ASM Sc. J., 20(1), 2025

https://doi.org/10.32802/asmscj.2025.1794

The Relative Risk Estimation of Pneumonia in Malaysia using Besag, York and Mollie (BYM)

I. Mohd Diah[1*] and N. Aziz[2,3]

[1]*Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900, Tanjong Malim, Perak, Malaysia.*

[2]*Department of Mathematics and Statistics, School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010, Sintok, Kedah Darul Aman, Malaysia.*

[3]*Institute of Strategic Industrial Decision Modelling (ISIDM), Universiti Utara Malaysia, 06010, Sintok, Kedah Darul Aman, Malaysia.*

Good relative risk values are important in the production of a disease risk map. This disease risk map can clearly show the risk areas of the disease occurrence. In order to get good relative risk values, choosing a good statistical model is a significant part that needs to be considered when studying disease occurrence. Hence, the objective of this paper is to estimate the relative risk values for pneumonia transmission based on the Besag, York and Mollie (BYM) model. The analysis involved pneumonia data in Malaysia for the year 2010 until the year 2019. Results show that Putrajaya has been classified as a very high-risk area with a 4.546 risk value, while Pulau Pinang shows the lowest risk area of contracting pneumonia.

Keywords: BYM model; disease mapping; pneumonia; relative risk

I. INTRODUCTION

Pneumonia is a lung illness that occurs when alveoli get infected with inflammatory cells. Microorganisms such as parasites, bacteria, viruses and fungi can cause someone to get the disease (WHO, 2019). Viruses and bacterial pneumonia are the most prevalent causes of pneumonia in humans. These viruses include SARS-COV-2, the virus that causes novel Coronavirus-2019 (COVID-19), which has already caused a worldwide outbreak and can cause pneumonia (International Vaccine Access Center, 2020). A number of variables, including age, the kind of organism that caused the inflammation, and general health, affect how severe the symptoms are. People who have had pneumonia typically experience chills, sweating, a high-grade fever, and green, yellow, or blood-stained sputum when coughing. For minor symptoms, it appears as if you have the flu or a cold, but it lasts longer. Pneumonia can be transmitted in a variety of ways. If viruses or bacteria detected in a child's nose or throat are breathed, they can infect the lungs. Pneumonia can also be transferred via air when an infected person sneezes or coughs. This is due to the tiny droplets that contain the bacteria. Furthermore, pneumonia can spread through the blood, particularly during and immediately after birth. More study is needed on the many bacteria that cause pneumonia and how they spread, as this is crucial for treatment and prevention.

Pneumonia can range in severity from mild to potentially fatal. The most vulnerable groups are infants and young children, especially those under five, as well as people 65 years of age and above. This is because the two groups are with weak immune systems, which may become more dangerous if infected with pneumonia. According to UNICEF (2021), even though pneumonia is a preventable infection, it causes more child deaths than any other disease. In 2019, based on the report by the Department of Statistics Malaysia (2019), pneumonia was one of the top three causes of mortality in Malaysia with 7, 542 deaths from 145,419 total pneumonia cases reported. Figure 1 shows the number of pneumonia cases reported in 2019 for every state in Malaysia.

*Corresponding author's e-mail: ijlalmd@gmail.com

From Figure 1 below, Labuan has been recognised as having the lowest number of pneumonia cases, with 592 cases, while Selangor has the highest number of pneumonia cases with 19 481 cases, among others. It is vital to investigate this issue since complications for persons infected with influenza and COVID-19 can lead to an increase in the number of pneumonia cases. This disease may become an outbreak if not controlled, as the number of pneumonia patients may increase because of other diseases.

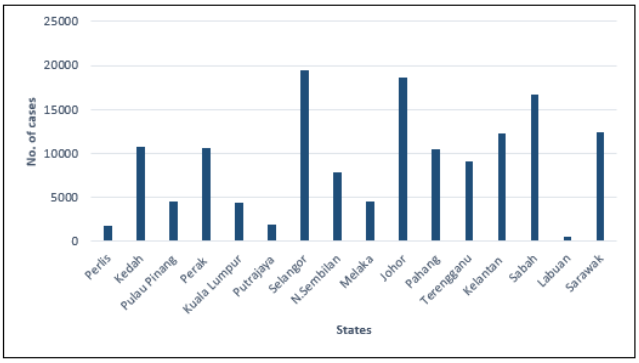

Figure 1. Number of pneumonia cases in 2019 based on every state in Malaysia

In Malaysia, current monitoring of high and low risk areas is still based on the numbers of pneumonia incidences throughout the regions. Large numbers of pneumonia cases, for example, in specific locations, represent high risks of pneumonia occurrence without taking into account other factors such as population size or region size. This technique simply provides general information and does not consider the location or area of disease transmission. Usually, before classifying an area as a high or low-risk area for disease transmission, the population size and the size of the area need to be considered. According to Samat and Percy (2012), disease mapping can be used to monitor infectious diseases.

Disease mapping is an important tool in public health research since it may be used for both disease monitoring and prevention (Samat & Percy, 2012). It may be used as a descriptive image to represent the prevalence of pneumonia in certain geographical areas. According to Diah, Aziz and Ahmad (2016), a good risk map is based on mathematical models that estimate the relative risk. When studying geographical distribution, one of the most important issues is relative risk estimation. A good value of relative risk can help to get good disease mapping. Relative risk is the ratio of the exposed group that will contract disease to the unexposed group that will contract the same disease. This paper first briefly describes the methodology used to estimate the relative risk using the Besag, York and Mollie (BYM) model. This is followed by an illustration of the application, the study's results, and a discussion. Finally, the conclusion is drawn.

II. MATERIALS AND METHOD

A. Besag, York and Mollie (BYM)

Clayton and Kaldor (1987) were the first to use empirical Bayesian inference for relative risks, which was further extended by Besag *et al.* (1991). This model for relative risks decomposes area-specific random effects into two components: one that considers effects that vary in a structured manner in space (clustering or correlated heterogeneity) and another that considers effects that vary in an unstructured manner across areas (uncorrelated heterogeneity). The inclusion of these random effects allows for the smoothing of relative risk at every level (global and local).

The observed number of pneumonia cases (y_i) in area i is assumed to follow a Poisson distribution with mean $e_i\theta_i$, and is stated as follows:

$$y_i \sim Poisson\ (e_i\theta_i) \quad (1)$$

Here, e_i refers to the expected number of cases in area i and θ_i refers to the "true" but unknown relative risk in area i. At the next level of the model, the variability of the log relative risk, $\log\ \theta_i$, is divided into three components:

$$log\ \theta_i = \alpha + u_i + v_i. \quad (2)$$

Here, α is the overall level of relative risk, u_i refers to the spatial random effect indicating correlated heterogeneity, and v_i stands for the random effect reflecting uncorrelated heterogeneity.

However, because space-time data was employed in this study, this BYM model took into account the space-time distribution. Both the particular intercept and the temporal trend are modelled as random effects in Bernardinelli *et al.*'s (1995) model. Different geographical locations and even

spatial structures are possible with this formulation. All temporal trends, however, are expected to be linear, which is a restrictive assumption. The relative risk model has the following form:

$$log\ \theta_{ij} = \alpha + u_i + v_i + \beta.t_j + \delta_i.t_j, \quad (3)$$

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

In Bayesian modelling, prior distributions for random effects must be specified. v_i is the uncorrelated heterogeneity distribution model,

$$v_i \sim N(0, \tau_v^2). \quad (4)$$

This distribution is considered to follow a normal distribution with a zero mean and a common variance (precision parameter), τ_v^2.

The clustering component uses a spatial correlation structure with risk estimates in each location reliant on neighbouring areas. Besag *et al.* (1991) introduced the conditional autoregressive (CAR) model, in which the value of a parameter in an area is influenced by the average value of the neighbourhood, with extra variability calculated with conditional variance depending on the number of neighbours.

$$[u_i | u_j, \mathrm{i} \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2), \quad (5)$$

$$\bar{u}_i = \frac{1}{\sum_j \omega_{ij}} \sum_j u_j \omega_{ij}, \quad (6)$$

$$\tau_i^2 = \frac{\tau_u^2}{\sum_j \omega_{ij}}, \quad (7)$$

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

Parameters τ_v^2 and τ_u^2 control the variability of random effects v and u. Here, ω_{ij} denotes the relationship between the areas i and j. The prior mean of each u_i is defined as a weighted average of the other $u_j, \neq j$.

B. Data Set

Malaysia's Department of Statistics and the Ministry of Health both contributed to the data set used in this study. This pneumonia data has been applied using the BYM method, which is in the form of the number of cases for 13 states and 3 federal territories in Malaysia from 2010 until 2019. To compute the relative risk in this study, WinBUGS software is used, which implements the Markov Chain Monte Carlo (MCMC). From the relative risk results, a pneumonia risk map is constructed using ArcGIS software.

III. RESULT AND DISCUSSION

The outcomes of relative risk estimation for all states in Malaysia are shown in Figure 2. From the graph, the Federal Territory of Putrajaya and the states of Perlis, Kedah, Perak, Negeri Sembilan, Melaka, Johor, Pahang and Terengganu have a relative risk greater than one for most epidemiology years. This means that persons in these states are more likely to contract pneumonia compared to the persons in the overall population. In this study, we defined the relative risk to be the conditional probability that persons in a specific area contract pneumonia divided by the conditional probability that the whole population is infected with the disease. When a value of relative risk is greater than one, it implies that people in the area have more tendency to be infected with this disease than people in the entire population. If a value of relative risk is close to one, it indicates that there is no significant difference in the chance of people suffering from pneumonia in a certain area or in the entire population. In contrast, for a relative risk value of less than one, people in that area are less likely to contract the disease than those in the whole population.

Based on Figure 2, there are no posterior expected relative risk values for the epidemiology year 2010. This is due to the situation called moving average, where the results are calculated by averaging a number of past data points. From Figure 2, it can be seen that the relative risk values are too smooth and appear to have lost the jump discontinuities that would have been apparent in relative risk values if using other approaches.

Table 1 shows the results of the posterior expected relative risk for the epidemiology year 2019. From Table 1, Perlis, Putrajaya, Negeri Sembilan, and Terengganu have relative risk values greater than 1.50. This condition suggests that individuals in these states are more likely to contract pneumonia than the rest of Malaysia's population, whereas people in other states are less likely to contract pneumonia than the rest of the population in Malaysia. Pulau Pinang has

the smallest value of the relative risk of contracting pneumonia, with 0.530, when compared to other states in Malaysia, while Putrajaya has the highest risk value of pneumonia infection, with 4.546. These findings are equivalent to the number prevalence of pneumonia, which was predicted in 2019 to be 25 per 10,000 persons in Pulau Pinang and 176 per 10,000 persons in Putrajaya.

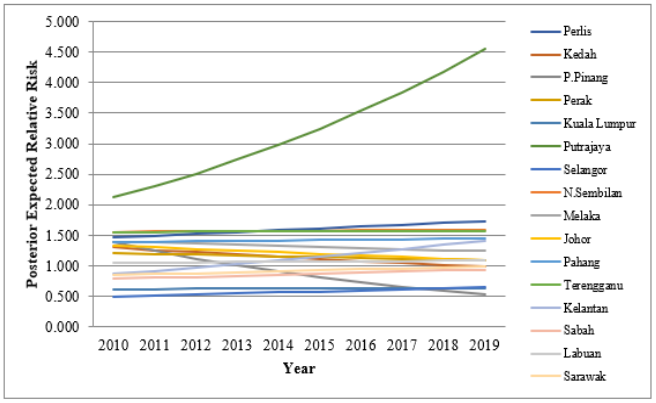

Figure 2. Plot of time series for relative risk values based on BYM model

Table 1. Relative risk values for the epidemiology year 2019 based on the BYM model

States	Relative Risk
Perlis	1.730
Kedah	0.981
P. Pinang	0.530
Perak	1.090
Kuala Lumpur	0.629
Putrajaya	4.546
Selangor	0.653
Negeri Sembilan	1.592
Melaka	1.241
Johor	1.090
Pahang	1.448
Terengganu	1.572
Kelantan	1.414
Sabah	0.938
Labuan	1.083
Sarawak	0.993

A thematic map is used in this study to distinguish between high and low-risk areas for the incidence of pneumonia cases in Malaysia. Each state is assigned one of five relative risk levels: very low, low, medium, high and very high risks with respective intervals of [0.0, 0.5), [0.5, 1.0), [1.0, 1.5), [1.5, 2.0) and [2.0,∞). The brightest shade colour represents the lowest risk area, while the darkest colour represents the extremely high-risk area. Figure 3 shows the risk map of pneumonia disease for the year 2019.

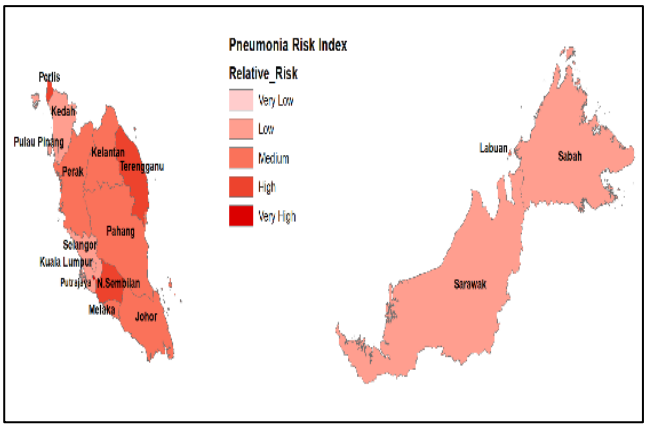

Figure 3. Risk map estimated using the BYM model for the year 2019

From Figure 3, Putrajaya has been identified as a very high-risk area of contracting pneumonia disease. This is followed by high-risk areas which are Perlis, Terengganu and Negeri Sembilan. For the states with medium risk of contracting pneumonia, Perak, Melaka, Johor, Kelantan, Pahang and Labuan have been classified into this category. Another seven states have been classed as low-risk areas. There is no state classified as a very low-risk area. According to statistics from the Annual Report 2019 from the Ministry of Health Malaysia (2019), Putrajaya had the highest annual population growth rate of 6.55% compared to other states, Putrajaya is Malaysia's smallest region (Department of Survey and Mapping Malaysia, 2017). Hence, susceptible people in Putrajaya have the highest risk of getting infected with pneumonia since Putrajaya has the smallest area and the highest yearly population growth rate. From this information, it can be concluded that the size of the area and the number of people in the area can influence the value of relative risk estimation.

IV. CONCLUSION

Estimating relative risk values is important in monitoring and controlling pneumonia. Researchers used the BYM model to estimate the relative risk. In conclusion, Putrajaya has been identified as the highest-risk area with the highest value of risk with 4.546, while Pulau Pinang has been recognised as the lowest-risk area with the lowest risk value of 0.530. This method, however, has drawbacks. BYM model is a smoothing model that is not intended specifically for cluster detection (Lawson, 2006). Hence, this situation has prompted other academics to propose novel ways of estimating relative risk.

V. ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Health Malaysia and the Department of Statistics Malaysia for providing the data and Universiti Utara Malaysia for the support of this study.

VI. REFERENCES

Bernardinelli, L, Clayton, D, Pascutto, C, Montomoli, C, Ghislandi, M & Songini, M 1995, 'Bayesian analysis of space—time variation in disease risk', Statistics in medicine, vol. 14, no. 21-22, pp. 2433-2443.

Besag, J, York, J & Mollie, A 1991, 'Bayesian image restoration with two applications in spatial statistics', Annals of the Institute of Statistical Mathematics, vol. 43, pp. 1-20.

Clayton, DG & Kaldor 1987, 'Empirical Bayes estimates of age-standardised relative risks for use in disease mapping', Biometrics, vol. 43, pp. 671-691.

Diah, IM, Aziz, N & Ahmad, N 2016, 'Relative Risk Estimation of Tuberculosis with Standardized Morbidity Ratio in Malaysia', Global Journal of Pure and Applied Mathematics (GJPAM), vol. 12, no. 5, pp. 4011-4019.

Department of Statistics Malaysia 2019, Statistics on Causes of Death, Malaysia, 2019, viewed on 20 November 2021, <https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=RUxlSDNkcnRVazJnakNCNVN2VGgrdz09&menu_id=LopheU43NWJwRWVSZklWdzQ4TlhUUT09>.

Department of Survey and Mapping Malaysia 2017, Keluasan Malaysia 2017, viewed on 27 November 2021, <https://www.data.gov.my/data/ms_MY/dataset/keluasan-malaysia/resource/a04e46de-5044-4081-b6cc-7853a46439b>.

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

Lawson, AB 2006, Statistical methods in spatial epidemiology, Wet Sussex, UK: John Wiley & Sons, Ltd.

Lawson, AB, Browne, WJ & Rodeiro, CLV 2003, Disease mapping with WinBUGS and MLwiN. Statistics in Practise, Chichester: John Wiley & Sons, Ltd.

Meza, JL 2003, 'Empirical Bayes estimation smoothing of relative risks in disease mapping', Journal of Statistical Planning and Inference, vol. 112, pp. 43-62.

Ministry of Health Malaysia 2019, Laporan Tahunan Kementerian Kesihatan Malaysia 2019 (ISSN 1511-1520), viewed on 27 November 2021, <https://www.moh.gov.my/moh/resources/Penerbitan/Penerbitan%20Utama/ANNUAL%20REPORT/LAPORAN%20TAHUNAN%20KKM%202019/mobile/index.html#p=1>.

Samat, NA & Percy, DF 2012, 'Dengue disease mapping in Malaysia based on stochastic SIR models in human populations', in ICSSBE 2012: Proceedings of 2012 International Conference on Statistics in Science, Business and Engineering: "Empowering Decision Making with Statistical Sciences," pp. 623–627.

UNICEF, "Pneumonia: A child dies of pneumonia every 39 seconds", viewed on 25 November 2021, <https://data.unicef.org/topic/child-health/pneumonia/>

World Health Organization (WHO) 2019, Pneumonia Fact Sheet, viewed on 25 November 2021, <http://www.who.int/news-room/fact-sheets/detail/pneumonia>.