ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Spatial-temporal Data Analysis Application for COVID-19 Relative Risk Mapping and Modeling,the case of Alabama state, USA.*

Getachew Dejene[a]

[a]Faculty of Informatics, University of Debrecen, Hungary.
getlala27@gmail.com

Abstract

Since the end of 2019, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has been found in China and has spread to all over the other nations, including the United States. Investigating how the sickness varies over time and space, areas that are particularly at risk, developing etiological hypotheses, measuring disparities and advocating better resource allocation. In this study, we estimated, modeled, and mapped the relative risk of the COVID-19 disease in Alabama State, USA. The methodology Generalized Linear Model (GLM) were utilized to estimate the relative risk of COVID-19 diseases for each county and weeks of study period using county level new infections and Areal map data were obtained from the official public Kaggle repository site between 2021-01-01 and 2021-12-31. The software package called R-INLA for fitting model, *ggplot*2 and *plotly* R package for mapping has been applied.

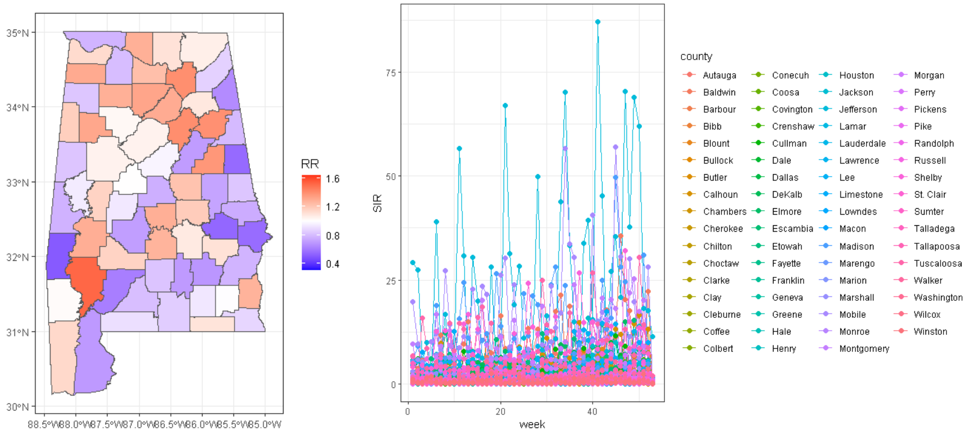
COVID-19 Disease risk varies in space and time due to variation in many factors and monitoring of sickness risk in public health has a long history[6, 8]. Spatio-temporal Disease maps are allow for a rapid visual review of geographic data and also allow for the detection of patterns that may be missed in tabular presentations[1]. These maps are critical for describing how illness varies over time and space, temporal development of spatial patterns of mortality or incidence hazards, identifying areas that are particularly at risk, developing etiological hypothe-

*PhD student, under supervision of : Prof. Dr. György Terdik, professor emeritus ses, measuring disparities, and advocating for better resource allocation and many more [4, 6]. These issues are frequently answered using information that summarizes illness incidence rate at the population level and has been spatio-temporally aggregated to a set of N non overlapping areal units for M consecutive time periods[2]. Despite the fact that the Standard Incidence Rate (SIR) is an unbiased estimate of relative risk, it may be misleading and untrustworthy in regions with a small population size or disease count, and smoothing could be employed to reduce the amount of variance caused by population size and heterogeneity. This has motivated researchers to estimate risk using the spatio-temporal model-based method instead. Since the end of 2019, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has been found in Wuhan city, China and has spread to all over the other nations including United States of America (USA)[3]. Due to the wealth of technology innovation and the accumulation of observed data through time, Spatio-temporal data like COVID-19, are plentiful and simple to get[5]. Data are called spatio-temporal as long as each one of them carries a location and a time stamp[7]. Examples include satellite images of certain parts of the earth, temperature readings from several nearby stations, human or animal travel patterns that may also include sensor readings, disease out breaks,etc. we do so by fitting a general linear model (GLM) with the family function specified as a Poisson

$$Y_{ij} \mid \theta_{ij} \sim \text{Poisson}(E_{ij}\theta_{ij})$$

where, Y_{ij} is the observed number of cases, E_{ij} is the expected number of cases in ith area and the jth time interval and θ_{ij} is the unknown true relative risk in area i and time period j.

For this study the data were obtained from the official Kaggle repository site https://www.kaggle.com/datasets/headsortails/covid19-us-county-jhu-data-demographics, Our data set consists of daily counts of COVID-19 new confirmed cases within each of the 67 spatial units (i.e. counties in state of Alabama, USA) from 1st January, 2021, 31th December, 2021 for Alabama state. We also compute the weekly aggregated confirmed cases as it may be handy in some analyses. Hence, we aim to estimated, mapped and model relative risk (RRs) of COVID-19 from aggregated weekly data at each of the county during the study period. The overall spatio-temporal and evolution of weekly RR of COVID-19 infections is depicted in Figure 1a and 1b as follows. In Figure 1a below, the map presents the SIR of COVID-19 for the 67 counties in Alabama. The colour shows the status of the COVID-19 diseases Standard Incidence rate (SIR) in a county. such that the SIR larger than one means it's colour "Red" that shows the corresponding county is highly risky for COVID-19 diseases, (i.e. Clark, St.Clair and Calhoun). "white" colour means, the SIR is mild in a county like Washington (i.e. Choctaw, Randolph, etc.). Figure 1b, shows a rapid increase in the number of confirmed incidences from Week 31 till Weeks 40–44 (with a dip in Week 40), followed by a substantial decline and leveling off from Week 1 till Week 8 with a short peak in Week 17.



(a) SIR of COVID-19 in Alabama counties.

(b)

Figure 1. Spatio- temporal and evolution of weekly RR of COVID-19 infections.

References

[1] P. Elliott, D. Wartenberg: *Spatial epidemiology: current approaches and future challenges*, Environmental health perspectives 112.9 (2004), pp. 998–1006, doi: https://doi.org/doi:10.2307/3838101.

[2] D. Lee: *A tutorial on spatio-temporal disease risk modelling in R using Markov chain Monte Carlo simulation and the CARBayesST package*, Spatial and Spatio-Temporal Epidemiology 34 (2020), p. 100353, doi: https://doi.org/doi:10.1016/j.sste.2020.100353.

[3] H. Lu, C. W. Stratton, Y.-W. Tang: *Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle*, Journal of medical virology 92.4 (2020), p. 401, doi: https://doi.org/doi:10.1002/jmv.25678.

[4] R. J. Marshall: *A review of methods for the statistical analysis of spatial patterns of disease*, Journal of the Royal Statistical Society: Series A (Statistics in Society) 154.3 (1991), pp. 421–441, doi: https://doi.org/doi:10.2307/2983152.

[5] M. F. McCabe, M. Rodell, D. E. Alsdorf, D. G. Miralles, R. Uijlenhoet, W. Wagner, A. Lucieer, R. Houborg, N. E. Verhoest, T. E. Franz, et al.: *The future of Earth observation in hydrology*, Hydrology and earth system sciences 21.7 (2017), pp. 3879–3914, doi: https://doi.org/doi:10.5194/hess-21-3879-2017.

[6] B. Ritz, I. Tager, J. Balmes: *Can lessons from public health disease surveillance be applied to environmental public health tracking?*, Environmental health perspectives 113.3 (2005), pp. 243–249, doi: https://doi.org/doi:10.2307/3436034.

[7] S. K. Sahu: *Bayesian modeling of spatio-temporal data with R*, Chapman and Hall/CRC, 2022.

[8] P. J. Villeneuve, M. S. Goldberg: *Methodological considerations for epidemiological studies of air pollution and the SARS and COVID-19 coronavirus outbreaks*, Environmental health perspectives 128.9 (2020), p. 095001, doi: https://doi.org/doi:10.1289/ehp7411.