THE EVALUATION OF STOCHASTIC MOSQUITO-BORNE INFECTIOUS DISEASE MAPPING MODEL IN MALAYSIA

SYAFIQAH HUSNA BINTI MOHD IMAM MA’AROF

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF MASTER OF SCIENCE (APPLIED STATISTICS)
(MASTER BY RESEARCH)

FACULTY OF SCIENCE AND MATHEMATICS
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2016

ABSTRACT

The main aim of this research is to propose an alternative method for estimating relative risk of mosquito-borne infectious disease mapping in Malaysia. This estimation method is based on stochastic SIR-SI model (*susceptible-infective-recovered* for human populations; *susceptible-infective* for mosquito populations). Based on previous research, relative risk estimation used total number of new infective cases at the denominator of the relative risk equation. In this research, relative risk estimation with the used of total posterior mean number of new infective cases at the denominator of the relative risk equation is introduced. This model is then been applied using data of mosquito-borne infectious disease from Malaysia. Data used in this research is provided by the Ministry of Health Malaysia which include 43339 dengue cases, 7 chikungunya cases and 3850 malaria cases for 16 states in Malaysia during year 2013. The findings revealed that the alternative method can overcome the problem of relative risk estimation, especially when there is no observed cases in certain regions. This is because this model considers spatial correlation between adjacent areas, takes into account the disease transmission model, as well as enable covariates adjustment which are known will contribute to the increasing number of disease cases. Pattern displayed on disease risk map proposed in this research can be used as reference by the authorities in preventing and reducing the cases of dengue, chikungunya and malaria in Malaysia.

PENILAIAN MODEL PEMETAAN STOKASTIK PENYAKIT BERJANGKIT BAWAAN NYAMUK DI MALAYSIA

ABSTRAK

Kajian ini bertujuan memperkenalkan satu kaedah alternatif bagi menganggar risiko relatif untuk pemetaan penyakit berjangkit bawaan nyamuk di Malaysia. Model anggaran ini adalah berdasarkan model stokastik SIR-SI (*susceptible-infective-recovered* bagi populasi manusia; *susceptible-infective* bagi populasi nyamuk). Berdasarkan kajian terdahulu, anggaran risiko relatif menggunakan jumlah bilangan bagi kes jangkitan baru di penyebut persamaan risiko relatif. Dalam kajian ini, anggaran risiko relatif menggunakan jumlah min posterior bagi kes jangkitan baru di penyebut persamaan risiko relatif diperkenalkan. Model ini kemudiannya diaplikasikan menggunakan data penyakit berjangkit bawaan nyamuk di Malaysia. Data yang digunakan dalam kajian ini diperolehi daripada Kementerian Kesihatan Malaysia yang melibatkan 43339 kes denggi, 7 kes chikungunya dan 3850 kes malaria bagi 16 buah negeri di Malaysia sepanjang tahun 2013. Hasil kajian menunjukkan bahawa kaedah alternatif yang diperkenalkan ini boleh mengatasi masalah anggaran risiko relatif, terutama apabila tiada kes yang dilaporkan di kawasan tertentu. Ini kerana model ini mengambilkira korelasi ruang antara kawasan bersebelahan, mengambilkira model serakan penyakit, serta membenarkan penambahan kovariat yang diketahui menyumbang kepada kenaikan bilangan kes penyakit. Pola yang ditunjukkan pada peta risiko penyakit yang dihasilkan menerusi kajian ini boleh dijadikan panduan oleh badan berkaitan dalam mengelakkan serta mengurangkan kejadian kes denggi, chikungunya dan malaria di Malaysia.

TABLE OF CONTENT

CHAPTER 5 APPLICATION TO RELATIVE RISK ESTIMATION

CHAPTER 6 CONCLUSION AND RECOMMENDATIONS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASR	Age-Standardized Mortality Rate
CAR	Conditional Autoregressive
CDC	Centers for Disease Control and Prevention
CBC	Complete blood count
DIC	Deviance Information Criterion
GIS	Geographic Information System
MOH	Ministry of Health Malaysia
SEIR	Susceptible-Exposed-Infective-Removed or Recovered for human populations
SEIR-SI	Susceptible-Exposed-Infective-Removed or Recovered for human populations; Susceptible-Infective for vector populations
SI	Susceptible-Infective for human populations
SIR	Susceptible-Infective-Recovered for human populations
SIR-SI	Susceptible-Infective-Recovered for human populations; Susceptible-Infective for vector populations
SIRS	Susceptible-Infective-Removed or Recovered-Susceptible for human populations
SIRS-SI	Susceptible-Infective-Removed or Recovered-Susceptible for human populations; Susceptible-Infective for vector populations
SIS	Susceptible-Infective-Susceptible for human populations
SMR	Standardized Morbidity Ratio
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Research Background

The mapping of disease incidence and prevalence has long been a part of public health, epidemiology and the study of disease in human populations (Koch, 2005). Factors like booming population, rapid urbanization, environmental pollution and global warming may influence the condition of disease outbreaks. Disease studies have revealed that strong spatial aspects involve disease diffusion and disease case location. Thus, it could help people to understand some disease outbreak by using disease mapping.

Basically, disease mapping is a method that can be used to show the geographical distribution of disease occurrence. It presents the incidence of specified diseases in areas of interest which involves the usage and interpretation of coloured or shaded maps. Unlike the raw disease data, disease mapping can offer a visual means of identifying the cause of the disease. For instance, in the study conducted by Khormi and Kumar (2011), they produced a risk map that showed most of the population was located within two kilometers of potential malaria carrying mosquito breeding grounds. The use of relative risk maps helps decision makers to keep those areas under intensive treatment. The production of such maps relies on the modelling to estimate and predict the risks. Hence, better estimation and prediction of risk would subsequently produce more accurate maps of disease risk.

This research is concerned with the study of the geographical distribution of mosquito-borne infectious disease in Malaysia. According to Lawson (2001), there are three main areas of application in disease study, which are disease mapping, disease clustering and ecological analysis. In disease mapping, the purpose of the analysis is to estimate the true relative risk of a disease of interest across a geographical study area. The primary objective is to reduce the noise in a disease map. In contrast, the purpose of the analysis in disease clustering is to study possible environmental hazards based on whether a disease map reveals localized clusters of cases. It is also used to determine the locations of any such clusters, which are important in public health surveillance. Thirdly, ecological analysis focuses on the analysis of geographical distributions of disease in relation to explanatory variables, where it can identify various supplementary issues relating to disease mapping.

One of the most significant topics in the field of geographical distributions of disease occurrence or disease mapping is relative risk estimation. As stated by Thomas et al. (2004), the estimation of infectious disease risk is considered as the first order. Most of these articles involve exploratory analysis, which includes the study of geographic information systems to integrate patient-related information and the study of covariates and their effects on disease distribution. Therefore, the main aim of this research is to review and improve the model proposed by Samat and Percy (2012) by introducing an alternative method of relative risk estimation based on the stochastic SIR-SI model (Susceptible-Infective-Recovered for human populations; Susceptible-Infective for mosquito populations) for the transmission of mosquito-borne infectious diseases.

At the beginning of the study, the compartmental and deterministic SIR-SI models that are suitable for mosquito-borne infectious disease transmission are discussed. Then, this model is improvised to develop the corresponding stochastic SIR-SI model by using discrete time-space data. The stochastic models include the extensions of the discrete time-space stochastic SIR-SI model proposed by Samat and Percy (2012), which requires the theoretical construction and iterative evaluation of SIR-SI difference equations.

At the end of this research, a new improved stochastic mosquito-borne infectious disease transmission model for disease mapping is proposed. It is expected that this improved method of relative risk estimation could give more accurate disease maps for mosquito-borne infectious diseases. In Malaysia, three mosquito-borne diseases have been identified which are dengue, chikungunya and malaria. Detailed characteristics of all diseases will be discussed in Chapter 2. In this study, dengue and chikungunya might give a similar pattern of disease occurrences in maps since the viruses for both diseases are transmitted by the same type of mosquito called *Aedes Aegypti*.

1.2 Problem Statement

Numerous studies of disease mapping use a regression-type model, where observable (fixed effects) and unobservable (random effects) variables are included to give a clean map to display the true excess risk surface (MacNab and Dean, 2002). However, only several studies use disease transmission models for disease mapping (Gemperli et al., 2006). There are also studies that use a geographic information system (GIS) to incorporate patient-related information (Mohd Din et al., 2007). GIS provides an effective way of managing, storing, analyzing and mapping the information of disease.

Current approaches for the identification of the high and low risk areas in Malaysia are still based on the total numbers of disease occurrences across the regions. Large numbers of cases in certain areas correspond to the high risks of disease occurrences without considering other factors such as the population size or the land size of individual areas.

In the study of disease mapping, there are two common approaches used to produce disease maps, which are known as the SMR method and the Poisson-Gamma model. The SMR basically compares the observed incidence with the expected incidence, which has been used traditionally for the analysis of counts within tracts as mentioned by Lawson (2006). This method is used to estimate the relative risk, which may be interpreted as the probability that an individual within a specified region contracts the disease divided by the probability that an individual in the population contracts the disease. Details information of SMR method are discussed in Section 2.3.2. The drawbacks of SMR method have led many researchers to explore another method for relative risk estimation which include the use of Bayesian method. The earliest examples of Bayesian method known as the Poisson-Gamma model. Details information of Poisson-Gamma model are discussed in Section 2.3.2.

However, the Poisson-Gamma model also have its own weakness since there is no possibility for allowing spatial correlation between risks in neighboring areas. The weaknesses of SMR method and Poisson-Gamma model has motivated this study to introduce an alternative method which consider the stochastic model. Some studies have emphasized the effects that stochasticity can have on transmission of a disease (Keeling and Ross, 2008). Most studies show that the use of stochastic terms in the analysis of infectious diseases allows for a probabilistic statement about the progression of the disease to be made which is generally agreed that this renders the modelling as more realistic.

Therefore, this study is interested to introduce an alternative method of relative risk estimation of mosquito-borne infectious disease based on the method proposed by Samat and Percy (2012) in order to identify the risk pattern on the map. This study will improvise the model proposed by Samat and Percy (2012) in order to give better relative risk estimation method.

1.3 Research Objectives

The objectives of this research are as follows;

1. To compare the estimated relative risk for dengue, chikungunya and malaria in identifying the high and low risk distributions between diseases.
2. To map the disease risk for dengue and chikungunya in order to investigate whether the same type of mosquito will give the same risk distribution on map.
3. To propose a new alternative method of estimating the 'expected cases' of the denominator for the relative risk equation, which subsequently introduced a better model of relative risk estimation.
4. To propose risk maps for mosquito-borne infectious diseases in Malaysia.

1.4 Research Questions

There are five research questions that can be relate to the objectives of this research. The research questions include;

i. Is there any difference of estimated relative risk and disease risk appearance on maps between these three diseases?
ii. Does the same type of mosquito gives similar risk distribution on map?
iii. How to estimate the new alternative method of 'expected cases'?

iv. Does the new alternative method of ‘expected cases’ estimation can overcome the problem of the relative risk estimation as proposed by Samat and Percy (2012)?

v. What are the differences of disease risk map between these three diseases based on four different methods?

1.5 Significance of the Research

This research is concerned about three mosquito-borne infectious diseases in Malaysia, which are dengue, chikungunya and malaria. There are some similarities and differences of these three diseases as discussed in Chapter 2. At the end of this research, the reader may differentiate the types and symptoms of mosquito-borne infectious disease. The production of a new alternative method to estimate the relative risk is considered stochastic element in the formulation of disease transmission in human population. The significance of this stochastic model can overcome the drawbacks of the SMR method and the Poisson-Gamma model.

This study also concerns about an alternative method in calculating the expected cases of the denominator for the relative risk equation. This method can overcome the problem of the expected cases estimation, especially when there is no observed case in certain regions. This new approach in relative risk estimation is a better and improved method compared to the SMR method and the Poisson-Gamma model because it considers the randomness in the stochastic elements of disease transmissions.

In this study, the production of disease maps will display the high and low risk areas of disease occurrences. Consequently, the maps might help the authorities to identify the state that deserves closer scrutiny or more attention. They may also allocate the resources or identify hazards related to disease. As for citizens, they can use the maps to identify whether they are at high or low risk areas of disease occurrences. If their area has high possibilities of contracting by mosquito-borne infectious disease, they may take extra precaution to protect themselves from disease infection.

1.6 Organization of the Report

This report is organized into six chapters. Chapter 1 introduces the background of the study and the problem statement involved in this research. Then, the objectives and the research questions for this study are stated in this chapter. This report also includes the significance of the research.

Next, Chapter 2 gives a brief explanation about the three diseases involved in this study, which are dengue, chikungunya and malaria. Detailed characteristics of these three diseases are discussed, including the signs and symptoms, transmission of the disease, and the prevention and control of the disease. Then, the differences between these three diseases are explained. In addition, this chapter also discusses the previous study on modelling disease transmission and the analysis of disease mapping, which consider the case-event data analysis and track-count data analysis. The explanation of the most common method used in the study of disease mapping called the Standardized Morbidity Ratio (SMR) method is included. The discussion on the earlier Bayesian approach, which is the Poisson-Gamma model, is also included in this chapter. The measurement of the disease frequency are also be discussed in this chapter. These include the explanation of the prevalence and incidence, which are the main types of disease frequency. In addition, a brief explanation about the standardization of rates and ratios are also be included.

Chapter 3 explains the methodology of the research and the models used for the mosquito-borne infectious disease transmission. This includes the compartmental SIR-SI model, deterministic SIR-SI model for disease transmission and the discrete time-space stochastic SIR-SI model as proposed by Samat and Percy (2012).

After that, Chapter 4 begins with the discussion of the interpretation of the relative risk value. This is followed by the explanation of the relative risk estimation based on the SMR method and the Poisson-Gamma model, which are the earlier method used in the epidemiology studies. The last section in this chapter discusses the relative risk estimation based on the stochastic SIR-SI model. Two types of the stochastic SIR-SI models are discussed. The first model is the relative risk estimation based on the stochastic SIR-SI model with the use of total number of new infective cases at denominator. This model is previously discussed in details by Samat and Percy (2012). The second model is the relative risk estimation based on the stochastic SIR-SI model with the use of total posterior mean number of new infective cases at the denominator. This newly improved model is mainly be used in the methodology of this research.

Then, Chapter 5 focuses on the application to relative risk estimation for three types of mosquito-borne infectious diseases that has been the focus for this research. These include the number of dengue, chikungunya and malaria cases for the year of 2013 in Malaysia. The WinBUGS codes program for the relative risk estimation based on the stochastic SIR-SI model is also been presented in this chapter. The explanations for each parameter also included in this section. The next section discusses the numbering for CAR model of every state in Malaysia that has been used throughout this research for the purpose of data analysis. This chapter also presents and demonstrates the results from the analysis of relative risk estimation based on the SMR method, the Poisson-Gamma model, the stochastic SIR-SI Model 1 and the stochastic SIR-SI Model 2 using dengue, chikungunya and malaria disease data, respectively. All related findings are presented and compared in tables, graphs and maps.

Finally, the conclusions of the research are discussed in Chapter 6. This chapter includes the research contributions and recommendations for future works in order to improve this research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter is divided into four sections. The first section briefly explains the facts concerning three types of mosquito-borne infectious disease in Malaysia, which are dengue, chikungunya and malaria. Meanwhile, the second section discusses previous studies related to the modelling and analysis of disease transmission mechanism for disease mapping. Then, the third section discusses the existing methods of measuring the disease frequency which include prevalence, risk, odds and incidence rate. Finally, the last section presents the methods of standardization of rates and ratios that have been used by previous researchers.

2.2 Mosquito-Borne Infectious Disease

Mosquitoes cause more human suffering than any other organism. Over one million people worldwide died from mosquito-borne diseases every year. Generally, mosquitoes go through four stages in their life cycle development (Figure 2.1).

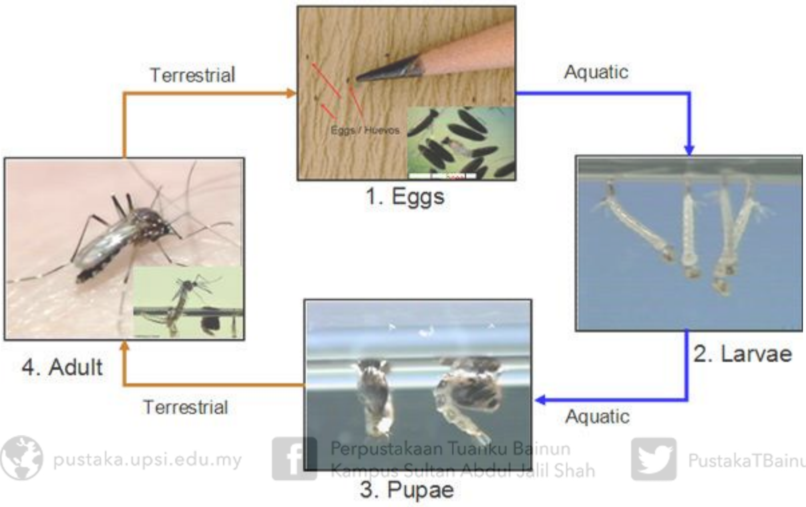

Figure 2.1. Mosquito life cycle (Source: CDC, 2012)

Female mosquitoes lay their eggs in the water container. The larva hatches when water inundates the eggs caused by rains or the addition of water from people. The larva will feed on microorganisms in the following days and then their skins will shed. The larva will change to pupa when it has acquired enough energy and size. Pupa does not feed. It is the body transformation until flying mosquito is formed. After breaking the pupa skin, the newly formed adult emerges from the water. Normally, if the weather conditions are suitable for mosquito development, it takes between 7 to 14 days for an egg to develop into an adult.