THE DEVELOPMENT OF STOCHASTIC MODELS FOR RELATIVE RISK ESTIMATION IN CONSTRUCTING PNEUMONIA DISEASE MAPPING IN MALAYSIA

IJLAL BINTI MOHD DIAH

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2022

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
(candidate for the degree of) **PhD**

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

"THE DEVELOPMENT OF STOCHASTIC MODELS FOR RELATIVE RISK ESTIMATION IN CONSTRUCTING PNEUMONIA DISEASE MAPPING IN MALAYSIA"

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 : ***16 Disember 2021.***
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:
16 December 2021.

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Shamshuritawati Sharif**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Ts. Dr. Norziha Che Him**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Assoc. Prof. Dr. Zakiyah Zain**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Assoc. Prof. Dr. Nazrina Aziz**	Tandatangan *(Signature)*

Tarikh:
(Date) ***16 December 2021***

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

Pneumonia merupakan salah satu penyebab utama kematian bagi penyakit berjangkit terutamanya di negara membangun. Secara konvensional, penyebaran penyakit ini hanya dipantau berdasarkan jumlah kes yang direkodkan tanpa mengambil kira taburan geografi. Alternatifnya, pemetaan penyakit boleh dibangunkan berdasarkan risiko relatif yang merangkumi taburan geografi. Pemetaan penyakit yang baik bergantung kepada ketepatan penganggaran risiko relatif daripada penyuaian terbaik model berstatistik. Oleh itu, kajian ini bertujuan untuk membangunkan satu kaedah alternatif dalam menganggarkan risiko relatif pneumonia berdasarkan empat model stokastik: Susceptible-Infected-Carrier (SIC), Susceptible-Infected-Recovered (SIR), Susceptible-Carrier-Infected-Recovered (SCIR) dan Susceptible-Vaccinated-Carrier-Infected-Recovered (SVCIR). Anggaran risiko relatif ini kemudiannya dibandingkan dengan kaedah sedia ada: Standardized Mortality Ratio (SMR), Poisson-gamma dan Besag, York dan Mollie (BYM Terdapat empat fasa dalam kajian ini. Pertama, empat model berketentuan yang bersesuaian untuk penyebaran penyakit pneumonia dipilih, yang mana melaluinya model stokastik dibangunkan. Seterusnya, keempat-empat model stokastik ini digunakan untuk menganggar risiko relatif penyakit pneumonia dengan menganalisis data pneumonia di Malaysia dari tahun 2010 sehingga tahun 2019. Prestasi keempat-empat model stokastik dan kaedah sedia ada dinilai dengan membandingkan nilai risiko relatif masing-masing. Akhir sekali, peta risiko pneumonia dibina berdasarkan nilai risiko relatif yang diperolehi. Dapatan menunjukkan bahawa terdapat jurang nilai penganggaran risiko relatif yang besar apabila menggunakan model stokastik SVCIR berbanding model lain. Nilai risiko relatif model stokastik SVCIR menurun dari tahap risiko tinggi kepada tahap risiko sederhana dan dari tahap risiko sederhana ke tahap risiko rendah. Situasi ini berlaku memandangkan model stokastik SVCIR mengambil kira korelasi reruang antara kawasan dan maklumat tambahan di dalam model seperti komponen vaksinasi dan komponen pembawa. Aplikasi model ke atas data Malaysia menunjukkan Putrajaya telah dikenalpasti sebagai kawasan berisiko tinggi dengan jangkitan pneumonia. Hal ini adalah kerana Putrajaya merupakan kawasan terkecil dengan kadar pertumbuhan populasi penduduk tertinggi di Malaysia. Kesimpulannya, model stokastik ini menunjukkan prestasi yang lebih baik berbanding model konvensional. Disamping itu, model ini boleh digunapakai untuk penyakit berjangkit lain yang mempunyai ciri penularan yang sama. Pemetaan penyakit ini dapat membantu kerajaan dalam mengutamakan kawasan yang memerlukan perhatian bagi mencapai sistem kesihatan yang lestari.

Kata kunci: Pemetaan penyakit, Penyakit berjangkit, Risiko relatif, Pneumonia, Model stokastik.

Abstract

Pneumonia is one of the leading causes of death for infectious diseases especially in developing countries. Conventionally, its spread is only being monitored based on the total number of cases recorded without considering geographical distribution. Alternatively, disease mapping can be constructed based on the relative risk that includes a geographical distribution. A good disease mapping relies on the accuracy of relative risk estimated from the best-fitted statistical model. Therefore, this study aims to develop an alternative method in estimating the pneumonia relative risk based on four stochastic models: Susceptible-Infected-Carriers (SIC), Susceptible-Infected-Recovered (SIR), Susceptible-Carrier-Infected-Recovered (SCIR), and Susceptible-Vaccinated-Carrier-Infected-Recovered (SVCIR).These estimated relative risks are then compared with those of the existing methods: Standardized Mortality Ratio (SMR), Poisson-gamma and Besag, York and Mollie (BYM) models. There are four phases in this study. Firstly, four deterministic models that are suitable for pneumonia disease transmission are selected, from which the stochastic models are developed. Next, these four stochastic models are applied to estimate the relative risk for pneumonia disease by analyzing pneumonia data in Malaysia from the year 2010 until the year 2019. The performance of these four stochastic models and existing methods is evaluated by comparing their relative risk values. Finally, the pneumonia risk maps are then constructed based on the relative risk values obtained. Findings show that there is a large gap in relative risk estimation values when using the stochastic SVCIR model compared to other models. The relative risk values when using stochastic SVCIR model decrease from high-risk level to medium risk level and from medium risk level to low-risk level. This situation occurs since stochastic SVCIR model allows for the spatial correlation between the areas and includes extra information in the model such as vaccination and carrier components. Application of the models on the Malaysian data shows that Putrajaya is identified as the highest risk of contracting pneumonia. This is because Putrajaya is the smallest area with the highest population growth rate in Malaysia. In conclusion, these stochastic models demonstrate better performance compared to the conventional models. Furthermore, these models are applicable to other infectious diseases with similar transmission characteristics. The disease mapping may assist the government in prioritizing areas that need further attention in gearing towards a sustainable health system.

Keywords: Disease mapping, Infectious disease, Relative risk, Pneumonia, Stochastic models.

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 supervisor as respect, Associate Professor Dr. Nazrina Aziz for the guidance and insight in the realization of this research, and also on proofread numerous drafts, suggestions and opinions given. I thank her for all advices; for the time she has given to this research; and most of all, for making me confident with my work. Without her, this thesis would not have been possible.

Heartfelt appreciation also goes to my beloved parents, Mohd Diah Hamdan and Fairuz Hassan and also my entire family (Fadhli, Husni, Ashraff, Syazna, Ijmal, Ihsan, Ridha, Umar and Zaheda) 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, Sufi Hafawati, Asma and my traveller buddies for the help and support given since the first day in my PhD study journey.

Finally, 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.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations and Mathematical Symbols

CAP	Community-acquired Pneumonia
CAR	Conditional Autoregressive
GBD	Global Burden of Disease
HAP	Hospital-acquired Pneumonia
IHME	Institute for Health Metrics and Evaluation
MCMC	Markov Chain Monte Carlo
MOH	Ministry of Health
NIP	National Immunisation Programme
OTC	Over-the-counter
RR	Relative Risk
SCIR	Susceptible–Carriers–Infected–Recovered
SEIR	Susceptible–Exposed–Infectious–Recovered
SIC	Susceptible–Infected–Carriers
SI_SCIc	Susceptible–Infected susceptible–Carriers–Infected carriers
SI_cI_iT	Susceptible–asymptomatic Infectives–symptomatic Infectives–treated Infective
SIR	Susceptible–Infected–Recovered

SMR	Standardized Mortality/Morbidity Ratio
SVCIR	Susceptible–Vaccinated–Carrier–Infected–Recovered
SVECI	Susceptible–Vaccinated–Exposed–Carrier–Infected
UNICEF	United Nations Children's Fund
VAP	Ventilator-associated Pneumonia
WHO	World Health Organization
ZPG	Zero Population Growth
$S_{i,j}$	Total number of susceptible persons for area *i*, at time *j*
$I_{i,j}$	Total number of infectious persons for area *i*, at time *j*
$C_{i,j}$	Total number of carrier persons for area *i*, at time *j*
$R_{i,j}$	Total number of recovered persons for area *i*, at time *j*
$V_{i,j}$	Total number of vaccinated 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*
μ	Birth rate and natural mortality rate (per year)
N_i	Population size for the study state
δ	Progression rate from latent period to infectious period
ϵ	Rate at which a carrier becomes infected

q	Rate at which the infectious individual recovers to the susceptible class
g	Recovery rate
$\alpha I_{i,j}$	Risk of a susceptible person's becoming infective in time *j*, where α is constant.
γ	Rate at which a carrier can infect a susceptible
β	Rate at which an infected person can infect a susceptible
μ_p	Mortality rate due to pneumonia of humans per year
p	Rate at which fraction of population has been vaccinated before the disease outbreaks
(*1-p*)	Fraction of susceptible population
φ	Waning rate of those individuals who are vaccinated but did not respond to vaccination
δ	Rate at which individuals lose their temporary immunity
ϑ	Vaccination rate
ε	Proportion of the serotype not covered by the vaccine
ρ	Probability of newly infected persons by the force of infection become carriers to join the carrier class
$(1-\rho)$	Probability of newly infected persons to join the infected class
g_n	Recovery rate by gaining natural immunity

β_0	Overall rate of the process
b_i	Random effect that absorbs residual spatial variation
t	Time

CHAPTER ONE
INTRODUCTION

1.1 Research Background

1.1.1 What is Pneumonia?

Pneumonia is an acute illness, which is a high-incident respiratory disease, categorized by the types of inflammation that is caused by microorganisms known as fungi, viruses, bacteria and parasites. Among these four microorganisms that potentially cause pneumonia, bacteria have been recognized to be the leading cause, especially Streptococcus pneumoniae (pneumococcus) (Otieno, Joseph & John, 2012). These viruses include the new virus that caused Corona Virus Disease 2019 (COVID-19) that is SARS-COV-2, which now had become outbreak worldwide (International Vaccine Access Center, 2020). Viruses are the most common cause of pneumonia in children under five years old. Even though viral pneumonia usually is just mild, however in some cases it can become worst.

Pneumonia is an infection that can cause lung air sacs inflammation. As the lungs become inflamed, the air sacs, or alveoli, will fill up with fluid (Normandin, 2021). The infection can be caused by viral, fungal, or bacterial. This disease can lead to a severe and life-threatening condition. Anyone may be infected by pneumonia without considering their age and body condition as it can occur even in young and healthy people. However, this disease will be riskier if it affects older individuals, infants, and those with weakened immune systems or those with other medical conditions (Normandin, 2021). People with pneumonia will suffer from these early signs: fever, chills, coughing that may produce mucus, chest pain, shortness of

breath and sweating. Depending on the bacteria or germs that caused pneumonia in the first place, the cough may be dry or with sputum (Ahmad, 2019). Nonetheless, other symptoms may differ for each person as it depends on the body condition, cause and severity of the disease (Wardlaw, Johansson & Hodge, 2006; Lynn, 2020).

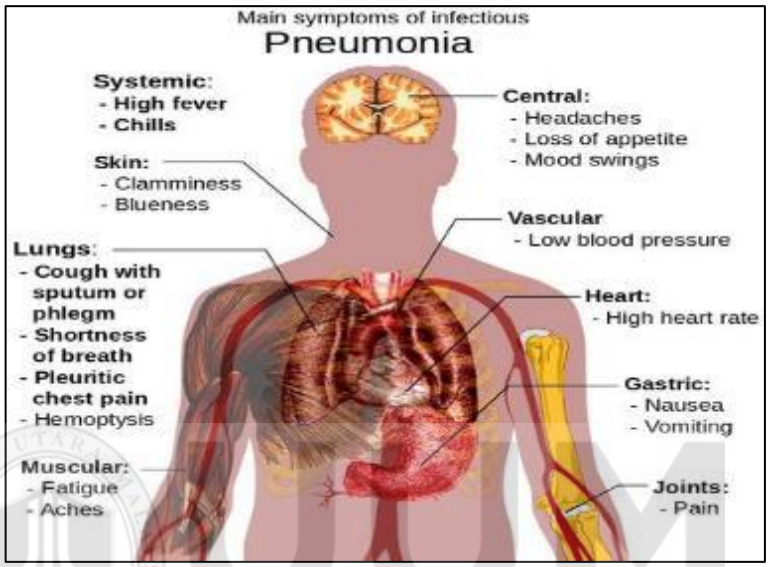

Figure 1.1. Main symptoms of pneumonia infection (Source: Ahmad, 2019)

Pneumonia is divided into a few categories. The categories are depending on (i) the kind of pathogens that infect, (ii) the location it is transmitted, (ii) the way it is obtained. Examples of organisms that can cause the infection are bacteria, virus, mycoplasma and fungi. The most common bacteria that cause pneumonia among children is Streptococcus pneumoniae. Meanwhile, viral pneumonia is not severe and only lasts for a shorter time compared to bacterial pneumonia. Besides classification via germs, pneumonia also has been classified by places it was acquired, such as hospital-acquired pneumonia (HAP) and community-acquired pneumonia (CAP) (Normandin, 2021). HAP is pneumonia that is acquired during a hospital stay, and it is more severe than CAP, as the bacteria involved may be more resistant to antibiotics.

Aside from that, HAP can also be divided into aspiration pneumonia and ventilator-associated pneumonia (VAP). According to WHO (2021), pneumonia can be transmitted to others when someone infected coughs or sneezes, and little droplets spread through the air. These droplets contain the infectious organism. As pneumonia can be caused by bacteria or viruses, anyone who has one of these in their nose or throat can develop pneumonia, which is commonly found in children (WHO, 2021). It can also be spread to others by contacting bacterium or virus-infected objects or surfaces. However, not all cases of pneumonia are caused by transmissible organisms. For example, someone can obtain aspiration pneumonia when they inhale an unwanted substance such as vomit into their lungs.

A person with suspected symptoms of pneumonia is advised to consult doctors for further steps of diagnosing. A chest x-ray may be conducted to examine the presence of the disease and also to determine whether the infection is restricted to one lobe or it is present bilaterally in both lungs. Apart from that, sputum culture, bronchoscopy, thoracentesis, pleural biopsy, and culture of pleural fluid may also help in diagnosing pneumonia. The most specific test to diagnose pneumonia is with open-lung biopsy which is almost never performed as several factors need to be considered when deciding on whether to proceed with open-lung biopsy such as disease progression and the diagnostic possibilities being considered (Ahmad, 2019).

A patient that is positively diagnosed with pneumonia, will undergo some of the following treatments depending on two major factors, which are the severity of the disease and type of pneumonia. A patient with mild pneumonia can be treated without being warded, but if the condition worsens, the patient needs to be warded in order to

receive treatments. Antiviral, antibiotic, and antifungal drugs are commonly prescribed by doctors for pneumonia treatment, based on the exact pathogen. Consuming oral antibiotics can usually treat bacterial pneumonia at home, which is very effective as most patients show response in just one to three days. On the other hand, for mild pneumonia, some over-the-counter (OTC) medication may be taken to reduce fever and pain with the supervision of doctor or pharmacist.

If the symptoms are severe, the patient might need to be hospitalized so that the doctors can keep track of the patient's heart rate, breathing and temperature. Warded patients may be treated with intravenous antibiotics injected into their vein. Besides, respiratory therapy will be conducted to deliver specific medications directly into the lungs, which helps the patients to perform breathing exercise to maximize the oxygenation. Patients with an extreme case may need oxygen therapy or ventilator to support their breathing (Ahmad, 2019).

1.1.2 World Pneumonia Scenario

According to the report by UNICEF (2018), pneumonia has been recognized as the top infectious disease causing death among children under the age of five, estimated around 2,400 children a day. As of 2014, the World Health Organization (WHO) recorded 18% of deaths of children under five years old due to pneumonia at 1.1 million deaths every year (WHO, 2016). Poor and rural communities, especially in the South Asia and sub-Saharan Africa, experience the most casualties caused by pneumonia. Due to the lower living standards in these countries, there is a high financial burden and insufficient budget to fund a mass health immunization programs for all populations in the countries afflicted by pneumonia.

Based on the fact sheet of WHO (2016), in the year 2015, about 920,136 children, who contribute around 16% of all deaths under-five, was caused by pneumonia. Figure 1.2 shows the percentage of death among children under-five attributed to pneumonia. It can be seen that sub-Saharan Africa, as well as South Asia, contributes the most significant proportion to pneumonia deaths. The percentage of fatalities in these regions has constantly increased from 77% in year 2000 to 82% in year 2015 (refer to Figure 1.3).

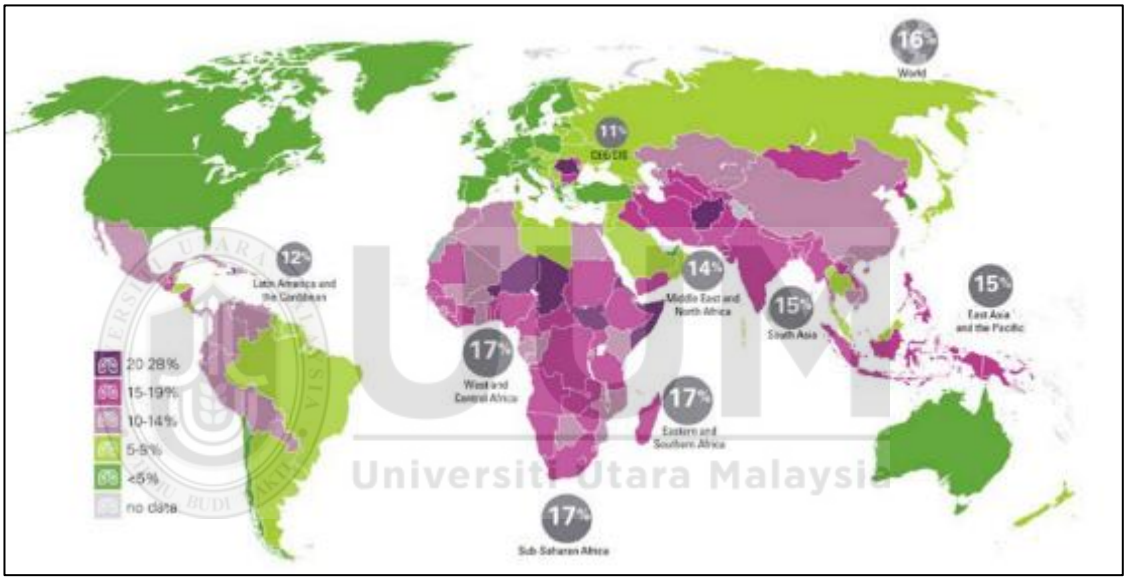

Figure 1.2. Percentage of death among children aged under-five caused by pneumonia (Source: UNICEF, 2016)

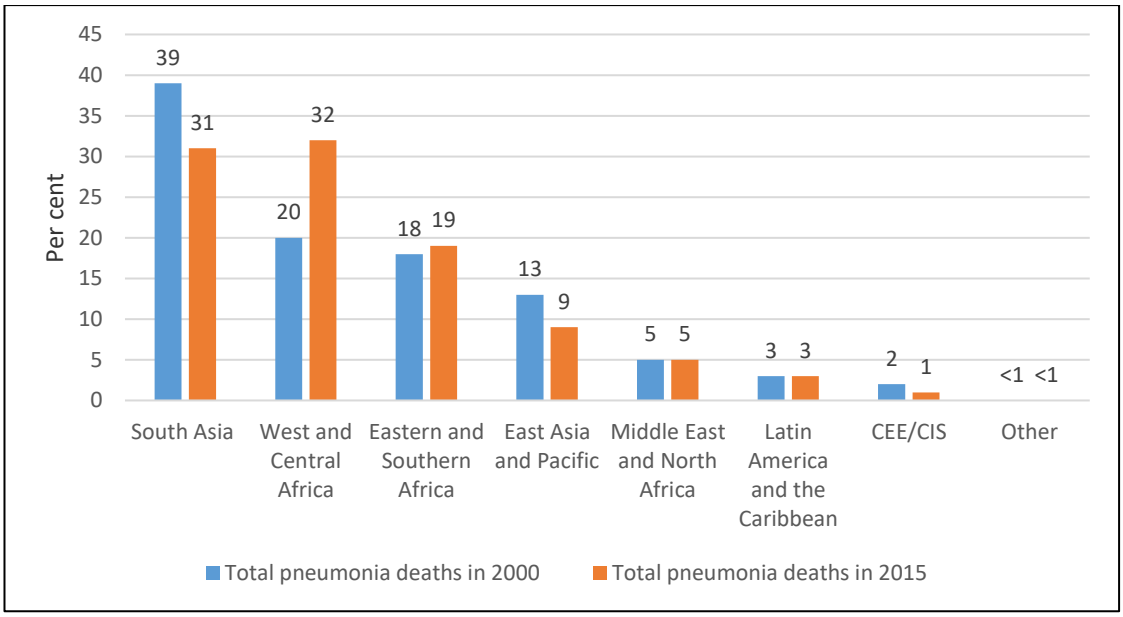

Figure 1.3. Distribution of under-five deaths caused by pneumonia in 2000 and 2015 by UNICEF region (Source: UNICEF, 2016)

According to UNICEF (2016), India recorded the highest under-five deaths caused by pneumonia in 2015 with 179,000 deaths, followed by Nigeria (132,200 deaths) and Pakistan (64,000 deaths). In 2016, about 16% (888,000) of the 5.6 million under-five deaths were due to pneumonia; most of them were less than two years old (UNICEF, 2018). Meanwhile, South Asia and Sub-Saharan Africa remain the highest proportion of death caused by pneumonia. Pneumonia is a poverty disease where mortality continued to be concentrated in the most impoverished populations. According to UNICEF (2016), low-income and lower-middle-income countries accounted for 62% of the world's under-five population but contributed to 90% of the global pneumonia deaths. Figure 1.4 shows the percentage distribution of deaths due to pneumonia disease classified by income levels. More than 50% of all pneumonia deaths concentrates among lower-middle-income countries, while 31% focuses on low-income countries. Higher disbursements were given to the lower-income countries. Middle-income countries with large populations received merely small

portion of disbursements, albeit amounting to almost half of the numbers of the disease (UNICEF, 2016).

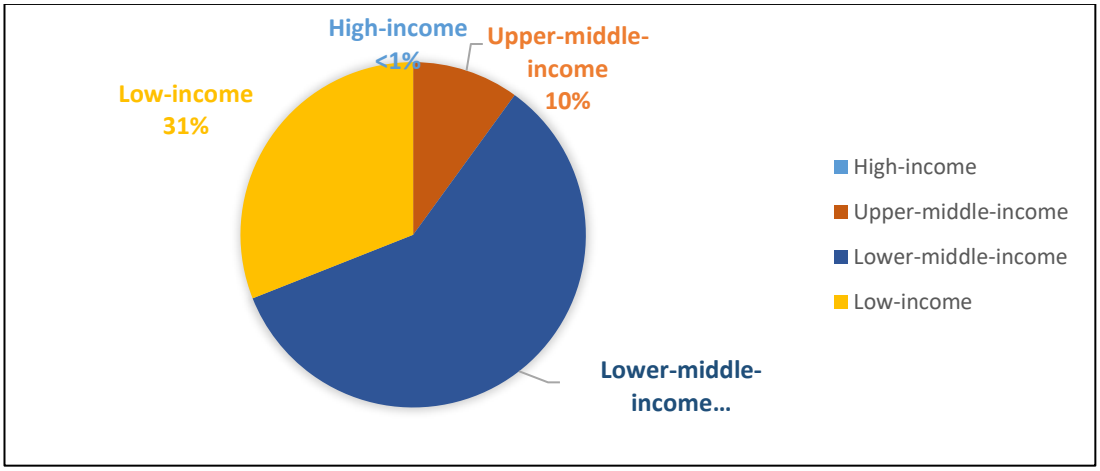

Figure 1.4. Percentage distribution of deaths attributed to pneumonia by income levels (Source: UNICEF, 2016)

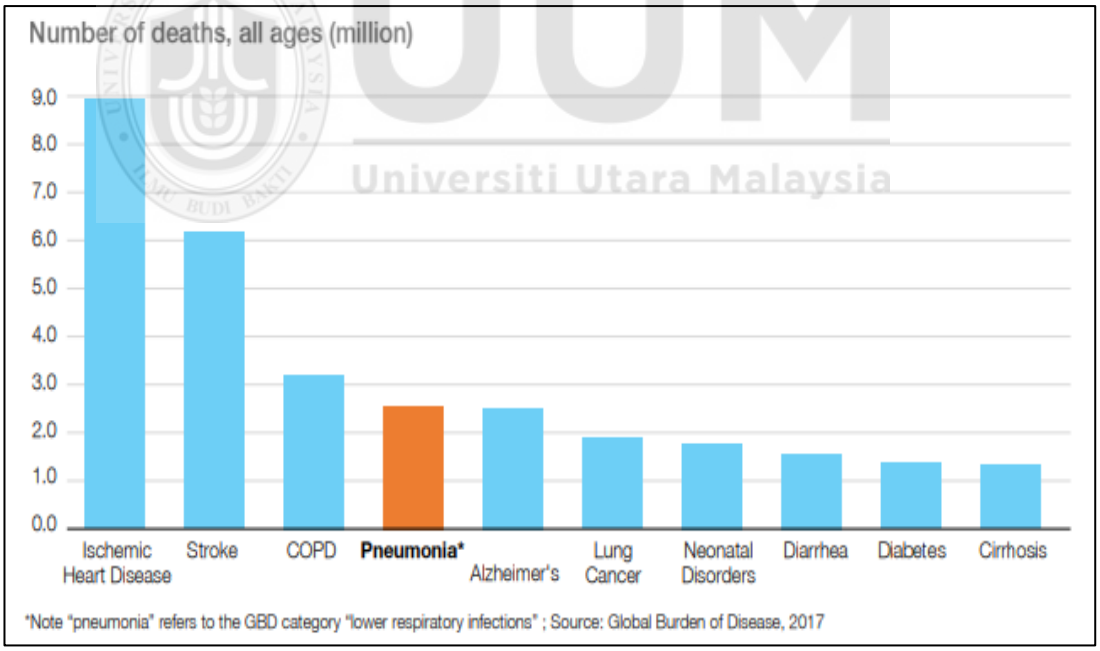

Figure 1.5. Top 10 causes of death globally in 2017 (Source: IHME, Global Burden of Disease (GBD), 2018)

Figure 1.5 shows that pneumonia is one of the top 10 causes of mortality globally. In 2017, it was estimated that 2.6 million deaths were caused by pneumonia,

according to the GBD (Greenslade, 2018) with 808,694 deaths among children under-five (WHO, 2018) and 1.13 million deaths occur among adults aged 70 years old and above (Figure 1.6). These 2.6 million fatalities concentrate in Saharan Africa, South-East Asia and South Asia (Greenslade, 2018). Among these countries, India recorded the highest mortality with 507,000 deaths due to pneumonia which is 20% of all global pneumonia deaths. Pneumonia causes less mortality in Central Europe, North Africa and the Middle East, Latin America and the Caribbean, Central Asia and Eastern Europe. According to the report, pneumonia deaths are also high in high-income countries (Greenslade, 2018). In high-income countries, the deaths among adults aged over 70 years, make up 86% of pneumonia deaths, while in low-income countries, especially in sub-Saharan Africa, deaths due to pneumonia among children is 60%.

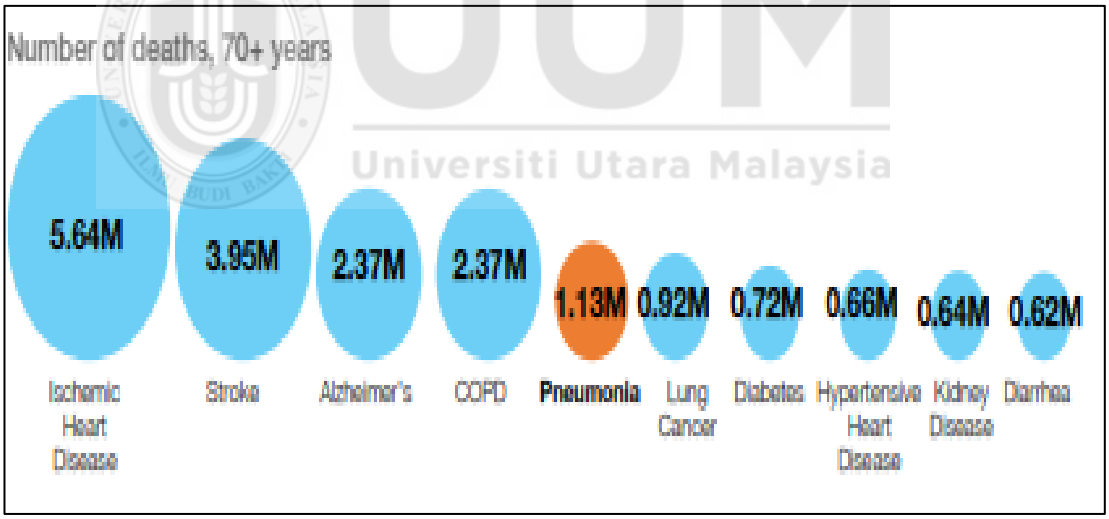

Figure 1.6. Top 10 causes of death among adults over 70 globally, 2017 (Source: Greenslade, 2018)

In contrast, many middle-income countries from South Asia and South-East Asia carry a "double burden" of pneumonia mortality, with significant deaths among both children and the elderly each year. For example, in South-East Asia, 14% of pneumonia deaths are among children under five and 60% are among adults over 70

years, while in South Asia, 40% of pneumonia deaths are among children, and 34% are among the elderly. Eastern Europe, Central Europe, North Africa and the Middle East, Latin America and the Caribbean, and Central Asia also carry a “double burden” of pneumonia (Figure 1.7). According to UNICEF (2019), pneumonia disease is still leading the number of deaths for children under age five, especially for infectious disease with 802,000 number of deaths (Figure 1.8).

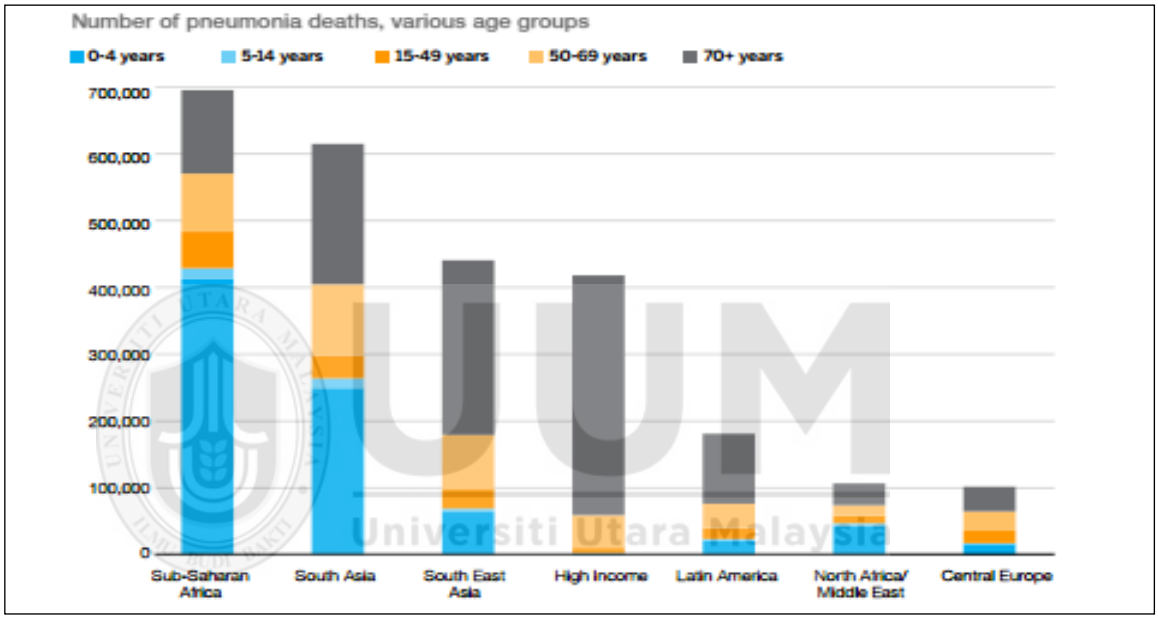

Figure 1.7. Pneumonia deaths concentrate in specific regions among specific age groups (Source: Greenslade, 2018)

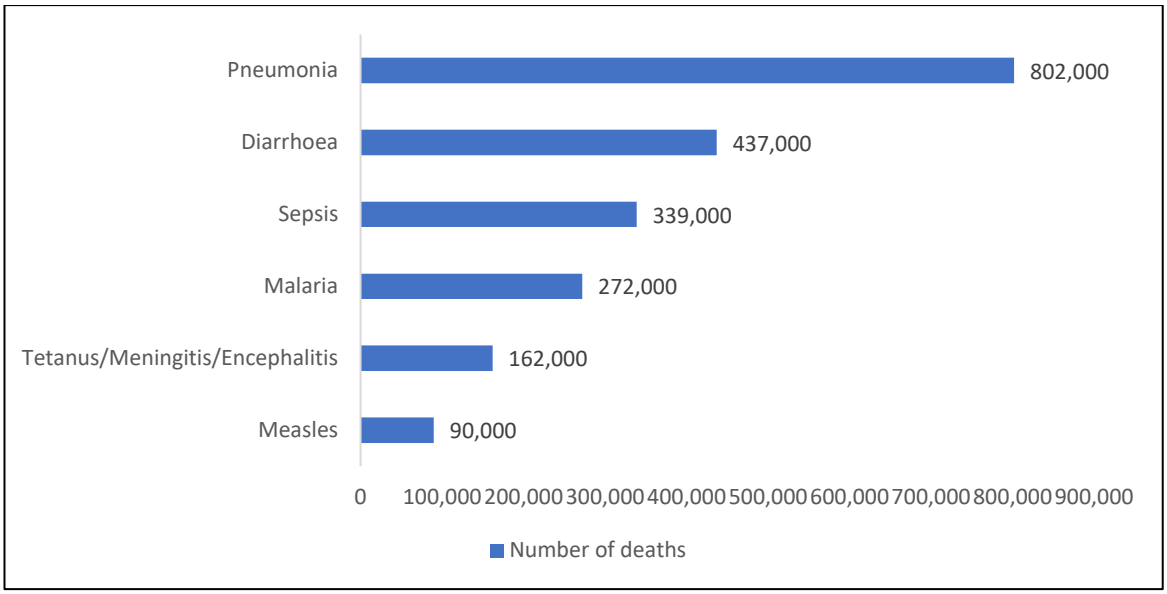

Figure 1.8. Death of children under-five by infectious disease in 2018
(Source: UNICEF, 2019)

In 2019, pneumonia is still becoming the single largest infectious killer of adults and children, killing over 2.5 million people, including 672,000 children (Stop pneumonia, 2020). While in 2020, there is a rising rate of pneumonia deaths from Corona Virus Disease 2019 (COVID-19), which has contributed 1.9 million to the death toll. This has the potential to rise 'all-cause' pneumonia fatalities by more than 75%. This burden of mortality is not caused by any other diseases (Stop pneumonia, 2020).

According to WHO (2020a), complication of someone having influenza and COVID-19 can result in pneumonia disease. These two diseases have the potential to lead an increase in the number of pneumonia cases (see Table 1.1). As the number of pneumonia cases rises because of the other diseases, it might become outbreak if it does not control. So, when the disease become outbreak for example as what happened in China especially Wuhan and also in Malaysia when COVID-19 becomes epidemic, it badly impacts the country. There will be broken global supply chain, closed offices

and factories. It also affects the import and export industries and impact the tourism industry. Not only the country's economy that will suffer, the people will also suffer. They need to be quarantined inside their house which may lead them to have psychological issues such as emotional stress, besides face difficulties in getting the food. The city is in lockdown which leads to the transportation system to be suspended.

Table 1.1

Influenza vs COVID-19 vs Pneumonia. Source: WHO (2020a) and CDC (2021)

	INFLUENZA	COVID-19 (CORONA VIRUS 2019)	PNEUMONIA
Example of viruses	Virus influenza A, B or C	Virus influenza A, B or C	Virus: adenoviruses, which can also cause the common cold and bronchitis, chickenpox (*Varicella zoster* virus), flu (influenza viruses), respiratory syncytial virus, which causes cold-like symptoms and virus influenza A and B
Transmission	Through direct contact when infected person cough or sneeze and through touch	Through direct and in-direct contact when infected person cough and got flu, through touch, and through unclean and unwell cooked food	Through direct contact when infected person cough or sneeze, or through the mouth or eyes when a person touches a surface that an infected person has coughed or sneezed on without washing their hands

Table 1.1 (continued)

Symptoms	Fever, chills, sweating, runny or stuffy nose, cough, or body aches, sore throat, headaches, muscle or body aches, sore throat, fatigue, nasal congestion	Fever, runny or stuffy nose, cough, headaches, shortness of breath, muscle or body aches, sore throat, pneumonia	Fever, chills, sweating, coughing that may produce mucus, chest pain, shortness of breath
Complications	**Pneumonia**, bronchitis, asthma flare-ups, heart problems, ear infections	**Pneumonia**, sepsis, high fever, severe cough, organ failure and death	Respiratory failure. Sepsis, a condition in which there is uncontrolled inflammation in the body, which may lead to widespread organ failure. Acute respiratory distress syndrome (ARDS), a severe form of respiratory failure. Lung abscesses, which are infrequent, but serious complications of pneumonia. Death.
Treatment	Conservative, anti-virus	Investigational treatments are available	Don't have any protocol treatment specialize yet
Vaccination	Have and need to take every year	Have and need to take 2 injections	Have and need to take 3 injections

1.1.3 Pneumonia in Malaysia

In Malaysia, pneumonia is one of the top three causes of mortality in adults, and the top three causes of death for children under the age of five (Department of Statistics Malaysia, 2019). Based on the Department of Statistics Malaysia, in 2014, the number of deaths due to pneumonia was 9,250 cases, which accounts for 12% of the total number of death causes (Figure 1.9). From that, 138 cases were among children under-five, which is recognized as the third leading cause for the deaths of children under-five after certain conditions originating in the prenatal period and congenital malformations, deformations, and chromosomal abnormalities (Figure 1.10). In 2014, it was reported that the number of deaths due to pneumonia recorded the highest percentage among female with 13.3%.

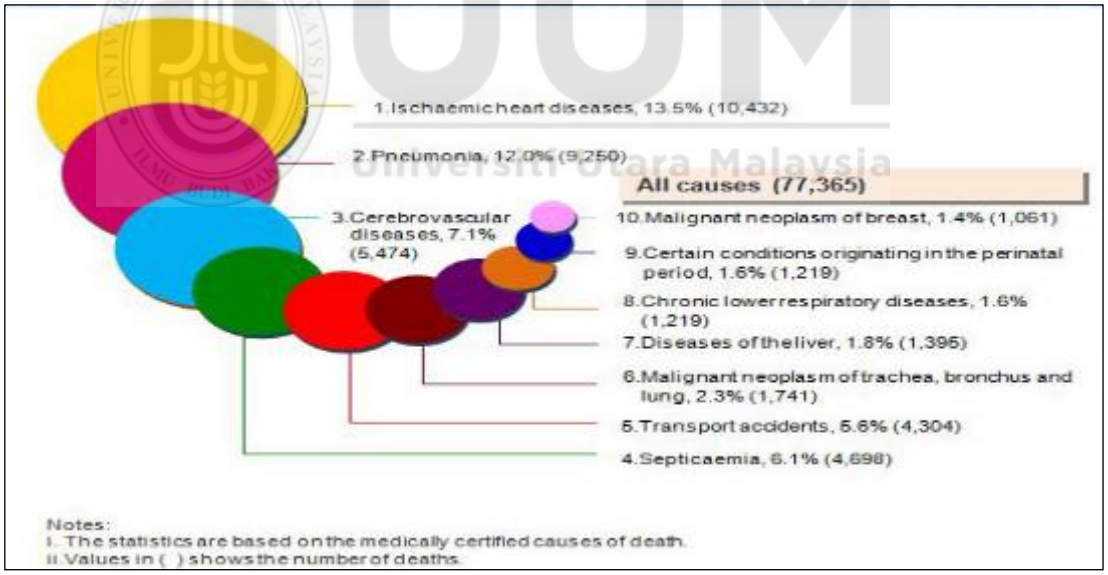

Figure 1.9. Ten principal causes of death in Malaysia in year 2014 (Source: Department of Statistics Malaysia, 2016)

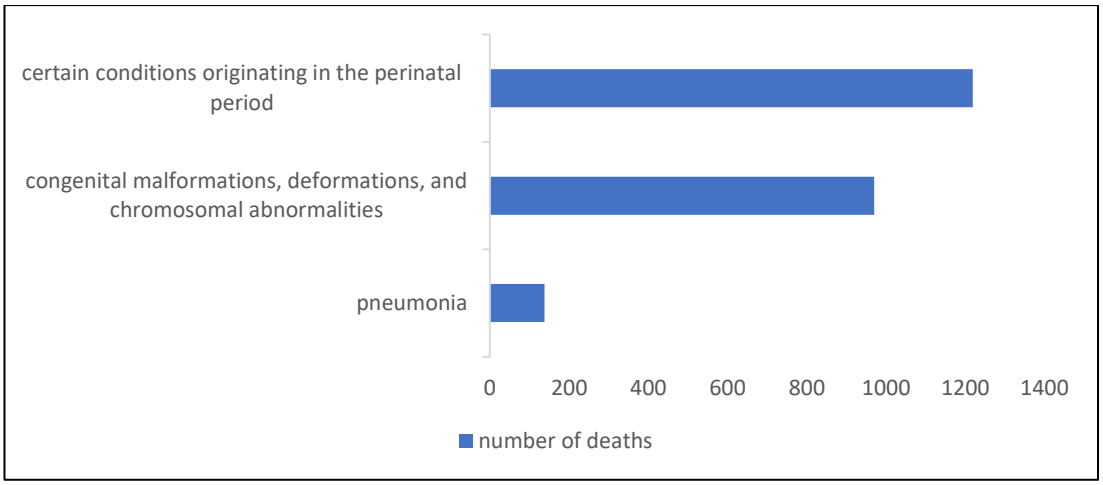

Figure 1.10. Top three causes of death for children under-five in Malaysia for the year 2014 (Source: Department of Statistics Malaysia, 2016)

It is essential to look into this problem as pneumonia is often associated with influenza, known as the top killer in Malaysia. In 2015 and 2016, it has been reported that pneumonia is the second-highest killer among Malaysians causing 10,200 and 10,678 deaths, respectively (Department of Statistics Malaysia, 2017). The number of deaths in children under-five years in the year 2015 and 2016 are 178 (5.2%) and 135 (3.8%), respectively (Figure 1.11). In 2015 and 2016, deaths due to pneumonia recorded the highest percentage among female with 14.1% (4,396 cases) and 14.0% (4,637 cases), respectively. Meanwhile, in 2017, pneumonia remained the second principal cause of death in Malaysia with 12.7%, after ischaemic heart diseases (Department of Statistics Malaysia, 2018). Most of these deaths concentrated in rural areas with 13.4%. In 2017 and 2018, it has been reported that pneumonia became the highest causes of death for age between 0 until 14 years old with 4.1% and 4.8%, respectively (Department of Statistics Malaysia, 2019). It was also reported that pneumonia remained the leading cause of death for female from year 2015 until 2019. The total number of pneumonia cases reported in Malaysia for year the 2019 is 145,419 cases with 7,542 number of deaths. In 2018 and 2019, pneumonia remained the second

principal cause of death in Malaysia (Figure 1.12) (Department of Statistic Malaysia, 2020).

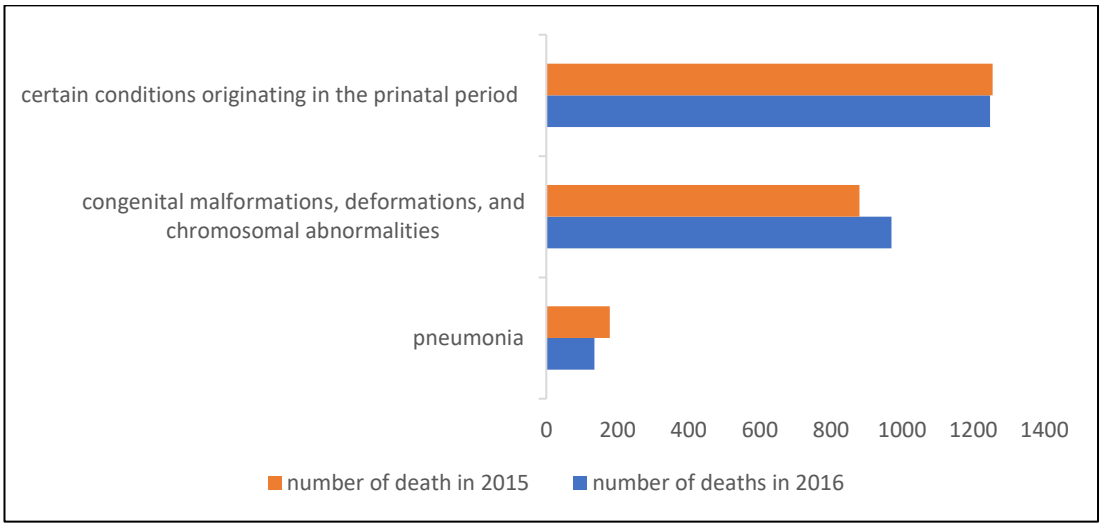

Figure 1.11. Top three causes of death for children under-five in Malaysia for year 2015 and 2016 (Source: Department of Statistics Malaysia, 2017)

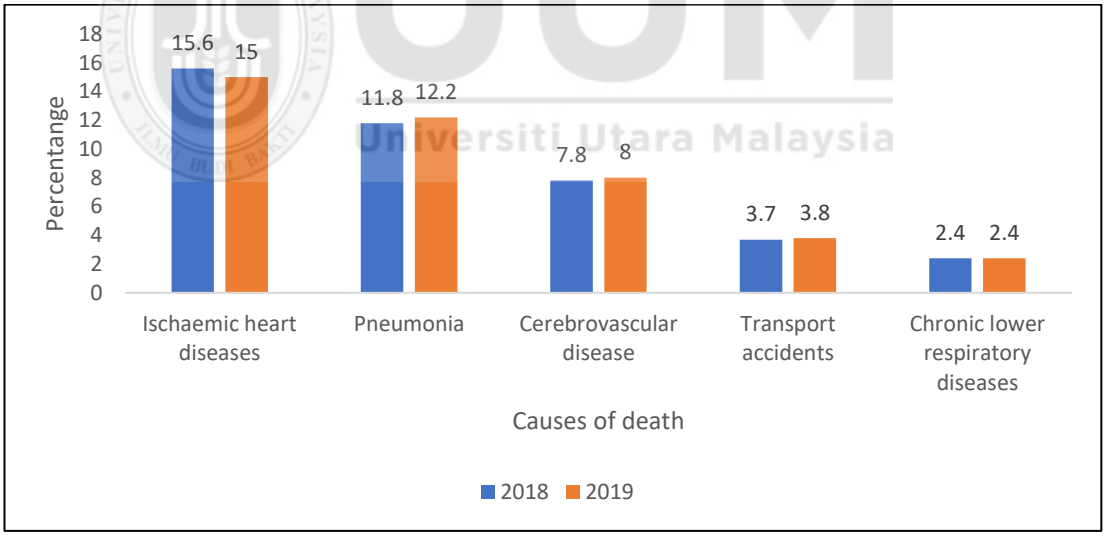

Figure 1.12. Five principal causes of deaths in Malaysia for the year 2018 and 2019 (Source: Department of Statistics Malaysia, 2020)

Based on the informations above, it can be concluded that pneumonia is one of the highest causes of death for children aged under five and also for elderly in Malaysia and worldwide. If this condition continues, it will have an impact on the number of

future population, whereby when the population decreases, this situation will affect the development of the country in the near future. The young generation is a country' asset who will lead the country in the future. A small population affects the country's productivity, thus affecting the country's economic growth.

To prevent pneumonia disease from worsening, Ministry of Health Malaysia (MOH) has taken action to create awareness among the citizen through a lot of campaigns regarding pneumonia and one of it is by celebrating Pneumonia Day on every 12th of November since 2009 to educate people about this disease. Our government also encourage parents to take pneumococcal vaccines for their baby. Prior to 2019, in Malaysia, this type of vaccine is not provided in our government clinic or hospital. Therefore, this vaccination is not made compulsory for all children under the age of two. However, the government encourages the parents of children under the age of two to take pneumococcal vaccines at any non-government clinics. According to former Health Minister, Dr. Dzulkefly Ahmad, the Health Ministry is considering making pneumococcal vaccines to be mandatory for children under the age of two (Chung, 2018). The vaccination is expected to be included in the National Immunisation Programme (NIP) by year 2020. Based on the presentation of the national budget for 2020, this pneumococcal vaccinations has been included in NIP starting year 2020. However, the government just cover the first injection only. According to Malaysia Health Director, General Tan Sri Dr. Noor Hisham, starting 1st December 2020, this pneumococcal vaccination now is provided for free and it is available at public health facilities (Bernama, 2020). This shows that the Malaysian government has lately begun to pay close attention to pneumonia disease by offering free Pneumococcal vaccinations, despite the fact that this type of vaccine is highly expensive. Besides, to monitor the number of pneumonia cases in Malaysia, the

Ministry of Health records the incidence cases yearly according to the locations which may helps to monitor the pneumonia cases from worsening.

Even if there is a cure for pneumonia, we need a method to monitor and limit the disease's spread in order to lower the risk. Hence, the next section will discuss the motivation that drives this study.

1.2 Motivation

1.2.1 Safeguarding the Community

Pneumonia is a disease, known as one of the numerous reasons, that causes millions of deaths throughout the world, especially among children below the age of five years (WHO, 2021). As mentioned earlier, many factors can lead a person to get pneumonia disease, for example, bacteria, viruses and other infectious agents. According to Leino, Auranen, Jokinen, Leinonen, Tervonen, and Takala (2001) and Hill et al. (2010), most studies suggest that children are the source of transmission of pneumonia disease to adults in the family. Moreover, Yoneyama (2010) stated that if this type of infection is uncontrolled, it can lead to an epidemic or a pandemic. An epidemic is a local spread of infectious disease within a region or a country, while pandemic is an infectious disease which spreads worldwide.

Since most children's death below the age of five are due to pneumonia disease, it will threaten the future, specifically on the population size in our country. According to Nor (2015) and Dr. Dzulkefli Ahmad, former health minister (in Hussin, 2019) stated that Malaysia is going to face ageing population in year 2030. It means that there are more people aged 70 years and older than there are children younger than five years

old. This will lead to zero population growth (ZPG), which occurs when there is no change in the number of people in a given time (birth and death rates are equal). There will be a change in the population pyramid structures. When the number of populations decreases, this will create a labor shortage, which then affect the country's economic growth in the future. There will be elderly aged 60 years and above in the future that are still in labor force. This situation can be seen in countries such as China and Japan where both countries have been recognized as super-ageing population countries with extremely low population share of children aged 0-14 years old (Johnston, 2019).

Having ageing population will negatively impact several aspects such as economy, policy challenges, tax and scientific and innovation products.

1) Impact on economy – when the population decreases it will create a labor shortage. Hence, we need to call labor from the outside (migrant workers). However, this will also open our country to illness. In Sabah, it is reported to has the highest number of TB occurrence and Sabah recorded the highest number of illegal immigrants in Malaysia (Diah, 2017).
2) Policy revision – government need to revise a lot of policy for example retirement policy and medical care policy. For retirement policy, government consider revising the new age for retirement. While for medical care for elderly, government need to increase expenditures for aged related programs, particularly healthcare expenditure. This will need high cost of providing healthcare and care homes since elderly tend to get ill more frequently.

3) Impact on tax – ageing population will lead to fewer working people. There will be reduced taxation income for the government as there are shrinking working population who are taxpayer.
4) Scientific and innovation products – when there are fewer young people in population, this will reduce of scientific and innovation products (Wang, 2019). This is because, usually young people have a lot of ideas of innovation. So, when these ageing population happen, there is a risk that productivity and economic growth will reduce.

Daley and Gani (2005) and Giodano et al. (2020) stated that modeling infectious of disease is a tool that can predict the future course of a disease outbreak and strategize the control of epidemics. In providing profound understanding on the importance of epidemiological processes, the transmission dynamics in a host population, formulation or implementation of disease control programs and the course of infection within a host, mathematical models have been acknowledged as powerful instruments (Hethcote, 2000; Mohammed Kizito & Tumwiine, 2018).

1.2.2 Transforming Traditional Approach of Case Monitoring

Currently, in Malaysia, the approaches used in estimating the low-risk and high-risk regions is still carried out by using the traditional approach which depend on the total number of pneumonia cases reported for each region. This approach only shows brief information without considering the disease transmitted vicinity. Before the area is marked as a low or high-risk area for disease occurrence, other factors such as the number of people in the population, the geographical area and so on need to be considered, otherwise it leads to biasness. This is because in some area, the number of

populations is small, and it might turn up that the number of cases is small too. Contrarily, if the geographic area is large, the number of populations will be substantial, and the number of cases might be high. For instance, the number of cases recorded in Perlis is assumed to be 100 people with 254,400 number of population while in Pahang is assumed to be 300 people with 1,670,000 number of populations, it is not correct to conclude Pahang as a high-risk area since the population size in Pahang is more than the population size in Perlis. Moreover, the region area of Pahang is much larger compared to Perlis.

Hence, by obtaining the geographic area information, precise information in terms of the number of incidences in one country can be obtained. This can be performed through disease mapping. According to Samat and Percy (2012), disease mapping can be utilized as a precautionary measure and disease prevention strategy. It can guide a more efficient allocation of resources in highly needed areas. A proper disease mapping depends on the modeling used when estimating the relative risk In Malaysia, there have been studies on dengue disease (Samat & Percy, 2012; Kristiani, & Samat, 2019), HIV and AIDS (Ideris, 2016), leptospirosis (Awang, 2017; Ideris et al., 2021), and tuberculosis (Diah, 2017). To-date, there is no work done on pneumonia disease using disease mapping in Malaysia. Most of the pneumonia studies are analyzing the stability of equilibrium points and understanding the dynamic transmission of pneumonia.

1.2.3 Incorporating Random Effects

A proper disease mapping can clearly illustrate the risk areas, depending on the relative risk estimation value. When investigating the geographical distribution, it is essential

to consider estimation of relative risk for disease mapping, which is still a continued study (Lawson, Browne & Rodeiro, 2003). Hence, there is a necessity to study the best approach. In the study of relative risk estimation of disease mapping, Standardized Morbidity/Mortality Ratio (SMR) has been used as the common way in estimating the relative risk value. Using of SMR directly may not be worth as it is incapable of detecting the small areas (Meza, 2003; Lawson et al., 2003; Awang & Samat, 2017) and does not take into consideration the high diversity of different regions and the spatial patterns of the regions under study (Lawson, 2003; Ideris, Malim & Shaadan, 2019). SMR is based on ratio estimators, which implies that using it might result in large changes in the estimate, and very small changes in the expected value (Lawson et al., 2003).

On the other hand, the Bayesian approach is highly suggested in the use of small area estimation since it smooths the relative risk and provides the measures of uncertainty associated with this relative risk estimation. For this method, the modeling takes into account the spatial autocorrelation. The approach to smoothing in the Bayesian method is by borrowing strength values from geographically referenced neighbouring values. Poisson-gamma model is one of the earliest Bayesian methods which is used by quite number of researchers (Lawson et al., 2013). However, there are still drawbacks of using this method as it cannot allow covariate for adjacent areas and the addition of a new parameter into the model. Also, it does not consider the disease transmission. Another early Bayesian model is Besag, York and Mollie (BYM) model which empirical Bayesian inference is apply for relative risks. This approach also has drawbacks where it sometimes smooths over large discontinuities for risk estimated, which may be crucial to maintain. This means that some risk surfaces

undergo over-smoothing and the risks produced might not depict the true risk (Lawson et al., 2013).

In order to overcome the drawbacks from SMR, Poisson-gamma model, and BYM model, researchers used transmission model or mathematical modeling (Samat, 2012). However, most of the mathematical modeling, especially for pneumonia disease, utilizes the deterministic model. As mentioned earlier, even though many researchers much prefer this type of model, it does not consider any possibility of random effects. A random effect is an additional of variance component that can be estimated on a map and associated with a specified probabilistic structure. In other words, the random effect is treated as a nuisance factor. Therefore, in this study, stochastic pneumonia disease transmission models which consider the parameters that are significant for pneumonia disease are proposed. Also, the random effect of spatial prior distribution is used, which allows dependency between adjacent areas.

According to Lawson et al. (2003), random effects are the effects that should be included within the analysis of the case that unobserved effects that might be thought to exist within the observed data. These effects are often termed as random effects, and their analysis has provided a large literature both in statistical methodology and in epidemiological applications (see, for example, Clayton 1991; Lawson, 2001). In the mapping context, a random effect could take a variety of forms. In its simplest form, a random effect is an extra quantity of variation or variance component which is estimable within the map, and which can ascribe a defined probabilistic structure. This component can affect individuals or can be associated with tracts or covariates. For

instance, individual susceptibility to disease varies and so individuals who become cases may have a random component relating to varied susceptibility. This is frequently referred to as weakness. Depending on the different level of background knowledge regarding the trait of this additional variation, the components would likely to possess different forms. For example, when observed counts, which are assumed to be regulated by the Poisson distribution, show greater variation than expected (i.e. variance > mean), it is sometimes described as overdispersion. There are many factors contributing to overdispersion such as at a particular scale where clustering happens in the counts and considerable numbers of cells have zero counts (sparseness). This could arise when rare diseases are mapped. Distinguishing the two basic forms of extra variations is vital in spatial applications. First, uncorrelated heterogeneity is a form of independent and spatially uncorrelated extra variation which happens as in aspatial case (see, for example, Besag, York & Mollie, 1991). Another example of where random effect form could arise is where a model is correlated with neighbouring spatial units (such as tracts, regions or case-events). This situation is often called correlated heterogeneity. In essence, extra variation means that there is spatial autocorrelation between spatial units. Various reasons can give rise to this autocorrelation. First, the disease of concern might be naturally clustered in its spatial distribution at the observation scale. Many infectious diseases display such spatial clustering, and a number of apparently non-infectious diseases also cluster. Second, autocorrelation can be induced in spatial disease patterns by the existence of unobserved environmental or frailty effects. As a result, extra variation observed in any application could arise from confounding variables that not included in the analysis. In disease mapping examples, this can easily arise when simple mapping methods are used on SMR with just basic age-sex standardization.

1.3 The Role of Mathematical and Statistical Modeling in Pneumonia Disease Control Research

Since SMR, Poisson-gamma models and BYM model possess some drawbacks, researchers have used transmission model or mathematical modeling to overcome the drawbacks. From previous studies, researchers came out with mathematical modeling to understand more about the transmission of pneumonia disease. According to Mcbryde (2006) and Assab et al. (2017) there are two reasons why mathematical models are helpful in controlling disease. Firstly, it can be used to forecast the development of an epidemic quantitatively, for instance, the peak of its total size, peak time, and also the repercussion of infection control interventions. This development includes non-linear interaction that occurs when multiple interventions are undertaken. Secondly, in order to refrain the assumption of serial independence and to handle the interval censoring as well as the unknown number of infectious diseases, mathematical model is used as it can provide information of trials design and statistical analysis structure. Besides that, mathematical models can also show how the progress of infectious disease happens. Susceptible–Infected–Carriers (SIC) model is the fundamental mathematical modeling for pneumonia.

Doura, Melendez-morales, Meyer and Perez (2000) presented two models known as the Susceptible–Infected–Carriers (SIC) model and the Susceptible–Infected susceptible–Carriers–Infected carriers (SI_SCIc) model which describes the dynamics of the population using streptococcus. From this basic SIC model, other researchers built other models such as the Susceptible–Carriers–Infected–Recovered (SCIR) by Otieno, Joseph and John (2012); the Susceptible–asymptomatic Infectives–symptomatic Infectives–treated Infective (SI_cI_iT) model by Ndelwa, Kgosimore, Massawe, and Namkinga (2015); Susceptible–Exposed–Infectious–Recovered (SEIR)

model by Kassa and Murthy (2016) as well as Susceptible–Vaccinated–Carrier–Infected–Recovered (SVCIR) model by Tilahun, Makinde and Malonza (2017). Recently, Soliman and Bueno (2018) used basic Susceptible–Infected–Recovered (SIR) model to analyze and predict transmission rate of pneumonia in Philippines. Mbabazi, Mugisha and Kimathi (2019) formulated Susceptible–Vaccinated–Exposed–Carrier–Infected model (SVECI model) which is a time delay model of pneumonia. However, most of the mathematical modeling, especially for pneumonia disease, uses deterministic models. As mentioned earlier, even though many researchers prefer this type of model, it does not consider any possibility of random effects. From previous studies, there are only two studies on pneumonia disease that used stochastic model (Smith et al, 1993 & Melegaro et al., 2002). Therefore, in this study, the stochastic pneumonia disease transmission models are proposed where random effects that are significant in pneumonia disease are considered in the model. The models considered in this study are SIC, SIR, SCIR, and SCVIR. These models are selected as they contain the appropriate parameters to be applied to Malaysia data. Details regarding all the models are explained in Chapter Two.

1.4 Research Objectives

The main objective of this study is to develop four stochastic models for pneumonia disease transmission in Malaysia. To accomplish this primary research objective, the following sub-objectives need to be achieved:

i. To construct the stochastic SIC, SIR, SCIR and SVCIR models.
ii. To estimate the relative risk based on the four stochastic models by using real data set.

iii. To compare the performance of the four stochastic models with existing methods by using relative risk estimation.

iv. To demonstrate pneumonia risk maps developed by using the relative risk estimation.

1.5 Significance of the Study

The significance of this study is to produce disease maps by using the alternative methods proposed. The maps are good references to assist the government in such a way to monitor the number of reported cases of pneumonia and the risk of this disease occurrence in Malaysia. Based on these maps, the government can identify whether further action such as the policy of government and financial aid are required.

Even though there is a treatment for pneumonia, we need a strategy to monitor the problems of this disease in order to reduce the high global burden of pneumonia. This can be done through disease mapping, where a clear picture of the hazard regions for disease occurrence can be shown (Diah, 2017). According to Thomas, Anthamatten, Root, Lucero, Nohynek, Tallo and Maleckar (2015), mapping can guide a more efficient allocation of resources. Once we identify the hazard areas, the government can focus on these areas and provide better facilities and services. Prevention programs can also be conducted efficiently. According to Lawson and Williams (2001), in disease mapping, usually mathematical modeling is used to describe the distribution of disease on the overall risk map. The analysis purpose is to determine the relative risk of diseases and to remove noise in the disease area.

Besides that, this study also proposes an alternative model in estimating risk for pneumonia disease which could overcome the drawbacks of Standardized

Mortality Ratio (SMR), Poisson-gamma model and BYM model. Stochastic terms are introduced into four disease transmission models proposed. These stochastic transmission models are SIC, SIR, SCIR and SVCIR which 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 carriers and vaccination elements in computing the relative risk. Most of the studies just considered the infectious element only. Hence, this study added value for carrier and vaccination elements as stochastic SVCIR model consider the infectious, carrier and vaccination components, while SIC and SCIR consider the infectious and carrier elements. These three proposed stochastic models are able to handle carrier cases.

CHAPTER TWO
MATHEMATICAL MODELING OF INFECTIOUS DISEASES

2.1 Introduction

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 basic mathematical modeling in order to gain better understanding of disease transmission, followed by previous studies on pneumonia disease using deterministic models and stochastic models. Other diseases that used stochastic model are also discussed in this chapter. Next, this chapter briefly describes the disease mapping analysis in section 2.3.

2.2 Basic Concepts of Mathematical Modeling

The mechanistic description of the infection transmission between two individuals (susceptible and infectious individuals) is the central idea of the transmission models, which is contradicted to statistical models. Description of this mechanistic allows to describe the evolution of time outbreaks in mathematical terms and in this way connects the individual-level processes that contribute to the transmission of the disease in detail. Thus, developing a mathematical model helps to concentrate on the important processes involved in shaping the epidemiology of infectious diseases and to uncover the most distinctive and acceptable parameters to control. Various disciplines are integrated in developing a mathematical modeling, from mathematics/physics, clinical sciences, microbiology, biology, environmental, zoology, to social science and computer science.

According to Aron (2007) and Choopojcharoen and Ali (2012), 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 it is vital for models to obtain the significant behavior of interest and take in the important processes. Creating a mathematical model that clearly explains the rationale and allows others to examine them Hence, the mathematical model allows in-depth analysis, accurate and quantitative forecast.

Mathematical modeling can be divided into two categories that are deterministic (or transmission) and stochastic (or statistical). The fundamental structure for a movement study between different systems can be shown by the deterministic models or also known as compartmental models which contain a limited number of regions and specified rules by which persons move from one region to another via a series of differential equations. In order to understand better about the deterministic model, here is a simple analogy by thinking each compartment as a room in a house. A person who lives in that house will travel from one room to another room over time. The person may be found in the bedroom, in the toilet, in the living room or in the kitchen at any given time. It is impossible that the person is found simultaneously in two different rooms at the same time.

A deterministic model can be used to describe and determine averages on a population scale, and it is ideal for larger population. To examine this type of model, a difference equation or differential equation can be used. For difference equation, discrete time steps are used to explain the transition between the different disease

compartments whereas differential equation focusing on the continuous case. On the other hand, for stochastic model, researchers are able to determine the probabilities movements between the populations thus will provide the probabilities of that particular model outcomes (Achterberg, 2009; Awang, 2017). This also allows the random variation in one or more inputs from time to time. Other than that, as there is alternation of potential variation in disease, exposure risk and other vital causes such as in a small population, the stochastic model is also used. According to Keeling and Rohani (2007), all diseases are subject to stochasticity in terms of the nature of transmission opportunities. Hence, in principle, a stochastic model is more realistic than a deterministic model.

Modeling usually consists of four steps. The first step is constructing a flow diagram representing the natural history and transmission of infection. Next, from this diagram, a set of mathematical equations is written to express the process of transmission. The third step is to determine the appropriate values for the parameters used in the equations, while the last step is to solve the algebraic equations numerically. Computer simulation programs are usually used to solve these equations.

2.2.1 Deterministic Models for Pneumonia Disease

In pneumonia transmission study, many researchers have developed and proposed several pneumonia models. For example, Doura et al. (2000) developed a mathematical model for pneumonia transmission dynamics. This model consists of three-dimensional non-linear differential equations system that defines three subpopulations known as susceptible (S), infected (I) and beta-hemolytic carriers (C). In their study, the dynamics of outbreaks in populations infected with a causative agent in strep throat known as Streptococcal pyogenes (S. pyogenes) is analyzed, using a

Susceptible–Infected–Susceptible (SIS) model which include additional class of infectious carriers. The disease's long-term dynamics in the population is also studied by the authors. Two transmission models were proposed in their study: Susceptible–Infected–Carriers (SIC) model and Susceptible–Infected susceptible–Carriers–Infected carriers (SI_SCI_C) model.

Besides that, Otieno, Joseph and John (2012) developed a deterministic model SICR (Susceptible–Infected–Carriers–Recovered) for pneumonia disease transmission for children under five. A mathematical explanation of the transmission dynamics of pneumonia is given in the study, considering the roles of carriers and recovery measures in the transmission. In their model, they used the theory of ordinary differential equations and dynamical systems. Basic reproduction number, R_0, is derived in their study as they evaluated the stability of the equilibrium points and analysis of bifurcation. A bifurcation is a qualitative change in the nature of the solution trajectories because of parameter change. This point of the change occurs is known as bifurcation point.

Apart from that, Ndelwa et al. (2015) formulated a mathematical model for dynamic transmission of pneumonia with treatment and screening intending to understand the dynamic transmission of pneumonia and the impact of these interventions. In their study, they used SI_cI_iT (Susceptible–asymptomatic Infectives–symptomatic Infectives–Treated Infectives model) model where the dynamic of pneumonia infection is divided into four subpopulations known as susceptibles (S), the asymptomatic Infectives (or simply carriers) (I_c), the symptomatic Infectives (I_i), the treated Infectives (T). Results from the study showed better outcomes when combining both screening and treatment than just considering one of the interventions.

On the other hand, Kassa and Murthy (2016) formulated a deterministic Susceptible–Exposed–Infectious–Recovered (SEIR) model in studying the pneumonia transmission. They used data from Boloso Sore Woreda, Wolaita Zone, SNNPR, Ethiopia, for their study to estimate the effects of control measures, especially vaccines, and the transmission of Streptococcus pneumonia disease. Also, Tilahun, Makinde and Malonza (2017) developed and determined a deterministic model for spread of pneumonia in various sizes of population. In their model, optimal control and cost-effectiveness strategies on a vaccinated group were considered, which leads to five compartments in their model called as SVCIR model.

Next, Soliman and Bueno (2018) analyzed and estimated the global transmission rates of pneumonia in the Philippines by using deterministic modeling. They used basic deterministic compartmental model, known as Susceptible–Infected–Recovered model (SIR model), to examine the population changes that have resulted from the changes of demographic. Besides, the study also compared the performance of using vaccination and quarantine methods in reducing the spread of pneumonia disease. Mbabazi, Mugisha and Kimathi (2019) proposed a mathematical model with time delays for pneumonia called Susceptible–Vaccinated–Exposed–Carrier–Infected model (SVECI model). They used this model to analyze hoft-bifurcation of pneumonia. In this model, they assumed that the exposed class (E) is the asymptomatic persons while carrier class (C) is the individuals with one serotype not covered by the vaccine. In their study, they explore the effect of two delays on pneumonia disease. The first-time delay is in the exposed class since there is a delayed time when a person is infected and when a person becomes infectious. The second time delay is in getting medical care which is included in the infectious class. In order to analyze the model, they used stability theory of delay differential equation.

2.2.1.1 The Susceptible–Infected–Carriers (SIC) Model

The Susceptible–Infected–Carriers (SIC) model categorizes the population into three compartments:

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

b) Currently infectious (Infected or Infectious) – *I*,

c) Already infected and does not require external inoculation (Carrier) – *C*.

Here, the number of persons in each compartment is represented through *S*, *I*, and *C*, while the total host population is $N = S + I + C$. Figure 2.1 shows the flow of transmission process.

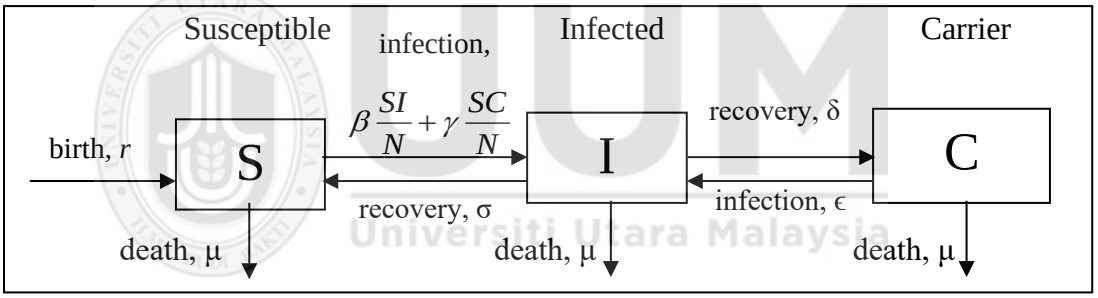

Figure 2.1. Flow of transmission process for SIC model.

From Figure 2.1, the arrows represent the movement between classes. If the arrow enters the compartment, it shows that the number of that compartment is increasing, while if the arrow leaves the compartment, it implies that there is a decreasing number in that compartment. In this model, the number of susceptible people (*S*) increases by the birth rate (*r*) and the infectious individuals recovered at the recovery rate (σ). On the other hand, the number of susceptible (*S*) decreases by natural (non-diseased) death rate (μ) and by infection transmission events rate of susceptible (β: the rate at which an infected person can infect a susceptible and γ: the rate at which a carrier can infect a susceptible). The number of infectious (*I*) increases by infection

events of susceptible and the number of individuals that become infected from carrier class (at rate ϵ). On the other hand, the number of infectious (I) decreases by the mortality rate (μ) and those who have recovered with the recovery rate (δ). Also, the number of carrier people (C) increases from the recovery rate (δ) and decreases by natural death rate (μ) and by carrier moving into the infectious class.

Based on the compartmental model in Figure 2.1, SIC model can be described mathematically in the form of differential equations as follow:

$$\frac{dS}{dt} = rN + \sigma I - (\beta \frac{SI}{N} + \gamma \frac{SC}{N}) - \mu S, \tag{2.1}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} + \gamma \frac{SC}{N} - (\sigma + \mu + \delta)I + \epsilon C, \tag{2.2}$$

$$\frac{dC}{dt} = \delta I - (\epsilon + \mu)C. \tag{2.3}$$

SIC model assumes that people begin with being susceptible towards the pneumonia disease, and then become infected by exposure to the infectious person. The susceptible class is completely free from streptococcus and only become infected through direct inoculation, which occurs in close contact with a carrier or infected categorized in the nasopharynx. A susceptible can get an infection by a strain found in a carrier on an infected. An infected person usually is more infectious than a carrier as the throat area carries and transmits a higher density of bacteria. A strain inside the infected person can be more infectious and harmful serotypes. Due to the lowered of virulence in the strain, individuals who carry streptococcus as part of their natural flora may be able to do so as their immune system can hold it in check. Infected or carrier

person, may be unaware of their condition, takes no actions or precautions to stop spreading the disease. Infected person in the infectious class moves to the susceptible class either by natural recovery (completely free from streptococcus) or by antibiotic treatment. However, natural recovery can result in colonies being established in the nasopharynx, which results in the status of the carrier.

In a case of a carrier is infected, he or she does not require external inoculation. When the immune system is affected by insufficient nutrition, stress, sleep, or an infection by a virus that diverts the immune system's recourses, a person in a carrier class will move into the infectious "sick" category. Within 24 hours, the colony can grow out of control and a person can move into the infectious class; or a person can cure without treatment back into the carrier class, or with treatment of antibiotics, a person moves into the susceptible class. In this model, Doura et al. (2000) made two assumptions. First, they assume a vast majority of recovered infectious to the susceptible class do so by antibiotic treatment. The second assumption is that even without antibiotics to the susceptible class, a carrier that moves into the infectious class can recover. However, a carrier must move through the infectious class first. In this SIC model, the movement of carrier directly into susceptible is not be studied.

2.2.1.2 The Susceptible–Infected susceptible–Carriers–Infected carriers (SI_SCIc) Model.

The Susceptible–Infected susceptible–Carriers–Infected carriers (SI_SCI_c) model is the second model proposed by Doura et al. (2000) in their study. This model used the same parameters as the Susceptible–Infected–Carriers (SIC) model. The only difference is that in this model, it has two more parameters *p* and *q* described as follows:

1) If p is the proportion of infected carriers, I_C that recovered to carriers' class, C, then 1-p is the proportion of infected carriers, I_C that recovered from susceptible class, S.

2) If q is the proportion of infected susceptible, I_S that recovered to susceptible class, S, then 1-q is the proportion of infected susceptible, I_S that recovered to carriers class, C.

There are four compartments in this model:

a) Susceptible to the disease (Susceptible) – S,
b) Infected susceptible – I_S,
c) Infected carriers – I_C,
d) Already infected and does not require external inoculation (Carrier) – C.

Here, the number of persons in each compartment represent by S, Is, Ic and C, while the total host population is $N = S + Is + Ic + C$. Figure 2.2 shows the flow of transmission process for this model.

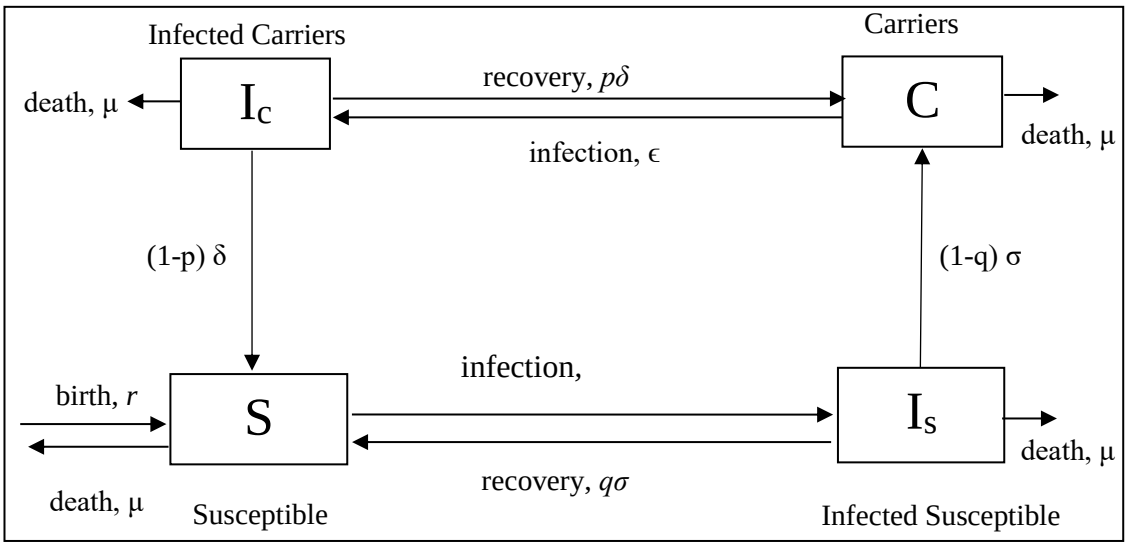

Figure 2.2. Flow of transmission process for SI_SCI_C model.

In SI_SCIc model, susceptible moves into and out of I_s after being infected by an infective or carrier. If a carrier's immune system is weak, an individual will move into the infectious class (I_c). When the signs of illness have been detected and a person is cured by antibiotics or recovers naturally (small percentage), the person move from I_c to susceptible class (S). If the infection goes untreated by antibiotics, a person can move from *Is* to carrier class (C). A few infections go completely unnoticed and this movement between classes happen quickly. This model assumed that the carriers (C) are not infectious.

The differential equations for SI_SCI_c model can be written as follows:

$$\frac{dS}{dt} = rN + q\sigma I_S + (1-p)\delta I_c - \left(\beta \frac{SI_S}{N} + \gamma \frac{SI_c}{N}\right) - \mu S \quad (2.4)$$

$$\frac{dI_s}{dt} = \beta \frac{SI_s}{N} + \gamma \frac{SI_c}{N} - (\sigma + \mu) I_s \quad (2.5)$$

$$\frac{dC}{dt} = (1-q)\sigma I_s + p\delta I_c - (\epsilon + \mu) C \quad (2.6)$$

$$\frac{dI_c}{dt} = \epsilon C - (\delta + \mu) I_c \quad (2.7)$$

2.2.1.3 The Susceptible–Infected–Carriers–Recovered (SICR) model

The Susceptible-Infected-Carriers-Recovered (SICR) model is proposed by Otieno, Joseph and John (2012). In this model, there are four compartments based on the disease status given as follows:

a) Susceptible – *S*,

b) Infectious – *I*,

c) Carriers – *C*,

d) Recovered – *R*.

The total population size, N at time t is $N = S + I + C + R$, while the per capita recruitment rate into the susceptible population is denoted by r. In this model, a new infection can be caused by an active contact with either a carrier or an infected individual at the force of infection of susceptible, denoted by λ. A newly infected person moves to carriers class (C) with a probability of ρ or moves into the infected class (I) with a probability of $(1 - \rho)$. Here, carriers refer to those individuals who carry the bacteria in their mouth or flora of nasopharynx of a person without causing any harm. When the immunity of the individual lowers, the bacteria in the carriers move to the lung, causing the infection. Carriers can become infected at the rate π, while infected individuals can recover at the rate η. The recovered individuals are clear from all the bacteria in their body and gain temporal immunity at proportion q, while $(1 - q)$ of them still carries the bacteria. The carriers can recover and gain temporary immunity at the rate β. According to Otieno, Joseph and Paul (2013), in this model, temporary immunity is an outcome of all possible ways that can lead to recovery from the illness. A person can get reinfected at the rate δ, a natural death rate (μ) and death caused by pneumonia disease at the rate α. The force of infection can be described as:

$$\lambda = \psi\left(\frac{I + \varepsilon C}{N}\right) : \psi = \kappa P \tag{2.8}$$

where

κ: rate of contact

P: probability that a contact is efficient to cause infection

Figure 2.3 represents the epidemiological of SICR model between classes of susceptible, carriers, infectious and recovered individuals in the population.

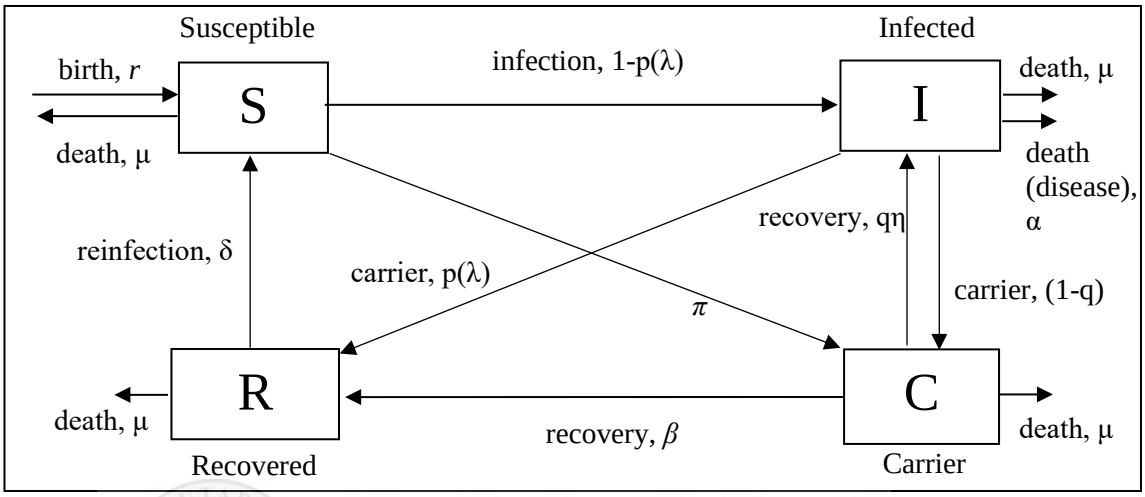

Figure 2.3. Flow of transmission process for SICR model

Based on Figure 2.3, the following system of ordinary differential equations are formulated to represent the model.

$$\frac{dS}{dt} = r + \delta R - (\lambda + \mu)S \tag{2.9}$$

$$\frac{dI}{dt} = (1 - p)\lambda S + \pi C - (\mu + \alpha + \eta)I \tag{2.10}$$

$$\frac{dC}{dt} = p(\lambda)S + (1 - q)\eta I - (\mu + \pi + \beta)C \tag{2.11}$$

$$\frac{dR}{dt} = q\eta I + \beta C - (\mu + \delta)R \tag{2.12}$$

2.2.1.4 The Susceptible-asymptomatic Infectives (or simply carriers) symptomatic Infectives-Treated Infectives (SI_CI_iT) Model

Ndelwa et al. (2015) described the dynamics of pneumonia disease by using four compartments namely

a) Susceptible – *S*,
b) Asymptomatic infectives (or simply carriers) – I_c,
c) Symptomatic infectives – I_i,
d) Treated infectives – *T*.

Figure 2.4 shows the flow of transmission in the SI_cI_iT model.

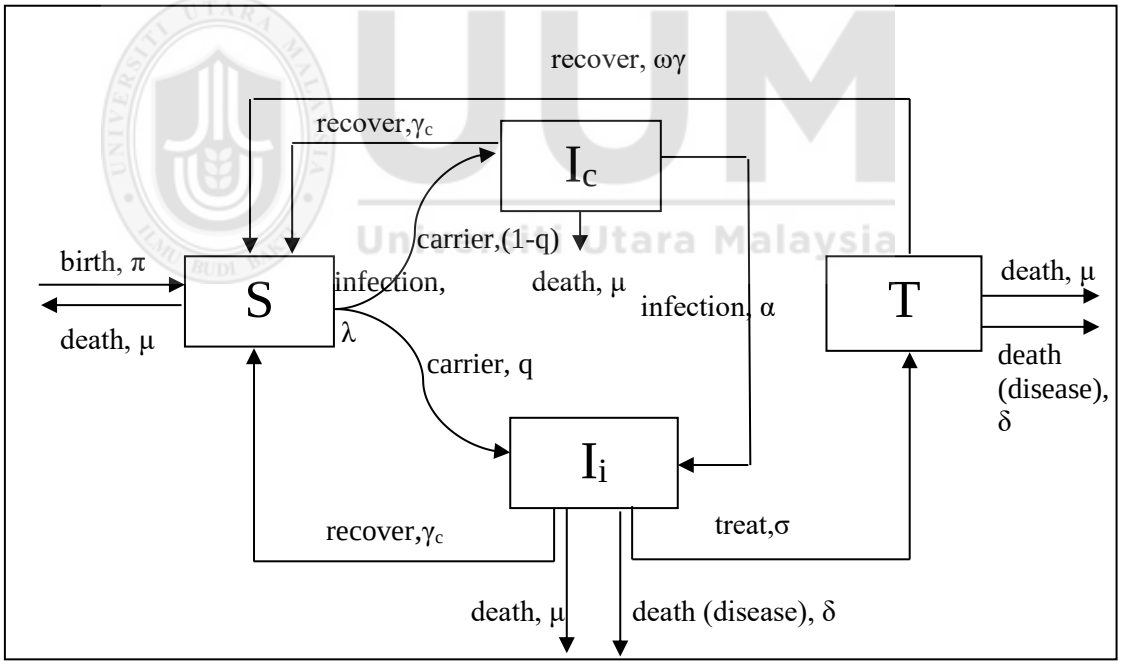

Figure 2.4. Flow of transmission process for SI_cI_iT model

Here, the number of individuals in each compartment is represented using *S*, I_C, I_i and *T*, where the total host population is $N = S + I_C + I_i + T$. It is assumed that the susceptible population is generated through birth at a constant rate, π and also by

treatment or by naturally infective recovery. The number of susceptible individuals (S) decreases by the infection rate with pneumonia (λ), where the proportion q becomes symptomatic infectious immediately and joins the symptomatic class (I_i). Here, the remaining proportion (1-q) becomes asymptomatic infectious (I_C). The number of susceptible people decreases through natural death at a constant rate (μ). Also, the number of asymptomatic infectious (I_C) decreases by screening method at a rate α where it progresses to symptomatic infectiousness (I_i). Besides, it minimises through recovery by gaining temporal immunity at rate γ_c and by natural mortality rate (μ). The number of symptomatic infectious (I_i) increases when the asymptomatic infectious change status and become symptomatic infectious through screening at rate α, and the infected proportion q of the susceptible individuals at the rate λ. It decreases by treating individuals at rate σ, by natural recovery rate (γ_i) and by mortality rate either due to pneumonia (δ) or by other causes (μ). The number of treated infective population (T) increases by the administration of treatment rate (σ), whereas treated infective recovers at a constant rate $\omega\gamma$ and joins the susceptible class. The number of treated infective (T) also decreases by both natural death rate (μ) and mortality rate due to pneumonia ($\tau\delta$). In SI_cI_iT model, ω and τ are the modification parameters with $\omega \geq 1$ implying that the treatment increases the recovery rate at $0 \leq \tau, q \leq 1$. A simple system of equations can be used to describe this model as follow:

$$\frac{dS}{dt} = \pi + \gamma_i I_i + \gamma_c I_c + \omega\gamma T - q\lambda S - (1-q)\lambda S - \mu S \tag{2.13}$$

$$\frac{dI_c}{dt} = (1-q)\lambda S - (\mu + \gamma_c + \alpha) I_c \tag{2.14}$$

$$\frac{dI_i}{dt} = q\lambda S + \alpha I_c - (\mu + \gamma_i + \sigma + \delta) I_i \tag{2.15}$$

$$\frac{dT}{dt} = \sigma I_i - (\mu + \omega\gamma + \tau\delta)T \quad (2.16)$$

2.2.1.5 The Susceptible-Exposed-Infectious-Recovered (SEIR) Model

Kassa and Murthy (2016) proposed two types of Susceptible-Exposed-Infectious-Recovered (SEIR) model in understanding the pneumonia transmission. The first one is without considering vaccination, while the other considers vaccination. Figure 2.5 shows the flow of the transmission process without vaccination.

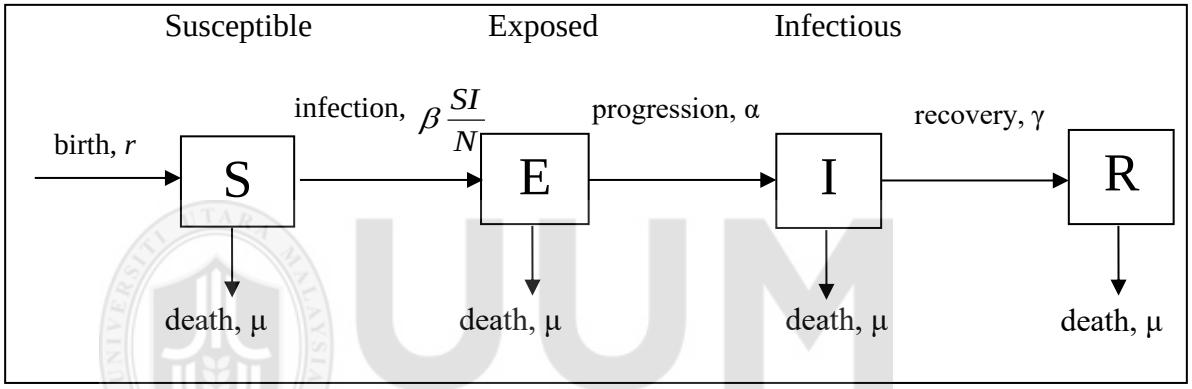

Figure 2.5. Flow of transmission process for SEIR model without vaccination.

In SEIR model, the population is divided into four compartments as follows:

a) Susceptible – *S*,
b) Exposed – *E*,
c) Infectious – *I*,
d) Recovered – *R*.

Here, the number of individuals in each compartment is represented by using *S*, *E*, *I*, and *R*, where the total host population is $N = S + E + I + R$. In this model, the number of susceptible individuals (*S*) increases by birth rate (*r*) while decreases by natural (non-diseased) death rate (μ) and by infection transmission events rate of susceptible

(β). In this study, it is assumed that the birth rate and death rate of human are equal. The number of exposed individuals (E) increases by the infection individual infected and decreases by the number of individuals who developed pneumonia slowly at rate α and natural death rate (μ). On the other hand, the number of infectious (I) increases by the number of individuals that develop pneumonia in the exposed state (at rate α), while the number of infectious (I) decreases by the natural mortality rate (μ) who recovered with the recovery rate (γ). Also, the number of recovered people (R) increased from the recovery rate (γ) and decreased by the natural death rate (μ).

The compartmental model in Figure 2.5 can be shown mathematically in the form of differential equations system as follows:

$$\frac{dS}{dt} = rN - \mu S - (\beta \frac{SI}{N}) \tag{2.17}$$

$$\frac{dE}{dt} = (\beta \frac{SI}{N}) - (\mu + \alpha)E \tag{2.18}$$

$$\frac{dI}{dt} = \alpha E - (\mu + \gamma)I \tag{2.19}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{2.20}$$

From the SEIR transmission model above, Kassa and Murthy (2016) extended the model by considering the effect of vaccination on the spread of pneumonia. In this model, the number of susceptible (S) increases by the birth rate (r). It decreases by testing and pneumonia therapy at rate δ and through infection transmission events of susceptible at rate β. The susceptible class also decreases by the rate of natural death (μ). The number of exposed individuals (E) increases through contact with the infected individuals at rate β and decreases by the number of individuals who developed

pneumonia slowly at rate α and natural death rate (μ). Apart from that, the number of infectious people (I) increases by the breakthrough of exposed individuals at rate α, while decreases by recovery rate (γ) and natural death rate (μ). This model assumes that both recovered susceptible individuals and recovered infected individuals become permanently immune to the disease. This generates the number of recovered people (R) who have complete protection against the pneumonia disease and decreases through the natural mortality rate (μ). Figure 2.6 shows the flow of transmission process for the SEIR model with vaccination.

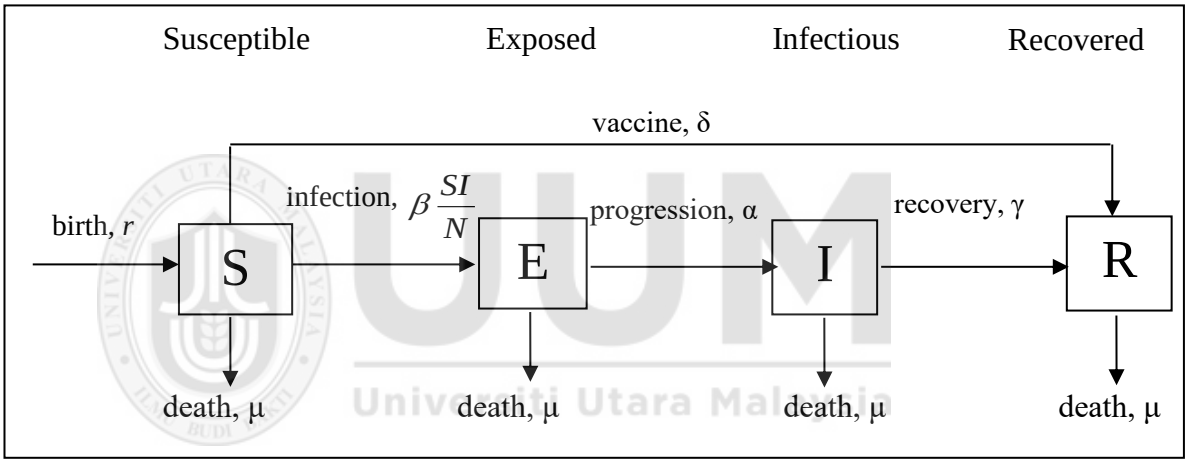

Figure 2.6. Flow of transmission for SEIR model with vaccination.

The form of differential equations system of the transmission process using mathematical expressions together with the parameters and population rates based on Figure 2.6, as follows:

$$\frac{dS}{dt} = rN - \mu S - (\beta \frac{SI}{N}) - \delta S \tag{2.21}$$

$$\frac{dE}{dt} = (\beta \frac{SI}{N}) - (\mu + \alpha)E \tag{2.22}$$

$$\frac{dI}{dt} = \alpha E - (\mu + \gamma)I \tag{2.23}$$

$$\frac{dR}{dt} = \gamma I + \delta S - \mu R \tag{2.24}$$

2.2.1.6 The Susceptible–Vaccinated–Carrier–Infected–Recovered (SVCIR) Model

Tilahun, Makinde and Malonza (2017) proposed a Susceptible-Vaccinated-Carrier-Infected-Recovered (SVCIR) model which divides the population into five sub-classes (compartments):

a) Susceptible – *S*,
b) Vaccinated – *V*,
c) Carrier – *C*,
d) Infected – *I*,
e) Recovered – *R*.

The total population at time *t* is given by $N = S + V + C + I + R$, where the SVCIR model assumed that before the disease outbreak, a proportion of the population was vaccinated at the rate of (p), while ($1 - p$) proportion is of susceptible population. The number of susceptible people (S) increases through individuals who are vaccinated but do not respond to vaccination with the waning rate (φ) and from individuals who lose their temporary immunity at rate δ. The susceptible class decreases by the individuals moving to vaccinated class with vaccination rate (ϑ) and through infection transmission events of susceptible either by carrier (C) or symptomatically infected individuals (I) with a force of infection $\lambda = \xi((I + \Upsilon C) / N)$ where $\xi = k\tau$. Here, k is a contact rate; τ is the probability that a contact is sufficient to cause infection and Υ is the transmission coefficient for the carrier. The model assumes that vaccination is not 100% effective, hence vaccinated class (V) has a probability to become a carrier or infectious with small fraction, where the infection force for the vaccinated class is

$\lambda v = \varepsilon\lambda$ where $0 \leq \varepsilon \leq 1$ and ε is the proportion of serotype not covered by the vaccine. Individuals who are newly infected by the force of infection become either carriers with a probability of ρ to join the carrier class (C) or infected with a likelihood of $(1 - \rho)$ to join the infected class (I). The carrier class can acquire symptoms of the disease or be able to screen themselves and join the infected class (I) at a rate of χ or recover by developing natural immunity at a rate of β. The number of infected persons (I) decreases by moving to recovered class (R) at a per capita rate of η, with treatment efficacy of q fraction of persons joining the recovered class (R) or joining the carrier class (C) with $(1 - q)$ fraction by adapting the treatment. The number of infected persons (I) also decreases through the mortality that caused either by pneumonia disease at rate (μ_p) or natural mortality at rate (μ).

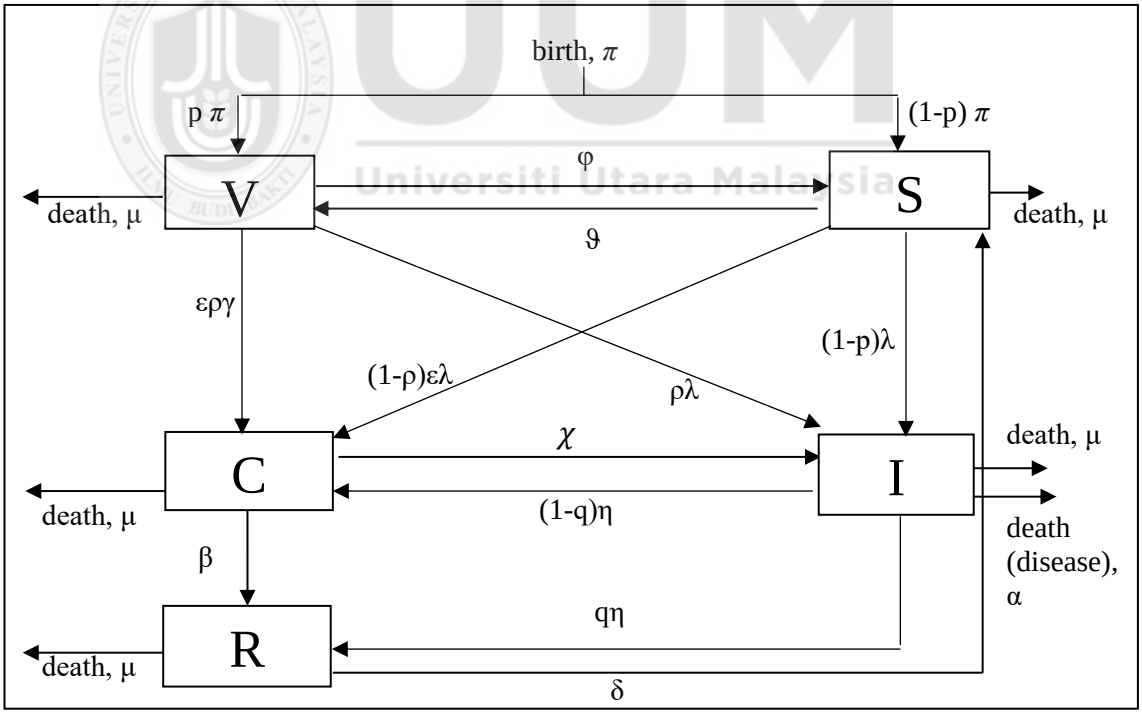

Figure 2.7. Flow diagram of SVCIR model

The form of differential equations system of the transmission process using mathematical expressions together with the parameters and population rates based on Figure 2.7, as follows:

$$\frac{dS}{dt} = (1-p)\pi + \varphi V + \delta R - (\mu + \lambda + \vartheta)S \tag{2.25}$$

$$\frac{dV}{dt} = p\pi + \vartheta S - (\mu + \varepsilon\lambda + \varphi)V \tag{2.26}$$

$$\frac{dC}{dt} = \rho\lambda S + \rho\varepsilon\lambda V + (1-q)\eta I - (\mu + \beta + \chi)C \tag{2.27}$$

$$\frac{dI}{dt} = (1-\rho)\lambda S + (1-\rho)\varepsilon\lambda V + \chi C - (\mu + \alpha + \eta)I \tag{2.28}$$

$$\frac{dR}{dt} = \beta C + q\eta I - (\mu + \delta)R \tag{2.29}$$

2.2.1.7 The Susceptible–Infectious–Recovered (SIR) Model

Basic mathematical modeling for infectious disease assumes that people being susceptible to infectious disease may become infected by exposure to contagious people, immediately they become infectious and most probably recover or die. Susceptible–Infective–Recovered (SIR) model is where the recovery empowers immunity to further infection, and they are said to be removed. Figure 2.8 shows the flow of transmission process for SIR model:

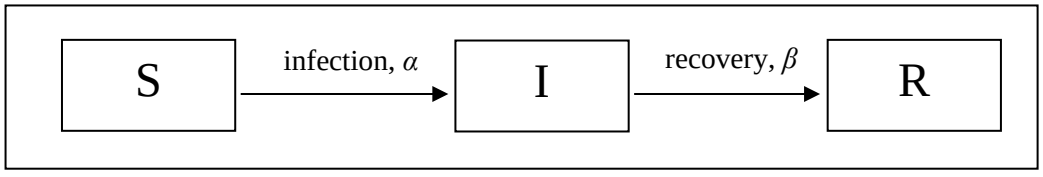

Figure 2.8. Flow of transmission process for SIR model.

The total number of host population is $N = S(t) + I(t) + R(t)$, where $S(t)$, $I(t)$ and $R(t)$ are the numbers in the population who are susceptible, infectious and removed at time t. Let $S(t)$, $I(t)$ and $R(t)$ be the numbers of the members of each class. In this model, Soliman and Bueno (2018) stated that the first parameter that was used in this model is the assumption of the constant and linear rate of change in the infection rate, α and the recovery rate, β of all compartments in the population. The second parameter implemented is the usage of two different population models to calculate the demographic changes due to instances of births and deaths in the population. Based on Figure 2.8, the three compartments are represented by the following mathematical equations:

$$\frac{dS}{dt} = -\alpha SI \tag{2.30}$$

$$\frac{dI}{dt} = \alpha SI - \beta I \tag{2.31}$$

$$\frac{dR}{dt} = \beta I \tag{2.32}$$

2.2.1.8 The Susceptible–Vaccinated–Exposed–Carrier–Infected (SVECI) Model

Figure 2.9 shows the flow of transmission process for SVECI model. Mbabazi, Mugisha and Kimathi (2019) described the dynamics of pneumonia disease by using five compartments namely

a) Susceptible – *S*,
b) Vaccinated – *V*,
c) Exposed – *E*,
d) Carrier – *C*,
e) Infected – *I*

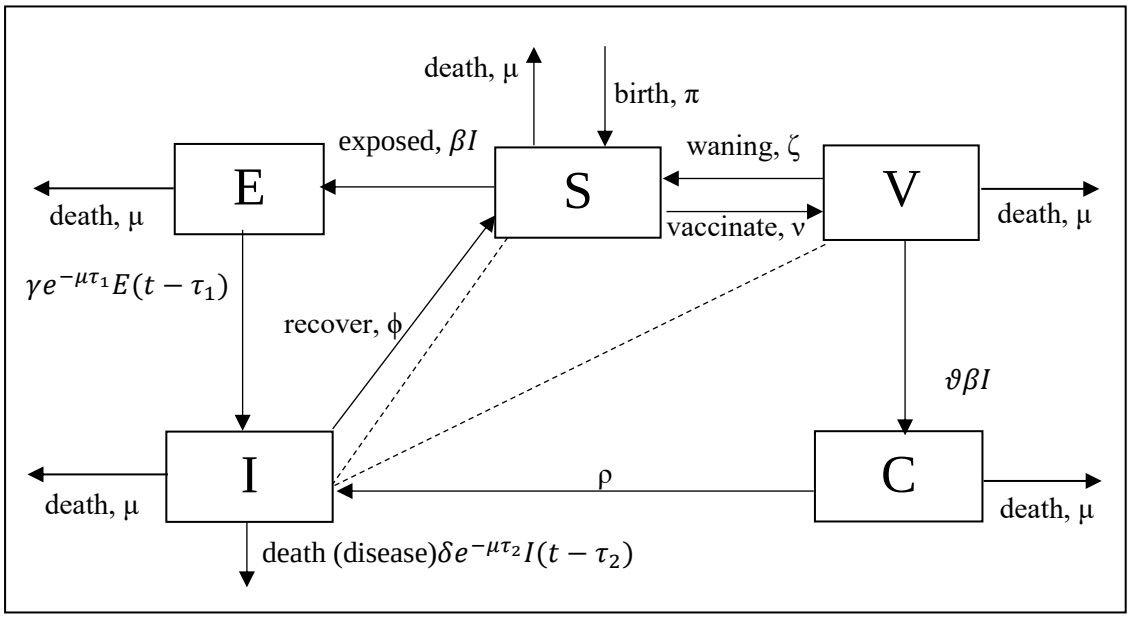

Figure 2.9. Flow of transmission process for SVECI model

The total population at time *t* is given by $N = S(t) + V(t) + E(t) + C(t) + I(t)$. SVECI model assumes that susceptible persons received a continuous vaccination strategy at rate *v* and the vaccination does not affect the infectious. Mbabazi et al. (2019) also assumed that the vaccination is not 100% effective where there is a chance for becoming infectious or becoming carrier in small proportions. Notation $\vartheta\beta I$ is the infection force for the vaccinated class where $0 \leq \vartheta \leq 1$ is the proportion of the serotype not covered by vaccine. The number of susceptible persons (*S*) increased through a constant recruitment which are from birth or migration and recovered persons. The number of susceptible persons (*S*) also increased when vaccinated persons (*V*) return to the susceptible class (*S*) when they did not respond to the vaccination at a waning rate (ζ) and when the infected persons upon recovery join the susceptible class (*S*) at rate (ϕ). The susceptible class (*S*) decreased by persons moving to exposed class (*E*) when become infected through an infection force ($\beta I(t)$). The exposed class (*E*), accounts for a time delay $\tau_1 > 0$ of the exposed persons, for example, the period between the time of an infection start and the time of developing pneumococcal

clinical symptoms, assume that a person is infected with exposure to influenza A disease that promotes severe pneumococcal pneumonia. The probability of a person surviving the natural death through the exposed period $[t-\tau_1, t]$ is $e^{-\mu\tau_1}$ and persons in exposed class (E) move to the infected class (I) at a rate (γ). The infected class (I) also increased when persons from the carrier class (C) become symptomatic and move to the infected class (I) at rate (ρ).

The infected class (I) accounts for a time delay $\tau_2 > 0$: the time taken by infected persons in order to seek medical care. In SVECI model, Mbabazi et al. assumed that the infected persons who survived the natural death through the infectious period $[t-\tau_2, t]$ have a survivorship function $e^{-\mu\tau_2}$. Besides, the infected class (I) decreased when infected persons that delay in seeking medical care die due to pneumonia disease at rate (δ). When infected people recover, they transfer to the susceptible class (S) at a rate of ϕ. All class exhibit a per capita natural death rate (μ). Based on Figure 2.9, the compartments are represented by the following mathematical equations:

$$\frac{dS}{dt} = b + \zeta V(t) + \phi I(t) - (\nu + \mu + \beta I(t))S(t), \tag{2.33}$$

$$\frac{dV}{dt} = \nu S(t) - (\mu + \zeta)V(t) - \vartheta\beta I(t)V(t), \tag{2.34}$$

$$\frac{dE}{dt} = \beta I(t)S(t) - \gamma e^{-\mu\tau_1}E(t-\tau_1) - \mu E(t), \tag{2.35}$$

$$\frac{dC}{dt} = \vartheta\beta I(t)V(t) - (\rho + \mu)C(t), \tag{2.36}$$

$$\frac{dI}{dt} = \rho C(t) + \gamma e^{-\mu\tau_1}E(t-\tau_1) - \delta e^{-\mu\tau_2}I(t-\tau_2) - (\mu + \phi)I(t). \tag{2.37}$$

2.2.2 Stochastic Models of Diseases

Smith et al. (1993) have conducted a study of pneumonia disease by using a stochastic model to estimate the acquisition and invasiveness of different *S.* pneumoniae serotypes by adapting the stochastic model to longitudinal pneumococcal carriage data. Children data from Papua New Guinea has been used in the study. The frequent invasion was associated with a high acquisition rate, high frequency and prolonged duration of carriage.

Apart from that, Melegaro et al. (2002) used a stochastic model of pneumococcal carriage, which consists of two groups, children and adults. In this study, they considered the susceptible and infected states only. Transition probabilities from susceptible to infected and vice versa were calculated. Results from the study shows that adults had lower pneumococcal carriage prevalence compared to children. Further results indicated that household acquired transmission made an important contribution to transmission of pneumococcal carriage within the family.

Stochastic model also has been used by other researchers in studying other type of diseases. Samat (2012) used stochastic model in studying the relative risk for dengue disease. She introduced an alternative approach for estimation of relative risk based on discrete time-space stochastic SIR-SI models (susceptible-infective-recovered for human populations; susceptible infective for vector populations) for the vector-borne infectious diseases transmission, specifically dengue disease. Findings from the study shows that discrete time-space stochastic SIR-SI models give better results compared to common approaches that are SMR and Poisson-gamma model.

Lawson (2013) proposed a discrete time-space stochastic susceptible-infective-removed (SIR) model for influenza. Based on Lawson (2013), the product of susceptible humans, $S_{i,j}$ and infective human, $I_{i,j}$ can give a large value, hence some equation seems unusable. Therefore, a constant of proportionality is placed before the product of the susceptible, $S_{i,j}$ and infectives, $I_{i,j}$ for human population. 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 individuals who become infected will 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. In this study, Lawson 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.

Liang, Greenhalgh, and Mao (2016) introduced stochasticity into the deterministic differential equation model for the spread of HIV amongst people who inject drugs (PWIDs) studied by Greenhalgh and Hay (1997). They investigated how environmental stochasticity affects the dynamical behaviour of the modified Kaplan model. Results of the study show that the introduction of stochastic noise into a model for the spread of HIV amongst PWIDs can cause the disease to die out in scenarios where deterministic models predict disease persistence.

Ming, Liu, Cheung and Wan (2016) used stochastic model to study the role of the heterogeneity in the analysis of the spread dynamics of infectious diseases in heterogeneous populations from temporal-spatial surveillance data especially for influenza A virus (H1N1) disease. The proposed model was evaluated using both simulated data and the real data from the year 2009 H1N1 epidemic in Hong Kong. They also used the data to test the prediction accuracy of their stochastic model. In the simulation experiment, their model achieved high accuracy in parameter estimation, with less than 10.0% mean absolute percentage error. In terms of the forward prediction of case incidence, the mean absolute percentage errors is 17.3% for the simulation experiment and 20.0% for the experiment on the real surveillance data.

Wang, Jiang, Hayat and Ahmad (2017) formulated stochastic model of HIV infection with T-cell proliferation and CTL immune response. The stochastic model is used to study the effect of environmental fluctuations on the HIV viral dynamics. Result from the study shows that the model solution is positive and global. The study also analyzed the extinction of the model. They also investigated the effect of white noises on model behavior.

Awang (2017) formulated stochastic model in studying the risk estimation for leptospirosis disease. She proposed age-structured compartment model which consider children and adult human. Results from the study show that the high-low risk area of leptospirosis is quite similar between the group of children and adult population. Based on the risk maps using the stochastic model, children are recognized to have very high-risk of being contracted with leptospirosis in many states compared to adults.

Vyambwera and Witbooi (2018) proposed a stochastic compartmental model for the tuberculosis (TB) population dynamics. Their study focuses on the analysis of TB in prisons, which have been identified as facilities with a significantly higher TB prevalence than the general population. Their main theorem is that the stochastic perturbation improves the stability of the underlying deterministic model's disease-free equilibrium. Results from the study shows that the inflow of infectious cases has an impact on the number of TB infected cases in the prison system. By screening the inflow on admission and providing a separate accommodation, TB infection in a prison system can be greatly reduced.

Ideris, Malim and Shaadan (2019; 2021) also used stochastic model to study the relative risk for leptospirosis disease. Their study proposed a discrete-time discrete-space stochastic model for relative risk estimation of leptospirosis in Malaysia based on SIR-SI (Susceptible-Infected-Recover for human; Susceptible-Infected for rat) transmission model. Result shows that Terengganu and Kelantan are the two most vulnerable states of leptospirosis for every epidemiology year from 2012 to 2016.

Based on the information above, it can be concluded that there are only two studies on pneumonia disease that used stochastic model: Smith et al. (1993) and Melegaro et al. (2002). This has led my study to formulate stochastic models for pneumonia disease.

2.3 Disease Mapping

This section discusses the history of disease mapping, analysis of disease mapping using three most common approaches to produce disease maps and relative risk estimation.

2.3.1 History of Disease Mapping

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 disease mapping refers to the geographical distribution shown in the visual form. Disease mapping has various names, some of which are: environmental epidemiology, small health studies and spatial epidemiology (Lawson, 2013). Clement (2014) defined 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. According to Samat 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, for example, studies by Samat (2012) and Diah (2017).

Among the first studies of disease mapping in epidemiology was conducted by John Snow in 1854 (Lawson & Williams, 2001). It was the map of the cholera victims that address links to the water supply locations in London. In this case, every street address of the cholera victim had been recorded. Snow (1854) was the first person to clearly show that cholera could spread through the contaminated water supply, where he used dot maps to show the residences of the victims. In Figure 2.9, his ‘dot maps’

of cholera patients' residence in the Golden Square, London showed a different group of cases around the water pump in Broad Street.

Figure 2.9. 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 helped in preventing the cholera epidemics in the United Kingdom since it identified those pipes which were contaminated by faecal material from cases of cholera. This type of map was also used by other researchers such as Deneke (1895) in studying the cholera cases in the city of Hamburg and the adjoining suburb of Altona. Apart from that, Holden (1880, as cited in Lawson & Williams, 2001) used 'dot maps' to demonstrate the value of disease mapping. His study recognized that the absence of sewage systems was more significant compared to the presence of unavoidable topographical features such as watercourses and elevation 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 between 1922-1950. Nowadays, this type of map is less preferred by researchers since the dot size influences the visual impression. The dots may not be seen if the sizes are too small; otherwise, it overlaps 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 the 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 gender, age and urbanization. Stock (1939, as cited in Lawson & Williams, 2001) also introduced maps using the Standardized Mortality Ratio (SMR) for cancer. These maps were drawn using ratio data for areas called choropleth maps. The construction method is quite simple, where each shaded area is in accordance with the data value. Areas with the highest values usually will be filled with the darkest tones. Figure 2.10 shows the example of the first usage of this type 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.

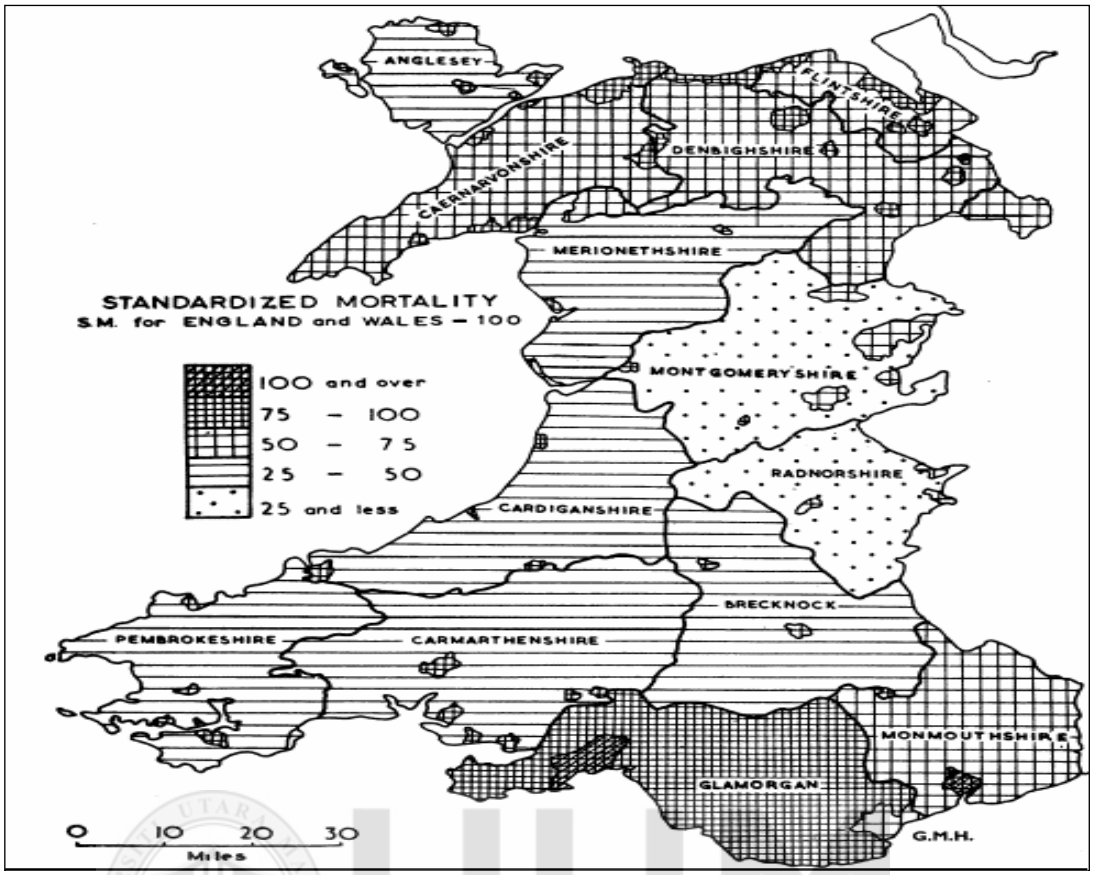

Figure 2.10. Cancer of lung and bronchus, Standardized Mortality for males, 1947 – 1953 (Source: Howe, 1959)

This study focuses on this type of map to determine the relative risk of pneumonia. The map appearance for choropleth map is affected by various factors such as the number of classes, tones and class intervals. Robinson and Sale (1969) and Korycka-Skorupa and Pasławsk (2017) stated that for the number of classes, it is best to have five or six shading categories. If the number of categories is less, the information may be lost, which causes the map to be over-generalized. However, if the number of categories is too large, it will not be effortless to interpret the data, as there is much information.

The tones for choropleth maps are utilized to represent the different values, which obviously give an impact on the appearance of the map (Robinson &

Sale, 1969). It is advisable to use bright, distinguishable tones spaced evenly from light to dark. The colour used should form a distinct sequence. Rather than using multiple colours, it is better to use different shades of one or two colours. Figure 2.11 shows the example of a map using shades of the same colour, which shows the low and high-risk areas based on SLIR method for tuberculosis. The class intervals, as mentioned earlier, also influence the appearance of the maps, which refers to the cut-off points. According to Robinson and Sale (1969), the choice of the class interval gives 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 leads to identify which area is of low or high fatality depending on the light shade or dark shade area.

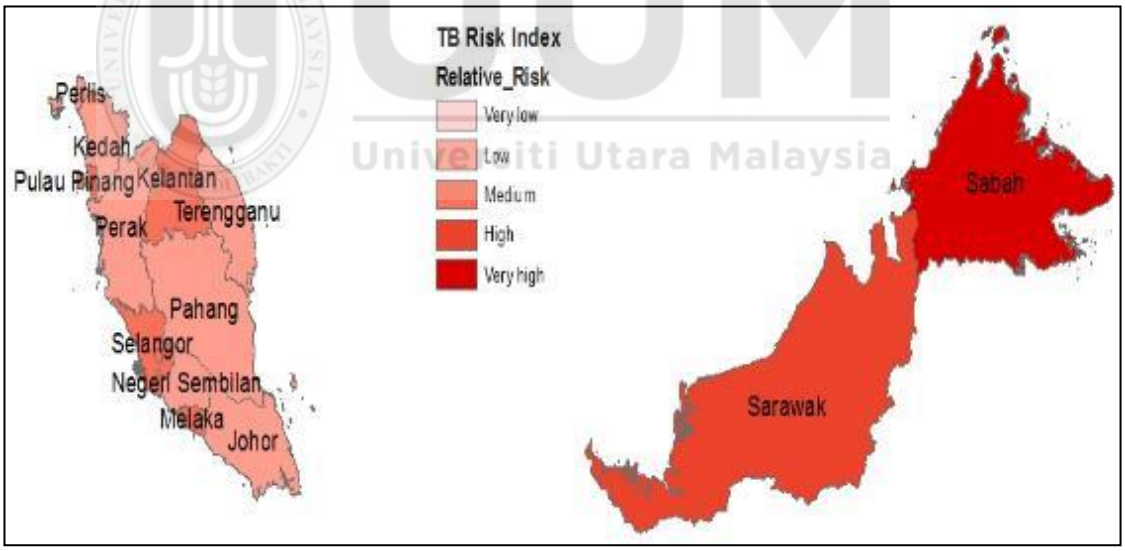

Figure 2.11. Disease map of estimated relative risks based on SLIR method (Source:Diah, 2017)

There are several ways of 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 a good impression of the underlying statistical distribution. Some may use quantiles to put an equal number of areas into each shading category.

According to Lawson (2013), there are two crucial characteristics of disease mapping: spatial or geographical distribution and disease distribution. In mapping the disease, the relative location of events is vital; hence geographical information system is needed. Another focus is disease distribution, where the main issue here is how to analyze disease incidence or prevalence when we have geographic information. This is sometimes called as geo-referenced disease data, by labelling the outcome with spatial tags. Disease mapping can indicate geographic areas with high and low incidence or mortality rates for specific diseases (Jin, Banerjee, & Carlin, 2007; Awang, 2017). Also, according to Koch (2005) and Waller and Carlin (2010), disease mapping for incidence and prevalence had been used in public health, the study of disease in human populations and epidemiology for a long time.

There are four main aims of using disease mapping (Shaddick, 2008). Maps are usually drawn for one of the two reasons either to analyze the data to determine which areas have a low or high incidence for the disease or illustration purpose so that main conclusions of analysis can be drawn for others' benefits. The analysis purpose in disease mapping is to estimate the true relative risk of a disease of interest throughout the geographical area of study and to reduce the noise in a disease map. Hence, in this study, disease mapping is the focus of attention.

To estimate the status of an area concerning 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 applied to case-event data. Case-event data are acknowledged as the street address of the disease cases recorded because it usually takes place within a fixed period, known as continuous data. In contrast, tract-count data correspond to observations, which consist of aggregated count cases within the tracts or small study areas over specified periods known as discrete data (Lawson, 2006). 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' population or background. Poisson-gamma model is another method 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. 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. BYM model also another method proposed by other researchers to find the relative risk estimation.

2.3.1.1 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 epidemiological terms, SMR is defined as Standardized Morbidity Ratio or Standardized Mortality Ratio. Here, morbidity refers to incidence while mortality refers to death. Note that Standardized Incidence Ratios (SIR) and Standardized Hospitalization Ratios (SHR) 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 while E_i is the expected number of cases.

In disease mapping, Samat (2012) assumed that the research area to be mapped is distributed into M mutually exclusive areas ($i=1,2,\ldots,M$). Every state has its number of observed cases, O_i and number of expected cases, E_i. Using O_i and E_i, obtained based on the existing data, may calculate one of the most common indices to estimate the relative risk $\hat{\theta}_i$ for state i. According to Lawson (2006) and Samat (2012), SMR model could be described as follows:

$$r_i = \hat{\theta}_i = \frac{O_i}{E_i}. \tag{2.33}$$

The observed number can be found from other resources such as the health indicator under the disease control department from the Ministry of Health Malaysia. On the other hand, the expected value can be counted by using a formula as in Equation (2.33) until (2.39) for the study region. For this research, based on Samat (2012), the expected number of cases, E_i is calculated as

$$E_i = N_i \frac{\sum O_j}{\sum N_j}, \tag{2.34}$$

where N_i is the population of region i and the summations (Σ) are for j=1, 2, . . ., M. Here, standardization of the total population at risk assumes that every person is equally at risk. Thus, relative risk can be estimated using the following formula:

$$\frac{O_i}{E_i} = \frac{O_i}{N_i \frac{\sum O_j}{\sum N_j}} \tag{2.35}$$

$$r_i = \frac{O_i}{N_i \frac{\sum O_j}{\sum N_j}} \tag{2.36}$$

$$= O_i \times \frac{\sum N_j}{N_i \sum O_j} \tag{2.37}$$

$$= \frac{O_i}{N_i} \times \frac{\sum N_j}{\sum O_j} \tag{2.38}$$

$$r_i = \hat{\theta}_i = \frac{\left(\frac{O_i}{N_i}\right)}{\left(\frac{\sum O_j}{\sum N_j}\right)} \tag{2.39}$$

where $\hat{\theta}_i$ defines the probability of a person within the state contracting the disease divided by the probability of a person in the population contracting the disease.

Despite the fact that SMR is often used to estimate the relative risks, it has several disadvantages. It is reliable to use SMR as a measure of relative risk for a wide area, such as a state or country. However, it is unreliable for small geographical areas such as 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 the ratio estimator. This has caused the use of SMR being criticized as it is challenging to interpret the data. In some cases where the areas of the expected number of cases are small, it can produce high values of relative risk and vice versa.

Moreover, SMR will be zero when there are no observed count data or cases (Awang & Samat, 2017). This causes the interpretation of SMR becomes difficult and needs to be conducted with caution. Poisson-gamma models were then introduced to solve these drawbacks which will discuss in the next section.

2.3.1.2 Poisson-gamma Model

Many researchers have investigated various methods for estimating the relative risk of disease due to the shortcomings of the SMR as a relative risk estimator. One of them includes the use of the initial applications of Bayesian method, namely the Poisson-gamma model (Lawson et al., 2003).

According to Iddrisu and Amoako (2016), in Poisson-gamma model, the unknown risk of disease in area i is given as θ_i where $i=1,2,\ldots,M$. 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 an area i can be written as $e_i=rN_i$, where the overall disease risk in the study population is given as follow:

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} N_i} \quad . \tag{2.40}$$

According to Samat and Ma'arof (2013), this model assumes that the number of new infectives, y_{ij} follow a Poisson distribution within a given period, with mean and variance $e_{ij}\theta_{ij}$, which can be write as $y_{ij} \sim$ Poisson $(e_{ij}\theta_{ij})$. Here, $i=1,2,\ldots,M$ denotes the research areas while $j=1,2,\ldots,T$ refers to periods of time, e_{ij} is the expected number of new infectives while θ_{ij} is the relative risk given by:

$$y_{ij}|e_{ij}, \theta_{ij} \sim Poisson\ (e_{ij}\theta_{ij}) \ . \tag{2.41}$$

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

$$\ell(\theta_{ij}) = \prod_{i=1}^{M} \prod_{j=1}^{T} \frac{\exp(-e_{ij}\theta_{ij})(e_{ij}\theta_{ij})^{y_{ij}}}{y_{ij}!} = P(y, e|\, \theta) \tag{2.42}$$

and

$$\ln\ell(\theta_{ij}) = -\sum_{i=1}^{M} \sum_{j=1}^{T} e_{ij}\theta_{ij} + \sum_{i=1}^{M} \sum_{j=1}^{T} y_{ij} \ln(e_{ij}\, \theta_{ij}) - \prod_{i=1}^{M} \prod_{j=1}^{T} y_{ij}. \tag{2.43}$$

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

$$\hat{\theta}_{ij} = \frac{y_{ij}}{e_{ij}}, \tag{2.44}$$

where $\hat{\theta}_{ij}$ represents the standardized mortality ratio in area *i* at time *j*. It was 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, given by:

$$\theta \sim Gamma(a, b). \tag{2.45}$$

The likelihood function for y_{ij} and prior distribution for θ_{ij} are respectively presented by

$$\ell(\theta) = \prod_{i=1}^{M} \prod_{j=1}^{T} \frac{(e_{ij}\theta)^{y_{ij}} \exp(-e_{ij}\theta)}{y_{ij}!} = P(y, e\,|\theta) \tag{2.46}$$

and

$$P(\theta \,\big|\, a, b) = \frac{(\theta)^{a-1}}{\Gamma(a)b^{a}} \exp(-\theta b), \theta, a, b > 0. \tag{2.47}$$

Based on this Poisson-gamma model, expected posterior relative risk is included in the analysis, where this risk is for all areas and periods. Bayesian hierarchical method is used for the estimation of parameters. Here, a and b are given prior distributions such that $a|\omega \sim P(a|\omega)$ and $b|\varphi \sim P(b|\varphi)$, where $P(a|\omega)$ and $P(b|\varphi)$ are the hyperprior distribution with hyper-parameters ω and $(\varphi_a, \varphi_b) \in \varphi$ for a and b, respectively. Parameters can be obtained by using this Bayesian hierarchical method. This is a second stage hierarchical modeling by using the Poisson-gamma model. In this study, $P(a|\omega) = \omega \exp(-\omega a)$ and $P(b|\varphi_a, \varphi_b)$ are defined as exponential distribution and gamma distribution, respectively. Hence, the posterior distribution is as follows:

$$P(\theta, a, b|y, e) \propto P(y, e|\theta, a, b)P(\theta)P(a|\omega)P(b|\varphi) \ . \quad (2.48)$$

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

Lawson et al. (2003) demonstrated the use of the Poisson-gamma model in their study, which gives a smoother map with less extreme values for estimating the relative risk than using the SMR. However, in this model, the adjustment of covariates is complicated, and there is no chance in dealing with spatial correlation between risks in neighboring locations. Thus, this encourages other researchers to come up with new approaches in estimating the relative risk.

2.3.1.3 Besag, York and Mollie Model

Besag, York and Mollie (BYM) model was introduced by Clayton and Kaldor (1987) and that was application to empirical Bayesian inference for relative

risks, which was later developed by Besag et al. (1991). Area-specific random effects are decomposed in this model for relative risks into a component that consider the effects that vary in a structured manner in space (clustering or correlated heterogeneity) and a component that consider the effects that vary in an unstructured manner across areas (uncorrelated heterogeneity) not the log-linear model. The incorporation of these random effects allows for the smoothing of relative risk at overall level (global and local).

For pneumonia disease, the observed number of pneumonia cases (y_i) in area *i* is assumed to follow Poisson distribution with mean $e_i\theta_i$, can be written as follows:

$$y_i \sim Poisson\ (e_i\theta_i), \tag{2.49}$$

where e_i stands for the expected number of cases in area *i* and θ_i stands for the "true" but unknown relative risk in area *i*. The variability of the log relative risk, log θ_i, is separated into three components at the next level of the model:

$$log\ \theta_i = \alpha + u_i + v_i, \tag{2.50}$$

where α represent an overall level of the relative risk, u_i is the spatial random effect which is the correlated heterogeneity and v_i is random effect reflecting the uncorrelated heterogeneity. However, since in this research, space-time data is used, hence this BYM model considered the space-time distribution. Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi and Songini (1995) proposed a model in which both specific intercept and temporal trend are modelled as random effects. This formulation allows for different spatial locations and even have spatial structure. However, all

temporal trends are assumed to be linear, which is a restrictive assumption. The model for the relative risk is of the form as below:

$$log\ \theta_{ij} = \alpha + u_i + v_i + \beta.t_j + \delta_i.t_j, \tag{2.51}$$

where $\beta.t_j$ is a linear trend term in time t_j and δ_i is an interaction random effect between space and time.

Prior distributions for random effects must be defined in Bayesian modeling. The distribution model for the uncorrelated heterogeneity, v_i is

$$v_i \sim N(0, \tau_v^2), \tag{2.52}$$

where this distribution does not rely on geographic location and is assumed to follow a normal distribution with zero mean and a common variance (precision parameter), τ_v^2.

A spatial correlation structure is used for the clustering component, where the risk estimation in any area relies on neighbouring areas. The conditional autoregressive (CAR) model proposed by Besag et al. (1991) is used where a parameter value in one area is influenced by its average neighbourhood value, with additional variability quantified by conditional variances depending on the number of neighbours.

$$[u_i | u_j, \mathrm{i} \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2), \tag{2.53}$$

where the mean of the areas bordering area *i*,

$$\bar{u}_i = \frac{1}{\sum_j \omega_{ij}} \sum_j u_j \omega_{ij}, \tag{2.54}$$

$$\tau_i^2 = \frac{\tau_u^2}{\sum_j \omega_{ij}}, \tag{2.55}$$

$\omega_{ij} = 1$ if i, j are adjacent (or 0 if they are not).

Here, ω_{ij} refer to the relationship between the area *i* and *j*. The prior mean of each u_i is defined as a weighted average of the other $u_j, \neq j$. Parameters τ_v^2 and τ_u^2 are used to control the amount of variability of random effects *v* and *u*. In a full Bayesian analysis, prior distributions must be specified for those parameters τ_v^2 and τ_u^2. These hyperprior distributions represent inverse-variance of random effects. Assume that these hyperprior distributions follow the gamma distribution. It is recommended to use a distribution with a large variance as there is no prior estimation for the random effects (Samat & Mey, 2017).

2.3.2 Relative Risk Estimation

Samat and Percy (2012) stated that the main issue when studying geographical distributions of disease incidence is the estimation of relative risk. The disease mapping construction depends on modeling to predict and estimate the risks. Therefore, better predictions and risk estimations would produce more precise disease risk maps. Relative risk is the risk of an event (or of developing a disease) relative to exposure. It 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.

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

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

The risk estimation of the disease may not be accurate when marking the area as low or high-risk area based on the number of cases occurred without considering other aspects like the size of region and the size of population as mentioned earlier in Section 1.2.2. This is proven in the study by Diah (2017). From the study, if it is based on the count data for tuberculosis disease occurrence in Malaysia, it shows that Johor had the highest number of cases compared to Pulau Pinang. However, when using stochastic transmission model, which considers other factors such as spatial correlation between risk in neighbouring area and number of populations, Pulau Pinang shows high risk of tuberculosis disease occurrence compared to Johor.

2.4 Summary

From previous studies, no studies have been conducted on the relative risk for pneumonia and most used a deterministic model to understand more about the spread of pneumonia. There are only two studies used stochastic model to understand the disease transmission for pneumonia. Hence, in my study, stochastic models for pneumonia disease are formulated and used. These alternative models proposed are used to predicting and estimating the low and high-risk areas based on the pneumonia transmission model, which includes an element of stochastic to monitor and control pneumonia disease.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

This chapter highlight the development of stochastic models proposed. There are four phases in conducting this study. The first phase focuses on the deterministic models, followed by the development of stochastic models. The third phase deals with the estimation of the relative risk and comparison of the results. Lastly, disease mapping for pneumonia disease is constructed. There are five steps in numerical modeling to develop the stochastic models. First is to capture the actual scenario quantitatively. Here, four deterministic models that are reliable with the data available are chosen. Next, from the chosen deterministic models, the stochastic element will be inserted to the equation. The stochastic element is insert in infectious class. Since these equations are difficult to solve, discretization and numerical methods adapted to approximate the governing equations. Later, a numerical analysis method is applied to calculate the solution for all chosen models, which Lawson's approach is adopted to solve the system of non-linear differential equations. Finally, algorithms in the computer will calculate the approximate solutions and it will end with interpretation from those solutions. Figure 3.1 shows the flow chart of the process in this study.

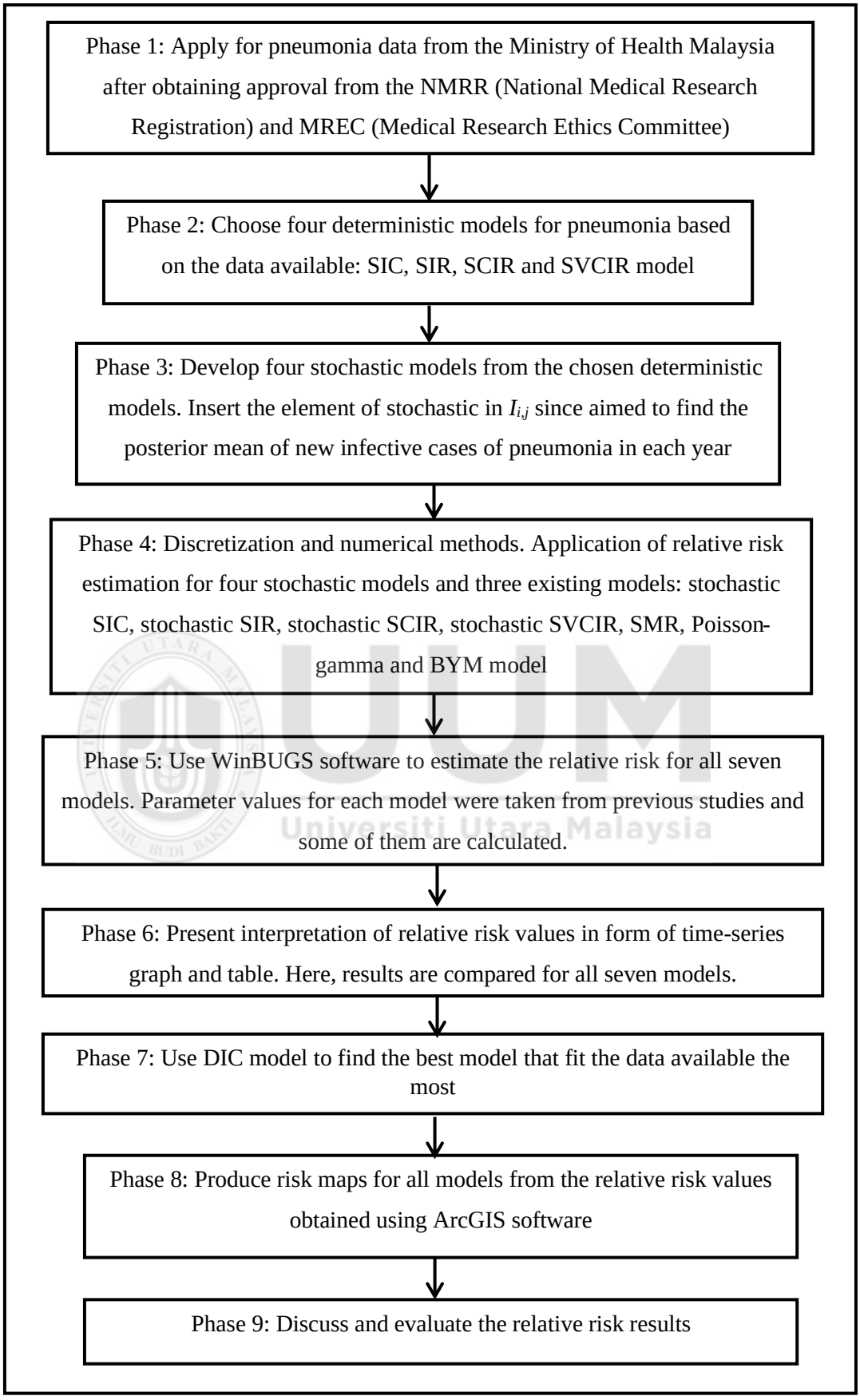

Figure 3.1. Flow chart of the research process

This chapter also provides a detail on the data sets utilized and then presents in depth the development of a stochastic model for direct disease transmission with particular application to pneumonia disease. The discussion initially considers four deterministic models, namely Susceptible–Infected–Carriers (SIC), Susceptible–Infectious–Recovered (SIR), Susceptible–Exposed–Infectious-Recovered (SEIR) or in this study, the exposed term is changed to the carriers' term as both have same meaning and so it will become Susceptible–Carriers–Infectious–Recovered (SCIR) in this study. The Susceptible–Vaccinated–Carriers–Infected–Recovered (SVCIR) is also discussed, specifically for pneumonia disease, followed by developing stochastic models SIC, SIR, SCIR and SVCIR for disease transmission. In this study, stochastic SIR model is adapted from stochastic SIR model proposed by Lawson (2006), and the other three models are the new models which difference from Lawson. These models are used later to estimate the relative risks for mapping pneumonia disease.

3.2 Deterministic SIC Model

The compartment model shown in Figure 3.2 is the basic model used in pneumonia transmission's study, which is a simple model proposed by Doura et al. (2000) (Subsection 2.2.1.1), known as the SIC model. 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

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

$C_{i,j}$ = total number of carrier persons for area i, at time j

The total host population for region *i* is $N_i = S_{i,j} + I_{i,j} + C_{i,j}$. Here, N_i is assumed to be a constant. Figure 3.2 shows the transmission process and the essential features of this model. The model derived has the same form as the deterministic SIC model used by Doura et al. (2000).

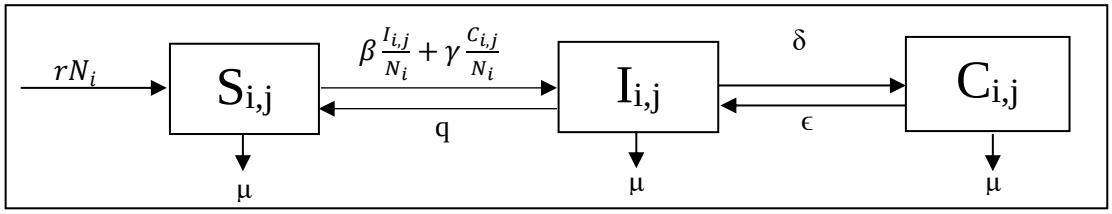

Figure 3.2. Flow diagram of SIC model for pneumonia disease.

All the parameters *r, μ, β, γ, q, δ,* and ϵ are per-capita rates. The flow rate of population is defined as the per-capita rate (*μ, β, γ, q, δ,* or ϵ) multiplied by the number of people subjected to that per-capita rate (N_i, $S_{i,j}$, $I_{i,j}$ or $C_{i,j}$). For example, the population for the birth, the in-flow "birth" to a compartment of "susceptible" is N_i, the per capita birth rate *r* times the total population N_i 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 births $= rN_i$,

Number of infectious from susceptible = $\beta \frac{S_{i,j}I_{i,j}}{N_i} + \gamma \frac{S_{i,j}C_{i,j}}{N_i}$,

Number of carriers = $\delta C_{i,j}$,

Number of infectious from carrier class = $\epsilon C_{i,j}$,

Number of infectious people recovers to the susceptible class = $qC_{i,j}$,

Number of deaths of $S_{i,j}$, = $\mu S_{i,j}$,

Number of deaths of $I_{i,j}$= $\mu I_{i,j}$,

Number of deaths of $C_{i,j}$= $\mu C_{i,j}$.

The compartmental model in Figure 3.2 can be shown mathematically in the form of differential equations system. The difference between the deterministic model used by Lawson and this SIC model used in this study is that for the pneumonia disease, model proposed consists of the carrier state instead of the recovered state.

Based on the differential equations in Subsection 2.2.1.1, this is non-linear due to the existence of the products of functions. According to Samat (2012), it is challenging to be solved analytically. On the other hand, analytic solution of linear ordinary differential equations system is generally elementary. Performing numerical analysis using computer programming can generally solve these non-linear equations. Hence, in this study, computer programming WinBUGS is used to carry out the numerical analysis to find out the solutions for the non-linear ordinary differential equations in (2.1) to (2.3) which are complicated.

In this study, the numerical analysis method is used to calculate the solution for SIC model, while Lawson's approach is adopted to solve the system of non-linear differential equations. The deterministic model for pneumonia disease transmission based on Figure 3.2 for i = 1, 2, …, M study area and j = 1, 2, …, T periods, reduces to this difference equation set:

$$S_{i,j} = rN_i + qI_{i,j-1} + (1-(\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i}) - \mu)S_{i,j-1} \quad (3.1)$$

$$I_{i,j} = (\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i})S_{i,j-1} + (1 - q - \mu - \delta)I_{i,j-1} + \epsilon C_{i,j-1} \quad (3.2)$$

$$C_{i,j} = \delta I_{i,j-1} + (1 - \epsilon - \mu)C_{i,j-1} \quad (3.3)$$

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.

3.2.1 Stochastic SIC Model

In the following analysis, the deterministic model is used to give an approximation to the stochastic means. According to Molzon (2009) and Samat and Percy (2012), deterministic models are good approximations to the stochastic models. A new notation is introduced, $\bar{I}_{i,j}$ which represents the numbers of newly infected in the study region i and period (j -1, j]. Since the pneumonia data observed are yearly, this study is aimed to find the posterior mean of new infective cases of pneumonia in each year.

For i = 1, 2, …, M study regions and j = 1, 2, …, T period, the stochastic models for pneumonia transmission is adapted from the deterministic SIC equations (3.1) – (3.3).

$$\bar{I}_{i,j} = (\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i})S_{i,j-1} \quad (3.4)$$

$$\bar{I}_{i,j} \sim Poisson(\lambda_{i,j}) \quad (3.5)$$

$$\lambda_{i,j} = \exp(\beta_0 + b_i)(\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i})S_{i,j-1} \quad (3.6)$$

$$\log(\lambda_{i,j}) = \beta_0 + b_i + \log(\frac{\beta I_{i,j-1}}{N_i} + \frac{\gamma C_{i,j-1}}{N_i}) + \log(S_{i,j-1})$$

$$S_{i,j} = \mu N_i + qI_{i,j-1} + (1 - \mu)S_{i,j-1} - \bar{I}_{i,j} \quad (3.7)$$

$$I_{i,j} = \bar{I}_{i,j} + (1 - q - \mu - \delta)I_{i,j-1} + \epsilon C_{i,j-1} \tag{3.8}$$

$$C_{i,j} = \delta I_{i,j-1} + (1 - \epsilon - \mu)C_{i,j-1} \tag{3.9}$$

From Equation (3.6) above, β_0 is a constant term that stands for the overall rate of the process across the study region, while b_i refers to the random effect that absorbs residual spatial variation, which allows dependencies among adjacent regions; this is the main stochastic element. In any analysis that involves the geographical region, spatial random variation is a crucial element. The Poisson distribution is used to represent the new number of infectives. In order to match the deterministic part of the Equation (3.2) with a positive multiplicative factor to show the spatial correlation, mean, $\lambda_{i,j}$ is chosen. The new number of infectives ($\bar{I}_{i,j}$) is assumed to follow a 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 SIC model proposed.

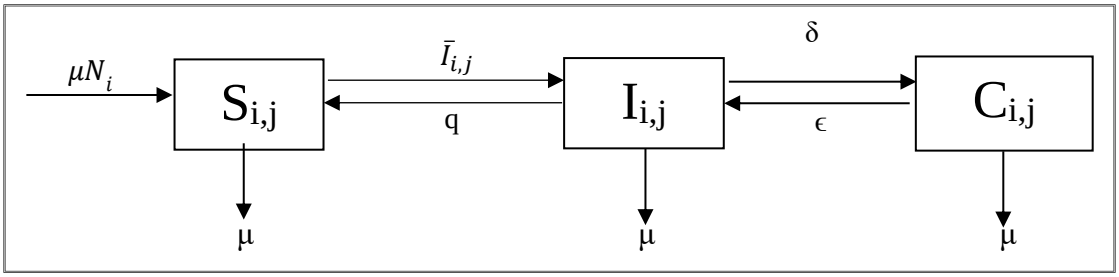

Figure 3.3. Flow diagram of the stochastic SIC model for pneumonia transmission

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*

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

$C_{i,j}$: total number of carrier persons for area i, at time j

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

N_i : the population size for the study state

δ : rate at which infectious move into carrier class

q : rate at which an infectious individual recovered moves to the susceptible class

ϵ : rate at which a carrier moves into the infectious class

β_0 : the overall rate of the process

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

This study is interested in determining the number of new infective persons ($\bar{I}_{i,j}$) which is the only stochastic element in this model, while the other terms are non-stochastic elements. In future research, these models can be extended by considering other stochastic terms such as the new number of carriers ($\bar{C}_{i,j}$). It is assumed that the number of new infectives ($\bar{I}_{i,j}$) follows the Poisson distribution where the expected number of new infective persons take in several elements of the transmission. This new infective case is measured by $\lambda_{i,j}$, which represents the rate of transmission. Poisson distribution is used as a starting point for the analysis of count data (Lawson, 2006; Samat, 2012; Awang, 2017). The number of new infective cases refers to the number of cases. In this study, it is assumed that the birth and death rates are equal as there is no change in the amount of people in a given time (zero population growth (ZPG)). Hence, the symbol of the birth rate, r is now changed to μ.

Based on Figure 3.3, the main aspect of the transmission equation is a linear predictor term that considers a random effect or any covariates and a simple direct dependence of current infective count on the same previous spatial unit, such that

$(\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i}) S_{i,j-1}$. Therefore, a number of new infective persons is given by $\bar{I}_{i,j} = (\beta \frac{I_{i,j-1}}{N_i} + \gamma \frac{C_{i,j-1}}{N_i}) S_{i,j-1}$. The equation for $S_{i,j}$, $I_{i,j}$, and $C_{i,j}$ are non-stochastic or fixed. Poisson distribution cannot be tested in isolation as these counts are conditional upon other variables. The constant term β_0 is also included here to express the overall rate of the process while the 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.

This study uses the intrinsic conditional autoregressive (CAR) prior distribution as a family of prior distributions for the random effect b_i. According to Lawson et al. (2003) and Samat and Percy (2012), the most common way; dealing with the spatial correlation between neighbouring regions is by using CAR model, where the probability density values at any particular place are conditional on the neighbouring regions. The model was introduced by Besag, York and Mollie (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) by temporarily dropping the SIC model for convenience,

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

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

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

Here, $w_{i,s}$ is an appropriate weight while $\sigma_c{}^2$ 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, the gamma distribution is considered as suggested by Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi and Songini (1995); Samat and Mey (2017). A range of CAR models is supported by GeoBUGS extension in the WinBUGS package.

The stochastic SIC transmission model for pneumonia disease suggested here is used in estimating the relative risk value. These procedures are repeated and applied to the other three models: SIR, SCIR and SVCIR.

3.3 Deterministic SIR Model

Figure 3.3 shows the compartment model for SIR model, which is the basic model used in influenza study. This model was proposed by Lawson (2013) and used by Soliman and Bueno (2018) in studying pneumonia disease (Subsection 2.2.1.7). To make it relevant and clearer, a different notation is used from the one used by Lawson, Soliman and Bueno. However, in this study, natural death rate and death rate due to pneumonia disease are considered in this model. For birth rate and natural death rate of humans per year in the model and are assumed to be equal. In this model, death

caused by pneumonia, μ_p, also be considered. Hence, for *i = 1, 2, ..., M* study areas and *j = 1, 2, ..., T* periods, the equations are defined as follows:

$$S_{i,j} = \mu N_i + (1 - \alpha - \mu) S_{i,j-1}, \tag{3.10}$$

$$I_{i,j} = \alpha I_{i,j-1} S_{i,j-1} + (1 - g - \mu - \mu_p) I_{i,j-1}, \tag{3.11}$$

$$R_{i,j} = g I_{i,j-1} + (1 - \mu) R_{i,j-1}, \tag{3.12}$$

where definitions of the compartments are given as follows:

$S_{i,j}$: total number of susceptible 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*

g : the hazard of an infectious person's being removed (recovery rate)

α : the risk of a susceptible person's becoming infective in time period *j*, where α is constant.

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

μ_p : death rate due to pneumonia per year

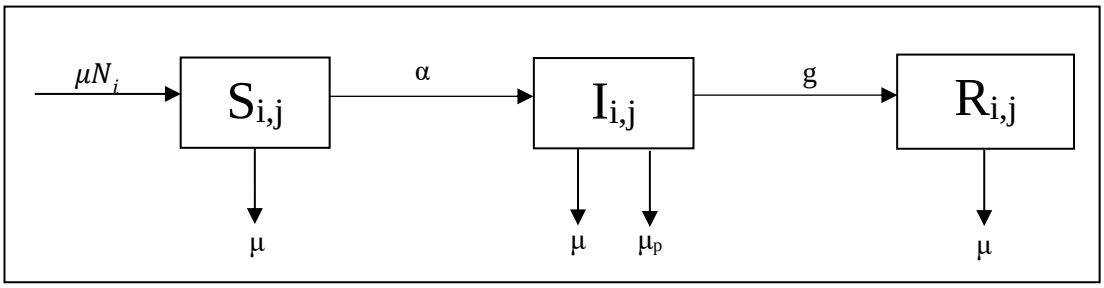

Figure 3.3. Flow diagram of the deterministic SIR model

The total host population for region i is given by $N_i = S_i + I_i + R_{i,j}$. Here, N_i is assumed to be a constant. The deterministic SIR model above 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.

3.3.1 Stochastic SIR Model

Figure 3.4 shows the flow of the proposed stochastic SIR model. New notations are introduced, $\bar{I}_{i,j}$ and $\Re_{i,j}$ which represent the numbers of newly infected and newly recovered respectively, in the study region *i* and period (*j* -1, *j*]. Since the pneumonia data are observed yearly, this study is aimed to determine the posterior mean of new infective pneumonia cases in each year.

The discrete time stochastic SIR model, for $i = 1, 2, \ldots, M$ study regions and $j = 1, 2, \ldots, T$ periods are given as follows:

$$\bar{I}_{i,j} \sim Poisson(\lambda_{i,j}) \tag{3.13}$$

$$\lambda_{i,j} = \exp(\beta_0 + b_i).\alpha.S_{i,j} \tag{3.14}$$

$$\log(\gamma_{i,j}) = \beta_0 + b_i + \log\alpha + \log(S_{i,j})$$

$$S_{i,j} = \mu N_i + (1-\mu)S_{i,j-1} - \bar{I}_{i,j} \tag{3.15}$$

$$I_{i,j} = \bar{I}_{i,j} + (1-\mu-\mu_p)I_{i,j-1} - \Re_{i,j} \tag{3.16}$$

$$R_{i,j} = \Re_{i,j} + (1-\mu)R_{i,j-1} \tag{3.17}$$

$$\Re_{i,j} = gI_{i,j-1} \tag{3.18}$$

Based on Figure 3.4, a linear predictor term that can take in a random effect or covariates that is $\alpha S_{i,j}$. Therefore, the number of new infective people is $\bar{I}_{i,j} = \alpha S_{i,j}$.

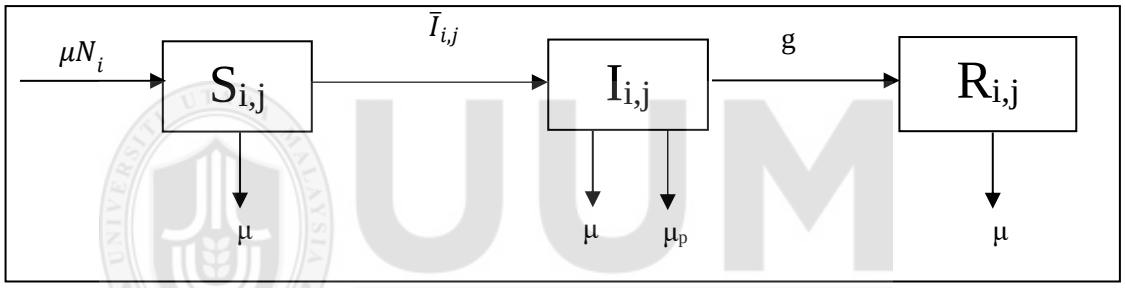

Figure 3.4. Flow diagram of the stochastic SIR model

3.4 Deterministic SCIR Model

In this study, the SEIR model without vaccination proposed by Kassa and Murthy (2016) will be used. However, the term "exposed" in the model will be changed to the term "carriers" in this study. Hence, the model will be called the Susceptible–Carriers–Infectious–Recovered (SCIR) model. In order to make it more relevant and clearer, different notations are used from the notations proposed by Kassa and Murthy (2016). In this study, for i = 1, 2, …, M study areas and j = 1, 2, …, T time periods, the definitions of the compartments and notations are given as follows:

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

$C_{i,j}$: total number of carriers 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*

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

γ : the transmission rate of infection events of susceptible

N_i : the population size for the study state

ϵ : rate at which carriers become infected

g : rate of recovery

The transmission process and the important features of this model are described in Figure 3.5.

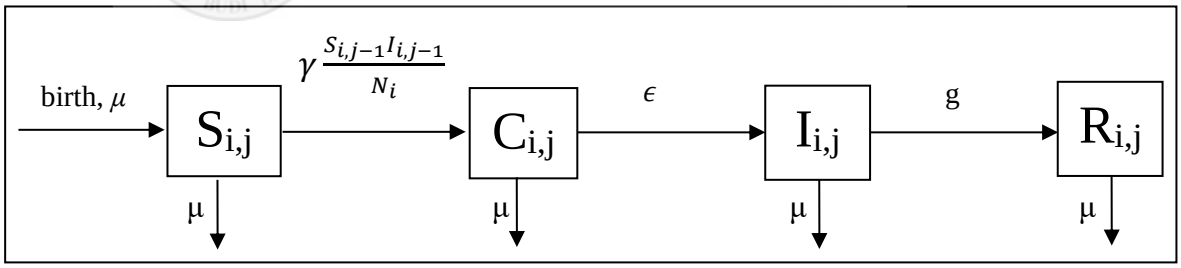

Figure 3.5. Flow diagram of deterministic SCIR model

The deterministic model for pneumonia disease transmission based on Figure 3.5 for $i = 1, 2, \ldots, M$ study areas and $j = 1, 2, \ldots, T$ periods, for pneumonia disease transmission from subsection 2.2.1.5 reduces to this set of difference equation:

$$S_{i,j} = \mu N_i + (1 - \mu - \left(\gamma \frac{I_{i,j-1}}{N_i}\right))S_{i,j-1} \tag{3.19}$$

$$C_{i,j} = \left(\gamma \frac{S_{i,j-1} I_{i,j-1}}{N_i}\right) + (1 - \mu - \epsilon) C_{i,j-1} \quad (3.20)$$

$$I_{i,j} = \epsilon C_{i,j-1} + (1 - \mu - g) I_{i,j-1} \quad (3.21)$$

$$R_{i,j} = g I_{i,j-1} + (1 - \mu) R_{i,j-1} \quad (3.22)$$

The total host population for region i is $N_i = S_i + C_i + I_i + R_i$. Here N_i is assumed to be constant. 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.

3.4.1 Stochastic SCIR Model

In this study, the stochastic SCIR model is developed based on the stochastic SIR model proposed by Lawson (2006). New notations $\bar{I}_{i,j}$ and $\Re_{i,j}$ are introduced, which represent the numbers of newly infected and newly recovered respectively, in the study region *i* and period (*j* -1, *j*]. For *i* = 1, 2, …, *M* study regions and *j* = 1, 2, …, *T* period, the stochastic models for pneumonia transmission is adapted from the deterministic SCIR equations (3.19) – (3.22) which gives

$$\bar{I}_{i,j} = \epsilon C_{i,j-1} \quad (3.23)$$

$$\bar{I}_{i,j} \sim Poisson\ (\lambda_{i,j}) \quad (3.24)$$

$$\lambda_{i,j} = \exp(\beta_0 + b_i) \epsilon C_{i,j-1} \quad (3.25)$$

$$\log(\lambda_{i,j}) = \beta_0 + b_i + \log \epsilon + \log C_{i,j-1}$$

$$S_{i,j} = \mu N_i + (1 - \mu - \left(\gamma \frac{I_{i,j-1}}{N_i}\right)) S_{i,j-1} \quad (3.26)$$

$$C_{i,j} = \left(\gamma \frac{S_{i,j-1} I_{i,j-1}}{N_i}\right) + (1-\mu) C_{i,j-1} - \bar{I}_{i,j} \quad (3.27)$$

$$I_{i,j} = \bar{I}_{i,j} + (1-\mu) I_{i,j-1} - \Re_{i,j} \quad (3.28)$$

$$R_{i,j} = \Re_{i,j} + (1-\mu) R_{i,j-1} \quad (3.29)$$

$$\Re_{i,j} = g I_{i,j-1} \quad (3.30)$$

Figure 3.6 shows the flow of the stochastic SCIR model that will be proposed.

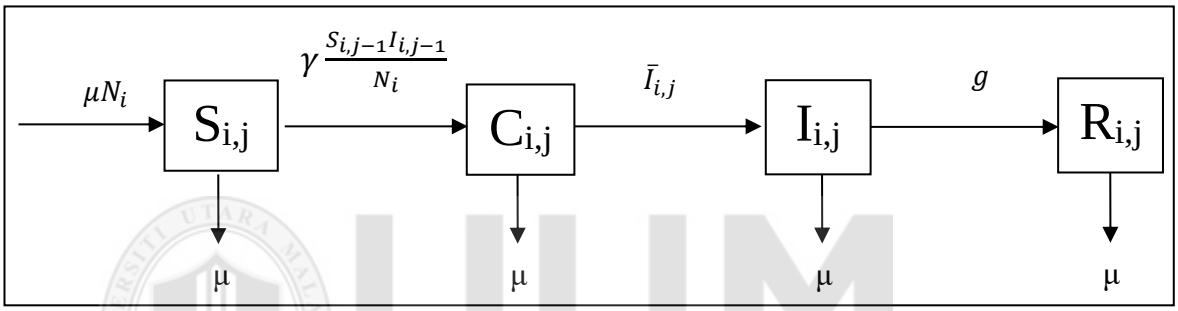

Figure 3.6. Flow diagram of stochastic SCIR model for pneumonia disease.

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*

$C_{i,j}$: total number of carrier 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)

γ : infection rate

N_i : the population size for the study state

g : rate of recovery

β_0 : the overall rate of the process

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

Based on Figure 3.6, a linear predictor term that can take in a random effect or covariates that is $\epsilon C_{i,j-1}$. Therefore, the number of new infective people is $\bar{I}_{i,j} = \epsilon C_{i,j-1}$. The equation for $S_{i,j}$, $C_{i,j}$, and $R_{i,j}$ are non-stochastic or fixed.

3.5 Deterministic SVCIR Model

The compartmental model shown in Figure 3.7 is adapted from Tilahun et al. (2017). However, in this study, different notation is used from the notation by Tilahun et al. (2017). In this study, $i = 1, 2, \ldots, M$ study areas and $j = 1, 2, \ldots, T$ periods, the definitions of the compartments and notations are given as follows:

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

$V_{i,j}$: total number of vaccinated persons for area *i*, at time *j*

$C_{i,j}$= total number of carrier 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*

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

μ_p : mortality rate due to pneumonia of humans per year

p : rate at which fraction of population has been vaccinated before the disease outbreaks

(1-p): fraction of susceptible population

φ : waning rate of those individuals who are vaccinated but did not respond to vaccination

δ : rate at which individuals loss their temporary immunity

ϑ : rate at which individuals from susceptible class move to vaccinated class (vaccination rate)

ε : proportion of the serotype not covered by the vaccine

ϵ : rate at which carriers become infected

ρ : probability of newly infected persons by the force of infection become carriers to join the carrier class

$(1-\rho)$: probability of newly infected persons to join infected class

g_n : recovery rate by gaining natural immunity

g : recovery rate by treatment, with treatment efficacy of q proportion of persons join the recovered class or join the carrier class with (1-q) proportion by adapting the treatment

γ : rate at which a carrier can infect a susceptible

N_i : the population size for the study state

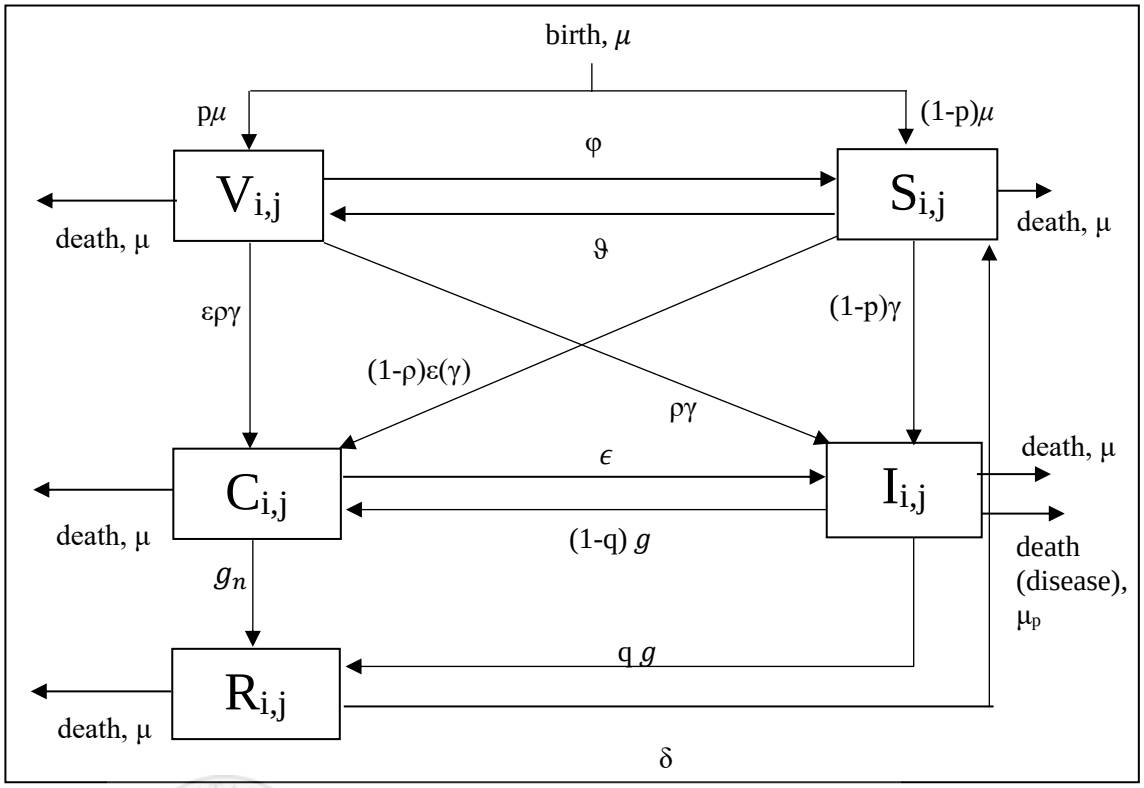

Figure 3.7. Flow diagram of deterministic SVICR model

The deterministic model for pneumonia disease transmission based on Figure 3.7 for $i = 1, 2, \ldots, M$ study areas and $j = 1, 2, \ldots, T$ periods, for pneumonia disease transmission from subsection 2.2.1.6 reduces to this set of difference equation:

$$S_{i,j} = (1-p)\mu N_i + \varphi V_{i,j-1} + \delta R_{i,j-1} + (1-\mu-\gamma-\vartheta)S_{i,j-1} \quad (3.31)$$

$$V_{i,j} = p\mu N_i + \vartheta S_{i,j-1} + (1-\mu-\varepsilon\gamma+\varphi)V_{i,j-1} \quad (3.32)$$

$$C_{i,j} = \rho\gamma S_{i,j-1} + \rho\varepsilon\gamma V_{i,j-1} + (1-q)gI + (1-\mu-g_n-\epsilon)C_{i,j-1} \quad (3.33)$$

$$I_{i,j} = (1-\rho)\gamma S_{i,j-1} + (1-\rho)\varepsilon\gamma V_{i,j-1} + \epsilon C_{i,j-1} + (1-\mu-\mu_p-g)I_{i,j-1} \quad (3.34)$$

$$R_{i,j} = g_n C_{i,j-1} + qgI_{i,j-1} + (1-\mu-\delta)R_{i,j-1} \quad (3.35)$$

The total host population for region i is $N_i = S_i + V_i + C_i + I_i + R_i$. Here N_i is assumed to be a constant. In the next section, from this formulation, stochastic mean populations will be provided.

3.5.1 Stochastic SVCIR Model

In this study, stochastic SVCIR model are developed. For i = 1, 2, …, M study regions and j = 1, 2, …, T period, the stochastic models for pneumonia transmission is adapted from the deterministic SVCIR equations (3.31) – (3.35). New notations $\bar{I}_{i,j}$ and $\Re_{i,j}$ are introduced, which represent the numbers of newly infected and newly recovered respectively, in the study region i and period (j -1, j].

$$\bar{I}_{i,j} = (1-p)\gamma S_{i,j-1} \tag{3.36}$$

$$\bar{I}_{i,j} \sim Poisson\ (\lambda_{i,j}) \tag{3.37}$$

$$\lambda_{i,j} = \exp(\beta_0 + b_i)(1-p)\gamma\, S_{i,j-1} \tag{3.38}$$

$$\log(\lambda_{i,j}) = \beta_0 + b_i + \log(1-p) + \log\gamma + \log S_{i,j-1}$$

$$S_{i,j} = (1-p)\mu N_i + \varphi V_{i,j-1} + \delta R_{i,j-1} + (1-\mu-\gamma-\vartheta)S_{i,j-1} \tag{3.39}$$

$$V_{i,j} = p\mu N_i + \vartheta S_{i,j-1} + (1-\mu-\varepsilon\gamma+\varphi)V_{i,j-1} \tag{3.40}$$

$$C_{i,j} = \rho\gamma S_{i,j-1} + \rho\varepsilon\gamma V_{i,j-1} + (1-q)gI_{i,j-1} + (1-\mu-g_n-\epsilon)C_{i,j-1} \tag{3.41}$$

$$I_{i,j} = \bar{I}_{i,j} + (1-\rho)\varepsilon\gamma V_{i,j-1} + \in C_{i,j-1} + \left(1-\mu-\mu_p\right)I_{i,j-1} - \Re_{i,j} \tag{3.42}$$

$$R_{i,j} = g_n C_{i,j-1} + q\Re_{i,j} + (1-\mu-\delta)R_{i,j-1} \tag{3.43}$$

$$\Re_{i,j} = \mathrm{g}I_{i,j-1} \tag{3.44}$$

A linear predictor term that can take in a random effect or covariates is $(1-p)\gamma S_{i,j-1}$. Therefore, the number of new infective people is $\bar{I}_{i,j} = (1-p)\gamma S_{i,j-1}$. In contrast, the equations for $S_{i,j}$, $V_{i,j}$, $C_{i,j}$, and $R_{i,j}$ are non-stochastic or fixed. Figure 3.8 shows the flow diagram of the stochastic SVCIR model.

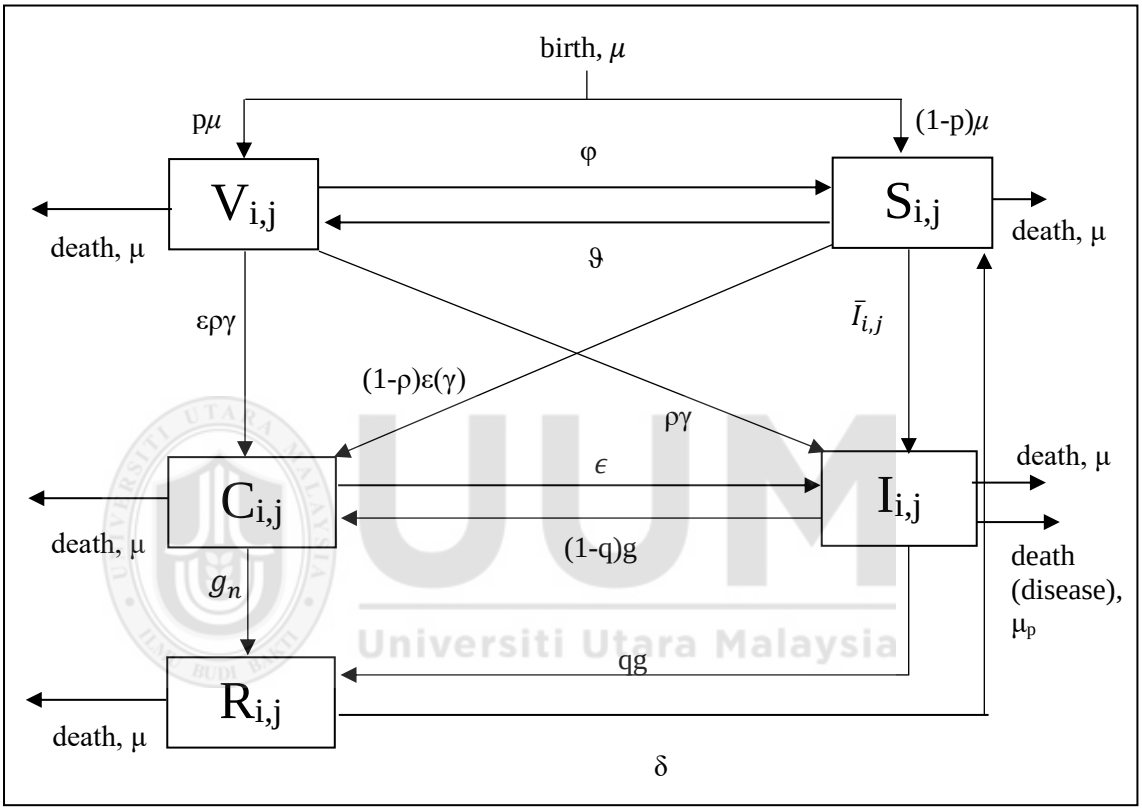

Figure 3.8. Flow diagram of stochastic SVCIR model for pneumonia disease

The stochastic SIC, SIR, SCIR and SVCIR models for transmission of pneumonia disease proposed in this study are used to estimate the relative risk.

3.6 Relative Risk Estimation for Disease Mapping

To estimate the relative risk, disease mapping studies mostly uses regression-type models. However, for this study, a different approach is applied in estimating the relative risk, which depends on the transmission model of the disease, built explicitly

for pneumonia. WinBUGS software is used to perform the computational analysis, which is a program intended to implement Bayesian inference on the statistical problem using Markov Chain Monte Carlo (MCMC) computations (Lawson et al., 2003; Ntzoufras, 2009). Further discussion and application for Bayesian analysis of disease mapping utilizing this software are available in Lawson et al. (2003) and Ntzoufras (2009).

In this study, the method used to estimate the relative risk is based on the method suggested by Samat and Percy (2012). In general, for $i = 1, 2, \ldots, M$ study regions and $j = 1, 2, \ldots, T$ periods, the posterior expected mean number of the infectives can be approximated using an unbiased sample mean given by

$$\tilde{\lambda}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \lambda_{ijk}, \tag{3.10}$$

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

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

$$\theta_{ij} = \frac{\lambda_{ij}}{e_{ij}}, \tag{3.11}$$

where e_{ij} stands for the expected number of new infective cases based on the population across the study areas. Hence, according to Samat and Percy (2012), by using an unbiased sample mean, the posterior expected relative risk can be estimated by

$$\tilde{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\theta_{ijk} = \frac{1}{n}\sum_{k=1}^{n}\frac{\lambda_{ijk}}{e_{ij}} = \frac{\tilde{\lambda}_{ij}}{\tilde{e}_{ij}} \quad . \tag{3.12}$$

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

$$\tilde{e}_{ij} = N_{ij}\frac{\sum_{i=1}^{m}\sum_{j=1}^{t}\bar{I}_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{t}\tilde{\lambda}_{ij}} \quad , \tag{3.13}$$

where N_{ij} is the population of area *i* at time *j*. In order to standardize the total number of populations at risk, it is assumed that everybody is equally at risk. Hence, based on Samat and Percy (2012), relative risk estimation formula can be written as follows

$$\tilde{\theta}_{ij} = \left(\frac{\tilde{\lambda}_{ij}}{N_{ij}}\right)\left(\frac{\sum_{i=1}^{M}\sum_{j=1}^{T}\tilde{\lambda}_{ij}}{\sum_{i=1}^{M}\sum_{j=1}^{T}\bar{I}_{ij}}\right). \tag{3.14}$$

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. The value for the relative risk in this study is interpreted as in Table 3.1.

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 implies that the people within the area are less likely to endure from this disease compared with the 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 pneumonia 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 Equation (3.14) is used in the next chapter to determine the relative risk estimation for the disease mapping via stochastic models proposed utilizing data from the former count cases for all tracts under consideration. Then, the performance of the relative risk estimation based on all the stochastic models with SMR, Poisson-gamma model and BYM model are measured. Based on these results of relative risk, pneumonia risk maps are constructed.

3.7 Real Data Set

Data used in this study is obtained from the Ministry of Health and the Department of Statistics in Malaysia. Firstly, approval is applied via the National Medical Research Registration (NMRR) website to gain the approval of Medical Research Ethics Committee (MREC) in requesting data of pneumonia in each state. The data consists of the number of populations, the number of pneumonia cases, the number of deaths, and the number of recovery. Each variable listed is required for each state in Malaysia which is 13 states including three federal territories that are Kuala Lumpur, Putrajaya, and Labuan. The duration of this study and data covers the year 2010 until year 2019. Data for the year 2020 is not available due to some hospitals still filling in the data for pneumonia cases in Malaysia. The recovery data are calculated based on previous study by Samat (2012) and Awang (2017) where

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

In application, the data available for analysis is the annual new infective pneumonia counts for states in Malaysia from the year 2010 until 2019.

3.7.1 Estimated Parameter Values for SIC, SIR, SCIR and SVCIR Models

In this study, the value for β that is the infection rate from susceptible class to infectious class and the rate at which a carrier can infect a susceptible (γ) are adopted from previous study by Kassa and Murthy (2016), with value of 0.5445 and 0.05 respectively. The rate at which recovered infectious moves into carrier class (δ) also adopt from Kassa and Murthy's study with value 1. The rate for recovered infectious moves into susceptible class (q) is estimated using formula from Kassa and Murthy's study which $q= 1\times 0.99999 = 0.99999$, where 1 is the recovery rate calculated using

formula as mentioned before and 0.99999 is assumed based on Mbabazi et al.'s study which refer to the probability of recovered infectious that join the susceptible class. Regarding the values for the progression rate from carrier period to infectious period ($\in$) and recovery rate by gaining natural or temporal immunity (γ_t), both are adopted from the previous study by Doura et al. (2000), Otieno et al. (2012) and Tilahun et al. (2017). The values for $\in$ and γ_t are 0.3059 and 0.0115 respectively. Force of infection for the vaccinated class (λ_v), is calculated using formula by Tilahun et al. (2017), $\lambda_v = 0.002 \times 0.5445$ where 0.002 is adopted from Ndelwa et al. (2015) and Tilahun et al. (2017) and 0.5445 is the value for β. For the recovery rate by gaining natural immunity (g_n), the value for waning rate of those individuals who are vaccinated but did not respond to vaccination (φ) and vaccination rate (ϑ) are adopt from Tilahun et al. (2017) with value of 0.0115, 0.0025 and 0.008, respectively. The probability value of newly infected become carriers, ρ and the value of rate at which fraction of population has been vaccinated before the disease outbreaks, p are 0.2 and 0.05 respectively, which adopt from Tilahun et al. (2017). N stands for total population size in Malaysia. While, in this study, N_i is the number of populations in each state. Moreover, the yearly rate values for μ, μ_P and g are 0.01334, 11.39, and 1 respectively. These rates are converted from daily rates to yearly rates which are derived from the Nishiura (2006) and Samat (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.

3.7.2 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 conversion 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	$(1-\mu_D)^4\mu_D$	$(1-\mu_D)^5$	$1-(1-\mu_D)^5$
⋮	⋮	⋮	⋮
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 (2019), the birth rates and death rates for human populations per day corresponding to the life expectancy of 74.5 years. Hence,

$$\mu_D = \frac{1}{365\ x\ 74.5} = 0.00003677$$

Converting this daily rate to a yearly rate produces

μ = 0.01334

The recovery rate, g corresponds to the minimum treatment period of seven days for the patient to recover. Therefore,

$g_D = \frac{1}{7}$ = 0.1429

Convert g_D to yearly rate gives

g = 1

For the deaths rate due to pneumonia disease, μ_P, 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, μ_p= 11.39.

This information is used to estimate the relative risk for pneumonia disease using stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model.

3.8 Summary

Four stochastic models for pneumonia proposed in this study have been explained in this chapter which accomplished the first objective of this study. In the next chapter, the outcomes of relative risk estimation based on these four stochastic models proposed and other existing methods (SMR, Poisson-gamma model and BYM model) are presented and explained.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the four proposed stochastic models that have been constructed in Chapter Three are used in estimating the relative risk values. The four proposed stochastic models are stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model. First, the diagnostic analysis such as CAR prior distribution is inserted in the model to find the relative risk. This CAR prior distribution is used to measure the random effect that is spatial distribution. Next, the diagnostic analysis involves by checking the posterior distribution through Markov Chain Monte Carlo (MCMC) method. This value also computed for three common methods that are used in disease mapping. They are SMR, Poisson-gamma model and BYM model using pneumonia data. Later, for measuring the performance of all these seven methods, the relative risk values are compared. Maps are then constructed based on the relative risk values results. The pneumonia data in Malaysia are used in this analysis from year 2010 until year 2019.

4.2 Application of Relative Risk Estimation for Pneumonia Disease Mapping

In this section, first, diagnostic analysis is analyzed through CAR prior distribution for spatial distribution, and then the diagnostic analysis is done by checking posterior distribution using MCMC method. MCMC method employs random sampling in a probabilistic space to estimate the posterior distribution of a parameter of interest. After all these diagnostics checking are clear, then the relative risk values obtained are used to produce maps. The results of four proposed models; stochastic SIC model,

stochastic SIR model, stochastic SCIR model and stochastic SVCIR model for pneumonia disease transmission were demonstrated and displayed. Then, the relative risk estimation outcomes are demonstrated and displayed, based on the three applications of existing relative risk estimation methods. These three methods are Standardized Morbidity Ratio (SMR), Poisson-gamma model and Besag, Mollie and York (BYM) model. All of these results are then compared based on relative risk values, DIC values and in term of classification range in maps so that the best-fitted model for relative risk estimation for pneumonia disease mapping in Malaysia can be found. The WinBUGS software is used to analyze the data and the ArcGIS software is used to produce the maps in this study.

Table 4.1 and Figure 4.1 display the number of pneumonia cases reported in Malaysia for year 2019. From Table 4.1 and Figure 4.1, Selangor recorded the highest number of pneumonia cases with 19,481 cases while Labuan present the lowest number of pneumonia cases with only 529 cases.

Table 4.1

Number of Pneumonia Cases Reported for 13 States and 3 Federal Territories in Malaysia for Year 2019.

States	**No. of Cases**
Perlis	1,703
Kedah	10,760
Pulau Pinang	4,447
Perak	10,550
Kuala Lumpur	4,325
Putrajaya	1,883

Table 4.1 (continued)

Selangor	19,481
Negeri Sembilan	7,844
Melaka	4,442
Johor	18,617
Pahang	10,434
Terengganu	9,020
Kelantan	12,254
Sabah	16,728
Labuan	529
Sarawak	12,402

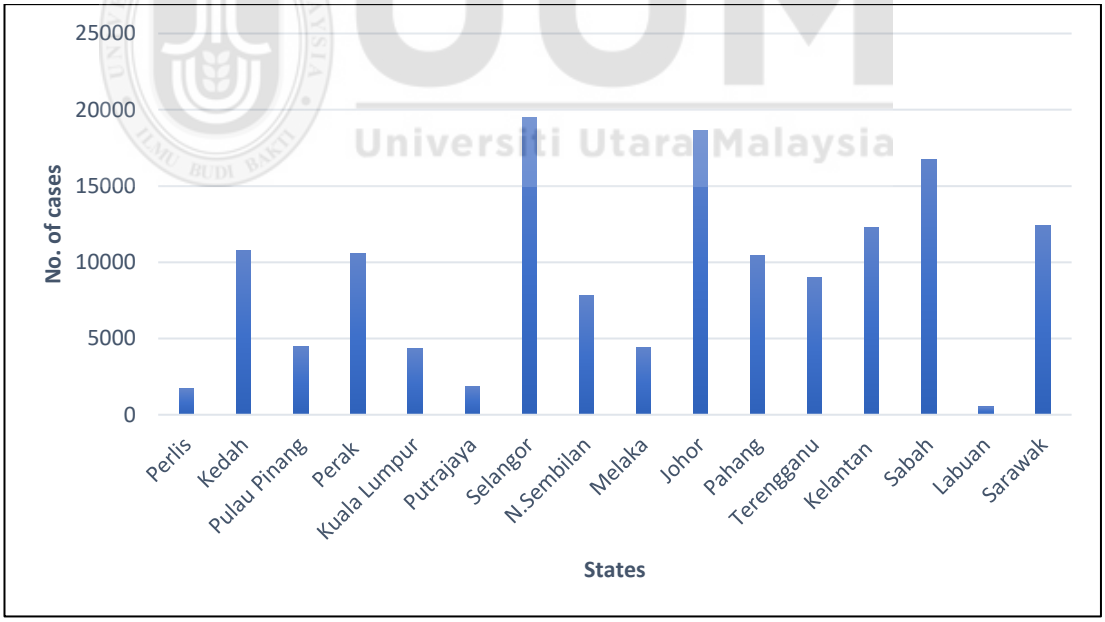

Figure 4.1. Number of pneumonia cases reported for every state in Malaysia in 2019

4.3 Relative Risk Estimation based on Stochastic SIC, Stochastic SIR, Stochastic SCIR and Stochastic SVCIR Models

This section discussed the results of relative risk estimation based on the four stochastic models proposed. Starting with discussing the WinBUGS codes for relative risk estimation for the four stochastic models proposed and description to assess the convergence of the MCMC method. The next subsections start with the relative risk estimation results for stochastic SIC model, followed by stochastic SIR model, stochastic SCIR model and stochastic SVCIR model.

4.3.1 WinBUGS Code for Estimation of Relative Risk based on the Stochastic SIC, Stochastic SIR, Stochastic SCIR and Stochastic SVCIR Models

Figure 4.2 until Figure 4.5 show the WinBUGS codes to compute a relative risk based on the four proposed stochastic models that have been constructed in Chapter Three.

```
#SIC MODEL -Stochastic newIh
Model{
for (i in 1:M){
Sh[i,1]<-Nh[i]-Ih[i,1]-Ch[i,1]
Ih[i,1]<-0.00003147*Nh[i]
Ch[i,1]<-0.1341*Nh[i]
}
for (i in 1:M){
for (j in 2:T){
  Sh[i,j]<-(muH*Nh[i])+sigma*Ih[i,j-1]-newIh[i,j]+(1-muH)*Sh[i,j-1]
  Ih[i,j]<-newIh[i,j] +(1-sigma-muH-delta)*Ih[i,j-1]+epsilon*Ch[i,j-1]
  newIh[i,j]~dpois(lambdanewH[i,j])
  log(lambdanewH[i,j])<-beta0+log((betaH*(Ih[i,j-1]+0.001)/Nh[i])+(gammac*(Ch[i,j-
  1]+0.001)/Nh[i]))+log(Sh[i,j-1]+0.001)+bh[i]
  Ch[i,j]<-(1-epsilon-muH)*Ch[i,j-1]+delta*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[1:16]~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
sigma<-0.002747
}
```


Figure 4.2. WinBUGS code for stochastic SIC model

```
#SIR MODEL -Stochastic newIh
Model{
for (i in 1:M){
      Sh[i,1]<-Nh[i]-Ih[i,1]-Rh[i,1]
      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)*Sh[i,j-1]-newIh[i,j]
          Ih[i,j]<-newIh[i,j]+(1-gamma-muH-muP)*Ih[i,j-1]
          newIh[i,j]~dpois(lambdanewH[i,j])
          log(lambdanewH[i,j])<-beta0+log(Sh[i,j]+0.001)+log(a)+bh[i]
          Rh[i,j]<-gamma*Ih[i,j-1]+(1-muH)*Rh[1,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[1:16]~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
gamma<-1
}
```

Figure 4.3. WinBUGS code for stochastic SIR model

```
#SCIR MODEL -Stochastic newIh
Model{
for (i in 1:M){
Sh[i,1]<-Nh[i]-Eh[i,1]-Ih[i,1]-Rh[i,1]
Ch[i,1]<-0.1341*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])/Nh[i]))*Sh[i,j-1]
  Ch[i,j]<-((betaH*Ih[i,j-1])/Nh[i])*Sh[i,j-1]+(1-muH)*Ch[i,j-1]-newIh[i,j]
  Ih[i,j]<-newIh[i,j] +(1-muH-gamma)*Ih[i,j-1]
  newIh[i,j]~dpois(lambdanewH[i,j])
  log(lambdanewH[i,j])<-beta0+log(a)+log(Ch[i,j-1]+0.001)+bh[i]
  Rh[i,j]<-(1-muH)*Rh[i,j-1]+newRh[i,j]
  newRh[i,j]<-(gamma*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[1:16]~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
gamma<-1
}
```

Figure 4.4. WinBUGS code for stochastic SCIR model

```
#SVCIR MODEL -Stochastic newIh
Model{
for (i in 1:M){
Sh[i,1]<-Nh[i]-Vh[i,1]-Ch[i,1]-Ih[i,1]-Rh[i,1]
Vh[i,1]<-0.5*Nh[i]
Ch[i,1]<-0.01341*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]<-(1-p)*(muH*Nh[i])+n*Vh[i,j-1]+sigma*Rh[i,j-1]+(1-muH-(((K*TA*Ih[i,j-
  1])+(Y*Ch[i,j-1]))/Nh[i])-m)*Sh[i,j-1]
  Vh[i,j]<-p*(muH*Nh[i])+m*Sh[i,j-1]+(1-muH-lamdaV-n)*Vh[i,j-1]
  Ch[i,j]<-o*(((K*TA*Ih[i,j-1])+(Y*Ch[i,j-1]))/Nh[i])*Sh[i,j-1]+o*(lamdaV)*Vh[i,j-1]+(1-
  q)*gamma*Ih[i,j-1]+(1-muH-gammaT-epsilon)*Ch[i,j-1]
  Ih[i,j]<-newIh[i,j] +(1-o)*lamdaV*Vh[i,j-1]+epsilon*Ch[i,j-1]+(1-muH-muP-
  gamma)*Ih[i,j-1]
  newIh[i,j]~dpois(lambdanewH[i,j])
  log(lambdanewH[i,j])<-beta0+log(1-o)+log(((K*TA*Ih[i,j-1]+0.001)+(Y*Ch[i,j-
  1]+0.001))/Nh[i])+log(Sh[i,j-1]+0.001)+bh[i]
  Rh[i,j]<-gammaT*Ch[i,j-1]+q*gamma*Ih[i,j-1]+(1-muH-sigma)*Rh[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[1:16]~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
gamma<-1
}
```

Figure 4.5. WinBUGS code for stochastic SVCIR model

It is to be noted from the WinBUGS codes in Figure 4.2 until Figure 4.5, the CAR prior distribution for diagnostic checking is used to represent the random effect and it is specified using the car.normal function. CAR prior distribution presently one of the most essential and commonly used models in disease mapping to represent spatial correlation (Sun, Tsutakawa & Speckman, 1999; Riebler, Sørbye, Simpson, Rue, Lawson, Lee, & MacNab, 2016; Aswi, Cramb, Duncan & Mengersen, 2020). Detailed description of notations used in the arguments to this function in Figure 4.2 until 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 weight 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.

SumNumNeighH: the length of adjH.

Next, diagnostic for assessing convergence of MCMC method in WinBUGS can be done by examining the 'history' plot of the samples at each iteration and looked for random scatter. Convergence is a term refers to whether the algorithm has reached its equilibrium (target) distribution. If this is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. A 'history' plot from this analysis is presented in Figure 4.6. The output of WinBUGS analysis refer to the results based on stochastic SIC model for epidemiology year 2010 until the epidemiology year 2019 for the states of Perlis. From Figure 4.6, it can be seen that generated observations of the 'history' plot is more convincing in terms of convergence, with all generated values within a parallel zone and no obvious

tendencies or periodicities. Other than 'history' plot, convergence of MCMC methods also can be confirmed through posterior summaries, quantiles graph and posterior densities. A complete WinBUGS output for the summary statistics for this estimation for all states and for other methods can be found in Appendix B until Appendix E.

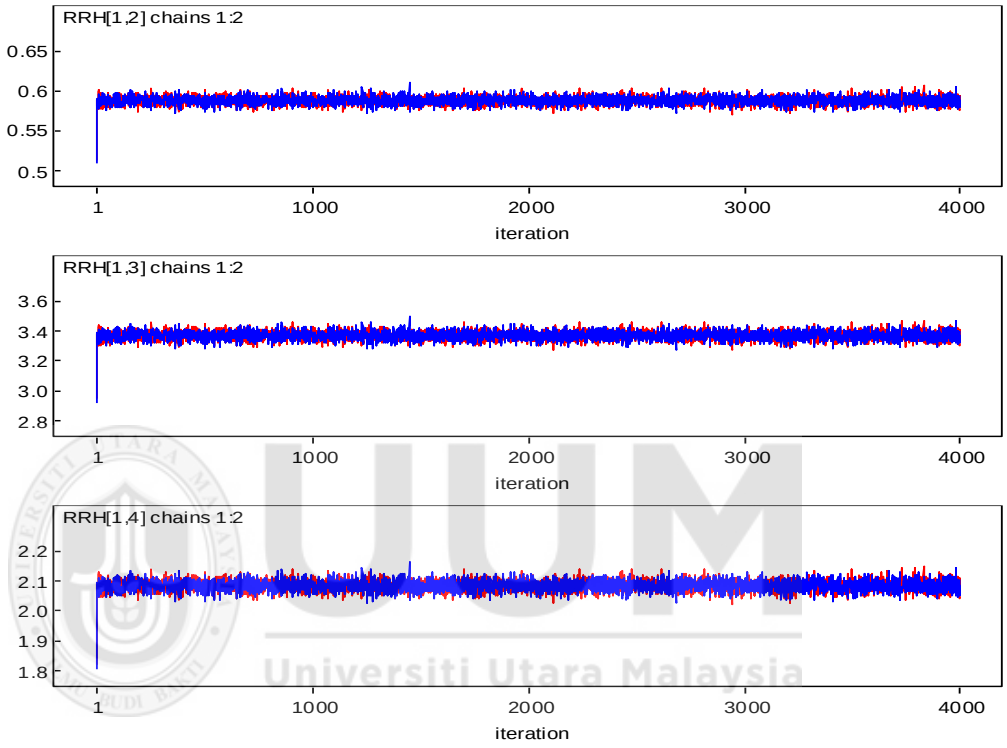

Figure 4.6. Example of WinBUGS output of the 'history' plot for convergence of the relative risk estimation based on stochastic SIC model for Perlis for year 2011 until year 2013

Some examples of WinBUGS output for the posterior summaries of the relative risk estimation based on stochastic SIC model are shown in Table 4.2. The simplest way to monitor convergence is to monitor the MC error (see Table 4.2) since small values of this error will indicate that the quantity of interest is calculated with precision (Ntzoufras, 2009; Peng, 2018). According to Ntzoufras (2009), if MC error is lower than 0.1% of the corresponding posterior standard deviation, hence it is convergence.

Table 4.2

Output of WinBUGS Results for Posterior Summaries of Relative Risk Estimation based on Stochastic SIC Model for the State of Perlis from the Epidemiology Year 2010 until the Epidemiology Year 2019

Node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	0.588	0.00496	6.174E-5	0.578	0.588	0.597
RRH[1,3]	3.369	0.02840	3.538E-4	3.314	3.369	3.432
RRH[1,4]	2.084	0.01757	2.189E-4	2.050	2.084	2.117
RRH[1,5]	1.856	0.01565	1.949E-4	1.826	1.856	1.886
RRH[1,6]	1.550	0.01307	1.628E-4	1.525	1.550	1.575
RRH[1,7]	1.603	0.01351	1.683E-4	1.577	1.603	1.628
RRH[1,8]	1.606	0.01354	1.687E-4	1.580	1.606	1.632
RRH[1,9]	1.424	0.01201	1.495E-4	1.401	1.424	1.447
RRH[1,10]	1.530	0.01290	1.607E-4	1.505	1.530	1.555

* RRH[x,y]: x refers to state and y refers to year

Based on Table 4.2, 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. The node in this study start with year 2011 because situation called moving average.

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 2.5th 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 Figure 4.7, it can be seen clearly that the red solid line representing mean and blue solid line representing median values are approximately the same. Hence, it is convergence.

From Table 4.2, the values of mean and median are approximately the same, this is also shown in posterior distribution density (see Figure 4.8). This shows that the posterior distribution density is reasonably symmetric, hence it is convergence. An example of the WinBUGS results for the posterior densities of the relative risk estimation based on the stochastic SIC model is shown in Figure 4.8.

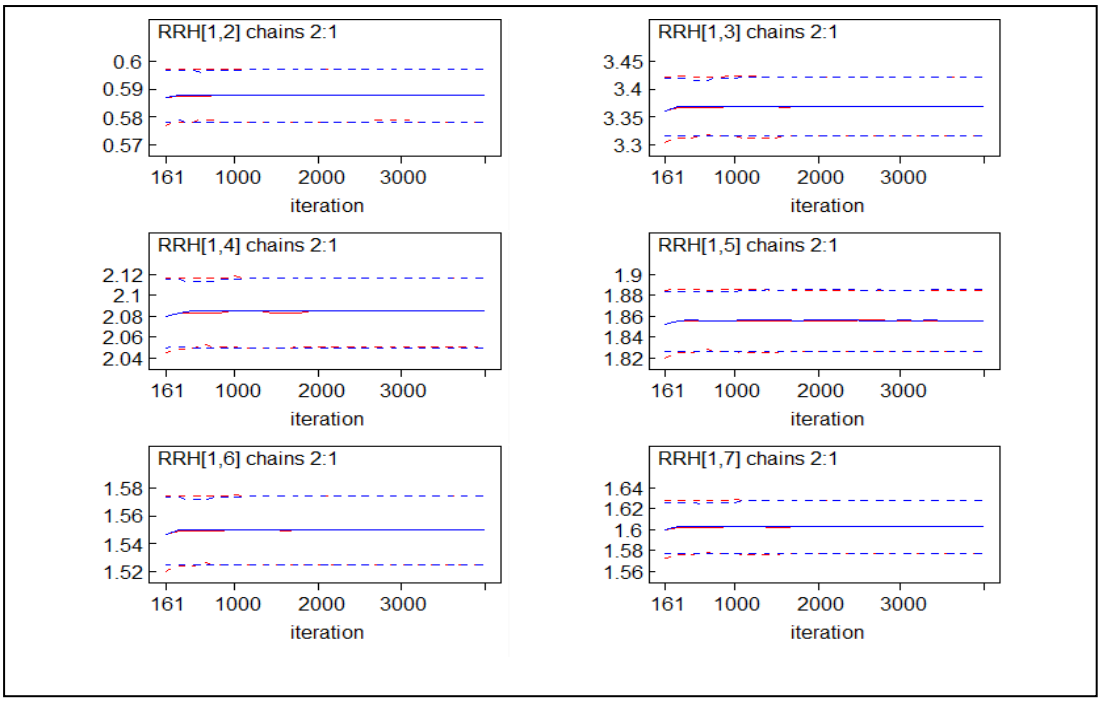

Figure 4.7. Example of WinBUGS output of the quantiles graph of the relative risk estimation based on stochastic SIC model

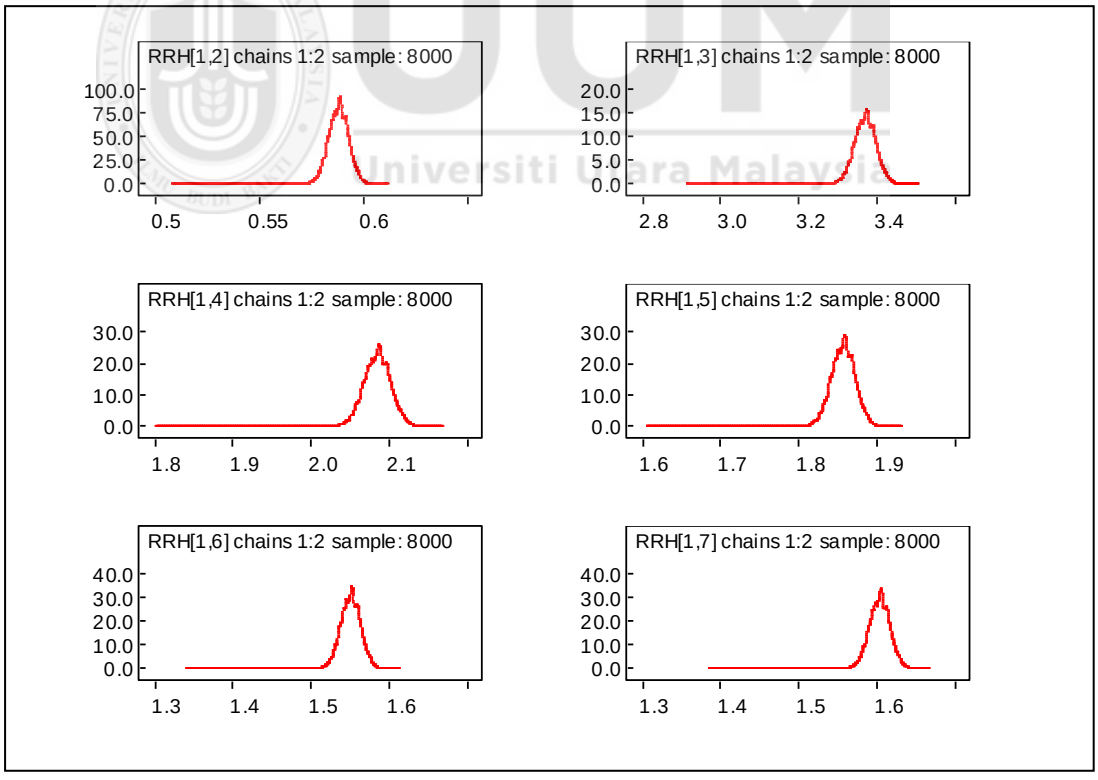

Figure 4.8. Example of WinBUGS output of the posterior densities of the relative risk estimation based on stochastic SIC model

4.3.2 Results of Relative Risk Estimation based on Stochastic SIC Model

The results of relative risk estimation based on the stochastic SIC model are 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 2010 even though the data is collected from 2010 until 2019. This situation is similar when using stochastic SIR, stochastic SCIR and stochastic SVCIR model. This is because of the situation called moving average where the results are calculated by averaging a number of past data points, hence the posterior expected relative risk values only occur at the second data points. Based on the Figure 4.9, it shows that there are five states with relative risk less than one for most epidemiology years which are the states of Pulau Pinang, Kuala Lumpur, Selangor, Sabah and Sarawak. This indicate that susceptible individuals within these five states are less likely to contract pneumonia compared to people in the overall population.

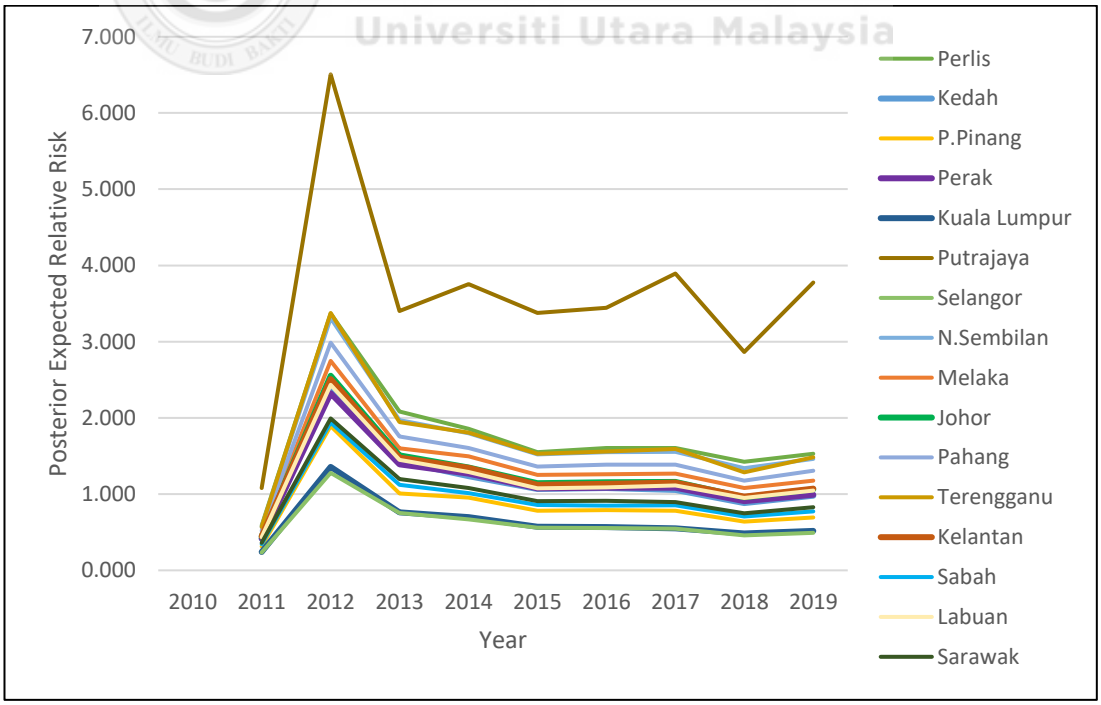

Figure 4.9. Time series plots of the relative risk estimation based on the stochastic SIC model for 13 states and 3 federal territories in Malaysia

4.3.3 Results of Relative Risk Estimation based on Stochastic SIR Model

Figure 4.10 shows the time series plot of the relative risk estimation based on the stochastic SIR model that was extended from basic SIR model proposed by Soliman and Bueno (2018). Based on Figure 4.10, susceptible people in Putrajaya have the highest risk of contracting pneumonia compared to susceptible people in overall population for most epidemiology years, while susceptible people in Selangor have the lowest risk of getting infect with pneumonia. Meanwhile, same with the stochastic SIC results where most of the states have relative risk more than one for all epidemiology years except for Pulau Pinang, Kuala Lumpur, Selangor, Sabah and Sarawak. This shows that these five states are less likely to contract pneumonia compared to the overall Malaysia population.

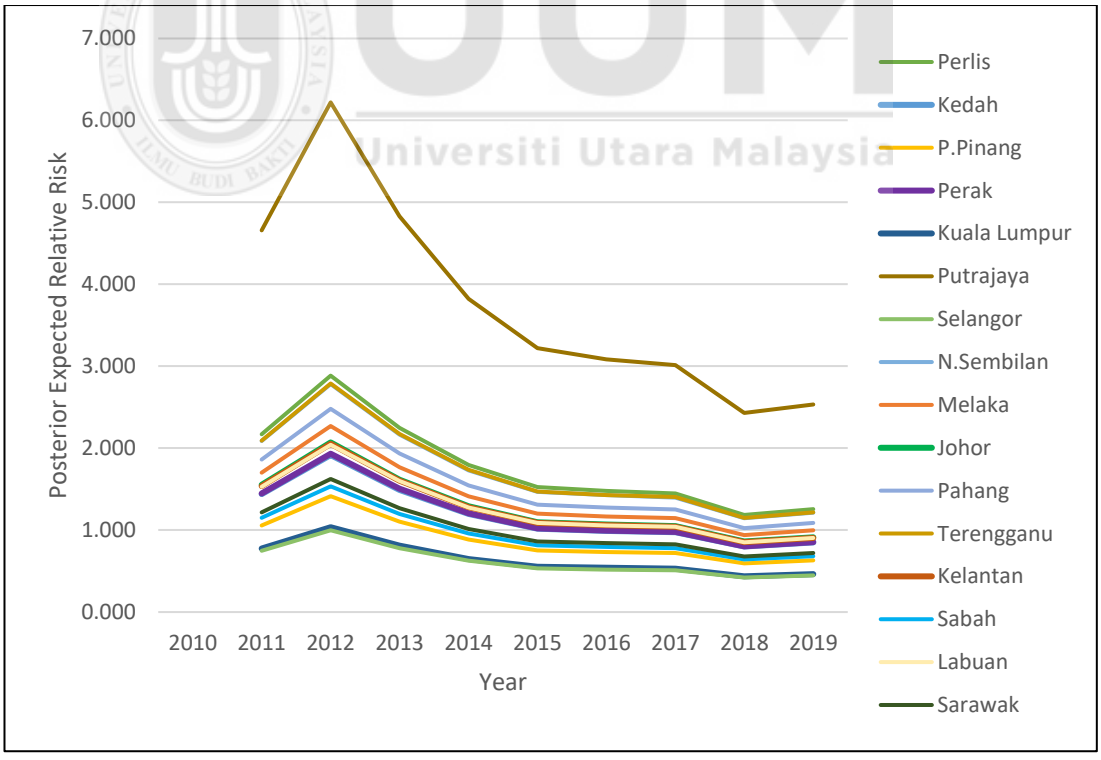

Figure 4.10. Time series plots of the relative risk estimation based on the stochastic SIR model for 13 states and 3 federal territories in Malaysia

4.3.4 Results of Relative Risk Estimation based on Stochastic SCIR Model

Figure 4.11 shows the time series plot of the relative risk estimation based on the stochastic SCIR model. From Figure 4.11, for all epidemiological years, in Putrajaya, susceptible people have the highest risk of contracting pneumonia relative to those in overall population, while susceptible people in Selangor have the lowest risk of contracting pneumonia. The second state with the highest risk is Perlis, followed by the state of Terengganu. The other states with relative risk more than one for most of epidemiology years are Kedah, Perak, Negeri Sembilan, Melaka, Johor, Pahang, Kelantan and Labuan. This shows that susceptible people in these eight states are more likely to contract pneumonia compared to the overall Malaysia population.

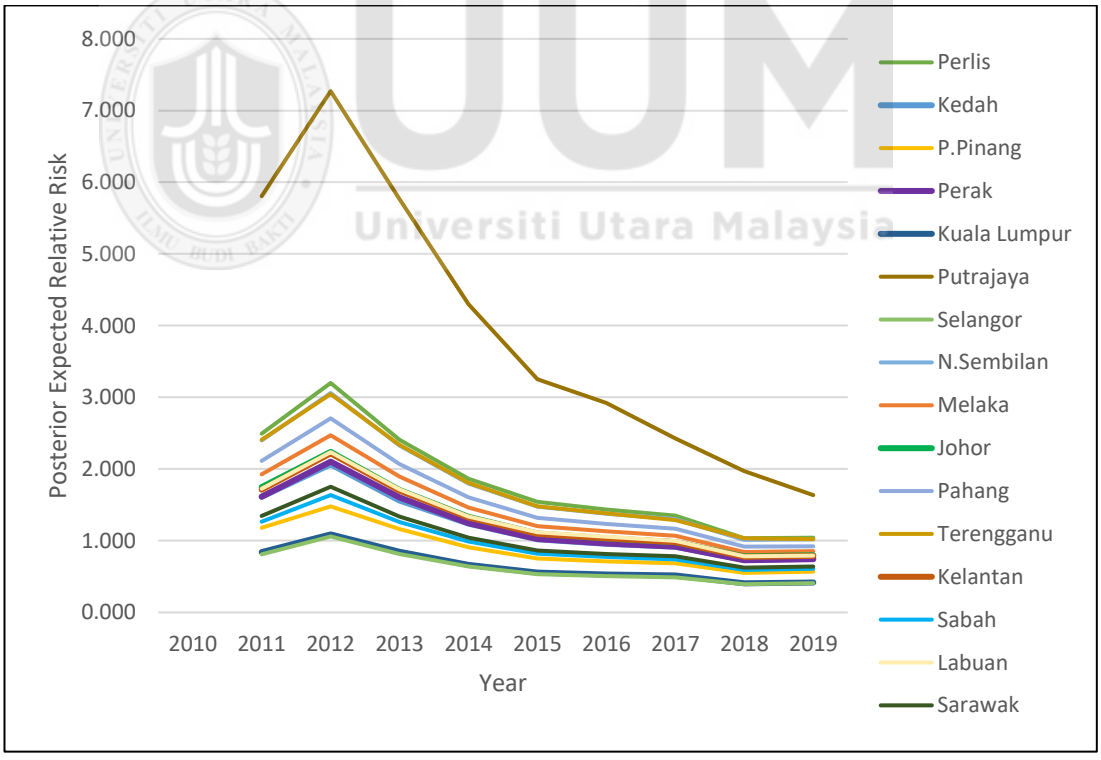

Figure 4.11. Time series plots of the relative risk estimation based on the stochastic SCIR model for 13 states and 3 federal territories in Malaysia

4.3.5 Results of Relative Risk Estimation based on Stochastic SVCIR Model

Figure 4.12 shows the time series plot of the relative risk estimation based on the stochastic SVCIR model. It has been recognized that susceptible people in Putrajaya are more likely to contract pneumonia and have the highest risk of contracting pneumonia compared to those in overall population. This is in contrast with susceptible people in Kuala Lumpur, where they have the lowest risk of contracting pneumonia. The other states with relative risk more than one for most of epidemiology years are Perlis, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu dan Kelantan. This shows that these eight states are more likely to contract pneumonia compared to the overall Malaysia population. The states of Kedah, Pulau Pinang, Perak, Kuala Lumpur, Selangor, Sabah, Labuan and Sarawak have been recognized with relative risk less than one for most of epidemiological years which indicates that susceptible people in these states are less likely to contract pneumonia compared to the overall Malaysia population.

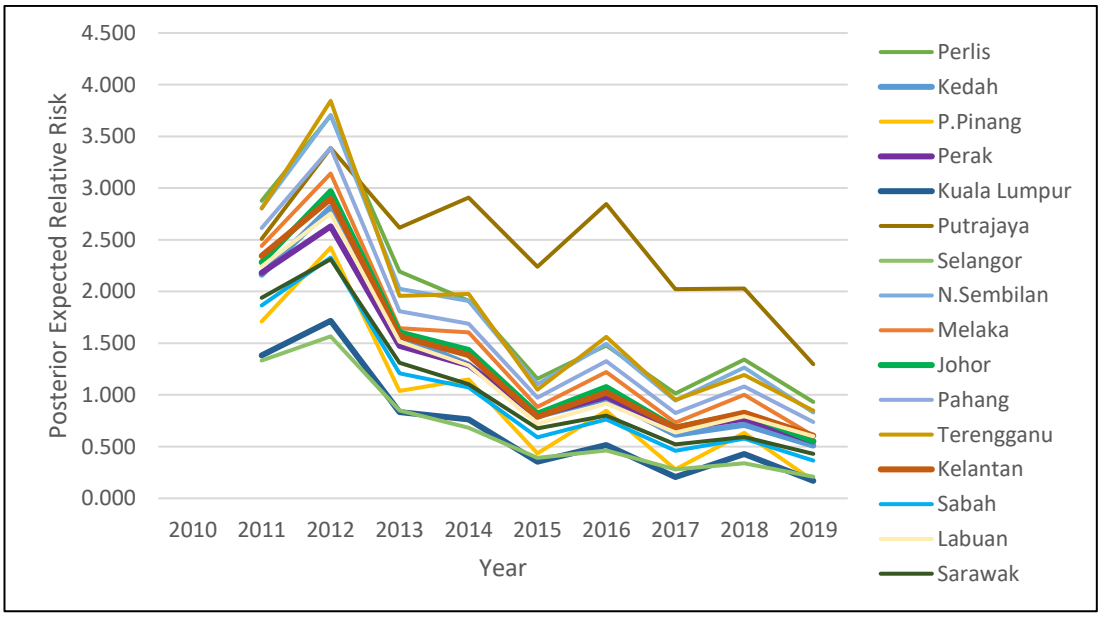

Figure 4.12. Time series plots of the relative risk estimation based on the stochastic SVCIR model for 13 states and 3 federal territories in Malaysia

4.4 Relative Risk Estimation Based on Standardized Morbidity Ratio (SMR) Method, Poisson-gamma Model and Besag, York and Mollie (BYM) Model

The methodologies for method SMR, Poison-gamma model and BYM model have been discussed earlier in Chapter 2. This section presents the relative risk results for these methods. The WinBUGS code for SMR method, Poisson-gamma model and BYM model are presented in Appendix F. Table 4.3 shows the example statistical output of relative risk estimation of pneumonia data using WinBUGS based Poisson-gamma model.

Table 4.3

Example Output for Posterior Expected Relative Risks in the State of Perlis, Malaysia based on Poisson-gamma Model for the Year 2010 until 2019

Node	Mean	Sd	MC error	2.5%	Median	97.5%
theta[1,1]	1.143	0.04331	5.03E-4	1.06	1.143	1.228
theta[1,2]	1.249	0.04129	4.736E-4	1.169	1.249	1.332
theta[1,3]	2.307	0.06597	7.306E-4	2.179	2.307	2.436
theta[1,4]	1.819	0.05177	6.34E-4	1.719	1.819	1.922
theta[1,5]	1.296	0.03928	4.258E-4	1.22	1.296	1.375
theta[1,6]	1.511	0.0389	4.679E-4	1.437	1.511	1.588
theta[1,7]	1.551	0.039	4.129E-4	1.474	1.551	1.628
theta[1,8]	2.183	0.04635	5.255E-4	2.091	2.183	2.273
theta[1,9]	1.683	0.0365	4.562E-4	1.61	1.683	1.755
theta[1,10]	1.46	0.03513	4.185E-4	1.392	1.46	1.53

Table 4.4 shows comparison of the numerical values for the relative risk estimations based on SMR method and Poisson-gamma model for year 2012. While, Table 4.5 show the numerical values for the relative risk estimations based on SMR method, Poisson-gamma model and BYM model for year 2019.

Table 4.4

Relative Risk Estimation based on SMR Method and Posterior Expected Relative Risk based on the Poisson-gamma Model for the Year 2012

States	SMR	Poisson-gamma
Perlis	2.3126	2.3070
Kedah	1.8674	1.8670
Pulau Pinang	0.3538	0.3545
Perak	1.0533	1.0540
Kuala Lumpur	0.1353	0.1363
Putrajaya	**0**	**0.01996**
Selangor	0.5208	0.5210
N. Sembilan	1.8074	1.8060
Melaka	1.1332	1.1330
Johor	1.5235	1.5240
Pahang	1.3795	1.3790
Terengganu	1.5513	1.5500
Kelantan	1.0067	1.0070
Sabah	0.7359	0.7360
Labuan	1.0439	1.0470
Sarawak	1.0585	1.0590

Table 4.5

Relative Risk Estimation based on SMR Method and Posterior Expected Relative Risk based on the Poisson-gamma Model and BYM Model for the Year 2019.

States	SMR	Poisson-gamma	BYM
Perlis	1.4605	1.4600	1.7300
Kedah	1.0939	1.0940	0.9814
Pulau Pinang	0.5586	0.5585	0.5300
Perak	0.9217	0.9218	1.0900
Kuala Lumpur	0.5189	0.5189	0.6289
Putrajaya	4.6379	4.4920	4.5460
Selangor	0.6820	0.6821	0.6533
N. Sembilan	1.5219	1.5220	1.5920
Melaka	1.0655	1.0660	1.2410
Johor	1.0985	1.0990	1.0900
Pahang	1.3747	1.3750	1.4480
Terengganu	1.6520	1.6520	1.5720
Kelantan	1.4908	1.4910	1.4140
Sabah	0.9781	0.9781	0.9375
Labuan	1.1782	1.1790	1.1083
Sarawak	0.9835	0.9835	0.9925

In this study, the prior parameter values, α and b for the Poisson-gamma model are unknown, but they are assumed to have an exponential prior distribution with hyper parameter values of 0.1 as proposed by Lawson et al. (2003) and Samat (2012). The prior expected relative risks are all equal to 1. In this analysis, a gamma prior distribution is used since it is the conjugate prior for the Poisson distribution. This distribution is the most common used for Poisson distribution. For BYM model, as recommended by Bernardinelli et al. (1995) and Samat and Mey (2017), in this study, gamma distributions (0.5,0.0005) are considered which yield a probability of 99% for both. This prior choice has less amount of information and allows the likelihood data to dominate the prior information; therefore, it will have least impact on the relative risk inference.

It can be seen from Table 4.4, the estimation by using SMR model shows susceptible people within the state of Perlis have the highest risk with 2.3126, while susceptible people within the federal territory of Putrajaya have the lowest risk of contracting pneumonia when compared with people in the entire population with 0 relative risk. Based on the Poisson-gamma model, susceptible people within the state of Perlis shows the highest risk, while Putrajaya have the lowest risk with 2.307 and 0.01996, respectively. This shows that the estimated relative risk will be zero when using SMR approach, for areas with no observed cases exist, as in the federal territory of Putrajaya. The relative risk equal to zero means that, there was no risk of contract pneumonia for the susceptible people within the state compared to people in the overall population. Logically, it is impossible for the susceptible people in certain state to have no chance of infection with pneumonia. This is one of the drawbacks of the SMR method. This disadvantage of SMR can be overcome by using Poisson-gamma model

where it provides positive value of relative risk estimation in the states that have no observed case.

Table 4.5 shows the relative risk estimation for the year 2019. It can be seen that susceptible people within the federal territory of Kuala Lumpur for these SMR and Poisson-gamma methods have the lowest risk with both have the same value of 0.5189, while using BYM model, Pulau Pinang has the lowest risk value with 0.5300. The susceptible people within the federal territory of Putrajaya have the highest risk of contracting the disease for these three methods with 4.6379, 4.4920, and 4.5460, respectively. From the result in Table 4.5, it shows the disadvantage of the SMR approach since it is based on ratio estimator by which the mean and variance of SMR are highly dependent on expected cases. The relative risk is very large in region where the expected numbers of pneumonia cases are small as shown for the federal territory of Putrajaya in epidemiology year 2019.

Results by these three methods presented in Table 4.5 give similar conclusions to those results illustrated in Figure 4.13, Figure 4.14, and Figure 4.15. Federal territory of Putrajaya, states of Perlis, Kedah, Perak, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan and Labuan are more likely to contract with the disease compared to people within the overall population, while the other states are less likely to contract pneumonia. Subsection 4.1.5 displays these results in a map that show high-low risk areas of pneumonia occurrences.

The findings of relative risk estimation for these three models are shown in Figure 4.13, Figure 4.14, and Figure 4.15. The graphs show that most states have

relative risk more than one for the year 2010 until 2019. From Figure 4.15, for the time series plots for the BYM model, the relative risk values are too smooth and appear to have lost the jump discontinuities which were apparent in relative risk values when using SMR and Poisson-gamma models. Some of the extreme values estimated from SMR and Poisson-gamma model have vanished. This is due to the model's components taking into account correlated and uncorrelated heterogeneity of area-specific random effects. The variability of risk estimated is determined by uncorrelated heterogeneity. However, in this case, variable *u* controlled by the parameter is small, resulting in lower risk variability. As a result, the risk of a specific area is similar to the risk in neighbouring areas. Putrajaya had the highest risk, which gradually increased from year 2012 to 2019. According to Lawson (2006) and Ideris, Malim and Shaadan (2021), the BYM model is not intended to design for cluster detection since it is a smoothing model.

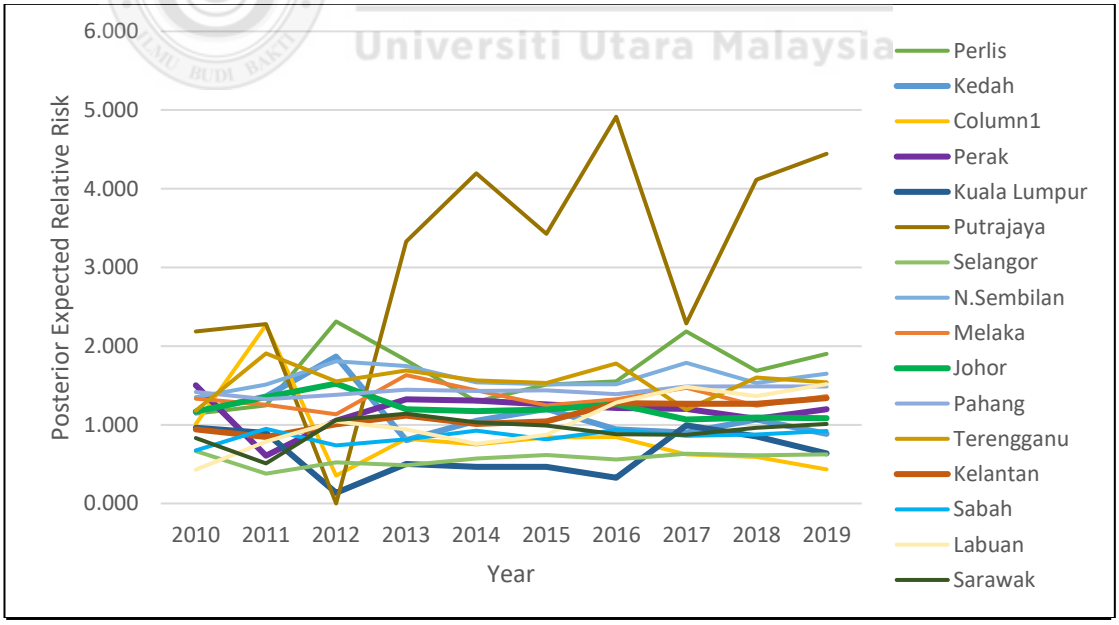

Figure 4.13. Time series plots of the relative risk estimation based on the SMR method for different states in Malaysia

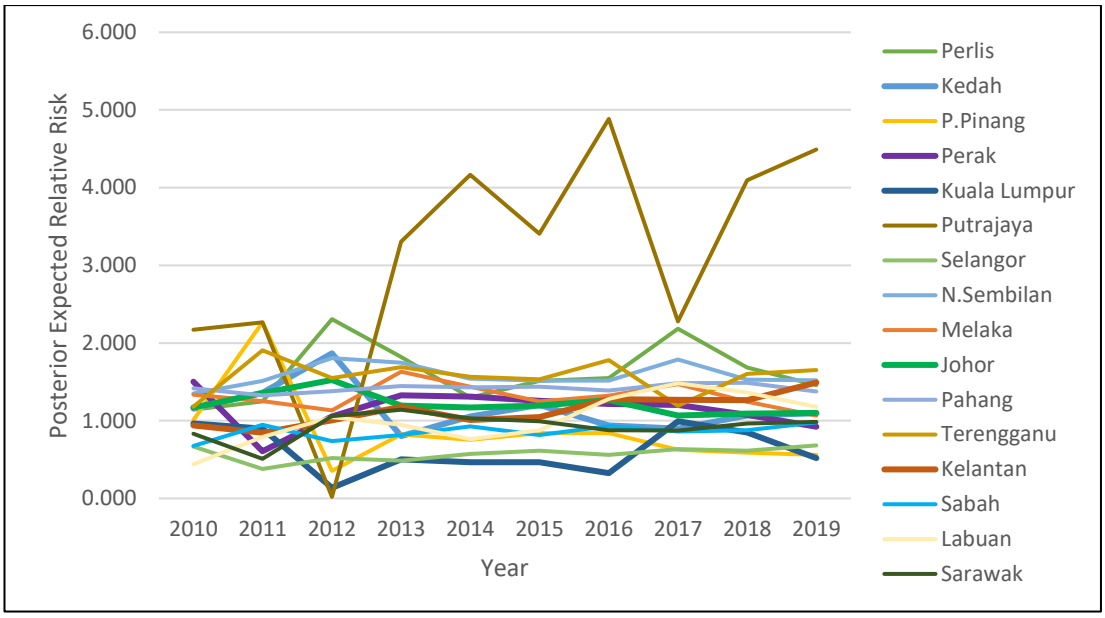

Figure 4.14. Time series plots of the relative risk estimation based on the Poisson-gamma model for different states in Malaysia

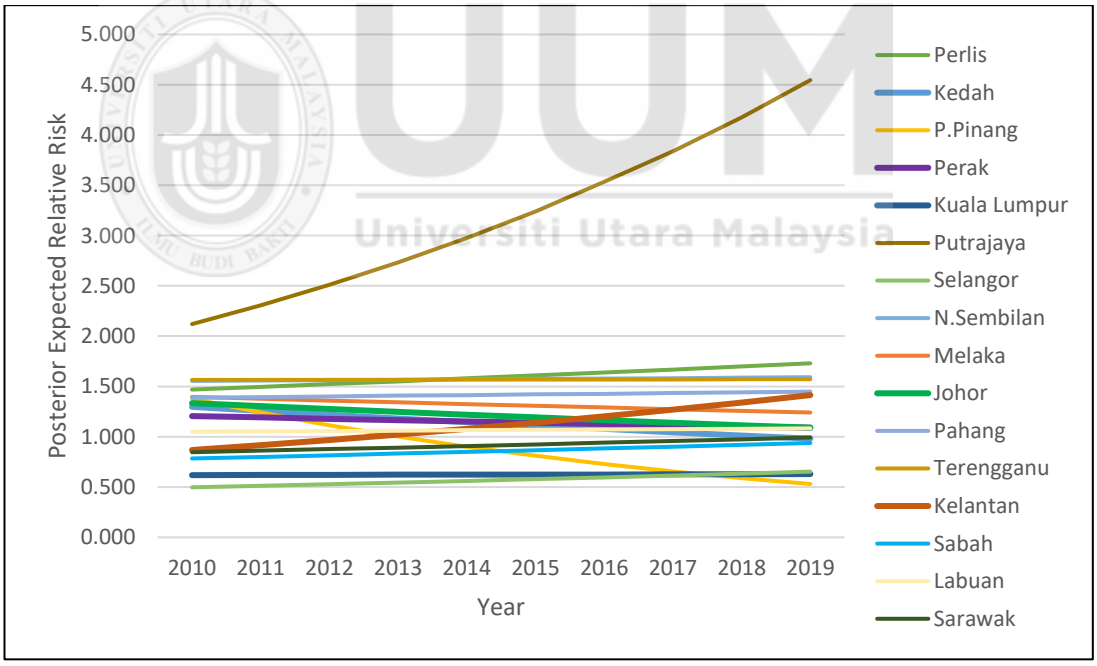

Figure 4.15. Time series plots of the relative risk estimation based on the BYM model for different states in Malaysia

Based on the definition of relative risk estimation values in Chapter Three, if the relative risk value is less than 1, it shows the possibility of contracting the disease has decreased, which means that susceptible people in the region are less likely to contract

with this disease than the people in the whole population. In contrast, for a relative risk value greater than 1, this implies that individuals within the region tend to contract from this disease than those individuals in the overall population in Malaysia. It can be seen clearly from the graph in Figure 4.13, Figure 4.14, and Figure 4.15 that the federal territory of Putrajaya, states of Perlis, Kedah, Perak, Negeri Sembilan, Melaka, Johor, Pahang and Terengganu have a relative risk greater than one for most epidemiology years.

4.5 Comparison of Posterior Expected Relative Risk based on SMR Method, Poisson-gamma Model, Besag, York and Mollie (BYM) Model, Stochastic SIC Model, Stochastic SIR Model, Stochastic SCIR Model and Stochastic SVCIR Model

The comparison results for estimation of relative risk of pneumonia disease mapping based on three existing methods that are SMR method, Poisson-gamma model, and BYM model and four proposed methods that are stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model are displayed in Table 4.6 and Table 4.7. These results correspond to the 13 states and 3 federal territories in Malaysia, specifically for the epidemiology year 2018 and 2019.

Table 4.6

Comparison between the Posterior Expected Relative Risks in the Epidemiology Year 2018 based on Seven Different Models

Relative Risk Estimations for Pneumonia Disease Mapping							
State	**SMR**	**Poisson-gamma**	**BYM**	**Stochastic SIC**	**Stochastic SIR**	**Stochastic SCIR**	**Stochastic SVCIR**
Perlis	**1.461**	**1.683**	**1.699**	**1.424**	**1.185**	**1.035**	**1.341**

Table 4.6 (continued)

Kedah	**1.068**	**1.068**	**1.012**	0.883	0.800	0.735	0.709
P.Pinang	0.591	0.591	0.589	0.641	0.594	0.552	0.633
Perak	**1.074**	**1.074**	**1.102**	0.910	0.802	0.728	0.769
Kuala Lumpur	0.850	0.850	0.628	0.486	0.436	0.406	0.429
Putrajaya	**4.113**	**4.094**	**4.176**	**2.866**	**2.430**	**1.969**	**2.029**
Selangor	0.612	0.612	0.634	0.457	0.421	0.394	0.338
N.Sembilan	**1.530**	**1.530**	**1.588**	**1.341**	**1.145**	**1.012**	**1.265**
Melaka	**1.244**	**1.244**	**1.257**	**1.080**	0.939	0.843	**1.002**
Johor	**1.091**	**1.091**	**1.114**	0.960	0.861	0.783	0.821
Pahang	**1.488**	**1.488**	**1.441**	**1.177**	**1.024**	0.918	**1.082**
Terengganu	**1.601**	**1.601**	**1.571**	**1.285**	**1.149**	**1.033**	**1.192**
Kelantan	**1.265**	**1.264**	**1.340**	0.966	0.853	0.776	0.829
Sabah	0.876	0.876	0.919	0.707	0.642	0.593	0.577
Labuan	**1.365**	**1.364**	**1.079**	0.948	0.851	0.783	0.795
Sarawak	0.962	0.962	0.975	0.750	0.678	0.625	0.593

Table 4.7

Comparison between the Posterior Expected Relative Risks in the Epidemiology Year 2019 based on Seven Different Models

Relative Risk Estimations for Pneumonia Disease Mapping							
State	**SMR**	**Poisson-gamma**	**BYM**	**Stochastic SIC**	**Stochastic SIR**	**Stochastic SCIR**	**Stochastic SVCIR**
Perlis	**1.461**	**1.460**	**1.730**	**1.530**	**1.255**	**1.041**	0.932
Kedah	**1.094**	**1.094**	0.981	0.980	0.849	0.747	0.510

Table 4.7 (continued)

P. Pinang	0.559	0.559	0.530	0.694	0.632	0.571	0.170
Perak	0.922	0.922	**1.090**	0.993	0.852	0.743	0.550
Kuala Lumpur	0.519	0.519	0.629	0.519	0.464	0.418	0.169
Putrajaya	**4.638**	**4.492**	**4.546**	**3.775**	**2.534**	**1.636**	**1.297**
Selangor	0.682	0.682	0.653	0.493	0.447	0.407	0.209
N. Sembilan	**1.522**	**1.522**	**1.592**	**1.463**	**1.213**	**1.018**	0.830
Melaka	**1.065**	**1.066**	**1.241**	**1.175**	0.997	0.857	0.605
Johor	**1.099**	**1.099**	**1.090**	**1.062**	0.913	0.797	0.552
Pahang	**1.375**	**1.375**	**1.448**	**1.308**	**1.086**	0.922	0.738
Terengganu	**1.652**	**1.652**	**1.572**	**1.485**	**1.216**	**1.023**	0.848
Kelantan	**1.491**	**1.491**	**1.414**	**1.071**	0.904	0.786	0.606
Sabah	0.978	0.978	0.938	0.775	0.682	0.608	0.365
Labuan	**1.178**	**1.179**	**1.083**	**1.045**	0.903	0.793	0.600
Sarawak	0.984	0.984	0.993	0.827	0.720	0.639	0.430

Based on Table 4.6 and Table 4.7, it can be concluded that by using these seven methods, Putrajaya has the highest risk area of contracting pneumonia. While for the lowest risk value of contracting pneumonia in the overall population for epidemiology year 2018 and year 2019 are not same for all seven methods. For SMR, Poisson-gamma and BYM models, in 2018, the state of Pulau Pinang has been recognized as the lowest risk value to contract pneumonia. However, by using the other four methods, Selangor has the lowest risk value to contract pneumonia even though it has different value for respective methods. In 2019, by using SMR and Poisson-gamma models, Kuala Lumpur has been recognized as the lowest risk value of contracting

pneumonia. Differ for BYM model, in 2019, Pulau Pinang has the lowest risk value to contract the disease. While, for the other four methods, Selangor has the lowest risk value except for stochastic SVCIR model where Kuala Lumpur has been recognized as the lowest risk value. In this analysis, there is a zero relative risk value when using SMR method as there is no observed count data in Putrajaya in year 2012 as mentioned earlier in Section 4.4.

From Table 4.6, five from sixteen states have relative risk estimation more than one when using all methods, and from Table 4.7, only one state has relative risk estimation more than one for all methods. This condition can be concluded that susceptible persons in this state are more likely to be infected with pneumonia compared to the overall population in Malaysia, whereas susceptible persons in the other states are less likely to be infected with pneumonia compared to persons in the entire population. Stochastic SVCIR model in Table 4.7 shows the smallest range of posterior expected relative risk across states when compared with other six methods with a minimum value of 0.169 and the maximum value of 1.297.

However, the most important is that these four models introduced are potentially more suitable than the three existing methods (SMR, Poisson-gamma model and BYM model). This is because these four stochastic models include a more detailed description of basic biology process and the nature of disease transmission. These four stochastic models 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 pneumonia incidence which are display in Section 4.7. These maps can be used by the government or other interested parties as a tool to recognize which states that need extra attention in terms of government policy and financial and also other supports in order to control and prevent pneumonia disease from becoming worse. To decide which method generates a smoother map, the maps are compared using these seven methods in Section 4.7. Before that, to find which proposed models fits the data available the best, model goodness-of-fit (GOF) measures is evaluated.

4.6 Model Goodness-of-fit (GOF) Measures

In this section, model goodness-of-fit (GOF) measures will be tested to help in making decision which models fit data the most. In statistical modeling, the use of GOF measures is frequently used in comparing fitted models. There are several methods that can be used as model GOF measures and these methods have been discussed by Lawson (2009) such as Akaike Information Criterion (AIC), Bayes Information Criterion (BIC), chi-square statistic and Deviance Information Criterion (DIC). BIC is widely used in Bayesian and hierarchical models as a model choice criterion. It asymptotically approximates a Bayes factor. However, in this study, DIC is used as model GOF measures since according to Lawson (2009) and Samat (2012), other measures, such as AIC and BIC have drawbacks when a model with several random effects is involved, while the posterior predictive loss in this analysis includes the predictive distribution. Besides, by using DIC, it can be evaluated easily in the WinBUGS software. The DIC is based on the posterior distribution of the log-likelihood and is easy to be applied to a wide range of statistic models. Spiegelhalter,

Best, Carlin and van der Linde (2002) proposed the DIC as model GOF measures. This DIC is defined as

$$DIC = 2E_{\theta|x}\{D(\theta)\} - D\{E_{\theta|x}(\theta)\}.$$

where $D(.)$ denotes the model's deviance and x referred to the observed data. In order to obtain an expected value of Θ, the average of the posterior samples of Θ is used. This value can be computed from a sample output from a chain. Based on Spiegelhalter et al. (2002) and Alhdiri, Samat and Mohamed (2017), the smaller the value for DIC means the better the model fits the data. The model with the smallest DIC is predicted to be the one that will ideally simulate a replicate data set with the similar form as the one currently observed. Although, Lawson et al. (2003) and Maryam et al. (2017) stated that all the model GOF measures are useful in assisting to pick the model and they offer little help in determining how well the model fits the data.

Table 4.8 displays the values of DIC for the newly infective persons in epidemiology year 2010 until 2019 based on four different stochastic models proposed in this study for pneumonia data for all states in Malaysia.

Table 4.8

Deviance information criterion (DIC) for relative risk estimation for pneumonia disease based on stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model

	Stochastic SIC	**Stochastic SIR**	**Stochastic SCIR**	**Stochastic SVCIR**
DIC	169,247.00	108,307.00	151,115.00	145,216.00

Based on Table 4.8, from the DIC values, stochastic SIR model fits the data best as it shows the smallest DIC value compared to the other models. This concludes that the stochastic SIR model is the best model to be used to analyze the pneumonia data available in this study. However, this selection of model may be not the 'true' model in estimating the relative risk. This will be discussed more on Section 4.2. In the next section, maps are generated to decide which method produces a smoother map.

4.7 Disease Mapping for Relative Risk Estimation in 2019

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 simplify the risk classification, the results of relative risk are classified into five different classes which are as follows in Table 4.9:

Table 4.9

Classes of Relative Risk Estimation (Source: Samat, 2012; Awang, 2017; Alhdiri, Samat & Mohamed, 2017)

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 to show and distinguish between the low and high risk areas of pneumonia occurrences with one-colour tones. The epidemiology year 2019 is chosen as an example and for comparative purposes only.

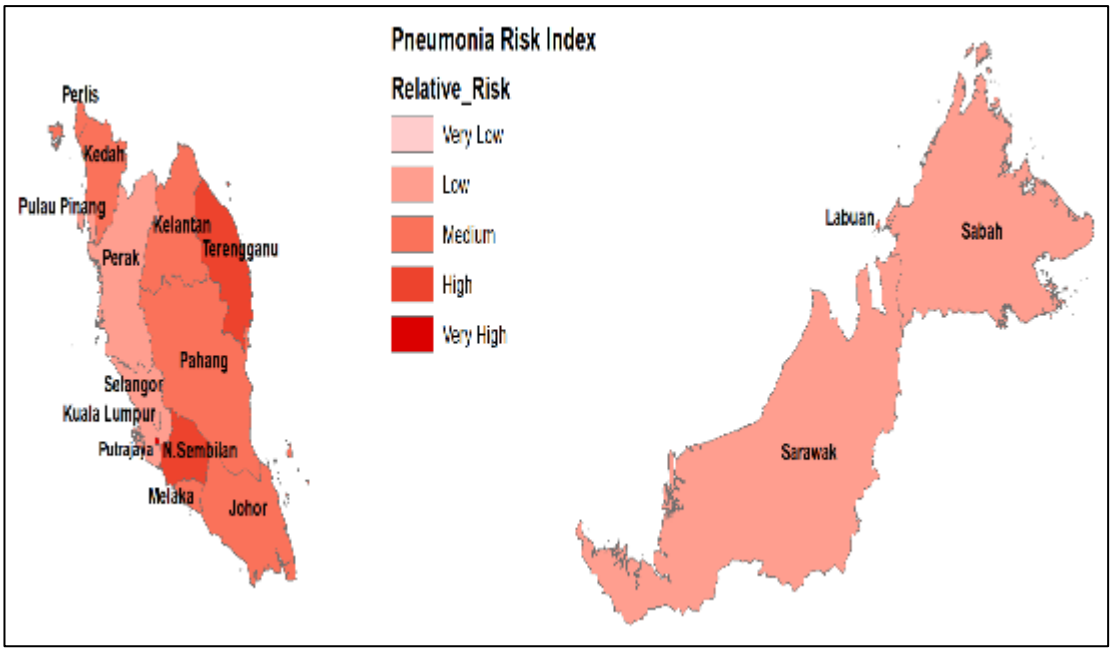

Figure 4.16. Disease map of relative risk estimation based on SMR method for the year 2019

Figure 4.16 shows the disease map based on the SMR method. Based on the figure, Putrajaya has been recognized as a very high-risk area for the epidemiology year 2019, followed by Negeri Sembilan and Terengganu states with high-risk areas. The other states have been categorized as medium risk except for the states of Pulau Pinang, Perak, Kuala Lumpur, Selangor, Sabah and Sarawak which have been categorized as low risk. No state categorized as very low risk area based on Figure 4.16.

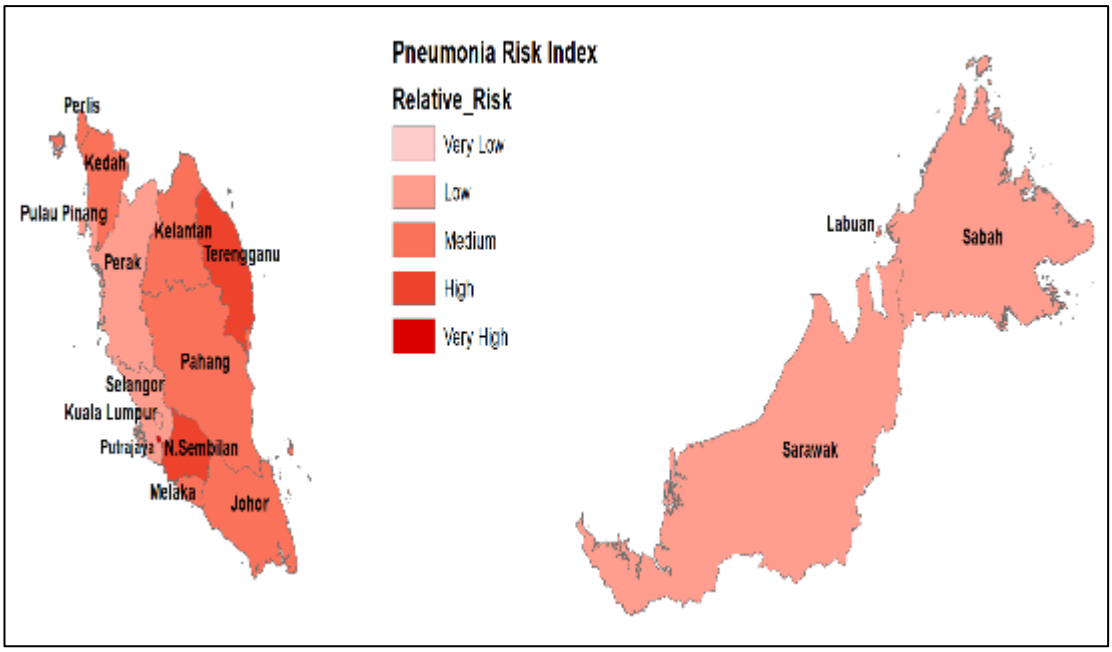

Figure 4.17. Disease map of relative risk estimation based on Poisson-gamma model for the year 2019

Similar to those results in the SMR map in Figure 4.16, Putrajaya has been classified as very high risk area for Poisson-gamma map, with no state classified as very low risk. Negeri Sembilan and Terengganu show the high risk while the other six states that are Perlis, Kedah, Melaka, Johor, Pahang and Kelantan with medium risk level. Another seven states are classified with low risk level. Findings are displayed in Figure 4.17.

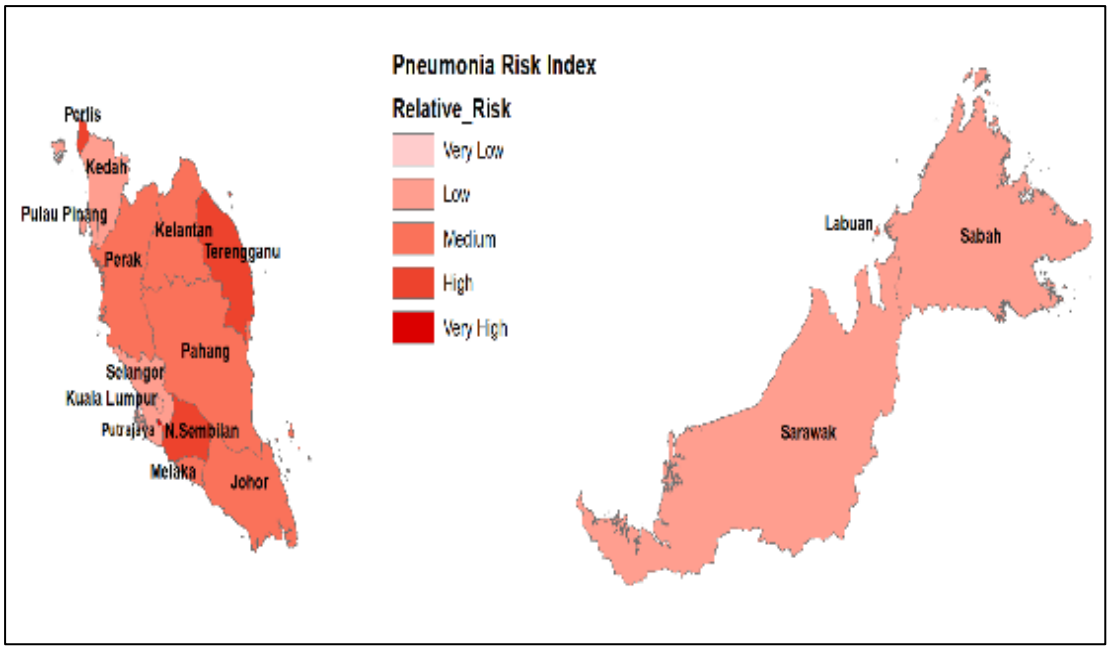

Figure 4.18. Disease map of relative risk estimation based on BYM model for the year 2019

In Figure 4.18, Putrajaya has been classified as very high risk area for BYM map, with no state classified as very low risk area. Perlis, Negeri Sembilan and Terengganu show the high risk while the other states that are Perak, Melaka, Johor, Pahang, Kelantan and the federal territory of Labuan are identified with medium risk level. Another seven states are classified with low risk level.

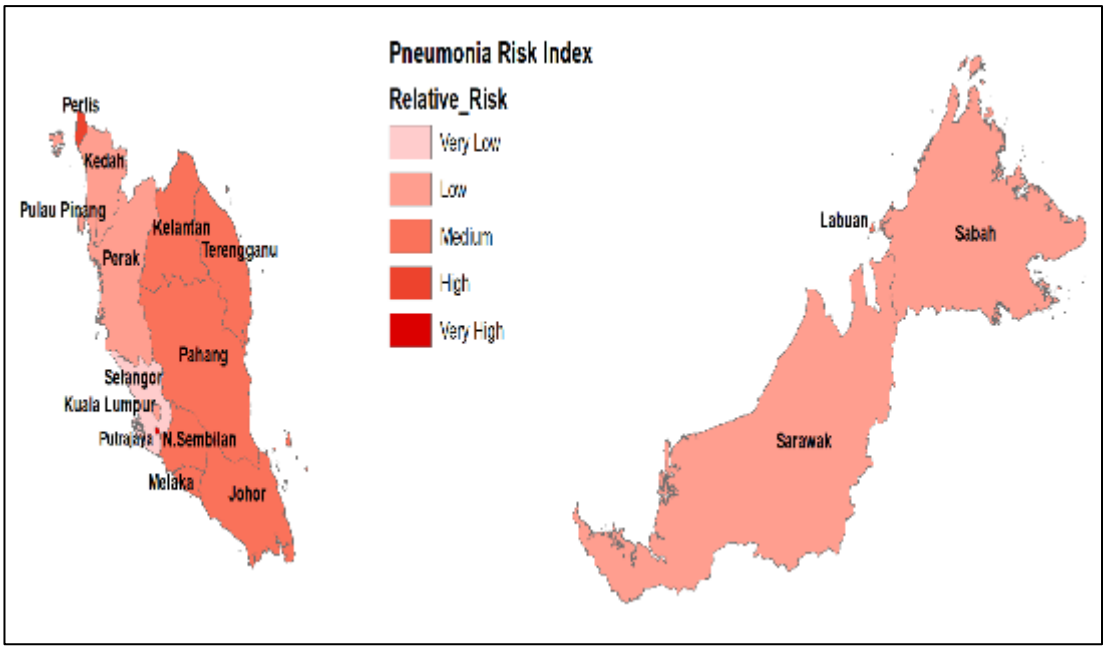

Figure 4.19. Disease map of relative risk estimation based on stochastic SIC model for the year 2019

Figure 4.19 presents the high-low risk area based on stochastic SIC model for the epidemiology year 2019. Similar to those results in the SMR map, Poisson-gamma map and BYM map, Putrajaya has a very high risk for pneumonia incidence for stochastic SIC map. The state of Perlis is the only state that has been recognized as high-risk area. These are followed by Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan and Labuan with medium risk. The other states are classified as a low risk for pneumonia occurrences except for Selangor which has very low risk area for stochastic SIC map.

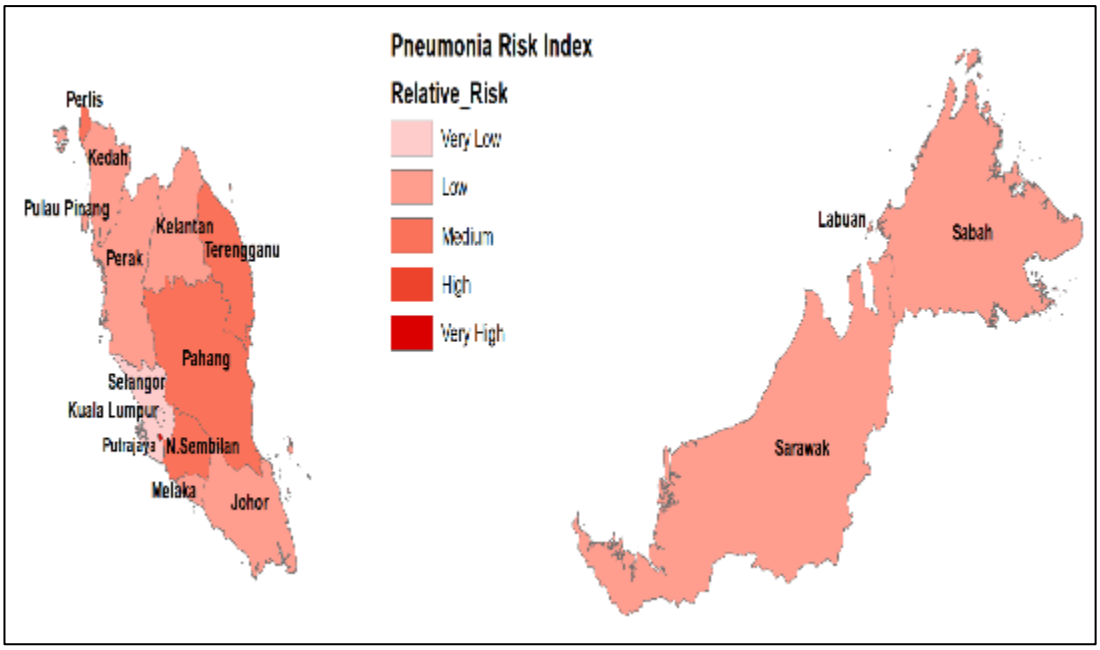

Figure 4.20. Disease map of relative risk estimation based on stochastic SIR model for the year 2019

Figure 4.20 displays the high-low risk area based on stochastic SIR model for the epidemiology year 2019. Putrajaya also has been identified to has a very high risk for occurrence of pneumonia, which has the same results for SMR map, Poisson-gamma map, BYM map and stochastic SIC map. However, for stochastic SIR map, there is no state identified with high risk. There are four states that have been identified with medium risk which are Perlis, Negeri Sembilan, Pahang and Terengganu. The other states have low risk except for Selangor and Kuala Lumpur with very low risk for stochastic SIR map.

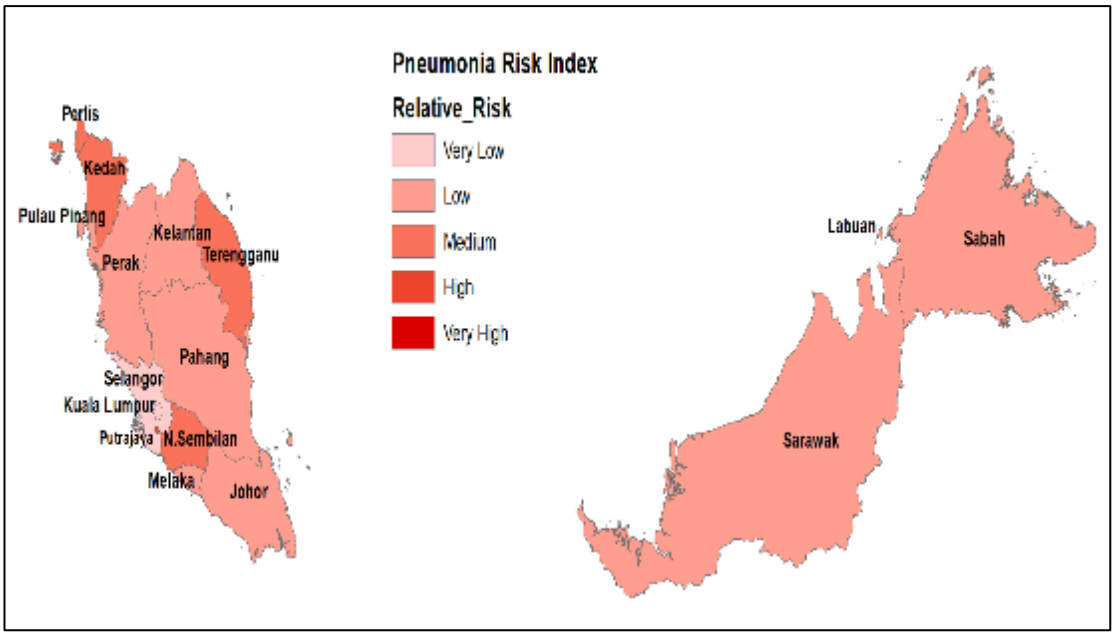

Figure 4.21. Disease map of relative risk estimation based on stochastic SCIR model for the year 2019

In contrast to the SMR map, Poisson-gamma map, BYM map, stochastic SIC map and stochastic SIR map, where the federal territory of Putrajaya has a very high risk of pneumonia occurrence for the five models, the stochastic SCIR map results in Figure 4.21 shows that Putrajaya is only identified as high risk. However, Putrajaya still has the highest risk value compared to the other states. There is no state classified as very high-risk area. The state of Perlis, Negeri Sembilan and Terengganu are categorized as areas with medium risk, followed by other states with low-risk areas except for Kuala Lumpur and Selangor with very low risk areas. These results for very low risk areas are similar with the stochastic SIR map.

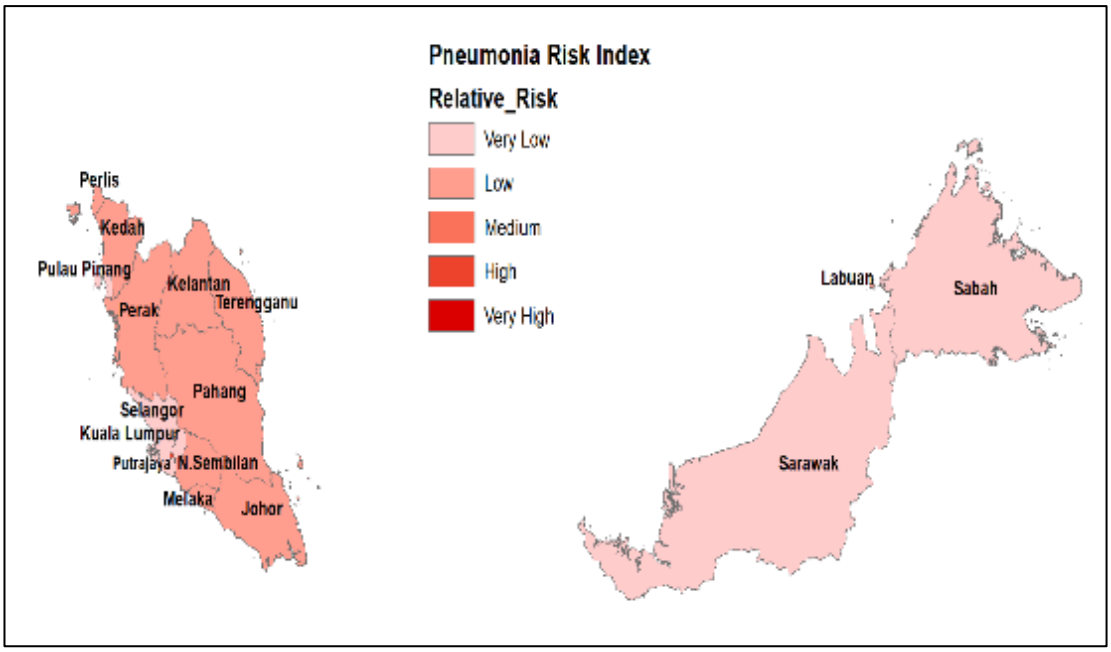

Figure 4.22. Disease map of relative risk estimation based on stochastic SVCIR model for the year 2019

Figure 4.22 shows the risk areas based on the stochastic SVCIR model. From the stochastic SVCIR map, it shows different result from the SMR map, Poisson-gamma map, BYM map, stochastic SIC map and stochastic SIR map for the federal territory of Putrajaya where it is classified as medium risk area of pneumonia occurrence. These are followed by ten states with low risk that are Perlis, Kedah, Perak, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan and Labuan. Another five states have been identified as very low risk of pneumonia occurrence. From the Figure 4.22, it can be seen that there is no state classified with very high or high risk of pneumonia occurrence.

4.8 Discussion on Relative Risk Estimation Models

Table 4.10

Posterior Expected Relative Risk based on Seven Different Methods for the States with Value of Relative Risk More Than One for Year 2019

	Relative Risk Estimations for Pneumonia Disease Mapping						
State	**SMR**	**Poisson - gamma**	**BYM**	**Stochastic SIC**	**Stochastic SIR**	**Stochastic SCIR**	**Stochastic SVCIR**
Perlis	**1.461**	**1.460**	**1.73**	**1.530**	**1.255**	**1.041**	0.932
Kedah	**1.094**	**1.094**	0.981	0.980	0.849	0.747	0.510
Perak	0.922	0.922	**1.090**	0.993	0.852	0.743	0.550
Putrajaya	**4.638**	**4.492**	**4.546**	**3.775**	**2.534**	**1.636**	**1.297**
N. Sembilan	**1.522**	**1.522**	**1.592**	**1.463**	**1.213**	**1.018**	0.830
Melaka	**1.065**	**1.066**	**1.241**	**1.175**	0.997	0.857	0.605
Johor	**1.099**	**1.099**	**1.090**	**1.062**	0.913	0.797	0.552
Pahang	**1.375**	**1.375**	**1.448**	**1.308**	**1.086**	0.922	0.738
Terengganu	**1.652**	**1.652**	**1.572**	**1.485**	**1.216**	**1.023**	0.848
Kelantan	**1.491**	**1.491**	**1.414**	**1.071**	0.904	0.786	0.606
Labuan	**1.178**	**1.179**	**1.083**	**1.045**	0.903	0.793	0.600

The value of relative risk decreases when using the stochastic SVCIR model for all states in Malaysia. Table 4.10 shows the list of states with value of relative risk more than 1. From Table 4.10, it is clearly seen that the value decrease when using the four stochastic models proposed especially for stochastic SVCIR model. This is because of the added elements, the carrier, recovered and vaccine, when using SVCIR model besides it considers the infectious element. For SMR, Poisson-gamma model, and BYM model, these three methods only consider infectious. Whilst for stochastic SIC model and stochastic SCIR model, these two methods only consider carrier and

infectious elements, and for stochastic SIC model, it does not consider the recovered element. Stochastic SIR model also considers two elements only that are infectious and recovered. The findings from Table 4.10 are clearly support the statement by Centers of Disease Control and Prevention (CDC) (2020) where the component of vaccine can give an effect to the risk for someone to contract pneumonia. WHO (2020b) also stated that vaccines can help to prevent pneumonia disease. According to the studies conducted, when using vaccines, it helps to protect against pneumonia although it cannot prevent all cases (CDC, 2020). The studies showed that when getting at least 1 shot of pneumococcal conjugate vaccine (PCV13), at least 8 in 10 babies from serious infections called invasive pneumococcal disease, 3 in 4 adults 65 years or older against invasive pneumococcal disease and 9 in 20 adults 65 years or older against pneumococcal pneumonia (CDC, 2020).

However, stochastic SCIR model can be considered as the second-best model to be used for estimating the relative risk since there is no data available for vaccine. In this study, the data for vaccine is assumed and adopt from previous study. Besides, stochastic SCIR model has added element, the carrier, recovered and also considers the infectious element. From Table 4.7, stochastic SCIR model shows the second smallest range of posterior expected relative risk across states when compared with other six methods with a minimum value of 0.407 and the maximum value of 1.636, respectively.

Result from Subsection 4.6 shows that the stochastic SIR model has the lowest DIC value compared to the other methods, followed by the stochastic SVCIR model. This result indicates that the stochastic SIR model is the best model to fit the data available. According to Deeth, Deardon and Gillis (2015), in reality, of course, more

than a single model selection criterion can be used in the process of selecting the most appropriate model to explain disease dynamics. For example, the posterior predictive capabilities of models could also be considered, as this enables researchers to analyze how different aspects of the disease system can be described by each model, such as the basic reproduction number, or other epidemic curve characteristics. These methods are also typically used in the model validation process, with which comparative measures such as the DIC are of course not helpful (Deeth, Deardon & Gillis, 2015). According to Spiegelhalter, Best, Carlin and van der Linde (2014) and Pooley and Marion (2018), even though DIC can help in selecting the best model, but the aim of DIC is not to find a 'true' model, but rather to accurately forecast future data sets in a world in where actual data generating is actually very high-dimensional (and unattainable). Based on the result's study from Pooley and Marion (2018), the used of DIC shows it is not suitable in selecting 'true' model especially for transmission models. Several researchers disputed DIC's ability to compare hierarchical or multi-level models when re-parameterisation happens (Pooley & Marion, 2018; Dey, Delampady & Gopalaswamy, 2019). According to Ideris, Malim and Shaadan (2021), finding a study that compares DIC values between different types of mathematical models is difficult. This is due to the fact that mathematical functions and operations differ amongst models with various operator scopes. As a result, comparing various types of mathematical models with different mathematical functions might lead to incorrect interpretation of DIC (Ideris, Malim and Shaadan, 2021).

Based on all the results provided in this chapter, it can be concluded that Putrajaya has the highest risk of pneumonia occurrence while Kuala Lumpur is the lowest risk area of pneumonia occurrence. These results are comparable with the

number of prevalence for pneumonia disease where in 2019, 176 per 10,000 residents number of prevalence been estimated in Putrajaya, whilst 24 per 10,000 residents in Kuala Lumpur. Based on the data from Ministry of Health Malaysia annual report 2019, Putrajaya recorded the highest annual population growth rate of 6.55%, while Kuala Lumpur recorded a declining annual population growth rate of 0.52% (Ministry of Health Malaysia, 2019). Apart from that, Putrajaya is the smallest region in Malaysia (Department of Survey and Mapping Malaysia, 2017). Therefore, susceptible people in Putrajaya have the highest risk of contracting pneumonia since it has the highest annual population growth rate and the smallest region. Hence, from this information, it can be concluded that the size of the population in the region and the size of the region itself can affect the value of relative risk estimation. Therefore, in this study, proposed stochastic models consider spatial elements and the number of populations for each state. Apart from that, as the federal administrative capital of Malaysia, Putrajaya is the area with high community transmission since people around the country or world come to this area all the time. From Figure 4.1, if it is based on the count data for pneumonia disease occurrence in Malaysia, it shows that Selangor has the highest number of cases compared to Putrajaya. However, when using stochastic transmission models, which considers other factors such as spatial correlation between risk in neighbouring area and number of populations, susceptible people in Putrajaya shows the highest risk of getting pneumonia disease compared to Selangor. For Selangor, even though it has the highest number of pneumonia cases reported, but when compared with its population, the number of prevalence for pneumonia disease only 28 per 10,000 residents. Based on the mortality rate, Putrajaya shows that it has the highest mortality rate compared to other states with 0.5 deaths per 1,000 persons, while for Selangor, the mortality rate is 0.14 deaths per 1,000 persons.

All models mentioned before which are SMR, Poisson-gamma model, BYM model, stochastic SIC model, stochastic SIR model, and stochastic SCIR model can produce maps. However, the map produced using the proposed model that is stochastic SVCIR model has its advantage. From Figure 4.22 which is pneumonia risk map based on stochastic SVCIR method for the epidemiology year 2019, 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 and consider the carrier, recovered and vaccine components. 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. These stochastic models allow to assign distributions to variables and parameters that represent the random nature of the disease transmission, which also included the space-time factor. Although these stochastic models proposed in this study are specifically relates to pneumonia disease, all of these methods extend readily to apply more generally to other direct airborne infectious diseases which have similar way in transmission of disease as pneumonia.

According to WHO (2021), pneumonia cases in children under the age of five are mostly caused of their weakened immune systems. Malnutrition or undernourishment can weaken a child's immune system, especially in newborns who are not exclusively breastfed. Pre-existing diseases, such as symptomatic HIV infections and measles, also raise a child's risk of contracting pneumonia (WHO, 2021). Based on the study conducted by Rahman, Ismail, Rahman, Yusof and Idris (2020), their study shows that the age group most impacted with measles is age 1 to <7 years with 50.8%. Adults in Malaysia are less affected by measles as a result of the initiation of the measles immunization programme in 1982. In the present study,

more than half of the measles epidemic cases involved unvaccinated children. Hence, this led them to get pneumonia disease also.

The results shown in Chapter Four presented the application of various models for estimation of relative risk and using these relative risks to generate the pneumonia risk maps in Malaysia. This study has proposed the four new models: stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model in estimating relative risk values. These proposed models can generate a more detailed disease map for relative risk of pneumonia in Malaysia. These models are also important since they can later help to identify the high-low risk areas of pneumonia. Figure 4.16 until Figure 4.18 demonstrate the disease maps for relative risk estimation that represent the high-low risk areas of pneumonia in Malaysia based on existing model: SMR, Poisson-gamma model and BYM model, while Figure 4.19 until Figure 4.22 illustrate the pneumonia risk maps based on the four proposed models: SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model stochastic for 13 states and 3 federal territories of Malaysia.

Based on the maps comparison above using these seven methods (SMR, Poisson-gamma model, BYM model, stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model) specifically for epidemiology year 2019, there are obvious differences in the estimated risks especially for the four proposed stochastic models. 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 models

described earlier, there are two models that can be considered as the best models that are better in estimating the relative risk: the stochastic SIR model and the stochastic SVCIR model. Even though the stochastic SIR model offers the best model that fit the real data available, but the stochastic SVCIR model offers a better way of describing the underlying behaviours compared to other six models. Hence, these SVCIR models should be used to predict the risks of pneumonia disease across the 13 states and 3 federal territories of Malaysia for short term forecasting. This method can be applied to others disease which have similar way of transmission of pneumonia disease.

4.9 Summary

Chapter Four present the data analysis and results of relative risk estimation for seven methods used in this study. Putrajaya has been recognized as highest risk area of contracting pneumonia using these seven methods. In the next chapter, conclusion about the findings of these relative risks is provided and some suggestions are proposed.

CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

In this chapter, conclusions are made based on the application of the four alternative methods that were proposed to estimate the relative risk of direct transmission infectious disease which include carrier and vaccine elements. The research focuses on pneumonia disease in Malaysia. Besides, some suggestions are included in this chapter so that further research can be held in the future. Limitation of this study is also included in this chapter.

5.2 Main Findings

Four stochastic models developed in this study are based on the transmission model of the disease. These four stochastic models represent current pneumonia scenario compared to the existing methods as they included other elements such as carrier and vaccine except for the stochastic SIR model that is proposed, is an extending model from basic model for infectious disease which included the birth rate, death rate and also death rate due to pneumonia disease.

Based on the results of relative risk comparison in Table 4.10, when using the proposed stochastic models, the value of the relative risk decreases compared to using SMR method, Poisson-gamma model and BYM model especially when using stochastic SVCIR model. This is because stochastic SVCIR model includes four elements; infectious, carrier, recovered and vaccine elements, while the other two methods; stochastic SIC and stochastic SCIR, just consider infectious and carrier elements, while for SCIR, it includes recovered element too. The existing methods

SMR, Poisson-gamma model, and BYM model and the stochastic SIR model just take into account infectious and recovered elements only. SMR and Poisson-gamma model showed that an area has medium risk, however when using stochastic SVCIR model, it becomes low risk as for example for the state of Perlis. This also can be proven by the maps obtained in Chapter Four as the map of relative risk based on stochastic SVCIR model shows large gap of colour compared to the other maps using the other six methods. Hence, in this study, from the four stochastic models that were proposed, the stochastic SVCIR model has been chosen as the best model to be used in estimating the relative risk since stochastic SVCIR model offers a more detailed description of the biological process which considers extra element compared to the other three proposed models that is vaccine element. Besides that, SVCIR model enables adjustment of covariate and allows for spatial correlation between risks in nearby regions. In addition, from December 2020, data for pneumococcal vaccine coverage will be available at MOH as this vaccine has been included in the National Immunization Program (NIP) in Malaysia.

Result also shows that Putrajaya has the highest risk values for all seven methods. This is because it has the highest annual population growth rate and also the smallest region in Malaysia. From this information, it can be concluded that population size and size of area can influence the risk of susceptible persons in the area contracting disease.

5.3 Conclusion

The primary objective of this research is to introduce alternative method for the estimation of relative risk for pneumonia in order to give a more detailed information regarding the current situation of pneumonia in Malaysia. This study considered some

important features such as the number of population and the transmission process of the disease to improve disease mapping. This study is done by studying the literature regarding the methods in estimating the relative risk of disease mapping and associating these methods to pneumonia transmission models. Several methods related to estimation of relative risk have been discussed in this study and most of them are new models that have been developed and driven by the basic model in relative risk estimation. These models are SMR method, Poisson-gamma model, Besag, York and Mollie (BYM) model, stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model.

In order to have a good disease mapping, it is a must to obtain 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 common models used in finding the relative risk which is the SMR method and the earliest of Bayesian methodology that is Poisson-gamma model. Another existing method is known as Besag, York and Mollie (BYM) model. However, these three methods have drawbacks. SMR cannot identify small areas because it ignores the high diversity of different areas as well as the spatial patterns of the studied areas. The used of SMR can yield large changes in estimate with relatively small changes in expected value since it is based on ratio estimation. SMR is very sensitive, particularly when there is no observed count in specific regions due to the estimated relative risk which becomes exactly zero when there are no observed cases. Meanwhile Poisson-gamma model does not allow the adjustment of the covariate and incapable to deal with spatial correlation between risks in adjacent areas. For BYM, the model has smoothened the relative risk estimation and appear to have lost the jump discontinuities, which may be crucial to maintain. By reviewing the literature, the method for estimating relative risk

for disease mapping could be improved by considering other significant variables in the model. From this information, new approaches have been developed for relative risk estimation based on four stochastic pneumonia transmission model. In detail, the development of models for relative risk estimation based on stochastic pneumonia transmission consists of modeling the stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model, which are then adapted in the relative risk estimation model.

From the map, it can be seen clearly which areas need more attention from the responsible authorities. Apart from that, in term of medical purpose, when the vaccine element is included, it helps to reduce the risk for the susceptible people in that area from getting pneumonia. Hence, this result can support the Ministry of Health Malaysia's decision to include the pneumococcal vaccination in NIP for children which this vaccine is provided for free and is available in public health facilities since 1st December 2020. According to Malaysia Health Director, General Tan Sri Dr. Noor Hisham, pneumococcal vaccine is safe and effective, which is similar with other vaccines used in the National Immunisation Programme Schedule (Bernama, 2020).

For estimating relative risk, the use of posterior expected new infective people to be determined by the fitted models, instead of the observed data used by SMR model. This helps to overcome the problems related with no observed cases reported in some regions and extreme sensitivity in regions with few observed cases. Overall, it can be concluded that from four stochastic models proposed in this study, the more elements that the model consider, the better relative risk estimation values.

5.4 Contribution

The contributions of this research are primarily to the advancement of knowledge on pneumonia disease mapping, particularly for the purpose of estimating relative risk.

5.4.1 Pneumonia Disease Transmission Model

Previous studies considered deterministic models in the analysis, for example studies by Doura et al. (2000) and Mbabazi et al. (2019). However, in this research, stochastic term is introduced into four disease transmission models that are SIC, SIR, SCIR and SVCIR model. For the deterministic and stochastic SIR model used in this research, the model is extended which include birth rate, death rate and death rate due to pneumonia. Moreover, these four stochastic models proposed consider the transmission mechanism of the disease, allowing the spatial correlation between risks in nearby areas and allow for covariate adjustments.

5.4.2 Relative Risk Estimation Method of Pneumonia Disease Mapping

Based on literature review, there is no study in Malaysia focusing on the method to estimate the relative risk for pneumonia disease. Most studies in Malaysia are based on the clinical and case study which does not focusing the observed count data to estimate relative risk, for example studies by Liam (2005), Azmi et al. (2012) and Azmi et al. (2016). Therefore, this research proposed alternative methods to estimate relative risk for pneumonia that are stochastic SIC model, stochastic SIR model, stochastic SCIR model and stochastic SVCIR model.

5.4.3 Application of Relative Risk Estimation of Disease Mapping to Pneumonia Data in Malaysia

Currently, there is no pneumonia map for Malaysia. Current approaches in identifying high-low risk areas or hot spot areas are based on the total number of pneumonia cases across the regions reported. A few studies use the geographic information system (GIS) method in the analysis to identify environmental factors that are related to pneumonia occurrence patterns, to predict potential high-risk areas for a pneumonia outbreak, for example study by Abdul Rahman et al. (2017). Hence, this research introduced pneumonia disease risk maps for Malaysia using the proposed relative risk estimation method based on four stochastic models for pneumonia disease transmission. The findings may assist the government in prioritizing locations which need further attention in gearing towards a sustainable health system, optimizing financial resources for healthcare and strengthening population health.

5.5 Limitation of the Study

In this study, there is a limitation that is pneumococcal vaccination data. This is because MOH does not have the data for pneumococcal vaccine coverage as this type of vaccine before December 2020 only available at private hospitals and clinics. For now, pneumococcal vaccine is included in NIP for babies.

5.6 Recommendations

There are few suggestions for the further research of this study. The first suggestion is to investigate the transmission of pneumonia 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 four stochastic models for direct infectious disease transmission. For future studies, stochastic terms can be assigned to new infective persons from vaccine state to infective state, to study the stochasticity of these variables besides improving the measurement of the relative risk.

Besides, in terms of application of the proposed methodology, these proposed stochastic models can be applied to other direct infectious diseases that have a similar biological transmission as pneumonia disease, such as COVID-19. This leads to further evaluations and a new study on how different types of disease data affect the proposed models.

REFERENCES

Abdul Rahman SR, Ismail, S. N. S., Sahani, M., Ramli, M. F. & Latif, M. T. (2017). A case crossover analysis of primary air pollution association on acute respiratory infection (ARI) among children in urban region of Klang valley, Malaysia. *Annals of Tropical Medicine and Public Health*, *10*(1), 44-55.

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

Ahmad, R. (2019, 152etrieve 15). Pneumonia: Symptoms, Causes, Treatment & Vaccine. *Bristol Medical Center*. 152etrieved from http://bristolmedicalcentre.com/pneumonia/

Alhdiri, M. A., Samat, N. A. and Mohamed, Z. (2017). Risk Estimation for Lung Cancer in Libya: Analysis Based on Standardized Morbidity Ratio, Poisson-Gamma Model, BYM Model and Mixture Model. *Asian Pac Journal Cancer Prev., 18*, 673–679. https://doi.org/10.22034/APJCP.2017.18.3.673

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

Assab, R., Nekkab, N., Crépey, P., Astagneau, P., Guillemot, D., Opatowski, L. and Temime, L. (2017). Mathematical models of infection transmission in healthcare settings: recent advances from the use of network structured data. *Current opinion in infectious diseases, 30*(4), doi: 10.1097/QCO.0000000000000390.

Aswi, A., Cramb, S., Duncan, E. & Mengersen, K. (2020). Evaluating the impact of a small number of areas on spatial estimation. *Int J Health Geogr, 19*, 39. https://doi.org/10.1186/s12942-020-00233-1

Awang, A. C. (2017). The Stochastic SIR-SI Age-Structured Model for Leptospirosis Mapping in Malaysia. MSc thesis, Universiti Pendidikan Sultan Idris, Tanjung Malim. Perak, Malaysia.

Awang, A. C. and Samat, N. A. (2017). Standardized morbidity ratio for leptospirosis mapping in Malaysia. *AIP Conference Proceedings*, *1847*(1), 020006. doi: 10.1063/1.4983861.

Azmi, S., Aziz, SH. A., and Adrian, G. (2012). An Analysis of Reported Pneumonia Deaths in Malaysia. *Value in Health Journal*. *15*(7), 561. Retrieved from https://www.valueinhealthjournal.com/article/S1098-3015(12)03722-9/fulltext

Azmi, S., Aljunid, S. M., Maimaiti, N., Ali, A., Nur, A. M., Rosas-valera, M. De, Encluna, J., Mohamed, R., Wibowo, B., Komaryani, K., and Roberts, C. (2016). Assessing the burden of pneumonia using administrative data from Malaysia, Indonesia, and the Philippines. *International Journal of Infectious Diseases*, *49*, 87–93. https://doi.org/10.1016/j.ijid.2016.05.021.

Bernama. (2020, Nov 24). Dr. Noor Hisham: Pneumococcal vaccine available at public health facilities by December. Malay Mail, Malaysia. Retrieved from https://www.malaymail.com/news/malaysia/2020/11/24/dr-noor-hisham-pneumococcal-vaccine-available-at-public-health-facilities-b/1925697

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.

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*, 1-20.

Centers for Disease Control and Prevention (CDC). (2020, August 7). Pneumococcal Vaccination: What Everyone Should Know. Retrieved on October 20th, 2020, from https://www.cdc.gov/vaccines/vpd/pneumo/public/index.html.

Centers for Disease Control and Prevention (CDC). (2021, June 7). Similarities and Differences between Flu and COVID-19. Retrieved on October 20th, 2021, from https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm.

Choopojcharoen, T. and Ali Magzari. (2012). *Mathematics behind System Dynamics*. Bachelor thesis, Worcester Polytechnic Institute, Worcester, Massachusetts.

Chung, C. (2018, October 6). Pneumococcal vaccination to be mandatory for those under two. The Star Online. Retrieved from https://www.thestar.com.my/news/nation/2018/10/06/pneumococcal-vaccination-to-be-mandatory-for-those-under-two#ltexPlYBk65dWUx4.99

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.

Clayton, D. G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. *Biometrics, 43*, 671-691.

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

Deeth, L. E., Deardon, R., & Gillis, D. J. (2015). Model Choice Using the Deviance Information Criterion for Latent Conditional Individual-Level Models of Infectious Disease Spread. *Epidemiol. Methods 2015*, *4*(1), 47–68. https://doi.org/10.1515/em-2014-0001

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

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

Department of Statistics Malaysia. (2016). *Statistics on Causes of Death, Malaysia, 2014*. Retrieved on August 1st, 2019 from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=eTY2NW00S3BLb1dldWJmVFNMWmphQT09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Department of Statistics Malaysia. (2017). *Statistics on Causes of Death, Malaysia, 2017*. Retrieved on August 1st, 2019 from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=Y3psYUI2VjU0ZzRhZU1kcVFMMThGUT09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Department of Statistics Malaysia. (2018). *Statistics on Causes of Death, Malaysia, 2018*. Retrieved on August 1st, 2019 from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=aWg2VjJkZHhYcDdEM3JQSGloeTVlZz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Department of Statistics Malaysia. (2019). *Statistics on Causes of Death, Malaysia, 2019*. Retrieved on November 11th, 2019, from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=RUxlSDNkcnRVazJnakNCNVN2VGgrdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Department of Statistics Malaysia. (2020). *Statistics on Causes of Death, Malaysia, 2020*. Retrieved on December 18th, 2020, from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=QTU5T0dKQ1g4MHYxd3ZpMzhEMzdRdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09

Department of Survey and Mapping Malaysia. (2017). *Keluasan Malaysia, 2017*. Retrieved from https://www.data.gov.my/data/ms_MY/dataset/keluasan-malaysia/resource/a04e46de-5044-4081-b6cc-7853a46439b

Dey, S., Delampady, M., and Gopalaswamy, A. M. (2019). Bayesian model selection for spatial capture–recapture models. *Ecology and evolution*, *9*(20), 11569-11583. https://doi.org/10.1002/ece3.5551

Diah, I. M. (2017). *Relative Risk Estimation of Tuberculosis Disease Mapping with Stochastics SLIR Models.* Master thesis, Universiti Utara Malaysia, Malaysia.

Doura, K., Melendez-morales, J. D., Meyer, G. G., and Perez, L. E. (2000). An S-I-S Model of Streptococcal Disease with a Class of Beta-Hemolytic Carriers. *Biometric Department*, *MTBI Cornell University Technical Report*, 493–518.

Global Burden of Disease Collaborative Network. (2018). Global Burden of Disease Study 2017 Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME).

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.

Giodano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippo, A. D., Matteo, A. D. and Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, *26*, 855-860. doi: 10.1038/s41591-020-0883-7.

Greenhalgh, D. and Hay, G. (1997). Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users. *IMA Journal of Mathemathics Applied in Medicine and Biology*, *14*(1), 11–38. doi: 10.1093/imammb/14.1.11.

Greenslade, L. (2018). *The Missing Piece: Why continued neglect of pneumonia threatens the achievement of health goals*. JustActions Report. https://stoppneumonia.org/wp-content/uploads/2018/11/The-Missing-Piece_-0611_Spread.pdf

Health Status. (2000). Leeds, Grenville & Lanark District Health Profile. Retrieved on August 5th, 2018, 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. (2000). The mathematics of infectious diseases. *SIAM Review*, *42*(4), 599–653.

Hill, P. C., Townend, J., Antonio, M., Akisanya, B., Ebruke, C., Lahai, G., Greenwood, B. M. and Adegbola, R. A. (2010). Transmission of Streptococcus pneumonia in Rural Gambian Villages: A Longitudinal Study. *Clinical Infectious Diseases*, *50*(11), 1468–1476.

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.

Hussin, K. A. (2019). Malaysia perlu bersedia hadapi status negara tua. *Berita Harian Online*, Nasional. Retrieved from https://www.bharian.com.my/berita/nasional/2019/04/547578/malaysia-perlu bersedia-hadapi-status-negara-tua

Iddrisu, A. K. 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

Ideris, S. H. (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.

Ideris, S. H., Malim, M. R. and Shaadan, N. (2019). A stochastic model for relative risk estimation of leptospirosis in Malaysia. *Journal of Physics: Conference Series*, *1366*(2019). https://doi.org/10.1088/1742-6596/1366/1/012118

Ideris, S. H., Malim, M. R. and Shaadan, N. (2021). Relative Risk Estimation for Human Leptospirosis Disease in Malaysia Based on Existing Models and Discrete Space-Time Stochastic Sir Model. *Journal of Science & Technology, 29*(2), 1071–1095.

International Vaccine Access Center (IVAC). (2020). *Pneumonia and Diarrhea Progress Report 2020*. Johns Hopkins Bloomberg School of Public Health. Retrieved from https://www.jhsph.edu/ivac/wp-content/uploads/2020/11/IVAC_PDPR_2020.pdf.

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.

Johnston, L. (2019, Oct 1). *What fast-ageing countries such as China tell us about our economic future*. Retrieved from https://www.weforum.org/agenda/2019/10/what-ageing-china-and-japan-teach-us-about-our-economies/

Kassa, M., and Murthy, S. N. (2016). Pneumonia Control Measures Under Five Year Children. *IOSR Journal of Mathematics (IOSR-JM), 12*(3), 64–70. https://doi.org/10.9790/5728-1203036470

Keeling, M. J. and Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press.

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

Korycka-Skorupa, J. and Pasławsk, J. (2017). The beginnings of the choropleth presentation, *49*(4), 187–198. https://doi.org/10.1515/pcr-2017-0012.

Kristiani, F., and Samat, N. A. (2019). The mathematical modelling of the SIR-SI contagion model of dengue disease which considers the blood type O factor human compartment: A theoretical simulation A theoretical simulation. *Journal of Statistics and Management Systems, 22*(3), 0510. https://doi.org/10.1080/09720510.2018.1550032

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

Lawson, A. B. (2009). *Bayesian disease mapping: hierarchical modelling in spatial epidemiology*. US: CRC Press.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modelling 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. and Williams, F. (2001). *An introductory guide to disease mapping*. Wet Sussex, UK: John Wiley & Sons, Ltd.

Leino, T., Auranen, K., Jokinen, J., Leinonen, M., Tervonen, P. and Takala, A. K. (2001). Pneumococcal carriage in children during their first two years: Important role of family exposure. *The Pediatric Infectious Disease Journal*, *20*(11), 1022–1027.

Liam, C. K. (2005). Community Acquired Pneumonia - A Malaysian Perspective. *Medical Journal Malaysia, 60*(2), 249-266.

Liang, Y., Greenhalgh, D., and Mao, X. (2016). A Stochastic Differential Equation Model for the Spread of HIV amongst People Who Inject Drugs. *Computational and Mathematical Methods in Medicine, 2016*, 6757928.

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

Lynn, N. (2020, Nov 12). *World Pneumonia Day 2020*. Retrieved from https://www.asbmb.org/asbmb-today/science/111220/world-pneumonia-day-2020

Mbabazi, F. K., Mugisha, J. Y. T., and Kimathi, M. (2019). Hopf-Bifurcation Analysis of Pneumococcal Pneumonia with Time Delays. *Abstract and Applied Analysis, 2019*(January). https://doi.org/10.1155/2019/3757036

Mcbryde, E. (2006). Mathematical and Statistical Modelling of Infectious Diseases in Hospitals. *Sciences-New York*, *10*(November), 777–782. https://doi.org/10.1093/nar/gkr403

Melegaro, A., Gay, N. J., and Medley, G. F. (2004). Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiology and Infection, 132*(3), 433–441. https://doi.org/10.1017/S0950268804001980

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

Ming, R., Liu, J., Cheung, W.K.W. and Wan, X. (2016). Stochastic modelling of infectious diseases for heterogeneous populations. *Infect Dis Poverty*, *5*, 107. https://doi.org/10.1186/s40249-016-0199-5

Ministry of Health Malaysia. (2019). *Laporan Tahunan Kementerian Kesihatan Malaysia 2019* (ISSN 1511-1520). https://www.moh.gov.my/moh/resources/Penerbitan/Penerbitan%20Utama/ANNUAL%20REPORT/LAPORAN%20TAHUNAN%20KKM%202019/mobile/index.html#p=1

Mohammed Kizito and Tumwiine, J. (2018). A Mathematical Model of Treatment and Vaccination Interventions of Pneumococcal Pneumonia Infection Dynamics. *Journal of Applied Mathematics*. https://doi.org/10.1155/2018/2539465

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

Ndelwa, E. J., Kgosimore, M., Massawe, E. S., and Namkinga, L. (2015). Mathematical Modelling and Analysis of Treatment and Screening of Pneumonia, *Mathematical Theory and Modeling*, *5*(10), 21–40.

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

Nor, N. M. (2015, July 20). *Population Ageing: Can We "Live Long and Prosper"*. Retrieved from http://www.krinstitute.org/What_We_Are_Reading-@-Population_Ageing_Can_We_Live_Long_and_Prosper.aspx

Normandin, B. (2021, November 10). *Everything you need to know about pneumonia* Healthline. Retrieved January 7, 2022, from https://www.healthline.com/health/pneumonia

Ntzoufras, I. (2009). *Bayesian Modeling using WinBUGS*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Rahman, N. A. A., Ismail, W. R. W., Rahman, R. A., Yusof, M. P. and Idris, I. B. (2020). Who Would Get Measles in Petaling District? A Trend Analysis of Measles Outbreak from 2014-2018. *Malaysian Journal of Medicine and Health Sciences*. *16*(3), 67-72. Retrieved from https://medic.upm.edu.my/upload/dokumen/2020090809444711_MJMHS_0048.pdf.

Otieno, O. J., Joseph, M., and John, O. (2012). Mathematical Model for Pneumonia Dynamics among Children. *The 2012 Southern Africa Mathematical Sciences Association Conference*, (Samsa), 26–29. Retrieved from https://su-plus.strathmore.edu/bitstream/handle/11071/3615/Mathematical Model for Pneumonia Dynamics paper.pdf?sequence=1&isAllowed=y

Peng, R. D. (2018). *Advanced Statistical Computing*. Bookdown. Retrieved from https://leanpub.com/advstatcomp/

Pooley, C. M., and Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society open sci. 5*:171519. http://dx.doi.org/10.1098/rsos.171519.

Riebler, A., Sørbye, S. H., Simpson, D., Rue, H., Lawson, A. B., Lee, D., & MacNab, Y. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, *25*(4), 1145–1165. https://doi.org/10.1177/0962280216660421

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

Samat, N. A. (2012). Mathematical Models for Vector-Borne nfectious Disease Mapping with Application to Dengue Disease in Malaysia. PhD thesis, University of Salford, Manchester, UK.

Samat, N. A. and Mey, L. W. (2017). Malaria Disease Mapping in Malaysia based on Besag-York-Mollie (BYM) Model. *Journal of Physics: Conference Series*, *890*(1). https://doi.org/10.1088/1742-6596/890/1/012167

Samat, N. A. and 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

Samat, N. A. and Ma'arof, S. H. M. I. (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.

Shaddick, G. (2008). Spatial Epidemiology. Retrieved on 2019, August 10 from https://www.stat.ubc.ca/~gavin/SpatEpiPart2.pdf

Smith, T., Lehmann, D., Montgomery, J., Gratten, M., Riley, I. D., and Alpers, M. P. (1993). Acquisition and invasiveness of different serotypes of Streptococcus pneumoniae in young children. *Epidemiology and Infection*, *111*(1), 27–39. https://doi.org/10.1017/S0950268800056648

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.

Soliman, B. W. M., & Bueno, A. C. F. (2018). Modelling the spread of pneumonia in the Philippines using Susceptible-Infected-Recovered (SIR) model with demographic changes. *Journal of Technology Management and Business*, *5*(1), 28–35.

Spiegelhalter, D. J., Best N. G., Carlin B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Royal Statistical Society B. 64,* 583–616. doi:10.1111/1467-9868.00353

Spiegelhalter D. J., Best N. G., Carlin B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal Royal Statistics Society B. 76*, 485–493. doi:10.1111/rssb.12062

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.

Stop pneumonia. (2020, November 12). *Championing the fight against pneumonia.* Retrieved on May 31st, 2021, from https://stoppneumonia.org/latest/world-pneumonia-day/

Sun, D., Tsutakawa, R., & Speckman, P. (1999). Posterior Distribution of Hierarchical Models Using CAR (1) Distributions. *Biometrika, 86*(2), 341-350. Retrieved June 27, 2021, from http://www.jstor.org/stable/2673516

Thomas, D. S. K., Anthamatten, P., Root, E. D., Lucero, M., Nohynek, H., Tallo, V., & Maleckar, J. (2015). Disease mapping for informing targeted health interventions: childhood pneumonia in Bohol, Philippines. *Tropical Medicine and International Health*, *20*(11), 1525–1533. https://doi.org/10.1111/tmi.12561

Tilahun, G. T., Makinde, O. D., and Malonza, D. (2017). Modelling and optimal control of pneumonia disease with cost-effective strategies. *Journal of Biological Dynamics*, *11*, 400–426. https://doi.org/10.1080/17513758.2017.1337245

UNICEF. (2016). *One is too many: Ending child deaths from pneumonia and diarrhea,* UNICEF, New York.

UNICEF. (2018). *Pneumonia: Pneumonia claims the lives of the world's most vulnerable children.* Retrieved on August 5th, 2018 from https://data.unicef.org/topic/child-health/pneumonia/

UNICEF. (2019). *Pneumonia: A child dies of pneumonia every 39 seconds.* Retrieved on November 11th, 2019, from https://data.unicef.org/topic/child-health/pneumonia/

Vyambwera, S. M., and Witbooi, P. (2018). A Stochastic TB Model for a Crowded Environment. *Journal of Applied Mathematics, 2018*. https://doi.org/10.1155/2018/3420528

Waller, L. A., and Carlin, B. P. (2010). Disease mapping. *Chapman & Hall/CRC handbooks of modern statistical methods*, *2010*, 217–243. https://doi.org/10.1201/9781420072884-c14.

Wang, S. (2019). *How population ageing affects technological innovation in perspective of human capital*. Master thesis. Retrieved from http://www.diva-portal.org/smash/get/diva2:1354181/FULLTEXT01.pdf.

Wang, Y., Jiang, D., Hayat, T. and Ahmad, B. (2017). A stochastic HIV infection model with T-cell proliferation and CTL immune response. *Applied Mathematics and Computation*, *315*,477-493. https://doi.org/10.1016/j.amc.2017.07.062.

Wardlaw, T., Johansson, E. W., and Hodge, M. (2006). Pneumonia: The Forgotten Killer of Children. WHO/UNICEF. Retrieved from http://www.who.int/maternal_child_adolescent/documents/9280640489/en/

World Health Organization. (2016).*Pneumonia Fact Sheet*. Retrieved on July 28th, 2018, from http://www.who.int/news-room/fact-sheets/detail/pneumonia

World Health Organization. (2018).*Pneumonia Fact Sheet*. Retrieved on August 18th, 2018, from http://www.who.int/news-room/fact-sheets/detail/pneumonia

World Health Organization. (2021).*Pneumonia Fact Sheet*. Retrieved on October 11th, 2021, from http://www.who.int/news-room/fact-sheets/detail/pneumonia

World Health Organization. (2020a).*Global research on coronavirus disease (COVID-19)*. Retrieved on May 28th, 2020, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov

World Health Organization. (2020b). *Pneumonia*. Retrieved on December 9th, 2020, from https://www.who.int/health-topics/pneumonia#tab=tab_1

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=4259

Appendix A
Knowledge Dissemination

1) Diah, I. M. (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) Diah, I. M., Aziz, N. and Ahmad, N. (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) Diah, I. M., Aziz, N., Ahmad, N. & Kasim, M. M. (2016). Tuberculosis disease mapping with stochastic equation. *IACE' 2016- Proceeding of the 3rd Innovation and Analytics Conference & Exhibition, 77-82.*

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

5) Diah, I. M., Aziz, N., & Kasim, M. M. (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) Diah, I. M., Aziz, N., & Kasim, M. M. (2017). Tuberculosis disease mapping in Kedah using standardized morbidity ratio. Presentation Session in *ICAST' 2017- Proceeding of the 2nd International Conference on Applied Science and Technology,* 10th April – 12th 2017.

7) Diah, I. M., Aziz, N., & Kasim, M. M. (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.

8) Diah, I. M., & Aziz, N. (2019). Stochastic modelling for pneumonia incidence: A conceptual framework. *The 4Th Innovation and Analytics Conference & Exhibition (IACE 2019)*, *2138*(August), 050010. https://doi.org/10.1063/1.5121115

9) Diah, I. M., & Aziz, N. (2019). Stochastic modelling for pneumonia incidence: A conceptual framework. Presentation Session in IACE 2019- *The 4Th Innovation and Analytics Conference & Exhibition,* 25th – 28th Mac 2019.

10) Diah, I. M., & Aziz, N (2021). The Relative Risk Estimation of Pneumonia in Malaysia using Standardized Morbidity Ratio (SMR). *ASM Science Journal. 16,* 1-5.

11) Diah, I. M., & Aziz, N. (2021). Mapping of Pneumonia Disease in Malaysia Using Poisson-Gamma Model. *Annals of Roman Society for Cell Biology. 25*(1), 2062-2067.

Appendix B

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

AB-1-1: Summary Statistics for the State of Perlis

node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	0.5879	0.004957	6.174E-5	0.5784	0.588	0.5973
RRH[1,3]	3.369	0.0284	3.538E-4	3.314	3.369	3.423
RRH[1,4]	2.084	0.01757	2.189E-4	2.05	2.084	2.117
RRH[1,5]	1.856	0.01565	1.949E-4	1.826	1.856	1.886
RRH[1,6]	1.55	0.01307	1.628E-4	1.525	1.55	1.575
RRH[1,7]	1.603	0.01351	1.683E-4	1.577	1.603	1.628
RRH[1,8]	1.606	0.01354	1.687E-4	1.58	1.606	1.632
RRH[1,9]	1.424	0.01201	1.495E-4	1.401	1.424	1.447
RRH[1,10]	1.53	0.0129	1.607E-4	1.505	1.53	1.555

AB-1-2: Summary Statistics for the State of Kedah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[2,2]	0.4148	0.001757	1.831E-5	0.4118	0.4148	0.418
RRH[2,3]	2.391	0.01013	1.055E-4	2.373	2.391	2.409
RRH[2,4]	1.433	0.006071	6.328E-5	1.423	1.433	1.444
RRH[2,5]	1.237	0.005239	5.461E-5	1.228	1.237	1.247
RRH[2,6]	1.068	0.004521	4.713E-5	1.06	1.068	1.076
RRH[2,7]	1.08	0.004574	4.768E-5	1.072	1.08	1.088
RRH[2,8]	1.055	0.004467	4.657E-5	1.047	1.055	1.063
RRH[2,9]	0.8828	0.003739	3.897E-5	0.8763	0.8828	0.8895
RRH[2,10]	0.9801	0.004151	4.327E-5	0.9729	0.9801	0.9876

AB-1-3: Summary Statistics for the State of Pulau Pinang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[3,2]	0.3141	0.001579	1.719E-5	0.3113	0.3141	0.317
RRH[3,3]	1.893	0.009517	1.036E-4	1.876	1.893	1.911
RRH[3,4]	1.008	0.005066	5.517E-5	0.9988	1.008	1.017
RRH[3,5]	0.9538	0.004794	5.221E-5	0.9452	0.9538	0.9626
RRH[3,6]	0.782	0.003931	4.281E-5	0.775	0.782	0.7892
RRH[3,7]	0.7918	0.00398	4.334E-5	0.7847	0.7918	0.7991
RRH[3,8]	0.7825	0.003933	4.283E-5	0.7755	0.7825	0.7897
RRH[3,9]	0.6408	0.003221	3.508E-5	0.635	0.6408	0.6467
RRH[3,10]	0.6939	0.003488	3.799E-5	0.6877	0.6939	0.7003

AB-1-4: Summary Statistics for the State of Perak

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[4,2]	0.4192	0.001713	1.873E-5	0.4162	0.4192	0.4222
RRH[4,3]	2.323	0.009497	1.038E-4	2.307	2.323	2.34
RRH[4,4]	1.391	0.005686	6.215E-5	1.381	1.391	1.401
RRH[4,5]	1.268	0.005183	5.665E-5	1.259	1.268	1.277
RRH[4,6]	1.075	0.004394	4.803E-5	1.067	1.075	1.083
RRH[4,7]	1.086	0.004439	4.852E-5	1.078	1.086	1.094
RRH[4,8]	1.083	0.004427	4.839E-5	1.075	1.083	1.091
RRH[4,9]	0.9096	0.003718	4.064E-5	0.9032	0.9096	0.9161
RRH[4,10]	0.9925	0.004057	4.435E-5	0.9856	0.9925	0.9969

AB-1-5: Summary Statistics for Kuala Lumpur

node	mean	Sd	MC error	2.5%	median	97.5%
RRH[5,2]	0.2411	0.001368	1.561E-5	0.2385	0.2411	0.2436
RRH[5,3]	1.357	0.007698	8.784E-5	1.342	1.357	1.371
RRH[5,4]	0.7617	0.004322	4.932E-5	0.7538	0.7617	0.7698
RRH[5,5]	0.6973	0.003957	4.515E-5	0.69	0.6973	0.7047
RRH[5,6]	0.5724	0.003248	3.706E-5	0.5664	0.5724	0.5785
RRH[5,7]	0.5692	0.00323	3.686E-5	0.5633	0.5692	0.5753
RRH[5,8]	0.552	0.003132	3.574E-5	0.5462	0.552	0.5578
RRH[5,9]	0.4859	0.002757	3.146E-5	0.4808	0.4859	0.491
RRH[5,10]	0.5187	0.002943	3.359E-5	0.5133	0.5187	0.5242

AB-1-6: Summary Statistics for Putrajaya

node	mean	sd	MC error	2.5%	median	97.5%
RRH[6,2]	1.08	0.01048	1.192E-4	1.06	1.08	1.101
RRH[6,3]	6.507	0.06309	7.18E-4	6.386	6.507	6.63
RRH[6,4]	3.401	0.03298	3.753E-4	3.338	3.401	3.465
RRH[6,5]	3.757	0.03643	4.146E-4	3.688	3.757	3.828
RRH[6,6]	3.378	0.03275	3.727E-4	3.315	3.378	3.441
RRH[6,7]	3.443	0.03338	3.799E-4	3.379	3.443	3.508
RRH[6,8]	3.893	0.03775	4.296E-4	3.821	3.893	3.967
RRH[6,9]	2.866	0.02779	3.163E-4	2.813	2.866	2.92
RRH[6,10]	3.775	0.03661	4.166E-4	3.705	3.775	3.846

AB-1-7: Summary Statistics for the State of Selangor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[7,2]	0.2336	8.894E-4	1.002E-5	0.2321	0.2336	0.2351
RRH[7,3]	1.279	0.004869	5.483E-5	1.271	1.279	1.287
RRH[7,4]	0.753	0.002867	3.228E-5	0.7481	0.753	0.7577
RRH[7,5]	0.6679	0.002543	2.863E-5	0.6635	0.6679	0.6721
RRH[7,6]	0.5585	0.002126	2.394E-5	0.5549	0.5585	0.562
RRH[7,7]	0.5566	0.002119	2.386E-5	0.553	0.5566	0.5602
RRH[7,8]	0.5466	0.002081	2.343E-5	0.543	0.5466	0.55
RRH[7,9]	0.4568	0.001739	1.959E-5	0.4539	0.4568	0.4597
RRH[7,10]	0.4932	0.001878	2.115E-5	0.4539	0.4932	0.4964

AB-1-8: Summary Statistics for the State of Negeri Sembilan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[8,2]	0.5696	0.002686	2.861E-5	0.5649	0.5696	0.5745
RRH[8,3]	3.309	0.0156	1.662E-4	3.282	3.309	3.337
RRH[8,4]	1.969	0.009286	9.891E-5	3.282	1.969	1.986
RRH[8,5]	1.799	0.009286	9.036E-5	1.784	1.799	1.815
RRH[8,6]	1.522	0.007176	7.643E-5	1.509	1.522	1.535
RRH[8,7]	1.551	0.007312	7.788E-5	1.538	1.551	1.564
RRH[8,8]	1.556	0.007337	7.815E-5	1.543	1.556	1.569
RRH[8,9]	1.341	0.006323	6.735E-5	1.33	1.341	1.352
RRH[8,10]	1.463	0.006901	7.35E-5	1.451	1.463	1.476

AB-1-9: Summary Statistics for the State of Melaka

node	mean	sd	MC error	2.5%	median	97.5%
RRH[9,2]	0.4795	0.00264	3.204E-5	0.4746	0.4795	0.4843
RRH[9,3]	2.748	0.01513	1.836E-4	2.72	2.748	2.776
RRH[9,4]	1.6	0.008811	1.069E-4	1.584	1.6	1.616
RRH[9,5]	1.496	0.008237	9.996E-5	1.481	1.496	1.511
RRH[9,6]	1.254	0.006908	8.383E-5	1.242	1.255	1.267
RRH[9,7]	1.262	0.006948	8.432E-5	1.249	1.262	1.275
RRH[9,8]	1.271	0.006997	8.491E-5	1.258	1.271	1.284
RRH[9,9]	1.08	0.005946	7.216E-5	1.258	1.08	1.091
RRH[9,10]	1.175	0.006469	7.85E-5	1.163	1.175	1.187

AB-1-10: Summary Statistics for the State of Johor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[10,2]	0.4431	0.001604	1.708E-5	0.4405	0.4431	0.4458
RRH[10,3]	2.553	0.009241	9.839E-5	2.538	2.553	2.569
RRH[10,4]	1.509	0.005463	5.816E-5	1.5	1.509	1.518
RRH[10,5]	1.351	0.004892	5.208E-5	1.344	1.351	1.36
RRH[10,6]	1.145	0.004144	4.412E-5	1.138	1.145	1.152
RRH[10,7]	1.157	0.00419	4.461E-5	1.151	1.157	1.165
RRH[10,8]	1.162	0.004206	4.478E-5	1.155	1.162	1.169
RRH[10,9]	0.9602	0.003476	3.701E-5	0.9547	0.9602	0.9662
RRH[10,10]	1.062	0.003844	4.092E-5	1.056	1.062	1.068

AB-1-11: Summary Statistics for the State of Pahang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[11,2]	0.5191	0.002232	2.185E-5	0.5153	0.5191	0.523
RRH[11,3]	2.987	0.01284	1.257E-4	2.964	2.987	3.009
RRH[11,4]	1.755	0.007547	7.387E-5	1.742	1.755	1.768
RRH[11,5]	1.605	0.006901	6.755E-5	1.593	1.605	1.617
RRH[11,6]	1.363	0.005861	5.737E-5	1.353	1.363	1.373
RRH[11,7]	1.388	0.005966	5.839E-5	1.377	1.388	1.398
RRH[11,8]	1.385	0.005956	5.83E-5	1.375	1.385	1.396
RRH[11,9]	1.177	0.005062	4.955E-5	1.169	1.177	1.186
RRH[11,10]	1.308	0.005624	5.505E-5	1.298	1.308	1.318

AB-1-12: Summary Statistics for the State of Terengganu

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[12,2]	0.5701	0.002621	3.021E-5	0.5655	0.5701	0.5747
RRH[12,3]	3.376	0.01552	1.789E-4	3.349	3.376	3.404
RRH[12,4]	1.946	0.008948	1.032E-4	1.931	1.946	1.962
RRH[12,5]	1.808	0.008312	9.582E-5	1.793	1.808	1.823
RRH[12,6]	1.528	0.007022	8.096E-5	1.515	1.528	1.54
RRH[12,7]	1.557	0.00716	8.254E-5	1.545	1.557	1.57
RRH[12,8]	1.592	0.00732	8.439E-5	1.579	1.592	1.605
RRH[12,9]	1.285	0.005908	6.812E-5	1.275	1.285	1.296
RRH[12,10]	1.485	0.006828	7.872E-5	1.473	1.485	1.497

AB-1-13: Summary Statistics for the State of Kelantan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[13,2]	0.4471	0.00199	2.065E-5	0.4436	0.4471	0.4506
RRH[13,3]	2.509	0.01117	1.159E-4	2.49	2.509	2.529
RRH[13,4]	1.481	0.006593	6.841E-5	1.47	1.481	1.493
RRH[13,5]	1.343	0.005976	6.201E-5	1.332	1.343	1.353
RRH[13,6]	1.124	0.005003	5.191E-5	1.115	1.124	1.133
RRH[13,7]	1.136	0.005059	5.249E-5	1.128	1.136	1.145
RRH[13,8]	1.153	0.005132	5.325E-5	1.144	1.153	1.162
RRH[13,9]	0.9657	0.004299	4.461E-5	0.9582	0.9657	0.9734
RRH[13,10]	1.071	0.004768	4.947E-5	1.063	1.071	1.08

AB-1-14: Summary Statistics for the State of Sabah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[14,2]	0.3437	0.001349	1.355E-5	0.3414	0.3437	0.346
RRH[14,3]	1.939	0.00761	7.647E-5	1.927	1.939	1.952
RRH[14,4]	1.122	0.004404	4.425E-5	1.115	1.122	1.13
RRH[14,5]	1.015	0.003984	4.003E-5	1.009	1.015	1.022
RRH[14,6]	0.8555	0.003357	3.373E-5	0.8499	0.8555	0.8613
RRH[14,7]	0.85	0.003335	3.352E-5	0.8445	0.85	0.8558
RRH[14,8]	0.8534	0.003349	3.365E-5	0.8478	0.8534	0.8592
RRH[14,9]	0.7065	0.002772	2.786E-5	0.7019	0.7065	0.7113
RRH[14,10]	0.7754	0.003043	3.057E-5	0.7703	0.7754	0.7806

AB-1-15: Summary Statistics for Labuan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[15,2]	0.4351	0.001016	1.14E-5	0.434	0.4351	0.4362
RRH[15,3]	2.434	0.005684	6.376E-5	2.427	2.434	2.44
RRH[15,4]	1.448	0.003383	3.795E-5	1.445	1.448	1.452
RRH[15,5]	1.292	0.003018	3.385E-5	1.289	1.292	1.295
RRH[15,6]	1.075	0.002512	2.818E-5	1.073	1.075	1.078
RRH[15,7]	1.087	0.002539	2.848E-5	1.084	1.087	1.09
RRH[15,8]	1.115	0.002605	2.922E-5	1.113	1.115	1.118
RRH[15,9]	0.9476	0.002213	2.483E-5	0.9453	0.9476	0.9501
RRH[15,10]	1.045	0.00244	2.737E-5	1.042	1.045	1.047

AB-1-16: Summary Statistics for the State of Sarawak

node	mean	sd	MC error	2.5%	median	97.5%
RRH[16,2]	0.3614	0.001529	1.525E-5	0.3588	0.3614	0.3641
RRH[16,3]	1.993	0.008431	8.405E-5	1.978	1.993	2.007
RRH[16,4]	1.199	0.005075	5.059E-5	1.191	1.199	1.208
RRH[16,5]	1.079	0.004566	4.552E-5	1.071	1.079	1.087
RRH[16,6]	0.9061	0.003834	3.822E-5	0.8995	0.9061	0.9128
RRH[16,7]	0.9105	0.003853	3.841E-5	0.9039	0.9105	0.9172
RRH[16,8]	0.8974	0.003797	3.786E-5	0.8909	0.8974	0.904
RRH[16,9]	0.7504	0.003175	3.165E-5	0.7449	0.7504	0.7559
RRH[16,10]	0.8271	0.0035	3.489E-5	0.8211	0.8271	0.8332

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

AB-2-1: Summary Statistics for the State of Perlis

node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	2.169	0.01857	2.374E-4	2.134	2.169	2.204
RRH[1,3]	2.885	0.02471	3.159E-4	2.839	2.885	2.932
RRH[1,4]	2.248	0.01925	2.462E-4	2.212	2.248	2.285
RRH[1,5]	1.796	0.01538	1.967E-4	1.767	1.796	1.825
RRH[1,6]	1.524	0.01305	1.669E-4	1.499	1.524	1.549
RRH[1,7]	1.48	0.01267	1.621E-4	1.456	1.48	1.504
RRH[1,8]	1.448	0.0124	1.586E-4	1.425	1.448	1.472
RRH[1,9]	1.185	0.01015	1.298E-4	1.166	1.185	1.205
RRH[1,10]	1.255	0.01075	1.375E-4	1.235	1.255	1.276

AB-2-2: Summary Statistics for the State of Kedah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[2,2]	1.438	0.006415	6.69E-5	1.428	1.438	1.449
RRH[2,3]	1.916	0.008545	8.91E-5	1.902	1.916	1.93
RRH[2,4]	1.492	0.006653	6.937E-5	1.481	1.492	1.503
RRH[2,5]	1.197	0.005338	5.566E-5	1.188	1.197	1.206
RRH[2,6]	1.017	0.004535	4.729E-5	1.009	1.017	1.025
RRH[2,7]	0.9905	0.004418	4.607E-5	0.9831	0.9905	0.9979
RRH[2,8]	0.9739	0.004344	4.53E-5	0.9668	0.9739	0.9813
RRH[2,9]	0.7995	0.003566	3.719E-5	0.7936	0.7995	0.8056
RRH[2,10]	0.8486	0.003785	3.947E-5	0.8424	0.8486	0.855

AB-2-3: Summary Statistics for the State of Pulau Pinang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[3,2]	1.058	0.005516	6.308E-5	1.049	1.058	1.068
RRH[3,3]	1.414	0.00737	8.428E-5	1.401	1.414	1.427
RRH[3,4]	1.104	0.005755	6.581E-5	1.094	1.104	1.114
RRH[3,5]	0.8843	0.004609	5.271E-5	0.8764	0.8843	0.8925
RRH[3,6]	0.7525	0.003922	4.485E-5	0.7457	0.7525	0.7594
RRH[3,7]	0.7332	0.003821	4.37E-5	0.7266	0.7332	0.7399
RRH[3,8]	0.7218	0.003762	4.302E-5	0.7153	0.7218	0.7284
RRH[3,9]	0.5939	0.003096	3.54E-5	0.5886	0.594	0.5994
RRH[3,10]	0.632	0.003294	3.767E-5	0.6263	0.6321	0.6378

AB-2-4: Summary Statistics for the State of Perak

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[4,2]	1.448	0.006267	6.775E-5	1.438	1.448	1.459
RRH[4,3]	1.932	0.008363	9.041E-5	1.919	1.932	1.946
RRH[4,4]	1.507	0.00652	7.049E-5	1.496	1.507	1.517
RRH[4,5]	1.204	0.005211	5.633E-5	1.196	1.204	1.213
RRH[4,6]	1.023	0.004427	4.786E-5	1.016	1.023	1.03
RRH[4,7]	0.9951	0.004306	4.656E-5	0.9882	0.9951	1.002
RRH[4,8]	0.9774	0.00423	4.573E-5	0.9705	0.9774	0.9844
RRH[4,9]	0.8022	0.003472	3.753E-5	0.7966	0.8022	0.808
RRH[4,10]	0.8523	0.003688	3.987E-5	0.8463	0.8523	0.8584

AB-2-5: Summary Statistics for Kuala Lumpur

node	mean	Sd	MC error	2.5%	median	97.5%
RRH[5,2]	0.7742	0.004511	5.099E-5	0.7661	0.7742	0.7823
RRH[5,3]	1.035	0.006031	6.816E-5	1.024	1.035	1.046
RRH[5,4]	0.8089	0.004714	5.327E-5	0.8005	0.8089	0.8175
RRH[5,5]	0.6483	0.003778	4.27E-5	0.6416	0.6483	0.6552
RRH[5,6]	0.5525	0.003219	3.639E-5	0.5468	0.5525	0.5583
RRH[5,7]	0.5394	0.003143	3.553E-5	0.5339	0.5394	0.5451
RRH[5,8]	0.5302	0.00309	3.492E-5	0.5247	0.5302	0.5358
RRH[5,9]	0.4357	0.002539	2.869E-5	0.4311	0.4357	0.4403
RRH[5,10]	0.4637	0.002702	3.054E-5	0.4589	0.4637	0.4686

AB-2-6: Summary Statistics for Putrajaya

node	mean	sd	MC error	2.5%	median	97.5%
RRH[6,2]	4.658	0.04553	5.14E-4	4.572	4.658	4.746
RRH[6,3]	6.22	0.0608	6.863E-4	6.105	6.22	6.337
RRH[6,4]	4.831	0.04722	5.33E-4	4.742	4.831	4.922
RRH[6,5]	3.822	0.03736	4.217E-4	3.751	3.822	3.894
RRH[6,6]	3.221	0.03149	3.554E-4	3.162	3.221	3.282
RRH[6,7]	3.083	0.03014	3.402E-4	3.027	3.083	3.141
RRH[6,8]	3.012	0.02944	3.323E-4	2.956	3.012	3.068
RRH[6,9]	2.43	0.02375	2.681E-4	2.385	2.43	2.476
RRH[6,10]	2.534	0.02477	2.796E-4	2.487	2.534	2.581

AB-2-7: Summary Statistics for the State of Selangor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[7,2]	0.7473	0.003032	3.45E-5	0.7425	0.7473	0.7521
RRH[7,3]	0.9983	0.00405	4.609E-5	0.9919	0.9983	1.005
RRH[7,4]	0.7802	0.003165	3.602E-5	0.7752	0.7802	0.7852
RRH[7,5]	0.6251	0.002536	2.886E-5	0.6211	0.6251	0.6291
RRH[7,6]	0.5324	0.00216	2.458E-5	0.5289	0.5324	0.5358
RRH[7,7]	0.5193	0.002107	2.397E-5	0.516	0.5193	0.5226
RRH[7,8]	0.5112	0.002074	2.36E-5	0.5079	0.5112	0.5144
RRH[7,9]	0.4206	0.001706	1.942E-5	0.4179	0.4206	0.4232
RRH[7,10]	0.4472	0.001814	2.065E-5	0.4444	0.4472	0.4501

AB-2-8: Summary Statistics for the State of Negeri Sembilan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[8,2]	2.088	0.01027	1.115E-4	2.071	2.088	2.106
RRH[8,3]	2.782	0.01368	1.485E-4	2.76	2.782	2.806
RRH[8,4]	2.167	0.01066	1.157E-4	2.149	2.167	2.186
RRH[8,5]	1.73	0.008509	9.237E-5	1.716	1.73	1.745
RRH[8,6]	1.468	0.00722	7.838E-5	1.456	1.468	1.481
RRH[8,7]	1.427	0.007016	7.616E-5	1.415	1.427	1.439
RRH[8,8]	1.398	0.006874	7.462E-5	1.386	1.398	1.41
RRH[8,9]	1.145	0.005629	6.111E-5	1.135	1.145	1.155
RRH[8,10]	1.213	0.005962	6.473E-5	1.203	1.212	1.223

AB-2-9: Summary Statistics for the State of Melaka

node	mean	sd	MC error	2.5%	median	97.5%
RRH[9,2]	1.702	0.009656	1.165E-4	1.685	1.702	1.719
RRH[9,3]	2.27	0.01288	1.554E-4	2.247	2.27	2.293
RRH[9,4]	1.769	0.01004	1.211E-4	1.751	1.769	1.787
RRH[9,5]	1.413	0.008018	9.675E-5	1.399	1.413	1.428
RRH[9,6]	1.201	0.006811	8.219E-5	1.188	1.201	1.213
RRH[9,7]	1.167	0.006624	7.993E-5	1.156	1.168	1.179
RRH[9,8]	1.145	0.006496	7.839E-5	1.133	1.145	1.157
RRH[9,9]	0.9392	0.005329	6.43E-5	0.9297	0.9393	0.9487
RRH[9,10]	0.9971	0.005657	6.826E-5	0.987	0.9972	1.007

AB-2-10: Summary Statistics for the State of Johor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[10,2]	1.553	0.00603	6.393E-5	1.544	1.553	1.562
RRH[10,3]	2.07	0.008037	8.522E-5	2.058	2.07	2.083
RRH[10,4]	1.614	0.006269	6.647E-5	1.605	1.614	1.624
RRH[10,5]	1.291	0.005013	5.315E-5	1.283	1.291	1.299
RRH[10,6]	1.097	0.004259	4.516E-5	1.091	1.097	1.104
RRH[10,7]	1.067	0.004143	4.393E-5	1.061	1.067	1.073
RRH[10,8]	1.048	0.004071	4.317E-5	1.042	1.048	1.055
RRH[10,9]	0.8605	0.003342	3.543E-5	0.8555	0.8605	0.8659
RRH[10,10]	0.9134	0.003547	3.761E-5	0.9081	0.9134	0.919

AB-2-11: Summary Statistics for the State of Pahang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[11,2]	1.86	0.008421	8.58E-5	1.846	1.86	1.874
RRH[11,3]	2.48	0.01123	1.144E-4	2.462	2.48	2.499
RRH[11,4]	1.934	0.008753	8.918E-5	1.919	1.934	1.948
RRH[11,5]	1.545	0.006992	7.124E-5	1.533	1.545	1.556
RRH[11,6]	1.311	0.005935	6.047E-5	1.301	1.311	1.321
RRH[11,7]	1.275	0.00577	5.879E-5	1.265	1.275	1.284
RRH[11,8]	1.25	0.00566	5.767E-5	1.241	1.25	1.26
RRH[11,9]	1.024	0.004636	4.723E-5	1.016	1.024	1.032
RRH[11,10]	1.086	0.004915	5.008E-5	1.078	1.086	1.094

AB-2-12: Summary Statistics for the State of Terengganu

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[12,2]	2.093	0.01007	1.158E-4	2.076	2.093	2.11
RRH[12,3]	2.789	0.01343	1.543E-4	2.766	2.789	2.811
RRH[12,4]	2.172	0.01046	1.202E-4	2.155	2.172	2.19
RRH[12,5]	1.735	0.008353	9.599E-5	1.721	1.735	1.749
RRH[12,6]	1.472	0.007086	8.144E-5	1.46	1.472	1.484
RRH[12,7]	1.429	0.006878	7.905E-5	1.417	1.429	1.44
RRH[12,8]	1.403	0.006756	7.764E-5	1.392	1.404	1.415
RRH[12,9]	1.149	0.005531	6.357E-5	1.14	1.149	1.158
RRH[12,10]	1.216	0.005855	6.729E-5	1.206	1.216	1.226

AB-2-13: Summary Statistics for the State of Kelantan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[13,2]	1.538	0.007176	7.462E-5	1.526	1.538	1.55
RRH[13,3]	2.053	0.009576	9.958E-5	2.037	2.053	2.069
RRH[13,4]	1.602	0.007471	7.769E-5	1.589	1.602	1.614
RRH[13,5]	1.281	0.005977	6.215E-5	1.271	1.281	1.291
RRH[13,6]	1.089	0.005082	5.284E-5	1.081	1.089	1.098
RRH[13,7]	1.059	0.004942	5.139E-5	1.051	1.059	1.068
RRH[13,8]	1.04	0.004852	5.046E-5	1.032	1.04	1.048
RRH[13,9]	0.853	0.003979	4.138E-5	0.8464	0.853	0.8597
RRH[13,10]	0.9037	0.004216	4.384E-5	0.8967	0.9037	0.9108

AB-2-14: Summary Statistics for the State of Sabah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[14,2]	1.149	0.004785	4.805E-5	1.141	1.149	1.156
RRH[14,3]	1.534	0.006388	6.416E-5	1.524	1.534	1.544
RRH[14,4]	1.198	0.004988	5.01E-5	1.19	1.198	1.206
RRH[14,5]	0.9583	0.003992	4.009E-5	0.952	0.9583	0.9648
RRH[14,6]	0.8155	0.003397	3.411E-5	0.8101	0.8155	0.821
RRH[14,7]	0.7943	0.003309	3.323E-5	0.7891	0.7943	0.7996
RRH[14,8]	0.7812	0.003254	3.268E-5	0.776	0.7812	0.7864
RRH[14,9]	0.6419	0.002674	2.685E-5	0.6377	0.6419	0.6462
RRH[14,10]	0.6817	0.002839	2.852E-5	0.6772	0.6817	0.6862

AB-2-15: Summary Statistics for Labuan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[15,2]	1.531	0.004175	4.73E-5	1.527	1.531	1.535
RRH[15,3]	2.041	0.005568	6.307E-5	2.036	2.041	2.046
RRH[15,4]	1.599	0.004362	4.941E-5	1.595	1.599	1.603
RRH[15,5]	1.277	0.003483	3.946E-5	1.274	1.277	1.28
RRH[15,6]	1.089	0.00297	3.365E-5	1.086	1.089	1.092
RRH[15,7]	1.058	0.002887	3.27E-5	1.056	1.058	1.061
RRH[15,8]	1.039	0.002833	3.209E-5	1.036	1.039	1.041
RRH[15,9]	0.8507	0.00232	2.629E-5	0.8486	0.8507	0.8529
RRH[15,10]	0.9026	0.002462	2.789E-5	0.9004	0.9027	0.905

AB-2-16: Summary Statistics for the State of Sarawak

node	mean	sd	MC error	2.5%	median	97.5%
RRH[16,2]	1.217	0.005432	5.411E-5	1.208	1.217	1.226
RRH[16,3]	1.624	0.007248	7.22E-5	1.612	1.624	1.636
RRH[16,4]	1.267	0.005654	5.632E-5	1.258	1.267	1.276
RRH[16,5]	1.014	0.004523	4.506E-5	1.006	1.014	1.021
RRH[16,6]	0.8619	0.003847	3.832E-5	0.8557	0.8619	0.8683
RRH[16,7]	0.8397	0.003748	3.733E-5	0.8337	0.8397	0.8459
RRH[16,8]	0.8258	0.003685	3.671E-5	0.8198	0.8258	0.8319
RRH[16,9]	0.6782	0.003027	3.015E-5	0.6733	0.6782	0.6832
RRH[16,10]	0.7202	0.003214	3.202E-5	0.715	0.7202	0.7256

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

AB-3-1: Summary Statistics for the State of Perlis

node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	2.493	0.0215	2.722E-4	2.453	2.493	2.533
RRH[1,3]	3.2	0.0276	3.493E-4	3.148	3.2	3.251
RRH[1,4]	2.408	0.02077	2.629E-4	2.369	2.408	2.447
RRH[1,5]	1.865	0.01608	2.035E-4	1.834	1.865	1.895
RRH[1,6]	1.542	0.0133	1.683E-4	1.517	1.542	1.567
RRH[1,7]	1.434	0.01237	1.566E-4	1.411	1.434	1.458
RRH[1,8]	1.352	0.01166	1.476E-4	1.33	1.352	1.373
RRH[1,9]	1.035	0.00893	1.13E-4	1.019	1.035	1.052
RRH[1,10]	1.041	0.008976	1.136E-4	1.024	1.041	1.058

AB-3-2: Summary Statistics for the State of Kedah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[2,2]	1.613	0.00737	7.694E-5	1.601	1.613	1.625
RRH[2,3]	2.065	0.009436	9.851E-5	2.05	2.065	2.081
RRH[2,4]	1.562	0.007139	7.453E-5	1.551	1.562	1.574
RRH[2,5]	1.238	0.005656	5.904E-5	1.229	1.238	1.247
RRH[2,6]	1.02	0.004662	4.866E-5	1.013	1.02	1.028
RRH[2,7]	0.9575	0.004376	4.568E-5	0.9504	0.9576	0.9647
RRH[2,8]	0.9189	0.004199	4.384E-5	0.9121	0.9189	0.9258
RRH[2,9]	0.7347	0.003357	3.505E-5	0.7293	0.7347	0.7403
RRH[2,10]	0.7465	0.003411	3.561E-5	0.741	0.7465	0.7521

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

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[3,2]	1.179	0.006257	7.039E-5	1.169	1.179	1.19
RRH[3,3]	1.479	0.007846	8.827E-5	1.465	1.479	1.492
RRH[3,4]	1.163	0.006172	6.944E-5	1.153	1.163	1.174
RRH[3,5]	0.9069	0.004812	5.414E-5	0.8987	0.9069	0.9152
RRH[3,6]	0.7543	0.004003	4.503E-5	0.7475	0.7543	0.7613
RRH[3,7]	0.7142	0.00379	4.264E-5	0.7078	0.7142	0.7208
RRH[3,8]	0.6839	0.003629	4.083E-5	0.6778	0.6839	0.6902
RRH[3,9]	0.5518	0.002928	3.294E-5	0.5469	0.5518	0.5569
RRH[3,10]	0.5713	0.003031	3.41E-5	0.5661	0.5713	0.5765

AB-3-4: Summary Statistics for the State of Perak

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[4,2]	1.615	0.007175	7.997E-5	1.604	1.615	1.627
RRH[4,3]	2.103	0.009344	1.041E-4	2.089	2.103	2.118
RRH[4,4]	1.606	0.007133	7.95E-5	1.594	1.606	1.617
RRH[4,5]	1.245	0.005533	6.167E-5	1.237	1.245	1.254
RRH[4,6]	1.025	0.004554	5.076E-5	1.018	1.025	1.032
RRH[4,7]	0.9638	0.004281	4.772E-5	0.957	0.9638	0.9707
RRH[4,8]	0.9165	0.004071	4.538E-5	0.9101	0.9165	0.9231
RRH[4,9]	0.7277	0.003232	3.603E-5	0.7226	0.7277	0.7329
RRH[4,10]	0.7432	0.003301	3.68E-5	0.7379	0.7432	0.7485

AB-3-5: Summary Statistics for Kuala Lumpur

node	mean	Sd	MC error	2.5%	median	97.5%
RRH[5,2]	0.8409	0.004975	5.636E-5	0.8322	0.8409	0.8498
RRH[5,3]	1.088	0.006436	7.291E-5	1.076	1.088	1.099
RRH[5,4]	0.8461	0.005006	5.671E-5	0.8373	0.8461	0.8551
RRH[5,5]	0.6635	0.003925	4.447E-5	0.6566	0.6635	0.6705
RRH[5,6]	0.5546	0.003281	3.717E-5	0.5488	0.5546	0.5605
RRH[5,7]	0.53	0.003136	3.552E-5	0.5245	0.53	0.5356
RRH[5,8]	0.5141	0.003041	3.446E-5	0.5087	0.5141	0.5195
RRH[5,9]	0.4059	0.002401	2.721E-5	0.4017	0.4059	0.4102
RRH[5,10]	0.4183	0.002475	2.804E-5	0.414	0.4183	0.4227

AB-3-6: Summary Statistics for Putrajaya

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[6,2]	5.808	0.05709	6.452E-4	5.701	5.808	5.917
RRH[6,3]	7.271	0.07147	8.078E-4	7.137	7.271	7.408
RRH[6,4]	5.765	0.05666	6.404E-4	5.659	5.765	5.873
RRH[6,5]	4.229	0.04157	4.699E-4	4.151	4.229	4.309
RRH[6,6]	3.253	0.03197	3.614E-4	3.193	3.253	3.314
RRH[6,7]	2.922	0.02872	3.246E-4	2.868	2.922	2.977
RRH[6,8]	2.427	0.02385	2.696E-4	2.382	2.427	2.472
RRH[6,9]	1.969	0.01935	2.187E-4	1.933	1.969	2.006
RRH[6,10]	1.636	0.01608	1.817E-4	1.606	1.636	1.666

AB-3-7: Summary Statistics for the State of Selangor

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[7,2]	0.8091	0.003379	3.879E-5	0.8039	0.8091	0.8143
RRH[7,3]	1.059	0.004423	5.076E-5	1.052	1.059	1.066
RRH[7,4]	0.8137	0.003399	3.901E-5	0.8085	0.8137	0.8189
RRH[7,5]	0.6404	0.002675	3.07E-5	0.6363	0.6404	0.6444
RRH[7,6]	0.5336	0.002229	2.558E-5	0.5302	0.5336	0.537
RRH[7,7]	0.5081	0.002122	2.436E-5	0.5049	0.5081	0.5114
RRH[7,8]	0.49	0.002047	2.349E-5	0.4868	0.49	0.4931
RRH[7,9]	0.3935	0.001644	1.886E-5	0.391	0.3935	0.396
RRH[7,10]	0.4074	0.001701	1.953E-5	0.4047	0.4074	0.41

AB-3-8: Summary Statistics for the State of Negeri Sembilan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[8,2]	2.398	0.01204	1.309E-4	2.378	2.398	2.418
RRH[8,3]	3.06	0.01536	1.67E-4	3.035	3.06	3.086
RRH[8,4]	2.328	0.01169	1.271E-4	2.309	2.328	2.348
RRH[8,5]	1.799	0.009032	9.818E-5	1.784	1.799	1.814
RRH[8,6]	1.476	0.00741	8.055E-5	1.464	1.476	1.488
RRH[8,7]	1.378	0.006919	7.521E-5	1.367	1.378	1.39
RRH[8,8]	1.3	0.006526	7.094E-5	1.289	1.3	1.311
RRH[8,9]	1.012	0.005079	5.521E-5	1.003	1.012	1.02
RRH[8,10]	1.018	0.005111	5.556E-5	1.01	1.018	1.027

AB-3-9: Summary Statistics for the State of Melaka

node	mean	sd	MC error	2.5%	median	97.5%
RRH[9,2]	1.925	0.0111	1.35E-4	1.906	1.925	1.945
RRH[9,3]	2.471	0.01425	1.732E-4	2.446	2.471	2.496
RRH[9,4]	1.897	0.01094	1.33E-4	1.878	1.897	1.916
RRH[9,5]	1.462	0.00843	1.025E-4	1.447	1.462	1.477
RRH[9,6]	1.202	0.006932	8.426E-5	1.19	1.202	1.214
RRH[9,7]	1.132	0.006529	7.935E-5	1.121	1.132	1.144
RRH[9,8]	1.071	0.006178	7.508E-5	1.06	1.071	1.082
RRH[9,9]	0.8428	0.00486	5.907E-5	0.8342	0.8428	0.8513
RRH[9,10]	0.8569	0.004942	6.006E-5	0.8482	0.8569	0.8655

AB-3-10: Summary Statistics for the State of Johor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[10,2]	1.748	0.007009	7.51E-5	1.738	1.748	1.759
RRH[10,3]	2.239	0.008975	9.617E-5	2.226	2.239	2.253
RRH[10,4]	1.709	0.00685	7.34E-5	1.699	1.709	1.719
RRH[10,5]	1.334	0.005348	5.731E-5	1.326	1.334	1.342
RRH[10,6]	1.1	0.004411	4.727E-5	1.094	1.1	1.107
RRH[10,7]	1.035	0.004147	4.444E-5	1.029	1.035	1.041
RRH[10,8]	0.9807	0.003931	4.213E-5	0.9751	0.9807	0.9868
RRH[10,9]	0.7834	0.00314	3.365E-5	0.7789	0.7834	0.7882
RRH[10,10]	0.797	0.003195	3.424E-5	0.7925	0.797	0.802

AB-3-11: Summary Statistics for the State of Pahang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[11,2]	2.112	0.009781	1.004E-4	2.096	2.112	2.128
RRH[11,3]	2.707	0.01253	1.286E-4	2.686	2.707	2.727
RRH[11,4]	2.071	0.009588	9.84E-5	2.055	2.071	2.086
RRH[11,5]	1.605	0.007434	7.629E-5	1.593	1.605	1.617
RRH[11,6]	1.318	0.006102	6.262E-5	1.308	1.318	1.328
RRH[11,7]	1.233	0.005707	5.857E-5	1.223	1.233	1.242
RRH[11,8]	1.167	0.005405	5.546E-5	1.158	1.167	1.176
RRH[11,9]	0.9183	0.004252	4.364E-5	0.9115	0.9183	0.9252
RRH[11,10]	0.922	0.004269	4.382E-5	0.9151	0.922	0.9289

AB-3-12: Summary Statistics for the State of Terengganu

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[12,2]	2.408	0.01183	1.376E-4	2.389	2.408	2.428
RRH[12,3]	3.045	0.01496	1.74E-4	3.02	3.045	3.07
RRH[12,4]	2.337	0.01148	1.336E-4	2.318	2.337	2.356
RRH[12,5]	1.804	0.008866	1.031E-4	1.79	1.804	1.819
RRH[12,6]	1.478	0.007264	8.449E-5	1.466	1.478	1.49
RRH[12,7]	1.38	0.00678	7.886E-5	1.369	1.38	1.391
RRH[12,8]	1.288	0.00633	7.362E-5	1.278	1.288	1.299
RRH[12,9]	1.033	0.005075	5.903E-5	1.025	1.033	1.041
RRH[12,10]	1.023	0.005029	5.849E-5	1.015	1.023	1.032

AB-3-13: Summary Statistics for the State of Kelantan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[13,2]	1.713	0.008171	8.633E-5	1.7	1.713	1.727
RRH[13,3]	2.219	0.01058	1.118E-4	2.202	2.219	2.236
RRH[13,4]	1.7	0.008107	8.565E-5	1.686	1.7	1.713
RRH[13,5]	1.324	0.006314	6.671E-5	1.314	1.324	1.334
RRH[13,6]	1.096	0.00523	5.526E-5	1.088	1.096	1.105
RRH[13,7]	1.034	0.004934	5.213E-5	1.026	1.034	1.043
RRH[13,8]	0.9782	0.004666	4.93E-5	0.9706	0.9782	0.9859
RRH[13,9]	0.7757	0.0037	3.909E-5	0.7697	0.7757	0.7818
RRH[13,10]	0.7859	0.003749	3.96E-5	0.7798	0.7859	0.7921

AB-3-14: Summary Statistics for the State of Sabah

node	mean	sd	MC error	2.5%	median	97.5%
RRH[14,2]	1.267	0.005432	5.491E-5	1.258	1.267	1.275
RRH[14,3]	1.637	0.007019	7.096E-5	1.626	1.637	1.648
RRH[14,4]	1.261	0.005407	5.466E-5	1.253	1.261	1.269
RRH[14,5]	0.9866	0.004231	4.277E-5	0.9801	0.9866	0.9932
RRH[14,6]	0.8166	0.003502	3.54E-5	0.8112	0.8166	0.8221
RRH[14,7]	0.7758	0.003327	3.363E-5	0.7707	0.7758	0.7811
RRH[14,8]	0.7403	0.003175	3.21E-5	0.7355	0.7403	0.7453
RRH[14,9]	0.5931	0.002543	2.571E-5	0.5892	0.5931	0.5971
RRH[14,10]	0.6084	0.002609	2.637E-5	0.6044	0.6084	0.6125

AB-3-15: Summary Statistics for Labuan

node	mean	sd	MC error	2.5%	median	97.5%
RRH[15,2]	1.724	0.005007	5.682E-5	1.72	1.724	1.729
RRH[15,3]	2.235	0.00649	7.365E-5	2.229	2.235	2.24
RRH[15,4]	1.716	0.004983	5.655E-5	1.712	1.716	1.72
RRH[15,5]	1.339	0.003888	4.412E-5	1.335	1.339	1.342
RRH[15,6]	1.117	0.003245	3.682E-5	1.114	1.117	1.12
RRH[15,7]	1.056	0.003067	3.481E-5	1.053	1.056	1.059
RRH[15,8]	0.9964	0.002894	3.284E-5	0.994	0.9964	0.999
RRH[15,9]	0.7834	0.002275	2.582E-5	0.7815	0.7834	0.7854
RRH[15,10]	0.7931	0.002303	2.614E-5	0.7911	0.7931	0.7952

AB-3-16: Summary Statistics for the State of Sarawak

node	mean	sd	MC error	2.5%	median	97.5%
RRH[16,2]	1.344	0.006139	6.108E-5	1.334	1.344	1.354
RRH[16,3]	1.753	0.008011	7.97E-5	1.741	1.753	1.766
RRH[16,4]	1.337	0.006109	6.078E-5	1.327	1.337	1.347
RRH[16,5]	1.041	0.004759	4.734E-5	1.034	1.041	1.049
RRH[16,6]	0.8624	0.003941	3.921E-5	0.8562	0.8624	0.8688
RRH[16,7]	0.8155	0.003726	3.707E-5	0.8096	0.8155	0.8215
RRH[16,8]	0.7818	0.003572	3.554E-5	0.7762	0.7818	0.7877
RRH[16,9]	0.6253	0.002857	2.843E-5	0.6208	0.6253	0.6299
RRH[16,10]	0.6388	0.002919	2.904E-5	0.6342	0.6388	0.6435

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

AB-4-1: Summary Statistics for the State of Perlis

node	mean	sd	MC error	2.5%	median	97.5%
RRH[1,2]	2.877	0.02399	2.993E-4	2.831	2.877	2.923
RRH[1,3]	3.702	0.03088	3.852E-4	3.643	3.702	3.762
RRH[1,4]	2.193	0.01829	2.282E-4	2.158	2.193	2.229
RRH[1,5]	1.91	0.01593	1.987E-4	1.879	1.91	1.94
RRH[1,6]	1.157	0.009648	1.203E-4	1.138	1.157	1.175
RRH[1,7]	1.479	0.01233	1.539E-4	1.455	1.479	1.503
RRH[1,8]	1.015	0.008469	1.056E-4	0.999	1.015	1.032
RRH[1,9]	1.341	0.01118	1.395E-4	1.319	1.341	1.362
RRH[1,10]	0.9322	0.007775	9.698E-5	0.9171	0.9322	0.9472

AB-4-2: Summary Statistics for the State of Kedah

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[2,2]	2.163	0.008761	9.01E-5	2.147	2.163	2.18
RRH[2,3]	2.815	0.0114	1.172E-4	2.794	2.815	2.837
RRH[2,4]	1.57	0.006357	6.537E-5	1.558	1.57	1.582
RRH[2,5]	1.292	0.005231	5.38E-5	1.282	1.292	1.302
RRH[2,6]	0.7941	0.003216	3.307E-5	0.7882	0.7941	0.8001
RRH[2,7]	0.964	0.003904	4.015E-5	0.9569	0.9641	0.9713
RRH[2,8]	0.6138	0.002486	2.556E-5	0.6092	0.6138	0.6184
RRH[2,9]	0.7089	0.002871	2.952E-5	0.7036	0.7089	0.7142
RRH[2,10]	0.5102	0.002066	2.125E-5	0.5064	0.5102	0.514

AB-4-3: Summary Statistics for the State of Pulau Pinang

node	Mean	sd	MC error	2.5%	Median	97.5%
RRH[3,2]	1.711	0.008353	9.151E-5	1.696	1.712	1.727
RRH[3,3]	2.425	0.01184	1.297E-4	2.403	2.425	2.448
RRH[3,4]	1.038	0.005067	5.551E-5	1.029	1.038	1.048
RRH[3,5]	1.15	0.005613	6.15E-5	1.14	1.15	1.161
RRH[3,6]	0.4353	0.002124	2.327E-5	0.4313	0.4353	0.4393
RRH[3,7]	0.8461	0.004129	4.524E-5	0.8385	0.8461	0.8539
RRH[3,8]	0.279	0.001362	1.492E-5	0.2765	0.279	0.2816
RRH[3,9]	0.6326	0.003087	3.382E-5	0.6269	0.6326	0.6384
RRH[3,10]	0.1703	6.847E-4	7.501E-6	0.169	0.1703	0.1716

AB-4-4: Summary Statistics for the State of Perak

node	Mean	sd	MC error	2.5%	Median	97.5%
RRH[4,2]	2.182	0.008498	9.212E-5	2.167	2.182	2.198
RRH[4,3]	2.63	0.01024	1.11E-4	2.612	2.63	2.649
RRH[4,4]	1.474	0.005739	6.222E-5	1.464	1.474	1.485
RRH[4,5]	1.286	0.005006	5.427E-5	1.277	1.286	1.295
RRH[4,6]	0.8006	0.003117	3.38E-5	0.795	0.8006	0.8064
RRH[4,7]	0.983	0.003828	4.149E-5	0.9761	0.983	0.99
RRH[4,8]	0.6739	0.002624	2.845E-5	0.6692	0.6739	0.6787
RRH[4,9]	0.7691	0.002995	3.247E-5	0.7637	0.7691	0.7746
RRH[4,10]	0.5502	0.002142	2.322E-5	0.5463	0.5502	0.5541

AB-4-5: Summary Statistics for Kuala Lumpur

node	mean	Sd	MC error	2.5%	Median	97.5%
RRH[5,2]	1.381	0.007651	8.711E-5	1.367	1.381	1.396
RRH[5,3]	1.716	0.009505	1.082E-4	1.698	1.716	1.734
RRH[5,4]	0.8351	0.004626	5.267E-5	0.8264	0.8351	0.844
RRH[5,5]	0.7635	0.004229	4.815E-5	0.7555	0.7635	0.7716
RRH[5,6]	0.3554	0.001968	2.241E-5	0.3517	0.3554	0.3591
RRH[5,7]	0.5135	0.002844	3.238E-5	0.5081	0.5135	0.5189
RRH[5,8]	0.2045	0.001133	1.289E-5	0.2023	0.2045	0.2066
RRH[5,9]	0.429	0.002376	2.705E-5	0.4245	0.429	0.4336
RRH[5,10]	0.1693	9.378E-4	1.068E-5	0.1675	0.1693	0.1711

AB-4-6: Summary Statistics for Putrajaya

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[6,2]	2.51	0.04329	4.975E-4	2.427	2.51	2.595
RRH[6,3]	3.388	0.06132	7.047E-4	3.27	3.388	3.509
RRH[6,4]	2.618	0.02513	2.888E-4	1.57	2.618	2.667
RRH[6,5]	2.907	0.0375	4.31E-4	1.835	2.907	2.981
RRH[6,6]	2.238	0.02148	2.468E-4	1.196	2.238	2.28
RRH[6,7]	2.845	0.03691	4.242E-4	1.774	2.845	2.917
RRH[6,8]	2.023	0.02902	3.335E-4	1.968	2.023	2.08
RRH[6,9]	2.029	0.02907	3.341E-4	1.973	2.029	2.086
RRH[6,10]	1.297	0.03165	3.637E-4	1.236	1.297	1.359

AB-4-7: Summary Statistics for the State of Selangor

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[7,2]	1.332	0.004804	5.351E-5	1.323	1.332	1.34
RRH[7,3]	1.566	0.00565	6.293E-5	1.556	1.566	1.576
RRH[7,4]	0.8503	0.003067	3.416E-5	0.8448	0.8503	0.8557
RRH[7,5]	0.6832	0.002464	2.745E-5	0.6788	0.6832	0.6875
RRH[7,6]	0.3934	0.001419	1.581E-5	0.3909	0.3934	0.3959
RRH[7,7]	0.4633	0.001671	1.861E-5	0.4603	0.4633	0.4662
RRH[7,8]	0.28	0.00101	1.125E-5	0.2782	0.28	0.2817
RRH[7,9]	0.3381	0.001219	1.358E-5	0.3359	0.3381	0.3402
RRH[7,10]	0.2087	7.529E-4	8.387E-6	0.2074	0.2087	0.2101

AB-4-8: Summary Statistics for the State of Negeri Sembilan

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[8,2]	2.806	0.01274	1.358E-4	2.783	2.806	2.83
RRH[8,3]	3.707	0.01683	1.794E-4	3.676	3.707	3.739
RRH[8,4]	2.026	0.009197	9.805E-5	2.009	2.026	2.043
RRH[8,5]	1.908	0.008662	9.234E-5	1.892	1.908	1.924
RRH[8,6]	1.104	0.005013	5.344E-5	1.095	1.104	1.114
RRH[8,7]	1.494	0.006783	7.231E-5	1.482	1.494	1.507
RRH[8,8]	0.9469	0.004299	4.583E-5	0.9391	0.9469	0.955
RRH[8,9]	1.265	0.005744	6.123E-5	1.255	1.265	1.276
RRH[8,10]	0.8296	0.003767	4.016E-5	0.8228	0.8296	0.8367

AB-4-9: Summary Statistics for the State of Melaka

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[9,2]	2.44	0.0131	1.601E-4	2.415	2.44	2.464
RRH[9,3]	3.141	0.01687	2.061E-4	3.11	3.141	3.173
RRH[9,4]	1.646	0.008838	1.08E-4	1.629	1.646	1.663
RRH[9,5]	1.604	0.00861	1.052E-4	1.587	1.604	1.62
RRH[9,6]	0.885	0.004752	5.808E-5	0.8761	0.885	0.894
RRH[9,7]	1.22	0.006552	8.009E-5	1.208	1.22	1.233
RRH[9,8]	0.732	0.00393	4.804E-5	0.7246	0.732	0.7394
RRH[9,9]	1.002	0.005381	6.577E-5	0.9921	1.002	1.012
RRH[9,10]	0.6047	0.003247	3.968E-5	0.5986	0.6047	0.6108

AB-4-10: Summary Statistics for the State of Johor

node	mean	sd	MC error	2.5%	median	97.5%
RRH[10,2]	2.283	0.007768	8.217E-5	2.27	2.283	2.298
RRH[10,3]	2.97	0.0101	1.069E-4	2.953	2.97	2.988
RRH[10,4]	1.604	0.005456	5.771E-5	1.594	1.604	1.614
RRH[10,5]	1.435	0.004883	5.165E-5	1.427	1.435	1.444
RRH[10,6]	0.8196	0.002788	2.949E-5	0.8149	0.8196	0.8247
RRH[10,7]	1.077	0.003664	3.876E-5	1.071	1.077	1.084
RRH[10,8]	0.6827	0.002323	2.457E-5	0.6788	0.6827	0.687
RRH[10,9]	0.8211	0.002793	2.955E-5	0.8164	0.8211	0.8262
RRH[10,10]	0.5521	0.001878	1.987E-5	0.5489	0.5521	0.5555

AB-4-11: Summary Statistics for the State of Pahang

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[11,2]	2.614	0.01075	1.035E-4	2.594	2.614	2.633
RRH[11,3]	3.39	0.01395	1.342E-4	3.365	3.39	3.416
RRH[11,4]	1.809	0.007442	7.163E-5	1.796	1.809	1.823
RRH[11,5]	1.69	0.006952	6.691E-5	1.677	1.69	1.703
RRH[11,6]	0.9729	0.004002	3.852E-5	0.9656	0.9729	0.9802
RRH[11,7]	1.325	0.005451	5.247E-5	1.315	1.325	1.335
RRH[11,8]	0.824	0.003389	3.262E-5	0.8178	0.824	0.8301
RRH[11,9]	1.082	0.004451	4.284E-5	1.074	1.082	1.09
RRH[11,2]	2.614	0.01075	1.035E-4	2.594	2.614	2.633

AB-4-12: Summary Statistics for the State of Terengganu

node	Mean	sd	MC error	2.5%	median	97.5%
RRH[12,2]	2.804	0.01242	1.429E-4	2.781	2.804	2.827
RRH[12,3]	3.844	0.01702	1.959E-4	3.813	3.844	3.875
RRH[12,4]	1.959	0.008675	9.985E-5	1.943	1.959	1.975
RRH[12,5]	1.977	0.008754	1.008E-4	1.961	1.977	1.993
RRH[12,6]	1.05	0.004653	5.355E-5	1.042	1.05	1.059
RRH[12,7]	1.563	0.006925	7.97E-5	1.551	1.563	1.576
RRH[12,8]	0.9511	0.004212	4.848E-5	0.9434	0.9511	0.9588
RRH[12,9]	1.192	0.005278	6.074E-5	1.182	1.192	1.201
RRH[12,10]	0.8476	0.003754	4.321E-5	0.8407	0.8476	0.8545

AB-4-13: Summary Statistics for the State of Kelantan

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[13,2]	2.344	0.01002	1.036E-4	2.325	2.344	2.362
RRH[13,3]	2.896	0.01238	1.28E-4	2.873	2.896	2.919
RRH[13,4]	1.569	0.006706	6.934E-5	1.557	1.569	1.582
RRH[13,5]	1.384	0.005914	6.115E-5	1.373	1.384	1.395
RRH[13,6]	0.7883	0.003369	3.484E-5	0.7822	0.7883	0.7945
RRH[13,7]	1.022	0.004368	4.516E-5	1.014	1.022	1.03
RRH[13,8]	0.6805	0.002908	3.007E-5	0.6752	0.6805	0.6858
RRH[13,9]	0.8288	0.003542	3.663E-5	0.8224	0.8288	0.8354
RRH[13,10]	0.606	0.00259	2.678E-5	0.6013	0.606	0.6108

AB-4-14: Summary Statistics for the State of Sabah

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[14,2]	1.864	0.006939	6.927E-5	1.852	1.864	1.877
RRH[14,3]	2.328	0.008665	8.65E-5	2.313	2.328	2.344
RRH[14,4]	1.209	0.004499	4.492E-5	1.201	1.209	1.217
RRH[14,5]	1.074	0.003998	3.992E-5	1.067	1.074	1.081
RRH[14,6]	0.5912	0.0022	2.197E-5	0.5873	0.5912	0.5951
RRH[14,7]	0.7629	0.00284	2.835E-5	0.7579	0.7629	0.7681
RRH[14,8]	0.4604	0.001714	1.711E-5	0.4574	0.4604	0.4635
RRH[14,9]	0.5767	0.002147	2.143E-5	0.5729	0.5767	0.5806
RRH[14,10]	0.3645	0.001357	1.355E-5	0.3621	0.3645	0.367

AB-4-15: Summary Statistics for Labuan

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[15,2]	2.243	0.004437	4.889E-5	2.237	2.243	2.249
RRH[15,3]	2.754	0.005448	6.003E-5	2.747	2.754	2.761
RRH[15,4]	1.514	0.002995	3.3E-5	1.51	1.514	1.518
RRH[15,5]	1.285	0.002543	2.802E-5	1.282	1.285	1.289
RRH[15,6]	0.7206	0.001425	1.571E-5	0.7188	0.7206	0.7224
RRH[15,7]	0.9182	0.001816	2.002E-5	0.9159	0.9182	0.9206
RRH[15,8]	0.6365	0.001259	1.387E-5	0.6349	0.6365	0.6381
RRH[15,9]	0.7947	0.001572	1.732E-5	0.7927	0.7947	0.7967
RRH[15,10]	0.6003	0.001188	1.309E-5	0.5988	0.6003	0.6018

AB-4-16: Summary Statistics for the State of Sarawak

node	mean	sd	MC error	2.5%	Median	97.5%
RRH[16,2]	1.939	0.007851	7.828E-5	1.925	1.939	1.953
RRH[16,3]	2.312	0.009364	9.337E-5	2.295	2.312	2.329
RRH[16,4]	1.312	0.005313	5.298E-5	1.303	1.312	1.322
RRH[16,5]	1.102	0.004461	4.449E-5	1.094	1.102	1.11
RRH[16,6]	0.6751	0.002734	2.726E-5	0.6702	0.6751	0.6801
RRH[16,7]	0.7995	0.003238	3.229E-5	0.7938	0.7995	0.8054
RRH[16,8]	0.5222	0.002115	2.109E-5	0.5185	0.5222	0.5261
RRH[16,9]	0.5932	0.002403	2.396E-5	0.589	0.5932	0.5976
RRH[16,10]	0.43	0.001741	1.736E-5	0.4269	0.43	0.4332

Appendix C

History Plot for Convergence of the Relative Risk Estimation based on Stochastic SIC Model, SIR Model, SCIR Model and SVCIR Model.

AC-1: 'History' plot for convergence of the relative risk estimation based on stochastic SIC model, SIR model, SCIR model and SVCIR model for the state of Perlis only.

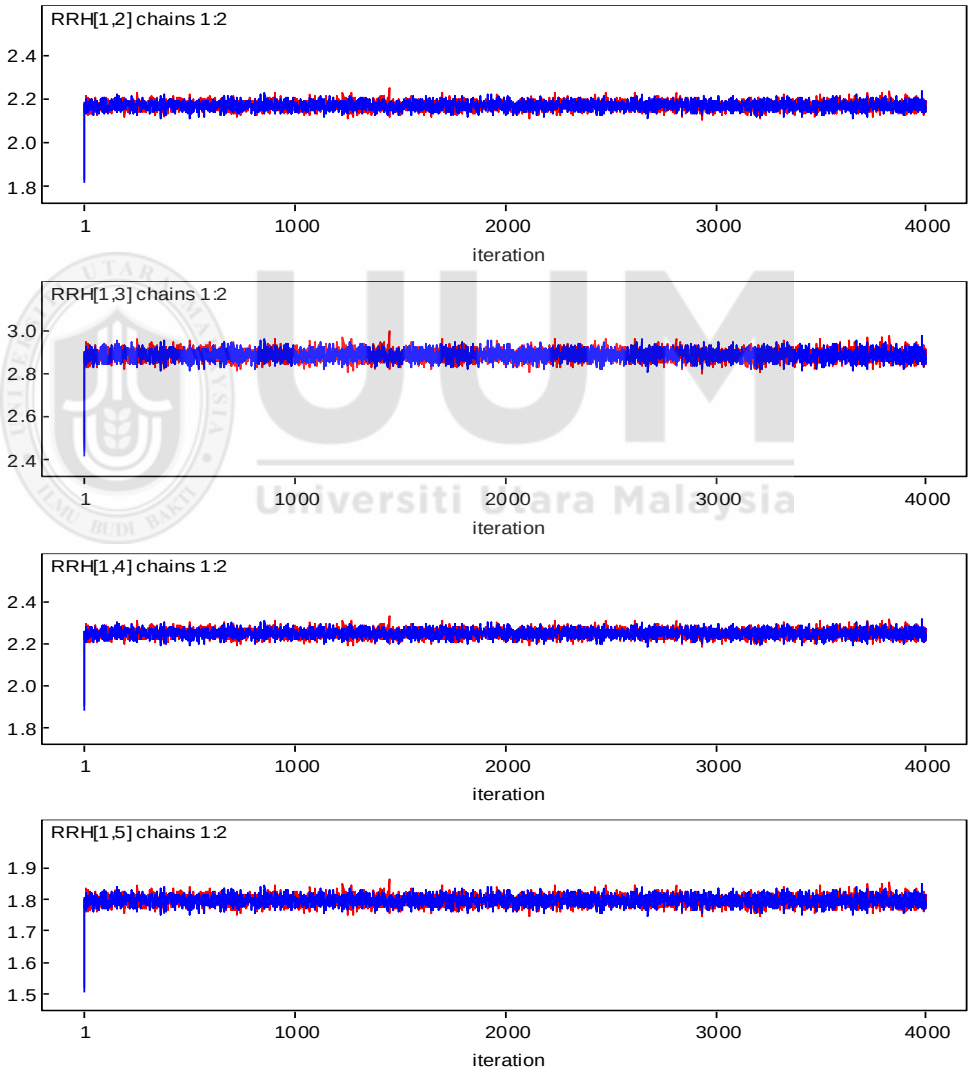

Figure AC.1. 'History' plot based on stochastic SIC model for the state of Perlis.

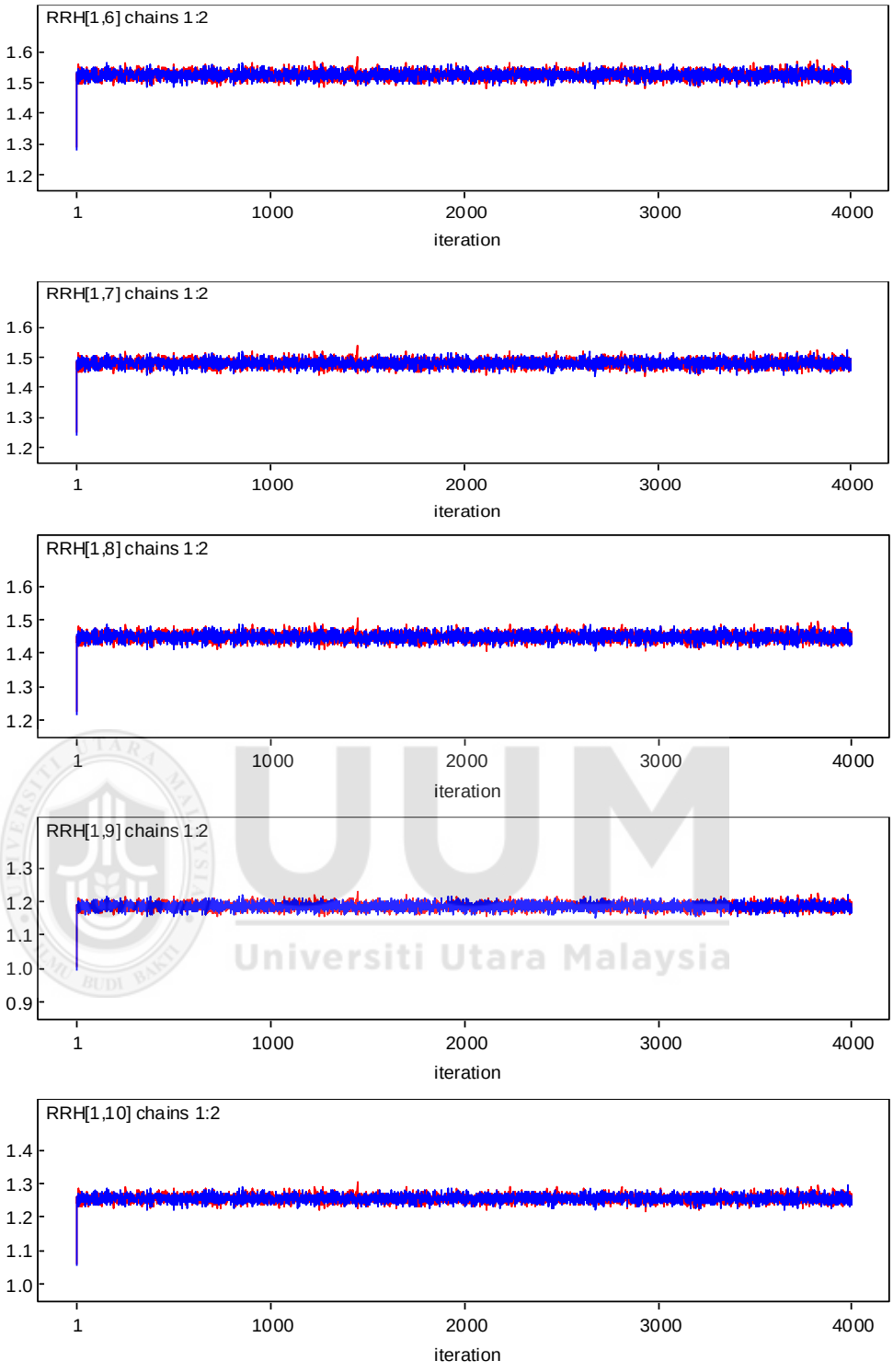

Figure AC.1. (continued)

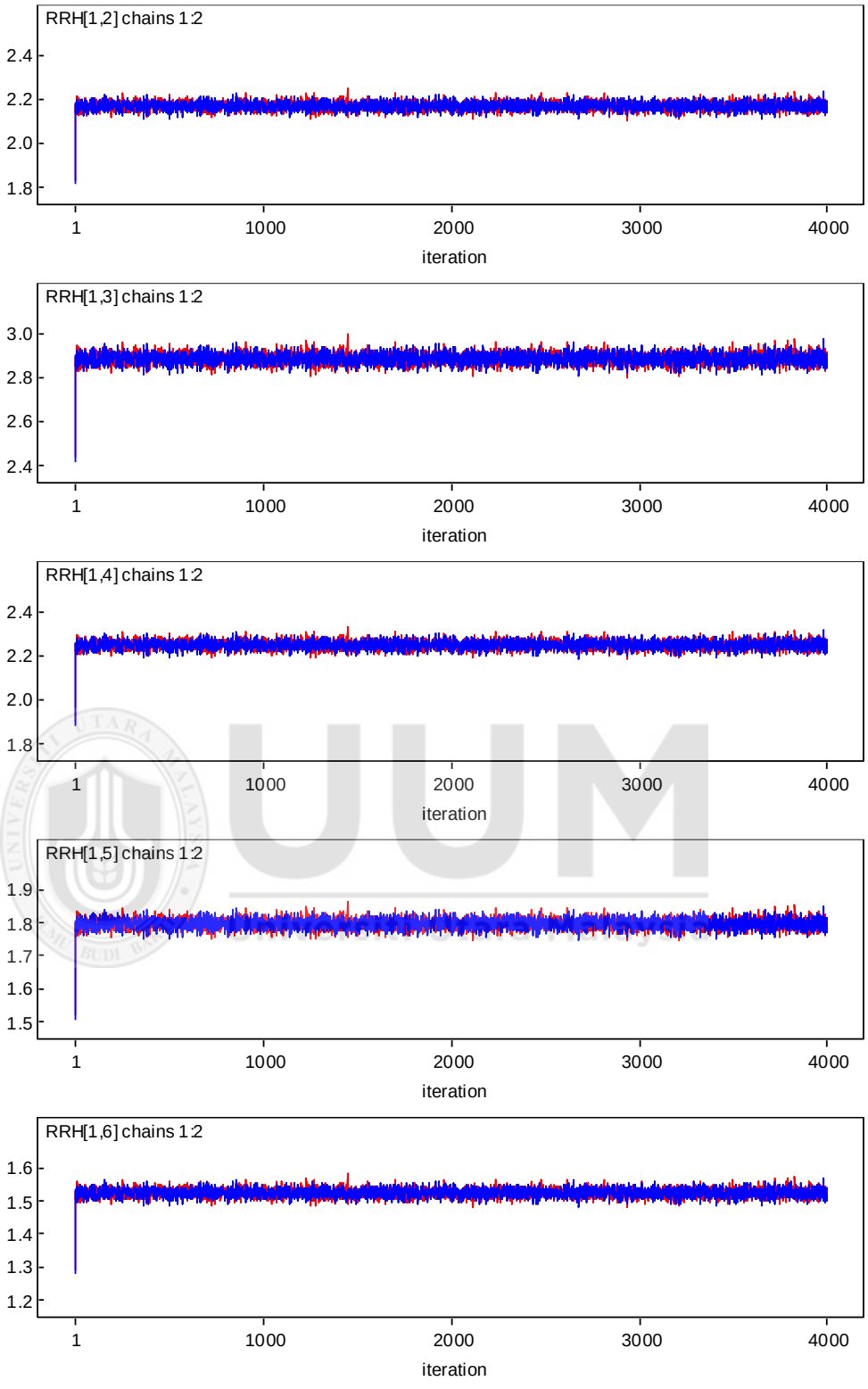

Figure AC.2. 'History' plot based on stochastic SIR model for the state of Perlis.

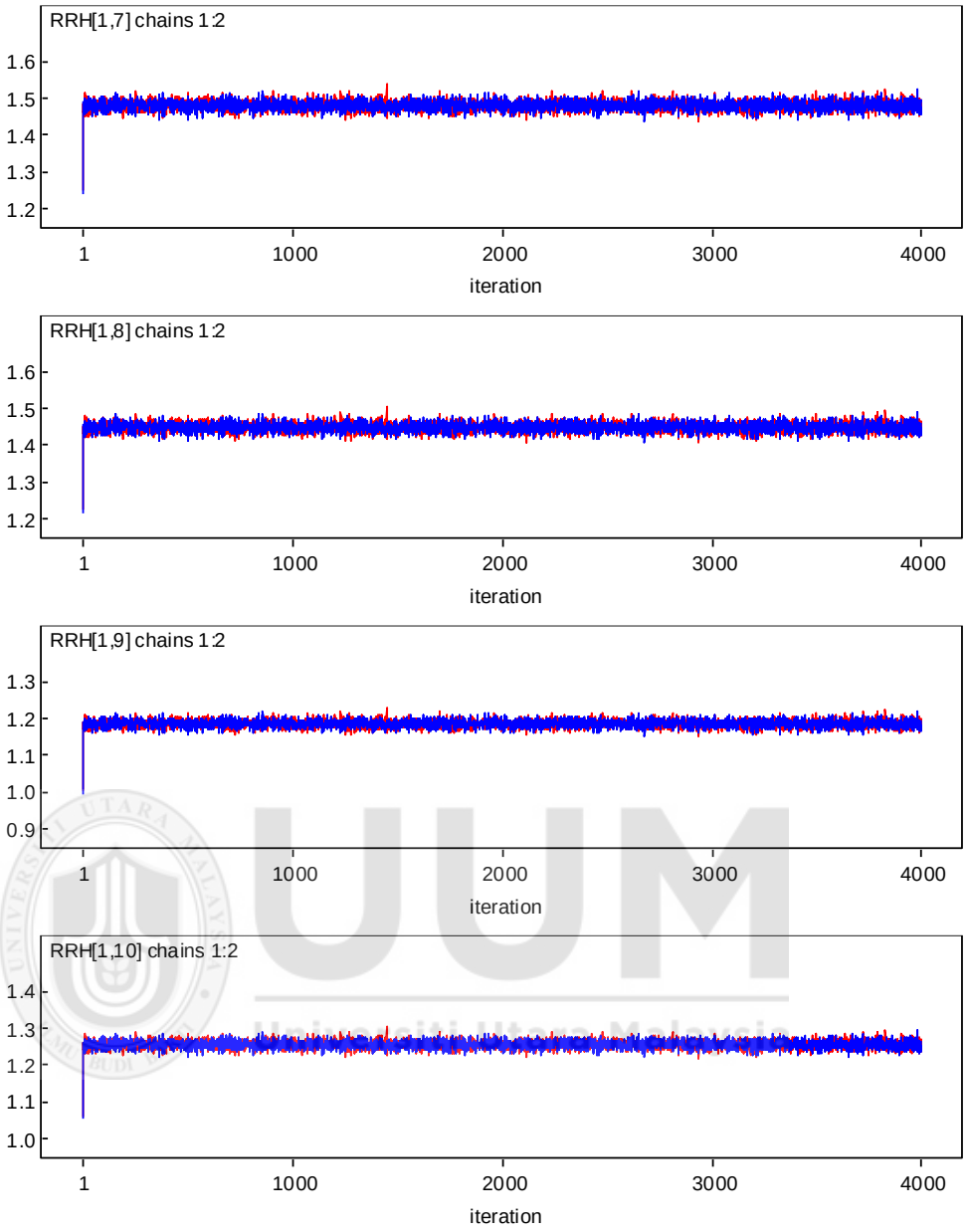

Figure AC.2. (continued)

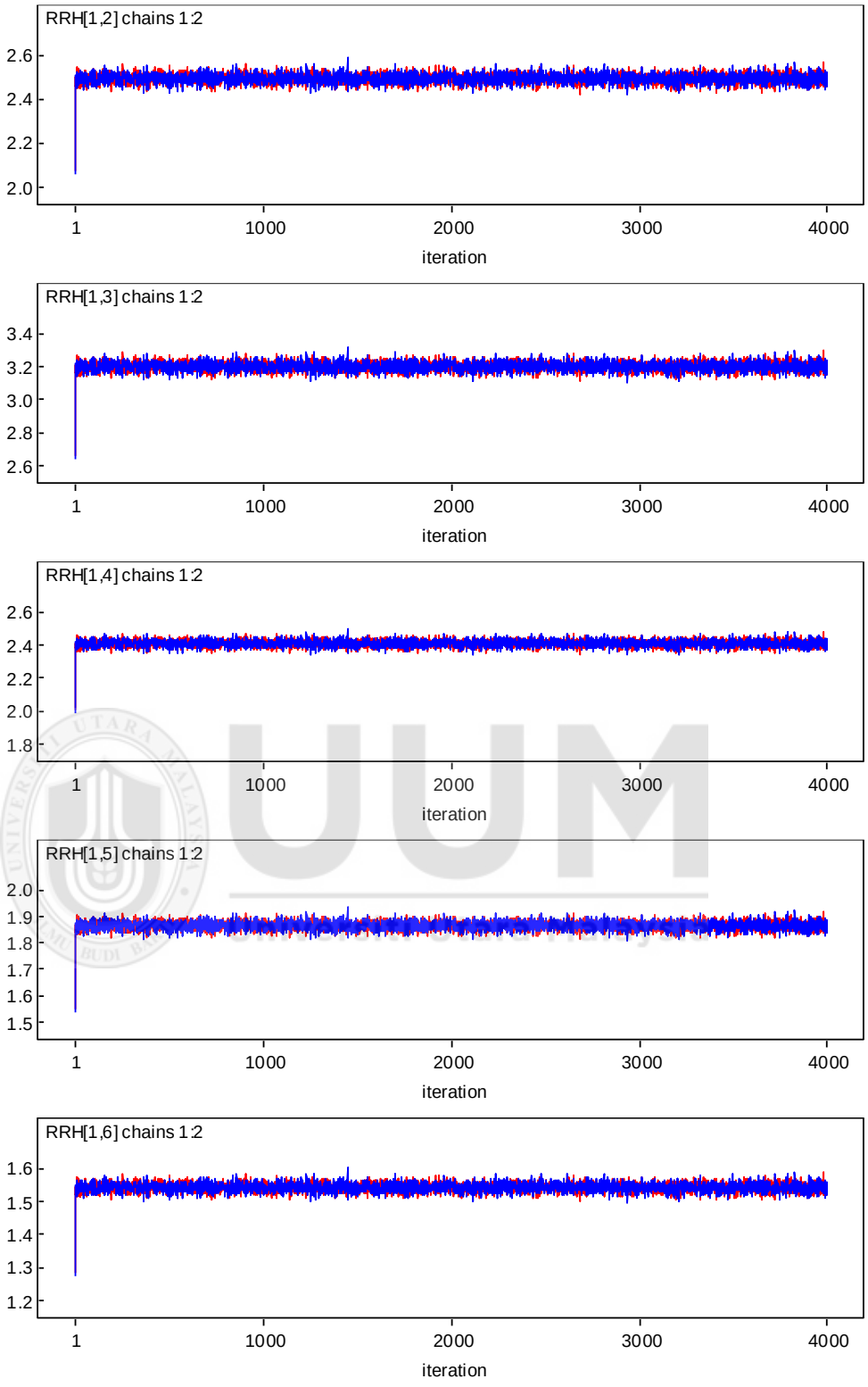

Figure AC.3. 'History' plot based on stochastic SCIR model for the state of Perlis.

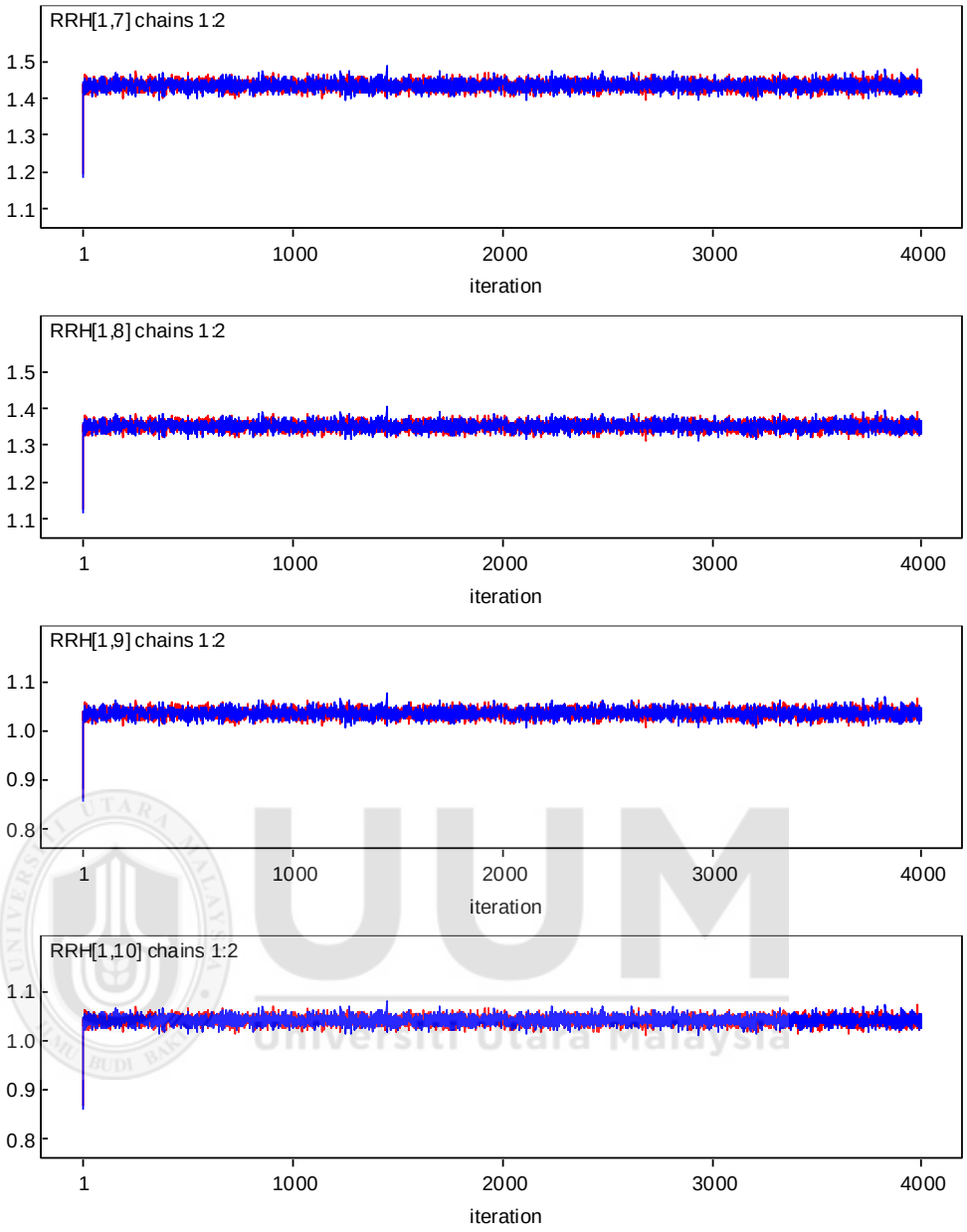

Figure AC.3. (continued)

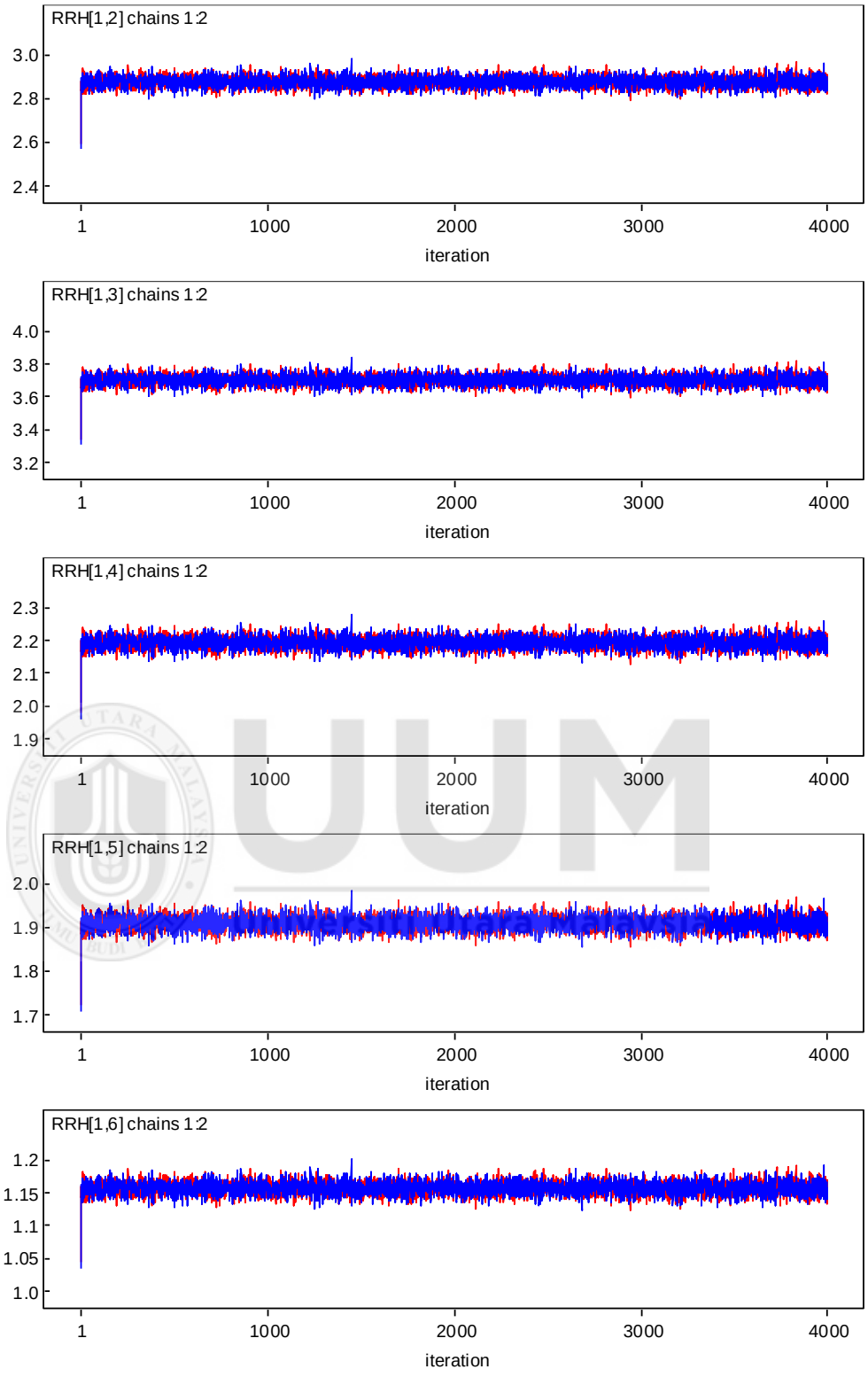

Figure AC.4. 'History' plot based on stochastic SVCIR model for the state of Perlis.

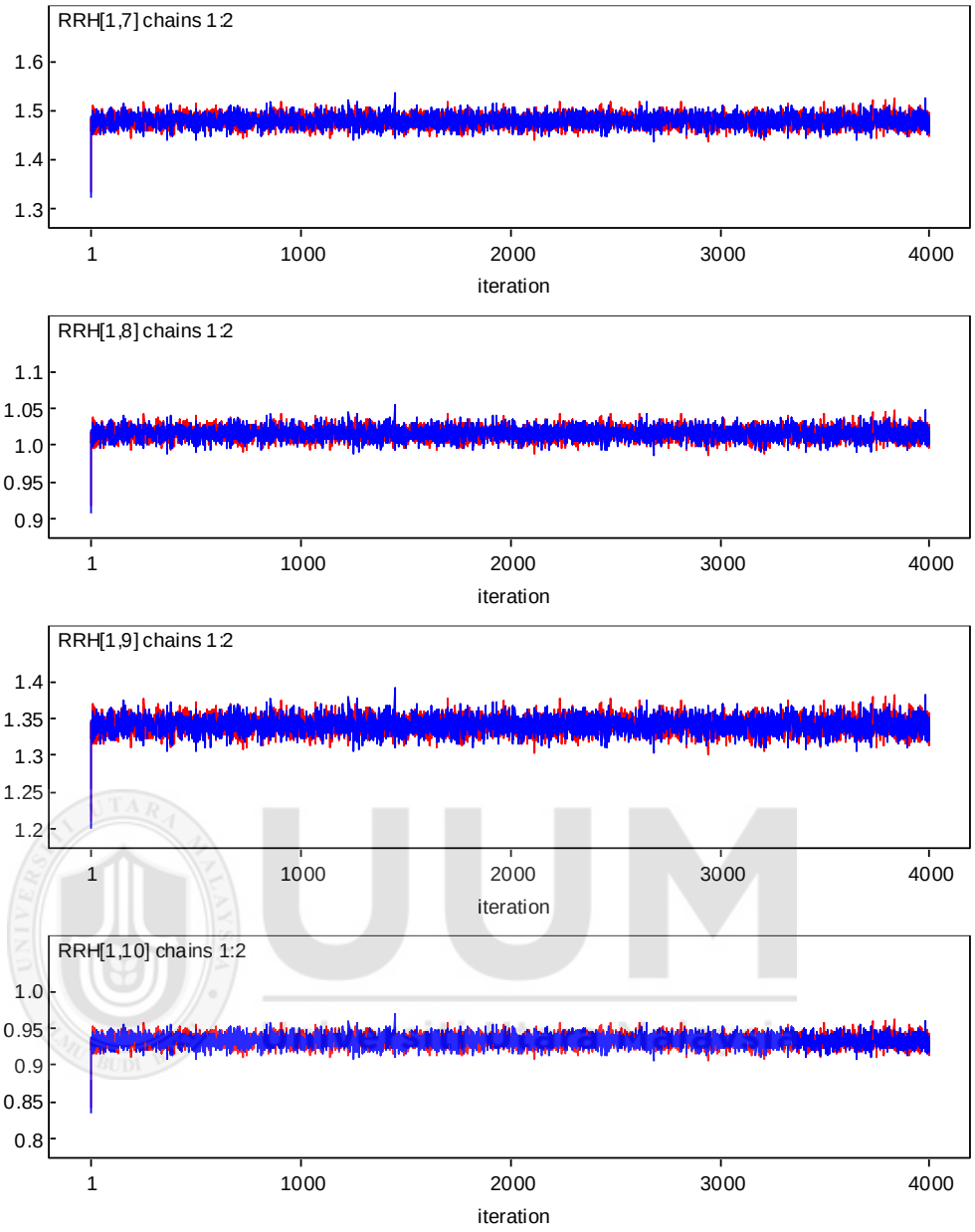

Figure AC.4. (continued)

Appendix D

WinBUGS Outputs of the Quantiles Graph of the Relative Risk Estimation

AD-1: WinBUGS outputs of the quantiles graph based on stochastic SIC model, SIR model, SCIR model and SVCIR model for the state of Perlis only.

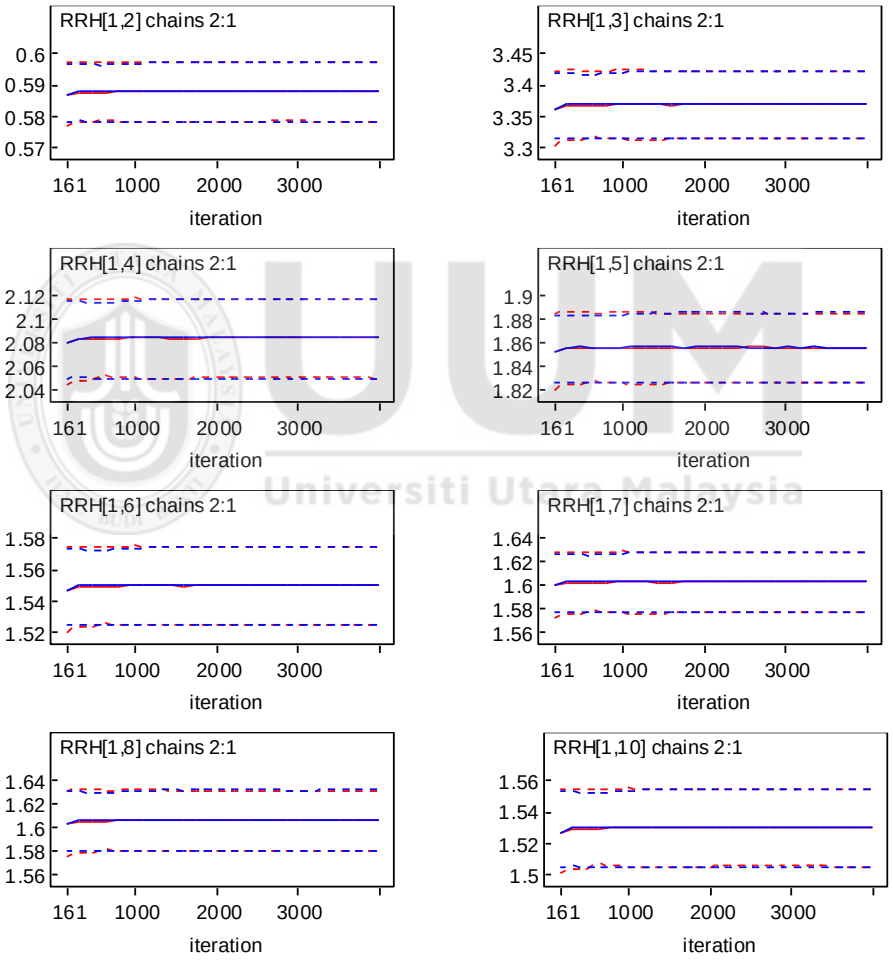

Figure AD.1. Quantile graph based on stochastic SIC model for the state of Perlis.

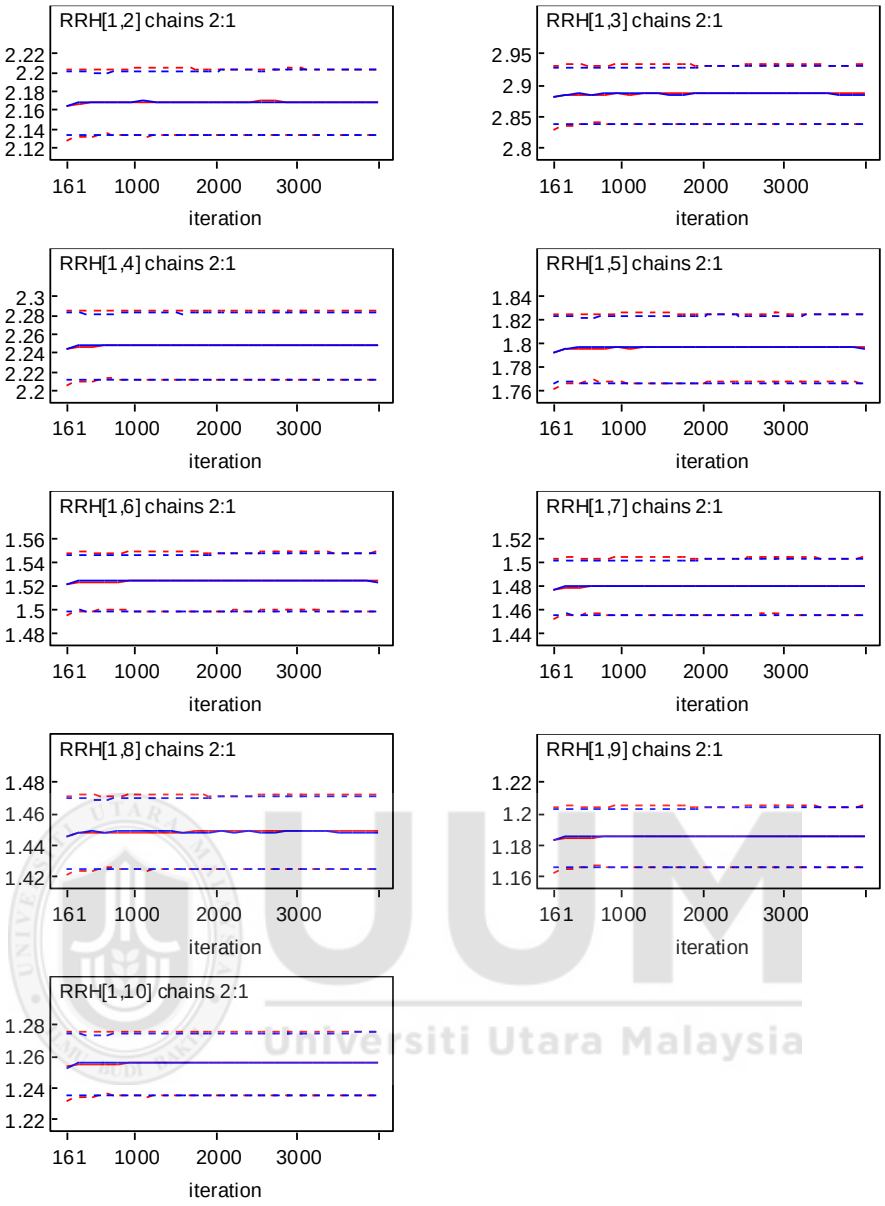

Figure AD.2. Quantile graph based on stochastic SIR model for the state of Perlis.

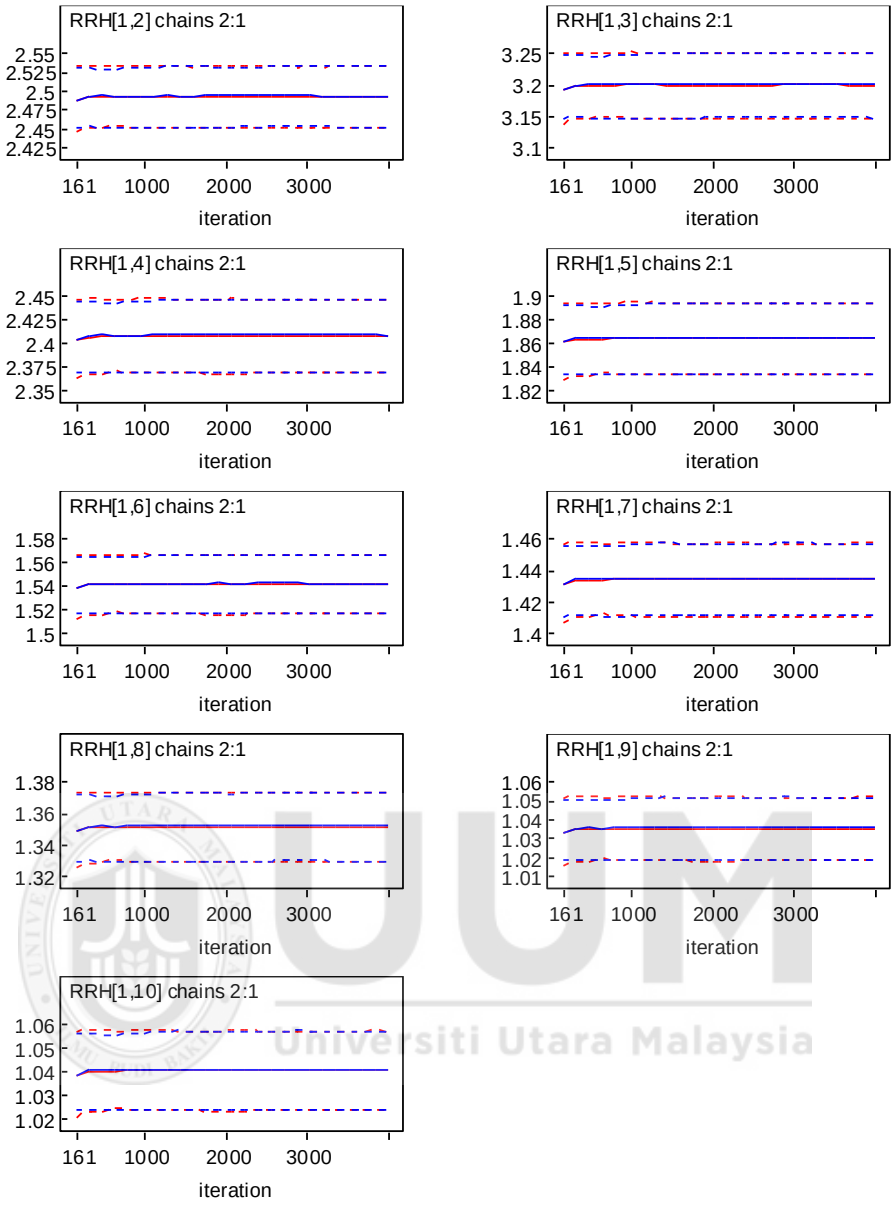

Figure AD.3. Quantile graph based on stochastic SCIR model for the state of Perlis.

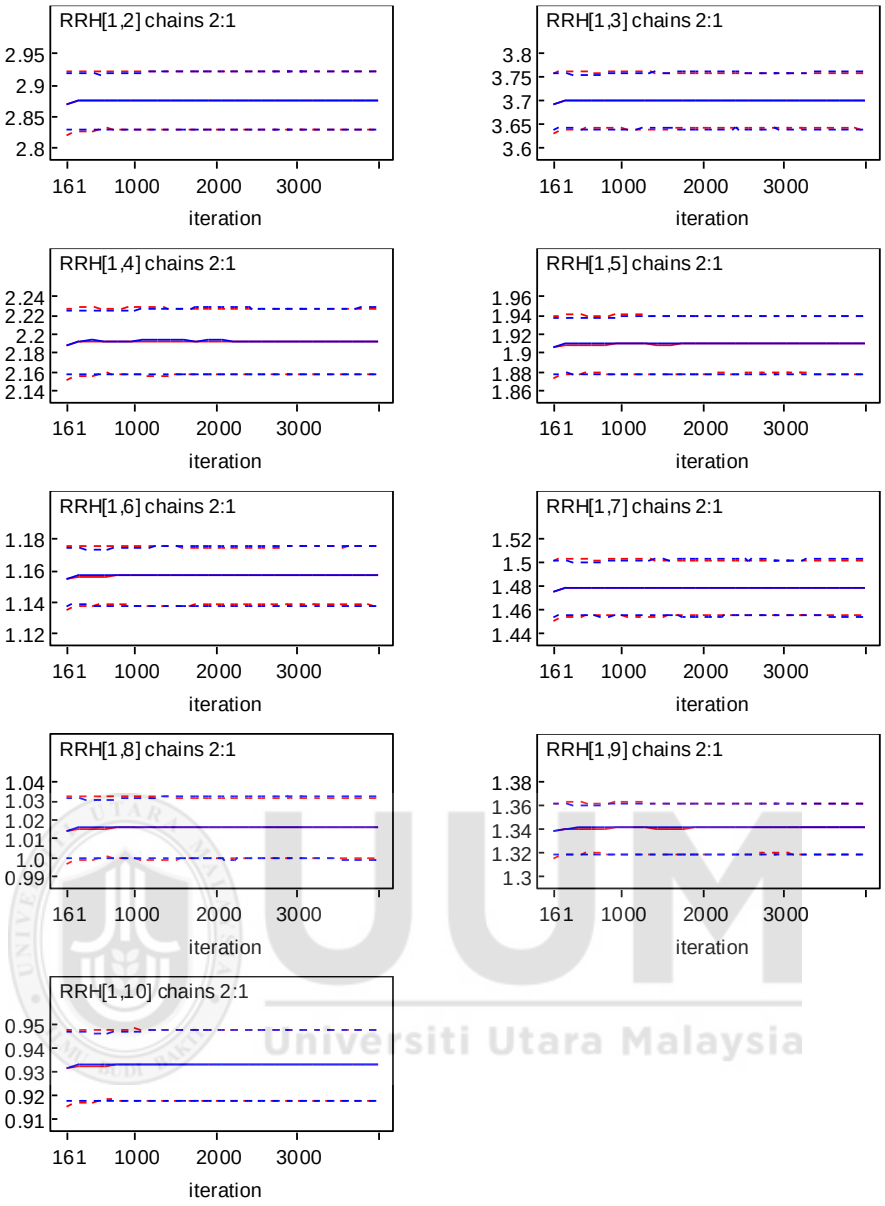

Figure AD.4. Quantile graph based on stochastic SVCIR model for the state of Perlis.

Appendix E

WinBUGS Outputs of the Posterior Densities of the Relative Risk Estimation

AE-1: WinBUGS outputs of the posterior densities based on SIC model, SIR model, SCIR model and SVCIR model for the state of Perlis only.

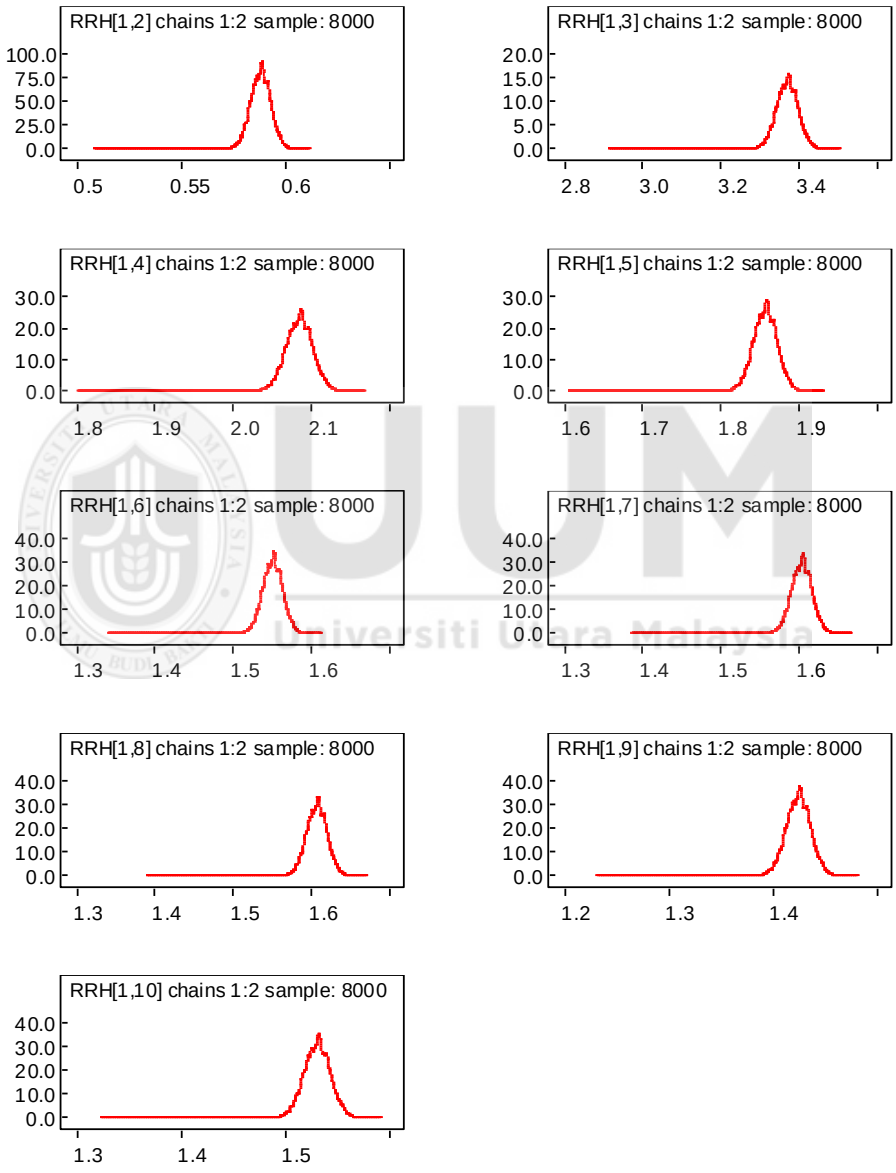

Figure AE.1. Posterior densities based on stochastic SIC model for the state of Perlis.

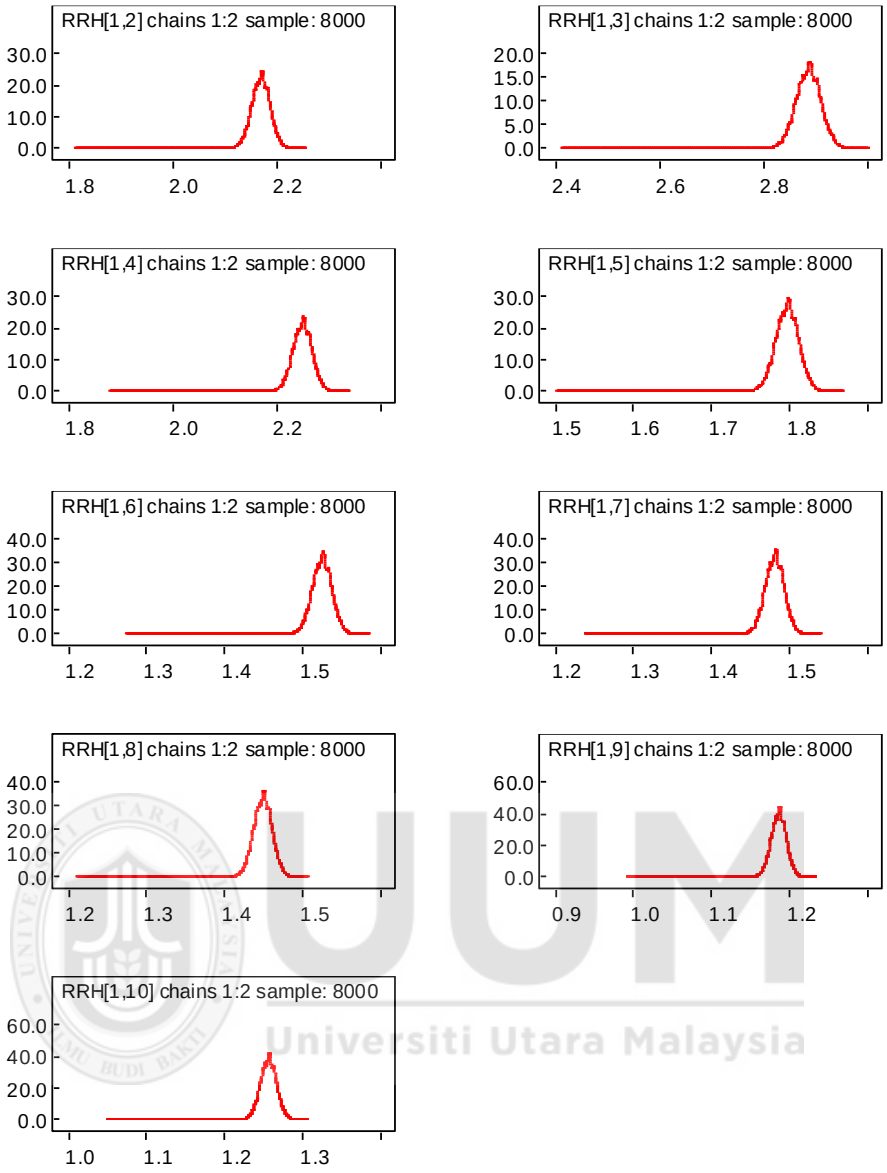

Figure AE.2. Posterior densities based on stochastic SIR model for the state of Perlis.

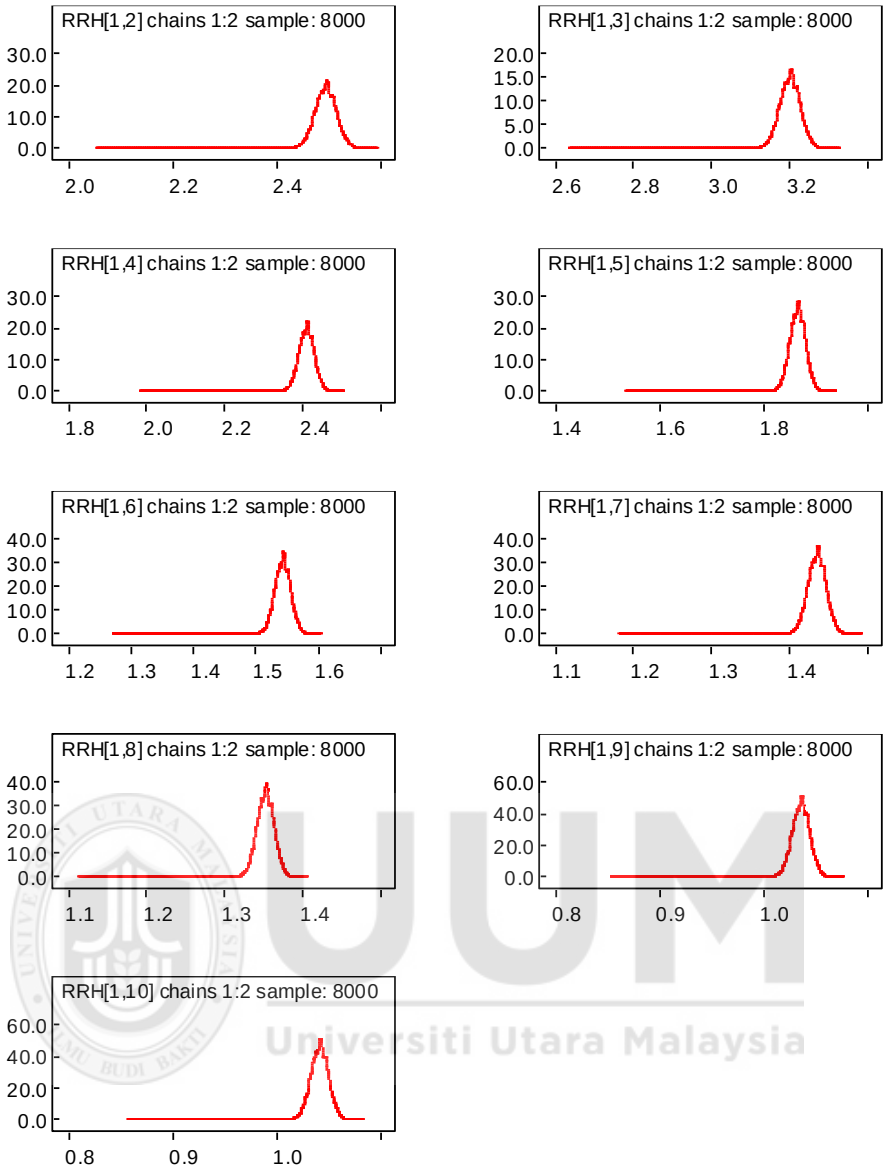

Figure AE.3. Posterior densities based on stochastic SCIR model for the state of Perlis.

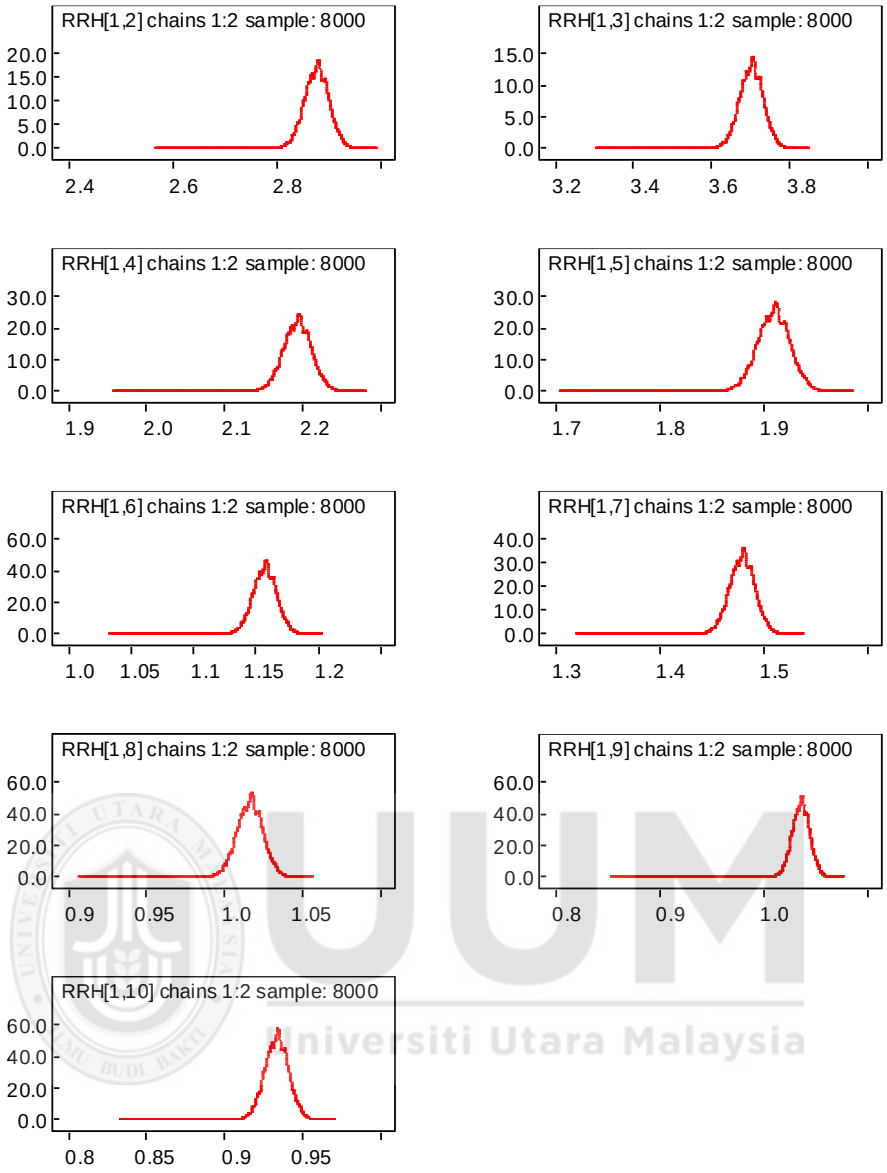

Figure AE.4. Posterior densities based on stochastic SVCIR model for the state of Perlis.

Appendix F

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

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

```
model
{
for (i in 1:M){
        for (j in 1:T)
        {
        # Poisson likelihood for observed counts
        y[i,j]~dpois(mu[i,j])
        log(mu[i,j])<-log(e[i,j])+alpha+u[i]+v[i]+beta*t[j]+delta[i]*t[j]
        # Relative Risk
        theta[i,j]<-exp(alpha+u[i]+v[i]+beta*t[j]+delta[i]*t[j])
        }
        theta_area[i]<-exp(u[i]+v[i])
        TT[i]<-exp(beta+delta[i])
}
#CAR distribution for the spatial correlated heterogeneity
u[1:M]~car.normal(adjH[], weightsH[], numH[], tau.u)
delta[1:M]~car.normal(adjH[], weightsH[], numH[], tau.delta)

# Posterior distributions for the uncorrelated heterogeneity
for(i in 1:M)
{
v[i]~dnorm(0,tau.v)
}
#weights
for (k in 1:SumNumNeighH)
{
weightsH[k]<-1
}
#improper distribution for the mean relative risk in the study region
alpha~dflat()
mean<-exp(alpha)

#hyperprior distributions on inverse variance parameters of random effects
beta~dnorm(0,1.0E-5)
tau.u~dgamma(0.5,0.0005)
tau.v~dgamma(0.5,0.0005)
tau.delta~dgamma(0.5,0.0005)
}
```

Figure AC.3. BYM model in WinBUGS

The code for SMR method and Poisson-gamma model were adapted from Samat and Ma'arof (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. BYM model's code also was written in Lawson et al. (2003) and was applied to analyze respiratory cancer mortality in Ohio.