Disease Mapping of Leprosy in Maluku Province, 2019 using Hierarchical Bayes

I F Mahdy1, R Fitriani1, W D Revildy1
1Departement of Statistics, Universitas Padjadjaran

*E-mail: Ilham19016@mail.unpad.ac.id

Abstract. Leprosy is a chronic disease caused by *Mycobacterium leprae* that affects the nerves of the extremities, skin, nasal mucosa, and upper respiratory tract. As of today, some countries include Indonesia facing difficulties to fight leprosy. The health ministry of Indonesia noted that the majority of leprosy cases occurred in the eastern region, like Maluku Province. Therefore disease mapping is needed to develop a map of leprosy so that the handling of this disease could operate optimally, especially in Maluku. One of the disease mapping techniques is Hierarchical Bayes (HB), which involves a measure of uncertainty compared to other methods such as Standardized Mortality Ratio (SMR). The results showed that the relative risk of leprosy in Maluku was almost the same in each region. It indicates that the disease is prevalent in Maluku, with the highest value in Ambon City and the lowest value in the Aru Islands.

1. Introduction

Disease mapping has a long history in various fields, especially epidemiology. Disease mapping is used to determine disease risk by region. The disease mapping methodology is based on the Bayes Theorem aiming at calculating the estimated parameter values using data on the number of occurrences of a disease and several assumptions related to these occurrences. The main problem that is always faced in disease mapping is the choice of measurement techniques that allow to measure the relative risk of a disease. One of the techniques commonly used in disease mapping is the standardized mortality ratio (SMR). However, there is a weakness in the SMR, in which the resulting variance is large for areas with small expected values and small for areas with high expected values, so that the SMR becomes unreliable. This complicates the interpretation of SMR [1]

Some researchers have tried to overcome the weaknesses of SMR, one of them is by using information from other regions through the Empirical Bayes (EB) method to reduce the total mean-square error, as was done by Clayton and Kaldor [2], which revealed that relative risk parameters of $\theta_i$ are considered as random variables, whose prior form is from the Gamma distribution $(a, b)$. The relative risk estimator is the average of $\hat{\theta}$, namely the combination of the SMR estimator and the correction of the mean of the priors. Parameters $a$ and $b$ are estimated from $O$, so that $g(0_i)$ are negative binomial distributions with parameters $a$ and $p$, so that the maximum likelihood for $a$ and $b$ is obtained as an iterative solution of the log likelihood equation. Derivative log likelihood function must be solved using iterative method, such as Newton Raphson. The main obstacle to the EB approach is the unmeasurable uncertainty problem of the MLE estimator obtained. The variance of the posterior estimate is insensitive
to changes caused by model estimates.

Because of the weaknesses of EB method, another method is used, namely the Hierarchical Bayes (HB) approach to estimate the relative risk and its uncertainty size. In HB a parameter is typically estimated by its posterior mean and its posterior variance is used as a measure of error in estimation. In its application the HB method is easier to understand, but often requires high dimensional integration. To determine the posterior density, the Gibbs sampling method is used.

Research on disease mapping in Indonesia has done a lot. For the example, Jaya et. al. discussed about dengue disease risk [3] and clustering diarrhea [4] in their research. Moreover, Yupiana et. Al [5] talked about human rabies cases in Bali. However, research about leprosy phenomenon, especially in the eastern region of Indonesia were rarely done. Therefore, this study will study the disease mapping of leprosy in Maluku. Thus, the policies could then be taken to overcome the spread of leprosy in Maluku.

2. Methodology

The data used in this analysis are data on the number of people with leprosy and population data according to districts/cities in Maluku. The data were obtained from Maluku Province in Figures in 2020 [6]. The number of areas observed was 11 districts/cities.

2.1 Disease Mapping

The basic concept of disease mapping is a description of the distribution of disease mapping. As knowledge develops, the concept of disease mapping can describe the distribution of disease maps in a wide area and show the diversity of diseases by measuring the value of relative risk. Disease mapping for a region $i$ is done by mapping the number of patients in region $i$ ($O_i$) against the expected rate in that region ($E_i$), which is called the Standardized Mortality Ratio (SMR) with the following measurements [1]:

$$ SMR_i = \frac{O_i}{E_i} \tag{1} $$

Where the $E_i$ value is the multiplication of the number of sufferers in an area with a population of $N$ with the proportion of the population of that area. The information generated from this ratio is information that explains in which areas the number of disease sufferers is distributed excessively or smaller than expected. SMR is an estimator of $\theta$ which represents a relative risk in the region and its magnitude is unknown. SMR is also a maximum likelihood estimator if the observation $O_i$ is Poisson distributed with parameters ($E_i\theta_i$). The likelihood function that is formed is as follows:

$$ L(\theta | O) = \prod_{i=1}^{m} \frac{e^{E_i\theta_i} (E_i\theta_i)^{O_i}}{O_i!} \tag{2} $$

The solution to equation (2) is:

$$ SMR = \hat{\theta} = \frac{O_i}{E_i} \tag{3} $$

With the mean and variance of the SMR are $\theta$ and $\frac{\theta^2}{E_i}$. Based on the description above, it shows the weaknesses of the SMR. The variance of the SMR is large for areas with low expectations and vice versa, thus it complicates the interpretation of the SMR. To overcome this weakness a number of researchers tried various ways, one of them is by using the Empirical Bayes (EB) method.

2.2 Empirical Bayes

Suppose that relative risk $\theta_i$ is considered as a random variable, with the prior form of the Gamma distribution ($a$, $b$), so the EB model developed can be written in the following form:

$$ y_i \sim P(E_i\theta_i) $$

$$ \theta_i \sim G(a, b) $$
so the posterior form of $\theta_i$ is:

$$
(\theta|\mathbf{O}) \propto \prod_{i=1}^{m} \frac{e^{-E_i\theta_i}}{\theta_i^{a_i}} (E_i\theta_i)^{a_i} \times \frac{b^a}{\Gamma(a)} \theta_i^{a-1} e^{-\theta_ib} 
$$

(4)

The simplification of the form above produces a posterior form in the form of a Gamma distribution $(a^*, b^*)$ with $a^* = o_i + a$ and $b^* = E_i + b$. So that the estimators of relative risk in empirical bayes are as follows:

$$
\hat{\theta} = \frac{o_i+a}{E_i+b}
$$

(5)

This estimator is a combination of the SMR estimator and the correction of the mean of the priors. Furthermore, in this approach the parameters a and b are estimated through the marginal of O, namely $g(o)$ with the following function:

$$
g(o_i) = \left(\frac{o_i + a - 1}{a - 1}\right) p^a (1-p)^{o_i} 
$$

(6)

This function is a negative binomial distribution with the parameters $a$ and $p = \left(\frac{b}{E_i+b}\right)^{o_i}$. So the likelihood functions for $a$ and $b$ are:

$$
\ell(a,b|o) = -\sum_{i}^m o_i \log(E_i+b) + ma \log(b) - a \sum_{i}^m \log(E_i+b) + \sum_{i}^m \log(\Gamma(o_i+a)) - m\log(\Gamma(a)) + c
$$

(7)

The above equation is solved using the iterative method, namely Newton Raphson.

2.3 Hierarchical Bayes

The Empirical Bayes method has a weakness, in which the uncertainty of the Maximum Likelihood Estimator estimator is obtained. So that Maiti [7] developed the Hierarchical Bayes method approach by looking at the parameters of the priors as random variables called hyperparameters. Based on the $a$ and $b$ values obtained from the results of Newton Raphson’s iterative method, the parameter estimator of the Hierarchical Bayes (HB) method uses log-normal prior. This estimator is defined by the following model [7]:

$$
y_i \sim P(E_i\theta_i) \\
\log(\theta_i)|\mu, r \sim N(\mu, r^{-1}) \\
\mu \sim U(-\infty, \infty) \\
r \sim G(\frac{a}{2}, \frac{b}{2})
$$

Posterior for parameter $\theta$, $\mu$, $r$ conditional $y$ data and parameter $E$, obtained based on the rule of multiplication and conditional probability, assuming $g(\theta, \mu, r|y,E)$ where:

$$
g(\theta, \mu, r|y,E) = P(\theta|\mu, r, y,E) \times P(\mu|r, y,E) \times P(r|y,E) \times P(y|E)
$$

(8)

Then substitute each component in the above equation to obtain:

$$
\begin{align*}
g(\theta, \mu, r|y,E) & \propto \prod_{i=1}^{m} \frac{e^{-E_i\theta_i}}{y_i!} (E_i\theta_i)^{y_i} \times \frac{(b/2)^{a/2}}{\Gamma(a/2)} r^{-a/2-1} e^{-rb/2} \\
& \times \frac{r}{\theta_i \sqrt{2\pi}} e^{\exp\left(-\frac{r}{2}(\log(\theta_i) - \mu)^2\right)} \\
& \times 1
\end{align*}
$$

(9)

The posterior above shows the involvement of a high-dimensional integral in doing inference related to the related parameters, while the shape of the posterior is unknown. Therefore, inference work is carried out through MCMC (Markov Chain Monte Carlo) simulations using the Gibbs sampling and/or Metropolis Hasting algorithm. For doing this, it is necessary to find the full conditional distribution of
the posterior formed, assuming \( h(\theta | \mu, r), k(\mu | \theta, r) \) dan \( l(r | \theta, \mu) \) with

\[
h(\theta | y, \mu, r) \propto \prod_{i=1}^{m} \theta_i^{y_i-1} \times \exp \left(-E_i \theta_i - \frac{r}{2} (\log(\theta_i) - \mu)^2 \right)
\]

\[
k(\mu | \theta, y, r) \propto \prod_{i=1}^{m} \exp \left(-\frac{r}{2} (\log(\theta_i) - \mu)^2 \right)
\]

\[
l(r | \theta, y, \mu) \propto \prod_{i=1}^{m} r^{1/2(a+1)-1} \exp \left(-r \frac{1}{2} ((\log(\theta) - \mu)^2 + b) \right)
\]

The \( k(.) \) Distribution above is known as a normal distribution with parameters \[\frac{1}{D} \sum_{i=1}^{D} \log \theta_i, (D \tau)^{-1}\] and \( l(.) \) is known as the gamma distribution with parameters \[\frac{1}{2} (D + a), \frac{1}{2} \left[ \sum_{i=1}^{D} (\log \theta_i - \mu)^2 + b \right]\], but \( h(.) \) is not known for its shape. Therefore, MCMC simulation is done using the Metropolis-Hasting algorithm for \( h(.) \), while \( k(.) \) and \( l(.) \) use the Gibbs algorithm.

2.4. Markov Chain Monte Carlo (MCMC)

The MCMC method is an approach method for Bayesian inference. According to Walsh [8], MCMC is used to obtain twenty parameters of the estimated value. This is done by simulating direct sampling of a complex posterior distribution. The main concept in MCMC is to sample an approximation of the posterior distribution of parameters by generating a Markov chain that has a distribution limit close to the posterior distribution of the parameter. The posterior distribution of parameters is obtained by determining the prior distribution first.

In MCMC, there are two kinds of algorithms, namely Metropolis-Hasting and Gibbs sampling. The Metropolis-Hasting algorithm is an algorithm for generating sample sequences using an acceptance and rejection mechanism. The Metropolis-Hasting algorithm is used when there is one unknown parameter value. While the Gibbs sampling algorithm is a special case of the Metropolis-Hasting algorithm. This algorithm requires all conditional distributions of the parameters to be sought. The Gibbs sampling algorithm is used when there is more than one unknown parameter.

2.5 Metropolis Hasting Algorithm

At first, Metropolis et al. formulated the Metropolis algorithm by introducing the Markov chain based on simulation methods used in natural science. Then Hastings generalized the method known as the Metropolis-Hastings algorithm [9]. The Metropolis-Hasting algorithm is as follows:

1) Specify the initial value of the parameter, eg \( \theta^{(0)} \)
2) For \( t = 1, 2, \ldots, T \):
   a) Determine \( \theta = \theta^{(t-1)} \)
   b) Determine a new candidate \( \theta' \) from a proposal distribution
   c) Calculate:
   \[
   \alpha = \min \left( 1, \frac{f(\theta')g(\theta | \theta')}{f(\theta)g(\theta' | \theta)} \right)
   \]
   d) \( \theta^{(t)} = \theta' \) with chance \( \alpha \) and \( \theta^{(t)} = \theta^{(t-1)} \) with chance \( 1 - \alpha \)

2.6 Gibbs Sampling Algorithm

Gibbs Sampling was introduced by Geman and Geman [10]. This algorithm is a special case of the single-component Metropolis-Hastings algorithm which uses the density of the proposal. Often, these conditional distributions have a known form, so number of random values can be easily simulated using standard functions in statistical and computational software. Gibbs sampling always moving to new values and does not require the specification of the proposal distributions [10]. The Gibbs sampling algorithm is as follows:

1) Determine the initial value of the parameters, for example \( \theta^{(0)}, \mu^{(0)}, \) and \( r^{(0)} \)
2) Calculate the parameters of each distribution

\[ \mu \sim N \left( \frac{1}{D} \sum_{i=1}^{D} \log \theta_i, (Dr)^{-1} \right) \]

\[ \tau \sim \text{Gamma} \left[ \frac{1}{2} (D + a), \frac{1}{2} \sum_{i=1}^{D} (\log \theta_i - \mu)^2 + b \right] \]

3) Generate parameter values from:

\[ \mu \text{ from } f(\mu|\theta^*, \mu^*) \]

\[ \tau \text{ from } f(\tau|\theta^*, \mu^*) \]

3. Result and Discussion

In this study, Hierarchical Bayes Estimation was used to analyze data on leprosy sufferers in Maluku Province in 2019. The highest number of leprosy sufferers was in the Tanimbar Islands with 64 people while the lowest number of people with leprosy was 18 people in Southwest Maluku.

Figure 1. Distribution of Leprosy Patients by Regency/City in Maluku Province 2019

Leprosy patient data is discrete count data that follows a Poisson distribution or is also called distribution rare events. The expected value and SMR are obtained with the results shown in table 1.

| No. | Regency/City       | Leprosy Patients | Total population | Expectation Value | SMR |
|-----|--------------------|------------------|------------------|-------------------|-----|
| 1   | Kepulauan Tanimbar | 64               | 113012           | 30,34             | 0,92|
| 2   | Maluku Tenggara   | 46               | 99790            | 26,79             | 0,71|
| 3   | Maluku Tengah     | 46               | 373378           | 100,24            | 0,80|
| 4   | Buru               | 62               | 143688           | 38,57             | 1,72|
| 5   | Kepulauan Aru      | 37               | 96114            | 25,80             | 1,61|
No. | Regency/City   | Leprosy Patients | Total population | Expectation Value | SMR |
---|----------------|------------------|------------------|-------------------|-----|
6   | Seram Bagian Barat | 37               | 171586           | 46.06             | 1.53 |
7   | Seram Bagian Timur  | 22               | 114677           | 30.79             | 0.68 |
8   | Maluku Barat Daya  | 18               | 73103            | 19.63             | 1.43 |
9   | Buru Selatan       | 26               | 63328            | 17.00             | 0.46 |
10  | Ambon             | 87               | 478616           | 128.49            | 2.11 |
11  | Tual              | 39               | 75578            | 20.29             | 1.92 |
**Total** |                | **484**          | **1802870**      |                   |     |

From the SMR value obtained, the next step is to calculate the initial value of the parameter that will be used in the full conditional distribution of the posterior which is shown in table 2. The initial values for the parameters \( a \) and \( b \) are obtained using Newton Raphson's iteration, while \( \mu \) dan \( r \) each is the mean and inverse of the variance from the SMR log.

**Table 2. Parameter Initial Value**

| Parameter | Initial Value |
|-----------|---------------|
| \( a \)   | 9.13          |
| \( b \)   | 5.19          |
| \( \mu \) | 0.13          |
| \( r \)   | 3.90          |

Based on the available information, parameter estimation \( \theta \) is carried out using Hierarchical Bayes (HB). In HB, a posterior form is formed in the form of a high-dimensional integral so that the solution is through MCMC (Markov Chain Monte Carlo) simulation. The parameter estimates \( a, b, \mu, \) dan \( r \) were obtained using the Gibbs algorithm. Furthermore, the parameter estimation is used in estimating the parameters \( \theta \) through the Metropolis Hasting algorithm.

**Table 3. Estimated Parameters \( \theta \)**

| No. | Regency/City   | \( \hat{\theta} \) |
|-----|----------------|---------------------|
| 1   | Kepulauan Tanimbar | 1.50                |
| 2   | Maluku Tenggara  | 1.54                |
| 3   | Maluku Tengah    | 1.52                |
| 4   | Buru             | 1.55                |
| 5   | Kepulauan Aru    | 1.47                |
| 6   | Seram Bagian Barat | 1.51              |
| 7   | Seram Bagian Timur | 1.49              |
| 8   | Maluku Barat Daya | 1.51              |
| 9   | Buru Selatan     | 1.54                |
| 10  | Ambon            | 1.57                |
| 11  | Tual             | 1.53                |

Estimated parameters \( \theta \) show the relative risk of leprosy in Maluku Province. The highest incidence rate of leprosy occurred in Ambon City at 1.57 and the lowest incidence rate of leprosy occurred in the Aru Islands at 1.47. The average incidence rate of leprosy in Maluku Province is 1.52. Then disease mapping based on relative risk of leprosy in Maluku shown in figure 2.

**Figure 2.** Disease Mapping of Leprosy’s Relative Risk in Maluku Province 2019
4. Conclusion

The relative risk that is formed from the Hierarchical Bayes method shows that the incidence rate of leprosy in Maluku Province is evenly distributed in each Regency/City, with the highest value in Ambon City and the lowest value in Aru Islands. Therefore, to handling leprosy’s phenomenon in Maluku, the government could pay attention to area with highest relative risk first in order to reduce the spread of the disease. This is the joint responsibility of both the government and the community so that this disease is immediately eliminated, such as in several areas of western Indonesia.

References

[1] Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, et. al 2000 Disease Mapping Models: An Empirical Evaluation. Stat Med 19: 2217-2241
[2] Clayton, J. and Kardor J 1987 Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping Biometrics Vol. 43 No.3, 671-681. International Biometric Society
[3] Jaya, I.G.N.M., Folmer, H 2020 Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung Indonesia J Geogr Syst 22 105–142 https://doi.org/10.1007/s10109-019-00311-4
[4] Jaya, I.G.N.M, Ruchjana, B.N., Andriyana, Y., and Agata, R 2019 Clustering with Spatial Constraints: The Case of Diarrhea in Bandung City, Indonesia. J. Phys.: Conf. Ser. 1397 012068
[5] Yupiana,Y., Basri, C., Ernawati, Sihombing, S., Zenal, F., et. al 2019 Risk Mapping of Risk Factors Associated with Human Rabies Cases in Bali Province, Indonesia Proceedings of the Conference of the International Society for Economics and Social Sciences of Animal Health - South East Asia 2019 (ISESSAH-SEA 2019) https://doi.org/10.2991/isessah-19.2019.36
[6] Badan Pusat Statistik 2020 Provinsi Maluku Dalam Angka 2020 Ambon Badan Pusat Statistik
[7] Maiti, T 1997 Hierarchical Bayes Estimation of Mortality Rates for Diseased Mapping Journal of Statistical Planning and Inference 69 (1998) 339-348 Elsevier
[8] Walsh, B 2004 Markov Chain Monte Carlo and Gibbs Sampling Ilceture Notes for FEB 581 version 26 April 2004
[9] Ntzoufras I 2009 Bayesian Modeling Using WinBUGS Hoboken New Jersey: A John Wiley & Sons Inc Publication
[10] Geman, S and Geman, D 1984 Stochastic Relaxation Gibbs Distribution and the Bayesian Restoration of Images IEE Transaction on Pattern Analysis and Machine Intelligence

Acknowledgments
The authors thank reviewer team for their input on improving this journal, colleagues for their support and BPS who have provided leprosy data especially in Maluku Province.