Relative Risk Estimation for Malaria Disease Mapping in Malaysia based on Stochastic SIR-SI Model

Penganggaran Risiko Relatif bagi Pemetaan Penyakit Malaria di Malaysia Menggunakan Model Stokastik SIR-SI

Syafiqah Husna Mohd Imam Ma'arof [1] & Nor Azah Samat [2]
Department of Mathematics, Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia
e-mail: [1] shusna89@gmail.com, [2] norazah@fsmt.upsi.edu.my

Abstract

Disease mapping is a study on the geographical distribution of a disease to represent the epidemiology data spatially. The production of maps is important to identify areas that deserve closer scrutiny or more attention. In this study, a mosquito-borne disease called Malaria is the focus of our application. Malaria disease is caused by parasites of the genus Plasmodium and is transmitted to people through the bites of infected female Anopheles mosquitoes. Precautionary steps need to be considered in order to avoid the malaria virus from spreading around the world, especially in the tropical and subtropical countries, which would subsequently increase the number of Malaria cases. Thus, the purpose of this paper is to discuss a stochastic model employed to estimate the relative risk of malaria disease in Malaysia. The outcomes of the analysis include a Malaria risk map for all 13 states and 3 federal territories in Malaysia, revealing the high and low risk areas of Malaria occurrences.

Keywords malaria disease, disease mapping, relative risk estimation, SIR-SI model, stochastic model

Abstrak

Pemetaan penyakit ialah suatu bidang kajian terhadap taburan penyakit secara geografi bagi memaparkan data epidemiologi. Peta risiko penyakit adalah penting dan boleh digunakan dalam pengenalpastian kawasan yang perlu pemerhatian dan perhatian lanjut. Dalam kajian ini, penyakit bawaan nyamuk yang dikenali sebagai Malaria adalah merupakan fokus aplikasi yang akan dijalankan. Penyakit Malaria adalah sejenis penyakit yang disebabkan oleh parasit dari jenis Plasmodium dan disebarkan menerusi gigitan nyamuk Anopheles betina. Langkah berjaga-jaga perlu diambil bagi mengelakkan virus Malaria daripada merebak ke seluruh dunia, khususnya bagi negara tropika dan sub-tropika, yang boleh menyumbang kepada peningkatan dalam bilangan kes Malaria. Oleh itu, kajian ini adalah bertujuan untuk membincangkan model stokastik yang boleh digunakan untuk menganggar risiko relatif bagi penyakit Malaria di Malaysia. Hasil analisis merangkumi peta risiko malaria bagi kesemua 13 negeri dan 3 wilayah persekutuan di Malaysia yang akan mendedahkan kawasan yang mempunyai risiko tinggi dan rendah bagi kejadian Malaria.

Kata kunci penyakit malaria, pemetaan penyakit, penganggaran risiko relatif, model SIR-SI, model stokastik

INTRODUCTION

Generally, disease mapping is a method to show the geographical distribution of disease occurrence. Disease map displays the high and low disease risk incidences of the specified diseases in the regions of interest. This paper is concerned with the geographical distribution of malaria disease in Malaysia. Relative risk estimation is one of the most significant topics in the field of geographical distributions of disease occurrence or disease mapping. According to Thomas et al. (2004), the estimation of infectious disease risk is considered as the first order. The aim of this paper is to review the model proposed by Samat and Percy (2012) and apply the stochastic SIR-SI model to the malaria data in Malaysia in 2013.

Malaria Disease and Its Situation in Malaysia

Based on World Health Organization (WHO) (2013), in 2012, there were about 207 million cases of malaria, which accounted for 62700 deaths over the world. The eminence of the death comes from children who live in Africa, where a child dies every minute due to malaria disease. Four types of parasite species caused the malaria disease in human, known as *Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae and Plasmodium ovale. Plasmodium falciparum*is is the most dangerous parasite, while *Plasmodium falciparum* and *Plasmodium vivax* are the most common parasites. Figure 1 displays the trends of reported malaria incidence around the world from the year 2000 to 2011. In 2011, about 3.3 billion people were at risk of malaria disease, with the populations in sub-Saharan Africa having the highest risk for malaria infection.

Malaria is caused by parasites that are sent to people through the bites of infected Anopheles mosquitoes. These malaria vectors bite mainly between dusk and dawn. The intensity of disease transmission depends on elements related to the parasite, the vector, the human host, and the surroundings. The parasites multiply in the liver and then infect the red blood cells. The symptoms of malaria in non-immune individual usually appear between 10 and 15 days after the infectious mosquito bite. Early symptoms are fever, vexation, and vomiting. If not treated, malaria can quickly become life-threatening by disrupting the blood supply to vital organs, and causing death. The incubation period, which is the time between being bitten by the mosquito and developing the symptoms varies. Normally, it takes between two or three weeks.

In parliamentary law, to cut down the malaria transmission, vector control is the best way to be seen. As for the individual, the prevention step involves protecting ourselves against mosquito bites and taking anti-malarial medicines. Pregnant women and young children are strongly advised not to travel to areas where malaria viral is endemic. People must remain in the house when it is gloomy day outside to avoid mosquitoes' bites. Long pants and long-sleeved shirts must be worn for extra bodily protection. Other prevention steps include using repellent on the skin, bed nets and flying-insect spray around sleeping areas.

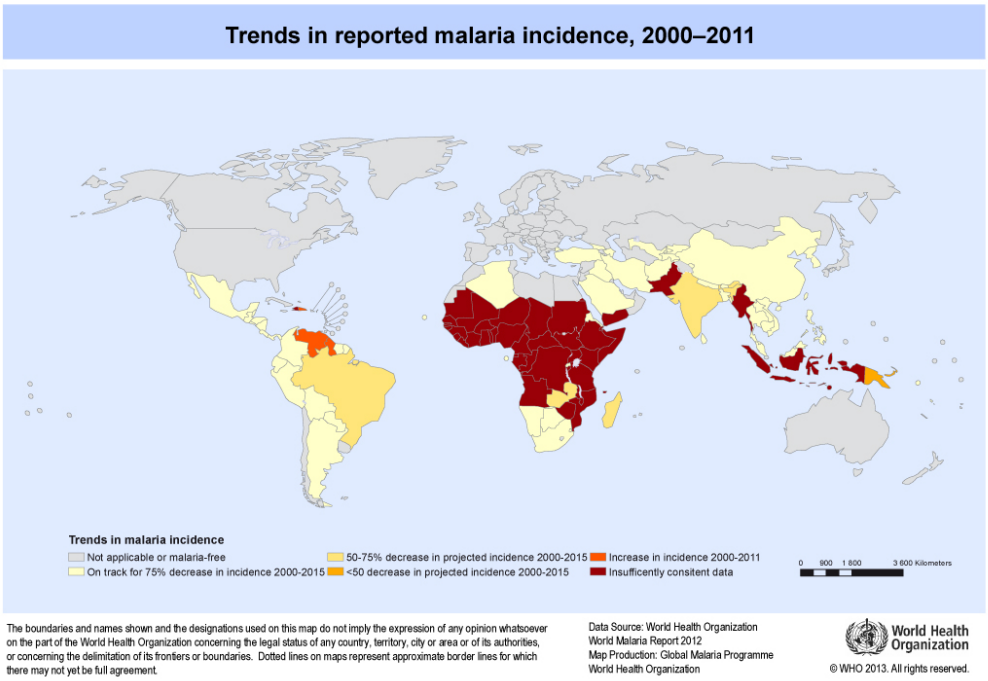

Figure 1 Trends in reported malaria incidence from 2000 to 2011 (Source: WHO, 2013)

In this paper, the data were provided by the Ministry of Health Malaysia, consisting of the observed count malaria data for every state in Malaysia from epidemiology week 1 to epidemiology week 52 in the year of 2013. Figure 2 represents the total number of malaria cases for all 13 states and 3 federal torritories in Malaysia from epidemiology week 1 to epidemiology week 52 in the year of 2013. Sabah had the highest total of malaria cases, which was 1606 cases in 2013, followed by Sarawak with 1004 cases. Two states had zero case, which were Putrajaya and Labuan. There is only one malaria case in Perlis during 2013.

Furthermore, Figure 3 depicts the time series plot of malaria disease based on the number of cases for each state in Malaysia from epidemiology week 1 to epidemiology week 52. From the plot, we can see in every epidemiology week, Sabah and Sarawak had the highest number of cases. Most of the cases were above 10 for every epidemiology week for both states. As for Kelantan, there were sudden increases in the number of cases registered between epidemiology weeks 20 and 27. The number of cases went down in epidemiology week 22 with 11 cases and increased again in epidemiology week 23 with 48 cases. While for other states, most had less than 10 cases for most of the epidemiology weeks in 2013.

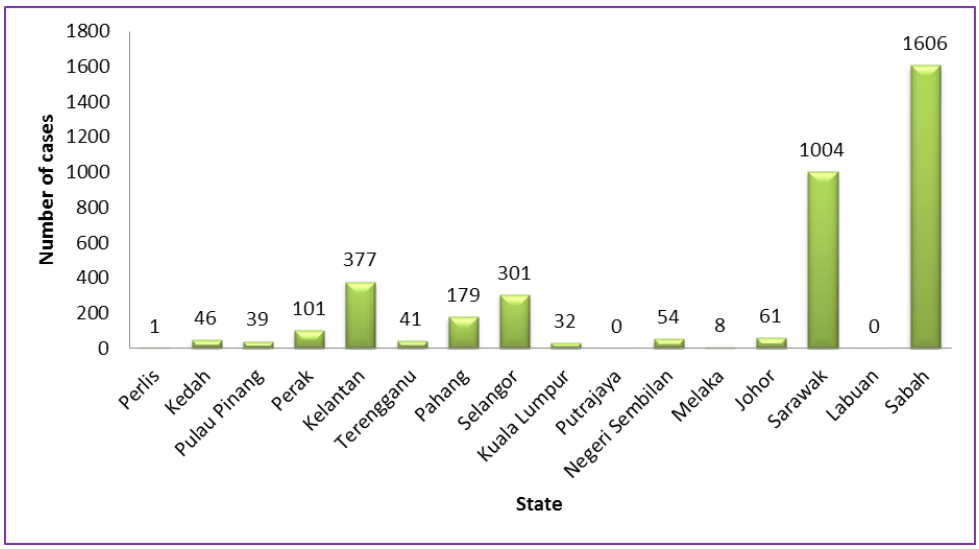

Figure 2 Total number of malaria cases in all states in Malaysia

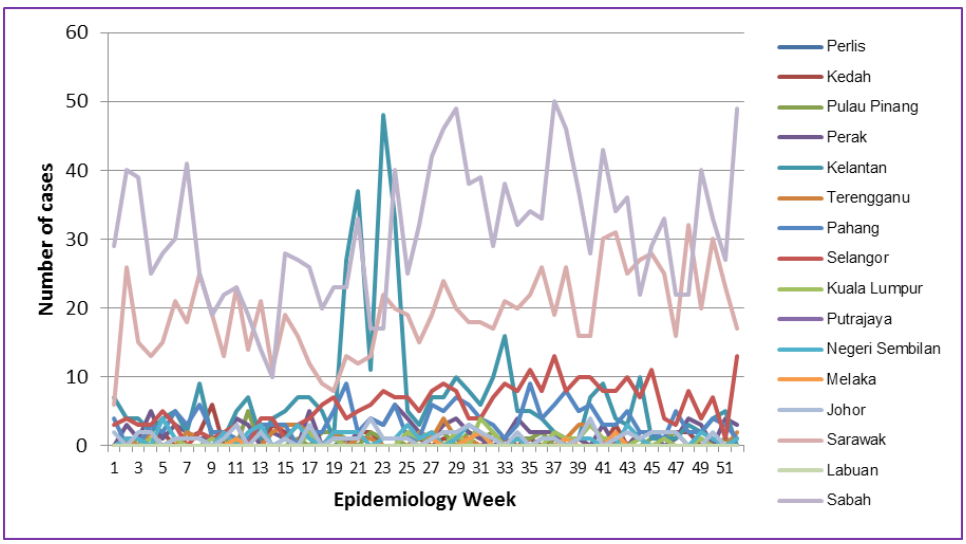

Figure 3 Time series plot based on number of cases for each state

Compartmental Model

The compartmental model displayed in Figure 4 represents the most common compartmental model used in the study of vector-borne infectious disease transmission as proposed by Samat and Percy (2012), which was adapted from Esteva and Vargas (1998) and Nishiura (2006). In this study, $i = 1, 2, ..., M$ represents study regions and $j = 1, 2, ..., T$ represents time periods.

The definition of notations in Figure 4 are as follows:

$S_{i,j}^{(h)}$: The total number of susceptible humans at time *j*,

$I_{i,j}^{(h)}$: The total number of infective humans at time *j*,

$R_{i,j}^{(h)}$: The total number of recovered humans at time *j*,

$S_{i,j}^{(v)}$: The total number of susceptible mosquitoes at time *j,*

$I_{i,j}^{(v)}$: The total number of infective mosquitoes at time *j,*

$\mu^{(h)}$ and $\mu^{(v)}$: The (assumed equal) death and birth rates of humans per week and the (assumed equal) death and birth rates of mosquitoes per week, respectively,

$\gamma^{(h)}$: The rate at which humans recover per week,

b : The biting rate per week,

m : The number of alternative hosts available as the blood source,

A : The constant recruitment rate for the mosquito vector,

$\beta^{(h)}$: The transmission probability from mosquitoes to humans,

$\beta^{(v)}$: The transmission probability from humans to mosquitoes,

$N_i^{(h)}$: The human population size,

$N_i^{(v)}$: The mosquito population size.

Deterministic Model

The compartmental model in Figure 4 can be written as a deterministic model. There is no stochastic element in the deterministic model and the entire input and output relation of the model is conclusively determined. The deterministic model for the vector-borne infectious disease transmission in human populations based on Figure 4 that are used in this study are as follows,

$$S_{i,j}^{(h)} = \mu^{(h)} N_i^{(h)} + \left\{ 1 - \mu^{(h)} - \left(\frac{\beta^{(h)} b}{N_i^{(h)} + m} \right) I_{i,j-1}^{(v)} \right\} S_{i,j-1}^{(h)} \tag{1}$$

$$I_{i,j}^{(h)} = (1 - \mu^{(h)} - \gamma^{(h)}) I_{i,j-1}^{(h)} + \left(\frac{\beta^{(h)} b}{N_i^{(h)} + m} \right) I_{i,j-1}^{(v)} S_{i,j-1}^{(h)} \tag{2}$$

$$R_{i,j}^{(h)} = (1 - \mu^{(h)}) R_{i,j-1}^{(h)} + \gamma^{(h)} I_{i,j-1}^{(h)} \tag{3}$$

While, the deterministic model for vector-borne infectious disease transmission in vector populations are as follows,

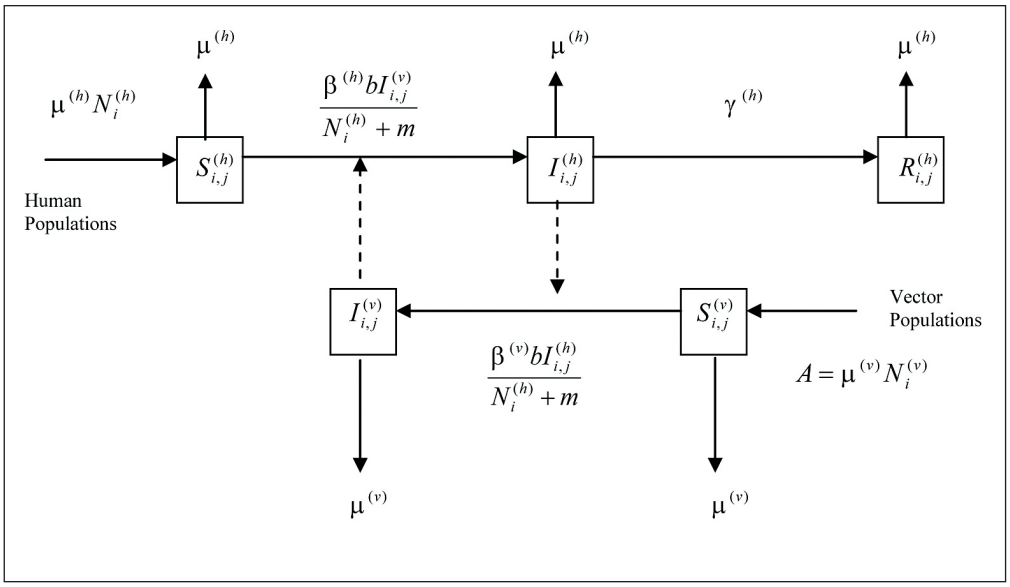

Figure 4 Compartmental SIR-SI model for vector-borne infectious disease transmission (Source: Samat and Percy, 2012)

$$S_{i,j}^{(v)} = \mu^{(v)}N_i^{(v)} + \left\{1-\mu^{(v)} - \left(\frac{\beta^{(v)}b}{N_i^{(h)}+m}\right)I_{i,j-1}^{(h)}\right\}S_{i,j-1}^{(v)} \tag{4}$$

$$I_{i,j}^{(v)} = (1-\mu^{(v)})I_{i,j-1}^{(v)} + \left(\frac{\beta^{(v)}b}{N_i^{(h)}+m}\right)I_{i,j-1}^{(h)}S_{i,j-1}^{(v)} \tag{5}$$

Then, these deterministic models are improved by assigning a probability distribution to reflect the randomness in the data, known as the stochastic model. This is discussed further in the next section.

Stochastic SIR-SI Model

According to Isham (2005), a deterministic analysis provides a good approximation for the stochastic means for a major outbreak when the sample size is large. Thus, in this report, deterministic models formulated above were applied to provide an estimate for the stochastic means. These stochastic models were applied on the malaria disease data in Malaysia to estimate the relative risks.

In the study conducted by Samat and Percy (2012), they included the terms $\mathfrak{I}_{i,j}^{(h)}$, $\mathfrak{R}_{i,j}^{(h)}$ and $\mathfrak{I}_{i,j}^{(v)}$ to represent the number of newly infective humans, newly recovered humans and newly infective mosquitoes, respectively, all in the interval or time period $(j-1, j]$ and the study region (i). In this paper, the same notations are used.

For $i = 1, 2, ..., M$ representing the study regions and $j = 1, 2, ..., T$ representing the time periods, the stochastic SIR-SI model for vector-borne infectious disease transmission in human population adapted the Equations (1) to (5) and included a probability distribution

to reflect the randomness inherent in the data as follows:

$$S_{i,j}^{(\mathrm{h})} = \mu^{(\mathrm{h})}\, N_i^{(\mathrm{h})} + (1 - \mu^{(\mathrm{h})})\, S_{i,j-1}^{(\mathrm{h})} - \Im_{i,j}^{(\mathrm{h})}, \tag{6}$$

$$\Im_{i,j}^{(\mathrm{h})} \sim \text{Poisson}\ \left(\lambda_{i,j}^{(\mathrm{h})}\right), \tag{7}$$

$$\lambda_{i,j}^{(\mathrm{h})} = \exp\left(\beta_0^{(\mathrm{h})} + c_i^{(\mathrm{h})}\right)\left(\frac{\beta^{(\mathrm{h})}\, b}{N_i^{(\mathrm{h})} + m}\right) I_{i,j-1}^{(\mathrm{v})}\, S_{i,j-1}^{(\mathrm{h})}, \tag{8}$$

$$I_{i,j}^{(\mathrm{h})} = \left(1 - \mu^{(\mathrm{h})}\right) I_{i,j-1}^{(\mathrm{h})} + \Im_{i,j}^{(\mathrm{h})} - \Re_{i,j}^{(\mathrm{h})}, \tag{9}$$

$$R_{i,j}^{(\mathrm{h})} = \left(1 - \mu^{(\mathrm{h})}\right) R_{i,j-1}^{(\mathrm{h})} + \Re_{\mathrm{i\,j}}^{(h)}, \tag{10}$$

$$\Re_{i,j}^{(\mathrm{h})} = \gamma^{(\mathrm{h})}\, I_{i,j-1}^{(\mathrm{h})}. \tag{11}$$

Due to the unavailability of sufficient data for vectors, the stochastic SIR-SI model for malaria disease transmission in vector populations were assumed as non-stochastic as shown below:

$$S_{i,j}^{(\mathrm{v})} = \mu^{(\mathrm{v})}\, N_i^{(\mathrm{v})} + \left(1 - \mu^{(v)}\right) S_{i,j-1}^{(\mathrm{v})} - \Im_{\mathrm{i\,j}}^{(v)}, \tag{12}$$

$$\Im_{i,j}^{(\mathrm{v})} = \left(\frac{\beta^{(\mathrm{v})}\, b}{N_i^{(\mathrm{h})} + m}\right) I_{i,j-1}^{(\mathrm{h})}\, S_{i,j-1}^{(\mathrm{v})}, \tag{13}$$

$$I_{i,j}^{(\mathrm{v})} = \left(1 - \mu^{(\mathrm{v})}\right) I_{i,j-1}^{(\mathrm{v})} + \Im_{\mathrm{i\,j}}^{(\mathrm{v})}. \tag{14}$$

Relative Risk Estimation for Disease Mapping

The computational analysis was performed using WinBUGS software, which is a package designed to carry out Markov chain Monte Carlo computation for a wide variety of Bayesian Models as stated by Spiegelhalter *et al.* (2003). Samat and Percy (2012) introduced an alternative method in order to estimate the relative risk based on the disease transmission model. For the purpose of this paper, $i = 1, 2, ..., M$ represented the study regions and $j = 1, 2, ..., T$ represented the time periods. The respective relative risk parameter $\theta_{ij}^{(h)}$was defined by:

$$\theta_{ij}^{(h)} = \frac{\lambda_{ij}^{(h)}}{e_{ij}^{(h)}} \tag{15}$$

where $\lambda_{ij}^{(h)}$ is the expected mean number of infective human and $e_{ij}^{(h)}$ is the expected number of new infective. A pseudo-random sample of observations, $\lambda_{ijk}^{(\mathrm{h})}$ for $k = 1, 2, ..., n$ was generated from the posterior distribution for the mean number of infective human, $\lambda_{ij}^{(h)}$. The posterior expected mean number of infective human could be approximated using unbiased samples as follows:

$$\tilde{\lambda}_{ij}^{(h)} = \frac{1}{n}\sum_{k=1}^{n}\lambda_{ijk}^{(h)} \tag{16}$$

Hence, the posterior expected relative risk for human populations could also be approximated using unbiased samples mean as follows:

$$\tilde{\theta}_{ij}^{(h)} = \frac{1}{n}\sum_{k=1}^{n}\theta_{ijk}^{(h)} = \frac{1}{n}\sum_{k=1}^{n}\frac{\lambda_{ijk}^{(h)}}{e_{ij}^{(h)}} = \frac{\tilde{\lambda}_{ij}^{(h)}}{e_{ij}^{(h)}} \tag{17}$$

Simply, the posterior expected relative risk, $\tilde{\theta}_{ij}^{(h)}$ was equal to the posterior expected mean number of infective, $\tilde{\lambda}_{ij}^{(h)}$ divided by the corresponding expected number of new infective, $e_{ij}^{(h)}$ based on human population across all studied regions. These formulations were then applied to calculate the relative risk for disease mapping based on the stochastic SIR-SI model for disease transmission using the case counts for all tracts under consideration.

Application of Relative Risk Estimation for Malaria Disease Mapping

This section displays the result of relative risk estimation based on the application of stochastic SIR-SI model for malaria disease transmission in Malaysia. Figure 5 represents the time series plot for the estimated relative risk based on stochastic SIR-SI model for malaria disease in Malaysia in 2013. Figure 5 shows that Sabah had the highest number of relative risk. The second highest number of relative risk appeared in Sarawak. Both states had relative risks above 2 for all epidemiology weeks. These mean that people

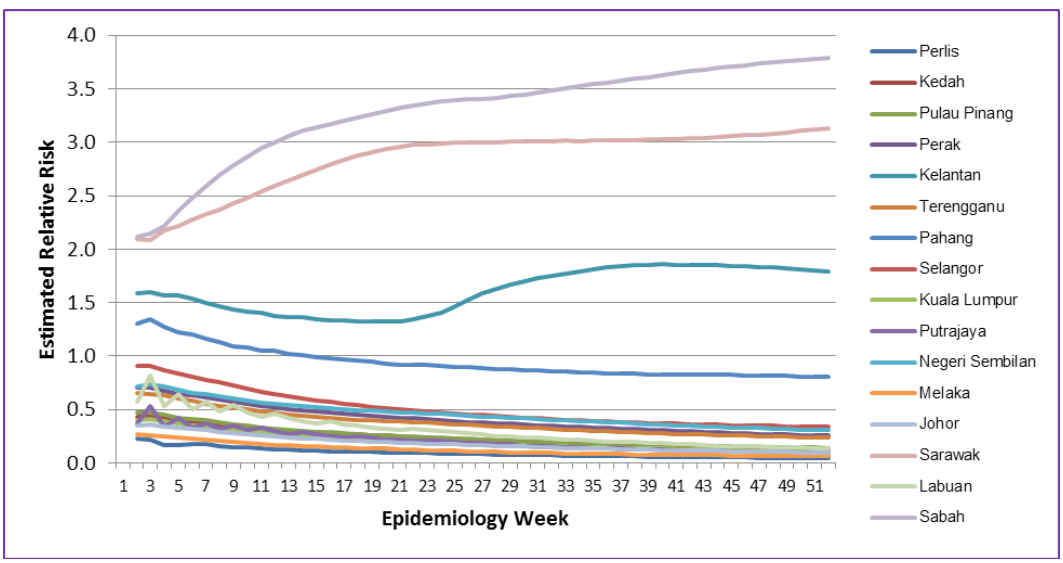

Figure 5 Time series plot of the estimated relative risk based on the Stochastic SIR-SI model

in these states are more likely to contract with malaria disease compared to people in overall population. Furthermore, the third state with the highest relative risk was Kelantan, followed by Pahang with risks close to 1 for most of epidemiology weeks. These mean that there is no real different in terms of risk between people in these states compared to people in overall population. While, the other 12 states had relative risks below 0.75 which means that people in these states are less likely to contract with malaria disease compared to people in overall population.

Risk Map for Malaria Disease Mapping in Malaysia

Five different levels of risks were used to categorize the relative risk values, which were very low, low, medium, high and very high with the intervals of (<0.5], (0.5-1.0], (1.0-1.5], (1.5-2.0] and (2.0>), respectively. The intervals were then used to display the state as having high or low risk area in a map. A very high risk area was represented by the darkest shade, while the lightest shade represented a very low risk area. Figure 6 depicts that Sarawak and Sabah were in very high risk area of malaria occurrences, which means that people in these states had very high risk to contract malaria disease compared to the overall population in Malaysia.

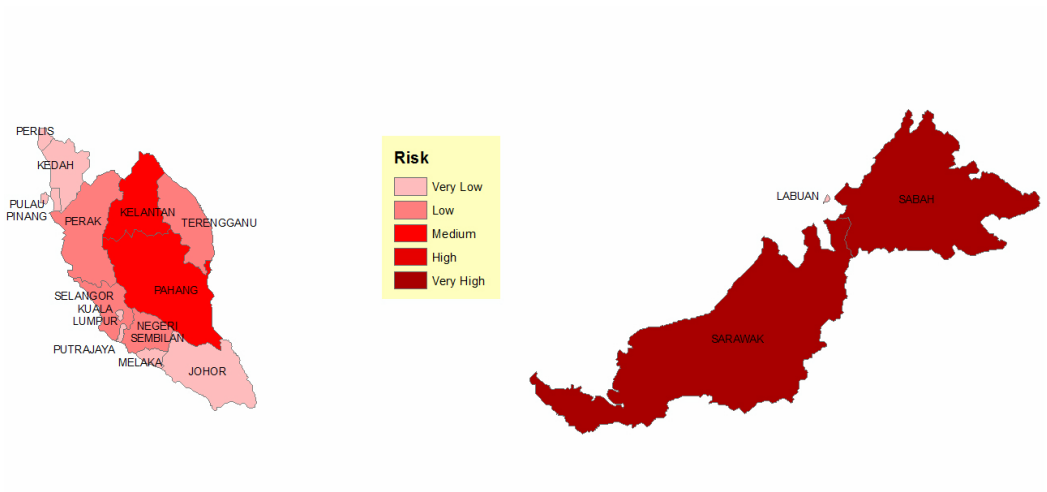

Figure 6 Risk map based on Stochastic SIR-SI model for Epidemiology Week 10

CONCLUSION AND FUTURE WORKS

The production of good disease maps relies on modelling to estimate and predict risks. Better risk estimations and predictions will produce more exact maps of disease risk. In this paper, it was shown that stochastic SIR-SI model was one of the methods that could be used to estimate the relative risk of malaria disease. The results of the analysis showed that Sabah and Sarawak were very high risk areas. Both states are located in the Borneo Island. It is because the malaria virus easily spreads in areas with dirt and has a climate that is not satisfactory. Moreover, the primary cause of the outbreak was fish ponds, which is a major

breeding place for malaria mosquito eggs. Peninsular Malaysia could be categorized as low risk area for malaria virus. In order to avoid the spread of malaria virus, the surrounding area should be clean and free of water reservoir. As stated by Kafadar (1999), the purpose of exploratory disease mapping was to provide insight, as opposed to precise estimates of location, spread or trends. Hence, the purpose of disease risk maps is as a tool for prevention and a control strategy for malaria disease.

REFERENCES

Esteva, L. and Vargas, C. (1998). Analysis of a dengue disease transmission model. *Mathematical Biosciences*, 150, pp. 131-151.

Isham, V. (2005). *Stochastic models for epidemics: Current issues and developments*. In Celebrating Statistics, Eds A. C. Davison, Y. Dodge and N. Wermuth. Oxford University Press, pp. 27-54.

Kafadar, K. (1999). Simultaneous smoothing and adjusting mortality rates in U.S. counties: Melanoma in white females and white males. *Statistics in Medicine.*18:3167-3188.

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

Samat, N.A. and Percy, D.F. (2012).Vector-borne infectious disease mapping with stochastic difference equations: An analysis of dengue disease in Malaysia. *Journal of Applied Statistics*, vol. 39(9), pp. 2029-2046. DOI: 10.1080/02664763.2012.700450.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003).*WinBUGS User Manual Version 1.4*, MRC Biostatistics Unit, Cambridge, UK.

Thomas, D.M., Desch, A.D., Gaff, H.D., Scheele, S.K., Jordan, R.K. and Davis, J.R. (2004). *Estimating infectius disease risk in the absence of incidence Data*. ESRI International Heatlh GIS Conference, Washington, DC.

World Health Organization (WHO). (2013). *Fact Sheets: Malaria*. Retrieved 13 January 2014, from http://www.who.int/mediacentre/factsheets/fs094/en/index.html

World Health Organization (WHO).(2013). *Trends in reported malaria incidence, 2000-2011*. World Health Organization Map Production: Public Health Information and Geographic Information Systems (GIS) World Health Organization. Retrieved 24 April 2013, from http://www.who.int/gho/malaria/malaria_003.jpg?ua=1