Pertanika J. Sci. & Technol. 29 (2): 1071 - 1095 (2021)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Relative Risk Estimation for Human Leptospirosis Disease in Malaysia Based on Existing Models and Discrete Space-Time Stochastic Sir Model

Sufi Hafawati Ideris*, Muhammad Rozi Malim and Norshahida Shaadan

Faculty of Computer and Mathematical Sciences, MARA University of Technology, Shah Alam, 40450 Selangor, Malaysia

ABSTRACT

The disease leptospirosis is known to be endemic in Malaysia, and it significantly impacts human wellbeing and the national economy. Current surveillance systems are based on morbidity and mortality leptospirosis national data from the Ministry of Health and remain inadequate due to the number of unreported and misdiagnosed cases. A robust surveillance system is needed to monitor temporal and spatial changes which yield improvements in terms of identifying high-risk areas and disease behaviour. The objective of this study is to identify high-risk areas by estimating relative risk using existing models which are the Standardized Morbidity Ratio (SMR), Poisson-gamma, log-normal, Besag, York and Mollié (BYM) and mixture models. An alternative model is also proposed which involves transmission systems and stochastic elements, namely the stochastic Susceptible-Infected-Removed (SIR) transmission model. This estimation of risk is expected to assist in the early detection of high-risk areas which can be applied as a strategy for preventive and control measures. The methodology in this paper applies relative risk estimates to determine the infection risk for all states in Malaysia based on monthly data from 2011 to 2018 using WinBUGS 1.4 software. The results of relative risks are discussed and presented in tables and graphs for each model to disclose high-risk areas across the country. Based on the risk estimates, different models used have different risk interpretations and drawbacks which make each model different in its use depending on the objectives of the

ARTICLE INFO

Article history:
Received: 9 September 2020
Accepted: 25 January 2021
Published: 30 April 2021

DOI: https://doi.org/10.47836/pjst.29.2.20

E-mail addresses:
sufi.hafawati@gmail.com (Sufi Hafawati Ideris)
rozi@tmsk.uitm.edu.my (Muhammad Rozi Malim)
shahida@tmsk.uitm.edu.my (Norshahida Shaadan)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

study. As a result, the deviance information criteria (DIC) values obtained do not differ greatly from each expected risk which was estimated.

Keywords: BYM, DIC, leptospirosis, log-normal, mixture model, poisson-gamma, relative risk, SIR transmission model, SMR

INTRODUCTION

Leptospirosis is a primarily zoonotic disease originating from the pathogenic leptospires species and living widely throughout nature. Small mammals like rodents are the reservoir hosts in which pathogenic leptospires are carried and maintained, reproducing itself in their hosts' kidneys. Excretion of leptospires through the urine of those infected animals causes acquisition in new hosts, whether other animals or humans (Haake & Lavett, 2014). Transmission may occur via two major routes: direct contact with urine or other bodily fluids of infected animals through the skin and intact mucus membranes such as the eyes, nose or mouth; and indirect contact with broken skin or through being scratched with soil, food or water contaminated with the urine of infected animals (Wahab, 2015). Leptospirosis is widespread and can be a serious public health issue especially in temperate or tropical climates (Haake & Levett, 2014). Malaysia is known to be a location in which leptospirosis is endemic among the south-east Asia region and significantly impacts human wellbeing and the national economy (Garba et al., 2017a). This is because the humid weather and climate conditions in Malaysia are conducive for the growth of the leptospires bacteria. The abundance of leptospires has therefore caused death and suffering in Malaysia since it was first diagnosed by Fletcher (1928). Based on Figure 1, the frequency of incidence and the mortality rate due to leptospirosis increased between 2004 and 2015.

Despite the dramatic increase in the number of cases and deaths, leptospirosis is a seriously neglected tropical disease. Factors like misdiagnosis of clinical symptoms which resemble other similar diseases (such as dengue and hepatitis) and the limited or lack of rapid diagnosing facilities causes the disease to continually be under-reported. This is also may create barriers to government efforts to determine the true incidence rate for the human population. The elimination of leptospirosis is therefore crucial to minimise the risk of infection. This prevention action can be performed through appropriate controls, including a robust and recognised vaccination programme, promotion of the use of hygiene and personal protective equipment as well as awareness of high incidence areas by the authorities and society. This is crucial for the risk estimates in geographical areas and revealing the hotspot areas of leptospirosis infection.

In Malaysia, the surveillance systems currently and typically rely on available leptospirosis national databases, which are the morbidity and mortality reports from

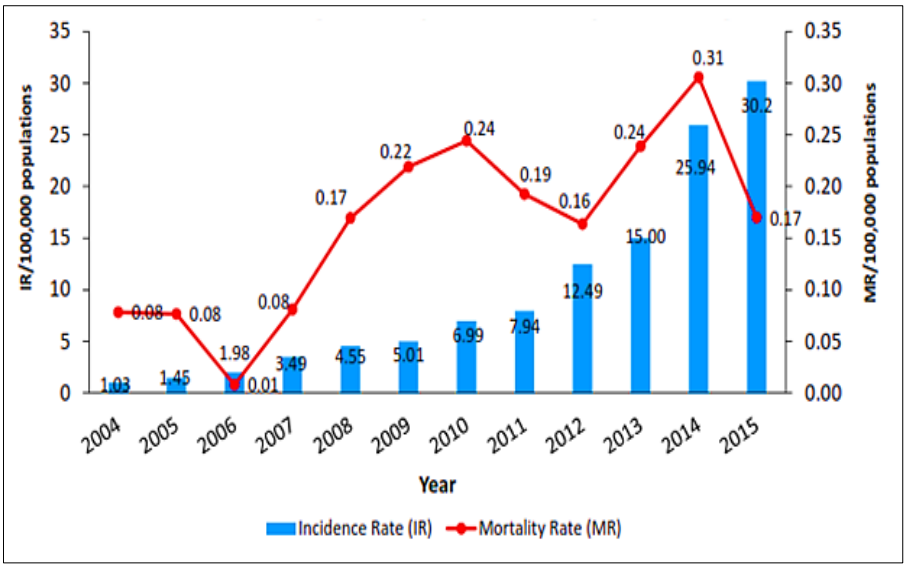

Figure 1. Incidence and mortality rate of leptospirosis in Malaysia, 2004 to 2015 (Wahab, 2015)

Ministry of Health Hospitals. However, the existing surveillance system is still inadequate to combat this disease due to the number of unreported and underestimated cases (Benacer et al., 2016). A robust surveillance system which can monitor temporal and spatial changes is needed to yield improvements in identifying at-risk areas and disease behaviour. This at the same time can assist the early detection of high-risk areas which can be applied as a strategy for preventive measures in Malaysia and other countries.

Estimation of relative risk is very common in the study of epidemiology and is defined as the estimation of the probability of infection after exposure to the risk variable compared with the likelihood of disease occurrence in a sample study group or population (Andrade, 2015). Relative risk (also known as risk ratio) calculation is usually expressed as a ratio to a denominator instead of a percentage. Relative risk in this study is defined as the probability that a person within a specific region would contract the disease against the likelihood that a person in the general population would contract the disease. To determine the disease risk status of an area, the first action is to assess the local supposed 'expected' incidence of the disease, followed by a comparison with local observations of the incidence value of the disease. Several common methods are used in the epidemiology field to estimate the ratio of risk by using data for counts within tracts. These are the SMR model, the Poisson-gamma model, the Log-normal model, the Besag, York and Mollié (BYM) model and the mixture model. These approaches are known as classic methods of estimating disease risk which are commonly used in disease mapping. However, this paper will not discuss in detail

the representatives of the disease risk map. The focus is on the estimation of relative risk approaches to determine the risk of leptospirosis disease across the Malaysian region. The methodology is discussed from a basic Poisson likelihood to more complex risk structure models which include spatial correlation and spatiotemporal.

The goal of epidemiology research is to gain an insight into the nature or history of transmission model development (Brauer, 2017). Biologically, leptospirosis disease involves direct and indirect processes of transmission from infectious to susceptible individuals. Therefore, a model that can incorporate this transmission system and an alternative method for estimating the relative risk of the leptospirosis SIR transmission model are required in this paper for the human population using monthly data from Malaysia. The results of relative risk estimated from both the existing and proposed transmission models are discussed and the performance is evaluated using deviance information criteria (DIC) to assess model complexity and best fitting model to the data.

METHODOLOGY

Data and Study Location

The relative risk estimation in this paper used secondary data for observed cases of leptospirosis infection in the form of count cases for all 13 states and 3 federal territories in Malaysia. These states and federal territories are Perlis, Kedah, Penang, Perak, Selangor, Federal Territory of Kuala Lumpur, Federal Territory of Putrajaya, Negeri Sembilan, Malacca, Johor, Pahang, Terengganu, Kelantan, Sarawak, Sabah, and Federal Territory of Labuan. In Malaysia, leptospirosis was gazetted as a notifiable disease in December 2010 (Garba et al., 2017a; Sulong et al., 2011). Thus, this study was designed to accommodate the monthly observed cases starting from January 2011 up to December 2018 so cumulative data could be taken at annual intervals (starting in January and ending in December) for all states and federal territories in Malaysia.

This secondary data was obtained from the Ministry of Health with ethic approval from the National Medical Research Registration (NMRR): NMRR-18-3006-43002 (IIR). Analyses are based on space and time dimensions since data from the public health authorities (MOH) is often available for time windows for several years. Here, the observed counts for the new infected human cases are based on the time $(j - 1, j)$ for $j = 1, 2, \ldots, 96$ months per calendar month and study region i for all 13 states and three federal territories in Malaysia. A one-month observation time was chosen based on the development of the leptospira bacteria in the human body once a person is infected (Triampo et al., 2008). In this study, the observed number for humans infected with leptospirosis was categorised as the dependent variable and time (monthly) was the independent variable.

Analyses Framework

This paper involves several analytical steps such as data collection, calculation, estimation, presentation of results as well as comparison. This progress flow is explained in more detail in the form of a flow chart in Figure 2.

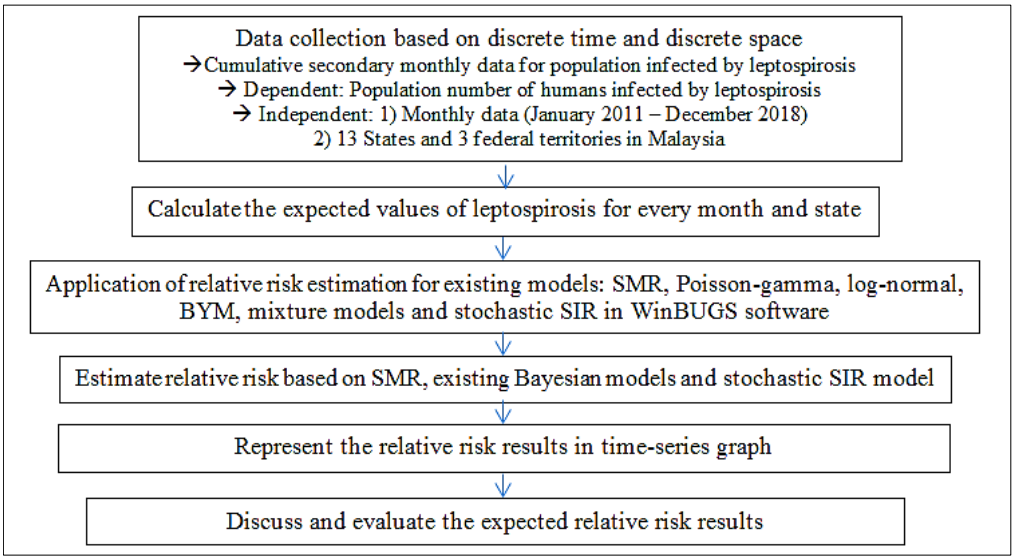

Figure 2. Flow of the process

Relative Risk Estimator for Leptospirosis using Existing Models

In statistical analysis, inferences are made by selecting a value or pattern for a certain phenomenon from empirical data, a procedure known as an estimator (Long, 2012). An estimator aims to measure or derive such risk estimates on a statistical spatial method. A basic summary measurement for an estimator often computes a local relative risk which is taken to measure excess risk within a population (Lawson, 2006). This risk estimate has been commonly used in spatial epidemiology and is calculated by comparing the observed incidence to that expected from the background population. This assessment of risk status of the ratio of observed to expected counts is called standardised mortality or morbidity ratio (SMR), a ratio which is an estimate of relative risk within each area or region. Usually, cofounding, or unobserved variables can be found within observed data in spatial units affecting individuals or associated tracts or covariates, these are known as random effects. Due to its simple estimates, this type of primary descriptive method often ignores such variables. Thus, more substantial models are used to include these effects in the analysis, which is discussed further in this study.

Standardized Morbidity Ratio (SMR). In broad epidemiological terms, SMR can be defined as either Standardised Mortality Ratio or Standardised Morbidity Ratio. The term mortality refers to death and the mortality rate is the number of deaths divided by the size of the composite population. Meanwhile, morbidity means disease or illness. It can range from mild illnesses such as common flu to chronic diseases such as cancer or traumatic brain injury. In addition, a study of nonfatal disease is sometimes known as the Standardised Morbidity Ratio (Richardson et al., 2004) where the status burden of a disease in an area respective to the disease is the ratio of observed to expected counts.

Suppose that the study area to be mapped is divided into M mutually exclusive states (i = 1, 2, …, M) and T time (j = 1, 2, …, T). Each state has its own observed number of cases 0_i and expected number of cases E_i, as shown in Equation 1 and 2.

$$E_i = \text{Population of state } i \text{ in time } j \times \frac{\sum number\ of\ observed\ cases\ for\ all\ states\ in\ time\ j}{\sum number\ of\ population\ in\ time\ j} \quad [1]$$

$$E_i = N_{i,j}\left(\sum_{i=1}^{M} O_j \Big/ \sum_{i=1}^{M} N_j\right) \quad [2]$$

Where Ni is the population of state i and the summations are for i = 1, 2, …, M. Here the standardisation is performed based on the total population at risk, assuming everybody is equally at risk. Using 0_i and E_i as obtained from the available data, the derivation of relative risk value, which is denoted as $\hat{\theta}_i$, can be calculated using Equation 3. The relative risk, $\hat{\theta}_{i,j}$ for state i at time j is defined as the probability that a person within the state i contracts the disease divided by the probability that a person in the population contracts the disease (Samat & Percy, 2012).

$$\text{Relative risk} = \frac{Observed\ number\ of\ cases\ in\ state\ i\ at\ time\ j}{Expected\ number\ of\ cases\ in\ state\ i\ at\ time\ j}$$

$$\hat{\theta}_{i,j} = \frac{O_{i,j}}{E_{i,j}}$$

$$\hat{\theta}_{i,j} = \frac{O_{i,j}}{N_{i,j}\left(\frac{\sum_{i=1}^{M} O_j}{\sum_{i=1}^{M} N_j}\right)} \quad [3]$$

$$\widehat{\theta}_{i,j} = \frac{\left(\frac{O_{i,j}}{N_{i,j}}\right)}{\left(\frac{\sum_{i=1}^{M} O_j}{\sum_{i=1}^{M} N_j}\right)}$$

However, this method has several drawbacks that make interpretation of the SMR difficult. According to Kismiantini (2009), the variance is large in a region that has small

expected values, especially an area with a small population size or region with rare diseases cases. This extra variation in a small region with a small population size can cause artefacts or errors in the estimated risk (Lawson & Williams, 2003). The SMR is measured by observed over expected incident numbers in the form of a ratio. If the ratio is more than 1 it may suggest excess in risk, but the SMR will become more than 1 or even large if the expected number of incident cases is too small. For instance, the observed count number for Labuan in April 2016 was 11 with the expected number counted as 1, giving an SMR of 11. Otherwise, if the expected value is too large or there is no observed value of cases in a region, then the SMR is very small or equal to zero (Awang & Samat, 2017). In this case, the zero value of relative risk indicates that there is no risk for the people in the study region infected by the disease even though everybody has the possibility of catching the disease. In addition, if the expected number is zero in a count tract, then the estimation ratio of the SMR is infinite. All these types of artefacts for the SMR often occur in a region with a small population size and area and with a small count of cases (such as Labuan, Perlis, and Putrajaya). Thus, the SMR is a reliable measure of relative risk for large geographical regions such as countries but may be unreliable for small areas such as counties (Meza, 2003). To overcome these drawbacks, many researchers have sought alternative solutions by introducing better alternative methods for relative risk estimation (Kismiantini, 2009). These include the use of Bayesian methods such as the Poisson-gamma model.

Poisson-Gamma Model. The Poisson-gamma model is one of the earliest examples of a Bayesian approach used to estimate disease relative risk especially when the data is typically small (Richardson et al., 2004). By using Bayes' rule, a combination of observed data and some prior knowledge would produce the posterior distribution. In Bayes' method, relative risk estimation is based on the posterior distribution which is defined as the product of likelihood function and prior distribution in Equation 4.

$$Posterior \propto Likelihood \times Prior \quad [4]$$

For $i = 1, 2, \ldots, M$ study states and $j = 1, 2, \ldots, T$ time, $y_{i,j}$ is the observed count of cases in state i and at time j. Likelihood is defined as the probability distribution for data that conditionally acts upon unknown parameter, $\theta_{i,j}$. This probability distribution for likelihood is often assumed to follow a Poisson distribution within a given period, with mean and variance $e_{i,j}\theta_{i,j}$ as follows:

$$y_{i,j}|e_{i,j}, \theta_{i,j} \sim Poisson\left(e_{i,j}\theta_{i,j}\right) \quad [5]$$

Based on Equation 5, $e_{i,j}$ is the expected number of cases and the parameter of interest, $\theta_{i,j}$ is assumed to be a parameter of relative risk for state i and treated as a random variable. The random variable distribution of parameter $\theta_{i,j}$ has a gamma prior distribution with parameters α and β as written in Equation 6.

$$\theta_{i,j} \sim Gamma(\alpha, \beta) \quad [6]$$

This gamma prior provides for a measure of extra variation affected by random effects within the estimation of risk (Lawson, 2006). Hence, the final output of the Bayes estimate includes the posterior expected of relative risk for all states and time periods. Previous studies by Awang and Samat (2017) and Ideris and Samat (2015) suggested that Poisson-gamma gave a better estimation compared to the SMR when dealing with zero or clustered counts data. This is due to the modelling of random effects applied within the Bayesian framework in the Poisson-gamma model.

Log-Normal Model. Although Poisson-gamma is convenient, it may be too complicated to allow spatial correlation between risk in nearby areas, while the covariate adjustment is also difficult when heterogeneity is involved (Lawson & Williams, 2003; Nurmalasari & Pramana, 2014). Spatial random effects may arise when the regions are believed to have site-dependent covariates. Thus, the log-normal model is introduced which allows spatial correlation between risk in nearby areas. A log-normal model can be written as Equation 7:

$$\begin{aligned} y_i &\sim Poisson(e_i\theta_i) \\ log\theta_i &= \alpha + v_i \\ v_i &\sim N(0, \tau_v^2) \end{aligned} \quad [7]$$

For a log-normal model, y_i are assumed to be Poisson distributed with parameter $e_i\theta_i$. The observed number of cases θ_i is described in logarithmic terms, with α as a semiparametric model of improper prior distribution for an intercept mean of relative risk in the study region (Lawson, 2006), while v_i represents the correlation parameter between the risk of neighbouring areas. This correlation is assumed to be normally distributed with mean 0 and variance τ_v^2 .

Besag, York and Mollié (BYM) Model. In the BYM model, random effects are specified into the above components for the model of relative risk estimation that considers effects that vary within a cluster or correlated heterogeneity, as well as effects that vary in uncorrelated heterogeneity. The relative risk for the BYM model is formulated as Equation 8 where u_i is the correlated heterogeneity while v_i is the uncorrelated heterogeneity. Meanwhile α is assigned as an improper uniform prior for the conditional autoregressive model (CAR) distribution to ensure that the model is identifiable.

$$\begin{aligned} y_i &\sim Poisson(e_i\theta_i) \\ log\theta_i &= \alpha + u_i + v_i \\ v_i &\sim N(0, \tau_v^2) \end{aligned} \quad [8]$$

In the form of extra variation, there is a spatial autocorrelation between spatial units due to the unobserved environmental effect, and a naturally clustered scale observed for the disease for its distribution as discussed before, where estimation of risk in any area depends on neighbouring regions. Therefore, the conditional autoregressive (CAR) model is used to unveil and describe the quantities for spatial relations between the data and spatial variation in the form of sub-regions aggregation as well as to detect the cluster areas. The CAR model is defined as Equation 9:

$$[u_i|u_k, i \neq k, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2) \quad [9]$$

where,

$$\bar{u}_i = \frac{1}{\sum_j \omega_{ik}} \sum_k u_k \omega_{ik}$$

$$\tau_1^2 = \frac{\tau_u^2}{\sum_k \omega_{ik}} \quad [10]$$

$\omega_{ik} = 1$ if *i, k* are adjacent (or 0 otherwise)

By referring to Equation 10, ω_{ik} represents the weightage for neighbouring areas. Summation of weightage ω_{ik} is equal to 1 if *i, k* is adjacent and 0 when *i*, *k* are not adjacent. Meanwhile, variables *v* and *u* are controlled by parameters τ_v^2 and τ_v^2 representing the inverse-variance of random effects. These hyperprior distributions are assumed to follow gamma distributions.

Mixture Model. To estimate true relative risk, an appropriate model should be able to detect and maintain discontinuities and variations between neighbouring areas. This means that an area should not be too dependent on its nearby areas, which causes the estimated risk to become almost identical to the neighbouring areas. To overcome this problem of the BYM model, the mixture model was proposed by contributing discrete jumps components to maintain the variation risk for each tract. The relative risk model is composed of three components: *v* as unstructured heterogeneity and two mixing components; and (u, φ) denoted as a combination of a component for spatial correlation and a component that models the discrete jumps between neighbouring areas.

$$log\theta_i = \alpha + v_i + p_i u_i + (1 - p_i)\varphi_i \quad [11]$$

Equation 11 depends on the value of p_i as a cluster assignment which subdivides the data points into clusters (Long, 2012). When p_i = 1 the standard BYM model arises for the smoothing, and when p_i = 0, a pure jump model arises which is distributed as Equation 12.

$$p_i \sim Beta(0.5,0.5)\ \forall i \quad [12]$$

When the discrete jump model is considered (for $p_i = 0$), the spatial correlation follows a chosen prior which examines the total difference of two independent random variables between neighbours. The results of Lawson and Williams (2003) show that estimated risk through the mixture model performs better than the BYM by maintaining the areas with jumps, and at the same time displaying smoothing in the other areas.

Based on the discussion of all the risk estimation methods above, the output risk varies according to the approaches used. Each method has different interpretations and elements used in estimating the spatial random variables. A more effective model is often assumed to produce true risk estimation which takes random effects into account (Lawson, 2006). In the next subtopic, a goodness-of-fit method for the model comparison for the Bayesian model, which is a package from WinBUGS software, is discussed.

Deviance Information Criterion (DIC). Model comparisons for complexity and fit of the data to the model in Bayesian measurements can be conducted using deviance information criterion (DIC) monitored in WinBUGS, given as Equation 13.

$$\text{DIC} = \text{Dbar} + \text{pD} \quad [13]$$

Based on Equation 13 of DIC above, Dbar is used to calculate the goodness of fit for the model while component pD calculates the complexity of the model by estimating the effective number of parameters used in the model (Spiegelhalter et al., 2003). A model with a small DIC number is estimated to be the best model to fit with the data. In statistical models of analysis, there is no perfect or good model, and no mathematical model can accurately estimate detailed and real situations. DIC is used to compare the models which best fit the data. However, some studies disputed the capacity of DIC to compare hierarchical or multi-level models when re-parameterisation occurs (Pooley & Marion, 2018; Dey et al., 2019). On the other hand, it is very elusive to find a study which compares DICs between different forms of mathematical models. This is because models with different operator scopes have different mathematical functions and operations. Therefore, comparing different types of mathematical models with different mathematical functions can cause misleading interpretation of DICs.

Expected Relative Risk for Leptospirosis using Discrete Space-Time Stochastic SIR Model for Human Transmission

The focus of this paper is the application of mathematical and statistical approaches that quantify transmission models and data uncertainty for the estimation of relative risk.

Descriptions of mathematical transmission models are often based on the understanding of a disease spread between susceptible and infectious individuals (Brauer, 2017). This study led to the development of a compartmental model which has long been introduced and applied by mathematicians and physicians. The basic development of a compartmental model is based on the ideas behind the SIR model proposed by Kermack and McKendrick (1927). The simple compartmental model assumes that the mixing population in the compartments is homogeneous. However, a small number of individuals are infected at the beginning of the disease outbreak and the transmission is a probabilistic event based on the pattern of contact between susceptible and infective individuals.

Therefore, in this paper a basic compartmental stochastic SIR model is proposed specifically for humans infected by leptospirosis disease. This model involves a transmission system (known as a compartmental model) and a stochastic process which involves the probabilistic process of infection adapted from an early study by Reed-Frost in the 1920s. The population was divided into compartments with a time rate of transfer from one compartment to another. The terminology SIR is used to describe individuals from susceptible state *S* to infected state *I* and eventually to recovered state *R*. Susceptible is classified for individuals who will be the subject of the disease but are not yet infected. Individuals classified as infected are the ones who had been infected with the disease from contact with infected rats' urine. Once the infected individuals recuperated from the disease, they moved to the recovered state.

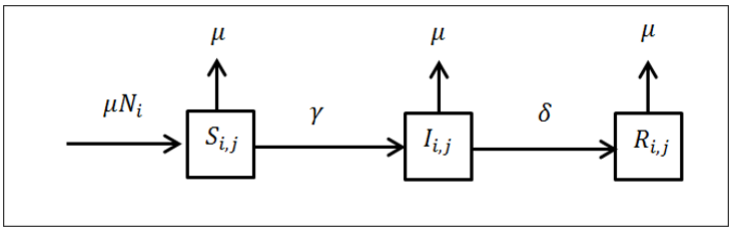

Figure 3. A compartmental SIR model for human leptospirosis transmission

Let *'S'* be the susceptible, *'I'* is the infected and *'R'* is recovered for i=1, 2,…, M study regions (for all states and federal territories in Malaysia) and j=1, 2,…, T time periods (which is monthly data from January to December per calendar month). From Figure 3, the compartment $S_{i,j}$ represents the total number of susceptible persons for leptospirosis for state i at j^{th} month, $I_{i,j}$ refers to the total number of persons infected with leptospirosis at state i for month j, and $R_{i,j}$ is the total number of recovered persons from leptospirosis disease for state i at month j.

The arrow from susceptible to infected represents the new infections occurring at rate γ. While δ denotes the recovery rate for an infected person. Furthermore, μ represents

the rate for persons who leave the group because of natural death such as from an illness unrelated to leptospirosis, like diabetic complications, myocardial infarct, malignancy, and motor-vehicle accidents among the human population. For human population, N_i is assumed to have closed population comprises of susceptible, infected, and recovered individuals written as Equation 14.

$$N_i = S_i + I_i + R_i \quad [14]$$

The total susceptible human population comprises the existing susceptible population which increases with the rate of birth, μ, of which every person born is assumed to be susceptible to leptospirosis. At the same time, the susceptible population decreases following effective contact with the infectious individuals at rate γ. The susceptible human population is further decreased by natural death at rate μ. In this model, birth and death rates are assumed to be equal for simplicity and the rate of death μ is constant for all population subgroups of susceptible, infected, and recovered individuals. Thus, the deterministic difference model for susceptible humans can be written as Equation 15.

$$S_{i,j} = S_{i,j-1} + \mu N_i - \gamma I_{i,j} S_{i,j-1} - \mu S_{i,j-1} \quad [15]$$

Next, the population of infected humans is generated from the existing previously infected individuals and increase by newly infective individuals by new infections of susceptible humans' rate γ. The population of infectious humans is decreased by natural death at rate μ, as well as by recovery from infection at rate δ. The infectives human formula can be written as Equation 16.

$$I_{i,j} = I_{i,j-1} + \gamma I_{i,j} S_{i,j-1} - \mu I_{i,j-1} - \delta I_{i,j-1} \quad [16]$$

The population of individuals in the recovered compartment is formulated by the existing number of previously recovered persons and increased by the number of newly recovered persons from the infected class at rate δ while it is decreased by natural death μ which is constant for all compartments as written in Equation 17.

$$R_{i,j} = R_{i,j-1} + \delta I_{i,j-1} - \mu I_{i,j-1} \quad [17]$$

The system of transmission behaviour for human leptospirosis disease (Figure 3) can be written as a linear system of difference deterministic Equation 18.

$$\begin{aligned} S_{i,j} &= \mu N_i + \{1 - \mu - \gamma I_{i,j-1}\} S_{i,j-1} \\ I_{i,j} &= (1 - \mu - \delta) I_{i,j-1} + \gamma I_{i,j-1} S_{i,j-1} \\ R_{i,j} &= \delta I_{i,j-1} + (1 - \mu) R_{i,j-1} \end{aligned} \quad [18]$$

Based on this proposed compartmental model, the deterministic Equation 16 was derived to develop the corresponding discrete-time discrete-space stochastic SIR model specifically for leptospirosis disease transmission for human population. Initially there must be an infected human in the population; $I_{i,0} = 1$. Meanwhile the initial number of susceptible humans is $S_{i,0} = N - 1$. Both $I_{i,j}$ and $S_{i,j}$ denote the number of infected and susceptible humans for state *i* at time *j*, respectively. The probability of contact between a susceptible human with infectious individuals is *b*. The infection probability for a susceptible human to become infected is *β*. The transmission coefficient of infectious individuals to the human population is denoted as $\frac{I}{N_i}$. Thus, the transmission rate or probability of susceptible individuals being infected is summarised as Equation 19.

$$\gamma = \frac{\beta b I}{N_i} \qquad [19]$$

This stochastic component has been adapted from the original model designed by Reed-Frost in the 1920s for leptospirosis infection. It is reasonable to represent uncertainty or randomness in the leptospirosis infection model. This stochastic model includes a new term $I_{i,j}$ representing the number of newly infected leptospirosis cases in the human population. The time interval is Poisson stationary and the independent increment properties (*j–1* and *j*) in the study region are represented by *i*. A stochastic process is a Poisson distribution process with random likelihood that may be analysed statistically but may not be predicted precisely. In this analysis, a random Poisson distribution will be used to generate the probabilities of outcomes of a system called the Monte Carlo simulation which includes a probability distribution to reflect the inherent randomness within the data.

For human leptospirosis, $I_{i,j}$ = the total number of newly infected human leptospirosis cases,

$$S_{i,j} = \mu N_i + S_{i,j-1} - \mu S_{i,j-1} - \iota_{i,j}$$

$$\iota_{i,j} \sim Poisson(\lambda_{i,j})$$

$$\lambda_{i,j} = [exp(\beta_o + c_i)]\left(\frac{\beta b}{N_i} I_{i,j-1} S_{i,j-1}\right) \qquad [20]$$

$$I_{i,j} = (1 - \mu - \delta) I_{i,j-1} + \iota_{i,j}$$

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

Based on Equation 20, β_o represents the constant term to describe the overall rates of the process. The intrinsic conditional autoregressive (CAR) priors are applied to fit the model for random effects c_i presented among the observed data and random effects for the unobserved variables in nearby areas. The discrete time discrete space stochastic SIR

model for human leptospirosis proposed in this study thereby used the estimation of relative risk, the results of which are discussed in the next section.

In this study, the intrinsic conditional autoregressive (CAR) priors are applied to fit the model for random effect c_i . The CAR-Normal distribution is used as the prior to allow for spatial dependence between the random effects c_i in nearby areas. These CAR priors were proposed by Besag et al. (1991), where the probability densities of values at any given location are conditional on the neighbouring areas. The CAR priors model used is as Equation 21.

$$\left[c_i \mid c_j, i \neq j, \sigma_c^2\right] \sim N\left(\bar{c}_i, \sigma_i^2\right) \quad [21]$$

where the mean

$$\bar{c}_i = \frac{1}{\sum_j \kappa_{i,j}} \sum_j c_i \kappa_{i,j} \quad [22]$$

and variance

$$\sigma_i^2 = \frac{\sigma_c^2}{\sum_j \kappa_{i,j}} \quad [23]$$

In this context, the κ_{ij} is Equation 22 and 23 are equal to 1 if *i,j* are adjacent and 0 if they are not adjacent. For a full Bayesian analysis, prior distributions must be specified for this hyperparameter. Here, this analysis considers gamma distribution for parameter σ_1^2 that controls the variability of *c*. The discrete time discrete space stochastic SIR model proposed in this study is used as an estimate of relative risk specifically for leptospirosis disease.

Generally, for $i = 1, 2,\ldots,M$ study region and $j = 1, 2,\ldots T$ time periods, the observation sample used is the pseudo-random sample where λ_{ijk} for $k = 1, 2,\ldots, n$ is generated from a posterior distribution for the mean number of leptospirosis infections, $\lambda_{i,j}$ in Equation 19. From this sample of the posterior mean number, the posterior expected mean is approximated using the Equation 24.

$$\tilde{\lambda}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \lambda_{ijk} \quad [24]$$

Meanwhile, the respective relative risk parameter is $\theta_{i,j}$ is defined as Equation 25.

$$Relative\ risk = \theta_{i,j} = \frac{Posterior\ mean\ number, \lambda_{i,j}}{expected\ number\ of\ new\ infectives, e_{i,j}} \quad [25]$$

Based on Equation 25, the pseudo-random sample generated for relative risk parameter is approximated as Equation 26:

$$\tilde{\theta}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \theta_{ijk} = \frac{1}{n} \sum_{k=1}^{n} \frac{\lambda_{ijk}}{e_{ijk}} = \frac{\tilde{\lambda}_{i,j}}{e_{i,j}} \quad [26]$$

Therefore, the posterior expected relative risk for leptospirosis, $\tilde{\theta}_{i,j}$ can be written as the posterior expected mean number of infections, $\tilde{\lambda}_{i,j}$ over the expected number of new infections, $e_{i,j}$ based on the human population of the study regions.

In this study, the value of relative risk is defined based on research by Samat and Percy (2012). A value of relative risk that is equal to zero means people within the study region have no relative risk or infection risk for the disease compared to the people in the population. However, the value of relative risk not ought to be zero since everybody in fact has the possibility or chance of contracting the disease. This is one major drawback of the SMR method that this present study aims to overcome. Next, when the value of relative risk is close to 1 this means that there is no significant difference between the conditional probability that a person in the study region contracts the disease and the conditional probability that a person in the general population contracts the disease.

On the other hand, if the value of relative risk is greater than 1, it indicates that the people within the study region are more likely to contract the disease compared to the people in the population. Conversely, a value of relative risk below 1 indicates a decrease in the likelihood of contracting the disease which means that the people in the study region are less likely to contract the disease compared to people in the population. All the traditional and proposed methods discussed in this study will be applied using WinBUGS software to process leptospirosis data from Malaysia. The results are presented in graphs and tables in the next section.

APPLICATION OF RELATIVE RISK ESTIMATION FOR LEPTOSPIROSIS DISEASE IN MALAYSIA

Application of Relative Risk Estimates based on Existing Models

This section discusses the results of the analyses using the SMR, Poisson-Gamma, Log-normal, BYM and mixture models and utilising leptospirosis data for all states in Malaysia from January 2011 to December 2018. Figure 4 depicts the time series plots of human leptospirosis cases reported for epidemiology months from January 2011 to December 2018 in four selected states with the highest cases throughout the period. It can be clearly seen that the state of Selangor had the highest number of leptospirosis cases in the early years. However, the trend changed in early 2015 as Kelantan recorded more cases than Selangor, after which point the numbers generally decreased in both states from 2016 to 2018. The average number of cases was between zero and 80 from 2011 until the end of 2014 and then rose sharply to 150 at the beginning of 2015, with Kelantan showing the highest number, 278 total cases, followed by Selangor with 143 cases in March 2015. Based on the frequency of reporting, it can be concluded that Selangor and Kelantan had the highest number of human leptospirosis cases. However, from a particular point of view, the higher number of cases reported in Figure 4 is most likely due to the greater

population density and the large size of the two states, which make the data inappropriate for representing true risk.

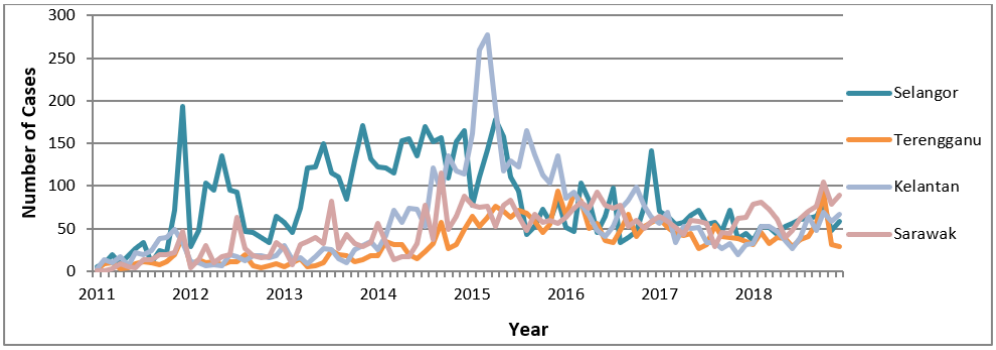

Figure 4. Numbers of human leptospirosis cases in Malaysia from 2011 to 2018

The results of relative risk for the selected highest-risk states using the SMR method from 2011 to 2018 are displayed in Figure 5. Most states have risk between zero to 2.5 except for Melaka (from the end of 2011 to early 2013), Kelantan (from 2014 to 2016) and Terengganu (from 2016 to 2018). This leads to a general conclusion that people in the state of Melaka, Kelantan, Terengganu, and Labuan are more likely to contract leptospirosis compared to the entire population. The risk of Labuan fluctuates over the years due to the high number of observed cases to expected value. Figure 4 shows Selangor frequently has a high number of human leptospirosis cases. However, by using the SMR model, Melaka has shown the highest risk, observed between early 2012 and early 2013 compared to the other states. This is followed by Kelantan and Terengganu in which states the risk began to increase in 2015 until 2018. Thus, the higher number of cases shown in Figure 4 does not reflect the risk of disease infection.

The time series pattern for posterior relative risk plotted for the Poisson-gamma model in Figure 6 is reasonably like the SMR in Figure 5. However, the SMR has a slightly higher range value of risk compared to Poisson-gamma risk. The range risk, which is the largest value minus the lowest value of the risk, in Figure 5 for the SMR is 9 (9 – 0). Meanwhile the range value of risk estimated from Poisson-gamma is 6.629 (6.664 – 0.035). The range is used to compare the spread of risk estimated from different models using a similar dataset. A large range of risk is assumed to have larger variability with more dispersion in the data and vice versa. However, a large range of a dataset may result from outliers. These outliers are unusual values of risk output probably caused by data entry errors, measurement errors or sampling problems. Outliers may also occur because of natural variation especially

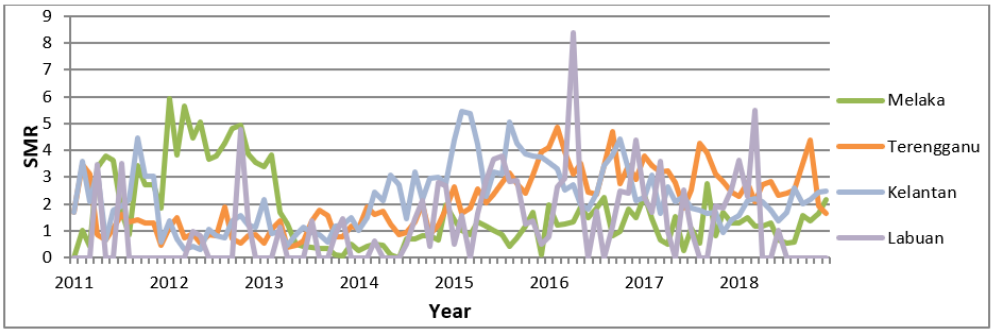

Figure 5. Estimates of relative risk for selected highest risk states based on the SMR of human leptospirosis, 2011 to 2018

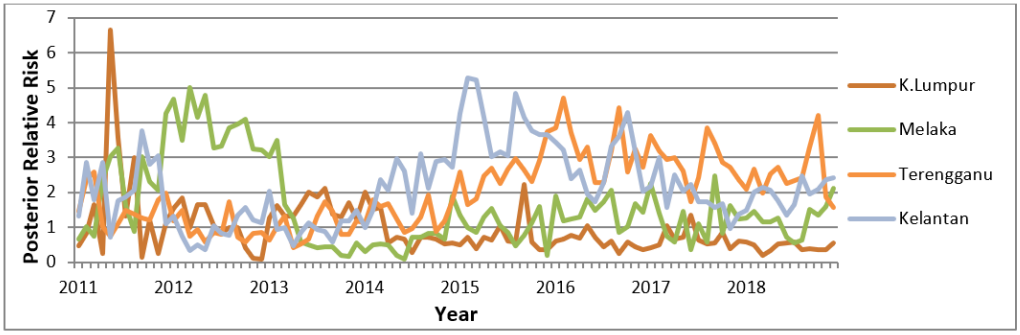

Figure 6. Estimates of relative risk for selected highest risk states based on the Poisson-gamma model of human leptospirosis, 2011 to 2018

when big data is involved. Outliers are useful for providing information about the highest and lowest variability risks inherent in the study areas. In this case, the estimated risk does not necessarily need to be precise towards a mean point. Risk is a variable that occurs naturally and removing outliers sometimes causes the subject areas to appear less variable than they are in reality.

It can be concluded that the use of the SMR offers a slightly higher variability than Poisson-gamma. Besides, the SMR has a drawback when dealing with zero cases. The risk estimated from a zero-case area will have zero risk estimated, as shown in Figure 5, which means that the area has no possibility for anyone there to catch the disease. But everyone has the possibility of being infected with the disease either low or high. Hence, Poisson-gamma in Figure 6 is used to overcome this shortcoming by estimating risk based on gamma posterior distribution that gives weightage for small areas with rare diseases (Lawson & Williams, 2003). The estimated risk shown in the time series graph is more realistic as there are no states with zero potential to contract the disease.

The time series pattern of posterior relative risk for the log-normal model in Figure 7 is quite like the Poisson-gamma model in Figure 6, with a slightly higher range value of risk. Melaka was still recorded as high risk from the end of 2011 to early 2013, while the risk for Kelantan and Terengganu rose gradually and consistently starting from 2014 up to 2018. This slight change of risk values was due to the spatial correlation in which the risk of an area will be affected by the risk from neighbouring areas such as between Kelantan and Terengganu. Furthermore, changes in range of risk may exist once the spatial random effects for correlated heterogeneity in the population are considered.

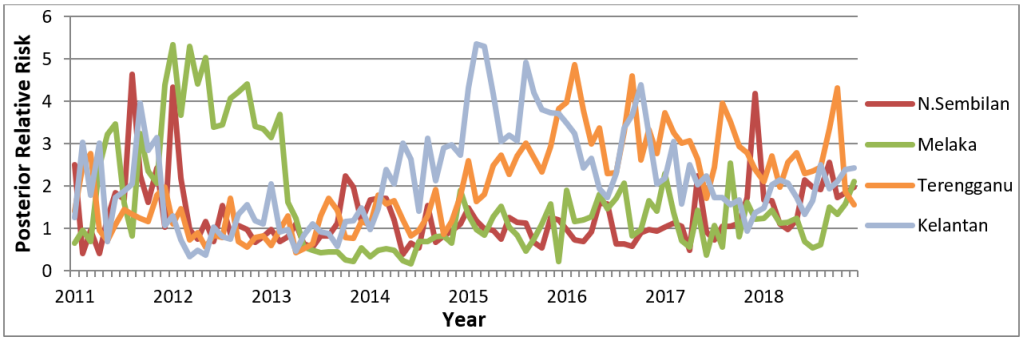

Figure 7. Estimates of relative risk for selected highest risk states based on the Log-normal model of human leptospirosis 2011 to 2018

Based on Figure 8, the time series plots for the BYM model are too smooth and the risk shows less variation than the observed risks for the SMR, Poisson-gamma and log-normal models. Some of the extreme values estimated from the SMR have disappeared. This is because the components in the model consider correlated and uncorrelated heterogeneity of area-specific random effects. The variability of risk estimated depends on uncorrelated heterogeneity. However, in this case variable u controlled by parameter τ_u^2 is small which causes lower variability of risk. Thus, the risk of a particular area is slightly like that of neighbouring areas. Melaka had the highest risk, which gradually decreased from 2011 to 2014, while the risk of Kelantan increased from 2012 to 2016. As the risk in Kelantan decreased, the risk in Terengganu increased from the end of 2016 until 2018.

As discussed before, extreme risk values estimated by the SMR often causes outliers. Without considering the natural variation, these outliers are believed to disfigure statistical analysis and perhaps contravene the assumptions. Subsequently, they produce unclear spatial patterns of risk if any mapping analysis is involved (Lawson & Williams, 2003). Therefore, many researchers have sought appropriate models (known as the smoothing approach) to overcome this problem by estimating the random effects. As a result, and using the same dataset, models such as Poisson-gamma, log-normal and BYM can overcome

the outliers and zero case problems of the SMR. Nevertheless, the BYM model shown in Figure 8 is smoothing 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. According to Lawson (2006), the BYM model is a smoothing model which is not designed specifically for cluster detection. Therefore, a mixture model was used to overcome this problem and simultaneously allow for smoothness and discontinuities.

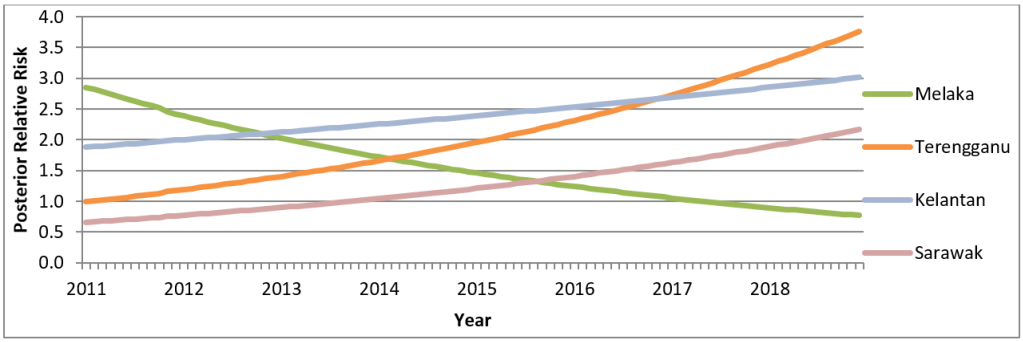

Figure 8. Estimates of Relative Risk for selected highest risk states based on the BYM model of Human Leptospirosis, 2011 to 2018

According to Figure 9, the state of Melaka recorded the highest risk from 2011 to early 2013 and was followed by Kelantan from 2014 to 2016 and Terengganu from 2015 to 2018. The pattern is still like the other observed method in which most states have a risk less than 2 with a small range of risk. For this method, the discontinuity of risks estimation is maintained, and the smoothness is employed at the same time.

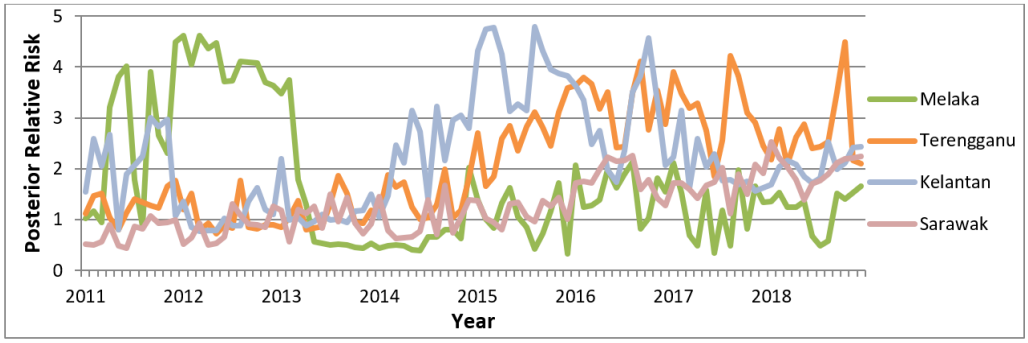

Figure 9. Estimates of relative risk for selected highest risk states based on the mixture model of human leptospirosis, 2011 to 2018

Application of Relative Risk based on Discrete Space-Time Stochastic SIR Transmission Model

The SIR transmission model proposed in this paper utilises the stochastic element in relative risk estimation for leptospirosis in Malaysia. The result of expected relative risk for posterior mean is plotted in the time-series graph in Figure 10.

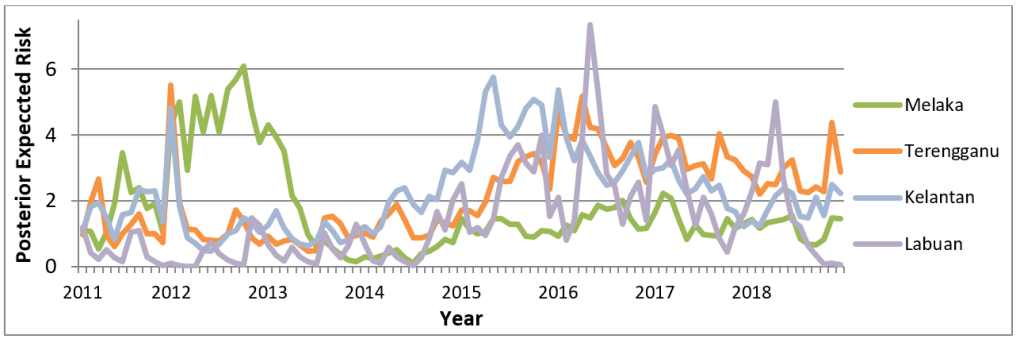

Figure 10. Expected relative risk for selected highest risk states based on the stochastic SIR model of human leptospirosis, 2011 to 2018

The average expected relative risk for most states based on the stochastic SIR transmission model for human leptospirosis is less than 2. It can be assumed that states with an average risk of under 2 are less likely to contract the disease compared to people in the overall population of Malaysia. According to the time-series graph plotted in Figure 10, Melaka, Kelantan, Labuan, and Terengganu have a higher risk compared to the other states. This indicates that these areas are more likely to contract leptospirosis compared to the total population of Malaysia. Besides that, in early 2011 and 2012, Terengganu and Kelantan had relatively high risk which gradually decreased over time. After that, starting in 2015 the risk increased again until 2018. An early sign, like this one from 2011 to 2012, should alert the authorities and communities to any potential increase in the risk of the disease occurring again in these areas.

Based on the expected relative risk between the existing and stochastic SIR models in this paper, it can be concluded that utilising different models produces different patterns of time-series for disease risk plotted. In line with the objective of this study, the proposed model is designed to yield information about areas of excess risk by using the transmission and stochastic approach naturally along with the uncertainty of leptospirosis infection behaviour.

Expected Relative Risk Results of Leptospirosis based on Existing Models and Discrete Space-Time Stochastic SIR Model for Human Leptospirosis Disease

Table 1 shows the comparison of estimated expected relative risk results for the epidemiology month of May 2014, based on the SMR, Poisson-gamma, log-normal, BYM, mixture models and the discrete time-space stochastic SIR model for human leptospirosis. This epidemiology month was chosen to clearly demonstrate the result of posterior expected mean of relative risk for the existing and stochastic SIR models of leptospirosis transmissions.

Table 1

Estimates relative risk for the epidemiology month of May 2014

No.	States	Number of Cases	SMR	Poisson-Gamma	Log-normal	BYM	Mixture	SIR
1	Perlis	0	0	0.3613	0.4097	0.4513	0.3352	0.3736
2	Kedah	1	0	0.0866	0.1303	1.1550	0.5979	0.6504
3	Pulau Pinang	10	0	0.3817	0.3831	0.2892	0.3589	0.1786
4	Perak	59	2	1.7080	1.7040	0.9622	1.4970	1.9580
5	Selangor	155	2	1.8320	1.8330	1.0580	1.5600	1.6380
6	Kuala Lumpur	17	1	0.7317	0.7148	0.9621	0.6821	1.2600
7	Putrajaya	0	0	0.6574	0.6385	0.6288	0.2548	0.0021
8	Negeri Sembilan	5	0	0.4025	0.4041	1.2220	0.7628	1.7090
9	Melaka	1	0	0.1954	0.2420	1.6290	0.4119	0.2621
10	Johor	20	0	0.4199	0.4201	0.2945	0.3993	0.3088
11	Pahang	10	0	0.4951	0.4882	0.8887	0.6389	0.9885
12	Terengganu	20	1	1.2360	1.2110	1.7550	1.2380	1.8860
13	Kelantan	74	3	2.9650	3.0060	2.2990	3.1500	2.3080
14	Sarawak	17	0	0.4831	0.4800	1.0970	0.6566	0.6304
15	Sabah	39	1	0.7736	0.7683	0.6225	0.8922	1.0860
16	Labuan	0	0	0.6573	0.6498	1.4050	1.3820	0.3078
	DIC	-	-	55.264	55.745	56.325	54.197	62.524

Table 1 indicates that Selangor had the highest number of leptospirosis cases in May 2014, 155, followed by Kelantan with 74 leptospirosis cases and Perak with 59 cases. However, expected relative risks estimated by all the existing and proposed models show that the state with the highest risk of catching the disease was Kelantan instead of Selangor. Intrinsically, the higher number of cases in the state of Selangor is because of the large population size and area, and not because of the high-risk occurrence. Meanwhile, the high risk for Kelantan is due to the flood season that affects the north-eastern states of Malaysia (Garba et al., 2017b).

From Table 1, the expected relative risk for the mixture model shows the state with the highest risk is Kelantan with relative risk value 3.15 followed by Selangor, 1.56 and Perak, 1.497. Since Kelantan is the highest risk, the nearest area; Terengganu is expected to have relatively high risk. However, in this case, Terengganu shows low risks of infection which is 1.238. Similarly, with the BYM model, Kelantan is at the highest risk of leptospirosis infection followed by Terengganu and Melaka. In this case, Kelantan neighbours Terengganu, while Melaka is located further away and does not have spatial correlation between those high-risk states. This indicates that these models do not appear to be robust in terms of the covariate element proposed when neighbouring areas have poor relatively influenced by the highest risk adjacent areas. The posterior expected relative risk for the stochastic SIR, on the other hand, shows Kelantan as at higher risk followed by Perak and Terengganu. Geographically, these three states abut against each other and have relatively correlated heterogeneity.

The smallest DIC value in Table 1 is the mixture model with a difference of approximately 1 to 2 with other models. However, it is difficult to conclude that the mixture model is the model most fitted to the data since the DIC value is not greatly dissimilar to the other existing Bayesian models. Next, the DIC value for the stochastic SIR model is 62.524 which is not too far from the others. However, as discussed before, comparing different mathematical models may not be appropriate and the interpretation of DIC values may be misleading. The DIC value comparison for the stochastic SIR model is more suitable compared to the same properties of transmission models. Therefore, regardless of the DIC value, the selection of the model is dependent on the objectives of the study and the availability of data.

CONCLUSION AND FUTURE WORK

In this study, the expected relative risk estimates from existing models have demonstrated high-risk areas contracting leptospirosis across Malaysia from 2011 to 2018. Each model gives different estimates of relative risk based on the components used for estimates of random variables in spatial units, such as an effort to overcome the drawbacks of the SMR. Next, in line with the objective of this study, expected relative risk is derived from

the transmission stochastic SIR model as an estimate of risk. The results disclosed areas at high risk of contracting leptospirosis disease for states in Malaysia which is significant for surveillance systems and control measures.

Compared with existing methods of relative risk estimates, the stochastic SIR model is more relevant and in line with infections due to leptospirosis transmission. Infectious disease is naturally transmitted to susceptible individuals from infectious individuals which finally recovered from the disease. In epidemiology studies, having this transition enables researchers to determine the target population for control measures. For example, susceptible individuals which comprise different age-groups will have different rates of infection. The high infection age-group/s can be easily targeted to stop the transmission. In another case, for different groups of the sub-population, only certain groups are vaccinated to prevent disease outbreak. This indicates the importance of transmission systems in the study of infectious disease. Next, the expected relative risk estimates from newly infective persons are based on the stochastic process. It is important to consider the probability or uncertainty of disease infection so the risk estimates can be more flexible depending on the behaviour of the disease and infection.

However, in this study, the stochastic SIR model only involves human transmission. Since leptospirosis disease originates from rodents, especially rats, it will be more specific and realistic if transmission from the rat population is considered in the model. An extension to this work consisting of a rat population in the transmission model should be the subject of future study. Furthermore, the application of risk would be more effective with the presentation of a risk map.

ACKNOWLEDGEMENT

The authors would like to thank and acknowledge the Institute of Research Management & Innovation (IRMI), Universiti Teknologi MARA (UiTM) for providing FRGS research fund (600-IRMI/FRGS 5/3 (388/2019)) and Ministry of Health Malaysia for the provision of human data infected by leptospirosis disease.

REFERENCES

Andrade, C. (2015). Understanding relative risk, odds ratio, and related terms: as simple as it can get. The Journal of clinical psychiatry, 76(7), 857-861. https://doi.org/10.4088/JCP.15f10150

Awang, A. C., & Samat, N. A. (2017). Standardized morbidity ratio for leptospirosis mapping in Malaysia. In *AIP Conference Proceedings* (Vol. 1847, No. 1, p. 020006). AIP Publishing LLC. https://doi.org/10.1063/1.4983861

Benacer, D., Thong, K. L., Verasahib, K. B., Galloway, R. L., Hartskeerl, R. A., Lewis, J. W., & Mohd Zain, S. N. (2016). Human leptospirosis in Malaysia: reviewing the challenges after 8 decades (1925-2012). *Asia Pacific Journal of Public Health, 28*(4), 290-302. https://doi.org/10.1177/1010539516640350

Besag, J., York, J., & Molli, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*(1), 1-20. https://doi.org/10.1007/BF00116466

Brauer, F. (2017, February 4). *Mathematical epidemiology: Past, present, and future*. Retrieved August 29, 2020, from https://www.ncbi.nlm.nih.gov/pubmed/29928732

Dey, S., Delampady, M., & 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

Fletcher, W. (1928). Recent work on leptospirosis, tsutsugamushi disease, and tropical typhus in the federated Malay States. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 21*(4), 265-288. https://doi.org/10.1016/s0035-9203(28)90019-x

Garba, B., Bahaman, A. R., Bejo, S. K., Zakaria, Z., Mutalib, A. R., & Bande, F. (2017a). Major epidemiological factors associated with leptospirosis in Malaysia. *Acta Tropica, 178*, 242-247. https://doi.org/10.1016/j.actatropica.2017.12.010.

Garba, B., Bahaman, A. R., Khairani-Bejo, S., Zakaria, Z., & Mutalib, A. R. (2017b). Retrospective Study of Leptospirosis in Malaysia. *EcoHealth, 14*(2), 389-398. https://doi.org/10.1007/s10393-017-1234-0

Haake, D. A., & Levett, P. N. (2014). Leptospirosis in humans. *Current Topics in Microbiology and Immunology,* (387), 65–97. https://doi.org/10.1007/978-3-662-45059-8_5

Ideris, S. H., & Samat, N. A. (2015). Comparison of HIV and AIDS diseases mapping in Malaysia based on standardized morbidity ratio and poisson-gamma model. *EDUCATUM Journal of Science, Mathematics and Technology, 2*(1), 69-81.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. *Series A, Containing Papers of a Mathematical and Physical Character, 115*(772), 700-721. https://doi.org/10.1098/rspa.1927.0118

Kismiantini. (2009, January 25). *Relative risk of disease using linear generalized mixed model*. Staff Site Universitas. Retrieved August 29, 2020, from http://staff.uny.ac.id/sites/defaultfiles/penelitian/Kismiantini,S.Si.,M.Si/A2009B1.pdf

Lawson, A. B. (2006). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lawson, A. B., & Williams, F. L. (2003). *An introductory guide to disease mapping*. John Wiley.

Long, N. X. (2012). *Clustering problems, mixture models and Bayesian nonparametrics*. Lecture presented at Vietnam Institute of Advanced Studies of Mathematics (VIASM), Hanoi. Retrieved December 05, 2020, from http://dept.stat.lsa.umich.edu/~xuanlong/Talks/nguyen_viasm12_part1.pdf

Meza, J. L. (2003). Empirical Bayes estimation smoothing of relative risks in disease mapping. *Journal of Statistical Planning and Inference, 112*(1-2), 43-62. https://doi.org/10.1016/S0378-3758(02)00322-1

Nurmalasari, M., & Pramana, S. (2014). Analysis of appendectomy in Belgium using disease mapping techniques. *Proceedings of the 13th Islamic Countries Conference on Statistical Sciences*, *27*, 417-430.

Pooley, C. M., & Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science, 5*(3), Article 171519. https://doi.org/10.1098/rsos.171519

Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives, 112*(9), 1016-1025. https://doi.org/10.1289/ehp.6740

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

Spiegelhalter, D., Thomas, A., Best, N., & Luun, D. (2003). *WinBUGS user manual version 1.4*. MRC Biostatistics Unit.

Sulong, M. R., Shafei, M. N., Yaacob, N. A, Hasan, H., Daud, A., Mohamad, W. M. Z. W., Zaliha, I., & Mohamed, R. A. (2011). Risk factors associated with leptospirosis among town service workers. *International Medical Journal, 18*(2), 83-88.

Triampo, W., Baowan, D., Tang, I. M., Nuttavut, N., Wong-Ekkabut, J., & Coungchawee, G. (2008). A simple deterministic model for the spread of leptospirosis in Thailand. *International Journal of Medical and Health Sciences, 2*(1), 22-26.

Wahab, Z. A. (2015, September 8-9). Epidemiology and current situation of leptospirosis in Malaysia. In *Persidangan Kesihatan Persekitaran Pihak Berkuasa Tempatan 2015*. WP