested to propose an alternative method to predict and estimate the high-low risk areas based on transmission models for TB disease which takes into account significant covariates in order to prevent and control TB disease.

1.3 Problem Statements

TB is a widespread disease which is one of several factors that causes millions of death worldwide. The numbers of TB cases reported keep on increasing from year to year. Among the factors that lead to the spread of TB disease widely are poor sanitation, crowded conditions and lack of access to medical treatment. Moreover, this type of disease can become epidemic and also pandemic if it is not controlled (Yoneyama, 2010). Epidemic is a local spread of infectious disease within a region or a country, while pandemic is an infectious disease which spread worldwide. Hence, there is a need to monitor on this type of disease before it becomes more serious as it can spread easily to the others. Daley and Gani (2005) stated that modeling the infectious disease is a tool that able to predict future course of a disease outbreak and strategy the control of epidemics.

In Malaysia, the current method use to estimate the high-low areas is still using traditional approach by monitoring based on the total number of TB cases for each region that have been reported. This method only shows general information

without considers the area of disease transmitted. Before the area is mark as high or low risk area for disease transmission, the number of people in population, gender and age need to be considered. According to Nor Azah and Percy (2012), disease mapping can be used in controlling and as the prevention strategies for a disease. It relies on the modeling used to estimate the relative risk to get a good disease map. In Malaysia, there have been works done on dengue disease (Nor Azah & Percy, 2012), HIV and AIDS (Sufi Hafawati, 2016) and Leptospirosis (Aznida, 2017), but there is no work has been done on TB which has latent period in disease transmission process.

Disease mapping can help to show a clear picture of the risk area. However it relies on the good value of relative risk estimation to get a good disease mapping. Relative risk estimation for disease mapping is still an ongoing study as it is an important issue that needs to be considered when investigating geographical distribution (Lawson et al., 2003). Since relative risk estimation for disease mapping is important as it can help to detect the critical areas of disease, there is a need to study the best method. In the study of TB disease, SMR has been used as the conventional way to estimate relative risk. Since direct use of SMR may not be worthy as it cannot detect the small areas (Meza, 2003; Lawson et al., 2003), as it does not take into consideration the high diversity of different regions and the spatial patterns of the areas under study. SMR is based on ratio estimators. Hence, the use of SMR can yield large changes in the estimate, and relatively small changes in the expected value (Lawson et al., 2003). Bayesians approach is highly recommended in the use of small area estimation as it smooths the relative risk and provides the measures of uncertainty associated with this relative risk estimation and the modeling

can take into account the spatial autocorrelation. The approach to smoothing in Bayesian approach is by borrowing strength values from geographically referenced neighboring values. Poisson-gamma model is one of the earliest Bayesian approach used by many researchers (Lawson et al., 2013). However, there are still drawbacks of using this method as it cannot allow covariate for adjacent areas and it does not consider the disease transmission. Moreover, SMR and Poisson-gamma model do not consider latent component in the model.

In order to overcome the drawbacks from SMR and Poisson-gamma model, researchers used transmission model or mathematical modeling. However, most of the mathematical modeling especially for TB disease use deterministic model. As mentioned earlier, even though this type of model is much preferred by many researchers, it does not take into account any possible random effects. In simplest form, a random effect refers to an extra quantity of variance component which is estimable within the map and which can be ascribed a defined probabilistic structure. Hence in this study, stochastic model for TB disease transmission was proposed which this model considers the covariates that are significant for TB disease and also include the latent component in the model.

1.4 Research Objectives

This study embarks on the following objectives:

i. To develop the stochastic SLIR models for TB disease transmission in Malaysia.

ii. To compute relative risk estimation for TB disease mapping based on stochastic SLIR models.

iii. To evaluate and compare the performance of relative risk estimation based on stochastic SLIR models of TB disease transmission with other existing methods that are SMR and Poisson-gamma model for disease mapping using TB data from Malaysia.

1.5 Research Questions

The research questions of this research are as follow:

i. How to develop the stochastic SLIR models for TB disease transmission in Malaysia?

ii. How to compute relative risk estimation for TB disease mapping based on stochastic SLIR models?

iii. Are there any differences in results of relative risk estimation based on stochastic SLIR models of tuberculosis disease transmission with other existing methods that are SMR and Poisson-gamma model for disease mapping using tuberculosis data from Malaysia?

1.6 Significance of the Study

This study produce disease maps by using new alternative method proposed. The maps are good references to assist the government in the way to monitor the number of TB disease cases that was reported in Malaysia. Based on the maps, states that need further attention in terms of government policy and financial support can be identified.

Besides that, this study also proposes a better model in estimating risk for TB disease which could overcome the drawbacks of Standard Mortality Ratio (SMR)

and Poisson-gamma models. Stochastic terms are introduced into the SLIR disease transmission model. This stochastic SLIR model considered the transmission mechanism of the disease, besides it allows for spatial correlation between risks in nearby regions and it enables covariate adjustments.

Moreover, previous studies do not take into account the latent element in computing the relative risk. Most of the studies just considered the infectious element only. Hence, this study added value for latent element as stochastic SLIR model takes into account the infectious and latent components. This proposed stochastic SLIR model is able to handle latent cases. Furthermore, the disease maps produced by the proposed model can help government or specifically the Ministry of Health in identifying areas need more attention in order to provide proper medical treatments.
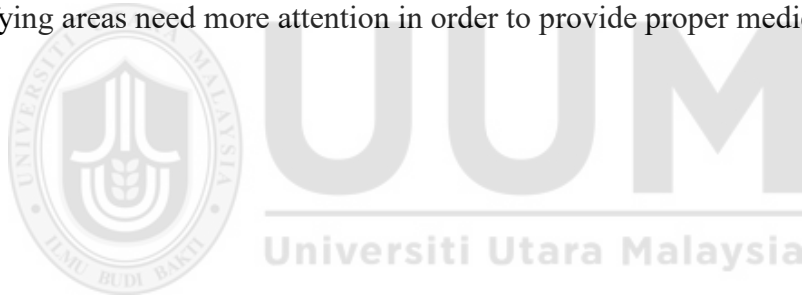
CHAPTER TWO
LITERATURE REVIEW

This chapter explains and discusses previous studies related to the modeling and analysis of disease transmission mechanisms for disease mapping. This chapter starts with a discussion about statistical models for TB, followed by disease mapping analysis in section 2.2. In order to understand more on the transmission of disease, discussion on basic concept of mathematical modeling for the infectious diseases are explained in section 2.3. This is followed by simple SIR model and detail on stochastic SIR model proposed by Lawson (2006) in section 2.4 and section 2.5, respectively.

2.1 Statistical Model for TB

Many cohort of studies used SMR model in order to estimate the relative risk of disease. Symons and Taulbee (1980) used SMR model as an approximation to relative risk. While, Kolappan, Subramani, Karunakaran and Narayanan (2006) used SMR model to calculate the mortality rate and excess general mortality among cohort of TB patients in India. However, according to Jones and Swerdlow (1998), there may be bias when using SMR model as it can cause substantial underestimation of the true relative risk when either the SMR or the prevalence of exposure in the general population are large.

Poisson-gamma model is then introduced as one of Bayesian methods that often used to estimate relative risk in disease mapping (Abdul Karim & Amoako, 2016). In the study by Abdul Karim and Amoako (2016), they applied

Poisson-gamma model (non-spatial model) and other suitable spatial models such as Conditional Autoregressive (CAR) model in order to identify which best fitting model for modeling and mapping TB relative risk in Kenya. Based on their findings, Poisson-gamma model gave good information about the TB prevalence in Kenya. However, it cannot corporate with covariate and unable to handle spatial correlation problem.

TB has also been analyzed by mathematical models from many aspects. One of the aspects that many researchers studied on is the dynamics of TB transmission among the individual people in a population. According to Hethcote (2007), dynamic models which describe change over time have been used for the spread of infectious disease. Dynamic model is defined as a simplification of reality in which it has been used throughout its long history in epidemiology and has been used in a variety of diseases. TB statistical models can either be deterministic or stochastic (Achterberg, 2009). States of individuals have to be defined within a population, such as by categorizing individuals into subpopulation groups based on criteria such as ‗immune to' or ‗infected with' TB.

In the 1960s, the earliest simple TB models were proposed by Waaler, Geser and Andersen (1962), Brogger (1967) and ReVelle (1967). In the study of Waaler et al. (1962), they built a linear model for TB which is based on three compartments: infected non-cases, infectious cases and susceptible. ReVelle (1967) used differential equations to formulate the relationship between TB prevalence and the rate of infection in his model. Most of the earlier proposed models focused on gaining a better understanding of TB dynamics and to forecast the effects of particular interventions by finding numerical solutions. The researchers did not look

at the long term behavior of TB as shown by the models, or qualitative analysis of the models and results. There were very few works on theoretical tuberculosis done from the 1970s until the early 1990s although the early research on TB models dated back in the 1960s. This situation may be due to the decrease of TB prevalence in developed nations and the increase of successful dependence on antibiotic treatment in the long term. However, in mid-1990s to present, there is an increase in TB modeling works due to the evolution of drug-resistant TB strains and co-infection with HIV/ AIDS.

Blower, Mclean, Porco, Small, Hopewell, Sanchez and Moss (1995) introduced the simplest Susceptible – Infected – Recovered (SIR) model for TB endemic. Endemic refers to the continuous presence of common diseases or infectious agent in a population within a geographic region. For example, the flu is endemic in Chicago winter but not in the summer. In their study, they developed a theoretical SIR based on age-implicit transmission models in order to determine common behavior of TB endemic. Corresponding to the study, Blower, Small, and Hopewell (1996) built more mathematically complex models, Susceptible – Exposed – Infected – Recovered (SEIR) model in which the infectious class of I is divided into two classes: I_I which is defined as infected and infectious individuals, and I_N refers to infected and non-infectious individuals. They extended the SIR models that include the population-level effects of chemoprophylaxis and treatment. In their study, it contained five non-linear ordinary differential equations. However, their models did not take into account the contribution of exogenous re-infection to the overall disease incidence. Vynnycky and Fine (1997) modelled an age-structured deterministic TB transmission dynamic model which has some unique aspects.

Dynamic approach is one of those and it can estimate re-infection contribution to age-specific TB incidence. They used Uninfected – Infected – Reinfected (UII $_r$) model. Their model has overcome the drawback from the previous Blower et al. (1996)'s model.

Porco and Blower (1998) used formulated transmission models of TB that is SEIR model from previous study to quantitatively understand the transmission dynamics of untreated TB epidemic in the absence of treatment. They applied time-dependent uncertainty and sensitivity analyses to the SEIR model. In their study, they also focus on identifying key parameters that contribute to the variability in the epidemiology results rather than to fit predictions to specific data.

Aparicio, Capurro, and Castillo (2000) formulated the deterministic cluster model using SEIR model. The model particularly explored the effect of intense and long exposure to persons with active TB on population level transmission dynamics. This model was in contrast to the model by Porco and Blower (1998) as it does not consider average number of individuals infected per year from one infectious case. In their study, the model differentiates between epidemiological active clusters and casual infections.

2.1.1 Modeling and Analysis of Disease Transmission

Disease transmission is divided into two types that are direct and indirect. There are two forms of direct disease transmission: direct contact and droplet spread and three forms of indirect that are airborne, vehicle borne and vectors borne (living organism) (U.S. Department of Health & Human Services, 2012).

Physical contact of an infected victim and susceptible person is called direct contact transmission whether transmission is through blood or body fluids such as saliva (‟Direct and Indirect Disease Transmission", 2011). Direct droplet transmission is a disease transmission caused by infected droplets that spread through the air contacting surfaces of nose, mouth or eye. U.S. Department of Health and Human Services (2012) stated that this type of direct transmission can occur when the infected person is in close proximity to another person when coughing or sneezing especially in an enclosed place. Droplets are relatively larger than 5 microns in size and projected up to about one meter distance. This is slightly different from indirect transmission.

Airborne transmission is one form of indirect disease transmission. Airborne transmission happens when dust or droplet nuclei suspended in the air carry infectious agents (U.S. Department of Health and Human Services, 2012). Airborne dust comprises material that has settled on surfaces and become re-suspended by air currents as well as infectious particles blown from the soil by the wind. It also can be spread via ventilation systems. Droplet nuclei are dried residue of less than 5 microns in size. In contrast to droplets that fall to the ground within a few feet, droplet nuclei may remain suspended in the air for long periods of time and may be blown over great distances. According to Tellier (2009), these solid or liquid particles will remain in the air for a long period of time because of their low settling velocity. This type of transmission can occur when the infected person is in close proximity to another person when coughing or sneezing especially in an enclosed place. Measles, for example, may occur in children who came into a doctor‛s office after a child with

measles had left, because of the measles. However, for TB disease, it has been classified as direct airborne transmission since it transferred from a person to another person (―Direct and Indirect Disease Transmission", 2011).

Vehicle borne are another type of indirect disease transmission. According to U.S. Department of Health and Human Services (2012), vehicle borne refers to type of disease that may indirectly transmit an infectious agent include water, food, biologic products (blood) and fomites (inanimate objects such as bedding, surgical scalpels and towels). A pathogen may passively be carried by a vehicle borne, for example hepatitis A virus may be carried by water or food. Alternatively, the vehicle borne may provide a setting in which the agent grows, multiplies or produces toxin as improperly canned foods provide an environment that supports production of botulinum toxin by *Clostridium botulinum*.

Another type of indirect disease transmission is a vector borne. For vector borne, an infectious agent may be carried by vectors such as mosquitoes, fleas, and ticks through purely mechanical means or may support growth or variations in the agent. Examples of mechanical transmission are fleas carrying *Yersinia pestis*, the causative agent of plague, in their gut and flies carrying *Shigella* on their appendages (U.S. Department of Health and Human Services, 2012). In contrast, in biologic transmission, the causative agent of malaria or guinea worm disease experiences maturation in an intermediate host before it can be transmitted to humans.

Many researchers used basic mathematical modeling to understand and predict the spread of a disease. Basic mathematical modeling for this infectious

disease is Susceptible-Infective-Recovered (SIR) model. Murali Haran (2009) stated that basic SIR models assume that individuals are born into the susceptible class. Those who have never been infected by the disease however are able to catch the disease if they move to the infected class called susceptible individuals. As the infected individuals spread the disease to susceptible, they will remain in the infected class before moving into the recovered class. He also stated that for individuals that are in the recovered class are assumed to be immune for life.   Figure 1.9 shows the SIR model flow:
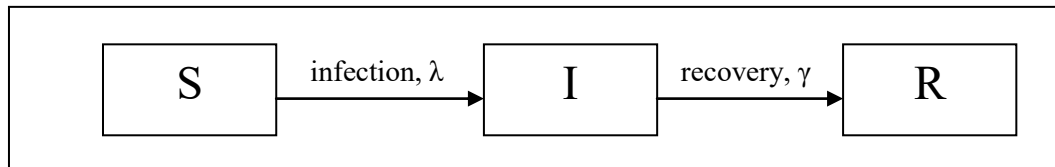
Figure 2.1. SIR model flow (source: Apenteng, 2013)

From Figure 2.1, the arrows represent the movement between classes. If the arrow enters the compartment, it shows the number for that compartment is increasing. While if the arrow leaves the compartment, it means that there is decreasing number in that compartment. For a closed population which means that there are no births, deaths and migration, Murali Haran (2009) stated that the total number of host population is $N = S + I + R$ so if S and I are known, R is known. Let $S(t)$, $I(t)$ and $R(t)$ be the numbers of the members of each class. Abramson (2001) made a few assumptions concerning the transmission and incubation periods. First, he assumed that the number of infected increases at a similar rate to both the numbers of susceptible and the infected: λS, where $\lambda = \beta I$ with $\beta > 0$, where β is the infection rate. According to Apenteng (2013), an infected individual makes contact and is able to transfer the disease to others with βN others per unit time. The

fraction of contacts by an infected with a susceptible is S/N. The law of mass action is applied to epidemics which states that: the rate of new infection of the disease is proportional to both the number of susceptible and the number of infectives: $\beta N \times S/N \times I = \beta SI = \lambda S$. Second, he assumed that the infected removal rate to the R class is proportionate to the number of infected only: αI, with $\alpha > 0$, where α refers to the removal rate. He also assumed that susceptible individual that catches the disease becomes infectious immediately as the incubation time is negligible.

Following the changes rate by Abramson (2011), the mean field model can be written as follows:

$$dS/dt = -\lambda S \tag{2.1}$$

$$dI/dt = \lambda S - aI \tag{2.2}$$

$$dR/dt = aI \tag{2.3}$$

For initial conditions: $S(0) = S_0 > 0,\ I(0) = I_0 > 0,\ R(0) = 0$ (2.4)

Based on equations (2.1)-(2.3), some important and general conclusions can be derived. Several models were developed over the years by changing the assumptions of the SIR model for example SIS.

SIS model stands for susceptible individuals (S) and infectious individuals (I). This type of model is used for non-immunizing disease (Hethcote, 2007). In this model, recovery compartment (R) is missing because infective individuals can become susceptible again after they had recovered from the infection. For example those who caught a common cold can become infected again

33

as they cannot acquire long lasting immunity for the disease. Figure 2.2 shows the flow of SIS model.

recovery, γ

| S | infection, λ | I |
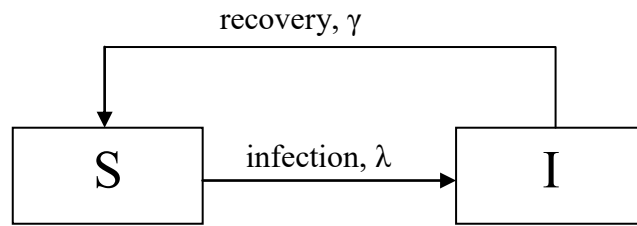
Figure 2.2. SIS model flow (source: Pluciński, Ngonghala, & Bonds, 2011)

From Figure 2.2, those who have never been infected by the disease however are able to contract the disease if they move to the infected compartment called susceptible individuals. As the infected individuals spread the disease to susceptible, they will remain in the infection compartment before moving to the susceptible compartment again. This model is different from the SIR model as there is no recovery compartment as person can become infected again to the disease. The equations of this model can be written as follow:

$$dS/dt = \gamma I - \lambda S \qquad (2.5)$$

$$dI/dt = \lambda S - \gamma I \qquad (2.6)$$

$$N = S + I \qquad (2.7)$$

2.2 Disease Mapping Analysis

Disease mapping is an epidemiological technique which is used to describe the geographic variation of disease and to generate etiological hypotheses about the possible cause for apparent differences in risks (Clement, 2014). According to Nor Azah and Percy (2012), in order to give a clean map (clean from extra noise), most of the diseases mapping studies use regression-type models that include observable

(fixed effects) and unobservable (random effects) variables besides showing the true excess risk. However, published studies for disease mapping which use structural disease transmission models are limited (Gemperli, Vounatsou, Sogoba & Smith, 2006). Some researchers used a stochastic process in finding the relative risk estimation for disease mapping especially for TB case. Many studies of TB are based on exploration of data analysis which includes the review of covariates and factors that contribute to TB disease. For example, studies by Narasimhan,Wood, Intyre and Mathai (2013); and Rafiza and Rampal (2012).

Generally for studies of disease mapping, there are two types of modeling based on the forms of data which available to the researchers. These two types of data are known as case-event data analysis and tract-count data analysis. According to Lawson (2006), case-event data analysis is commensurate with the disease occurrence data that representing the exact cases locations in defined study regions at particular points in time. These types of data also may be referred as continuous data. In other words, the case-event data are expressed as point locations that usually consist of residential streets of address or zip code and time associated infections, reporting or diagnosis. According to Lawson et al. (2003), these types of data are hard to obtain due to the inherent medical confidentiality requirements. Hence, tract-count data analysis is used to overcome this problem.

The tract-count data is a type of discrete data, which correspond to observations that consist of aggregate count cases within the tracts or small study areas such as districts, counties or census tracts, over specified periods of time.

2.2.1 Tract-Count Data Analysis

There are much published research used this type of data for example study by Berke (2004) and Wakefield (2007). The risk in a map can be represented as simple as crude counts where a map shows the high and low risk areas which is based on the numbers of disease incidence, without considering other factors such as the population size in relevant areas. The classical method is the most common statistic used in disease mapping which is based on the standardized morbidity ratio (SMR). However, there are several drawbacks in using the SMR in disease mapping as mentioned earlier in Chapter One. Many other methods have been introduced to overcome it which included informal methods, basic methods and advanced Bayesian models. Lawson et al. (2003) had discussed these methods in their study.

The following section describes the SMR and Poisson-gamma models in detail how the method estimates the relative risk of disease mapping. These two methods are the most common approaches used to produce maps of the disease and the latter represents one of the earliest applications of Bayesian methods in this context.

2.2.1.1 Standardized Morbidity or Mortality Ratios (SMR)

The SMR is a method commonly used by researchers in order to measure relative risk in disease mapping. In board epidemiological terms, SMR can be defined as Standardized Morbidity Ratio or Standardized Mortality Ratio. Morbidity refers to incidence while mortality refers to death. Standardized Incidence Ratios (SIR) and Standardized Hospitalization Ratios have the same concept as SMR (Health Status, 2000). Lawson (2006) described that the SMR method essentially compares the observed occurrence with the expected occurrence that has been

traditionally used to analyze the counts within regions and calculated as $\hat{\theta}_i = O_i / E_i$, where O_i is the observed number of deaths or incident cases of disease in the area and E_i is the expected number of cases.

In disease mapping, based on Nor Azah (2012), assume that the research area to be mapped is distributed into M mutually exclusive areas ($i=1,2,...,M$). Every state has its own number of cases observed O_i and number of cases expected, E_i. Using O_i and E_i as obtained based on the existing data, this can calculate one of the most common indices to estimate the relative risk $\hat{\theta}_i$ for state i, where according to Lawson (2006) and Nor Azah (2012), SMR model is defined as follows:

$$r_i = \hat{\theta}_i = \frac{O_i}{E_i} \tag{2.1}$$

The observed number could be found from other resources such as the health indicator from the Ministry of Health Malaysia which is under disease control department. For the expected value, it could be counted by using a particular formula as in Equation (2.2) until (2.7) for the study region. For the purpose of this research, based on Nor Azah (2012), the expected number of cases, E_i is calculated as

$$E_i = N_i \frac{\sum O_i}{\sum N_j} \tag{2.2}$$

where N_i is the population of region i and the summations (Σ) are for $j=1, 2, . . ., M$. Here, standardization the total population at risk, assume that everybody is equally at risk. Consequently, relative risk can be estimate using formula;

$$\frac{O_i}{E_i} = \frac{O_i}{N_i \frac{\sum O_i}{\sum N_j}} \tag{2.3}$$

$$r_i = \frac{O_i}{N_i \frac{\sum O_j}{\sum N_j}} \tag{2.4}$$

$$= O_i \times \frac{\sum N_j}{N_i \sum O_j} \tag{2.5}$$

$$= \frac{O_i}{N_i} \times \frac{\sum N_j}{\sum O_j} \tag{2.6}$$

$$r_i = \hat{\theta}_i = \frac{\left(\dfrac{O_i}{N_i}\right)}{\left(\dfrac{\sum O_j}{\sum N_j}\right)} \tag{2.7}$$

which defines it as the probability that a person within the state contracts the disease is divided by the probability that a person in the population contracts the disease.

Even though SMR was used commonly to assess the relative risks, it has some drawbacks where it is reliable to use SMR as a measurement of relative risk for wide areas as for example states or countries, however it is unreliable for small geographical areas like districts (Meza, 2003). According to Lawson et al. (2003), the SMR's mean and variance are highly based on E_i, as it depends on a ratio estimator. This has caused the use of SMR has been critiqued as it is very difficult to interpret the data. In some cases where the areas of the expected numbers of cases are small, it can give very large numbers and vice versa. Moreover, SMR will necessarily be zero when there are no observed count data or cases. This causes the

interpretation of SMR becomes difficult and needs to be done with caution. Poisson-gamma models were then introduced to solve these drawbacks. Next section will discuss about this Poisson-gamma models.

2.2.1.2 Poisson-gamma Model

Many researchers have investigated various methods for estimating the relative risk of a disease due to the shortcoming of the SMR as a relative risk estimator. One of it includes the use of the initial applications of Bayesian methods, namely the Poisson-gamma model (Lawson et al., 2003). Poisson distribution is used as this is the fundamental model for count data.

According to Abdul Karim and Amoako (2016), in this model, for the unknown risk of disease in area i is given as θ_i where $i=1,2,...,M$. Let y_i and N_i represent the number of new infectives and the population at risk respectively in area i. The expected number of disease cases in area i can be written as $e_i=rN_i$, where the overall disease risk in the study population as follow:

$$r_i = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} N_i} \tag{2.8}$$

According to Nor Azah and Syafiqah Husna (2013), in this model, it is assumed that the numbers of new infectives, y_{ij} follow a Poisson distribution within a given time period, with mean and variance $e_{ij}\theta_{ij}$. Here, $i=1,2,...,M$ for research areas and

$j=1,2,\ldots,T$ refer to periods of time, e_{ij} is the expected number of new infective while θ_{ij} is the relative risk:

$$y_{ij}|e_{ij},\ \theta_{ij} \sim Poisson\ (e_{ij}\theta_{ij}) \qquad (2.9)$$

Based on the sample $y=\{y_{11},\ldots,y_{MT}\}$, Abdul Karim and Amoako (2016) stated that the likelihood function and the corresponding log-likelihood function are expressed as follows respectively:

$$\ell(\theta_{ij}) = \prod_{i=}^{M}\prod_{j=1}^{T}\frac{\exp(-e_{ij}\theta_{ij})(e_{ij}\theta_{ij})^{y_{ij}}}{y_{ij}!} = P(y,e\,|\,\theta) \qquad (2.10)$$

and

$$\ln \ell(\theta_{ij}) = -\sum_{i=1}^{M}\sum_{j=1}^{T}e_{ij}\theta_{ij} + \sum_{i=1}^{M}\sum_{j=1}^{T}y_{ij}\,\ln(e_{ij}\theta_{ij}) - \prod_{i=1}^{M}\prod_{j=1}^{T}y_{ij}! \qquad (2.11)$$

The maximum likelihood estimator $\hat{\theta}_{ij}$ of θ_{ij} is obtained by $\dfrac{\partial(\ln \ell(\theta_{ij}))}{\partial\theta_{ij}}=0$ and is given by

$$\hat{\theta}_{ij} = \frac{y_{ij}}{e_{ij}} \qquad (2.12)$$

where $\hat{\theta}_{ij}$ represent the standardized mortality ratio in area i at time j. According to Abdul Karim and Amoako (2016), it is assumed that under the Bayesian paradigm, $y_{ij} \sim Poisson\ (e_{ij}\theta)$, where μ is $\mu_{ij}=e_{ij}\theta$ is the Poisson mean and the parameter of relative risk has a gamma distribution with shape and scale parameters α and b respectively:

$$\theta \sim Gamma(a,b) \tag{2.13}$$

The likelihood function for y_{ij} is presented by

$$\ell(\theta) = \prod_{i=1}^{M} \prod_{j=1}^{T} \frac{(e_{ij}\theta)^{y_{ij}} \exp(-e_{ij}\theta)}{y_{ij}!} = P(y,e|\theta) \tag{2.14}$$

and prior distribution for θ_{ij} is showed by

$$P(\theta|a,b) = \frac{(\theta)^{a-1}}{\Gamma(a)b^a} \exp(-\theta b), \theta, a, b > 0 \tag{2.15}$$

(Abdul Karim & Amoako, 2016)

Based on this Poisson-gamma model, expected posterior relative risk is included in the analysis. This risk is for all areas and time periods. In this Poisson-gamma model, Bayesian hierarchical method is used for the parameters estimation. According to Abdul Karim and Amoako (2016), a and b are given prior distributions such that $a|\omega \sim P(a|\omega)$ and $b|\phi \sim P(b|\phi)$, where $P(a|\omega)$ and $P(b|\phi)$ are the hyperprior distribution with hyper-parameters ω and $(\phi_a, \phi_b) \in \phi$ for a and b respectively. Parameters can be obtained using this Bayesian hierarchical method. This is a second stage hierarchical modeling using the Poisson-gamma model. In this study, $P(a|\omega) = \omega \exp(-\omega a)$ and $P(b|\phi_a, \phi_b)$ are defined as exponential distribution and gamma distribution respectively. Hence, the posterior distribution is as follow

$$P(\theta, a, b|y, e) \propto P(y,e|\theta,a,b)P(\theta)P(a|\omega)P(b|\phi) \tag{2.16}$$

The estimation of Poisson-gamma model's parameters were carried out using WinBUGS software.

Lawson et al. (2003) demonstrates the use of Poisson-gamma model in their study which gives a smoother map with less extreme values for estimating the relative risk as compared to using SMR. However, in this model, the adjustment of the covariate is difficult and there is no likely to deal with spatial correlation between risks in adjacent regions. Thus, this encourages other researchers to propose new methods to estimate the relative risk.

2.2.2 Case-Event Data Analysis

Analysis of case-event data involves the use of point process models and related methodology. This can be seen in the studies by Cox (1972) and Diggle (2003). According to Lawson (2006), most of the earliest development of case-event data can be found in ecological application mostly in a homogenous environment. Some basic models for the point processes to ecological applications have been applied in these early studies. Among these models, complete spatial randomness (CSR), spatial cluster processes and spatial inhibition processes are the most important models in applications (Lawson, 2006). This work has been reviewed by Ripley (1981), Cressie (1993) and Diggle (2003). In order to model spatial epidemiological data, these basic point process models have been extended and modified.

This model basically concerned with individuals having an independent probability of becoming a case. Heterogeneous background and non-stationary or long-range spatial trend components are include in the models by applying a special type of Poisson process model called the heterogeneous Poisson process (HEPP)

model. HEPP is a simple extension of the homogeneous Poisson process where the first-order intensity is allowed to be spatially dependent (Lawson, 2006).

A number of reviews by Gatrell, Bailey, Diggle, Rowlingson, and Rowlingsont (1996) on non-stochastic methods for the exploring and modeling of spatial point patterns with particular reference to geographical epidemiology has been made. This paper described how spatial point patterns can be represented statistically. There are four methods mentioned by the authors that are Kernel estimation, extensions to Kernel estimation that involve the ratio of Kernel estimates to both cases and controls, the K function, and extensions to the K function that involve bivariate and space-time patterns. In particular, the use of ratios of Kernel estimates has been proposed as a way to assess spatial variation in disease risk. According to Gatrell et al. (1996), K function explains the extent to which there is spatial dependence in the arrangement of events in this study. All of these methods mentioned in this study are then applied to environmental epidemiology to examine the important issue of spatial clustering. They considered space-time before they examined a spatial point process model that appears to be considerable value in geographical epidemiology.

Many of disease mapping studies with case-event data consider space time point process models in the analysis due to the occurrence of disease incidence which commonly related to the location and time which a person contracts a disease (see for example, Fishman and Synder (1976); Lawson (2000); Diggle (2006); Diggle, Kaimi, & Abellana (2010); Schrodle & Held (2011a); Schrodle & Held, (2011b)). This space time point process models also known as spatio-temporal point process.

Most of the spatio-temporal point process theory is carried over from the spatial point process.

For TB case, Srinivasan and Venktesan (2013) use the geographic information system (GIS) and Kringing method in their space-time analysis to find the spatial pattern and spatial dependence of TB within a Chennai ward population in the year 2004 until 2006.

2.3 Basic Concepts in Mathematical Modeling of Infectious Disease

The central idea about transmission models, as opposed to statistical models, is a mechanistic description of the transmission of infection between two individuals (infectious and susceptible persons). This mechanistic description makes it possible to describe the time evolution of an epidemic in mathematical terms and in this way connects the individual level processes that contribute to disease transmission in much detail. Hence, developing mathematical model helps to focus thoughts on the essential processes involved in shaping the epidemiology of an infectious disease and to reveal the parameters that are most influential and amenable for control. Mathematical modeling is then also integrative in combining knowledge from very different disciplines from microbiology, clinical sciences, biology, mathematics/physics, computer science, zoology, to environmental and social science.

According to Aron (2007), mathematical modeling has a clear mathematical description of the simplified dynamics of a system. The advantage of using the models is they aid understanding by simplifying the reality. Mathematical models can show how the progression of infectious diseases and can help inform public

health interventions. Helmersson (2012) stated that models need to capture the important behavior of interest and take in the important processes. Creating a mathematical model that clearly explains the rationale and allows others to check them out. Thus, the mathematical model allows accurate, in-depth analysis and quantitative forecast.

According to Helmersson (2012) there are two types of mathematical modeling that are deterministic (or transmission) and stochastic (or statistical). Deterministic models contain a finite number of regions and specify rules by which individuals move from one region to another region through a series of differential equations. The deterministic models also known as compartment model. It can be used to determine and describe the average on the population scale and it is suitable to be used for large population. To examine this type of model, differential equation or different equation can be used. Differential equation focuses on the continuous cases while difference equation explains the transition between the different disease compartments using discrete time steps. Conversely, stochastic models determine the probabilities of movements between the populations and therefore can give probabilities of particular model outcomes (Achterberg, 2009). Stochastic models is used when alteration of possible variation in exposure risk, disease and other factors are important, for example in a small population.

Normally there are four steps in modeling. First step is a flow diagram that represents the natural history and transmission of infection. Based on this diagram, write a set of mathematical equations to express the transmission process. The third step is to find proper values for the parameters used in the equations. Finally, to solve the equations algebraically or numerically, computer simulation programs are usually

used. Due to short time limit of this thesis, the focus of this study is on the first three steps and the last step of calculation is carried out for simple cases only.

2.4 Simple Epidemic SIR Model for Infectious Disease (Human Only)

First, let consider a simple infectious disease case of transmission process such as measles. As stated previous, from simple models, some of the essential quantities can be learned. By analyzing the model, it can help to have better understanding on the phenomenon of outbreak spread and disappearance.

Here, persons in the population can be categorized into three compartments:

a) Susceptible to the disease (Susceptibles) – S,

b) Currently Infectious (Infectious or Infectives) – I,

c) Recovered and immune (Recovered or Removals) – R.

Here, the number of persons in each compartment is represented through S, I, and R. The total host population is $N = S + I + R$. Figure 2.1 shows the flow of transmission process. This model is known as SIR model. Here, the details including vectors are ignored, and considering only the population of the human.
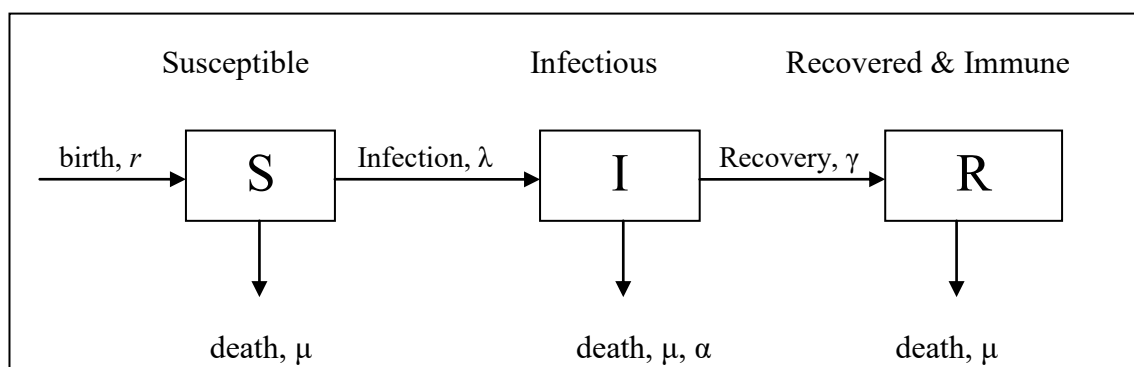
Figure 2.3. Flow diagram of SIR model

The flow rate is known as the incidence rate is represented by each arrow in Figure 2.3. This incidence rate shows the movement of individuals that enter or leave a compartment per unit time. All the parameters ($r, \mu, \alpha, \lambda, \gamma$) in this model are per capita rate where

r : birth rate

μ : natural death rate

α : death rate because of the disease

λ : the transmission of infection rate

γ : recovery rate

From Figure 2.3, the number of susceptible (S) is increased by the birth rate (r) and decreased by the natural death that is not caused by the disease at rate μ. The number of infectious persons (I) is increased by the infection events rate (λ) from the susceptible stage and decreased by recovery rate (γ) of infected persons and by the natural death rate (μ) plus deaths rate caused by the disease (α). The number of recovered and immune persons (R) increased by the recovered infectious persons and decreased by the natural death rate (μ).

To find the flow rate of population, the per-capita rate ($r, \alpha, \lambda, \gamma, \mu$) is multiplied with the number of persons subjected to that per-capita rate (S, I, R or N). For example, rN is for the "birth" flow to "susceptible" compartment, where r is per-capita birth rate and is multiplied by the totals number of the system's population N. It is assumed that all persons giving birth at the same rate r where males and females

are averaged over. Another example is λS for the population for infection rate, where the per-capita infection rate λ multiplied the number of susceptible S. It can be summarized the population flow rate for each arrow as follow:

Number of birth = rN,

Number of infection = λS,

Number of recoveries = γI,

Number of deaths of $S = \mu S$,

Number of deaths of $I = (\mu + \alpha)\,I$,

Number of deaths of $R = \mu R$.

Next, based on the flow diagrams, the equation is written to state the change of state variables: dS/dt, dI/dt, dR/dt where t refers to time. Based on Helmersson (2012):

rate of change of susceptible with respect to time = Number of birth – number of infection – number of deaths (of S)

rate of change of infectious with respect to time = Number of infection – number of recoveries – number of deaths (of I)

rate of change of recovered with respect to time = Number of recoveries – number of deaths (of R)

According to Helmersson (2012), since every term in the equations is a rate quantity and has the unit of 1/time. Based on the compartmental model in Figure 2.1 and the equations written above, using the parameters and population flow rates, it can also be showed mathematically in the form of differential equations as follow

$$dS/dt = rN - \lambda S - \mu S \tag{2.17}$$

$$dI/dt = \lambda S - \gamma I - (\mu + \alpha) I \tag{2.18}$$

$$dR/dt = \gamma I - \mu R \tag{2.19}$$

$$N = S + I + R \tag{2.20}$$

Based on the equations above, λ (i.e. λ is the force of infection) is the only non-constant parameter which is the per-capita rate of susceptible infection and explain the risk that individuals who are exposed to get infected per unit of time. It relies on the contact rate with other persons, c, the probability of transmission when infectious person contacts a susceptible, pT, and the infectious proportion in the population, I/N. It is assumed that $\lambda = p$TcI/N or $\lambda = \beta I/N$, where $\beta = p$Tc is the total rate of transmission in the population.

Here N, S, I, R are dependent variables that change with time. In this case, the variables describe the epidemiological system – population situation in every compartment. Their dependence on different intrinsic time and can be simulated in the model by using a computer program based on the given equations (2.17) - (2.20). The values are manipulated indirectly. However, $r, \mu, \gamma, \alpha,$ and β (or c & pT) are parameters that do not change with time. When the values are set, they remain constant during the calculation as the computer program performs. The parameters are chosen based on the assumptions or estimates from epidemiological data.

After making the assumption that the parameters values are particularized based data and it is made for a particular infectious disease and a particular population, able to move to solving the equations which is the final step of the modeling. The time evolution of state variables is obtained when the equations are solved. As most models cannot be solved algebraically, computer numerical integration is normally a standard method. The initial state should be specified and the computer program will solve the set of equations (2.17) – (2.20) as an iterative procedure over time. Nevertheless, some of the analytical solutions can be achieved under special conditions. After all the steps, the analytical solutions are listed, particularly those connected to threshold conditions. Next section discusses on the stochastic SIR model that was proposed by Lawson (2006) which this study adopts this approach to solve the nonlinear differential equations system and implement this approach to the stochastic models.

2.5 Susceptible-Infected-Recovered (SIR) Model for Direct Disease Transmission

2.5.1 Deterministic SIR Model

The deterministic model for direct disease transmission as in equation (2.21) is proposed by Lawson (2013) in his study of discrete time-space stochastic susceptible-infective-removed (SIR) model for influenza. In order to make it more relevant and clear, different notation is used from the notation used by Lawson.

In Lawson's model, for $i = 1, 2, \ldots, M$ study regions and $j = 1, 2, \ldots, T$ time periods, the equations are defined as follows:

$$S_{i,j+1} = S_{i,j} - I_{i,j} - R_{i,j} \tag{2.21}$$

$$I_{i,j} = S_{i,j} I_{i,j-1} \tag{2.22}$$

$$R_{i,j} = g I_{i,j} \tag{2.23}$$

where

$S_{i,j}$: susceptible population for region i at the start of period j

$I_{i,j}$: number of new infective for region i during the time period jth

$R_{i,j}$: number of removed cases for region i during the jth time period

g : the hazard of an infectious person's being removed.

According to Trottier and Phillippe (2002), the model presented are generally based on the purpose of the study or the availability of the data as it is different from other typical deterministic model used in direct disease transmission study.

Based on Lawson (2013), the product of susceptible humans, $S_{i,j}$ and infective human, $I_{i,j}$ can give a large value, hence Equation (2.22) seems unusable. Therefore, a constant of proportionality should be placed before the product of the susceptible, $S_{i,j}$ and infectives, $I_{i,j}$ for human population. This can be proven by using the common deterministic SIR model as shown in Figure 2.4. The definition for $S_{i,j}, I_{i,j}, R$ and g are shown in Figure 2.4 are similar to those in Lawson's

deterministic model described previously, with the addition of $\alpha I_{i,j}$. $\alpha I_{i,j}$ that refers to the risk of a susceptible person's becoming infective in time period j, where α is constant.
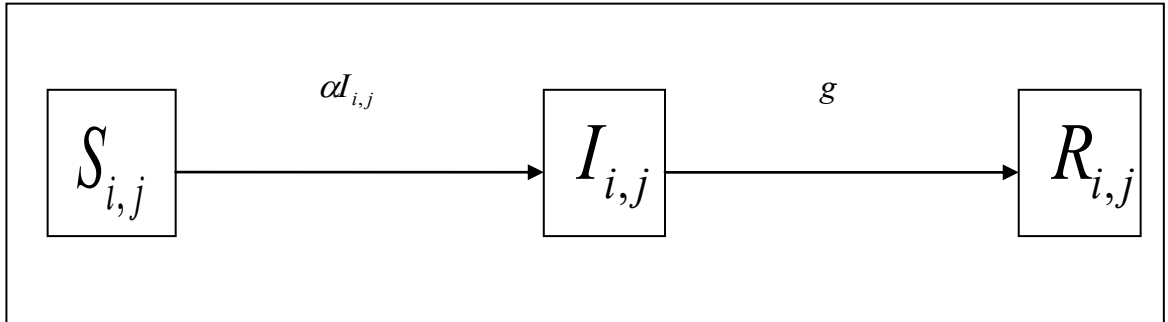
Figure 2.4. Compartmental SIR model for direct disease transmission

Lawson (2013) stated that for discrete time steps, the compartmental model shown in Figure 2.4 can be represented mathematically by using difference equations system as follows:

$$S_{i,j+1} = S_{i,j} - \alpha I_{i,j} S_{i,j} \tag{2.24}$$

$$I_{i,j} = I_{i,j-1} + \alpha I_{i,j-1} S_{i,j-1} - g I_{i,j-1} \tag{2.25}$$

$$R_{i,j} = g I_{i,j} \tag{2.26}$$

The assumption for the removed category in this case only includes the deceased due to infection and does not include people who recover from infection. The latter are in a new category that is not susceptible, infected or removed. The model also assumes that infective either die or infect susceptible within the same time period that they will become infective again or they recover. Corresponding to disease incubation, however, any susceptible those are infected, only become infective again in the next

time period. Lastly, in any particular time interval, the reduction in susceptible is assumed proportional to the product of susceptible and new infective, and the increase in death is proportional to the number of new infective.

Equation (2.25) shows that the product of susceptible and infective humans in Equation (2.22) should involve proportionally using a constant or a rate. The product in Equation (2.25) is multipled by a constant of α. From the deterministic SIR model above, it is extended to generate a stochastic SIR model that also assigns proportionality to the product of susceptible and infective. The details about these features are explained in the next section.

2.5.2 Stochastic SIR Model

The discrete time stochastic SIR model, for i = 1, 2, …, M study regions and j = 1, 2, …, T time periods, proposed by Lawson (2013) as follows:

$$I_{i,j} \sim Poisson(\mu_{i,j}) \tag{2.27}$$

$$\mu_{i,j} = \exp(\beta_0 + b_i).S_{i,j}.I_{i,j-1} \tag{2.28}$$

$$\log(\mu_{i,j}) = \beta_0 + b_i + \log(S_{i,j}) + \log(I_{i,j-1})$$

$$S_{i,j+1} = S_{i,j} - I_{i,j} - R_{i,j} \tag{2.29}$$

$$R_{i,j} = gI_{i,j} \tag{2.30}$$

where

$I_{i,j}$: number of new infective for region i at period j (assumed to follow Poisson distribution as this is the fundamental model for count data). The Poisson distribution mean has a simple mechanistic model for the transmission of infection and a linear predictor term consisting of covariate or random effect ($\mu_{i,j} = \exp(\beta_0 + b_i).S_{i,j}.I_{i,j-1}$).

β_0 : overall rate of the process which is a constant term mean

b_i : spatial random effect for region i, which absorbs residual spatial variation.

The number of new infective is assumed to follow Poisson distribution because according to Lawson (2006), Poisson distribution is often used as a starting point for analysis of count data. Even though there are many types of distributions available such as Normal, Binomial, Gamma, Exponential, one must be aware that there are two types of distributions which are discrete and continuous. In this Lawson' study, he interested to find the number of new infective cases which refers to the number of cases. Since the number of cases is a discrete number, discrete distribution is used and Poisson distribution is one of the most appropriate distributions to be used since it has been used by many researchers for count data analysis.

The simple direct dependence of current infective count, $S_{i,j}.I_{i,j-1}$ is the main element of the transmission equation. It is also a simple form of space-time interaction. The equation for $S_{i,j+1}$ and $R_{i,j}$ are fixed or non-stochastic. $I_{i,j}$ is the only stochastic element, where Equation (2.22) is replaced by Equation (2.27) and Equation (2.28). Moreover, Lawson assumed that the removal rate g to be fixed. In

the next chapter, this SIR model is extended to derive a SLIR model that has generic applicability but is specifically relevant to the TB disease study since TB disease transmission has latent period before someone become an active TB.

CHAPTER THREE
METHODOLOGY

This chapter discusses the methodologies that are used in this study. This chapter also provides discussion on the data sets that are utilized. The chapter provides in detail the development of a stochastic SLIR model for direct disease transmission with particular application to TB disease. There are four phases in conducting this study as displayed in Figure 3.1. The first phase focuses on the deterministic model, followed by stochastic model is in the second phase. Then, the third phase deals with the estimation of the relative risk and comparison of the results. Lastly, the disease mapping for TB disease is constructed. The discussion initially considers the simple deterministic SLIR model for direct disease transmission proposed by Blower et al. (1995) with the combination of basic SLIR model (Subsection 1.2.1), specifically for TB disease and then followed by developing a stochastic SLIR model for direct disease transmission from it. This will be used later to estimate the relative risks for mapping TB disease.
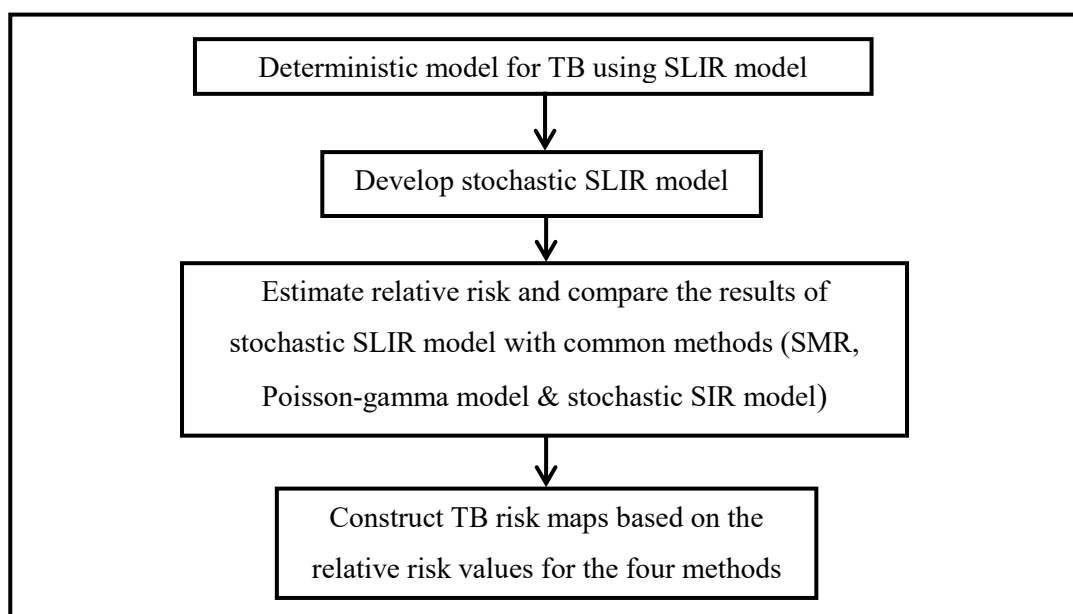
Figure 3.1. Flow of the research methodology

3.1 Susceptible-Latently Infected-Infected-Recovered (SLIR) Models for Direct Disease Transmission

In this section, a compartmental model and the development of the deterministic Susceptible-Latently infected-Infected-Recovered (SLIR) models for direct disease transmission were described, followed by stochastic SLIR models. For the purposes of this research, TB disease is used as case study. However, these models can be used for other direct disease transmissions, especially for a disease where the virus is transmitted by airborne and has latent period. The analysis of this study is started by extending the discrete time SIR model proposed by Lawson (2006) in order to form SLIR for TB disease transmission.

3.1.1 Deterministic SLIR Model

The compartmental model displayed in Figure 3.2 is the most common model used in the study of TB transmission which is a simple model adapted from Blower, Mclean, Porco, Small, Hopewell, Sanchez and Moss (1995) and combined with basic SLIR model (Subsection 1.2.1). In this study, for $i = 1, 2, \ldots, M$ study areas and $j = 1, 2, \ldots, T$ time periods, the definitions of the compartments are given as follows:

$S_{i,j}$ = total number of susceptible persons for area i, at time j

$L_{i,j}$ = total number of latently infected persons for area i, at time j

$I_{i,j}$ = total number of active infectious persons for area i, at time j

$R_{i,j}$ = total number of recovered persons for area i, at time j

The total host population for region i is $N_i = S_i + L_i + I_i + R_i$. The transmission process and the important features of this model are described in Figure 3.2. This model is called SLIR model.

According to Blower et al. (1995), TB models that assumed to be infectious are divided into two types: primary progressive TB (i.e., fast TB) and latently infected TB (i.e., slow TB). This is because infected individuals can develop active TB either by pathogenic mechanisms that are direct progression which the disease develops quickly after infection or by endogenous reactivation where the disease develops many years after the infection.
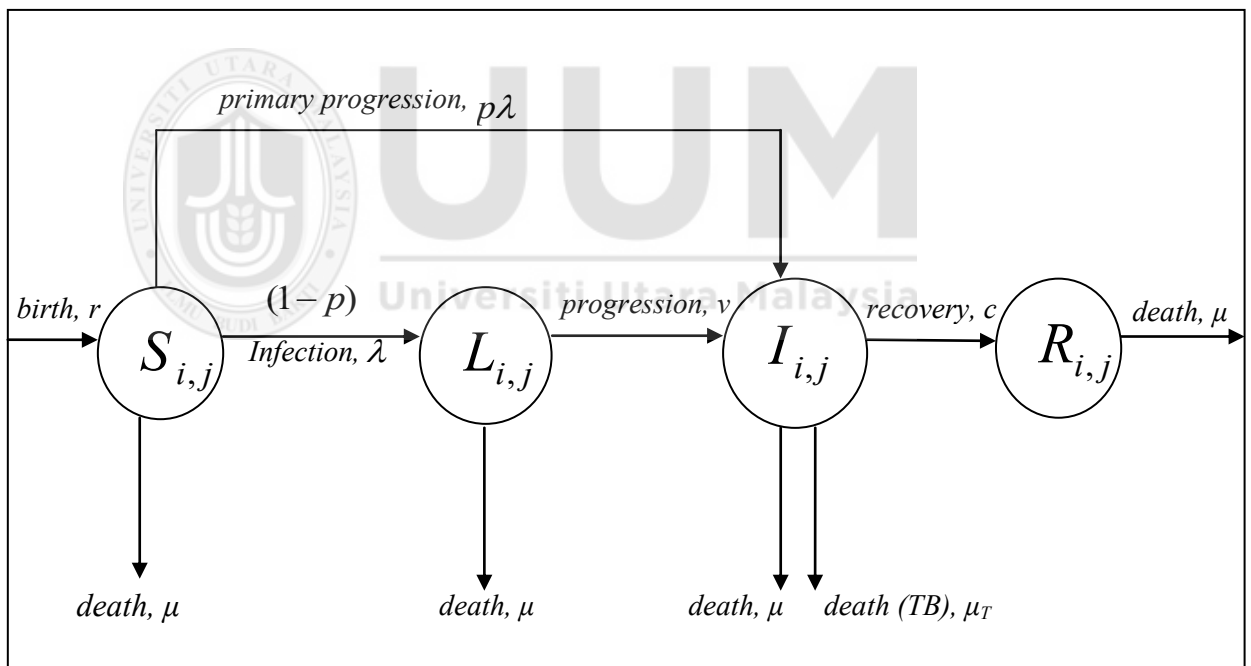
Figure 3.2. Flow diagram of the SLIR model for TB transmission

Each arrow in Figure 3.2 indicates the flow rate that is known as the incidence rate where people enter or leave a compartment per unit time. In this model, the number of susceptible people (*S*) is increased by birth rate (*r*) and decreased by natural (non-diseased) death rate (μ) and by infection transmission

events rate of susceptible (λ, the force of infection). This occurrence of infection rate is calculated as the product of the numbers of the susceptible present at time t ($S(t)$) and the per capita force of infection at time t ($\lambda(t)$); where $\lambda(t) = \beta I(t)$. Here, $\lambda(t)$ is defined as the per-susceptible risk of becoming infected with the virus and is calculated as the product of the number of infected individuals at β and time t ($I(t)$), the transmission coefficient, which indicates the likelihood that an infectious case will transmit the infection to a susceptible individuals successfully. The number of latently infected persons (L) is increased by the remaining fraction ($1-p$) of infection individual and decreased by the number of individuals who develop TB slowly at an average rate v and natural death rate (μ). The number of infectious (I) is increased by infection events of susceptible (direct progression soon after infection, fast TB, at a probability of p) and the number of individuals that develop TB through endogenous reactivation in the latently infected state (slow TB) (at rate v). The number of infectious (I) is decreased by the mortality rate either because of TB (μ_T) or by other causes (μ) and have been recovered with the recovery rate (c). The number of recovered people (R) is increased from the recovery rate (c) and decreased by natural death rate (μ).

All the parameters r, μ, μ_T, λ, v, p and c are per-capita rates. The flow rate of population is the per-capita rate (r, μ, μ_T, λ, v, p or c) multiplied by the number of people subjected to that per-capita rate (N, S, L, I or R). For example, the population for the birth, the in-flow —birth" to a compartment of —susceptible" is N, the per capita birth rate r times the total population N of the system, assuming all individuals give birth at the same rate which is averaged over males and females. Hence, the population flow rate in Figure 3.2 can be summarized as below:

Number of birth $= rN$

Number of infection $= \lambda S$

Number of latently infection $= (1-p)\lambda S$

Number of latently to active TB $= vL$

Number of infectious $= p\lambda S + vL$

Number of recoveries $= cI$

Number of deaths of $S = \mu S,$

Number of deaths of $L = \mu L,$

Number of deaths of $I = (\mu + \mu_T)I$

Number of deaths of $R = \mu R$

rN is the recruitment rate of people per year, where r is the birth rate per year. In this study, it is assumed that the birth rate and death rate of human per year are equal. As described in Chapter 2, TB transmission occurs due to aerosol particles that contain the live tubercle bacilli released from the individuals with infectious TB disease (I). Then, a person could inhale these infectious particles which stay in the lungs of this susceptible individual. Thus, each of the susceptible will be undergoing

an immediate infection risk per year, λ. It is assumed that $\lambda = \beta I$ where βI is the rate

of hazard for infection which is greater when there are more infective. It is assumed

that β is the transmission rate per year (hazard rate per infective). Infected persons

stay non-infectious until they develop the disease by one of two pathogenic

mechanisms: direct progression (also known as primary progression) which occurs

immediately after infection with *M. tuberculosis* or endogenous reactivation which

takes several years after being infected. The two pathogenic mechanisms are

modeled by allowing a proportion p of individuals progresses to disease relatively

soon after infection; the remaining proportion 1- p of infected persons become

latently infected. The L latently infected individuals are not infective to the others.

The infection remains latent in some latently infected persons and the infection may

be continuous for life (Porco & Blower, 1998). However, in a small number of

latently infected persons, latent infection reactivation may lead to disease

(Porco & Blower, 1998). This is because the progression rate to disease is extremely

slow. The probability per unit time that a latently infected individual will undergo

reaction is representing by v. In other words, v refers to the progression rate from

latent TB to active TB per person per year. Active infectious TB cases die either due

to TB at an average μ_T (death rate due to TB per year) or because of other causes at

an average rate μ (natural mortality rate from other cause per year). The number of

naturally recovered cases is represented by R with rate c per person per year, which

the persons are no longer infectious, even they may retain infection.

The compartmental model in Figure 3.2 can be shown mathematically in the

form of differential equations system. The difference between the deterministic

model that used by Lawson and the one used in this study is that for TB disease,

model proposed consists the exposed or latent state. The equations are written based on the flow diagram (Figure 3.2) to show the change of state variables: dS/dt, dL/dt, dI/dt, dR/dt. Here t is defined as time.

rate of change of susceptible with respect to time = number of birth – number of infection – number of deaths (of S)

rate of change of latent with respect to time = number of latently infection – number of latency to active disease – number of deaths (of L)

rate of change of infectious with respect to time = number of infectious – number of recoveries – number of deaths (of I)

rate of change of recovered with respect to time = number of recoveries – number of deaths (of R)

Every term is a rate quantity and has the unit of 1/time. Below is the transmission process using mathematical expressions using the parameters and population flow rates based on the flow of diagram in Figure 3.2 and the equations written above:

$$S'(t) = rN - (\lambda + \mu)S(t) \tag{3.1}$$

$$L'(t) = (1-p)\lambda S(t) - (v+\mu)L(t) \tag{3.2}$$

$$I'(t) = p\lambda S(t) + vL(t) - (\mu + \mu_T + c)I(t) \tag{3.3}$$

$$R'(t) = cI(t) - \mu R(t) \tag{3.4}$$

Based on the above differential equations, this is non-linear because of the existence products of functions. According to Nor Azah and Percy (2013), it is difficult to solve analytically, while the analytic solution will be relatively easy if the system of linear ordinary differential equations will be relatively easy. These non-linear equations generally can be solved by performing numerical analysis using computer programming. Hence, in this study, computer programming that is WinBUGS is used to help to carry out numerical analysis as a means of determining solutions of complicated systems of non-linear ordinary differential equations such as those in (3.1) to (3.4).

In this study, numerical analysis method is used to calculate the solution for SLIR model. Lawson's approach is adopted as mentioned in Section 2.3 to solve the system of non-linear differential equations.

Hence, the deterministic model for TB disease transmission based on Figure 3.2 for $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, T$ time periods, for TB disease transmission reduces to this difference equation set:

$$S_{i,j} = rN + (1 - \lambda - \mu)S_{i,j-1} \tag{3.5}$$

$$L_{i,j} = (1 - p)\lambda S_{i,j-1} + (1 - v - \mu)L_{i,j-1} \tag{3.6}$$

$$I_{i,j} = p\lambda S_{i,j-1} + vL_{i,j-1} + (1 - \mu - \mu_T - c)I_{i,j-1} \tag{3.7}$$

$$R_{i,j} = cI_{i,j-1} + (1 - \mu)R_{i,j-1} \tag{3.8}$$

63

The model derived above has the same form as the deterministic SLIR model used by Blower et al. (1995). Here N_i is assumed to be constant, where $N_i = S_i + L_i + I_i + R_i$. In the next section, this formulation is used to provide links to stochastic mean populations. In this study, the birth and natural death rate of humans per year are assumed to be equal.

3.1.2 Stochastic SLIR Models

In the following analysis, the deterministic model is used to provide an approximation to the stochastic means. According to Molzon (2009), deterministic models are good approximations to the stochastic models. In this study, stochastic SLIR models are developed based on stochastic SIR model proposed by Lawson (2006). New notations are introduced, $\bar{I}_{i,j}$ and $\mathfrak{R}_{i,j}$ which represent the numbers of newly infected and newly recovered respectively, all in the study region i and time period $(j-1, j]$ since the TB data that observed are yearly new infective cases and this study interested to find the posterior mean of new infective TB cases in each year.

For $i = 1, 2, \ldots, M$ study regions and $j = 1, 2, \ldots, T$ time period, the stochastic models for TB transmission is adapted from the deterministic SLIR equations (3.5) – (3.8).

$$\bar{I}_{i,j} = p\beta I_{i,j-1} S_{i,j-1} \tag{3.9}$$

$$\bar{I}_{i,j} \sim Poisson(\gamma_{i,j}) \tag{3.10}$$

$$\gamma_{i,j} = \exp(\beta_0 + b_i)p\beta I_{i,j-1} S_{i,j-1} \tag{3.11}$$

$$\log(\gamma_{i,j}) = \beta_0 + b_i + \log p + \log \beta + \log(I_{i,j-1}) + \log(S_{i,j-1})$$

$$S_{i,j} = \mu N + (1 - \beta I_{i,j-1} - \mu)S_{i,j-1} \qquad\qquad (3.12)$$

$$L_{i,j} = (1 - p)\beta I_{i,j-1}S_{i,j-1} + (1 - v - \mu)L_{i,j-1} \qquad\qquad (3.13)$$

$$I_{i,j} = \bar{I}_{i,j} + vL_{i,j-1} + (1 - \mu - \mu_T)I_{i,j-1} - \Re_{i,j} \qquad\qquad (3.14)$$

$$R_{i,j} = \Re_{i,j} + (1 - \mu)R_{i,j-1} \qquad\qquad (3.15)$$

$$\Re_{i,j} = cI_{i,j-1} \qquad\qquad (3.16)$$

From the Equation (3.11) above, β_0 stand for the overall rate of the process and it is a constant term. b_i refers to the random effect that absorbs residual spatial variation which allows dependencies among adjacent regions and it is the main stochastic element. In any analysis that involves the geographical region, spatial random variation is an important element. Poisson distribution is used to model the number of new infectious as this is the basic model for count data. Its mean, $\gamma_{i,j}$ is chosen to match part of the deterministic form in Equation (3.7) with a positive multiplicative factor to show the spatial correlation. The number of new infectious ($\bar{I}_{i,j}$) is assumed to follow Poisson distribution where the expected number of new infective persons includes several elements of the disease transmission. Figure 3.3 shows the flow of the stochastic SLIR model that was proposed.
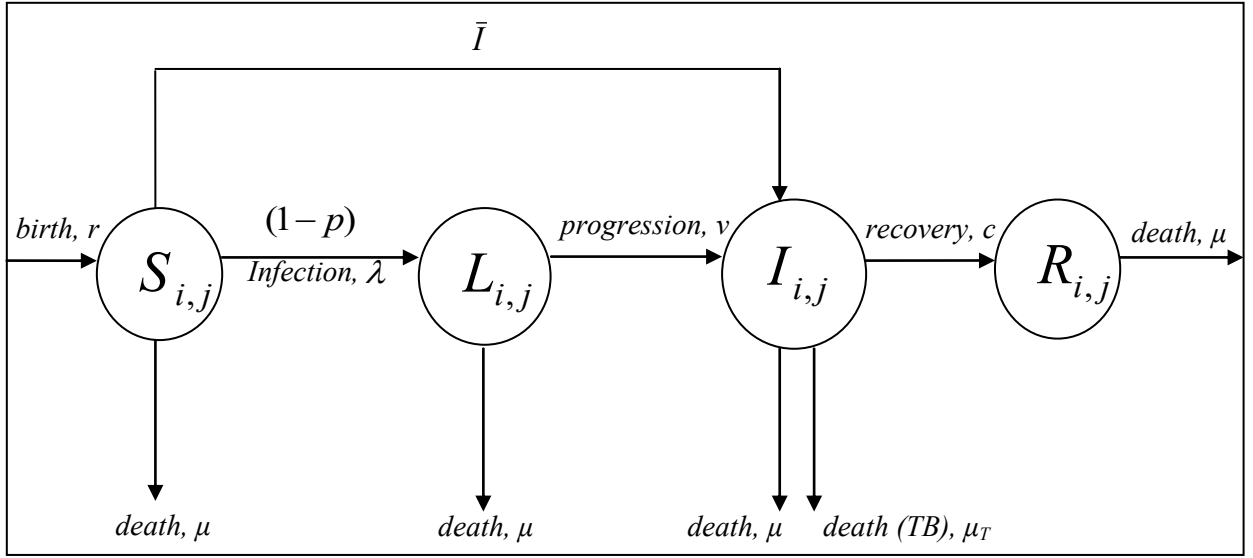
Figure 3.3. Flow diagram of the stochastic SLIR model for TB transmission

where

$\bar{I}_{i,j}$: the number of new infective persons for area i, at time j

$S_{i,j}$: total number of susceptible for area i, at time j

$L_{i,j}$: total number of latently infected persons for area i, at time j

$I_{i,j}$: total number of infectious persons for area i, at time j

$R_{i,j}$: total number of recovered persons for area i, at time j

$\Re_{i,j}$: the number of newly recovered persons for area i, at time j

μ : birth and natural death rate of humans per year (assumed equal)

μ_T : mortality rate because of TB of humans per year

λ : the transmission of infection events of susceptible

N : the population size for the study state

v : the progression rate from latent period to infectious period

p : probability of new infections that develop progressive primary active

TB

(1-p): probability of infected persons become latently infected

c : the rate of recovery

β_0 : the overall rate of the process

b_i : the random effect for region *i,* that absorbs residual spatial variation

This study interested to know the number of new infective person. So that, the number of new infective $(\bar{I}_{i,j})$ is the only stochastic element in this model. The other term is a non-stochastic element. In future research, these models can be extended due to the interest of other researchers by considering other stochastic terms. It is assumed that the number of newly infectious $(\bar{I}_{i,j})$ follows the Poisson distribution where the expected numbers of new infective persons take in several elements of the transmission. This new infective case is measured by $\gamma_{i,j}$ where it is the rate of transmission. Poisson distribution is used as a starting point for analysis of count data (Lawson, 2006). This study is interested to find the number of new infective cases which refers to the number of cases. Since the number of cases is a discrete number, discrete distribution and Poisson distribution are used in this study as Poisson distribution is one of the most appropriate distribution to be used many researchers used it for count data analysis. In this study, it is assumed that the birth rate and death rate to be equal. Hence, the symbol of the birth rate that is *r* is changed to be *μ*.

Based on Figure 3.2, the main element of the transmission equation is the simple direct dependence of current infective count on previous the same spatial unit and a linear predictor term that can take in a random effect or covariates that is $p\beta I_{i,j-1}S_{i,j-1}$. Therefore, for a number of new infective people is

$\bar{I}_{i,j} = p\beta I_{i,j-1} S_{i,j-1}$. The equation for $S_{i,j}$, $L_{i,j}$, and $R_{i,j}$ are non-stochastic or fixed. Poisson distribution cannot be tested in isolation as these counts are conditional upon other variables. As in Lawson's model in Subsection 2.5.2, the constant term β_0 is also included here to express the overall rate of the process and term b_i is a spatial random effect that stands for a random effect with a spatial prior distribution which allows dependency between adjacent areas.

The intrinsic conditional autoregressive (CAR) prior distribution is used in this study as a family of prior distributions for the random effect b_i. According to Lawson et al. (2003), the CAR model is a common way of dealing with the spatial correlation between neighbouring areas where the probability densities of values at any given place are conditional on the neighbouring areas. This model was introduced by Besag et al. (1991). According to De Oliveira (2012), the main objective of this CAR model is to unveil and quantify spatial relations present among the data, in particular, to quantify how quantities of interest vary with explanatory variables and to detect clusters of ‗hot spots'.

The conditional distribution of the spatially component in area i is specified as normal (Lawson et al., 2003) where temporarily dropping the SLIR model for the convenience,

$$c_i \big| c_s, i \neq s, \sigma_c^2 \sim Normal(\bar{c_i}, \sigma_i^2),$$

with mean $\bar{c_i} = \dfrac{1}{\sum\limits_s w_{i,s}} \sum\limits_s c_s w_{i,s}$ and variance $\sigma_i^2 = \dfrac{\sigma_c^2}{\sum\limits_s w_{i,s}}$. Stern and Cressie (1999) suggested that

$$
w_{i,s} = \begin{cases} w_{i,s} & \text{for neighbouring} \\ 0 & \text{for otherwise.} \end{cases}
$$

Here, $w_{i,s}$ is an appropriate weight and σ_c^2 is the unknown variance parameter which is the only hyperparameter in this distribution. Prior distribution must be specified for this hyperparameter for a full Bayesian analysis. Here, gamma distribution is considered as suggested by Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi and Songini (1995). A range of CAR models is supported by GeoBUGS extension to the WinBUGS package.

The stochastic SLIR model for transmission of TB disease that proposed here is used to estimate the relative risk.

3.2 Relative Risk Estimation for Disease Mapping

To estimate the relative risk, disease mapping studies are mostly used regression-type models. However, for this study, a different method to estimate the relative risk is used which is depend on the transmission model of disease specifically built for TB. WinBUGS software is used to perform the computational analysis. WinBUGS software is a program intended to implement Bayesian inference on the statistical problem using Markov Chain Monte Carlo (MCMC) computations (Lawson et al., 2003). Further discussion and application for Bayesian analysis of disease mapping using this software are available in Lawson et al. (2003).

In this study, the method used to estimate the relative risk is based on the method suggested by Nor Azah and Percy (2012). In general, for $i = 1, 2, \ldots, M$ study regions and $j = 1, 2, \ldots, T$ time periods, the posterior expected mean number of infectives can be approximated using an unbiased sample mean

$$\widetilde{\lambda}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \lambda_{ijk} \tag{3.17}$$

where λ_{ijk} for $k = 1, 2, \ldots, n$ is produced from the posterior distribution for the expected mean number of infective $\widetilde{\lambda}_{ij}$.

Then, the parameter of relative risk θ_{ij} is defined by

$$\theta_{ij} = \frac{\lambda_{ij}}{e_{ij}} \tag{3.18}$$

where e_{ij} stand for expected number of new infective cases based on the population across the study areas. Hence, by using an unbiased sample mean, according to Nor Azah and Percy (2012), the posterior expected relative risk can be estimated by

$$\widetilde{\theta}_{ij} = \frac{1}{n} \sum_{k=1}^{n} \theta_{ijk} = \frac{1}{n} \sum_{k=1}^{n} \frac{\lambda_{ijk}}{e_{ij}} = \frac{\widetilde{\lambda}_{ij}}{\widetilde{e}_{ij}} \tag{3.19}$$

In other words, the posterior expected relative risk is equal to the posterior expected mean number of infectives, $\widetilde{\lambda}_{ij}$ divided by the corresponding mean number of infectives based on the human population across all study areas, $\widetilde{e}_{ij}$. The posterior expected number of new infective cases, $\widetilde{e}_{ij}$ can be calculated as

$$\widetilde{e}_{ij} = N_{ij} \frac{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{t} \bar{I}_{ij}}{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{t} \widetilde{\lambda}_{ij}} \qquad (3.20)$$

where N_{ij} is the population of area i at time j. In order to standardize the total

population at risk, it is assume that everybody is equally at risk. Hence, based on

Nor Azah and Percy (2012), relative risk estimation formula can be written as follow

$$\widetilde{\theta}_{ij} = \left(\frac{\widetilde{\lambda}_{ij}}{N_{ij}} \right) \left(\frac{\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{T} \widetilde{\lambda}_{ij}}{\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{T} \bar{I}_{ij}} \right) \qquad (3.21)$$

In this analysis, the relative risk is defined to be the conditional probability

that a person within a region contracts the disease divided by the conditional

probability that a person in the population contracts the disease which this definition

is followed from the study by Nor Azah and Percy (2012). In this analysis, based on

the definition explained by Nor Azah and Percy (2012), a value for the relative risk is

interpreted as in Table 3.1.

71

Table 3.1

Interpretation of Relative Risk Value

Relative Risk	Interpretation
< 1	it shows that there is a decrease in the likelihood of getting the disease, which means that the people within the area are less likely to endure from this disease compared with people in the population.
= 1	no real difference between the conditional probability of a person within the specific region and the general population to contract with the disease. This means that there is no significant difference in terms of the likelihood that the people affected with TB disease in a region and within the whole population.
>1	this shows that people within the area are tending to suffer from this disease compared with people in the population.

This formulation is used in the next chapter to seek the relative risk estimation for disease mapping via stochastic SLIR model utilizing data in the former of counts cases for all tracts under consideration. Then, the performance of relative risk estimation based on the stochastic SLIR model with SMR and Poisson-

gamma model are measured. Based on these results of relative risk, TB risk maps are constructed.

3.3 Data collection

Data retrieved in this study are provided from the Ministry of Health, the Institute for Medical Research and the Department of Statistics in Malaysia. Firstly, the researcher needs to apply for the approval from the NMRR (National Medical Research Registration) website in order to get the approval from MREC (Medical Research Ethics Committee) to request data of TB in each state. The data consist of the number of population, number of TB cases, and number of deaths. Each variable listed is required for each district in 14 states in Malaysia. The duration of this study and data will cover the year 2008 until the year 2015. These data are from the Ministry of Health and the Institute for Medical Research and these data are confidential. Data related to population number in each district in Kedah and the life expectancy in Malaysia are from Department of Statistics. The recovery data are calculated based on previous study by Nor Azah (2012) where

recovery rate = 1/ minimum treatment period for the patient to recover

In application, the data available for analysis is the yearly new infective TB counts for states and districts in Malaysia from the year 2008 until 2015.

3.3.1 Data Set for SLIR Model

In this study, the value for β is calculated based on formula introduced by Porco and Blower (1998) that is $\beta = ECR/N$ where ECR stands for the estimated effective contact rate estimated which also defined as the number of secondary infectious per

73

unit of time produced by an infective in a susceptible community. ECR value is assumed to be the same as that found in other developed countries ($\approx 13\%$ per annually) which is based on Stýblo et al. (1969), Blower et al. (1995) and also has been used by Liao et al. (2012). In this study, N stands for total population size in Malaysia. Hence,

$$\beta = \frac{0.13}{29121054} = 0.000000045$$

Regarding the values for the probability of new infections that develop primary active TB progression (p) and progression rate from latent period to infectious period (v), these are adopt from the previous study by Liao et al. (2012). The values for p and v are chosen to be 0.08 and 0.00392 respectively. For the value of latent infected person is also adapted from Liao et al. (2012), $L = 0.02 \times 0.95 \times N_i$ where 0.02 is adopted from Yeh et al. (2005) and 0.95 is based on 95% CI. While, in this study, N_i is the number of population in each state. Moreover, the yearly rate values for μ, μ_T and c are 0.01417, 0.05411, and 0.86914 respectively. These rates are converted from daily rates to yearly rates which are derived from the Nishiura (2006) and Nor Azah (2012), since the observed counts in this analysis correspond to yearly observations while the rates that were obtained from the literature are daily rates. The following section explains about the calculation of these rates.

3.3.1.1 Converting Daily Rates to Yearly Rates

Assume that μ_D is the daily rate of death, where subscript D represents ―daily‖. In order to understand more on how to convert the daily rate into a yearly rate, it is shown in Table 3.2. The rate is needed to synchronize with the available yearly

observed data of new infective. The basic assumption is that every event (infection or death perhaps) has a constant probability of happening on any given day. Even though it is simple, this assumption at least allows us to make a reasonable conversation if there is no information on yearly rates.

Table 3.2

Converting Daily Rates to Yearly Rates

Day	Daily Deaths	Surviving	Cumulative Deaths
1	μ_D	$(1-\mu_D)$	$1-(1-\mu_D)=\mu_D$
2	$(1-\mu_D)\,\mu_D$	$(1-\mu_D)^2$	$1-(1-\mu_D)^2$
3	$(1-\mu_D)^2\,\mu_D$	$(1-\mu_D)^3$	$1-(1-\mu_D)^3$
4	$(1-\mu_D)^3\,\mu_D$	$(1-\mu_D)^4$	$1-(1-\mu_D)^4$
5	$(1-\mu_D)^4\,\mu_D$	$(1-\mu_D)^5$	$1-(1-\mu_D)^5$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
365	$(1-\mu_D)^{364}\,\mu_D$	$(1-\mu_D)^{365}$	$1-(1-\mu_D)^{365}$

From Table 3.2, it can be concluded that if μ_D is the daily rate of death, then $1-(1-\mu_D)^{365}$ is the yearly rate of death. Assume that the birth rates and death rates are equal in this study. Based on Department of Statistics Malaysia (2017), the birth rates and death rates for human populations per day corresponding to the life expectancy of 75.05 years. Hence,

$$\mu_D = \frac{1}{365\times75.05} = 0.0000365$$

Converting this daily rate to a yearly rate produces

$\mu = 0.01323$

The recovery rate, c_D corresponds to the minimum treatment period of six months for the patient to recover. Therefore,

$$c_D = \frac{1}{(30 \times 6)} = 0.005556.$$

Convert c_D to yearly rate gives

$c = 0.86914.$

For the deaths rate due to TB disease, μ_T, it is calculated by using formula from the U.S Department of Health and Human Service (2012). The cause-specific death rate formula is as follow:

Cause–specific death rate = (number of deaths for a specific cause) ÷ (total population) ×100000

Hence, $\mu_T = 0.05411.$

This information is used to estimate the relative risk for TB disease using stochastic SLIR model. In the next chapter, the outcomes of relative risk estimation based on

the stochastic SLIR model proposed and other existing methods (SMR, Poisson-gamma model and stochastic SIR model) are presented.

CHAPTER FOUR
STOCHASTIC SLIR AND APPLICATION TO RELATIVE RISK ESTIMATION FOR TB DISEASE MAPPING IN MALAYSIA

This chapter discusses the application of relative risk estimation to TB data based on the proposed stochastic SLIR model, SMR, Poisson-gamma model and stochastic SIR model. The data used in the analysis are observed by year (2008-2015) count TB disease data in Malaysia.

4.1 Application of Relative Risk Estimation for TB Disease Mapping

In this section, the results of three applications of existing relative risk estimation methods were demonstrated and displayed. These three methods are Standard Morbidity Ratio (SMR), Poisson-gamma model and the SIR model that was proposed by Lawson (2006). Then, the relative risk estimation outcomes are demonstrated and displayed, based on the proposed SLIR model for TB disease transmission. All of these results are then compared and displayed in tables, graphs and maps so that the best-fitted model for relative risk estimation for TB disease mapping in Malaysia can be found.

Figure 4.1 displays the number of TB cases reported in Malaysia for year 2015. From Figure 4.1, it can be seen that Sabah has the highest number of TB cases while Perlis has the lowest number of cases of TB among others.
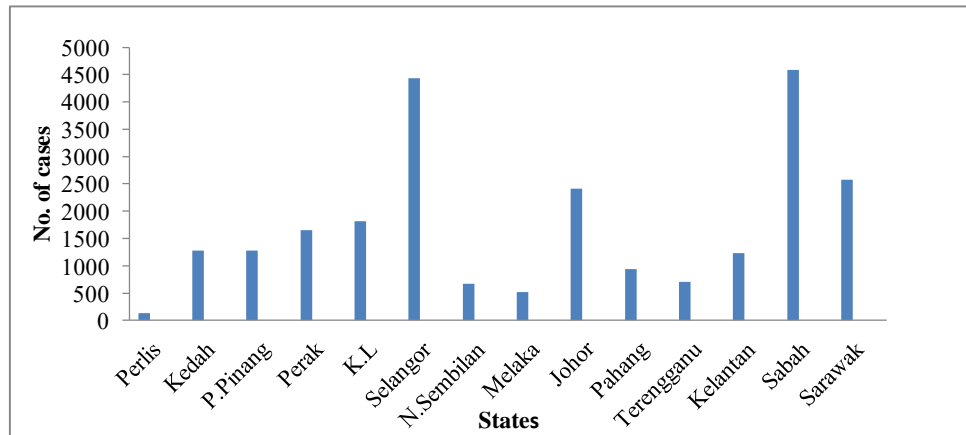
Figure 4.1. Number of TB cases reported for every state in Malaysia in 2015

4.1.1 Relative Risk Estimation Based on Standardized Morbidity Ratio (SMR) Method and Poisson-gamma Model for TB Disease Mapping

The methodologies for method SMR and Poison-gamma model have been discussed earlier in Chapter 2. This section presents the relative risk results for both methods. The WinBUGS code for SMR method and Poisson-gamma model are presented in Appendix C. Table 4.1 shows the statistical output of relative risk estimation of TB data using WinBUGS based Poisson-gamma model.

Table 4.1

Output for Posterior Expected Relative Risks in the State of Perlis, Malaysia based on Poisson-gamma Model for the Year 2008 until 2015

Node	Mean	Sd	MC error	2.5%	Median	97.5%
theta[1,1]	0.9007	0.07455	0.001719	0.7635	0.8982	1.0540
theta[1,2]	0.8136	0.07050	0.001745	0.6841	0.8112	0.9555
theta[1,3]	0.7765	0.06791	0.001188	0.6481	0.7738	0.9156
theta[1,4]	0.8406	0.06703	0.001275	0.7172	0.8386	0.9806
theta[1,5]	0.9882	0.06915	0.001698	0.8529	0.9869	1.1290
theta[1,6]	0.8247	0.06385	0.001486	0.7034	0.8230	0.9540
theta[1,7]	0.611	0.05255	0.001183	0.5144	0.6090	0.7211
theta[1,8]	0.6736	0.05640	0.001168	0.5719	0.6716	0.7862

Table 4.2 and Table 4.3 show the numerical values for the relative risk estimation

based on SMR method and Poisson-gamma model.

Table 4.2

Relative Risk Estimation based on SMR Method and Posterior Expected Relative Risk based on the Poisson-gamma Model for the Year 2015

States	SMR	Poisson-gamma
Perlis	0.653	0.674
Kedah	0.768	0.770
Pulau Pinang	0.962	0.961
Perak	0.826	0.827
Kuala Lumpur & Putrajaya	1.229	1.226
Selangor	0.964	0.964
N. Sembilan	0.766	0.770
Melaka	0.747	0.751
Johor	0.845	0.845
Pahang	0.724	0.727
Terengganu	0.774	0.776
Kelantan	0.904	0.906
Sabah (including Labuan)	1.607	1.603
Sarawak	1.220	1.219

Table 4.3

Relative Risk Estimation based on SMR Method and Posterior Expected Relative Risk based on the Poisson-gamma Model for the Year 2015 in Kedah

States	SMR	Poisson-gamma
Baling	0.805	0.838
Bandar Baharu	0.852	0.907
Kota Setar	1.400	1.380
Kuala Muda	0.816	0.824
Kubang Pasu	0.938	0.945
Kulim	0.756	0.774
Langkawi	1.355	1.283
Padang Terap	0.976	0.988
Pendang	1.230	1.189
Sik	0.953	0.968
Yan	0.818	0.864

In this study, the prior parameter value, α and b for the Poisson-gamma model are unknown but they are assumed to have exponential prior distribution with hyper parameter values of 0.1 as suggested by Lawson et al. (2003). The prior expected relative risks are all equal to 1. A gamma prior distribution is used in this analysis as this distribution is the conjugate prior for the Poisson distribution. This distribution is the most common used for Poisson distribution.

It can be seen from Table 4.2, the estimation by using SMR model shows susceptible people within the state of Sabah have the highest risk with 1.607, while susceptible people within the state of Perlis have the lowest risk of contracting TB when compared with people in the overall population with 0.653. Based on the Poisson-gamma model, susceptible people within the state of Sabah shows the highest risk, while Perlis have the lowest risk with 1.603 and 0.674, respectively.

Table 4.3 shows the relative risk estimation based on districts in Kedah for the year 2015. It can be seen that susceptible people within the district of Kulim for both methods have the lowest risk with 0.756 and 0.774, respectively, while susceptible people within the district of Kota Setar have the highest risk of contracting the disease for both methods with 1.4 and 1.38, respectively.

Results by both methods presented in Table 4.2 give similar conclusions to those results illustrated in Figure 4.2 and Figure 4.3. Federal of Kuala Lumpur and Putrajaya, the states of Kelantan, Sarawak and Sabah (including Labuan) are more likely to contract with the disease compared to people within the overall population,

while the others states are less likely to contract TB. Subsection 4.2.4.1 displays these results in a map that show high-low risk areas of TB occurrences.

The results of relative risk estimation for both models are shown in Figure 4.2 and Figure 4.3. Both graphs show that most states have relative risk less than one for the year 2008 until 2015. This is a necessary consequence of the positively skew distribution inherent of the positive valued relative risk.
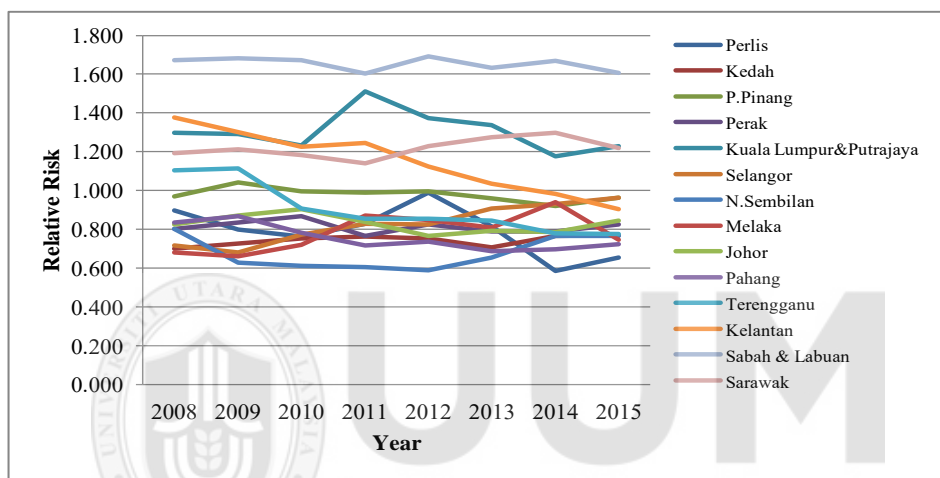
Figure 4.2. Time series plots of the relative risk estimation based on the SMR method for different states in Malaysia
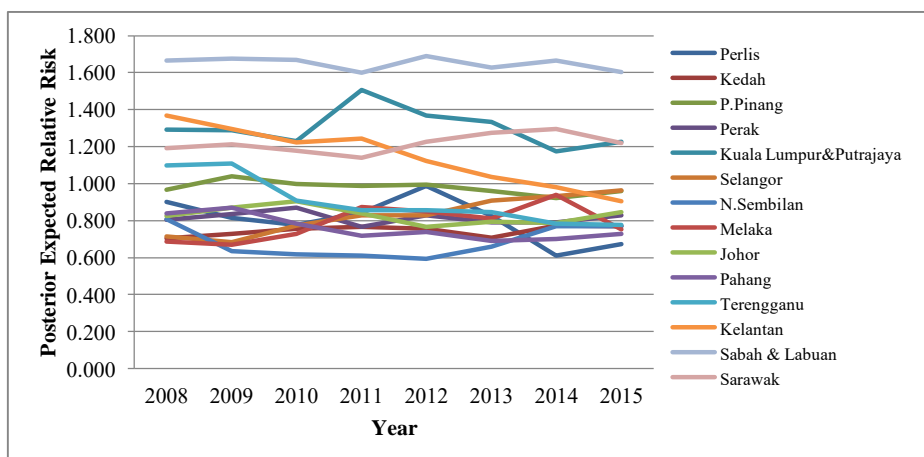
Figure 4.3. Time series plots of the relative risk estimation based on the Poisson-gamma model for different states in Malaysia.

Based on the definition of relative risk estimation values in Chapter 3, if the relative risk value is less than 1, it shows that there is a decrease in the likelihood of getting the disease, which means that susceptible people within the region are less likely to contact with this disease compared with people in the population. Conversely, for a value of relative risk greater than 1, this indicates that people within the region tend to suffer from this disease compared with people in the overall population in Malaysia. It can be seen clearly from the graph in Figure 4.2 and Figure 4.3 that the federal of Kuala Lumpur and Putrajaya, the states of Kelantan, Sarawak and Sabah (including Labuan) have a relative risk greater than 1 for most epidemiology years.

4.1.2 Relative Risk Estimation based on Stochastic SIR Model for TB Disease Mapping

The methodology to estimate the relative risk based on the stochastic SIR model has been explained in Section 2.5 in Chapter 2 which this model was proposed by Lawson (2006). The WinBUGS code for Stochastic SIR model is presented in Appendix C. This stochastic SIR model is apply to analyze TB data for 14 states in Malaysia. The output for the relative risk estimation based on the stochastic SIR model has been represented in Table 4.4, specifically for the epidemiology year 2015.

Table 4.4

Posterior Expected Relative Risks based on Stochastic SIR Model for the Epidemiology Year 2015

State	Posterior Expected Relative Risk
Perlis	0.279
Kedah	0.875
Pulau Pinang	1.032
Perak	0.894
Kuala Lumpur & Putrajaya	1.316
Selangor	1.094
Negeri Sembilan	0.862
Melaka	1.071
Johor	0.893
Pahang	0.769
Terengganu	0.838
Kelantan	1.047
Sabah	1.867
Sarawak	1.463

From Table 4.4, the number of states that have posterior expected relative risks less than one and greater than one is equal in the epidemiology year 2015. Sabah has the highest risk area of contracting TB with 1.867 and Perlis has the lowest risk value of contracting TB with 0.279 when compared with people in the overall population in Malaysia for the epidemiology year 2015. Based on the results shown in Table 4.4, a map is construct to give a clear picture of high-low risk areas

of TB occurrences which are display in Subsection 4.2.4.1. The relative risk estimation results based on the stochastic SIR model have been presented in Figure 4.4.
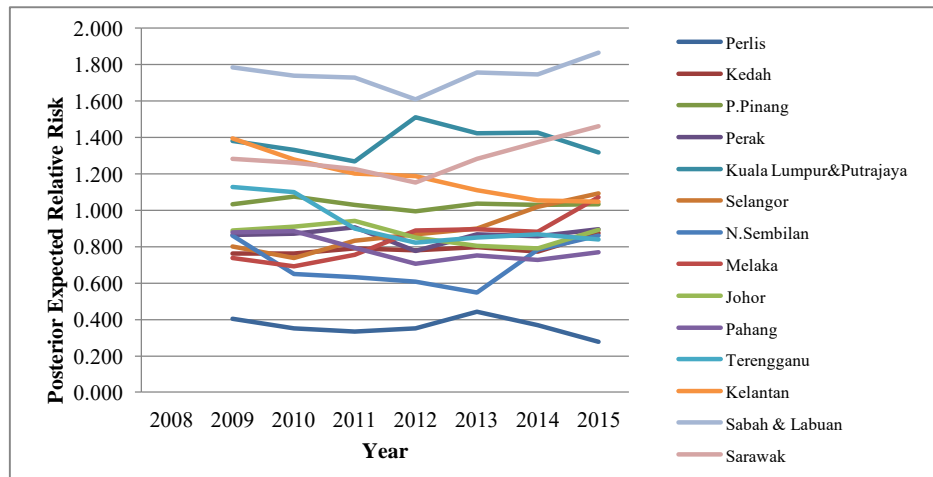
Figure 4.4. Time series plots of the relative risk estimation based on the stochastic SIR model for 14 states in Malaysia

Based on Figure 4.4, it can be seen that there are no posterior expected relative risk values for the epidemiology year 2008. This is because of the situation called moving average where the results are calculated by averaging a number of past data points. Figure 4.4 shows that most of the states have relative risk less than one for all epidemiology years except for the states of Pulau Pinang, Kuala Lumpur, Kelantan, Sabah and Sarawak. This give similar general conclusion as for results of relative risk estimation based on SMR method and Poisson-gamma model described above, where susceptible people in the states of Pulau Pinang, Kuala Lumpur, Kelantan, Sabah and Sarawak states are more tend to contract TB while the others are less tend to contract TB compared to overall population in Malaysia.

85

4.1.3 Relative Risk Estimation based on Stochastic SLIR Model Proposed for

TB Disease Mapping

4.1.3.1 WinBUGS Code for Estimation of Relative Risk based on the Stochastic

SLIR Model

Figure 4.5 shows the WinBUGS code to find relative risk based on the stochastic

SLIR model explained in Section 3.1.2 in Chapter 3 in this chapter.

```
#NEW MODEL -Stochastic newIh

Model{

for (i in 1:M){
Sh[i,1]<-Nh[i]-Lh[i,1]-Ih[i,1]-Rh[i,1]
Lh[i,1]<-0.019*Nh[i]
Ih[i,1]<-0.00003147*Nh[i]
Rh[i,1]<-0.00003147*Nh[i]
}

for (i in 1:M){
for (j in 2:T){

Sh[i,j]<-(muH*Nh[i])+(1-muH-(betaH*Ih[i,j-1]))*Sh[i,j-1]
Lh[i,j]<-((1-p)*(betaH*Ih[i,j-1])*Sh[i,j-1])+(1-v-muH)*Lh[i,j-1]
Ih[i,j]<-newIh[i,j] +(v*Lh[i,j-1])+(1-muH-muT-c)*Ih[i,j-1]
newIh[i,j]~dpois(lambdanewH[i,j])
log(lambdanewH[i,j])<-beta0+log(betaH*p)+log(Ih[i,j-1]+0.001)+log(Sh[i,j-1]+0.001)+bh[i]
Rh[i,j]<-(1-muH)*Rh[i,j-1]+newRh[i,j]
newRh[i,j]<-(c*Ih[i,j-1])

#RELATIVE RISK

RRH[i,j]<-lambdanewH[i,j]/eH[i,j]
}
}

#CAR prior distribution for random effects bh. The sum of bh is always 0

bh[1:14]~car.normal(adjH[], weightsH[], numH[], varbh)
for (k in 1:SumNumNeighH){weightsH[k]<-1}

#Other priors

beta0~dflat()                    #Flat prior for the intercept
varbh~dgamma(0.01,0.01)          #Prior on precision for spatial random effect bh
c<-0.86914

}
```

Figure 4.5. Stochastic SLIR model in WinBUGS

It can be seen clearly from the WinBUGS code in Figure 4.5 that the CAR prior distribution is specified using the car.normal function for the random effects. The description of notations used in the arguments to this function in Figure 4.5 are as follow:

adjH[]: a vector listing the identification (ID) numbers of the adjacent area for each area. This is a sparse representation of the full adjacency matrix for the study region.

weightsH[]:a vector the same length as adjH[] giving unnormalised weights associated with each pair of areas where ‗weightsH[]‘equal to 0 if areas i and j are not neighbours, while ‗weightsH[]‘equal to 1 if they are neighbours. A common choice to set all the ‗weightsH[]‘ equal to 1 since it gives the standard CAR model (Besag, York & Mollie, 1991).

numH[]:   the number of adjacent areas.

varbh: a scalar argument corresponding to the prior on precision for spatial random effects.

SumNumNeigh: the length of adjH.

The output of WinBUGS based on the stochastic SLIR model which leads to relative risk estimation are presented and discussed in detail. In WinBUGS, diagnostics for assessing convergence of the MCMC method can be done by examining the ‗history‘ plot of the samples at each iteration and looking for random scatter. A ‗history‘ plot from this analysis is presented in Figure 4.6. The output of WinBUGS analysis refer to the results based on stochastic SLIR model for epidemiology year 2010 until the epidemiology year 2015 for the states of Sabah

(including Labuan). These results include posterior summaries, quantiles graph and posterior densities. A complete WinBUGS output for the summary statistics for this estimation for all states can be found in Appendix B.
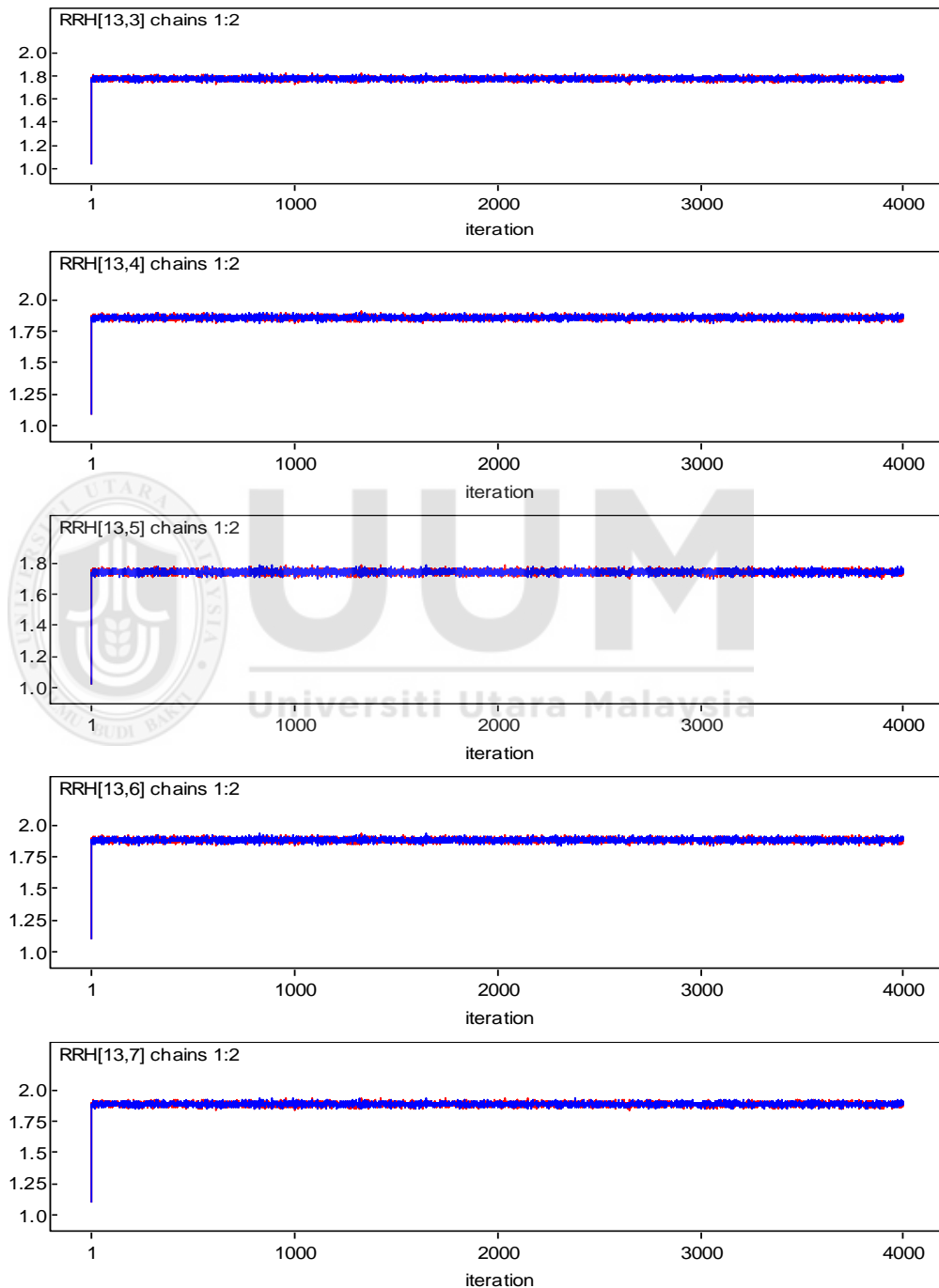
Figure 4.6. Example of WinBUGS output of the ‗history' plot for convergence of the relative risk estimation based on stochastic SLIR model.

Some example of WinBUGS output for the posterior summaries of the relative risk estimation based on stochastic SLIR model is shown in Table 4.5.

Table 4.5

Output of WinBUGS Results for Posterior Summaries of Relative Risk Estimation based on Stochastic SLIR Model for the State of Sabah from the Epidemiology Year 2010 until the Epidemiology Year 2015.

node	mean	sd	MC error	2.5%	median	97.5%
RRH[13,2]	0.053	5.307E-4	5.272E-6	0.0524	0.053	0.054
RRH[13,3]	1.775	0.01771	1.759E-4	1.749	1.775	1.801
RRH[13,4]	1.856	0.01852	1.84E-4	1.829	1.857	1.884
RRH[13,5]	1.741	0.01737	1.726E-4	1.715	1.741	1.767
RRH[13,6]	1.881	0.01877	1.864E-4	1.853	1.881	1.909
RRH[13,7]	1.885	0.01881	1.869E-4	1.858	1.885	1.913
RRH[13,8]	2.015	0.02010	1.997E-4	1.985	2.015	2.044

Based on Table 4.5, the terms used can be summarized as follow:

node : the name of the unknown quantity or interest variable. In this study, estimation of relative risk is the quantity that the researcher interested.

mean : the average of the simulations, which is an approximate mean of the unknown quantity, which is in this study, this is acknowledged as the posterior expected relative risk.

sd : the standard deviation of the simulations which is an approximation of the standard deviation of the posterior distribution.

MC error : short form for Monte Carlo error which is the computational accuracy of the mean. In order to get MC error as small as desired, this can be made by increasing the number of simulations.

2.5% : the 2.5[th] percentile of the simulations, which is approximate of the lower

endpoint of the 95% credible interval.

median : the 50th percentile of the simulations.

97.5% : the 97.5th percentile of the simulations which is an approximation of

the upper endpoint of the 95% credible interval.

The values of the mean, 2.5th percentile and 97.5th percentile of 95 % credible interval are summarized by WinBUGS results of the analysis as a quantiles graph as displayed in Figure 4.7.

From Table 4.5, it can be seen clearly that the mean and median values are approximately the same. This shows that the posterior distribution density is reasonably symmetric. An example of the WinBUGS results for the posterior densities of the relative risk estimation based on the stochastic SLIR model is shown in Figure 4.8.
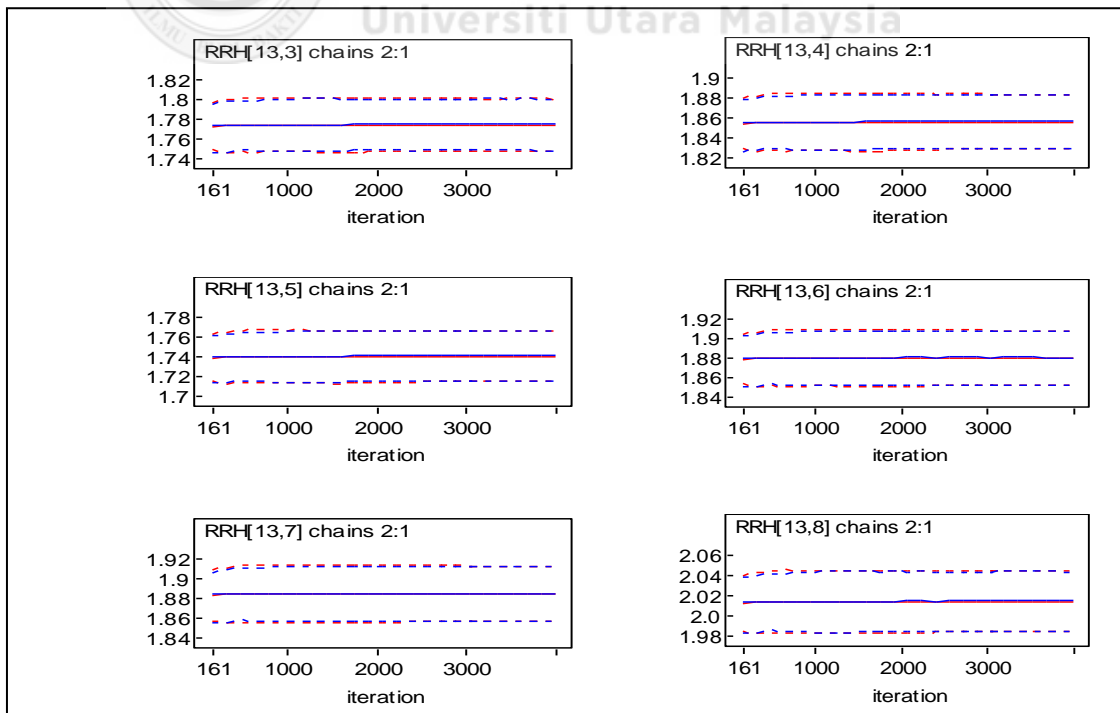
Figure 4.7. Example of WinBUGS output of the quantiles graph of the relative risk estimation based on stochastic SLIR model.
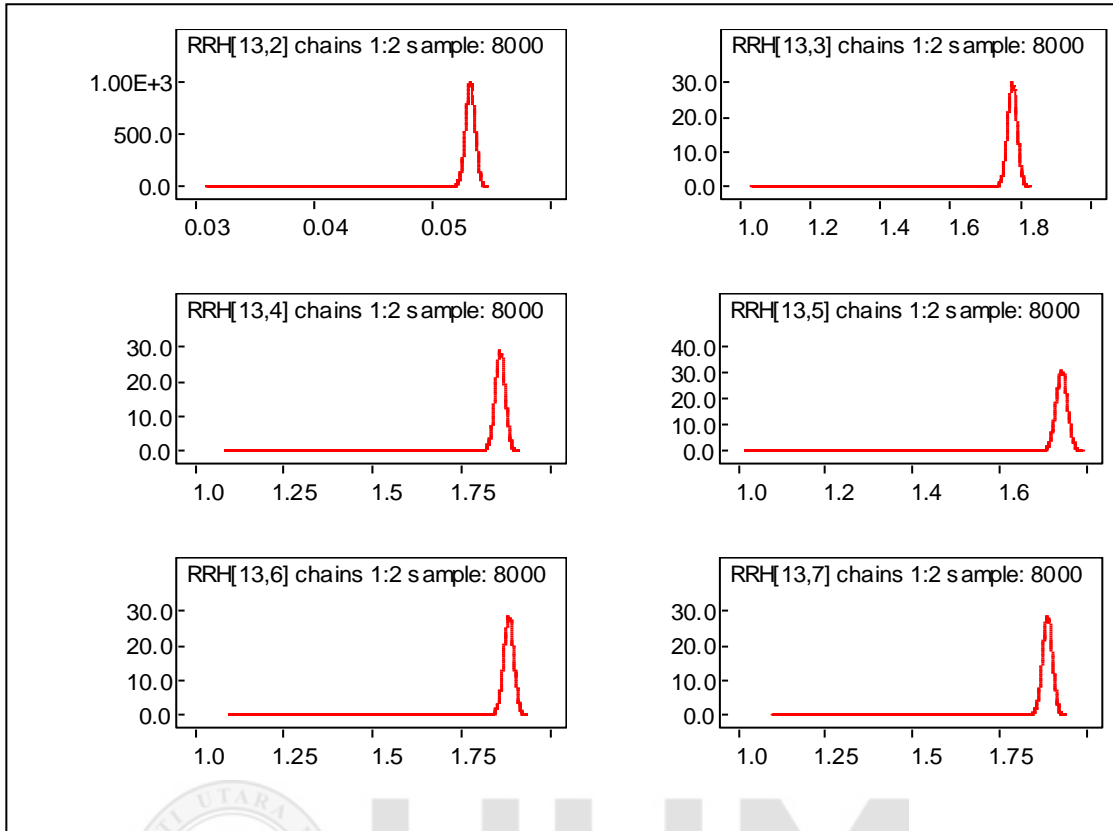
Figure 4.8. Example of WinBUGS output of the posterior densities of the relative risk estimation based on stochastic SLIR model.

4.1.3.2 Results of Relative Risk Estimation based on Stochastic SLIR model

The results of relative risk estimation based on the stochastic SLIR model displayed in Figure 4.9. Based on the Figure 4.9, it is clearly seen that there are no posterior expected relative risk values for the epidemiology year 2008. This situation is similar when using stochastic SIR model as this situation is called as moving average as mentioned before in Subsection 4.2.2. Based on the Figure 4.9, most of the states have relative risk less than one for all epidemiology years except for Pulau Pinang, Kuala Lumpur, Kelantan, Sabah and Sarawak. This gives similar general conclusion as for results of relative risk estimation based on SMR method, Poisson-gamma model and also stochastic SIR model as mentioned before. This shows that susceptible individuals in these five states are more probably to get TB while

susceptible individuals within other states are less probably to get TB compared to individuals within the overall population.
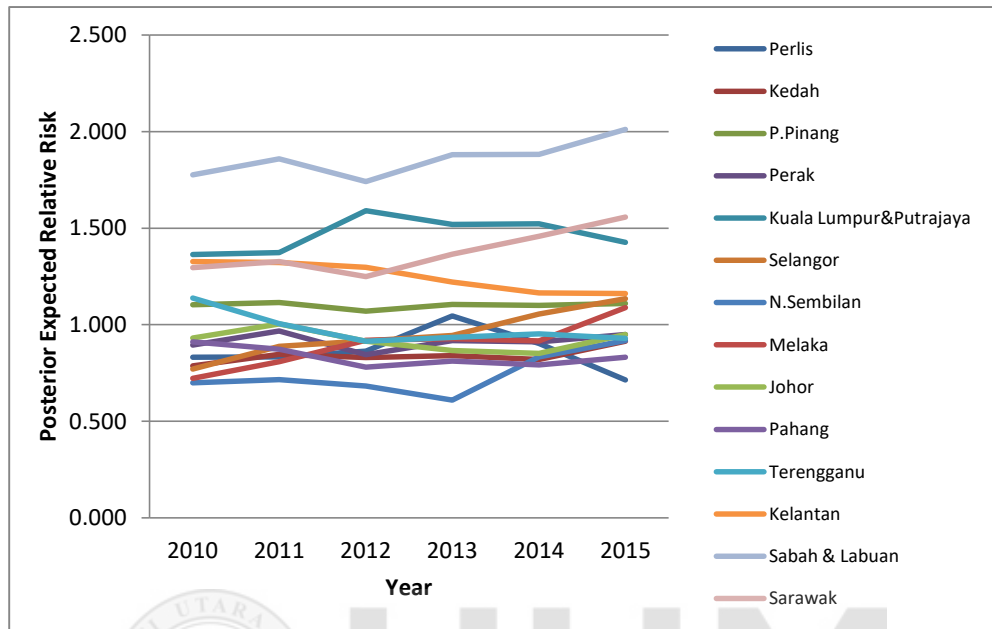
Figure 4.9. Time series plots of the relative risk estimation based on the stochastic SLIR model for 14 states in Malaysia

4.1.4 Comparison of Posterior Expected Relative Risk for TB Disease Mapping based on SMR Method, Poisson-gamma Model, Stochastic SIR Model and Stochastic SLIR Model

The comparison results for estimation of relative risk of TB disease mapping based on four methods that are SMR method, Poisson-gamma model, the stochastic SIR model and stochastic SLIR model are displayed in Table 4.6 and Table 4.7. These results correspond to the 14 states in Malaysia, specifically for the epidemiology year 2014 and 2015.

Table 4.6

Comparison between the Posterior Expected Relative Risks in the Epidemiology Year 2014 based on Four Different Models

	Relative Risk Estimations for TB Disease Mapping			
State	**SMR**	**Poisson-gamma model**	**Stochastic SIR model**	**Stochastic SLIR model**
Perlis	0.586	0.610	0.368	0.9028
Kedah	0.766	0.772	0.772	0.8203
Pulau Pinang	0.920	0.920	**1.028**	**1.102**
Perak	0.788	0.789	0.857	0.912
Kuala Lumpur & Putrajaya	**1.177**	**1.175**	**1.427**	**1.524**
Selangor	0.931	0.931	**1.020**	**1.057**
Negeri Sembilan	0.765	0.768	0.785	0.834
Melaka	0.940	0.941	0.880	0.9163
Johor	0.787	0.788	0.789	0.8537
Pahang	0.696	0.699	0.727	0.7925
Terengganu	0.780	0.783	0.869	0.9532
Kelantan	0.983	0.983	**1.054**	**1.166**
Sabah	**1.668**	**1.664**	**1.747**	**1.885**
Sarawak	**1.297**	**1.294**	**1.373**	**1.46**

Table 4.7

Comparison between the Posterior Expected Relative Risks in the Epidemiology Year 2015 based on Four Different Models

	Relative Risk Estimations for TB Disease Mapping			
State	**SMR**	**Poisson-gamma model**	**Stochastic SIR model**	**Stochastic SLIR model**
Perlis	0.653	0.674	0.279	0.7148
Kedah	0.768	0.770	0.875	0.9157
Pulau Pinang	0.962	0.961	**1.032**	**1.112**
Perak	0.826	0.827	0.894	0.9516
Kuala Lumpur & Putrajaya	**1.229**	**1.226**	**1.316**	**1.429**
Selangor	0.964	0.964	**1.094**	**1.139**
Negeri Sembilan	0.766	0.770	0.862	0.918
Melaka	0.747	0.751	**1.071**	**1.089**
Johor	0.845	0.845	0.893	0.9527
Pahang	0.724	0.727	0.769	0.8338
Terengganu	0.774	0.776	0.838	0.9307
Kelantan	0.904	0.906	**1.047**	**1.163**
Sabah	**1.607**	**1.603**	**1.867**	**2.015**
Sarawak	**1.220**	**1.219**	**1.463**	**1.56**

Based on Table 4.6 and Table 4.7, it can be concluded that by using these four methods, Sabah has the highest risk area of contracting TB and Perlis has the lowest risk value of contracting TB in the overall population for epidemiology year 2014 and year 2015 even though the values from different methods are different. In

this analysis, there is no zero relative risk value when using SMR method as there is observed count data in all states. Based on the results especially from Table 4.6, stochastic SLIR model shows the smallest range of posterior expected relative risk across states when compared with other three methods with a minimum value of 0.9028 and the maximum value of 1.885.

However, the most important is that the model introduced is potentially more suitable than other three methods including SIR model. This is because the stochastic SLIR model that includes a more detailed description of basic biology process and the nature of disease transmission. It also can overcome the drawbacks of the SMR and Poisson-gamma model by allowing adjustment of covariate and spatial correlation between risks in nearby regions. Models that do not consider spatially correlated heterogeneity are less strong (Lawson, 2006).

Based on the results shown in Table 4.6 and Table 4.7, maps are constructed to show a clear picture of low and high risk areas for incidence of TB which are display in Subsection 4.2.4.1. These maps can be used by the government or other interest parties as a tool to recognize which states that need extra attention in term of government policy and financial and also others support in order to control and prevent TB disease from becoming worse. In order to determine which method produces a smoother map, the maps are compared using these four types of method.

4.1.4.1 Disease Maps for the Relative Risk Estimation of TB Disease in Malaysia during Epidemiology Year 2014 and Epidemiology Year 2015

In this section, disease maps are used to present the statistical results for relative risk estimation as discussed in previous sections graphically. In order to facilitate the

classification of risk, the results of relative risk are divided into five different classes which as follow:

Table 4.8

Classes of Relative Risk Estimation

Risk Level	Value
Very low	[0.0,0.5)
Low	[0.5,1.0)
Medium	[1.0,1.5)
High	[1.5,2.0)
Very high	[2.0, ∞)

Choropleth maps (also known as thematic maps) are used with tones of one colour to present and distinguish between the low and high risk areas of TB incidences. The epidemiology year 2014 and year 2015 are chosen as an example and for comparison purpose only.

Figure 4.10 and Figure 4.11 show the disease maps based on the SMR methods. Based on both figures, Sabah has been recognized as a high risk area for the epidemiology year 2014 and 2015, followed by the state of Sarawak and Kuala Lumpur including Putrajaya with medium risk. Another eleven states have been categorized as low risk. There is no state classified as very high and very low risk area based on Figure 4.10 and Figure 4.11.

Figure 4.10. Disease map of relative risk estimation based on SMR method for the year 2014
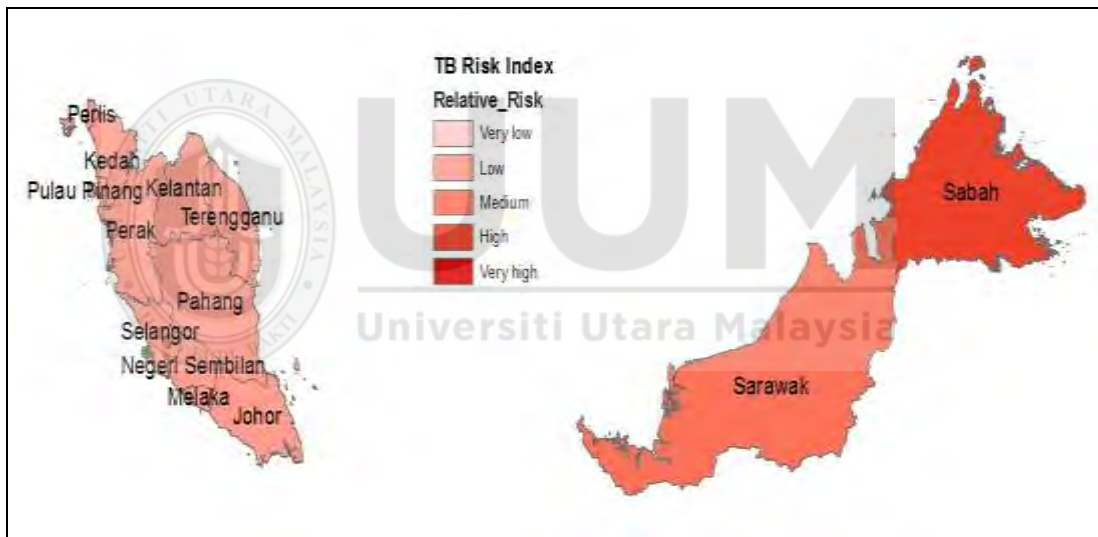
Figure 4.11. Disease map of relative risk estimation based on SMR method for the year 2015

Similarly to those results in the SMR maps in Figure 4.10 and Figure 4.11, the state of Sabah has been classified as a high risk area for both Poisson-gamma maps, with no state classified as very high and very low risk. Sarawak, Kuala Lumpur and Putrajaya show the medium risk while the other states with low risk for both years. Results are shown in Figure 4.12 and 4.13.

Figure 4.12. Disease map of relative risk estimation based on Poisson-gamma
method for the year 2014
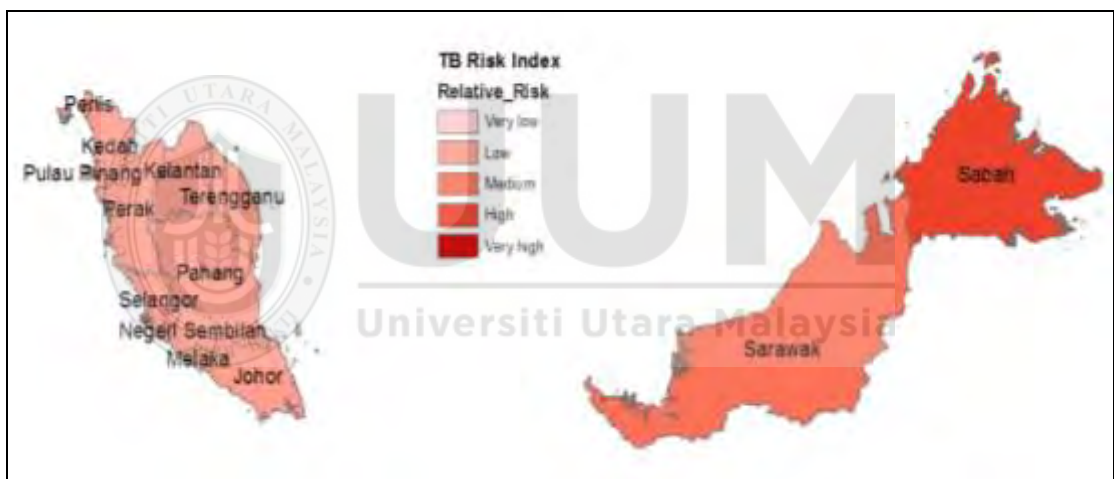
Figure 4.13. Disease map of relative risk estimation based on Poisson-gamma
method for the year 2015

Figure 4.14 and Figure 4.15 present the high-low risk area based on
stochastic SIR model for the epidemiology year 2014 and year 2015. Similar to those
results in the SMR maps and Poisson-gamma maps, Sabah has a high risk for
occurrence of TB for both stochastic SIR maps. These are followed by Pulau Pinang,
Kuala Lumpur and Putrajaya, Selangor, Kelantan and Sarawak with medium risk for
Figure 4.14, while for Figure 4.15 shows the same results as Figure 4.14 for medium

risk, but including Melaka as a medium risk for TB occurrences. The other states

have low risk except for Perlis with very low risk for both stochastic SIR maps.
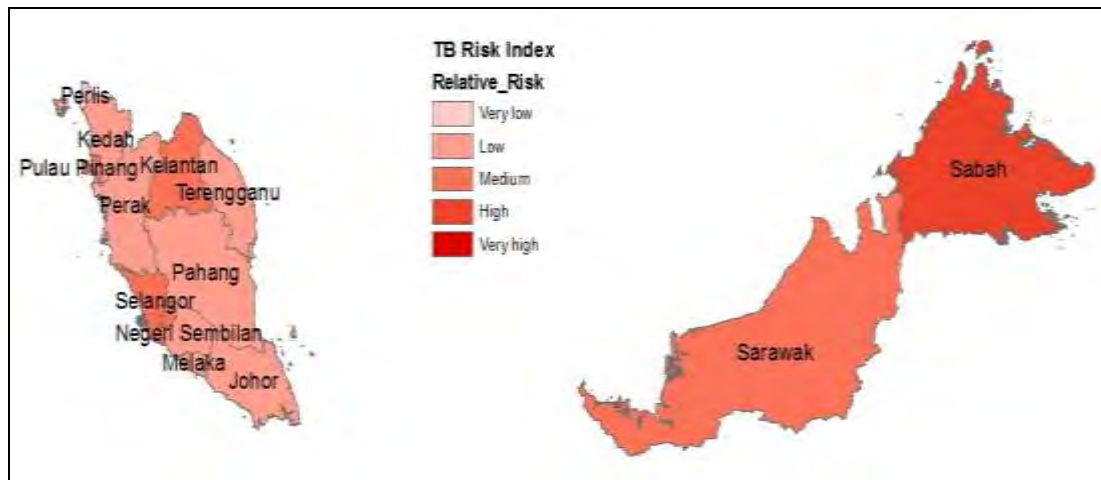
Figure 4.14. Disease map of relative risk estimation based on stochastic SIR model for the year 2014
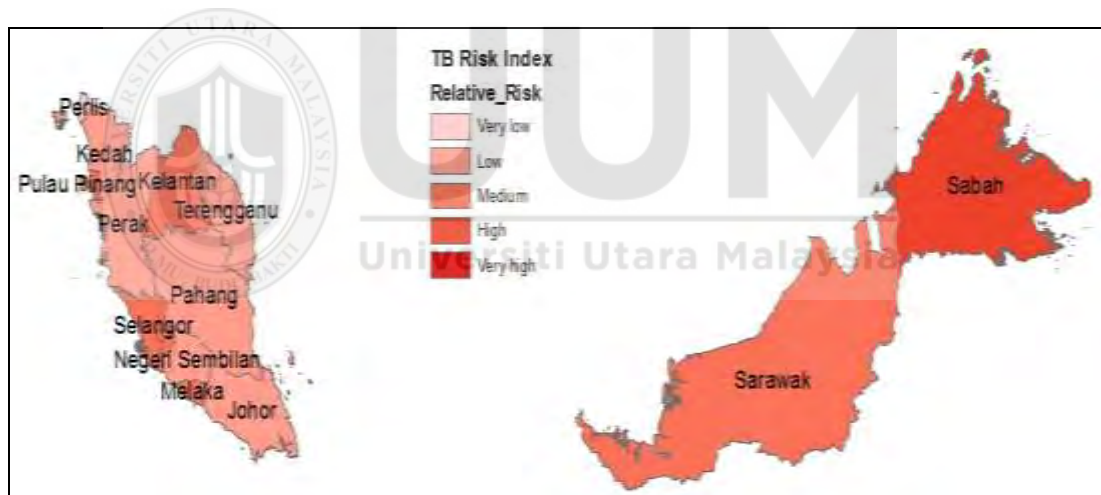
Figure 4.15. Disease map of relative risk estimation based on stochastic SIR model for the year 2015

Figure 4.16 and Figure 4.17 display maps for posterior expected relative risks

based on stochastic SLIR model especially for TB disease transmission. From Figure

4.16, Sabah has been classified as high risk in the epidemiology year 2014, with no

state with very high risk. While in Figure 4.17, Sabah has been recognized as a very

high risk for TB occurrences. Based on Figure 4.17, people in Sarawak has a high

risk for contracting TB in the year 2015 while in the year 2014, Sarawak has been categorized as medium risk. Then, followed by the state of Pulau Pinang, Kuala Lumpur and Putrajaya, Selangor, Melaka and Kelantan with medium risk for both stochastic SLIR maps. The other seven states for Figure 4.16 and Figure 4.17 are classified as low risk, with no state classified as a very low risk.
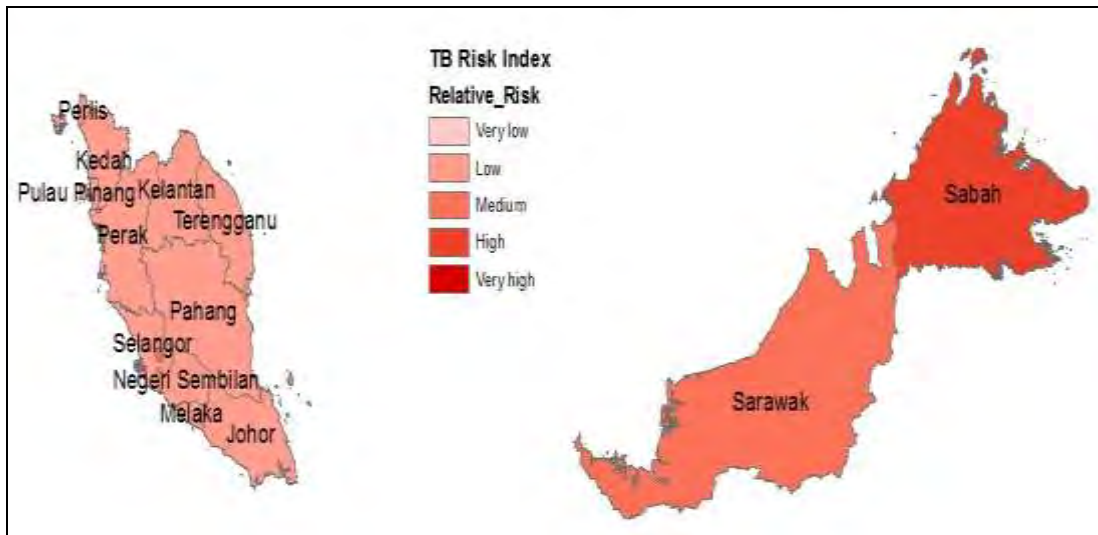

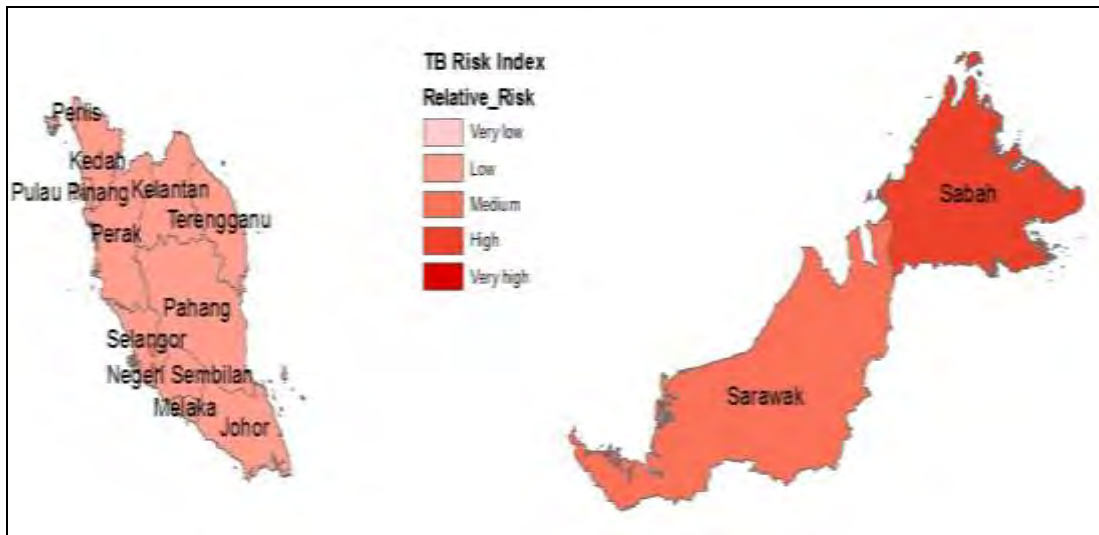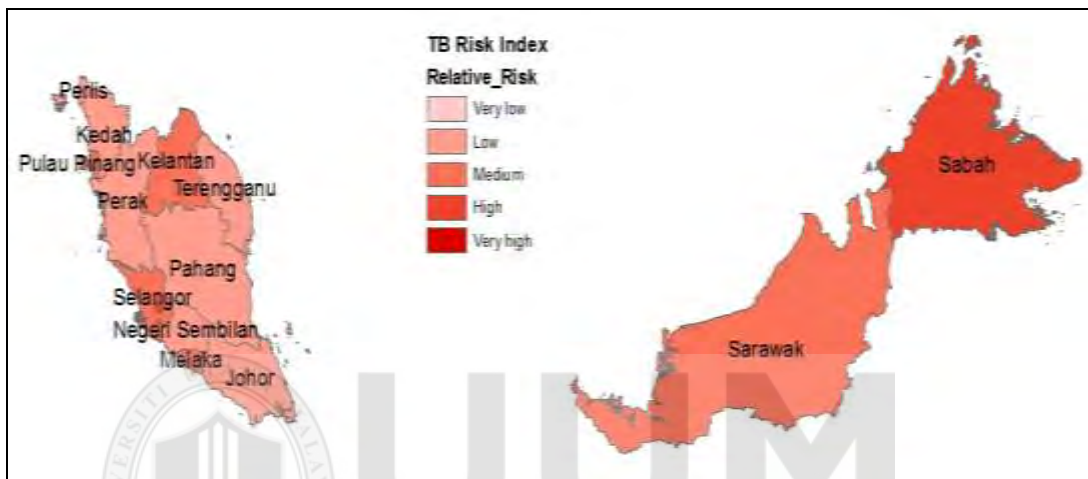*Figure 4.16.* Disease map of relative risk estimation based on stochastic SLIR model for the year 2014
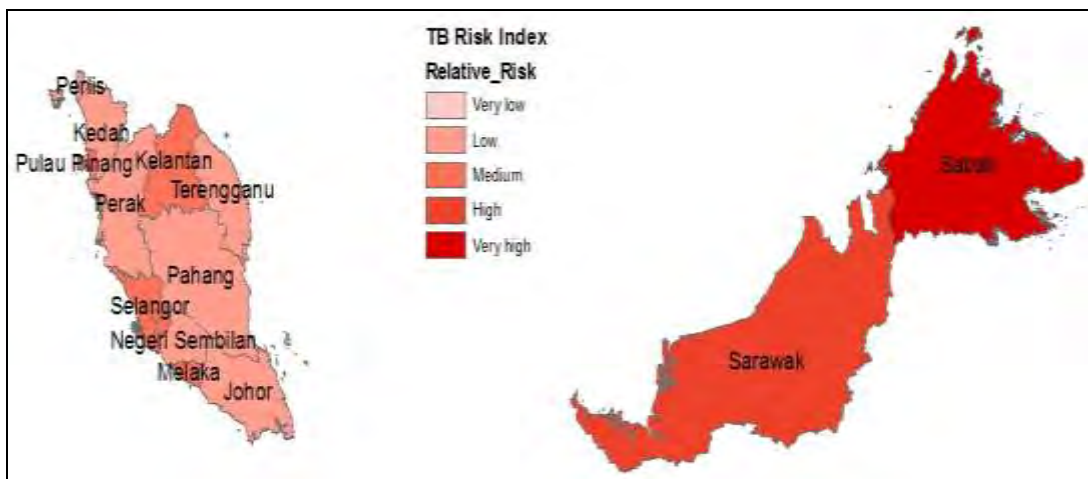
Figure 4.17. Disease map of relative risk estimation based on stochastic SLIR model for the year 2015

Based on the maps comparison above using these four methods (SMR, Poisson-gamma model, stochastic SIR model and stochastic SLIR model), there are no obvious differences in the estimated risks. However, the numerical analysis considered in the previous section can be used for purposes of inference, while the map is primarily intended to be a good tool performance to identify high risk areas so that more consideration can be given to the prioritized affected states. Here, it cannot easily conclude whether any map is smoother or more precise maps than others, but based on the details of the model described earlier, the evidence continues to support the stochastic SLIR model that it is better in estimate the relative risk. This stochastic SLIR model offers a better way of describing the underlying behaviours as well as estimating the relative risk compared to SMR, Poisson-gamma and stochastic SIR models. Hence, this model should be used to predict the risks of TB disease across the 14 states of Malaysia for short term forecasting. This method can be applied to others disease which have latent period in the transmission of the disease.

4.2 Discussion

Table 4.9

Posterior Expected Relative Risk based on Four Different Methods for the States with Value of Relative Risk More Than One

	Relative Risk Estimations for TB Disease Mapping			
State	**SMR**	**Poisson-gamma model**	**Stochastic SIR model**	**Stochastic SLIR model**
Pulau Pinang	0.962	0.961	**1.032**	**1.112**
Kuala Lumpur & Putrajaya	**1.229**	**1.226**	**1.316**	**1.429**
Selangor	0.964	0.964	**1.094**	**1.139**
Melaka	0.747	0.751	**1.071**	**1.089**
Kelantan	0.904	0.906	**1.047**	**1.163**
Sabah	**1.607**	**1.603**	**1.867**	**2.015**
Sarawak	**1.220**	**1.219**	**1.463**	**1.56**

The value of relative risk is increasing when using SLIR model for all states in Malaysia. Table 4.9 shows the list of states with value of relative risk more than 1. From Table 4.9, it is clearly seen that the value increase when using SLIR model. This is because of the added element, the latent, when using SLIR model besides it considers the infectious element. Compared to other three methods that are SMR, Poisson-gamma model and stochastic SIR model, these three methods only consider infectious element. This finding is comparable with input by Immigration Department where the number of illegal immigrant comes from those states in Table 4.9. As mentioned by Baussano, Mercadante, Pareek, Lalvani, and Bugiani (2013) location with high illegal immigrant will have higher risk of having TB prevalence.

Based on the all results as provided in Chapter 4, it can be concluded that Sabah has the highest risk of TB occurrence while Perlis is the lowest risk area of TB occurrence. According to the Malaysian Health Minister, Datuk Seri Dr. S. Subramaniam, one of the contributing factors for the rise of TB infection among locals in the country is the influx of foreigners ("One TB patient can potentially infect 10 people", 2014). These foreigners consist of refugees, migrant workers and even illegal migrants, who are not documented in the national statistics of foreigners in the country. This is because it is difficult to detect the bacteria as illegal immigrants were not screened before entering the country. Sabah recorded the highest number of illegal immigrants in Malaysia (Arvin & Nurfazlina, 2015). Since the distance of Sabah is so near to Philippines and Indonesia, this makes illegal immigrants easier to enter the state either using the land or sea route. Most of the foreign workers in Sabah come from these countries (The Philippines and Indonesia). Besides there also employed immigrants from others countries such as India,

Pakistan and China in Sabah (Arvin & Nurfazlina, 2015). Malaysia Health Ministry director of disease control division, Dr. Chong Chee Keong said that the Philippines and Indonesia have been identified as higher TB burden countries and these countries contribute to ‗TB migration' to our country (Benedict, 2014). There is a possibility that they were the cause of the increase risk of TB disease in Malaysia. Besides that, according to Arvin and Nurfazlina (2015), many illegal immigrants deported have family in Sabah. This is also one of the factors that they re-enter Sabah and perpetuate chain migration. Compared to other states, Perlis has the lowest risk of TB occurrence even though it is also located near another country such as Thailand. This may be because of the strict immigration to enter the state and less number of foreign workers.

All models mentioned before which are SMR, Poisson-gamma model and stochastic SIR model can produced maps. However, the map produced using the proposed model that is stochastic SLIR model has its advantage. From Figure 4.17 which is TB risk map based on stochastic SLIR method for the epidemiology year 2015, from the tones of the colour, it shows a large gap between the risk compared to others map using different methods since it includes extra information in the model such as the spatial correlation between the areas consider the latent component. Spatial correlation is one of the elements that need to be considered since the states are located side by side. Hence, the risk may be transferred to other states. In the next chapter, conclusion about the findings of these relative risks is provided and some suggestions are proposed for further research are also included.

CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

In this chapter, conclusions are made based on the application of the alternative method that was proposed that is stochastic SLIR model to estimate the relative risk of direct transmission infectious disease with a latent period. The research focuses on TB disease in Malaysia. Besides that, some suggestions are included in this chapter so that further research can be held in the future.

5.1 Conclusion

Previous studies showed that none of them include latent element in order to estimate the relative risk. Hence, stochastic SLIR model is developed which this model is based on the transmission model of the disease. This stochastic SLIR model represents current TB scenario compared to existing methods as it included latent element.

In order to have a good disease mapping, it is a must to get good relative risk estimation. In this study, several relative risk estimation methods have been discussed. Most of the new models that have been developed are driven by the commonest models used in finding the relative risk which is the SMR method and the earliest of Bayesian methodology that is Poisson-gamma model. There is also stochastic SIR model proposed by Lawson (2006). However, these three methods have drawbacks. SMR cannot detect the small areas because it does not take into account the high diversity of different areas it can yield large changes in estimate with relatively small changes in expected value since it is based on ratio estimation.

While, Poisson-gamma model does not allow the adjustment of the covariate and incapability to deal with spatial correlation between risks in adjacent areas. Stochastic SIR model does not consider latent component in the model. Hence, in order to overcome the drawbacks of these three methods, stochastic SLIR model is used to estimate the relative risk for TB disease.

Based on the relative risk comparison results, when using the proposed stochastic SLIR model, the value of the relative risk increases compared when using SMR method, Poisson-gamma model and stochastic SIR model. This is because stochastic SLIR model includes two elements, infectious and latent element, while the other three methods just consider infectious element only. Previous method showed that the area has a high risk, but when using stochastic SLIR model, it becomes very high risk as for example for the state of Sabah. This also can be proved by the maps obtained in Chapter 4 as the map of relative risk based on stochastic SLIR model shows large gap of colour compared to the other maps using other three methods. From the map, it can be seen clearly which areas need more attention from the responsible authorities.

5.2 Future Research

There are few suggestions for the further research of this study. The first suggestion is to investigate the transmission of TB disease based on other factors such as age and gender. Hence, future studies might generate a disease model transmission based on age and gender.

Another suggestion is to assign stochastic terms to other compartmental variables. In this research, the only stochastic term is new infective persons in the

SLIR model for direct infectious disease transmission. For future studies, stochastic terms can be assigned to new infective persons from latently infected state to infective state, in order to study the stochasticity of these variables besides improving the measurement of the relative risk.

Next, as the findings from the study showed that Sabah has the highest risk area, for possible extensions to this study could include the estimation of relative risk based on the stochastic SLIR model for the districts in Sabah in order to know which district in Sabah has critical area.

Besides that, in terms of application of the proposed methodology, these proposed stochastic SLIR models can be applied to other direct infectious diseases that have a similar biological transmission as TB disease, such as rubella. This lead to further evaluations and new study on how different types of disease data contribute to the proposed models.
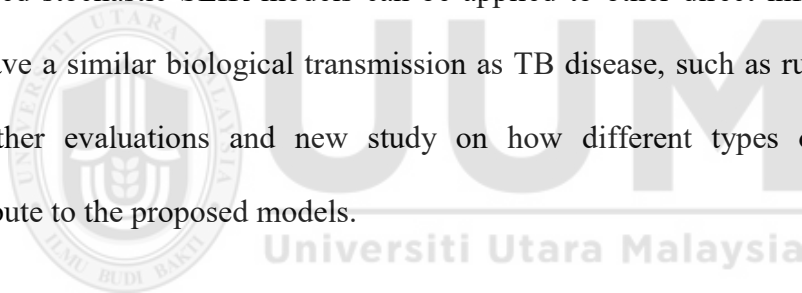
REFERENCES

Abdul Karim Iddrisu and Amoako, Y. A. (2016). Spatial Modeling and Mapping of Tuberculosis Using Bayesian Hierarchical Approaches. *Open Journal of Statistics*, *6*(June), 482–513. doi:10.4236/ojs.2016.63043

Abramson, G. (2001). Mathematical modeling of the spread of infectious diseases. *A series of lectures given at PANDA, University of New Mexico*. Retrieved from https://pdfs.semanticscholar.org/938d/a251d4b0152d7f27175d54a0e595dc525 8d9.pdf

Achterberg, J. T. (2009). *Computational models frame new and outstanding biological questions of Mycobacterium tuberculosis* (Working Paper No. 93). Retrieved from https://www.csss.washington.edu/files/working-papers/2009/wp93.pdf

Ai, J.-W., Ruan, Q.-L., Liu, Q.-H., & Zhang, W.-H. (2016). Updates on the risk factors for latent tuberculosis reactivation and their managements. *Emerging Microbes & Infections*, *5*(1), e10. doi:10.1038/emi.2016.10

Aparicio, J. P., Capurro, A.F., Castillo-Chavez, C. (2000). Transmission and dynamics of tuberculosis on generalized households. *Journal of Theoretical Biology*. 206:327-341.

Apenteng, O. O. (2013). Epidemiological cellular automata : A case study involving AIDS. Master thesis, Auckland University of Technology, New Zealand.

Aron, J. L. (2007). Mathematical Modeling Dynamics of Infection. PhD thesis, Johns Hopkins University, Maryland.

Arvin Tajari and Nurfazlina Affendi. (2015). Illegal Immigrant and Security Crisis in Sabah (Malaysia). *International Conference on Social Science Research, ICSSR*, *2015*(June), 1–13.

Aznida Che Awang. (2016). The Stochastic SIR-SI Age-Structured Model for Leptospirosis Mapping in Malaysia. MSc thesis, Universiti Pendidikan Sultan Idris, Tanjung Malim. Perak, Malaysia.

Baussano, I., Mercadante, S., Pareek, M., Lalvani, A., and Bugiani, M. (2013). High Rates of Mycobacterium tuberculosis among Socially Marginalized Immigrants in Low-Incidence Area, 1991–2010, Italy. *Emerging Infectious Diseases*, *19*(9), 1437-1445. https://dx.doi.org/10.3201/eid1909.120200.

Beggs, C. B., Noakes, C. J., Sleigh, P.A., Fletcher, L.A., and Siddiqi, K. (2003). The transmission of tuberculosis in confined spaces: an analytical review of alternative epidemiological models. *Int J Tuberc Lung*, *7*(11), 1015-1026.

Berke, O. (2004). Exploratory disease mapping: kriging the spatial risk function from regional count data. *International Journal of Health Geographics*, *3*, 18. doi:10.1186/1476-072X-3-18.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, *14*(21-22), 2433-2443.

Benedict, N. (2014, January 15).TB kills more annually than dengue, says ministry. *The Malay Mail Online*. Retrieved from http://www.themalaymailonline.com/malaysia/article/tb-kills-more-annually-than-dengue-says-ministry

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*, 1-59.

Bhowmik, D., Chandira, R. M., and Pradesh, U. (2009). Recent trends of drug used treatment of tuberculosis. *Journal of Chemical and Pharmaceutical Research*, *1*(1), 113–133.

Blower, S. M., Mclean, A.R., Porco, T.C., Small, P. M., Hopewell, P.C., Sanchez, M. A., Moss, A. R. (1995). The intrinsic transmission dynamic of tuberculosis epidemics. *Nature Medicine*, 1, 815-821.

Blower, S. M., Small, P. M., Hopewell, P. C. (1996). Control strategies for tuberculosis epidemics new models for old problems. *Science*, 273, 497-500.

Brogger, S. (1967). Systems analysis in tuberculosis control: A model. *American Review of Respiratory Disease*. 95, 419-434.

Clement, E. P. (2014). Small Area Estimation with Application to Disease Mapping. *International Journal of Probability and Statistics*, *3*(1), 15–22. doi:10.5923/j.ijps.20140301.03.

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

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons, Inc.

Daley, D. J. & Gani, J. (2005). *Epidemic Modeling: An Introduction.* NY: Cambridge University Press.

De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics, 64*, 107–133.

Deneke, T. (1895). Nachtraglches zur Hamburger Cholera-Epidemie von 1892. *Munchener Medcinische Wochenschrift,* 41, 957-961.

Department of Statistics Malaysia. (2017). Retrieved on August 17, 2017 from https://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=dDM2enN vM09oTGtQemZPVzRTWENmZz09

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns* (2nd ed.). London: Amold.

Diggle, P. J. (2006). Spatial-temporal point processes, partial likelihood, foot and mouth. Statistical Methods in Medical Research, 15, 325-336.

Diggle, P. J., Kaimi, I., Abellana, R. (2010). Partial-Likelihood Analysis of Spatio-Temporal Point-Process Data. *Biometrics*, 66 (2), 347-354.
doi: 10.1111/j.1541-0420.2009.01304.x

Direct and Indirect Disease Transmission. (2011). *Delaware Health and Social Services*, 5156. Retrieved from http://dhss.delaware.gov/dph/files/directindtranspi.pdf

Fishman, P.M., Snyder, D. (1976). The statistical analysis of space-time point processes. *IEEE Transactions on Information Theory,* 22, 257–274.

Gatrell, A. C., Bailey, T. C., Diggle, P. J., Rowlingson, B. S., & Rowlingsont, B. S. (1996). Point Spatial application pattern analysis geographical epidemiology. *Transactions of the Institute of British Geographers*, *21*(1), 256–274. Retrieved from https://www.msu.edu/~ashton/classes/866/papers/gatrell_ppa.pdf

Gemperli, A., Vounatsou, P., Sogoba, N., and Smith, T. (2006). Malaria mapping using transmission models: An application to survey data from Mali. *American Journal Epidemiology*, 163, 289-297.

Hauck, F. R., Neese, B. H., Panchal, A. S., & El-Amin, W. (2009). Identification and management of latent tuberculosis infection. *American Family Physician*, *79*(10), 879–886.

Health Information Centre, MOH, Report of Morbidity & Mortality for Inpatient for the year 2004-2012 (PER-PD206). Retrieved from www.**moh**.gov.my

Health Status. (2000). Leeds, Grenville & Lanark District Health Profile. Retrieved on Mac 5, 2016 from http://www.healthunit.org/chsr/data_def/def.htm

Helmersson, J. (2012). Mathematical Modeling of Dengue - Temperature Effect on Vectorial Capacity. Master thesis, Umeå University, Sweden.

Hethcote, H. W. (2007). The Mathematics of Infectious Diseases The Mathematics of Infectious Diseases *. *Society for Industrial and Applied Mathematics*, *42*(4), 599–653. Retrieved from http://www.jstor.org/discover/10.2307/2653135?uid=3739736&uid=2&uid=4& uid=3739256&sid=21104838342357

Howe, G. M. (1959). The geographical distribution of disease with special reference to cancer of the lung and stomach in Wales. *British Journal of Preventive & Social Medicine*, *13*, 204–210. doi:10.1136/jech.13.4.204.

Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *J R Stat Soc Series B Stat Methodol*, 69(5), 817–838. doi:10.1111/j.1467-9868.2007.00612.x.

Jiloris, D. F., Jamaliah Ahmad and Yap, K. T. (2004). Epidemiology of tuberculosis and leprosy, Sabah, Malaysia. In *Tuberculosis*, *84*(1-2), 8–18. doi:10.1016/j.tube.2003.08.002

Jones, M. E., and Swerdlow, A. J. (1998). Bias in the standardized mortality ratio when using general population rates to estimate expected number of deaths. *American Journal of Epidemiology*, *148*(10), 1012–1017.

Kethireddy, S. (2010). Tuberculosis-A Review of Clinical Features, Differential Diagnosis and Treatments Available. *International Journal of Pharmacy & Technology*, *2*(2), 206–245.

Koch, T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*, Redlands, CA: ESRI Press.

Konstantinos, A. (2010). Diagnostic tests Testing for tuberculosis. *Aust Prescr*, *33*(1), 12–18.

Kolappan, C, Subramani, R, Karunakaran, K, and Narayanan, P. R. (2006). Mortality of tuberculosis patients in Chennai, India. *Bulletin of the World Health Organization*, *84*(7), 555-560.

Kumar,V., Abbas, A.K., Fausto, N., Mitchell, R.N., (2007). *Robbins Basic Pathology* (8th ed.). Saunders Elsevier. 516–522. ISBN 978-1-4160-2973-1.

Kuppusamy, I. (2004). Tuberculosis in Malaysia: Problems and prospect of treatment and control. In *Tuberculosis*, *84*(1-2), 4–7. doi:10.1016/j.tube.2003.08.014

Larssen, E. M., Desai, S. S., Anahita Dua, and Shortell, C. E. K. (2013). *Phlebology, Vein Surgery and Ultrasonography:Diagnosis and Management of Venous Disease*. Switzerland: Springer Science & Business Media.

Lawson, A. B. (2000). Cluster modeling of disease incidence via RJMCMC methods: a comparative evaluation. *Statistics in Medicine*, 19, 2361-2376.

Lawson, A. B. (2006). *Statistical methods in spatial epidemiology*. Wet Sussex, UK: John Wiley & Sons, Ltd.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. (2nd ed.). New York: CRC Press.

Lawson, A. B., Browne, W. J., and Rodeiro, C.L.V., (2003). *Disease mapping with WinBUGS and MLwiN*. Statistics in Practise. Chichester: John Wiley & Sons, Ltd.

Lawson, A.B. & Williams, F. (2001). *An introductory guide to disease mapping*. Wet Sussex, UK: John Wiley & Sons, Ltd.

Liao, C.-M., Cheng, Y.-H., Lin, Y.-J., Hsieh, N.-H., Huang, T.-L., Chio, C.-P., Ling, M.-P. (2012). A Probabilistic Transmission and Population Dynamic Model to Assess Tuberculosis Infection Risk. *Risk Analysis*, *32*(8), 1420–1432. doi:10.1111/j.1539-6924.2011.01750.x

Lilienfeld, A. M. and Lilienfeld, D. E. (1981). *Foundations of Epidemiology* (2nd ed). Oxford: Oxford University Press.

Loh, F. F., Royce, T., Natasha, J., Geryl, O., Ivan, L., Tan, S. C., Lee, H. S., and Voon, S. M. (2017, 19 July). Infectious disease making comeback. *The Star Online*. Retrieved on September 12, 2017 from http://www.thestar.com.my/news/nation/2017/07/19/infectious-diseases-making-comeback-more-deaths-from-tuberculosis-than-dengue-and-hivrelated-complic/

Meza, J.L., (2003) Empirical Bayes estimation smoothing of relative risks in disease mapping, *Journal of Statistical Planning and Inference*, 112, 43-62

Ministry of Health Malaysia. (2012-2016). *Health Indicators 2003-2013*. Retrieved from http://vlib.moh.gov.my/cms/content.jsp?id=com.tms.cms.section.Section_62581998-c0a81049-d0570e00-98ce6828

Ministry of Health Malaysia. (2002). *2012 annual report*. Retrieved from http://www.moh.gov.my/images/gallery/publications/md/ar/2002-1.pdf

Ministry of Health Malaysia. (2015). *Global AIDS Response Progress Report Malaysia 2015*. Retrieved from http://www.moh.gov.my/index.php/file_manager/dl_item/554756755a584a69615852686269394d59584276636d46754c3031686247463563326c6858306442556c6c4253587a49774d5455756347526d

Molzon, R. (2009). Deterministic approximation of stochastic evolutionary dynamics. *Proceedings of the 2009 International Conference on Game Theory for Networks, GameNets '09*, 323–332.

Mukhtar A. Adeiza., Abdullah A. Abba, & Juliana U. Okpapi. (2014). HIV-Associated tuberculosis: A sub-saharan african perspective. *Sub-Saharan African Journal of Medicine*, *1*(1).

Murali Haran. (2009). An introduction to models for disease dynamic. Retrieved on October, 15 2016 from www.unc.edu>~rls>samsidisdyntut

Narasimhan, P., Wood, J., Intyre, C. R. M., Mathai, D. (2005). Risk factors for tuberculosis. *Monaldi Archives for Chest Disease - Pulmonary Series*, *63*(1), 37–46. doi:10.1155/2013/828939.

National Institute of Allergy and Infectious Disease. (2012). *Tuberculosis (TB): Prevention*. Retrieved on December 24, 2016 from http://www.niaid.nih.gov/topics/tuberculosis/Understanding/pages/prevention.aspx

New York State Department of Health Tuberculosis (TB). (2007). Retrieved from http://www.health.ny.gov/diseases/communicable/tuberculosis/fact_sheet.htm

Nishiura, H. (2006). Mathematical and statistical analysis of the spread of dengue *Dengue Bulletin*, *30*, 51-67.

Nor Azah Samat. (2012). Mathematical Models for Vector-Borne nfectious Disease Mapping with Application to Dengue Disease in Malaysia. PhD thesis, University of Salford, Manchester, UK.

Nor Azah Samat & Percy, D. F. (2012). Dengue disease mapping in Malaysia based on stochastic SIR models in human populations. *ICSSBE 2012 - Proceedings, 2012 International Conference on Statistics in Science, Business and Engineering: "Empowering Decision Making with Statistical Sciences,"* 623–627. doi:10.1109/ICSSBE.2012.6396640

Nor Azah Samat, & Syafiqah Husna Mohd Imam Ma'arof. (2013). Dengue Disease Mapping with Standardized Morbidity Ratio and Poisson-gamma Model : An Analysis of Dengue Disease in Perak, Malaysia. *World Academy of Science, Engineering and Technology*. 80(3), 571–575.

One TB patient can potentiallay infect 10 people: Malaysian health minister". (2014, April 1). *The Star/ Asia News Network*. Retrieved on Febuary 12, 2017 from http://health.asiaone.com/health/health-news/one-tb-patient-can-potentially-infect-10-people-malaysian-health-minister

Pluciński, M. M., Ngonghala, C. N., & Bonds, M. H. (2011). Health safety nets can break cycles of poverty and disease: a stochastic ecological model. *Journal of The Royal Society Interface*, *8*(65), 1796-1803.

Porco, T. C., & Blower, S. M. (1998). Quantifying the intrinsic transmission dynamics of tuberculosis. *Theoretical Population Biology*, *54*(2), 117–132. doi:10.1006/tpbi.1998.1366.

Rafiza, S., & Rampal, K. G. (2012). Serial testing of Malaysian health care workers with QuantiFERON??-TB gold in-tube. *International Journal of Tuberculosis and Lung Disease*, *16*(2), 163–168. doi:10.5588/ijtld.11.0364.

ReVelle, C. S. (1967). *The economics allocation of tuberculosis control activities in developing nations*. Cornell University.

Ripley, B.D. (1981). *Spatial Statistics*. New York: Wiley.

Robinson, A. H., and Sale, R. D. (1969). *Elements of Cartography* (3rd ed.). New York : Wiley.

Schrodle, B. and Held, L. (2011a). A primer on disease mapping and ecological regression using inla. *Computational Statistics*, 26, 241-258.

Schrodle, B. and Held, L. (2011b). Spatio-temporal disease mapping using inla. *Environmetrics*, 22, 725-734.

Shaddick, G. (2008). Spatial Epidemiology. Retrieved on 2017, August 10 from https://www.stat.ubc.ca/~gavin/SpatEpiPart2.pdf

Snow, J. (1854). *On the Mode of Communication of Cholera* (2nd ed.). London: Churchill Livingstone.

Snow, J. (1936). Snow on cholera: being a reprint of two papers. London: The Commonwealth Fund.

Srinivasan, R., & Venktesan, P. (2013). Bayesian Model For Spatial Dependance And Prediction Of Tuberculosis. *International Journal of Research In Medical and Health Sciences*, 3(2), 1–6.

Stern, H. S. and Cressie, N. A. (1999). Inference for extremes in disease mapping. In Lawson, A. B. Böhning, D., Lasaffree, E., Biggeri, A., Vel, J. F., and Bertolline, R. (Eds). *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley & Sons, Ltd.

Stýblo, K., Meijer, J., and Sutherland, I. (1969). Tuberculosis Surveillance Research Unit Report No. 1: the transmission of tubercle bacilli; its trend in a human population. *Bull Int Union Tuberc*, 42,1–104.

Sufi Hafawati Ideris. (2016). *The Development Of Stochastic Sir and S(IMIF)R Models for Heterosexual HIV and Aids Disease Mapping in Malaysia.* Master thesis, Universiti Pendidikan Sultan Idris, Malaysia.

Symons, M. J., & Taulbee, J. D. (1980). *Standardized Mortality Ratio as an Approximation to Relative Risk* (Institute of Statistic Series No. 1294). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.7305&rep=rep1&type=pdf

Tellier, R. (2009). Aerosol transmission of influenza A virus: a review of new studies. *Journal of The Royal Society Interface*, 6(Suppl_6), S783–S790. doi:10.1098/rsif.2009.0302.focus

Trottier, H., Philippe, P. (2002). Deterministic Modeling of Infectious Diseases: Measles Cycles and the Role of Births and Vaccination. *The Internet Journal of Infectious Diseases*, 2 (2).

Tuberculosis (TB, Consumption) – Diagnosis and Treatment. (2016, December 5). *Lecturio Medical Online Library*. Retreved from https://www.lecturio.com/magazine/tuberculosis/

U.S.Department of Health and Human Service. (2012). *Principle of Epidemiology in Public Health Practice: An Introduction to Applied Epidemiology and Biostatistics* (3rd ed.), 21. Retrieved on January 2, 2017 from https://www.cdc.gov/ophss/csels/dsepd/ss1978/ss1978.pdf

U.S.Department of Health and Human Service. (2016, May 11). *Tuberculosis Facts: TB and HV/AID*. Retrieved on August 2, 2017 from https://www.cdc.gov/tb/publications/factseries/tbandhiv_eng.htm

Vynnycky, E. and Fine, P. E. M. (1997). Eliminating human tuberculosis: the implications of age dependent of risks of disease and the role of reinfection. *Epidemiology and Infection*, *119*(2), 183-201.

Waaler, H., Geser, A., and Andersen, S. (1962). The use of mathematical models in the study of the epidemiology of tuberculosis. *American Journal of Public Health*. 52, 1002-1013.

Wakefield J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8,158–183. doi: 10.1093/biostatistics/kxl008.

World Health Organization. (2013). *Global tuberculosis report 2013*. Retrieved from http://apps.who.int/iris/handle/10665/91355

World Health Organization. (2014). *Global tuberculosis report 2014*. Retrieved from http://www.who.int/tb/publications/global_report/en/

World Health Organization. (2015). *Global tuberculosis report 2015*. Retrieved from http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua =1

World Health Organization. (2016). *Global tuberculosis report 2016*. Retrieved from http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394- eng.pdf?ua=1

World Health Organization. (2017). *Tuberculosis profile: Malaysia*. Retrieved on Febuary 19, 2017 from https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Repor ts%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=MY&LAN=EN &outtype=html

Yeh, Y. P., Luh, D. L., Chang, S. H., Suo, J., Chang, H. J., & Chen, T. H. H. (2005). Tuberculin reactivity in adults after 50 years of universal bacille Calmette-Gu??rin vaccination in Taiwan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *99*(7), 509–516. doi:10.1016/j.trstmh.2005.03.001

Yoneyama, T. (2010). Simulating the pandemic considering international traffic. *ProQuest Dissertations and Theses*, *3420944*(May 2010), 234. Retrieved from http://search.proquest.com.eserv.uum.edu.my/pqdtglobal/docview/751333529/27EEFAC35EA54D98PQ/10?accountid=42599

Appendix A
Knowledge Dissemination

1) Ijlal Mohd Diah (2016). Relative Risk Estimation of Tuberculosis with Standardized Morbidity Ratio in Malaysia, *Global Journal of Pure and Applied Mathematics*. ISSN 0973-1768, 12(5), 4011–4019.

2) Ijlal Mohd Diah, Nazrina Aziz and Nazihah Ahmad (2016). Tuberculosis Disease Mapping with Poisson-Gamma Model in Malaysia. *Research Journal of Applied Sciences, 11, 822-825.* doi:10.3923/rjasci.2016.822.825

3) Ijlal Mohd Diah, Nazrina Aziz, Nazihah Ahmad, & Maznah Mat Kasim (2016). Tuberculosis disease mapping with stochastic equation. *IACE' 2016-Proceeding of the 3rd Innovation and Analytics Conference & Exhibition, 77-82*.

4) Ijlal Mohd Diah, Nazrina Aziz, Nazihah Ahmad, & Maznah Mat Kasim (2016). Tuberculosis disease mapping with stochastic equation. Presentation Session in *IACE' 2016- 3rd Innovation and Analytics Conference & Exhibition,* 30th October – 1st November 2016.

5) Ijlal Mohd Diah, Nazrina Aziz, & Maznah Mat Kasim (2017). Tuberculosis disease mapping in Kedah using standardized morbidity rato. *ICAST' 2017-Proceeding of the 2nd International Conference on Applied Science and Technology, 1891*(1). https://doi.org/10.1063/1.5005429

6) Ijlal Mohd Diah, Nazrina Aziz, & Maznah Mat Kasim (2017). Tuberculosis disease mapping in Kedah using standardized morbidity rato. Presentation

Session in *ICAST' 2017- Proceeding of the 2nd International Conference on Applied Science and Technology,* 10th April – 12th 2017.

7) Ijlal Mohd Diah, Nazrina Aziz & Maznah Mat Kasim, (2017). A Comparison of Four Disease Mapping Techniques as Applied to TB Diseases in Malaysia. *Journal of Telecommunication, Electronic and Computer Engineering. 9* (2-11), 133–137.

Appendix B

WinBUGS Output of Summary Statistics for Relative Risk Estimation based on Stochastic SLIR Model

AB-1: Summary Statistics for the State of Perlis

node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	0.0487	0.00152	1.9E-5	0.04583	0.0487	0.05158
RRH[1,3]	0.831	0.02594	3.243E-4	0.7824	0.831	0.8807
RRH[1,4]	0.8337	0.02602	3.254E-4	0.7849	0.8337	0.8835
RRH[1,5]	0.863	0.02693	3.36E-4	0.8124	0.863	0.9145
RRH[1,6]	1.045	0.0361	4.078E-4	0.9836	1.045	1.107
RRH[1,7]	0.9028	0.02818	3.523E-4	0.85	0.9028	0.9567
RRH[1,8]	0.7148	0.02231	2.79E-4	0.673	0.7148	0.7575

AB-2: Summary Statistics for the State of Kedah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[2,2]	0.05011	6.724E-4	7.2E-6	0.04898	0.05011	0.05128
RRH[2,3]	0.7865	0.01055	1.148E-4	0.7688	0.7865	0.8048
RRH[2,4]	0.8452	0.01134	1.215E-4	0.8262	0.8452	0.8649
RRH[2,5]	0.8293	0.01113	1.192E-4	0.8106	0.8293	0.8486
RRH[2,6]	0.8401	0.01127	1.207E-4	0.8212	0.8401	0.8597
RRH[2,7]	0.8203	0.01101	1.179E-4	0.8018	0.8203	0.8394
RRH[2,8]	0.9157	0.01229	1.316E-4	0.8951	0.9157	0.937

AB-3: Summary Statistics for the State of Pulau Pinang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[3,2]	0.05117	6.701E-4	8.222E-6	0.05005	0.05117	0.05233

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[3,3]	1.103	0.01444	1.772E-4	1.079	1.103	1.128
RRH[3,4]	1.113	0.01458	1.789E-4	1.089	1.113	1.139
RRH[3,5]	1.071	0.01402	1.72E-4	1.047	1.071	1.095
RRH[3,6]	1.105	0.01447	1.775E-4	1.081	1.105	1.13
RRH[3,7]	1.102	0.01443	1.771E-4	1.078	1.102	1.127
RRH[3,8]	1.112	0.01456	1.786E-4	1.087	1.112	1.137

AB-4: Summary Statistics for the State of Perak

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[4,2]	0.05044	6.184E-4	6.015E-6	0.04944	0.05046	0.05147
RRH[4,3]	0.8932	0.01095	1.065E-4	0.8754	0.8935	0.9114
RRH[4,4]	0.9651	0.01183	1.151E-4	0.9458	0.9653	0.9846
RRH[4,5]	0.8438	0.01034	1.006E-4	0.827	0.844	0.8609
RRH[4,6]	0.918	0.01125	1.095E-4	0.8997	0.9182	0.9366
RRH[4,7]	0.912	0.01118	1.087E-4	0.8938	0.9122	0.9304
RRH[4,8]	0.9516	0.01166	1.135E-4	0.9326	0.9518	0.9709

AB-5: Summary Statistics for the State of Kuala Lumpur & Putrajaya

node	mean	Sd	MC error	2.5%	median	97.5%
RRH[5,2]	0.05207	6.128E-4	6.761E-6	0.05108	0.05207	0.05311
RRH[5,3]	1.362	0.01603	1.768E-4	1.336	1.362	1.389
RRH[5,4]	1.371	0.01614	1.781E-4	1.345	1.371	1.399
RRH[5,5]	1.59	0.01871	2,064E-4	1.56	1.59	1.622
RRH[5,6]	1.52	0.01789	1.974E-4	1.491	1.52	1.55
RRH[5,7]	1.524	0.01794	1.979-4	1.495	1.524	1.555
RRH[5,8]	1.429	0.01682	1.856E-4	1.402	1.429	1.458

AB-6: Summary Statistics for the State of Selangor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[6,2]	0.05156	5.249E-4	5.389E-6	0.05079	0.05156	0.05235
RRH[6,3]	0.7675	0.007814	8.022E-5	0.756	0.7675	0.7793
RRH[6,4]	0.8851	0.009011	9.252E-5	0.8719	0.8851	0.8987
RRH[6,5]	0.9156	0.009321	9.57E-5	0.9019	0.9156	0.9296
RRH[6,6]	0.945	0.00962	9.877E-5	0.9308	0.945	0.9595
RRH[6,7]	1.057	0.01076	1.105E-4	1.041	1.057	1.073
RRH[6,8]	1.139	0.0116	1.191E-4	1.112	1.139	1.157

AB-7: Summary Statistics for the State of Negeri Sembilan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[7,2]	0.05046	8.727E-4	9.799E-6	0.0489	0.05046	0.0521
RRH[7,3]	0.6989	0.01209	1.357E-4	0.6773	0.6989	0.7215
RRH[7,4]	0.7149	0.01236	1.388E-4	0.6929	0.7149	0.7381
RRH[7,5]	0.682	0.0118	1.324E-4	0.661	0.682	0.7041
RRH[7,6]	0.6101	0.01055	1.185E-4	0.5912	0.6101	0.6298
RRH[7,7]	0.834	0.01442	1.691E-4	0.8083	0.834	0.861
RRH[7,8]	0.918	0.01588	1.783E-4	0.8896	0.918	0.9477

AB-8: Summary Statistics for the State of Melaka

node	mean	sd	MC error	2.5%	median	97.5%
RRH[8,2]	0.04988	8.843E-4	1.001E-5	0.04827	0.04988	0.0515
RRH[8,3]	0.7227	0.01281	1.45E-4	0.6993	0.7227	0.7461
RRH[8,4]	0.8084	0.01433	1.622E-4	0.7822	0.8084	0.8346
RRH[8,5]	0.9199	0.01631	1.846E-4	0.8901	0.9199	0.9497

node	mean	sd	MC error	2.5%	median	97.5%
RRH[8,6]	0.9283	0.01646	1.863E-4	0.8982	0.9283	0.9583
RRH[8,7]	0.9163	0.01624	1.839E-4	0.8866	0.9163	0.946
RRH[8,8]	1.089	0.0193	2.185E-4	1.054	1.089	1.124

AB-9: Summary Statistics for the State of Johor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[9,2]	0.0505	5.696E-4	5.818E-6	0.0496	0.0505	0.05141
RRH[9,3]	0.9279	0.01047	1.069E-4	0.9114	0.9279	0.9446
RRH[9,4]	1.001	0.01129	1.154E-4	0.9835	1.001	1.019
RRH[9,5]	0.9137	0.01031	1.053E-4	0.8974	0.9137	0.9301
RRH[9,6]	0.8675	0.009784	9.994E-5	0.8521	0.8675	0.8831
RRH[9,7]	0.8537	0.009628	9.835E-5	0.8385	0.8537	0.869
RRH[9,8]	0.9527	0.01075	1.098E-4	0.9358	0.9527	0.9699

AB-10: Summary Statistics for the State of Pahang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[10,2]	0.04961	7.182E-4	7.493E-6	0.04838	0.04961	0.05089
RRH[10,3]	0.9093	0.01316	1.373E-4	0.8867	0.9093	0.9327
RRH[10,4]	0.8717	0.01262	1.317E-4	0.8501	0.8717	0.8942
RRH[10,5]	0.7801	0.01129	1.178E-4	0.7607	0.7801	0.8002
RRH[10,6]	0.813	0.01177	1.228E-4	0.7928	0.813	0.8339
RRH[10,7]	0.7925	0.01147	1.197E-4	0.7728	0.7925	0.8129
RRH[10,8]	0.8338	0.01207	1.259E-4	0.8131	0.8338	0.8553

AB-11: Summary Statistics for the State of Terengganu

node	Mean	sd	MC error	2.5%	median	97.5%

node	mean	sd	MC error	2.5%	median	97.5%
RRH[11,2]	0.04983	7.79E-4	8.561E-6	0.04848	0.04983	0.05126
RRH[11,3]	1.138	0.01779	1.955E-4	1.107	1.138	1.171
RRH[11,4]	1.005	0.01571	1.727E-4	0.9777	1.005	1.034
RRH[11,5]	0.9149	0.0143	1.572E-4	0.8901	0.9149	0.9412
RRH[11,6]	0.9339	0.0146	1.605E-4	0.9086	0.9339	0.9608
RRH[11,7]	0.9532	0.0149	1.638E-4	0.9273	0.9532	0.9806
RRH[11,8]	0.9307	0.01455	1.599E-4	0.9055	0.9307	0.9575

AB-12: Summary Statistics for the State of Kelantan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[12,2]	0.0503	6.347E-4	6.992E-6	0.04927	0.0503	0.05138
RRH[12,3]	1.325	0.01672	1.842E-4	1.298	1.325	1.353
RRH[12,4]	1.321	0.01667	1.836E-4	1.294	1.321	1.349
RRH[12,5]	1.297	0.01636	1.803E-4	1.27	1.297	1.325
RRH[12,6]	1.221	0.01541	1.697E-4	1.196	1.221	1.247
RRH[12,7]	1.166	0.01471	1.621E-4	1.142	1.166	1.191
RRH[12,8]	1.163	0.01468	1.617E-4	1.14	1.163	1.188

AB-13: Summary Statistics for the State of Sabah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[13,2]	0.05318	5.307E-4	5.272E-6	0.0524	0.05318	0.05397
RRH[13,3]	1.775	0.01771	1.759E-4	1.749	1.775	1.801
RRH[13,4]	1.856	0.01852	1.84E-4	1.829	1.856	1.884
RRH[13,5]	1.741	0.01737	1.726E-4	1.715	1.741	1.767
RRH[13,6]	1.881	0.01877	1.864E-4	1.853	1.881	1.909
RRH[13,7]	1.885	0.01881	1.869E-4	1.858	1.885	1.913
RRH[13,8]	2.015	0.0201	1.997E-4	1.985	2.015	2.044

AB-14: Summary Statistics for the State of Sarawak

node	mean	sd	MC error	2.5%	median	97.5%
RRH[14,2]	0.05232	5.733E-4	5.868E-6	0.051	0.05232	0.05323
RRH[14,3]	1.293	0.01416	1.45E-4	1.27	1.293	1.315
RRH[14,4]	1.326	0.01452	1.487E-4	1.302	1.326	1.349
RRH[14,5]	1.248	0.01367	1.399E-4	1.226	1.248	1.269
RRH[14,6]	1.367	0.01497	1.533E-4	1.343	1.367	1.39
RRH[14,7]	1.46	0.016	1.638E-4	1.435	1.46	1.486
RRH[14,8]	1.56	0.01709	1.749E-4	1.532	1.56	1.587

Appendix C

WinBUGS Code for Relative Risk Estimation based on SMR Method, Poisson-gamma Model, Stochastic SIR Model and Stochastic SLIR Model

AC-1: WinBUGS Code for Estimation of Relative Risk based on SMR Method and Poisson-gamma Model

```
model{
for (i in 1:M){
for (j in 1:T){

            #Relative Risk
            theta[i,j]<-y[i,j]/e[i,j]
}}}
```

Figure AC.1. SMR method in WinBUGS

```
model{
for (i in 1:M){
for (j in 1:T){
            #Poisson likelihood for observed counts
            y[i,j]~dpois(mu[i,j])
            mu[i,j]<-e[i,j]*theta[i,j]
            #Relative Risk
            theta[i,j]~dgamma(a,b)
}
}
#Prior distribution for "population" parameters
a~dexp(0.1)
b~dexp(0.1)
#Population Mean and Population variance
mean<-a/b
var<-a/pow(b,2)
}
```

Figure AC.2. Poisson-gamma model in WinBUGS

The code for SMR method and Poisson-gamma model were adapted from Nor Azah and Syafiqah Husna (2013) which was used to analyze dengue disease occurrence for

districts in Perak, Malaysia. Moreover, this Poisson-gamma model's code was written by Lawson et al. (2003) in their study which was applied to analyze influenza data from South Carolina.

AC-2: WinBUGS Code for Estimation of Relative Risk based on the Stochastic SIR Model

This code has been written by Lawson (2006) to analyze influenza seasons in 13 consecutive time periods in South Carolina for the year 2004-2005. Nevertheless, in order to suit the particular requirement of this study, we modified the notations and formulations used in that WinBUGS code.

```
Model{
for (i in 1:M){
    Rh[i,1]<-0
    Sh[i,1]<-Nh[i]
    muH[i,1]<-Sh[i,1]
    Ih[i,1]~dpois(muH[i,1])
}
for (i in 1:M){
for (j in 2:T){
    Rh[i,j]<-betaR*Ih[i,j]
    Sh[i,j]<-Sh[i,j-1]-Ih[i,j-1]-Rh[i,j-1]
    Ih[i,j]~dpois(muH[i,j])
    log(muH[i,j])<-beta0+log(Sh[i,j]+0.001)+log(Ih[i,j-1]+0.001)+b1[i]

#Relative Risk
theta[i,j]<-muH[i,j]/eH[i,j]
}
}
#CAR prior distribution for random effects, the sum of b1 is always zero
b1[1:14]~car.normal(adj[],weights[],num[],tau.b1)
for (k in 1:sumNumNeigh){
weights[k]<-1}

#Other priors
beta0~dflat()                    #Flat prior for the intercept
tau.b1~dgamma(0.01,0.01)         # Prior on precision
betaR<-0.001

}
```

Figure 4.6. Stochastic SIR model in WinBUGS