

# Bayesian hierarchical model for mapping positive patient Covid-19 in Surabaya, Indonesia

Cite as: AIP Conference Proceedings **2329**, 060003 (2021); <https://doi.org/10.1063/5.0042113>  
Published Online: 26 February 2021

Rudianto Artiono



View Online



Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

[Number of flood disaster estimation in Indonesia using local linear and geographically weighted regression approach](#)

AIP Conference Proceedings **2329**, 060006 (2021); <https://doi.org/10.1063/5.0042118>

[The Fourier series estimator to predict the number of dengue and malaria sufferers in Indonesia](#)

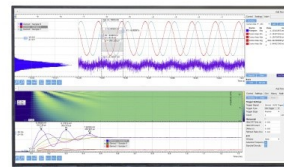
AIP Conference Proceedings **2329**, 060002 (2021); <https://doi.org/10.1063/5.0042115>

[Modeling bivariate Poisson regression for maternal and infant mortality in Central Java](#)

AIP Conference Proceedings **2329**, 060007 (2021); <https://doi.org/10.1063/5.0042142>

Challenge us.

What are your needs for periodic signal detection?



Zurich Instruments



# Bayesian Hierarchical Model for Mapping Positive Patient Covid-19 in Surabaya, Indonesia

Rudianto Artiono<sup>a)</sup>

*Department of Mathematics, Universitas Negeri Surabaya, Indonesia*

<sup>a)</sup>rudiantoartiono@unesa.ac.id

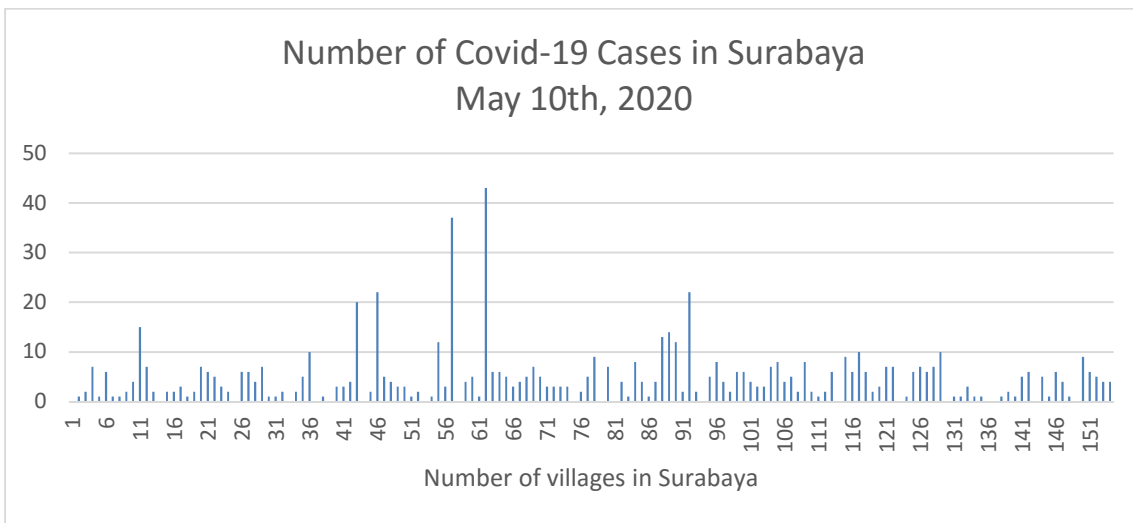
**Abstract.** Covid-19 is a disease caused by the coronavirus and has infected the world population not only in developing countries like Indonesia but also in developed countries like the United States of America. The spread of the disease originating from Wuhan in China is evenly distributed throughout the world. Surabaya, as one of the major cities in Indonesia was also affected by Covid-19. The aim of this study was to map the relative risk of the spread of positive Covid-19 patients in Surabaya using the Bayesian Hierarchical Model with spatial analysis to deal with regional dependencies. The object used was spatial object as many as 154 villages in Surabaya. Meanwhile, a number of positive patients Covid-19 as of May 10<sup>th</sup>, 2020 was as many as 708 patients. The method used to estimate the relative risk was the Bayesian method with the Integrated Nested Laplace Approximation (INLA) approach. It was used after several studies have shown that the INLA approach is more accurate in providing estimated values compared to the Maximum-Likelihood estimation. The mapping results showed that there is a spatial dependence on the spread of Covid-19 disease in Surabaya.

## INTRODUCTION

Coronavirus disease, known as Covid-19, is an importance-worldwide infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2). It is a mutation from the previous coronavirus which also infects humans, such as HCov-229E and HCov-OC43. Some researchers have found that the novel of coronavirus has hereditary similarities with the coronavirus found in bats [1-5]. It spreads in humans through droplets from someone who has been infected. Furthermore, over a period of 14 days, a virus will replicate in the respiratory tract up to the part of the human lung which will result in respiratory failure and death [6-8]. Some researchers also found that the virus is very dangerous when infecting people over the age of 50 years, even though in reality there are also children who are exposed to Covid-19 [9-10]. Someone with a high level of immunity can get rid of viruses which are within the body, whereas people with low levels of immunity will involvement some symptoms that are manifestations of this disease. In several studies, there are also asymptomatic Covid-19 patients which have been found reactive to the tests. Even though the person does not show any symptoms of Covid-19, the person has the virus already in his body [11-12]. It is the reason why test ought to be carried out to distinguish whether an individual has been uncovered to the coronavirus or not. This test not only can be used to determine the presence of viruses in the human body, but also can be done to determine the distribution of diseases in one country.

Since November 2019, the distribution of Covid-19 patient has spread almost throughout the world. The disease, which was first discovered in Wuhan China, has infected approximately 213 countries with a total number of patients up to the end of the third week of May 2020 as many as 4,338,658 people. The number of new cases added in this period was 90,269 people and the total number of cumulative deaths was 297,119 people [13]. The average number of these patients came from developed countries such as the United States of America, Spain, Russia, the United Kingdom, and Italy. In the United State of America, with a total population in 2020 as many as 331,002,651 people, there were Covid-19 patients as many as 1,361,522 people. Meanwhile, developing countries like Indonesia were also affected by the coronavirus.

Information obtained from <https://covid19.go.id/>, the official website of the Indonesian government for handling Covid-19, expressed that the number of positive patients in Indonesia as of May 16, 2020 was 16,496 individuals with a total number of deaths was around 1,076 people. This number expanded each single day with an average growth of new cases was 300-400 individuals per day. In spite of the fact that, the most elevated number number of Covid-19 patients in Indonesia came form the capital city of Jakarta, the infection has spread equitably to several districts/cities in Indonesia. Of the entire number of districts in Indonesia as numerous as 514, there were 383 districts that had been infected by corona virus. Surabaya as one of the biggest cities in Indonesia also had a number of patients. Data compiled from <https://lawancovid-19.surabaya.go.id/>, the official website of the surabaya government for dealing Covid-19, as of May 16, 2020 there were 945 positive patients in Surabaya which spread into 154 villages. In Fig. 1, it can be seen that the spread of Covid-19 illness that occurred in Surabaya was equitably dispersed in nearly all existing villages with the most noteworthy number of patients up to May 10, 2020 was 43 individuals in Kemayoran village, North Surabaya.



**FIGURE 1.** Data of Number of Positive Patient Covid-19 in Surabaya from March 2, 2020 to May 10, 2020

Considering the spread of Covid-19 that occurs rapidly and equitably from nation to another or from one region to another, from March 11, 2020 the World Health Organization announced Covid-19 as a global pandemic [14]. Spatial analysis method can be used to examine spatial distribution with variations between regions. It moreover can be used to construct a map of Covid-19 cases that accommodate aspects of regional dependencies by identifying cases through their relative risks.

However, in the study of disease transmission, the foremost important thing that we ought to concern is about how to inference the disease spread from the past information and it is not simple to urge total information in public health area. In some cases, it is still conceivable to develop model which can be utilized as a begining point to inference. Specifically, models endeavor to portray the instrument by which the observed data are generated. Inference at that point continues by endeavoring to estimate the parameter. One strategy that can be used to estimate the parameter is Bayesian. It comes from the expansion of the likelihood paradigm. In Bayesian, it is not only population which are random variable but parameters are also random variable. Therefore, the parameter also have distribution and it is known as prior distribution. By using the likelihood model (data) and prior distribution, the posterior distribution can be obtained. As the process update continuously then posterior distribution also could be prior distribution for the next algorithm.

Integrated Nested Laplace Approximation (INLA), one approach developed by Rue [15] can be used to find posterior probability and relative risk for public health data. Thus, this study was conducted to analyse spatial distribution and to build a map of positive Covid-19 patients in Surabaya with INLA approach. Since Covid-19 is known as a disease that spreads through droplets, thus this study assumed that villages which have a large population density will spread the disease quickly. The map can be used not only to assess the pattern of the disease spread but also to distinguish which villages have relatively high risks or relatively low risks. Calculation and plotting data are conducted entirely with R-INLA.

## METHOD AND MATERIAL

The most suitable model for bayesian disease mapping is regression. The correlation between dependent variable and independent variable can be determined easily through this model. Furthermore, regression model that focused on the spatial elements is called the spatial models. It is a subclass of structure additive regression models. It allows us to be able to include some covariates either directly related to spatial or not. This model is also more flexible to use [16]. One of the models that exist in the structure additive regression models is latent Gaussian models (LGM). Suppose  $y_i$  is the number of patient in each village area in Surabaya which is modeled a

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i); \\ \lambda_i &= \rho_i \cdot E_i \end{aligned} \quad (1)$$

with link function  $\eta_i = \log(\rho_i)$ , define latent Gaussian field as  $x = (\eta, \alpha, f, \beta)$ . Since  $\eta, \alpha, f$  and  $\beta$  are random variable then it should have parameter. Suppose  $f(i)$  is parameter for  $y_i$  and  $\psi$  is parameter for  $x$  then we can build a hierarchical structure for this.

In the first stage, we used latent Gaussian model as follows.

$$\eta_i = \log(\rho_i) = \alpha + v_i + v_i + \sum_{j=1}^m \beta_j Z_{ji} \quad (2)$$

where:

- $\lambda_i$  represents the mean
- $E_i$  represents the expected number of cases
- $\alpha$  represents the intercept term
- $v_i = f_1(i)$  represents spatially structured residual in the area of  $i$
- $v_i = f_2(i)$  represents spatially unstructured residual in the area of  $i$
- $\beta_j Z_{ji}$  represents linear effect of  $j$ th covariate in the area of  $i$

In the second stage, we considered spatial structure so then it gave two distributions as follows.

1.  $v_i = f_1(i)$  spatially structured residual with an intrinsic conditional autoregressive structure

$$\begin{aligned} v_i | v_{j \neq i} &\sim \text{Normal}(m_i, s_i^2) \\ m_i &= \frac{\sum_{j \in \mathcal{N}(i)} v_j}{\#\mathcal{N}(i)} \text{ and } s_i^2 = \frac{\sigma_v^2}{\#\mathcal{N}(i)} \end{aligned} \quad (3)$$

Where:  $\#\mathcal{N}(i)$  is number of area which share boundaries

2.  $v_i = f_2(i)$  spatially unstructured residual with an exchangeable prior

$$v_i \sim \text{Normal}(0, \sigma_v^2)$$

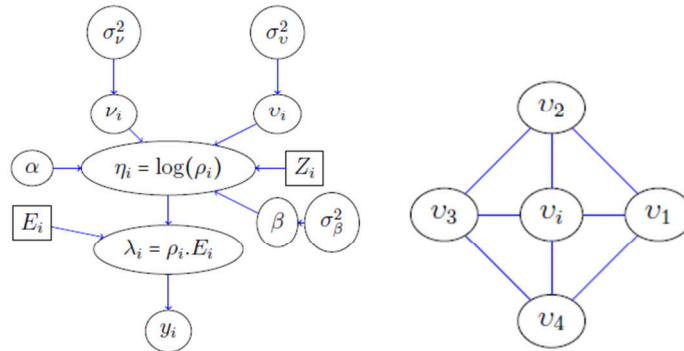
Furthermore, in the third stage, we defined hyperparameter as follows.

$$\psi = \{\sigma_v^2, \sigma_v^2\} \quad (4)$$

Informative prior are specified by INLA's default, as follows:

$$\begin{aligned} \sigma_v^2 &\sim \text{Gamma}^{-1}(1, 0.0005) \\ \sigma_v^2 &\sim \text{Gamma}^{-1}(1, 0.0005) \end{aligned}$$

Figure 2 represents the framework of spatial structure in each area which is satisfy Gaussian Markov random field (conditional independence) [17].

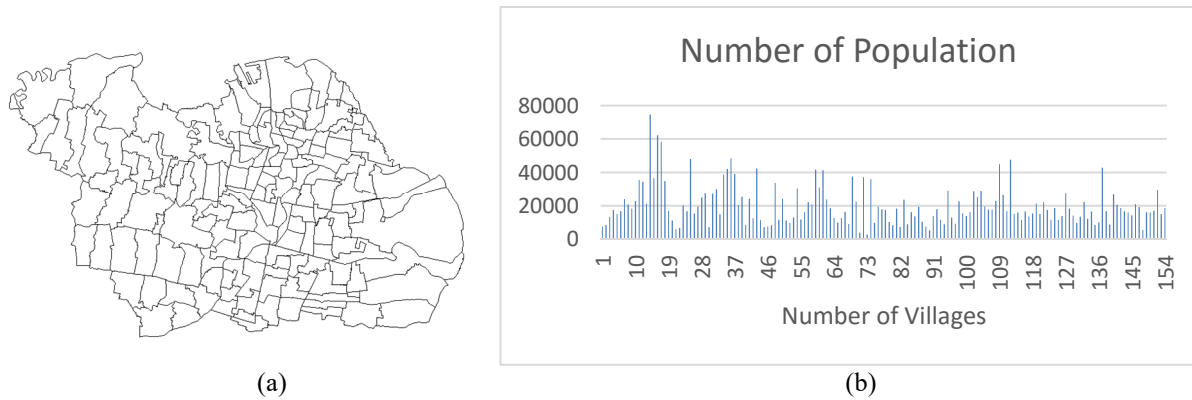


**FIGURE 2.** Framework of Hierarchical Structure

Furthermore, the data used in this study were positive patient of Covid-19 in every villages in Surabaya with the number of cases for each village was simbolized by  $y_i$  with a period ranging from March 2, 2020 to May 10, 2020. The data obtained from <https://lawancovid-19.surabaya.go.id/> there were 708 positive Covid-19 patients spread across 154 villages in Surabaya. Meanwhile, the population density for each village ranged from 2,644 to 74,665 people/km<sup>2</sup>.

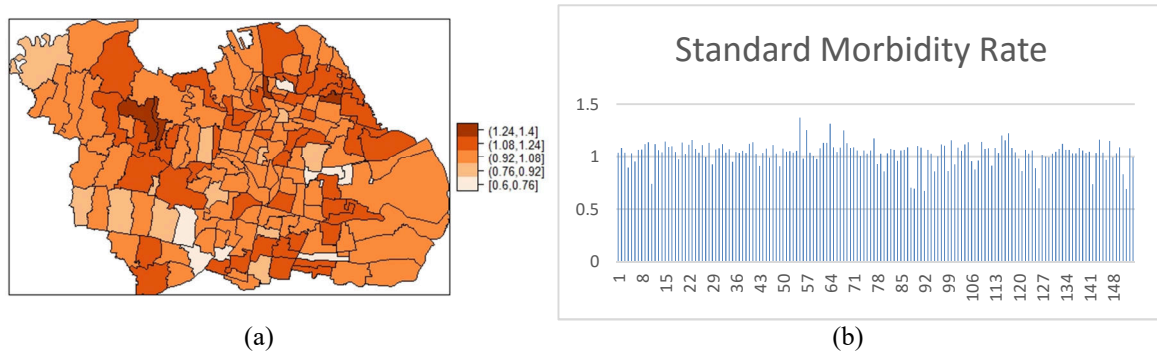
## RESULTS

Data obtained from the Badan Pusat Statistik of Surabaya, there are 154 urban villages spread across five regions, namely North Surabaya, East Surabaya, West Surabaya, South Surabaya and Central Surabaya. Meanwhile, the area of Surabaya city is 326.81 km<sup>2</sup> with the population of Surabaya city based on 2018 registration results as 3,102,089 people. Figure 3 shows a map of the city of Surabaya with 154 villages and number of population in each village.



**FIGURE 3.** (a) The map of Surabaya with 154 villages and (b) number of population in each village

Furthermore, Covid-19 positive patient data for each village was used to map the disease. Disease mapping is based on the results of the estimated relative risk values. Meanwhile, the relative risk measure that is most commonly used in disease mapping is the Standardized Morbidity Ratio (SMR). This is the ratio between cases in region  $i$  ( $y_i$ ) which is assumed to be distributed with Poisson and the expected value of cases occurring in region  $i$  ( $E_i$ ). Figure 4 shows the SMR for each village in Surabaya.



**FIGURE 4.** (a) The relative risk map of the postive Covid-19 patients and (b) The SMR in each village

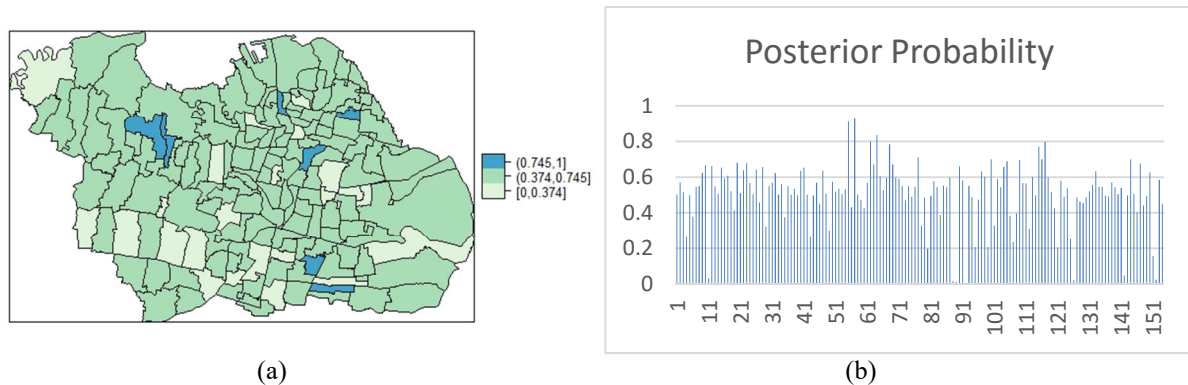
In Fig. 4, it can be seen that the villages with dark brown color have an SMR value between 1.24 to 1.4 or the highest SMR value. It means that these villages had the highest relative risk compared to other villages. These villages were the areas most at risk from the spread of Covid-19 disease. In figure 4, it also can be seen that Manukan Wetan, Balongsari, Dukuh Setro and Nyamplungan villages had the highest risk of the spread of Covid-19. Meanwhile, until May 10, 2020, the villages that were not too risky for the spread of Covid-19 were Sidotopo, Gebang Putih, Manyar, Rungkut Kidul, Kebonsari, Kebraon, and Wiyung villages. These villages only had SMR

values ranging from 0.6 to 0.76. It means that these area ware not too risky for the spread of Covid-19 disease in Surabaya or had a relatively low risk.

Moreover, by using INLA, posterior probability can be produced that represent the probability of finding cases of covid-19 in one village after population density is considered. The proportion of people which probably will be infected of Covid-19 can be obtained. The interpretation of the probability is divided into three levels [17]:

- The region with the probability value between 0.745-1. It implies that the region will have a high probability to the spread of Covid-19 or in the other word we will discover a lot of new cases of Covid-19 or we are going to find a new patient of Covid-19 easily in this region. The region with this probability value is defined as areas with high risk to the spread of the disease.
- The region with the probability value between 0.374-0.745. It implies that the region will have a medium probability to the spread of Covid-19 but it is not really dangerous or in the other word we will discover new cases of Covid-19 but it is not as many as the region with the probability value 0.745-1. The region with this probability value is defined as areas with medium risk to the spread of the disease.
- The region with the probability value between 0-0.374. It implies that the region will have a small probability to the spread of Covid-19 or in the other word we will not discover new cases of Covid-19 or it is not easy to find a new patient of Covid-19 in this region. The region with this probability value is defined as areas with low risk to the spread of the disease.

The posterior probability, which was obtained from arcGIS and INLA, can be seen in Fig. 5.



**FIGURE 5.** (a) The posterior probability map of postive Covid-19 patients and (b) The PP in each village

In Fig. 5, the darker colors on the map indicate that the posterior probability of finding new cases of Covid-19 in those region are higher than the other region. Some villages such as Tenggilis Mejoyo, Rungkut Tengah, Manukan Wetan, Balongsari, Dukuh Setro, Nyamplungan and Pacar Keling are villages with posterior probability values between 0.745 to 1. It means that these villages have a higher probability of finding new covid-19 patients. Meanwhile, some villages such as Sumberejo, Lakarsantri, Lidah Kulon, Babatan, Wiyung, Kebraon, Kebonsari, Ketintang, Margorejo, Siwalankerto, Rungkut Kidul, Wonorejo Rungkut, Gebang Putih, Manyar Sabrangan, Mojo, Simomulyo Baru, Jeparo, Kapasari, Sidotopo and Tegal Sari are villages in Surabaya which have a posterior probability value between 0 to 0.0374. It means that these villages have a low probability of finding new cases or new patients with covid-19.

## CONCLUSION

In this study, we concluded that the Bayesian hierachical model with the INLA approach can be used to determine the relative risk and posterior probability of the spread of Covid-19 in some villages in Surabaya. Regard to the results of the estimated relative risk through the calculation of the standard morbidity rate, the villages of Manukan Wetan, Balongsari, Dukuh Setro and Nyamplungan had the highest risk of the spread of Covid-19 disease. Meanwhile, from the calculation of posterior probability, villages such as Tenggilis Mejoyo, Rungkut Tengah, Manukan Wetan, Balongsari, Dukuh Setro, Nyamplungan, and Pacar Keling had a higher probailty of finding new cases or new positive patients infected with Covid-19. Of the total 154 urban villages in Surabaya, only around 4.5% have a high probability of finding positive new Covid-19 cases.

## REFERENCES

1. J. Niu, L. Shen, B. Huang, F. Ye, L. Zhao, H. Wang, Y. Deng, W. Tan. Non-Invasive Bioluminescence Imaging of Hcov-OC43 Infection and Therapy in The Central Nervous System of Live Mice, *Antiviral Research* **173**, (2020).
2. F. A. Rabi, M. S. Al Zoubi, G. A. Kasasbeh, D. M. Salameh, A. D. Al-Nasser, SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens* **9**(3), 231 (2020). <https://doi.org/10.3390/pathogens9030231>
3. M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research* **24**, 91-98 (2020).
4. Y. Fan, K. Zhao, Z. L. Shi, P. Zhou. Bat Coronaviruses in China. *Viruses* **11**, 210 (2019).
5. D. X. Liu, J. Q. Liang, T. S. Fung. Human Coronavirus-229E, -OC43, NL63 and -HKU1. *Reference Module in Life Sciences*. (2020). [10.1016/B978-0-12-809633-8.21501-X](https://doi.org/10.1016/B978-0-12-809633-8.21501-X).
6. C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Y. Song. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* **180**(7), 934-943 (2020).
7. R. H. Du, L. R. Lian, C. Q. Yang, W. Wang, T. Z. Cao, M. Li, G. Y. Guo, J. Du, C. L. Zheng, Q. Zhu, M. Hu, X. Y. Li, P. Peng, H. Z. Shi. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. *European Respiratory Journal* **55**, 2000524 (2020).
8. Y. C. Li, W. Z. Bai, T. Hashikawa, The Neuroinvasive Potential of SARS-CoV-2 May Play A Role in The Respiratory Failure of Covid-19 Patients. *Journal of Medical Virology*. <https://doi.org/10.1002/jmv.25728>
9. H. A. Rothan, S. N. Byrareddy. The Epidemiology and Pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity* **109**, (2020)
10. J. F. Ludvigsson. Systematic Review of Covid-19 in children shows milder cases and a better prognosis than adults. *ACTA PAEDIATRICA*, (2020). DOI: [10.1111/apa.15270](https://doi.org/10.1111/apa.15270)
11. Y. Bai, L. Yao, T. Wei. Presumed Asymptomatic Carrier Transmission of Covid-19. *JAMA* **323**(14), 1406-1407 (2020).
12. Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, J. Wang, Z. Hu, Y. Yi, H. Shen. Clinical Characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Sci China Life Sci.* **63**(5), 706-711 (2020).
13. WHO, Situation Report-116. World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
14. D. Cucinotta, M. Vanelli. WHO Declares Covid-19 a pandemic. *Acta Biomed* **91**, N. 1, 157-160 (2020).
15. H. Rue, S. Martino, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Statist. Soc. B* **71**, Part 2, pp. 319–392 (2009).
16. M. Blangiardo, M. Cameletti. *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley & Sons.
17. R. Artiono. Parameter Estimation Using INLA for Disease Mapping of Leptospirosis in Bantul Indonesia. *J. Phys.: Conf. Ser.* **1417** 012026