Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2020, 2020:97
https://doi.org/10.28919/cmbn/5134
ISSN: 2052-2541

MULTILEVEL MODEL OF DENGUE DISEASE TRANSMISSION IN WEST JAVA PROVINCE, INDONESIA BY MEANS INLA

I.G.N.M. JAYA[1,2,*], B.N. RUCHJANA[3], A. SETIAWAN ABDULLAH[4], N. SUNENGSIH[5]

[1]Department of Statistics, Padjadjaran University, Indonesia

[2]Faculty Spatial Science, Groningen University, the Netherlands

[3]Department of Mathematics, Padjadjaran University, Indonesia

[4]Department of Computer Science, Padjadjaran University, Indonesia

[5]Department of Statistics, Padjadjaran University, Indonesia

Abstract: Dengue is an infectious disease that severe health problems even death. The number of infections has been increased significantly worldwide. Dengue disease is transmitted through *Aedes Aegypti* vectors. Identifying high-risk areas is vital to control disease transmission. Population density is a crucial risk factor that can accelerate dengue transmission. The cities in general have higher population density than districts. Therefore, the modeling dengue risk is critical to take into account the district and city levels information. It can be done by developing multilevel models as a representation of the hierarchical models. The models were estimated by means of integrated nested Laplace approximation (INLA). We found there is a high different regression coefficients between district and city levels. Population density has a high positive impact on the city level.

Keyword: Bayesian; clustering; dengue disease; elevation; West Java; INLA; multilevel.

2010 AMS Subject Classification: 91D20.

---

[*]Corresponding author
E-mail address: mindra@unpad.ac.id

Received October 22, 2020

1. INTRODUCTION

Dengue is an infectious disease that severe health problems even death [1-3]. The number of infections has been increased significantly worldwide [4, 5]. The South-East Asia and Western Pacific regions are the most seriously affected. Indonesia is one of the countries in South-East Asia with high dengue case every year [4]. The female *Aedes Aegypti* mosquitoes is a vector of dengue virus. The dengue virus is transmitted through the bites of infected female *Aedes Aegypti. Aedes Aegypti* feeds all days particularly are early in the morning and in the evening before sunset [6].

Controlling dengue disease transmission is essential to minimize the health and socio-economic impact of this disease. However, the schema of dengue disease transmission was very complicated [7]. It involves many factors, such as socio-economic, environmental, and weather factors. Those variables influence mosquitos' life cycle, the transmission of the dengue virus, and human immunity [7, 8].

Population density, health behaviours and weather variables are the important risk factors for dengue diseases [3, 9, 10]. Population density has provided important insights into the epidemiology of dengue disease [10]. Areas with high population density facilitate disease transmission [2]. Weather variables, especially temperature and rainfall, significantly influence the breeding of the *Aedes aegypti* mosquitoes [2]. In this study, we focus on evaluating the effect of population density. We realized that considering a lot of number covariates in the model could be better in understanding the disease dengue disease transmission. However it could be lead to multicollinearity problem and produce unreliable estimate [11].

The complexity of the interactions between environmental factors, spatial dependency, and heterogeneity needs more attention to disease modeling to obtain more reliable risk prediction [1, 12, 13]. The disease incidence and environmental factors may represent the third dimension of epidemiological experiment: person, time and space [14]. The last dimension is widely measured by the administrative boundary. A small regions may be classified by the district and city levels. Although well identified, quantifying the relative influence of each of these level in dengue transmission would raise serious methodological difficulties [14]. The origin of variation between disease incidence could be explained by a complex combination of characterizing small areas (the individual level) or district and city levels (the group level). The policy a higher level will influence health outcomes in a given small area. So, relations between individual level and group level determinants are of particular interest, especially for investigating the reasons for variation between small areas: are the environmental factors, population density, healthy behaviours influenced by the policy at a higher level. It could be the small area in one group have similar characteristic and different from the other groups. For example, the small areas grouped at the city level may have higher population density and healthy behaviour compared to the small areas at the district level, with the spatial variation between small areas in one group that may still exist. Combining small areas and collective exposures necessitate analysis of several collective situations simultaneously, and in each one, several small areas. Gathered small areas in the same level (district or city levels), small areas in same level are more similar to each other than small areas from different levels. They are organized into groups of dependent data (also called clusters); small areas are said to be nested within geographical areas [14]. A statistical approach based on multilevel modeling has been developed to handle the analysis incorporating different data levels. A variety of names have been used synonymously for ‘multilevel model’: ‘hierarchical model’, ‘random effect model’, ‘variance component model’, or ‘mixed model’. The models produce varying regression coefficients. Geographically weighted regression (GWR) is widely used to modeling covariates that allow the regression coefficients vary over space [15-17]. In this study, we develop a multilevel model by means Bayesian hierarchical model based on Poisson log-linear model. The model is developed by considering spatial dependency, heterogeneity, and population density across small areas and levels.

We apply the model for identifying high-risk areas in West Java, Indonesia. West Java is the top rank in dengue disease incidence for every year. There are 18 districts and 9 cities in West Java with much different in population density.

The paper is structured as follows section 2 describe the method, section 3 figure out the application and discussion, and section 4 focus on the conclusion and discuss about conclusion.

2. MATERIA AND METHOD

Material

The data were obtained from West Java health department [18]. We used three variables are number of dengue cases, population at risk and population density and used the district and city levels as a group. The data are presented in Table 1.

Table 1. Statistics of number of cases, population, population density, and district-city group

id	Distric/Cityt	Cases	Population	Population Density	Group
1	Bogor	741	5,965,410	2.201	District
2	Sukabumi	233	2,466,272	0.595	District
3	Cianjur	113	2,263,072	0.589	District
4	Bandung	1,774	3,775,279	2.135	District
5	Garut	331	2,622,425	0.853	District
6	Tasikmalaya	196	1,754,128	0.688	District
7	Ciamis	29	1,195,176	0.845	District
8	Kuningan	352	1,080,804	0.973	District
9	Cirebon	215	2,192,903	2.227	District
10	Majalengka	108	1,205,034	1.001	District
11	Sumedang	408	1,152,400	0.759	District
12	Indramayu	911	1,728,469	0.847	District
13	Subang	143	1,595,825	0.843	District
14	Purwakarta	263	962,893	1.166	District
15	Karawang	60	2,353,915	1.425	District
16	Bekasi	181	3,763,886	3.073	District
17	Bandung Barat	419	1,699,896	1.302	District
18	Pangandaran	75	399,284	0.395	District
19	Bogor	727	1,112,081	9.385	City
20	Sukabumi	238	328,680	6.812	City
21	Bandung	2,826	2,507,888	14.957	City
22	Cirebon	24	319,312	8.547	City
23	Bekasi	626	3,003,923	14.539	City
24	Depok	891	2,406,826	12.017	City
25	Cimahi	288	614,304	15.643	City
26	Tasikmalaya	223	663,517	3.866	City
27	Banjar	97	183,110	1.613	City

Table 2. Total cases and prevalence rate per group

District	Total cases	Prevalence Rate (× 100,000)
District	6,552	17.162
City	5,940	53.323

We considered district and city levels as different groups because we found that the prevalence rate of dengue cases much different between both levels. There were 19 districts and 8 cities in West Java, Indonesia. The prevalence rate in district level is 17.162/100.000 population at risk habitants and at city level is 53.323/100.000 population at risk.

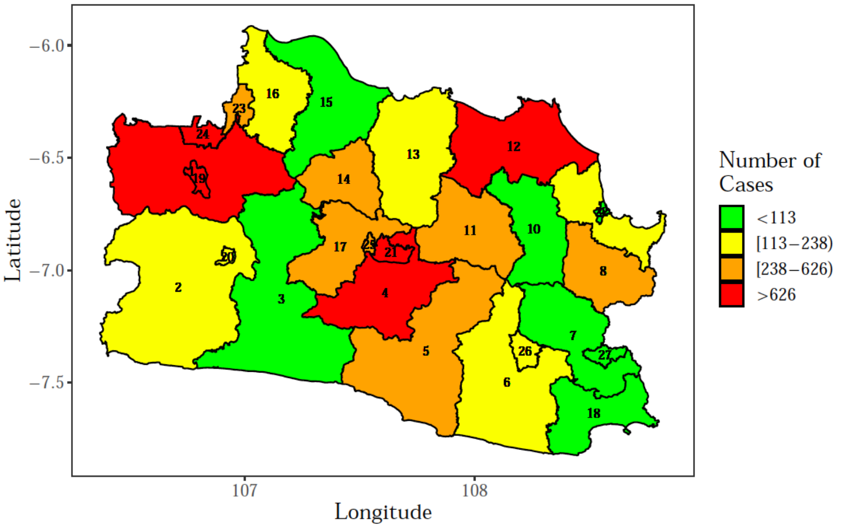


Figure 1. Spatial distribution of number of dengue cases for 18 districts and 9 cities in West Java

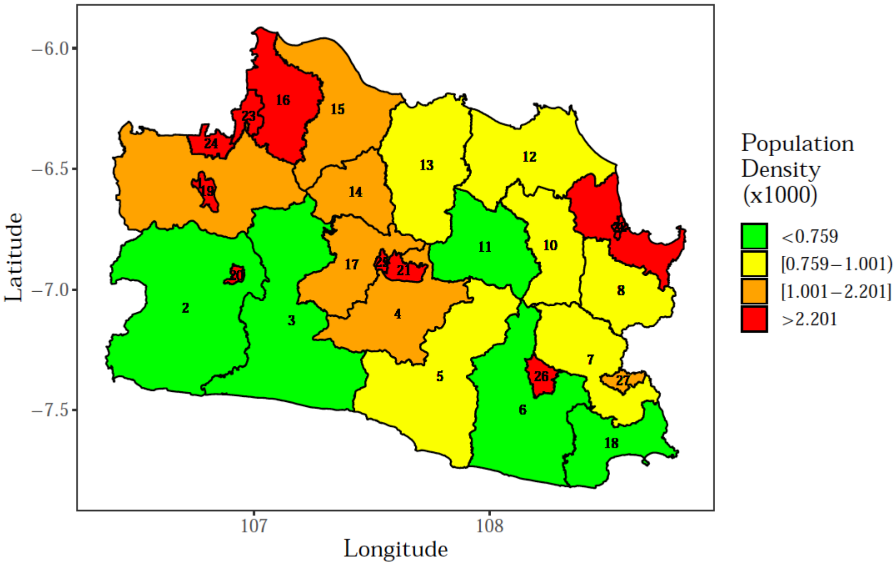


Figure 2. Population density for 18 districts and 9 cities in West Java

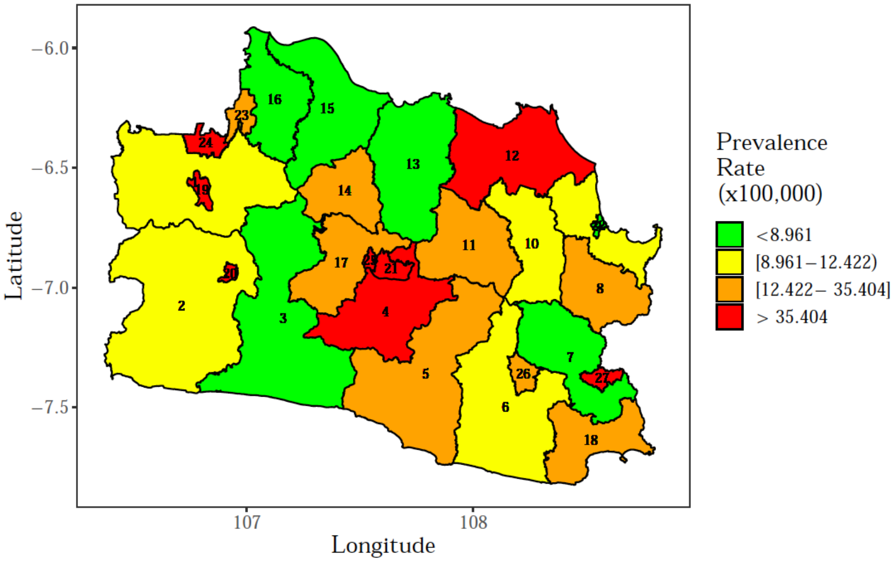


Figure 3. Spatial distribution of the prevalence rate of dengue disease for 18 districts and 9 cities in West Java

Figure 1 shows the spatial distribution of dengue fever cases in West Java. It can be seen that the red areas are dominated by city level. Figure 2 presents the population density. The areas with high population density are located at the city level. The prevalence distribution is presented in Figure 3. It supports that the prevalence rate at the city level is much higher than the district level. This condition validates the use of districts and cities as deep levels in multilevel modeling.

Standardized incidence ratio (SIR)

SIR is known as crude or unreliable risk estimate [19]. It is because SIR tends to be high for a small area and low for larger areas [1, 12, 13, 19]. The SMR is expressed as the ratio between number of cases (y_i) on the expected count (E_i) as follows [20]:

$$SIR_i = \frac{y_i}{E_i} \tag{1}$$

The expected number of cases E_i in each area is calculated as $E_i = N_i(\sum_{i=1}^{n} y_i / \sum_{i=1}^{n} N_i)$ where n denotes the number of areas and N_i is population at risk in area ith. The areas with a relative risk higher than one are indicated as a potentially high-risk area.

Poisson Loglinear model

Relative risk based models have been proposed to overcome the unreliable estimates of SIR. Using the model approach, we can introduce a random-effects model that smooth the SIR estimate to provide a more reliable risk estimate. In this study we use the Poisson model is given by [1, 2, 19]:

$$y_i|E_i,\theta_i \sim Poisson(E_i\theta_i);\ i=1,\dots,n \tag{2}$$

The Poisson regression model fits for count data. Log-linear models are widely used to explain the effects of the *K*-risk factors on the mean function. The model for *n* spatial units is given by:

$$\log \mathbb{E}(y_i) = \text{offset}(\log E_i) + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i;\ i=1,\dots,n \tag{3}$$

where $\text{offset}(.)$ that represents the regression coefficient of $\log E_i$ that is fixed to one, $\boldsymbol{x}_i = (1,\boldsymbol{x}_{i1},\dots,\boldsymbol{x}_{iK})'$is a $(K+1)\times 1$ is vector of area-level risk factors, $\boldsymbol{\beta}=(\beta_0,\beta_1,\dots,\beta_K)'$ is $(K+1)\times 1$ is vector of regression parameters, and $\epsilon_i=\omega_i+v_i$ denotes the random effect components used to accommodate the spatial dependency (ω_i) and heterogeneity (v_i). Here we use only single covariates so that $K=1$. In this study we consider population density as the covariate and intrinsic conditional autoregressive (iCAR) and Gaussian (*i.i.d*) priors to account spatial dependency (ω) and heterogeneity (v), respectively. iCAR model is defined follows Besag York and Molie model [21]:

$$\omega_i|\boldsymbol{\omega}_{-i},\tau_\omega,\mathbf{W} \sim \mathcal{N}\left(\frac{\sum_{j=1}^{n} w_{ij}\omega_j}{\sum_{i=1}^{n} w_{ij}}, \frac{\sigma_\omega^2}{\sum_{i=1}^{n} w_{ij}}\right) \tag{4}$$

where $\mathbf{W}=\left(w_{ij}\right)$ is the 'neighbourhood' matrix with $w_{ij}=1$ if i and j are neighbouring and $w_{ij}=0$ otherwise. The spatial heterogeneity (v_i) is widely assigned by an Gaussian identically independent prior distribution, that is:

$$v_i|\sigma_v^2 \sim \mathcal{N}\left(0,\frac{1}{\sigma_v^2}\right) \tag{5}$$

where σ_v^2 is the variance parameter of v_i.

The area is defined as a hot-spot (high-risk region) if the estimated exceedance probability is $\widehat{\Pr}(\theta_{it}>1|\mathbf{y})>0.95$. In a similar vein, district is defined as a cold-spot (low-risk region) if the estimated exceedance probability $\widehat{\Pr}(\theta_{it}>0.95|\mathbf{y})<0.05$.

Multilevel model

To evaluate the effect of the population density to the relative risk and get the reliable estimation of the relative risk we proposed five different model. In this model we consider the different district and city level particularly for M2-M5. The models are presented below:

M1: $$\eta_{ij} = \alpha + \beta x_{ij}; i = 1,..,27; j = 1,2 \tag{6}$$

M2: $$\eta_{ij} = \alpha_j + \beta x_{ij}; i = 1,..,27; j = 1,2 \tag{7}$$

M3: $$\eta_{ij} = \alpha_j + \beta_j x_{ij}; i = 1,..,27; j = 1,2 \tag{8}$$

M4: $$\eta_{ij} = \alpha_j + \beta_j x_{ij} + u_i \sim iid(0, \sigma_u^2); i = 1,..,27; j = 1,2 \tag{9}$$

M5: $$\eta_{ij} = \alpha_j + \beta_j x_{ij} + u_i \sim Besag; i = 1,..,27; j = 1,2 \tag{10}$$

where sub index i denote observation at the i-th area and j indicates the group levels. Model (M1) denotes the non-hierarchical model which does not take into account the fact the for each areas can be classified in two groups district and city levels. This model is known as pooled model. Models (M2-M5) presents the hierarchical. M2 assumes the intercept varies by district and city levels. M3 takes into account the interaction between population density (x) and the group levels. Model (M4-M5) consider the spatial heterogeneity and spatial dependency, respectively. The structured of hierarchical models can be seen in Figure 5 below:

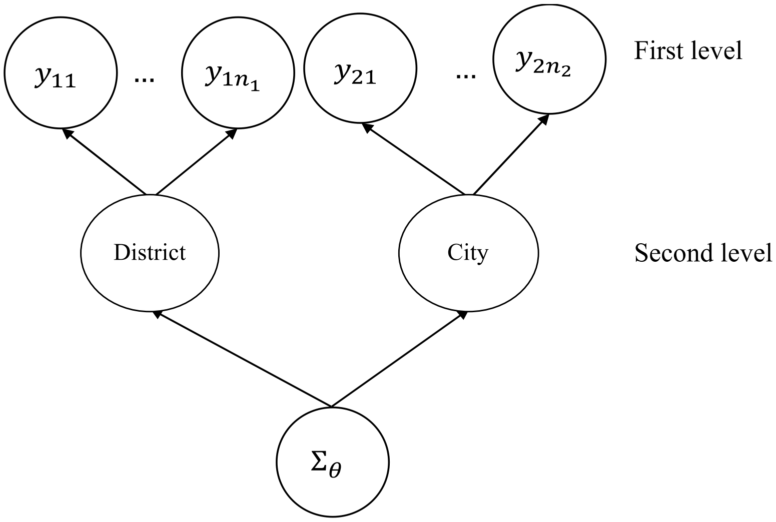


Figure 5 Multilevel structure

Models (1) to (5) can be estimated by mean Bayesian method via estimated by means of Integrated Nested Laplace Approximation (INLA) (see [1, 2] for detail). The best model is selected using deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC) [22]. For parameter interest we use Gaussian distribution and for the hyperparameter we use inverse gamma distribution. The models were estimated using R-software (the R-code available upon by request).

3. RESULT AND DISCUSSION

In the analysis we consider hierarchical level organization which included fixed and random effects, and cross-level interaction. Accurate estimation of the parameter estimates is the advantage of multilevel modeling [14].

Data exploration

Data exploration is the important part in building regression modeling. Using data exploration we can find the hidden structured in the data [23]. First step, we evaluate the relationship between log(SIR) and population density. We use log(SIR) as representation of the log-linear model. The relationship is presented in Figure 6.

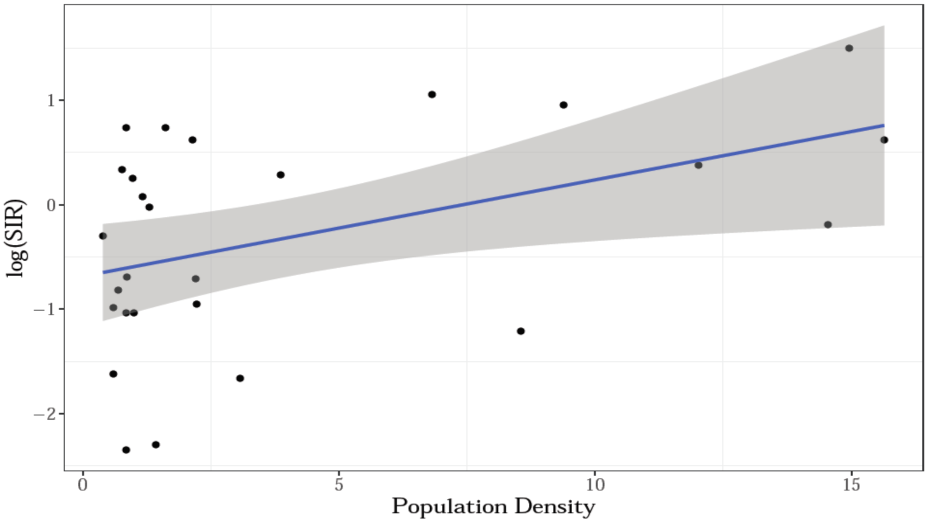


Figure 6. The relationship between population density and relative risk

Figure 6 shows the strong linear relationship between population density and log(SIR). However, it can be a misleading conclusion if we ignored the district and city levels. Figure 7 shows the relationship between population density and log(SIR) by considering the district and city level.

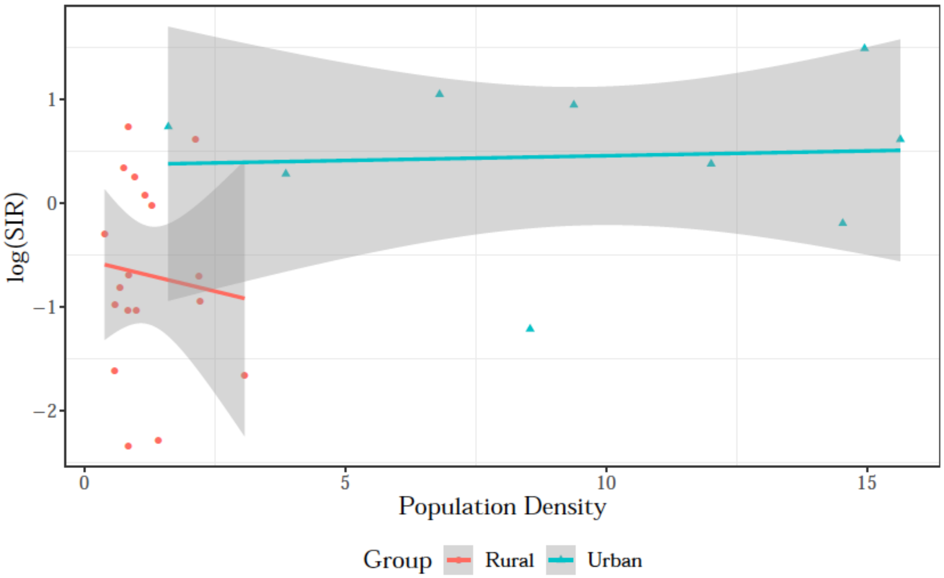


Figure 7. The relationship between population density and relative risk by group

Figure 7 shows a contradiction relationship between population density and log(SIR) for different levels. There is a positive relationship between population density and log(SIR) for the city-level and a negative association for the district level.

Models comparison

To select the best model for explaining the effect of the population density on West Java's relative risk, we present the model comparison based on DIC, WAIC, Pearson correlation between the predicted relative risk and observed relative risk in Table 3.

Table 3. Model comparison

Model	Model	DIC	WAIC	Pearson Correlation
M1	$\eta_{ij} = \alpha + \beta x_{ij}; i = 1,..,27; j = 1,2$	7163.093	8392.916	0.527
M2	$\eta_{ij} = \alpha_j + \beta x_{ij}; i = 1,..,27; j = 1,2$	6960.615	8433.937	0.594
M3	$\eta_{ij} = \alpha_j + \beta_j x_{ij}; i = 1,..,27; j = 1,2$	6932.656	8392.140	0.592
M4	$\eta_{ij} = \alpha_j + \beta_j x_{ij} + u_i \sim iid(0,\sigma_u^2); i = 1,..,27; j = 1,2$	253.290	245.723	1.000
M5	$\eta_{ij} = \alpha_j + \beta_j x_{ij} + u_i \sim Besag; i = 1,..,27; j = 1,2$	253.563	246.356	1.000

Table 4 shows the comparison of fit and complexity of the models based on the DIC and WAIC sizes. The results of the analysis found that model (M1) has the highest DIC and WAIC model (M4) is a model with the suitability and complexity of the model that best fits the data indicated by the lowest DIC and WAIC and the highest correlation value between prediction and observation (Pearson). In other words, the best model is model (M4). This result proves that the district and city groups in the data must be accommodated in the modeling of population density against relative risk.

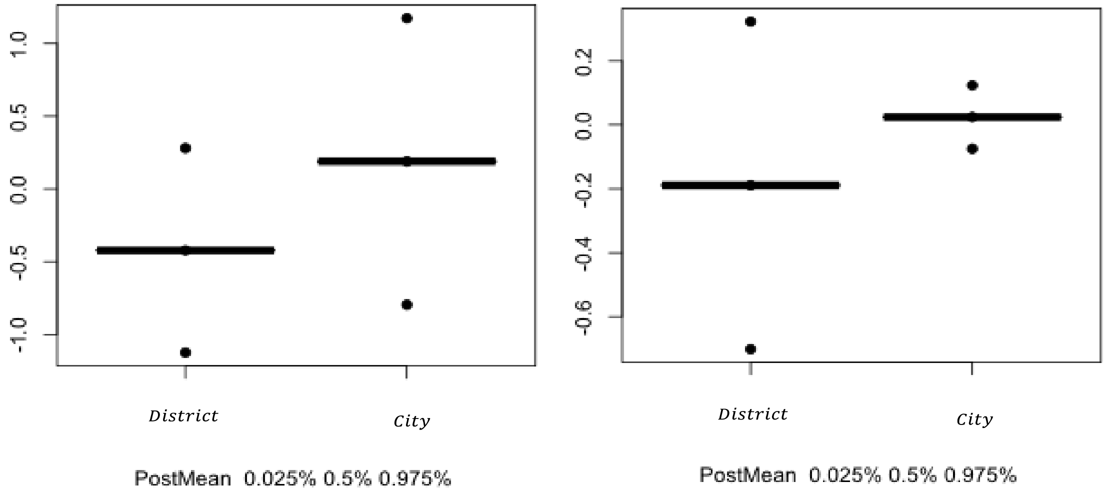
Models Interpretation

The posterior means and posterior standard deviation are presented in Tables 4-5.

Table 4. Posterior means of parameters interest for district and city levels

Levels	Intercept	Slope
District	-0.422	-0.189
City	0.189	0.024

Table 4 shows the different intercept and slope regression for the city district and city level. The district-level has a negative intercept and slopes that support the relative risk at the district level lower than the city level. The coefficients are plotted in Figure 8.



a. Varying Intercepts

b. Varying Slopes

Figure 8. Intercept and slope of district and city areas

Figure 8 shows the different intercept and slop coefficients between district and city levels.

Table 5. Posterior standard deviation (SD) of intercept, slope, and heterogeneity

Posterior SD	Mean	Relative Contribution
Posterior standard deviation group for intercept	1.00	35.17
Posterior standard deviation group for effect of population density	0.92	32.12
Posterior standard deviation group for heterogeneity	0.93	32.71

Table 5 shows the relative posterior standard deviation of each parameter model and its relative contribution to dengue risk variation in West Java. District and city groups give the most considerable contribution (35.17%). It indicates that district-city have different characteristics that cause the variability of dengue risk. Descriptive analysis shows that the risk in city areas higher than in district areas. It can be influenced by the population density and the other characteristics covered by heterogeneity random-effects components. Population density that interact with group give 32.12% contribution and the other components area 32.71% contribution.

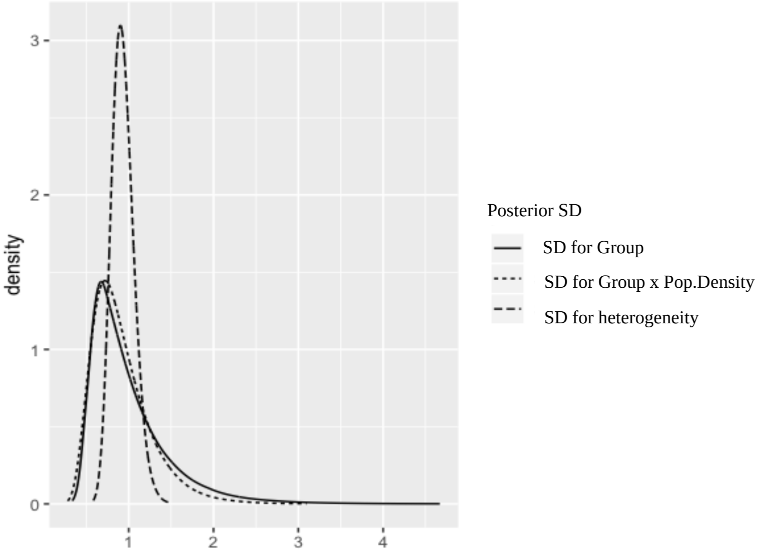


Figure 9. Posterior standard deviation of Group, Group × Population density, and heterogeneity

Based on the model (M4) we estimate the relative risk of each area and also calculate the exceedance probability value to show areas where the relative risk is significantly different from one.

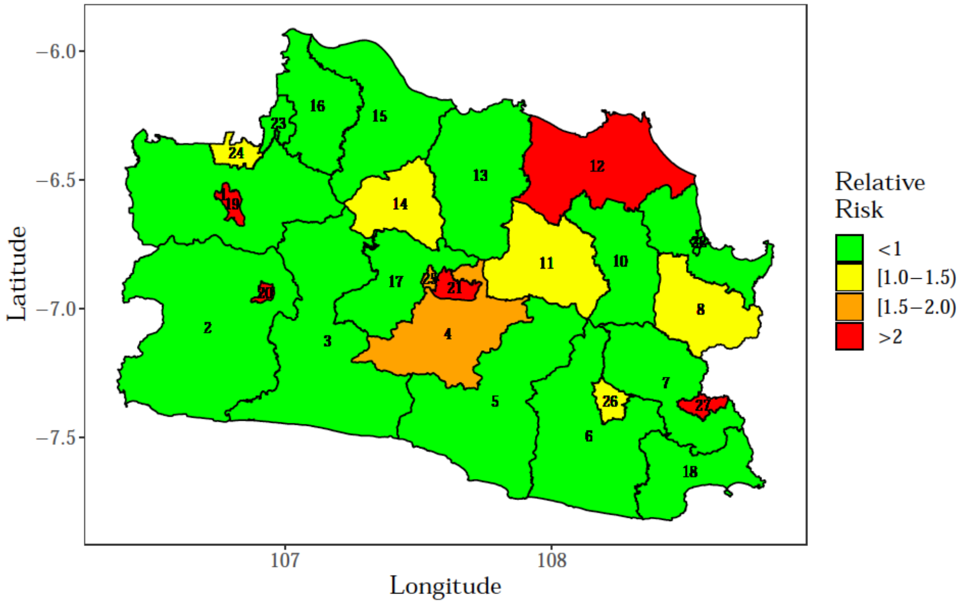


Figure 10. Relative risk for 27 districts and cities in West Java

Figure 10 shows the relative risk in 27 urban districts in West Java, Indonesia, in 2019. The city level tends to have a higher relative risk than the district level. To evaluate the significant cold and hot-spot risks, we use the exceedance probability approach. The area is categorized as a hot-spot if it has an exceedance probability higher than 0.950. Conversely, it is classified as a cold-spot if the exceedance probability is lower than 0.05 [2].

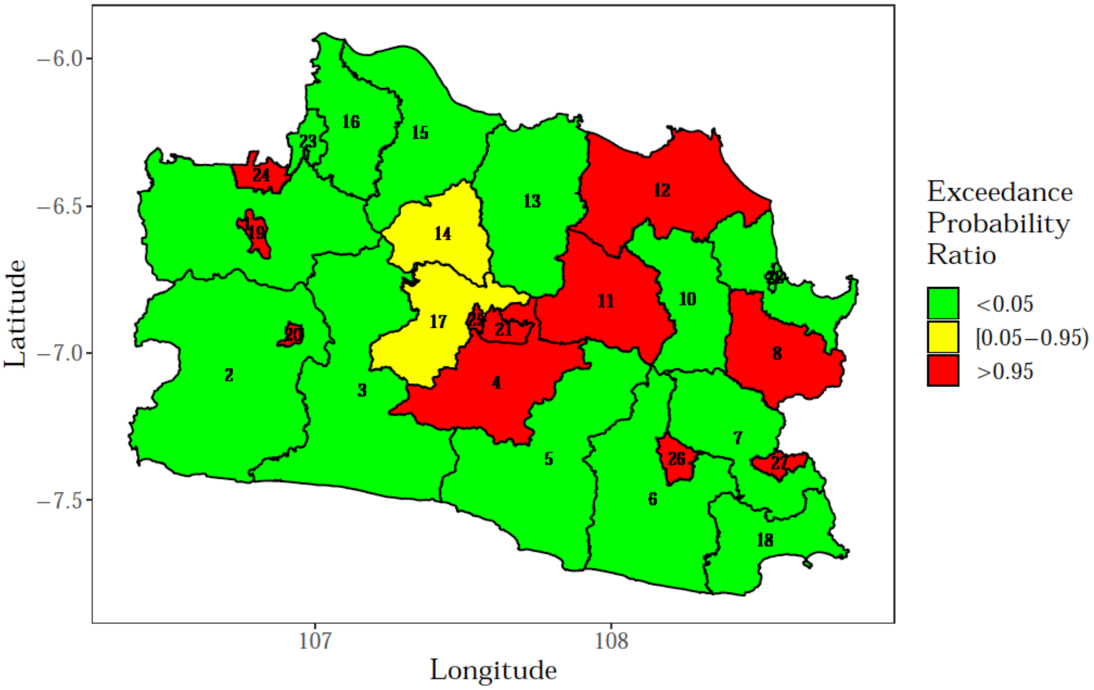


Figure 11. Exceedance Probability Ratio for 27 districts and cities in West Java

Figure 11 shows all city levels categorized as hotspots because of greatly influenced by population density and the high mobility of urban people. This condition accelerates the transmission of dengue fever from one person to another.

4. CONCLUSION

Identifying potential groups in modeling the effect of independent variables on response is very important. Ignoring potential groups may lead to misleading conclusions. Each group may have a different model and must be accommodated to obtain accurate and reliable results. Groups in the model can be treated as a higher-level concept in multilevel modeling or hierarchical modeling. The Bayesian method facilitates hierarchical modeling through the random effect model. Our analysis of modeling on the effect of population density on the relative risk of dengue fever in West Java found that the district and city clusters have different disease models. For city groups, population density is known to have a positive effect on increasing relative risk. On the other hand, in district areas, population density has a negative effect or in this case, it does not affect. This is because the district group relatively has a low population density. We also evaluate if we ignore these city and district groups. We found that at all levels, the effects of population density were positive. If we compare the quality of the model, we find that the model has a better quality of goodness. The model shows that group variation affects the impact of population density on the relative risk.

ACKNOWLEDGMENT

We thanks to Rector Universitas Padjadjaran. Financial support was received from ALG Unpad contract: 1427/UN6.3.1/LT/2020

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] I.G.N.M. Jaya, H. Folmer, Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia, J. Geogr Syst. 22 (2020), 105–142.

[2] I.G.N.M. Jaya, H. Folmer, Identifying Spatiotemporal Clusters by Means of Agglomerative Hierarchical Clustering and Bayesian Regression Analysis with Spatiotemporally Varying Coefficients: Methodology and Application to Dengue Disease in Bandung, Indonesia, Geogr. Anal. (2020) gean.12264. https://doi.org/10.1111/gean.12264.

[3] I.G.N.M. Jaya, Y. Andriyana, B. Tantular, Elevation as a proxy for dengue disease transmission in Indonesia, Commun. Math. Biol. Neurosci., 2020 (2020), Article ID 89.

[4] WHO, Dengue and severe dengue, WHO, 23 June 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. [Accessed 26 September 2020].

[5] O.J. Brady, P.W. Gething, S. Bhatt, et al. Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus, PLoS Negl. Trop. Dis. 6 (2012), e1760.

[6] M. Trips, G. McClelland, J. Gillet, C. Teesdale, T. Rao, Diel periodicity in the landing of Aedes aegypti on man, Bull. World Health Organ. 48(5) (1973), 623–629.

[7] K.L. Ebi, J. Nealon, Dengue in a changing climate, Environ. Res. 15(1) (2016), 115-123.

[8] A.G. Watts, J. Miniota, H.A. Joseph, et al. Elevation as a proxy for mosquito-borne Zika virus transmission in the Americas, PLoS ONE. 12 (2017), e0178211.

[9] O.J. Brady, N. Golding, D.M. Pigott, et al. Global temperature constraints on Aedes aegypti and Ae. albopictus persistence and competence for dengue virus transmission, Parasit Vectors. 7 (2014), 338.

[10] W.-P. Schmidt, M. Suzuki, V. Dinh Thiem, R.G. White, A. Tsuzuki, L.-M. Yoshida, H. Yanai, U. Haque, L. Huu Tho, D.D. Anh, K. Ariyoshi, Population Density, Water Supply, and the Risk of Dengue Fever in Vietnam: Cohort Study and Spatial Analysis, PLoS Med. 8 (2011), e1001082.

[11] I.G.N.M. Jaya, B. Ruchjana, A. Abdullah, Comparison Of Different Bayesian And Machine Learning Methods In Handling Multicollinearity Problem: A Monte Carlo Simulation Study, ARPN J. Eng. Appl. Sci. 15(18) (2020), 1998-2011.

[12] P. Yin, L. Mu, M. Madden, J. E. Vena, Hierarchical Bayesian modeling of spatio-temporal patterns of lung cancer incidence risk in Georgia, USA: 2000–2007, J. Geogr. Syst. 16(2014), 387–407.

[13] A. Lawson, Bayesian Disease Mapping, Hierarchical Modelling in Spatial Epidemiology, 2nd ed., Boca Raton: CRC Press Taylor and Francis Group, 2018.

[14] F. Mauny, J. Viel, P. Handschumacher, B. Sellin, Multilevel modelling and malaria: a new method for an old disease, Int. J. Epidemiol. 33(6) (2004), 1337–1344.

[15] R. Fitriani, I. Jaya, Spatial modeling of confirmed covid-19 pandemic in East Java province by geographically weighted negative binomial regression, Commun. Math. Biol. Neurosci. 2020 (2020), 58.

[16] E. Marhamah, I. Jaya, Modeling Positive Covid-19 Cases In Bandung City By Means Geographically Weighted Regression, Commun. Math. Biol. Neurosci. 2020 (2020), 77.

[17] I. Jaya, B. N. Ruchjana, A. S. Abdulah, E. Hermawan, Bayesian Spatial Modeling and Mapping of Dengue Fever: A Case Study of Dengue Fever in the City of Bandung, Indonesia, Int. J. Appl. Math. Stat. 54(3) (2016), 94-103.

[18] W. Java, Health Profile 2019, West Java: Health Office of West Java, 2020.

[19] I.G.N.M. Jaya, H. Folmer, B.N. Ruchjana, F. Kristiani, Y. Andriyana, Modeling of Infectious Diseases: A Core Research Topic for the Next Hundred Years, in: R. Jackson, P. Schaeffer (Eds.), Regional Research Frontiers - Vol. 2, Springer International Publishing, Cham, 2017: pp. 239–255.

[20] D. Clayton and J. Kaldor, Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping, Biometrics, 43(3) (1987), 671-681.

[21] J. Besag, J. York, A. Mollié, Bayesian image restoration, with two applications in spatial statistics, Ann. Inst. Stat. Math. 43 (1991), 1–20.

[22] S. Watanabe, Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, J. Mach. Learn. Res. 11 (2010), 3571-3594.

[23] M.C.F. de Oliveira, H. Levkowitz, From visual data exploration to visual data mining: A survey, IEEE Trans. Visual. Comput. Graphics. 9 (2003), 378–394.