

On stay at home orders: Using the power of data science for spatial and temporal modeling and visualization of COVID-19

Christopher R Prentice

University of North Carolina Wilmington

Rachel Carroll (✉ carrollr@uncw.edu)

University of North Carolina Wilmington

Research Article

Keywords: COVID-19, data science, RShiny, spatio-temporal, stay at home orders, Poisson

Posted Date: February 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-275352/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on July 6th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-93433-z>.

On stay at home orders: Using the power of data science for spatial and temporal modeling and visualization of COVID-19

Abstract

Coronavirus disease 2019 dominated and augmented many aspects of life beginning in early 2020. Related research and data generation developed alongside its spread. We developed a Bayesian spatio-temporal Poisson disease mapping model for estimating real-time characteristics of the coronavirus disease in the United States. We also created [several dashboards](#) for visualization of the statistical model for fellow researchers and simpler spatial and temporal representations of the disease for consumption by community members in our region. Findings suggest that the risk of confirmed cases is higher for health regions under partial stay at home orders and lower in health regions under full stay at home orders, when compared to before stay at home orders were declared. These results confirm the benefit of state-issued stay at home orders as well as suggest compliance to the directives towards the older population for adhering to social distancing guidelines.

Keywords: COVID-19; data science; RShiny; spatio-temporal; stay at home orders; Poisson

1. Introduction

The coronavirus disease 2019 (COVID-19) began in Wuhan, China in December of 2019 and quickly evolved into a global crisis (Trombetta et al. 2016). As of May 18, 2020, data suggested that the pandemic resulted in 1,492,822 confirmed cases and

89,101 deaths in the United States alone (Johns Hopkins University Center for Systems Science and Engineering 2019). Due to the widespread nature of this crisis and the need to develop data-driven insights into COVID-19 and its spread, data from various reputable sources were made publically available and updated daily.

In this project, data science tools are coupled with robust statistical modeling to offer interactive real-time exploration and visualization of COVID-19 at the local, regionally, nationally, and internationally (Chang et al. n.d.; Microsoft Corporation 2014; Salesforce 2003). Data visualization dashboards that are updated automatically and daily furnished an ideal environment for understanding the distribution and characteristics of COVID-19 across the United States over time. These dashboards allowed for the easy dissemination of information to a range of audiences in academe, government and nonprofit organizations, and members of the general populace. We display these dashboards alongside an array of related ones created by colleagues in the UNCW Data Science program, [here](#).

The first steps of this project involved compiling data from multiple sources, including: COVID-19 data from Johns Hopkins University Center for Systems Science and Engineering (Johns Hopkins University Center for Systems Science and Engineering 2019), population and demographic information from the U.S Census Bureau, airport and travel data from the US Transportation Department and US Bureau of Transportation Statistics, and unemployment information from the US Bureau of Labor Statistics (Population Estimates Program 2018; U.S. Bureau of Labor Statistics 2020;

U.S. Bureau of Transportation Statistics 2009; U.S. Transportation Department 2019). The stay at home orders variable was self-constructed with guidance from government mandates and the New York Times (Mervosh et al. 2020). After compiling the data, we constructed visualization dashboards for optimal viewing, focusing on those values that vary over time (e.g., confirmed COVID-19 cases). Next, we implemented the statistical model to spatially and temporally explore the prevalence of confirmed cases and death outcomes. The statistical model we employed was a Bayesian spatio-temporal Poisson disease mapping model that allowed for estimation of important fixed effect parameter estimates in addition to residual spatial and temporal variation (Lawson et al. 2016; Lawson and Lee 2017; Lesaffre and Lawson 2013). The dashboards linked above display the real-time implementation of this statistical model. The results reported in this article focus on the time of the pandemic when stay at home orders were largely in effect.

2. Results

An R Shiny COVID-19 regression dashboard with the results for the data presented in this manuscript is available [here](#). This limits the data to the referenced March 23, 2020 to May 18, 2020 window.

Figure 1 displays the standardized incidence or mortality ratios (y_{ij}/e_{ij}) for confirmed cases, active cases, new cases, and deaths for the first, third, fifth, seventh, and ninth Mondays across all health regions in the United States. These maps suggest that the spatial distribution of increased (greater than 1) vs. decreased (less than 1) risk remains largely consistent across time, which provides justification for the separate spatial and

temporal random effect terms in the model described above. Additional real time visualizations of these outcomes, including representation as a percent of the population, can be seen [here](#) at the county level via daily updates of the Johns Hopkins data. The time window considered in this manuscript is represented as days 1-57 on the slide bar.

Table 1 displays the statistical model parameter estimates associated with the confirmed case and death outcomes for the US as a whole and for the four separate US Census-defined regions. An inverse logarithm was applied to these parameter estimates to offer an interpretation directly related to the multiplicative change in relative risk (θ_{ij}) of the outcome considered. Consequently, an estimate with a 95% credible interval fully less than one suggests decreased risk while an estimate with an interval greater than one suggests increased risk. The estimates for the other two outcomes – active cases and new cases – are largely consistent across all models (these findings are made available in the linked Shiny dashboard). Generally speaking, the parameter estimates for the outcomes in Table 1 suggest increased risk under partial stay at home orders and decreased risk under full stay at home orders compared to before the stay at home orders were instated. Risk appears to decrease as the percent unemployment in a region rises. Similarly, findings suggest risk increases for areas with a higher percent of Black or African American population in the South. Some evidence exists that risk is also higher where there is a higher percent of smokers and a higher percent older population. The testing variable was least consistent among the predictors in its effect on confirmed cases and deaths outcomes. It appears that more positive tests, both

count of positive tests and percent positive tests, suggest more confirmed cases but not necessarily more deaths.

Figure 1: Standardized incidence or mortality ratio for A) confirmed cases, B) active cases, C) new cases, and D) deaths for the first, third, fifth, seventh, and ninth Mondays across all health regions in the United States.

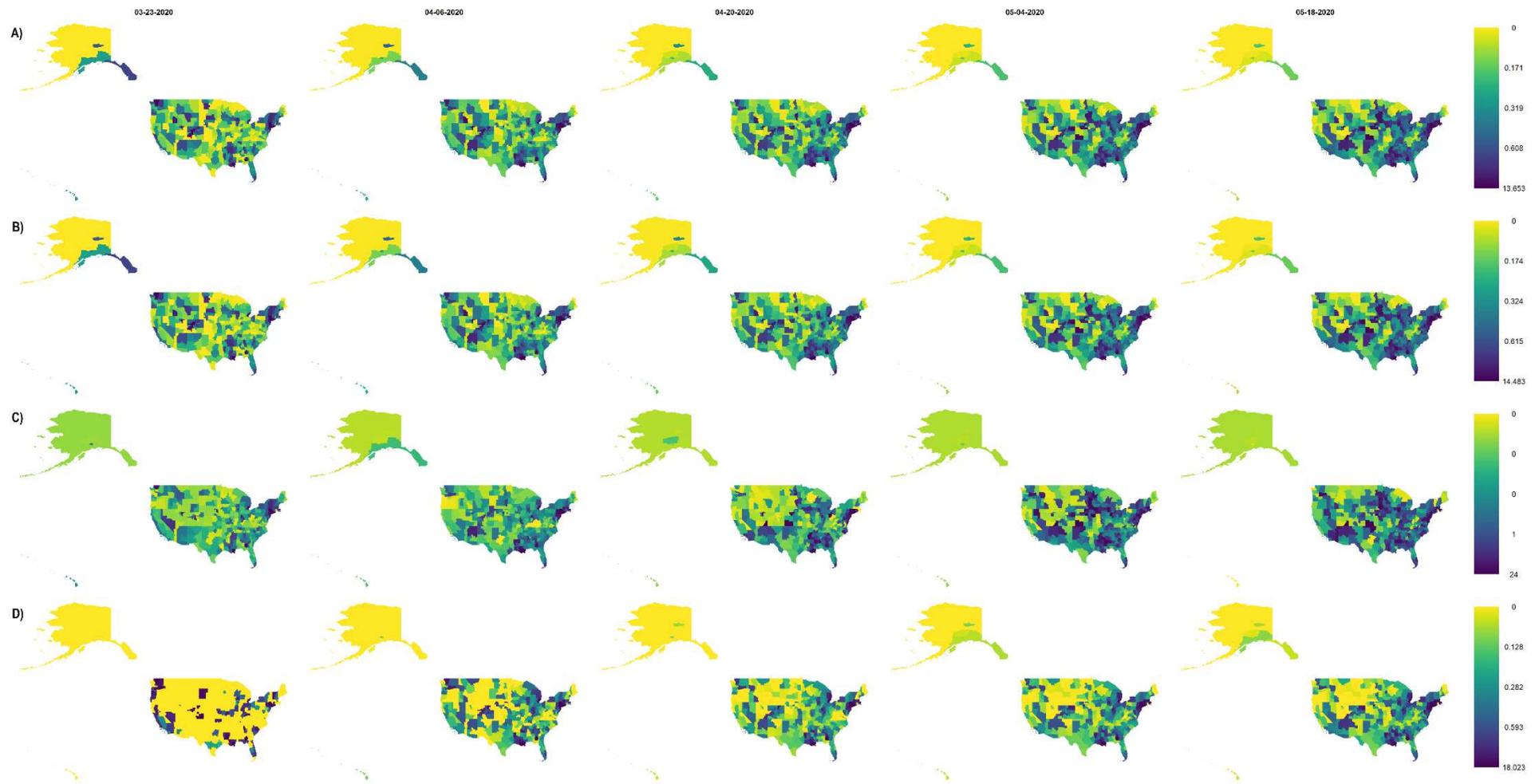


Table 1: Fixed effect parameter estimates for the confirmed cases and deaths outcomes represented as inverse logarithm mean estimates and associated 95% credible intervals.

Predictor	US	South	West	Midwest	Northeast
Confirmed Cases					
International Airport					
No (Ref)	1.000	1.000	1.000	1.000	1.000
Yes	0.928 (0.869, 0.990)*	1.044 (0.995, 1.095)~	0.973 (0.990, 1.051)	1.003 (0.912, 1.103)	1.007 (0.739, 1.371)
State Issued Stay at Home Orders					
Before (Ref)	1.000	1.000	1.000	1.000	1.000
During – Partial	1.005 (1.002, 1.007)*	1.052 (1.047, 1.056)*	0.900 (0.894, 0.906)*	1.014 (1.006, 1.022)*	1.023 (1.018, 1.028)*
During – Full	1.041 (1.039, 1.043)*	0.950 (0.946, 0.954)*	0.855 (0.849, 0.860)*	1.085 (1.081, 1.089)*	0.993 (0.989, 0.997)*
None	1.212 (1.115, 1.316)*	0.930 (0.851, 1.016)	0.850 (0.770, 0.939)*	1.063 (0.970, 1.165)	--- ^B
health region Land Area	1.001 (1.000, 1.001)*	0.997 (0.996, 0.999)*	1.000 (1.000, 1.000)	0.999 (0.996, 1.002)	0.997 (0.987, 1.008)
% Unemployed	0.961 (0.941, 0.982)*	0.978 (0.959, 0.997)*	1.010 (0.988, 1.032)	1.031 (0.993, 1.069)	0.996 (0.899, 1.104)
% Smokers	1.007 (0.998, 1.016)~	1.008 (0.997, 1.018)~	0.995 (0.985, 1.006)	0.993 (0.976, 1.010)	1.004 (0.938, 1.075)
% 65 and Older	1.008 (0.997, 1.018)~	1.001 (0.994, 1.009)	1.000 (0.988, 1.011)	1.000 (0.984, 1.016)	1.005 (0.947, 1.066)
% AA Population	0.996 (0.994, 0.999)*	1.001 (1.000, 1.003)*	1.004 (0.987, 1.021)	1.004 (0.996, 1.012)~	0.997 (0.972, 1.022)
Testing ^A	1.123 (1.121, 1.124)*	1.007 (1.004, 1.010)*	0.976 (0.972, 0.979)*	1.003 (1.003, 1.003)*	1.005 (1.004, 1.005)*
Deaths					
International Airport					
No (Ref)	1.000	1.00	1.000	1.000	1.000
Yes	0.925 (0.854, 1.001)~	1.066 (1.011, 1.124)*	0.929 (0.834, 1.035)	1.013 (0.920, 1.116)	1.015 (0.746, 1.380)
State Issued Stay at Home Orders					
Before (Ref)	1.000	1.00	1.000	1.000	1.000
During – Partial	1.030 (1.019, 1.042)*	1.182 (1.153, 1.211)*	0.875 (0.844, 0.906)*	0.844 (0.810, 0.879)*	1.062 (1.039, 1.085)*
During – Full	1.041 (1.031, 1.051)*	0.935 (0.914, 0.956)*	0.846 (0.820, 0.873)*	0.966 (0.950, 0.982)*	1.044 (1.023, 1.065)*
None	1.282 (1.158, 1.418)*	0.969 (0.877, 1.069)	0.842 (0.731, 0.969)*	1.009 (0.919, 1.109)	--- ^B
Health Region Land Area	1.001 (1.000, 1.001)*	0.998 (0.997, 1.000)~	1.000 (0.999, 1.000)	0.999 (0.996, 1.002)	1.000 (0.989, 1.010)
% Unemployed	0.944 (0.920, 0.968)*	0.973 (0.952, 0.993)*	1.020 (0.990, 1.052)	1.004 (0.967, 1.043)	0.998 (0.901, 1.106)
% Smokers	1.003 (0.991, 1.014)	1.004 (0.993, 1.015)	0.990 (0.976, 1.004)~	0.983 (0.966, 1.000)~	0.998 (0.933, 1.068)
% 65 and Older	1.021 (1.008, 1.034)*	1.005 (0.997, 1.014)~	0.992 (0.976, 1.009)	1.010 (0.993, 1.027)	0.999 (0.942, 1.060)
% AA Population	0.995 (0.992, 0.998)*	1.001 (1.000, 1.003)~	1.008 (0.984, 1.033)	0.996 (0.988, 1.005)	1.000 (0.975, 1.025)
Testing ^A	1.094 (1.089, 1.099)*	0.957 (0.944, 0.970)*	0.959 (0.943, 0.976)*	1.000 (1.000, 1.001)	0.999 (0.998, 0.999)*

^A Percent of positive tests was used for the Midwest and Northeast while log count of positive tests was used for all other models

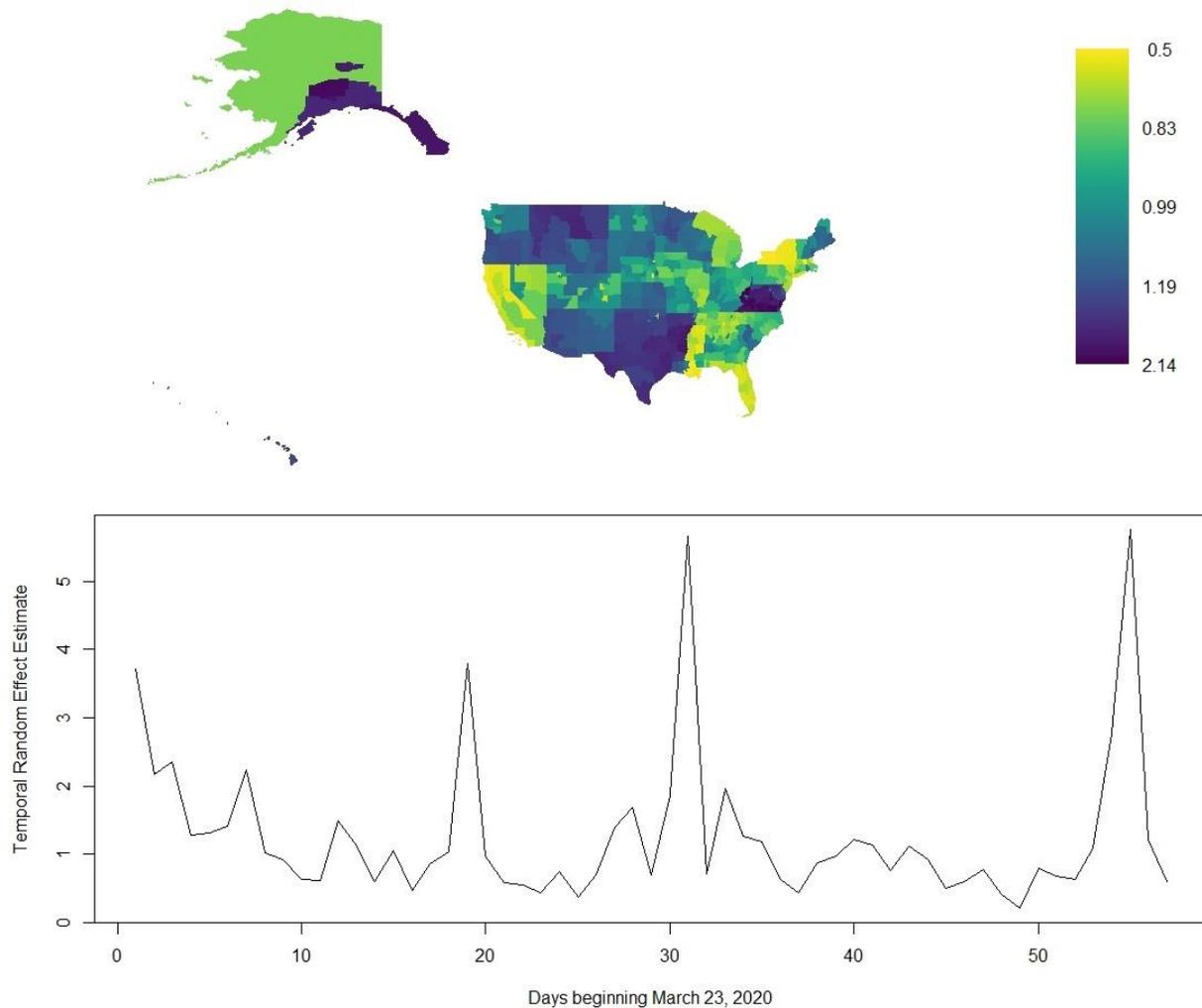
^B There is no estimate associated with the none category of the stay at home orders variable for the NE model because all states in the northeast issued a stay at home orders.

* Indicates well estimated Bayesian inference at the 0.05 level

~ Indicates borderline Bayesian inference

Figure 2 displays the spatial and temporal random effect estimates for confirmed cases across the United States. These visuals display the residual variation in the outcome after adjusting for the fixed effects in Table 1. Hence, increased (estimates greater than 1) or decreased (estimates less than 1) risk could arise from reasons pertaining to social distancing adherence or alternate population characteristics not considered in this study. State borders are largely well defined here indicating consistency in the residual within states. After adjustment, many of the health regions with increased risk include parts of Idaho, Louisiana, Montana, and Alaska, Texas, Oklahoma, Arkansas, West Virginia, Virginia, Maryland, Delaware, and Hawaii. Alternatively, areas of decreased risk include parts of Louisiana and Mississippi, California, Florida, and New York. Finally, the temporal random effect parameter largely accounts for delays and errors in the data (Jordan 2020). There are three sharp peaks at days 19 (April 10), 31 (April 22), and 55 (May 17).

Figure 2: Spatial and temporal random effect estimates for the confirmed cases outcome considering all health regions in the United States.



3. Discussion

The results presented here demonstrate the usefulness of these spatio-temporal statistical models for identifying several important risk factors for a range of COVID-19 outcomes. The use of Bayesian inference is both relevant to this inquiry and valuable as a benchmark and template for researchers to use for additional applications in data science. Additionally, the dashboards offer an ideal interactive environment for displaying these results for a range of interested stakeholders.

The two state-level spatio-temporal variables were among the most important fixed effects in the model, despite their different roles. The stay at home orders variable provided important predictive information to describe differences in risk for health regions under the four levels of the variable, while the testing variable served as a control variable for the number and positive rate of tests performed. For the stay at home orders predictor, all levels were well estimated for nearly all models, and there was general consistency across the models with respect to the direction of association, particularly when excluding the West models. Interestingly, these estimates show that there is generally less risk for health regions under full stay at home orders and health regions that were never under stay at home orders, and more risk for health regions under partial stay at home orders when compared to health regions before stay at home orders were implemented. It is possible that governors in states with full stay at home orders are appropriately maintaining these orders where health regions are at greater risk to maintain control of disease spread within their borders. For states with partial stay at home orders the findings suggest they may have prematurely removed restrictions or, perhaps it was a calculated risk and governors surmised that the hospital systems were prepared to accommodate the potential influx of cases. Less risk in states with partial or full stay at home orders could indicate that stay at home orders were successful and removal could be considered. Similarly, less risk for health regions that were never under stay at home orders could suggest that these governors made the appropriate decision for their states.

When considering results pertaining to testing, it is important to remember that these estimates do not suggest increased risk with more testing as a cause and effect relationship – merely that more comprehensive testing will necessarily increase the numbers of confirmed cases. Indeed, greater testing gives a better representation of the true disease presence in the country. Our examination time window and spatial aggregation may not be appropriate to examine the relationships between testing and the death outcomes. Examinations of these data in a recently published study suggest that there is a lag time of up to eight weeks between testing and death, potentially depending on the age of the individual (Testa et al. n.d.).

Other notable results uncovered here include less risk for health regions with more unemployment, more risk for health regions with more Black or African American population in the South, and less risk of new cases but higher risk of death for health regions with more older population. It is possible that with more unemployment, the frequency and volume of social interactions are diminished resulting in lower levels of disease transmission. Recall that these are unemployment estimates for March 2020, so health regions with more unemployment may also be health regions where more individuals were laid off or furloughed during the crisis. The finding that there is greater risk for health regions with more Black or African American population in the South aligns with anecdotal reporting and other findings regarding the disproportionate effects of this crisis on minority populations (Tai et al. 2020). Incidentally, the South is an ideal place for the examination of minority populations because the presence is greater there than in other parts of the county. Finally, the age-related results – for regions with

populations that have more residents aged 65 and older – suggest that older individuals adhere better to the social distancing standards but are still more susceptible to death, findings that reaffirm results in other studies (Bialek et al. 2020; Onder et al. 2020).

Although the fixed effect estimates were quite similar across models, the spatial and temporal random effects were different when the US Census-defined regions were compared to those generated for the entire United States. Some of these differences are a consequence of scaling the random effects in the different models, but several indicators suggest an opposite risk direction. This finding indicates that there are differences in the COVID-19 crisis across the country; thus, it is critical to consider these areas on their own. The US Census-defined region with the most consistent spatial random effect distribution to the one estimated for the United States as a whole was the Northeast. It is possible that the Northeast was driving much of the US-based model because the outbreak was centralized in that region for most of this time period. The West US Census-defined region was the most inconsistent with respect to the fixed effect estimates, and this is likely due to two factors: 1) the geographic and cultural heterogeneity within that region and 2) the strict nature of the stay at home orders in both California and Washington.

The statistical model implemented in this study has both advantages and disadvantages for understanding disease spread. First, this model appears robust to data delays that are inherent in real-time data (Jordan 2020). For example, there is a strong seasonal trend in the data where more new cases are reported later in the week versus earlier in

the week (see [this dashboard](#) for evidence). However, that same trend is not evident in the temporal random effect estimates and we observed no well estimated parameters when including a variable that attempted to adjust for this trend. The reason for the models' robustness lies in the expected count (e_{ij}) term, since that matrix by definition includes the same seasonal trend. One negative feature for these models involves the computational efficiency. While implementation via the INLA package in R speeds up the model fitting process, these complex models become more computationally inefficient with every day of new data generation. As such, real-time, or daily data as used here, updates become less feasible the longer the crisis continues. Our current real-time rendering of these models uses weeks as the temporal unit for this reason.

4. Conclusion

In sum, this article demonstrates how statistical models of this nature and caliber can illuminate important relationships with disease spread and identify relevant population-level characteristics. Most importantly, we believe that these results confirm the benefit of state-issued stay at home orders as well as suggest compliance with respect to the directives towards the older population for adhering to social distancing guidelines.

5. Materials and methods

5.1. Data

Although most of the spatial data employed in this analysis are available at the county level, spatio-temporal statistical modeling of these data at this aggregation is complicated for two major reasons: 1) zero inflation and 2) large dimension. Both of these complications would render our methods inappropriate and inefficient. In order to avoid the issues associated with county-level modeling, we used a broader spatial

aggregation – health regions, which are one or more counties serviced by the same health department (Day et al. 2019). By adopting health regions, the zero count of confirmed cases (as of May 18, 2020) dropped from 33,206 to 302 and the spatial dimension shrank from 3,142 (number of US counties) to 389 (number of US health regions). Correspondingly, we aggregated all county-level predictors and outcomes to the health region level. There is no evidence of loss of power with the aggregation of these data to the health region level (Day et al. 2019).

5.1.1. Outcome measures

Outcome data were sourced from the Johns Hopkins University Center for Systems Science and Engineering (Johns Hopkins University Center for Systems Science and Engineering 2019). Data for this study were geographically constrained to the United States and confined to the first eight weeks of recorded data (March 23, 2020 to May 18, 2020). Limiting the window of time to the first eight weeks was necessary and appropriate for three reasons. First, doing so guaranteed good representation in the two temporal varying predictors (state-issued stay at home orders and testing prevalence), particularly for the timeline pertaining to stay at home orders since several states had partially reopened but none had fully reopened. Second, by constraining the window of time it afforded the ability to examine data on a daily temporal scale as opposed to the real time models now that show data in weeks for computational efficiency. Finally, confining the data in this way allowed for the inclusion of testing information, which lags behind case counts in data release.

We examined four outcomes: number of confirmed cases, number of active cases, number of new cases, and number of deaths. Recovery data is not captured at the US county level and therefore is not incorporated as a stand-alone outcome or in the calculation of active cases. Additionally, there could be differences in covariate relationships based on the region of the country considered. As such, we made an estimation for the US as a whole and apportioned the numbers into the following regions as defined by the US Census Bureau (Geography Division. U.S. Department of Commerce Economics and Statistics Administration. U.S. Census Bureau. 1984): South (S), West (W), Midwest (MW), and Northeast (NE).

5.1.2. Covariates for adjustment

All models adjust for the same neighborhood-level covariates, which were selected based on two criteria: to describe the state-level responses to the pandemic and to represent population level characteristics that may contribute to the disparities in outcome counts between health regions. We elaborate on the selection of predictors in the paragraphs below. The statistical model indirectly adjusts for differences in population across the observed areas (see Section 2.2 below for more details). Images (gif or png) for all predictors considered are available for review in the supplemental files.

To understand and adjust for the state-level response to COVID-19, we considered two important temporal predictors, the state-by-state issuance of stay at home orders and the rate of testing. These variables are specified at the state-level because they are either mandated at the state-level or are only available at that level of aggregation,

respectively. The self-constructed predictor representing presence of statewide stay at home orders required daily updates when running in real-time. The most current, convenient, and reliable information for this categorical variable (levels: before, during – partial, during – full, and none) came from the New York Times (Mervosh et al. 2020). To ensure accuracy, we concurrently examined government mandates for more information. To construct the testing variable, we produced either a log transformation of positive test count or a percent of positive tests variable. The variable varied by US Census-defined region (US, S, W, MW, and NE) depending on what appeared best fitting according to goodness of fit criterion and most appropriate in terms of estimate directionality as a control variable (The Atlantic 2020). The region-specific testing variable was consistent across outcomes within that region.

Population characteristics were also included as predictors in this model, specifically the presence of an international airport, land area of the health region, percent of the population that is unemployed, percent of the population that is a smoker, percent of the population that is aged 65 and older, and percent of the population that is Black or African American. The binary variable for presence of an international airport was used to represent community connectivity to international locations, which could be an important driver of disease progression particularly in this time window at the start of the pandemic. This predictor was created by combining information on airport locations from the U.S. Bureau of Transportation Statistics (U.S. Bureau of Transportation Statistics 2009; U.S. Transportation Department 2019) and international flights for each of those airports from the US Department of Transportation.

Whereas urban areas typically comprise small health regions, rural areas are found in larger health regions. Size of a health region necessarily influences important factors like access to care, access to quality care, and capacity of health providers. Therefore, we adjusted for urbanicity by including information on the land area of each observation. Unemployment was estimated as the percent of the population that was unemployed in March 2020 in each health region using information from the Bureau of Labor Statistics (U.S. Bureau of Labor Statistics 2020). The rest of the population demographic variables – smoking, age, and race – are 2018 estimates from the US Census Bureau and were included as proxies for comorbidities and vulnerable populations (Population Estimates Program 2018; Zigler 2017).

5.2. Statistical model

The statistical model applied here is commonly used in disease mapping for aggregated count outcomes, and is a variation of a Bayesian spatio-temporal Poisson Knorr-Held model (Knorr-Held 2000; Knorr-Held and Besag 1998; Lawson 2013; Lawson et al. 2016; Lawson and Lee 2017; Lesaffre and Lawson 2013). This model can be defined as follows for health region i and day j :

$$y_{ij} \sim \text{Pois}(\mu_{ij})$$

$$\mu_{ij} = e_{ij}\theta_{ij}$$

$$\log(\theta_{ij}) = X'_{ij}\boldsymbol{\beta} + u_i + \gamma_j$$

where y_{ij} is the outcome of interest (count of confirmed cases, active cases, new cases, or deaths), μ_{ij} is the mean of the Poisson model, e_{ij} is the expected count, θ_{ij} is the relative risk, X'_{ij} is the design matrix for the predictors, $\boldsymbol{\beta}$ is the vector of fixed effect

estimates, u_i is the spatial random effect, and γ_j is the temporal random effect. e_{ij} is calculated as the rate of infection across all health regions on a given day times the population at risk for a given health region, which is assumed constant over time. As such, a unique e_{ij} is produced for each health region and day. Given the Bayesian methodology applied here, prior distributions were required for all parameters and they were defined as follows: $\beta_k \sim N(0, \tau_\beta^{-1})$ for each of the k fixed effect parameter estimates, $u_i \sim N(0, \tau_u^{-1})$ for an uncorrelated spatial random effect, $\gamma_j \sim N(\gamma_{j-1}, \tau_\gamma^{-1})$ for a temporal random walk effect, and all precisions were such that $\tau \sim \text{Gam}(2, 1)$. All these prior distributions are considered non- or weakly-informative and sensitivity analysis (data not shown) suggests little influence on the resulting parameter estimates. This model description applies to all outcomes (confirmed cases, active cases, new cases, and deaths) and US Census-defined regions (US, S, W, MW, and NE) considered.

5.3. Computational Details

R statistical software furnished much of the means for data processing and analysis in the work presented here. Specifically, the R packages `rgdal`, `INLA`, and `fillmap` were necessary for spatial data processing, statistical modeling, and spatial plotting respectively (Bivand et al. 2019; Blangiardo et al. 2013; Carroll 2016; Carroll et al. 2015; Martins et al. 2013; R Core Team 2015; Rue et al. 2009; Schrödle and Held 2010, 2011; Ugarte et al. 2014). We, along with several others from the UNCW Data Science program, produced data visualization dashboards with an array of popular software including R Shiny, Tableau, and Power BI (Chang et al. n.d.; Microsoft Corporation 2014; Salesforce 2003). In addition to the regression model results, these dashboards also display general spatial, temporal, and spatio-temporal tracking of COVID-19 cases

and deaths at the county, state, national, and international levels. Code for the statistical models and Shiny apps are available at [this GitHub repository](#).

Acknowledgements

[removed for review]

References

- Bialek, S., Boundy, E., Bowen, V., Chow, N., Cohn, A., Dowling, N., Ellington, S., Gierke, R., Hall, A., MacNeil, J., Patel, P., Peacock, G., Pilishvili, T., Razzaghi, H., Reed, N., Ritchey, M., and Sauber-Schatz, E. (2020), “Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020,” *MMWR. Morbidity and Mortality Weekly Report*, 69, 343–346. <https://doi.org/10.15585/mmwr.mm6912e2>.
- Bivand, R., Rowlingson, B., and Tim, K. (2019), “rgdal: Bindings for the ‘Geospatial’ Data.”
- Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013), “Spatial and spatio-temporal models with R-INLA,” *Spat Spatiotemporal Epidemiol*, 4, 33–49. <https://doi.org/10.1016/j.sste.2012.12.001>.
- Carroll, R. (2016), “fillmap: Create maps with SpatialPolygons objects.”
- Carroll, R., Lawson, A. B., Faes, C., Kirby, R. S., Aregay, M., and Watjou, K. (2015), “Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping,” *Spat Spatiotemporal Epidemiol*, 14–15, 45–54. <https://doi.org/10.1016/j.sste.2015.08.001>.

- Chang, W. ., Cheng, J. ., Allaire, J. J. ., Xie, Y. ., and McPherson, J. . (n.d.). “shiny: Web application framework for R.”
- Day, K., Carroll, R., and Zhao, S. (2019), “A comparison of spatial resolution in North Carolina,” *High School Journal of Mathematics*, In print.
- Geography Division. U.S. Department of Commerce Economics and Statistics Administration. U.S. Census Bureau. (1984), “Census regions and divisions of the United States,” Available at https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.
- Johns Hopkins University Center for Systems Science and Engineering (2019), “2019 Novel Coronavirus Visual Dashboard.”
- Jordan, M. I. (2020), “On Identifying and Mitigating Bias in the Estimation of the COVID-19 Case Fatality Rate,” *Harvard Data Science Review*.
<https://doi.org/10.1162/99608f92.f01ee285>.
- Knorr-Held, L. (2000), “Bayesian modeling of inseperable space-time variation in disease risk.,” *Stat Med*, 19, 2555–2567.
- Knorr-Held, L., and Besag, J. (1998), “Modelling risk from a disease in time and space,” *Stat Med*, 17, 2045–2060. [https://doi.org/10.1002/\(sici\)1097-0258\(19980930\)17:18<2045::aid-sim943>3.0.co;2-p](https://doi.org/10.1002/(sici)1097-0258(19980930)17:18<2045::aid-sim943>3.0.co;2-p).
- Lawson, A. B. (2013), *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*, Boca Raton, FL: CRC Press.
- Lawson, A. B., Banerjee, S., Haining, R., and Ugarte, M. D. (2016), *Handbook of Spatial Epidemiology*, Boca Raton, FL: CRC Press.
- Lawson, A., and Lee, D. (2017), “Bayesian Disease Mapping for Public Health,” pp.

443–481. <https://doi.org/10.1016/bs.host.2017.05.001>.

Lesaffre, E., and Lawson, A. B. (2013), *Bayesian Biostatistics*, West Sussex, U.K.:

Wiley. <https://doi.org/10.1002/978-1-119-94241-2>.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013), “Bayesian computing with INLA: New features,” *Comput Stat Data An*, 67, 68–83.

<https://doi.org/10.1016/j.csda.2013.04.014>.

Mervosh, S., Lee, J. C., Gamio, L., and Popovich, N. (2020), “See Which States Are Reopening and Which Are Still Shut Down,” *The New York Times*, Available at See Which States Are Reopening and Which Are Still Shut Down.

Microsoft Corporation (2014), “Power BI.”

Onder, G., Rezza, G., and Brusaferro, S. (2020), “Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy,” *JAMA*.

<https://doi.org/10.1001/jama.2020.4683>.

Population Estimates Program (2018), “Population and Housing Estimates,” US Census Bureau.

R Core Team (2015), “R: A language and environment for statistical computing.,” *R Foundation for Statistical Computing*, Vienna, Austria. .

Rue, H., Martino, S., and Chopin, N. (2009), “Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion),” *J R Stat Soc B*, 71, 319–392.

Salesforce (2003), “Tableau Software.”

Schrödle, B., and Held, L. (2010), “A primer on disease mapping and ecological regression using INLA,” *Computation Stat*, 26, 241–258.

<https://doi.org/10.1007/s00180-010-0208-2>.

Schrödle, B., and Held, L. (2011), “Spatio-temporal disease mapping using INLA,” *Environmetrics*, 22, 725–734. <https://doi.org/10.1002/env.1065>.

Tai, D. B. G., Shah, A., Doubeni, C. A., Sia, I. G., and Wieland, M. L. (2020), “The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States,” *Clinical Infectious Diseases*. <https://doi.org/10.1093/cid/ciaa815>.

Testa, C. C., Krieger, N., Chen, J. T., and Hanage, W. P. (n.d.). “Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22, 2020 – July 8, 2020),” *HCPDS working paper*, 10.

The Atlantic (2020), “The COVID Tracking Project.,” *CC BY-NC-4.0 license*, Available at <https://covidtracking.com/>.

Trombetta, H., Faggion, H. Z., Leotte, J., Nogueira, M. B., Vidal, L. R. R., and Raboni, S. M. (2016), “Human coronavirus and severe acute respiratory infection in Southern Brazil,” *Pathogens and Global Health*, 110, 113–118. <https://doi.org/10.1080/20477724.2016.1181294>.

U.S. Bureau of Labor Statistics (2020), “Unemployment Statistics.”

U.S. Bureau of Transportation Statistics (2009), “US Airports.”

U.S. Transportation Department (2019), “International Report Passengers.”

Ugarte, M. D., Adin, A., Goicoa, T., and Militino, A. F. (2014), “On fitting spatio-temporal disease mapping models using approximate Bayesian inference,” *Stat Methods Med Res*, 23, 507–530. <https://doi.org/10.1177/0962280214527528>.

Zigler, C. (2017), “County-Level Smoking Data,” Harvard Dataverse.

<https://doi.org/10.7910/DVN/VZ21KD>.

Figures

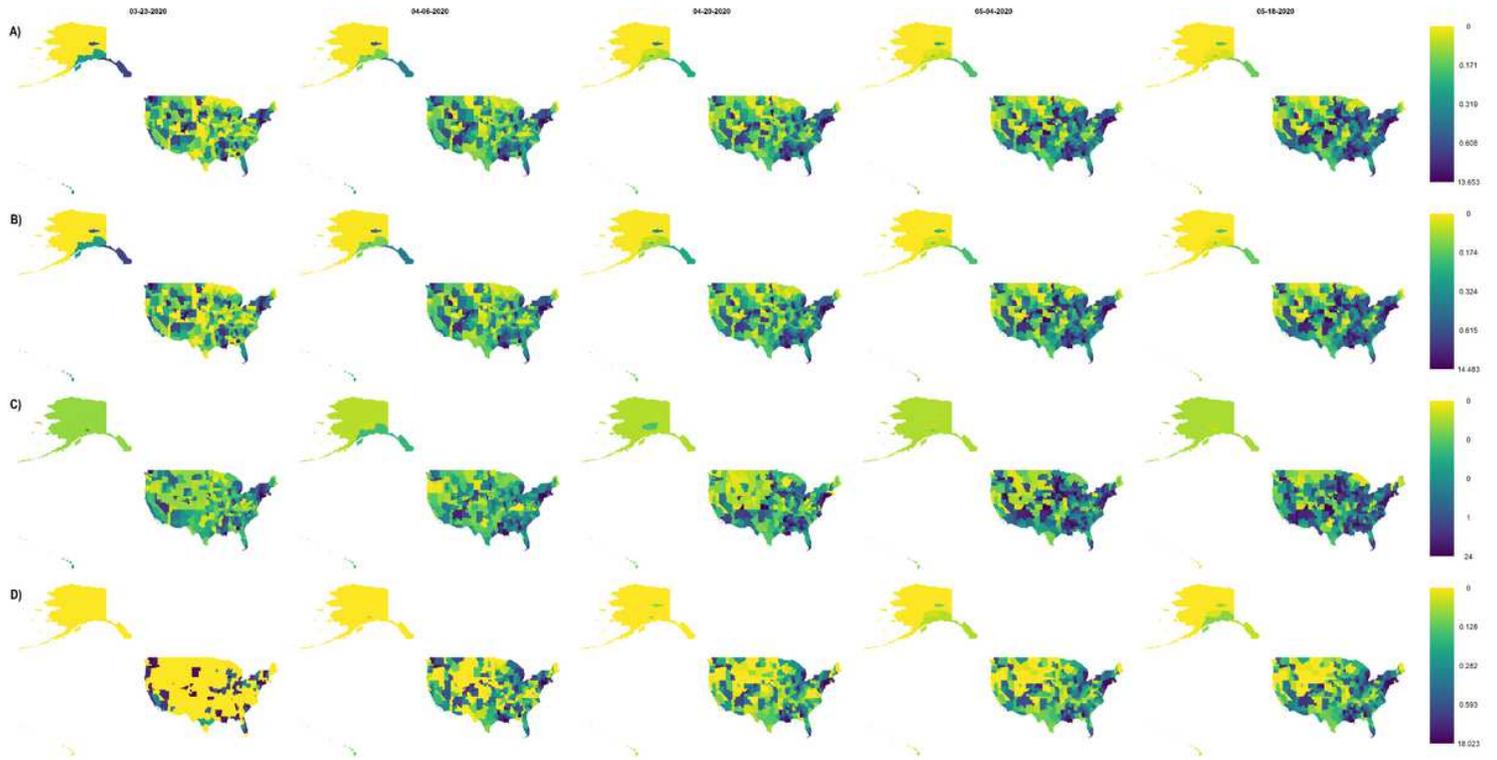


Figure 1

Standardized incidence or mortality ratio for A) confirmed cases, B) active cases, C) new cases, and D) deaths for the first, third, fifth, seventh, and ninth Mondays across all health regions in the United States.

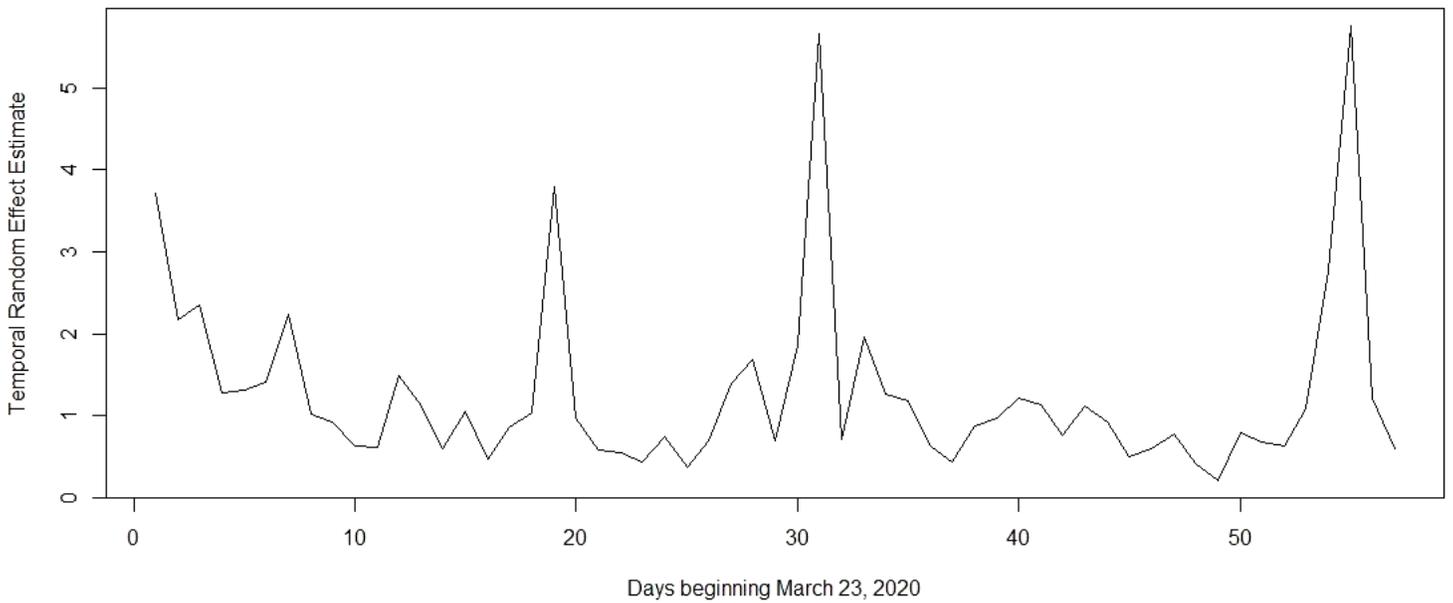
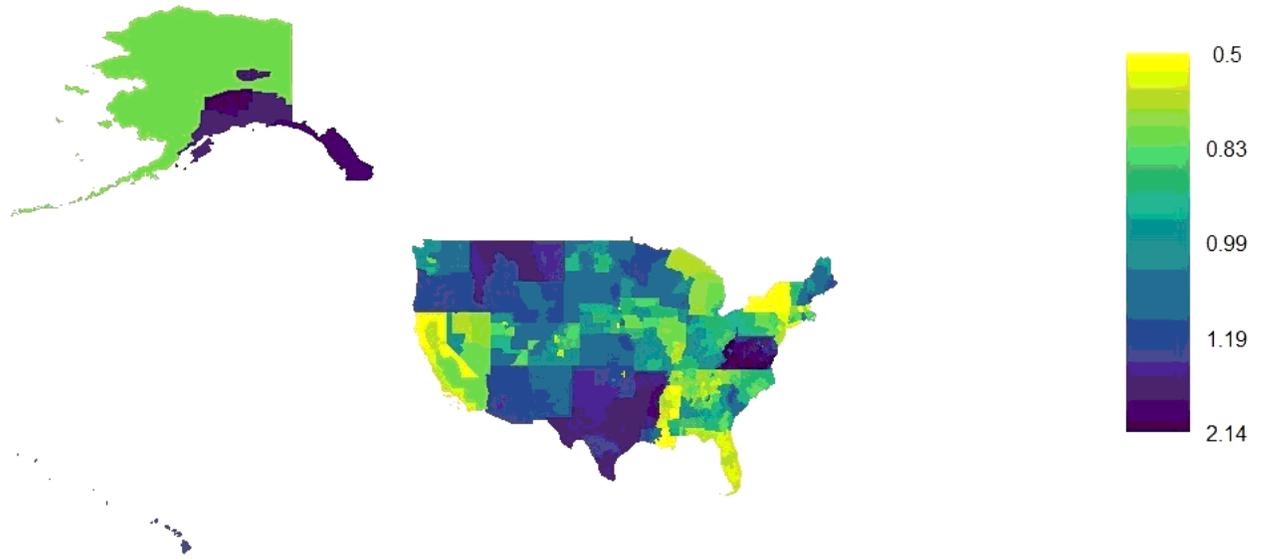


Figure 2

Spatial and temporal random effect estimates for the confirmed cases outcome considering all health regions in the United States.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.7z](#)