A Comparison of Four Disease Mapping Techniques as Applied to TB Diseases in Malaysia

Ijlal Mohd Diah, Nazrina Aziz, Maznah Mat Kasim
School of Quantitative Sciences, UUM College of Arts and Sciences,
Universiti Utara Malaysia, 06010, Sintok, Kedah Darul Aman, Malaysia
ijlalmd@gmail.com

***Abstract*—This paper discusses the results of relative risk estimation based on four different types of methods. The methods used in this study are Standard Morbidity Ratio (SMR), Poisson-gamma model, stochastic Susceptible-Infective-Recovered (SIR) model and new alternative method that we proposed, stochastic Susceptible-Latently infected-Infectious-Recovered (SLIR) model. All the results are comparing and presenting in the form of graphs, tables and maps. These relative risk estimations are applied to TB count data in Malaysia. The maps showed the high-low risk areas for TB occurrence and this can be useful to interest parties in terms of government policy and financial support.**

***Index Terms*—TB Disease; Relative Risk; Disease Mapping; Stochastic Model.**

I. INTRODUCTION

Tuberculosis (also known as TB) is an infectious air-borne disease that caused by bacteria called Mycobacterium tuberculosis. TB is spread through the air via tiny droplets that are released into the air when someone with infected TB sneezes or coughs. Commonly, TB attacks the lungs; however, it can also affect other parts of the body such as joints, kidneys, brain and also nervous system [1].

TB infection is classified into two types; active TB and latent TB. People with active TB usually show symptoms such as weight loss, night sweats, cough that lasts more than three weeks, fever and chest pain [2]. Only active TB person can spread the disease to others. This is different with people with latent TB infection as they do not show any symptoms and cannot spread the disease to others [3].

TB cases have been reported an increase from year to year in Malaysia [4]. TB can be controlled if detected early. Hence, the use of statistical models for the TB disease transmission and the relative risk estimation for disease mapping are significant contributions to prevention and control strategies for TB disease.

In this paper, we measure and compare the performance of relative risk estimation based on four different types of models in order to find the best-fitted model for estimation of relative risk for TB disease mapping in Malaysia. These four methods are Standardized Morbidity Ratio (SMR), Poisson-gamma model, the stochastic susceptible-infected-recovered (SIR) model proposed by [5] and new alternative method that we proposed, stochastic susceptible-latently infected-infectious-recovered (SLIR) model.

II. METHODOLOGY

A. SMR Method

SMR is the most common method used in estimating the relative risk. SMR method essentially compares the observed incidence with the expected incidence which has been used traditionally for the analysis of counts within tracts and calculated as:

$$\hat{\theta}_i = \frac{O_i}{E_i} \tag{1}$$

where O_i is the observed number of deaths cases of the disease in the area and E_i is the expected number of cases.

In disease mapping, suppose that the area of research to be mapped is divided into M mutually exclusive states ($i=1,2,...,M$). Each state has its own observed number of cases O_i and expected number of cases, E_i . Using O_i and E_i as obtained based on the available data, we can calculate one of the most common indices to estimate the relative risk $\hat{\theta}_i$ for state i, which is the SMR model defined as follows:

$$r_i = \hat{\theta}_i = \frac{O_i}{E_i} \tag{2}$$

The observed number of cases can be found from other resources such as health indicator from the Ministry of Health Malaysia which is under disease control department. The expected value can be count by using a particular formula as discussed detail in [6].

B. Poisson-gamma Model

Poisson-gamma model is one of the earliest approaches to Bayesian methodology in disease mapping [7]. It is assumed that the numbers of new infectives y_{ij} follow a Poisson distribution within a given period of time, with mean and variance $e_{ij}\theta_{ij}$. Here, $i=1,2,...,M$ for study regions and $j=1,2,...,T$ refers to the time period, e_{ij} is the expected number of new infective while θ_{ij} is the relative risk:

$$y_{ij}|e_{ij}, \theta_{ij} \sim \text{Poisson}\ (e_{ij}\theta_{ij}) \tag{3}$$

The relative risk parameter has a gamma prior distribution with parameters α and β:

$$\theta_{ij} \sim Gamma(\alpha, \beta) \tag{4}$$

Based on this Poisson-gamma model, the expected posterior relative risk will be included in the analysis. This risk is for

all regions and for all time periods. Further discussion based on this method can be found in [8].

C. Stochastic SIR Model

We applied the stochastic SIR models proposed by Lawson as in [5]. In this study, for i = 1, 2, ..., M study areas and j = 1, 2, ..., T periods of time:

$$I_{i,j} \sim Poisson(\lambda_{i,j}) \tag{5}$$

$$\lambda_{i,j} = \exp(\beta_0 + b_i).S_{i,j}.I_{i,j-1} \tag{6}$$

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

$$S_{i,j+1} = S_{i,j} - I_{i,j} - R_{i,j} \tag{7}$$

$$R_{i,j} = \gamma I_{i,j} \tag{8}$$

In this model, every notation is defined as follows:

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

$I_{i,j}$: total number of new infective persons at time j

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

$\lambda_{i,j}$: the transmission of infection events of susceptible

γ : hazard of an infectious person's being recovered

β_0 : constant term to describe overall rate of the process

b_i : spatial random effect (designed to residual spatial variation for population)

The total number of new infective people, $I_{i,j}$ are assumed to follow a Poisson distribution. The Poisson distribution mean which is called as mean number of infectives, $\lambda_{i,j}$ include a simple mechanistic model for the transmission of infection and a linear predictor term consist of covariate or random effect ($\lambda_{i,j} = \exp(\beta_0 + b_i).S_{i,j}.I_{i,j-1}$). A conditional autoregressive (CAR) prior is used as a family of prior distributions for the random effect in this study.

D. Stochastic SLIR Model

We proposed a new alternative method in estimating the relative risk for TB disease. We used simple deterministic model adapted from [9] and develop it to a stochastic model which was based on the stochastic SIR model as mentioned before. In this study, for i = 1, 2, …, M study areas and j = 1, 2, …, T periods of time, our stochastic models for TB transmission are:

$$\bar{I}_{i,j} \sim Poisson(\gamma_{i,j}) \tag{9}$$

$$\gamma_{i,j} = \exp(\beta_0 + b_i)\, p\lambda S_{i,j-1} \tag{10}$$

$$S_{i,j} = rN + (1-\lambda-\mu)S_{i,j-1} \tag{11}$$

$$L_{i,j} = (1-p)\lambda S_{i,j-1} + (1-v-\mu)L_{i,j-1} \tag{12}$$

$$I_{i,j} = \bar{I}_{i,j} + vL_{i,j-1} + (1-\mu-\mu_T)I_{i,j-1} - \Re_{i,j} \tag{13}$$

$$R_{i,j} = \Re_{i,j} + (1-\mu)R_{i,j-1} \tag{14}$$

$$\Re_{i,j} = cI_{i,j-1} \tag{15}$$

$$N = S + L + I + R \tag{16}$$

Every notation is defined as follows:

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

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

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

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

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

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

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

μ_T : mortality rate because of TB of humans per year

λ : the transmission of infection events of susceptible

N : the population size for the study region

v : the progression rate from latent period to infectious period

p : probability of new infections that develop progressive primary active TB

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

c : the rate of recovery

β_0 : the overall rate of the process

b_i : the random effect that absorbs residual spatial variation

III. Experimental

A. Relative Risk Estimation

In this study, in order to perform these models, we used WinBUGS software. It is a program designed to carry out Markov chain Monte Carlo (MCMC) computations for implement wide variety of Bayesian inference on the statistical problem [7].

First, we find the posterior distribution of the mean number of new infective people. From this information, we find the posterior mean of the relative risk. The estimation of relative risk in this study is based on a formula suggested by [10]. In general, for i = 1, 2, …, M study areas and j = 1, 2, …, T periods of time, the posterior expected mean number of infectives can be approximated using an unbiased sample mean.

$$\tilde{\lambda}_{i,j} = \frac{1}{n}\sum_{k=1}^{n} \lambda_{ijk} \tag{17}$$

where λ_{ijk} for k = 1, 2, …, n is produced from the posterior distribution for the expected mean number of infective $\tilde{\lambda}_{ij}$. The relative risk parameter θ_{ij} is defined by:

$$\theta_{ij} = \frac{\lambda_{ij}}{e_{ij}} \tag{18}$$

where e_{ij} refers to the expected number of new infective cases based on the population across the study areas. The posterior expected relative risk is equal to the posterior expected mean number of new infectives, $\tilde{\lambda}_{ij}$ divide by the mean number of infectives based on the human population across all study areas, $\tilde{e}_{ij}$.

$$\tilde{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \theta_{ijk} = \frac{1}{n}\sum_{k=1}^{n} \frac{\lambda_{ijk}}{e_{ij}} = \frac{\tilde{\lambda}_{ij}}{\tilde{e}_{ij}} \tag{19}$$

The relative risk in this study is defined to be the conditional probability that people within the area contract the disease divided by the conditional probability that people

in the population contract the disease. We condition upon no infection thus far in this context.

In this study, the relative risk is described as the conditional probability that a person within an area contracts the disease divided by the conditional probability that a person in the population contracts the disease. According to [10], for a value of relative risk above 1 means that people within the area are more tend to contract the disease compared with people in the overall population. For relative risk less than 1, this indicates that the people within the area are less likely to endure from the disease compared with people in the population. If the relative risk value is close to 1, it means that there is no real difference in terms of the likelihood that people become infected by TB within an area and within the whole population.

Data used in this study were provided by Ministry of Health, the Institute for Medical Research and the Department of Statistics in Malaysia. All methods mentioned here are applied to TB data in the form of counts of cases within 14 states in Malaysia for the year 2008 until the year 2015. All of these results are presented in graph and table to demonstrate overall relative risks value. Lastly, TB map is constructed based on these results to display the high-low risk area of TB disease occurrences

IV. RESULTS

A. Results of Relative Risk Estimation

Table 1
Comparison Between the Posterior Expected Relative Risks in The Epidemiology Year 2015 Based on Four Different Models

State	Relative Risk Estimations for TB Disease Mapping			
	SMR	Poisson-gamma model	Stochastic SIR model	Stochastic SLIR model
Perlis	0.653	0.674	0.279	0.714
Kedah	0.768	0.770	0.875	0.914
P. Pinang	0.962	0.961	1.032	1.110
Perak	0.826	0.827	0.894	0.949
K.L & Putrajaya	1.229	1.226	1.316	1.426
Selangor	0.964	0.964	1.094	1.135
N. Sembilan	0.766	0.770	0.862	0.917
Melaka	0.747	0.751	1.071	1.088
Johor	0.845	0.845	0.893	0.949
Pahang	0.724	0.727	0.769	0.832
Terengganu	0.774	0.776	0.838	0.929
Kelantan	0.904	0.906	1.047	1.161
Sabah	1.607	1.603	1.867	2.010
Sarawak	1.220	1.219	1.463	1.556

Based on Table 1, we can see that from all methods, Sabah has the highest risk area of contracting TB while Perlis has the lowest risk value of contracting TB when compared with people in the overall population for the epidemiology year 2015 even though the values from different methods are diverse. In this analysis, there is no zero-relative risk value when using SMR method as there is observed count data in all states.

Based on Figure 1 until Figure 4, most of the states have relative risk less than one for all epidemiology years except for the states of Pulau Pinang, Kuala Lumpur, Kelantan, Sabah and Sarawak. This shows that susceptible people in these five states are more likely to get TB while susceptible people within other states are less likely to get TB compared to people the overall population in Malaysia.

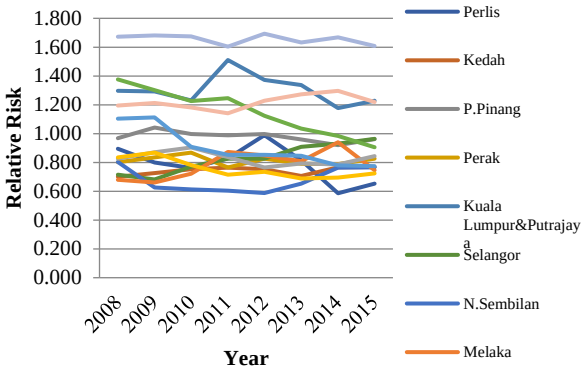

Figure 1: Time series plots of the relative risk estimation based on the SMR method for different states in Malaysia

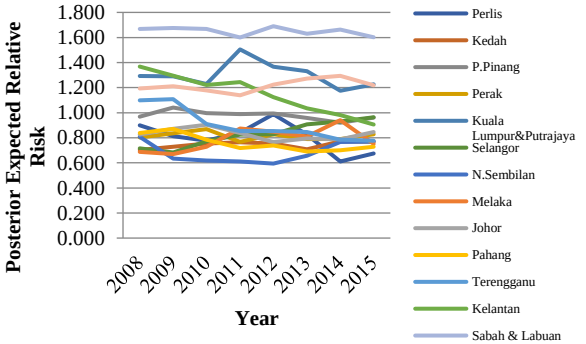

Figure 2: Time series plots of the relative risk estimation based on the Poisson-gamma model for different states in Malaysia

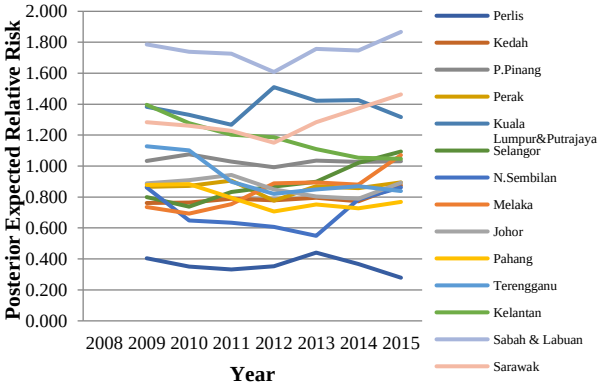

Figure 3: Time series plots of the relative risk estimation based on the stochastic SIR model for 14 states in Malaysia

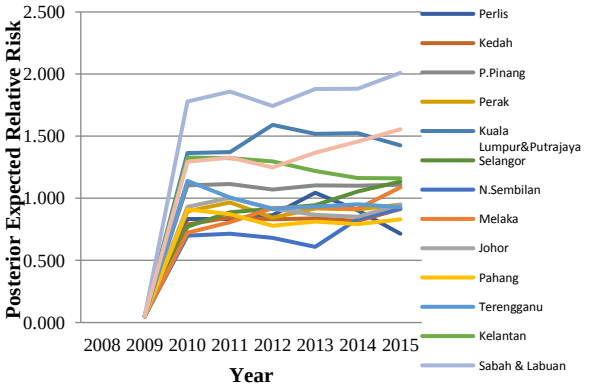

Figure. 4. Time series plots of the relative risk estimation based on the stochastic SLIR model for 14 states in Malaysia

From the results shown in Table 1, we construct maps to show a clear picture of low and high-risk areas of TB occurrences. These maps can be used by the government or other interest parties as a tool to recognize which states that need further attention in terms of government policy and financial and also others support in order to control and prevent TB disease from becoming worse. In order to determine which method produces a smoother map, we compared the maps using these four types of method.

B. Disease Maps of TB Occurrence

In order to facilitate the classification of risk, we divided the results of relative risk into five different levels which are very low, low, medium, high and very high with respective intervals of [0, 0.5), [0.5, 1), [1.0, 1.5), [1.5, 2.0) and [2.0, ∞). We used choropleth maps (also known as thematic maps) with tones of one colour to display and differentiate between the low and high-risk areas of TB occurrences. The epidemiology year 2015 is chosen as an example and for comparison purposes only.

Figure 5 shows the disease maps based on the SMR methods. Based on the figure, Sabah has been recognized as a high-risk area for the epidemiology year 2015, followed by the state of Sarawak and Kuala Lumpur including Putrajaya with medium risk. The other eleven states are classified as low risk. There is no state classified as very high and very low-risk area based on Figure 5.

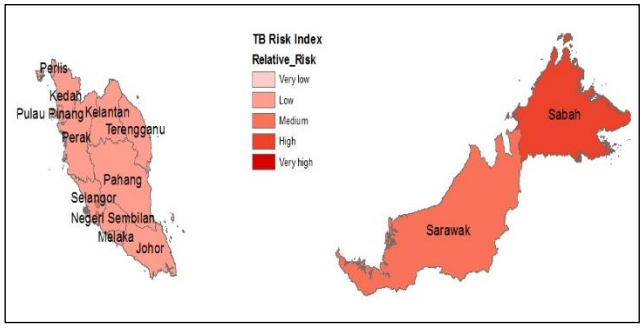

Figure 5: Disease map of relative risk estimation based on SMR method for the year 2015

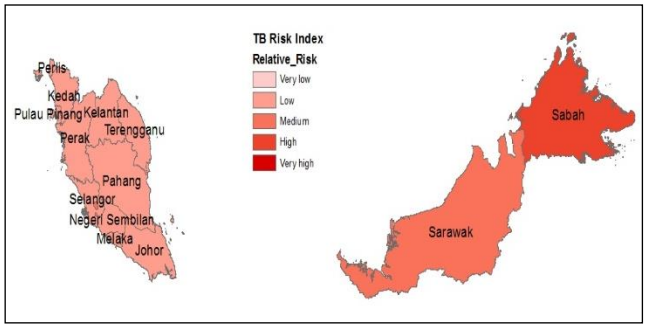

Figure 6: Disease map of relative risk estimation based on Poisson-gamma method for the year 2015

Similar to those results in the SMR map in Figure 6, the state of Sabah has been classified as a high-risk area for Poisson-gamma map, with no state classified as very high and very low risk. Sarawak, Kuala Lumpur and Putrajaya show the medium risk while the other states with low risk.

Figure 7 presents the high-low risk area based on stochastic SIR model for the epidemiology year 2015. Similar to those results in the SMR map and Poisson-gamma map, the state of Sabah has a high risk for TB occurrences. These are followed by Pulau Pinang, Kuala Lumpur & Putrajaya, Selangor, Melaka, Kelantan and Sarawak with medium risk for TB occurrences. The other states have low risk except for Perlis with very low risk.

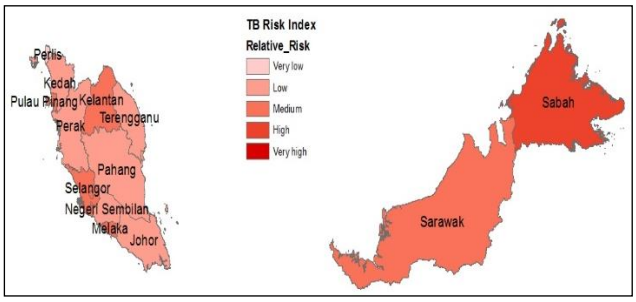

Figure 7: Disease map of relative risk estimation based on stochastic SIR model for the year 2015

Figure 8 displays map for posterior expected relative risks based on stochastic SLIR model especially for TB disease transmission. From Figure 8, Sabah has been recognized as a very high risk for TB occurrences. People in Sarawak have a high risk for contracting TB in the year 2015. These are followed by the state of Pulau Pinang, Kuala Lumpur & Putrajaya, Selangor, Melaka and Kelantan with medium risks. The other seven states are classified as low risk, with no state classified as a very low risk.

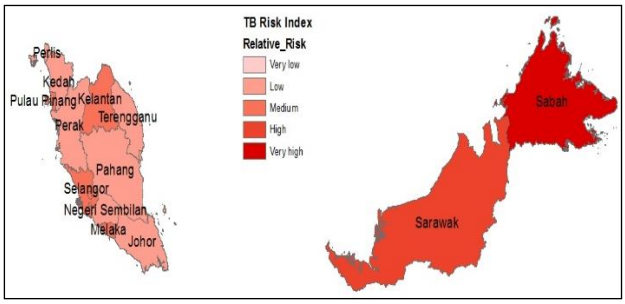

Figure 8: Disease map of relative risk estimation based on stochastic SLIR model for the year 2015

Based on the map's comparison above using these four methods (SMR, Poisson-gamma model, stochastic SIR model and stochastic SLIR model), there are no obvious differences in the estimated risks.

However, from Figure 8 which is TB risk map based on stochastic SLIR method for the epidemiology year 2015, we can see 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 states. Spatial correlation is one of the elements that need to consider since the states are located side by side. Hence, the risk may be transfer to other states.

V. CONCLUSION

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 areas of high risk so that more attention can be given to the priority states. Here, we cannot easily conclude whether any map is smoother or more precise maps than others, but based on the details of the model described earlier, the evidence

continues to support the stochastic SLIR model. This stochastic SLIR model offers a better way of describing the underlying behaviours as well as estimating the relative risk compared to SMR, Poisson-gamma and stochastic SIR models. Hence, we believe that this model should be used to predict the risks of TB disease across the 14 states of Malaysia for short-term forecasting. This method can be applied to others disease which has a latent period in the transmission of the disease.

ACKNOWLEDGMENT

The authors would like to acknowledge Universiti Utara Malaysia for the financial support in respect of this study.

REFERENCES

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

[2] D. Bhowmik, R. M. Chandira, and U. Pradesh, "Recent trends of drug used treatment of tuberculosis", *Journal of Chemical and Pharmaceutical Research*, 2009, vol. 1, no. 1, pp. 113–133.

[3] A. Konstantinos, "Diagnostic tests testing for tuberculosis", *Aust Prescr*, 2010, vol. 33, no. 1, pp.12–18.

[4] Ministry of Health Malaysia, Health Indicators 2003-2013. Retrieved from http://vlib.moh.gov.my/cms/content.jsp?id=com.tms.cms.section.Section_62581 998-c0a81049-d0570e00-98ce6828

[5] A. B. Lawson, *Statistical methods in spatial epidemiology*, 2nd ed., England: John Wiley & Sons, Ltd, 2006.

[6] I. M. Diah, N. Aziz and N. Ahmad, "Relative risk estimation of tuberculosis with standardized morbidity ratio in Malaysia," *Global Journal of Pure and Applied Mathematics*, 2016, vol. 12, no. 5, pp. 4011-4019.

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

[8] I. M. d Diah, N. Aziz and N. Ahmad, "Tuberculosis Disease Mapping with Poisson-Gamma Model in Malaysia," *Research Journal of Applied Sciences*, 2016, vol. 11, pp. 822-825. doi:10.3923/rjasci.2016.822.825

[9] S. M. Blower, A. R. Mclean, T. C. Porco, P. M. Small, P. C. Hopewell, M. A. Sanchez, and A. R. Moss, "The intrinsic transmission dynamic of tuberculosis epidemics", *Nature Medicine*, 1995, vol. 1, no. 8, pp. 815-
821.

[10] N. A. Samat & D.F. Percy, "Dengue disease mapping in Malaysia based on stochastic SIR models in human populations," in *ICSSBE 2012 – Proceedings, 2012 International Conference on Statistics in Science.*