Bayesian spatial survival modelling for dengue fever in Makassar, Indonesia

5 authors, including:

Sri Astuti Thamrin
Hasanuddin University
52 PUBLICATIONS   **301** CITATIONS

SEE PROFILE

Aswi Aswi
State University of Makassar
131 PUBLICATIONS   **402** CITATIONS

SEE PROFILE

Ansariadi Ansariadi
Hasanuddin University
94 PUBLICATIONS   **4,332** CITATIONS

SEE PROFILE

Andi Kresna Jaya
Hasanuddin University
27 PUBLICATIONS   **57** CITATIONS

SEE PROFILE

Bayesian spatial survival modelling for dengue fever in Makassar, Indonesia[☆]

Sri Astuti Thamrin [a,*], Aswi [b], Ansariadi [c], Andi Kresna Jaya [a], Kerrie Mengersen [d]

[a] *Department of Statistics, Hasanuddin University, Makassar 90245, Indonesia*
[b] *Statistics Study Program, Universitas Negeri Makassar, Makassar 90224, Indonesia*
[c] *Department of Epidemiology, Hasanuddin University, Makassar 90245, Indonesia*
[d] *School of Mathematical Science, Queensland University of Technology, Brisbane 4000, Australia*

ARTICLE INFO

ABSTRACT

Objective: To understand the spatial pattern of dengue fever (DF) patients' survival and investigated factors influencing DF patients' survival.
Method: A Bayesian spatial survival method via a conditional autoregressive approach was used to analyze the factors that influence DF patients' survival in 14 sub-districts from January 2015 to May 2017 in Makassar city, Indonesia. Bayesian spatial and a non-spatial model were compared by using deviance information criterion.
Results: The spatial model was more suitable than a non-spatial model. Under the Bayesian spatial model, there was a substantive relationship between age, grade and DF patients' survival time.
Conclusions: The relative risk map and related factors of DF patients' survival can indicate the health policy makers to give special attention to the high risk areas in order to faster and more targeted treatment.

© 2020 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Dengue fever (DF) is a disease that often causes outbreaks or epidemics. The recent report from the World Health Organization[1] reported that globally, there has been a significant increase in the number of DF every year. Indeed, Indonesia has the most cases of DF disease in Southeast Asia region with more than 129,650 people infected with the virus in 2015.[2] Makassar city is one of the areas in Indonesia that has DF virus each year. There is a real need to increase efforts to prevent transmission of the disease and reduce the mortality rate.

Efforts to reduce the death rate from DF virus can be either preventive or curative. An example of a preventive method is raising awareness and a healthy lifestyle to reduce the incidence and spread of DF.[3] An example of a curative method is providing the right medical treatment to patients. Here, the issue of survival of patients with DF becomes important to investigate.

Survival time of patients with DF disease can be influenced by many factors. So far in the literature, in general, the factors can be grouped into two major themes. The first theme is associated with physical or biological characteristics of DF patients such as age, gender, etc.[1] The second theme covers a number of non-biological factors such as location, local conditions, rainfall, socio-economic conditions and therapy or treatment given to patients with DF.[4-7]

Survival data are commonly analyzed using proportional hazards regression.[8-10] However, this method is based on a relatively simple model and it does not allow the incorporation of external information such as survival spatial information. Bayesian approaches to modelling and analysis of DF have become increasingly popular over the last decade. For example, Honorato et al.[11] investigated cases of DF which were spatially distributed in Brazil using a Bayesian spatial hierarchical model using conditional autoregressive (CAR) prior. Lowe et al.[12] used a general spatio-temporal mixed model with a Bayesian framework taking into account geographical and temporal resolution to evaluate the spread of DF in Brazil Southeast. Furthermore, Jaya et al.[13] used a Bayesian spatial approach to model and map the DF cases in Bandung, Indonesia.

As noted above, most of this literature focuses on incidence and there are fewer papers on DF survival. The literature on spatial modelling of DF survival is even more sparse and limited to small-scale studies.[14-18] For example, Thamrin and Taufik[14] used multilevel survival to study DF survival. Aswi et al.[18] considered DF survival in a single hospital in Makassar using Bayesian Weibull survival and a semiparametric Cox proportional hazards model. In this study, a Bayesian spatial survival method via a CAR approach was used to analyze DF patients' survival, and to investigate factors that influence DF patients' survival. This is investigated through a DF patients' survival in Makassar, Indonesia.
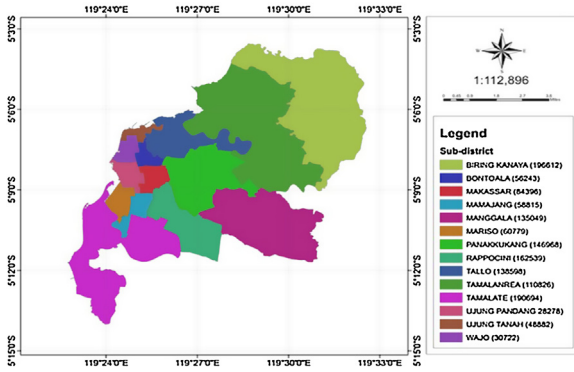
Methods

Study area and source of data

Makassar is the most populous city in South Sulawesi or even in the eastern part of Indonesia.[19] Makassar city consists of 14 sub-districts with 143 villages in these sub-districts (Figure 1).[20]
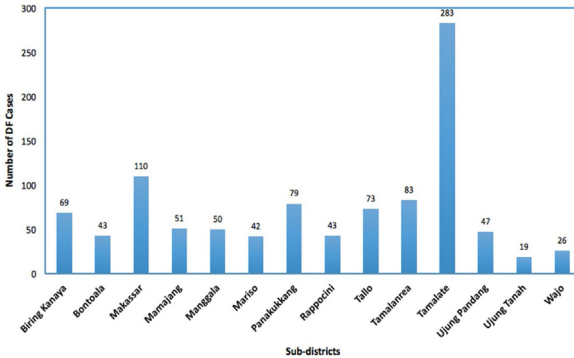

Figure 1. Map of Makassar city, Indonesia, with the population density of each sub-district.

Figure 2. The number of DF incidences in Makassar city, Indonesia based on sub-district for the period of 2015–2017.

Data used were collected from four big hospitals in Makassar city, that is, Dr. Wahidin Sudirohusodo, Labuang Baji, Pelamonia, and Haji Hospitals. In this dataset, the variables used were sex, haematocrit, leukocyte, haemoglobin, thrombocyte, level severity of DF type (grade), age, survival time, and address of DF patients. The number of patients diagnosed with DF were treated in these hospitals, in the period of January 2015 to May 2017 is 934.

In our models, the survival response variable t_{ij} is defined as the number of days that a patient stayed in the hospital until they recovered or got consent to leave the hospital or death. The DF patients self-discharge or transfer to another hospital and still alive at the end of the study period were treated as 'censored'. The DF incidence rates, which is the number of cases per 10,000 people per year, for each sub-district from the four big hospitals in Makassar city are presented in Figure 2. Data on population numbers for the same time period were collected from the Central Bureau of Statistics of Makassar city, Indonesia.[19] The population size was used to calculate and to map the hazard rate.

Spatial survival model

Following the description in Banerjee et al.,[17] let t_{ij} be the response described in the previous section, namely time to recovery for subject j in sub-district $i, j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, l$. Let x_{ij} be the corresponding vector of individual-specific covariates. Assume a proportional hazard $h(t_{ij}; x_{ij})$ that follows a Weibull distribution with $h_0(t_{ij}) = \rho t_{ij}^{\rho-1}$ as follows:

$$h(t_{ij}; x_{ij}) = h_0(t_{ij}) \exp(\beta^T x_{ij}) \tag{1}$$

$h_0(t_{ij})$ represents a baseline hazard. The exponential term involving the covariates affect the baseline hazard multiplicatively. Here, ρ is the shape parameter of the baseline hazard and the vector β

contains the intercept for the baseline hazard. The parameter ρ represents monotonicity of the hazard rate in the Weibull mode,[21] where $\rho = 1$ represents a constant hazard rate, and $\rho > 1$ and $\rho < 1$ indicate that the hazard will increase and decrease monotonically with time, respectively. Since $W_i = \log \omega_i$, in the frailty setting,[22] Eq. (1) is extended to:

$$h(t_{ij}; x_{ij}) = \rho t_{ij}^{\rho-1} \exp(\beta^T x_{ij} + W_i). \tag{2}$$

W_i is the sub-district-specific frailty term with distribution:

$$W|\lambda \sim CAR(\lambda). \tag{3}$$

$$t_{ij} \sim Weibull(\alpha, \mu),$$

$$\rho \sim Gamma(1, 1),$$

$$\beta \sim Normal(0.0, 0.001),$$

$$\beta_0 \sim Normal(0.0, 0.001)$$

where $\log(\mu) = \beta_0 + \beta_x$. CAR denotes a conditionally autoregressive structure $W_i|W_{i' \neq i} \sim N(\bar{W}_i, 1/\lambda m_i)$[17,22] and λ is the CAR parameter in Eq. (3) which states the

$$\lambda^{1/2} \exp \left[-\frac{\lambda}{2} \sum_{i=1}^{l} (W_i - W_i')^2 \right] \propto \lambda^{1/2}$$

$$\exp \left[-\frac{\lambda}{2} \sum_{i=1}^{l} m_i W_i (W_i - \bar{W}_i) \right] \tag{4}$$

where i adj i' denotes regions i' that are adjacent to region i, $\bar{W}_i$ is the average of the $W_{i \neq i'}$ that are adjacent to W_i, and m_i is the number of these adjacencies.

To complete the Bayesian model specification in Eq. (4), we use prior distributions on ρ, β and σ^2. We use a gamma hyperprior distribution for λ. With this parameter, the gamma distribution has mean 1 and the variance 1; see Banerjee et al.[17] In this paper, we include both spatial and non-spatial frailties.[17] Non-spatial frailties means that the model has an unstructured frailty term, expressed as independently and identically distributed (i.i.d) $N(0, \sigma^2)$. The model was fitted using the OpenBUGS software.[23] The MCMC algorithm was run with 150,000 iterations, discarding 10,000 as burn-in. Monitoring MCMC convergence was done through trace plots, autocorrelation plots, and plots of the posterior distributions of the model parameters.

Goodness of fit measures were used to test whether a spatial survival model adequately represents the data. The deviance information criterion (DIC)[24] is one of the criteria in evaluating the best model for several data sets using the Bayesian approach. A model is more suitable when it has the smallest DIC.[24]

Results

The model was fitted to DF survival dataset as described in Section "Spatial Survival Model". By using the different set of starting values, the chain was run. Convergence can be assessed whether it has been achieved or not through the MCMC chain. The results of the hazard rates obtained from Eq. (2) were mapped using the Map tool R version 4.0.2 package.

According to Figure 1, the city of Makassar is divided into 14 sub-districts so a spatial weights matrix of queen contiguity order one of 14×14 was formed. The corresponding adjacency matrix for Makassar city, Indonesia is shown in Figure 3. The Weibull assumption for the survival time was assessed by plotting $\log[-\log(S(t))]$ versus $\log(t)$, where the survival probability $S(t)$ was obtained using Kaplan–Meier estimation and t is the survival time. The

```
list(Nsubj=934,regions=14, nsum=58, adj=c(2,      3,      6,
1,      3,      4,      5,      6,
1,      2,      4,      6,
2,      3,      5,      11,     12,
2,      4,      6,      8,      10,     11,
1,      2,      3,      5,      7,      8,
6,      8,      9,
5,      6,      7,      9,      10,
7,      8,      10,
5,      8,      9,      11,     14,
4,      5,      10,     12,     14,
4,      11,     14,
14,
10,     11,     12,     13      ),
num = c(3, 5, 4, 5, 6, 6, 3, 5, 3, 5, 5, 3, 1, 4),
```

Figure 3. Adjacency matrix for Makassar city, Indonesia.

Weibull distribution was indicated by a straight line plot. This assumption was also verified by application of the Mann test (if $F_{approximation} = 0.784 < F_{criterion} = 1.17$ at the 5% level, then the null hypothesis is accepted). The Weibull distribution was thus considered to be appropriate for these data.

From Figures 4 and 5, the MCMC outputs for the significant parameters of the model in the form of posterior density and trace

Table 1
Summary of posterior estimated parameters for spatial survival for DF survival data.

Parameter	Mean	95% CI
Intercept (β_0)	−4.297	(−4.951, −3.652)
Sex (β_1)	0.082	(−0.033, 0.187)
Age (β_2)	−0.006	(−0.012, −0.0004)
Grade (β_3)	0.199	(0.079, 0.315)
Haemoglobin (β_4)	0.097	(−0.076, 0.269)
Haematocrit (β_5)	−0.088	(−0.177, 0.003)
Thrombocyte (β_6)	−0.022	(−0.091, 0.047)
Leucocyte (β_7)	−0.049	(−0.161, 0.061)
ρ	2.473	(2.342, 2.607)
λ	3.15	(1.176, 6.335)

plots show a well-behaved and fast mixing process. The posterior estimates of parameters and corresponding 95% confidence intervals (CI) are presented in Table 1. It can be inferred that age and grade substantially describe patients' survival times. The variables age and grade both have a negative effect on the expected survival time. For example, an increase of one year in patient's age increased the hazard at DF survival by $\exp(−0.006) = 0.994$ days.

Table 2 provides a summary of model fit with a corresponding non-spatial survival model as a comparison. In the latter, Eq. (3) is replaced by $W_i \sim N(0, \sigma^2)$. The spatial model is more suitable for these data. This is indicated by the larger size of the effective parameter pD and the smaller value of DIC.
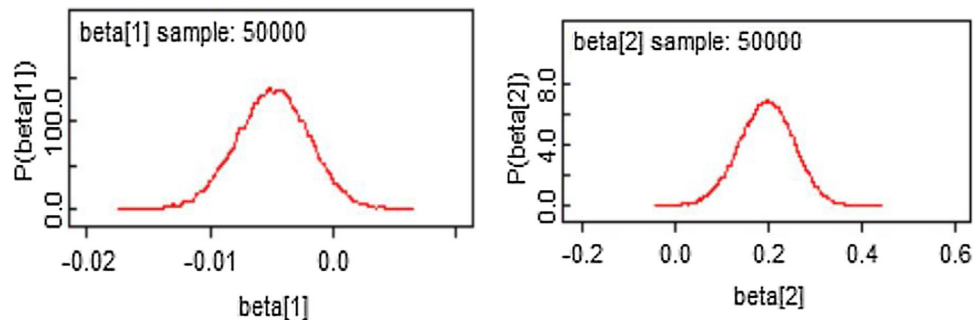

Figure 4. Posterior densities plots of the significant parameter for DF survival data in the Makassar city, Indonesia.
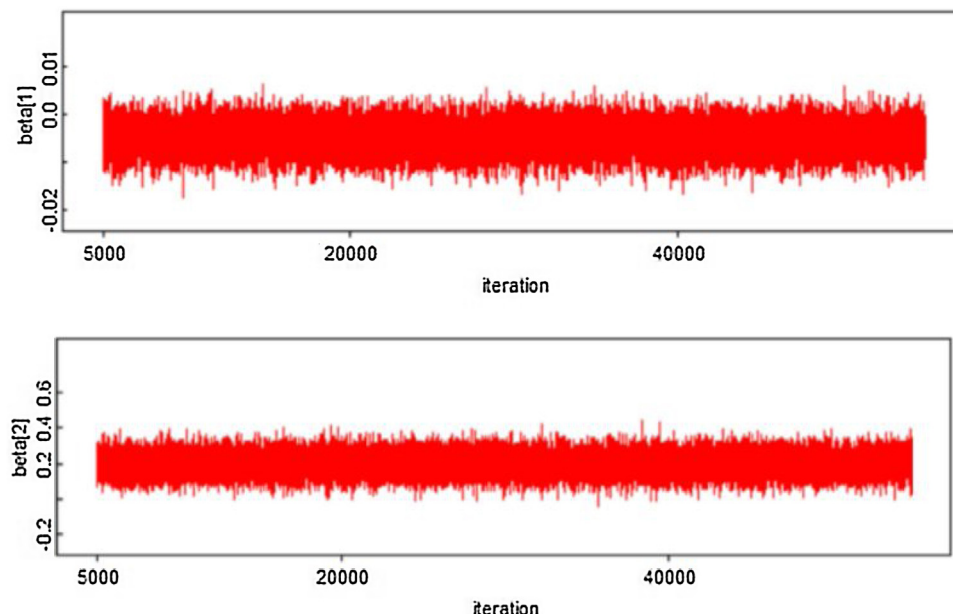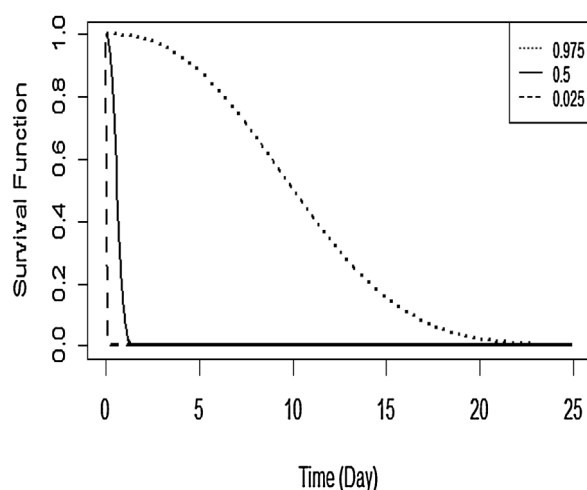

Figure 5. History iteration plots of the significant parameter for DF survival data in the city of Makassar, Indonesia.

Table 2
The values of pD and DIC for survival spatial and non-spatial models.

Model	pD	DIC
Spatial survival	13.5	3698
Non-spatial survival	9.115	3725

Figure 7. Estimated of the posterior survival function $S(t)$ for DF data with age and grade variables.

The mapped summary of the posterior mean survival times per sub-district is shown in Figure 6. The results indicate that regional sub-districts located around urban centres, namely Mariso had the lowest hazard rate. The constant hazard rates occurred in the sub-districts Manggala, Panakkukang, Bontoala, Rappocini, Tamalanrea, and Wajo. The higher hazard rate of DF patients occurred in the sub-districts Biringkanaya, Makassar and Mamajang.

The Kaplan–Meier curve for DF survival with covariates age and grade is shown in Figure 7. The covariates were established for sur-vival time in the first patient for DF. The probability of survival for this first patient is predicted to fall fairly rapidly towards 0.10 when the survival time is about 2 days, then declines very slowly thereafter.

Discussion

We have presented the modelling of the spatial correlated sur-vival data in this paper. To estimate spatial survival with a CAR model, a Bayesian model via a MCMC computational technique was used. The case study used the survival times of DF patients from hospitals in Makassar, Indonesia for the period 2015–2017 with covariates.

Overall, the results of this study indicated that age and grade fac-tors appear to influence the number of days of patients hospitalised with DF until they got recovery or were allowed to go home. The prediction of DF patients' survival time based on influencing vari-ables is also illustrated. This suggests that when the survival time is about 2 days, the probability of patient's survival time drops fairly rapidly towards 0.10 and decays very slowly thereafter. Mariso had the lowest hazard rate while Biringkanaya had the highest haz-ard rate, fooled by Makassar and Mamajang. These districts are the potential regions to target for health policy intervention in high risk areas and for more targeted, efficient allocation of the relevant resources. Notwithstanding this, further investigation is needed in order to consider socio economic, temperature, and humidity effects in modelling the survival factors.[5–7,12] The flexibility of the Bayesian spatial survival models invites extensions to more com-plicated scenarios such as a spatial clustering[6] and spatial-temporal model.[5,25]

Conclusions

The relative risk map and related factors of DF patients' sur-vival can indicate the health policy makers to give special attention to the high risk areas in order to faster and more targeted treatment.
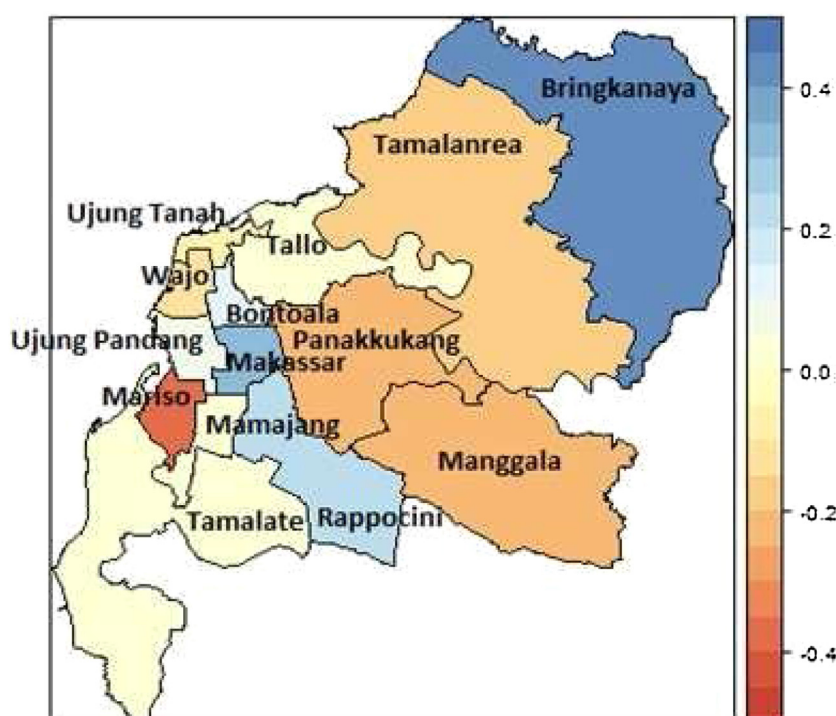
Figure 6. Posterior mean CAR frailties model with covariates for Makassar county level DF survival data.

Authorship contributions

SAT contributed to the concept and design of the study, carried out statistical analysis and wrote manuscripts. AS and KM interpreted the data, analyzed and wrote the manuscript. AKJ carried out the statistical analysis and AN collects the necessary data. All authors read and approved the final manuscript.

Acknowledgements

Conflicts of interest

The authors declare that there is no conflict of interests.

References

1. World Health Organization. Global strategy for dengue prevention and control 2012–2020; 2012.
2. Kementerian Kesehatan. Profil kesehatan Indonesia 2015.
3. Kautner I, Robinson MJ, Kuhnle U. Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention. J Pediatr. 1997;131:516–24.
4. Hales S, De Wet N, Maindonald J, et al. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. Lancet. 2002;360:830–4.
5. Sun W, Xue L, Xie X. Spatial-temporal distribution of dengue and climate characteristics for two clusters in Sri Lanka from 2012 to 2016. Sci Rep. 2017;7:1–12.
6. Huang C-C, Tam TYT, Chern Y-R, et al. Spatial clustering of dengue fever incidence and its association with surrounding greenness. Int J Environ Res Public Health. 2018;15:1869.
7. Kikuti M, Cunha GM, Paploski IAD, et al. Spatial distribution of dengue in a Brazilian urban slum setting: role of socioeconomic gradient in disease risk. PLoS Negl Trop Dis. 2015;9:e0003937.
8. Nguyen DV, Rocke DM. Partial least squares proportional hazard regression for application to DNA microarray survival data. Bioinformatics. 2002;18:1625–32.
9. Rosenwald A, Wright G, Chan W, et al. Molecular diagnosis and clinical outcome prediction in diffuse large B-cell lymphoma. N Engl J Med. 2002;346:1937–47.
10. Kleinbaum DG, Klein M. Survival analysis. Springer; 2010.
11. Honorato T, de Aquino Lapa PP, Sales CMM, et al. Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of Bayesian model. Rev Bras Epidemiol. 2014;17:150–9.
12. Lowe R, Bailey TC, Stephenson DB, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. Stat Med. 2013;32:864–83.
13. Jaya I, Abdullah AS, Hermawan E, et al. 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. 2016;54:94–103.
14. Astuti Thamrin S, Taufik I. Spatial random effects survival models to assess geographical inequalities in dengue fever using Bayesian approach: a case study. J Phys. 2018;979:12095.
15. Iriawan N, Astutik S, Prastyo DD. Markov Chain Monte Carlo-based approaches for modeling the spatial survival with conditional autoregressive (CAR) frailty. CiteSeerX; 2010.
16. Thamrin SA, Amran, Jaya AK, et al. Bayesian inference for spatial parametric proportional hazards model using Spatsurv R. In: AIP Conference Proceedings, vol. 1827. AIP Publishing LLC; 2017. p. 20015.
17. Banerjee S, Wall MM, Carlin BP. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. Biostatistics. 2003;4:123–42.
18. Aswi A, Cramb S, Duncan E, et al. Bayesian spatial survival models for hospitalisation of Dengue: a case study of Wahidin hospital in Makassar, Indonesia. Int J Environ Res Public Health. 2020;17:878.
19. Makassar BK. Penduduk Kota Makassar; 2017.
20. Affairs M of I. Registration Book for Area Code and Data of 2013, Indonesia; 2013.
21. Box-Steffensmeier JM, Jones BS. Event history modeling: A guide for social scientists. Cambridge University Press; 2004.
22. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43:1–20.
23. Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: evolution, critique and future directions. Stat Med. 2009;28:3049–67.
24. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. J R Statist Soc B. 2002;64.
25. Cramb SM, Mengersen KL, Baade PD. Spatio-temporal relative survival of breast and colorectal cancer in Queensland Australia 2001–2011. Spat Spatiotemporal Epidemiol. 2016;19:103–14.