



OPEN

Using spatial and temporal modeling to visualize the effects of U.S. state issued stay at home orders on COVID-19

Rachel Carroll¹✉ & Christopher R. Prentice²

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 analysts and data scientists in the policymaking community 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.

The coronavirus disease 2019 (COVID-19) began in Wuhan, China in December of 2019 and quickly evolved into a global crisis. 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 alone¹. 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 publicly 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 both regionally and nationally². Data visualization dashboards 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. Supporting code and links to these dashboards as well as an array of related visualizations created by colleagues in the UNCW Data Science program can be found 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 Engineering¹, population and demographic information from the US 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 Statistics³⁻⁶. The stay at home orders variable was self-constructed with guidance from government mandates and the New York Times⁷. 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 variation⁸⁻¹⁰. 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.

¹Department of Mathematics and Statistics, University of North Carolina Wilmington, Wilmington, NC 28403, USA. ²Department of Public and International Affairs, University of North Carolina Wilmington, Wilmington, NC 28403, USA. ✉email: carrollr@uncw.edu



Figure 1. A combination of the stay at home orders variable and standardized incidence ratios by health region for the confirmed cases outcome. (A) Before stay at home orders, (B) During full stay at home orders, (C) During partial stay at home orders, and (D) Never stay at home orders. Health regions are only present in panes where the stay at home orders category applied.

Results

An R Shiny COVID-19 regression dashboard with the results for the data presented in this manuscript is available here. Data and analysis are constrained to focus on the March 23, 2020 to May 18, 2020 window.

Figure 1 displays a combined visualization of the stay at home orders and a calculation of the median standardized incidence ratio (y_{ij}/e_{ij}) of confirmed cases for each health region across the days that they were in the given stay at home orders category (before, during—full, during—partial, and never). The same image is available for the deaths outcome in the Supplemental Figures as SAH SMR.png. Figure 1 illustrates the areas of the country where the different levels of the stay at home order exist, and suggests higher risk for areas under partial stay at home orders (darker shading indicates more risk). Medians of these medians represent risk across each category of stay at home orders, and are 0.23 (before stay at home orders), 0.33 (during full stay at home orders), 0.41 (during partial stay at home orders), and 0.25 (never stay at home orders). Results suggest most risk under partial stay at home orders.

An animation of the state-level stay at home orders variable (our main covariate of interest) and the individual images used to create the animation are provided in the supplemental materials as sah.gif and “SAH” (folder), respectively. This animation shows most of the country shut down and then gradually begin to reopen over the time considered in this study. Several states never issued statewide stay at home orders variables—Arkansas, Iowa, Nebraska, North Dakota, Oklahoma, South Dakota, Utah, and Wyoming. The Supplemental Figures in the “SIR SMR” folder display the standardized incidence or mortality ratios for confirmed cases, active cases, new cases, and deaths for each day across all health regions in the United States. Standardized incidence and mortality ratios 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. Despite this relative consistency, there is a slight increase in risk, particularly with respect to death, over time.

Table 1 displays the statistical model fixed effect 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 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 and Supplemental Table 1). 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 (darkest gray highlighted estimates). Specifically, the estimate from the South confirmed cases outcome model suggest a 5.2% higher risk in health regions under partial stay at home orders and a 5.0% lower risk in health regions under full stay at home orders. Risk appears to decrease as the percent unemployment in a region rises (medium gray highlighted estimates); the estimate suggests that for each one percent increase in unemployment, the risk of confirmed cases increases by 2.2% in the South confirmed cases outcome model. Similarly, one finding shows

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)*
Never	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	1.000	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	1.000	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)*
Never	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)*

Table 1. Fixed effect parameter estimates for the confirmed cases and deaths outcomes represented as inverse logarithm mean estimates and associated 95% credible intervals. ^APercent of positive tests was used for the Midwest and Northeast while log count of positive tests was used for all other models. ^BThere is no estimate associated with the never 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. Gray highlights indicate common estimate patterns mentioned in the text. Figure 2 displays the spatial and temporal random effect estimates for confirmed cases across the United States. These visuals display the residual spatial and temporal variation in the outcome after adjusting for the fixed effects in this table. 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. In the spatial random effect, 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 data¹¹. There are three sharp peaks at days 19 (April 10), 31 (April 22), and 55 (May 17).

risk increases in regions with higher percentages of Black or African American residents, and this finding is particularly noticeable in the South region. Some evidence exists that risk is also higher where there is a higher percent of smokers and a higher percent older population (light gray highlighted estimates). 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. Finally, all models include the same parameters for comparison, and they are well estimated at the 0.05 or 0.10 level for at least one model indicating an important relationship between that variable and the COVID-19 outcomes.

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 for all models except, notably, the West model. 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

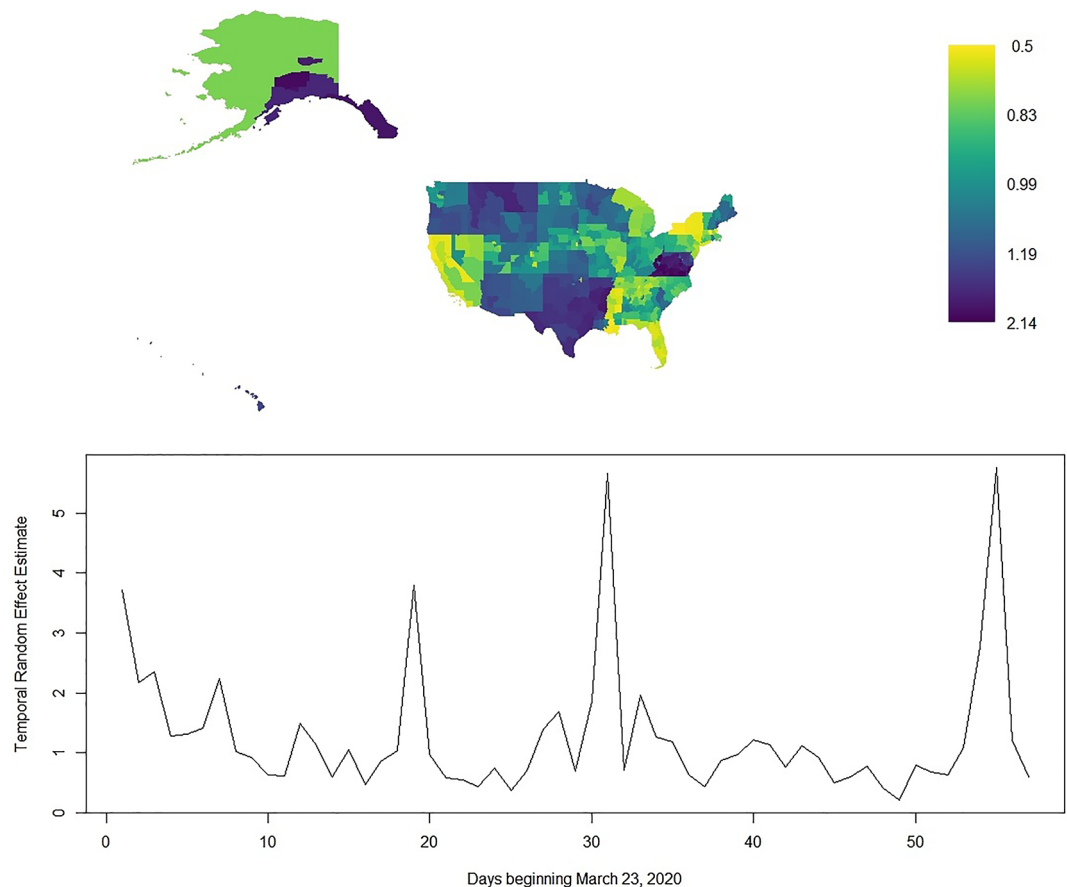


Figure 2. Spatial and temporal random effect estimates for the confirmed cases outcome considering all health regions in the United States.

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 individual¹².

Other notable results uncovered here include decreased risk for health regions with more unemployment, increased risk for health regions as percent of Black or African American population increases in the South, and lower 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 populations¹³. 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 country. 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^{14,15}.

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. We are not able to ascertain from these findings the exact cause of the disjuncture, but present two possible factors that could be explored in future academic inquiry: (1) the geographic and cultural heterogeneity within the West region and (2) the nature and duration of the stay at home orders in both California and Washington that appear different from other regions.

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 data¹¹. 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. 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 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. Finally, this is not a causal model, and we cannot determine the causes of higher or lower risk of COVID-19. Rather, we observe common characteristics among areas with higher or lower risk of COVID-19. While we include socio-demographic covariates, we cannot make claims of social determinants of health from our parameter estimates.

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.

Materials and methods

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 department¹⁶. 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 level¹⁶.

Outcome measures. Outcome data were sourced from the Johns Hopkins University Center for Systems Science and Engineering¹. 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 Bureau¹⁷: South (S), West (W), Midwest (MW), and Northeast (NE).

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 the [Statistical model](#) section 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 never) came from the New York Times⁷. 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 variable¹⁸. 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^{4,5} 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 Statistics⁶. 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^{3,19}.

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^{8–10,20–22}. This model can be defined as follows for health region i and day j :

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

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

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

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, β 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.

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^{23–32}. 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 BI². 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.

Received: 25 February 2021; Accepted: 23 June 2021

Published online: 06 July 2021

References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. *shiny: Web Application Framework for R*.
3. Population Estimates Program. Population and Housing Estimates. (2018).
4. U.S. Bureau of Transportation Statistics. US Airports. (2009).

5. U.S. Transportation Department. International Report Passengers. (2019).
6. U.S. Bureau of Labor Statistics. Unemployment Statistics. (2020).
7. Mervosh, S., Lee, J. C., Gamio, L. & Popovich, N. See Which States Are Reopening and Which Are Still Shut Down. *The New York Times* (2020). See Which States Are Reopening and Which Are Still Shut Down. Accessed 20 May 2020.
8. Lesaffre, E. & Lawson, A. B. *Bayesian Biostatistics*. (Wiley, 2013). <https://doi.org/10.1002/978-1-119-94241-2>
9. Lawson, A. B., Banerjee, S., Haining, R. & Ugarte, M. D. *Handbook of Spatial Epidemiology* (CRC Press, 2016).
10. Lawson, A. & Lee, D. Bayesian disease mapping for public health. *Harvard Data Sci. Rev.* **36**, 443–481 (2017).
11. Jordan, M. I. On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Sci. Rev.* <https://doi.org/10.1162/99608f92.f01ee285> (2020).
12. Testa, C. C., Krieger, N., Chen, J. T. & Hanage, W. P. Visualizing the lagged connection between COVID-19 cases and deaths in the United States: An animation using per capita state-level data (January 22–July 8, 2020). *HCPDS Work. Pap.* **10** (2020).
13. Tai, D. B. G., Shah, A., Doubeni, C. A., Sia, I. G. & Wieland, M. L. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciaa815> (2020).
14. Bialek, S. *et al.* Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **69**, 343–346 (2020).
15. Onder, G., Rezza, G. & Brusaferro, S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* <https://doi.org/10.1001/jama.2020.4683> (2020).
16. Day, K., Carroll, R. & Zhao, S. A comparison of spatial resolution in North Carolina. *High Sch. J. Math.* In print (2019).
17. Geography Division. U.S. Department of Commerce Economics and Statistics Administration. U.S. Census Bureau. Census regions and divisions of the United States. (1984). https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.
18. The Atlantic. The COVID Tracking Project. *CC BY-NC-4.0 license* (2020). <https://covidtracking.com/>.
19. Zigler, C. County-Level Smoking Data. V1 (2017). <https://doi.org/10.7910/DVN/VZ21KD>
20. Knorr-Held, L. Bayesian modeling of inseparable space-time variation in disease risk. *Stat. Med.* **19**, 2555–2567 (2000).
21. Knorr-Held, L. & Besag, J. Modelling risk from a disease in time and space. *Stat. Med.* **17**, 2045–2060 (1998).
22. Lawson, A. B. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology* (CRC Press, 2013).
23. Bivand, R., Rowlingson, B. & Tim, K. *rgdal: Bindings for the 'Geospatial' Data*. (2019).
24. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. Spatial and spatio-temporal models with R-INLA. *Spat. Spat. Epidemiol.* **4**, 33–49 (2013).
25. Ugarte, M. D., Adin, A., Goicoa, T. & Militino, A. F. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat. Methods Med. Res.* **23**, 507–530 (2014).
26. Martins, T. G., Simpson, D., Lindgren, F. & Rue, H. Bayesian computing with INLA: New features. *Comput. Stat. Data An.* **67**, 68–83 (2013).
27. Schrödle, B. & Held, L. A primer on disease mapping and ecological regression using INLA. *Comput. Stat.* **26**, 241–258 (2010).
28. Schrödle, B. & Held, L. Spatio-temporal disease mapping using INLA. *Environmetrics* **22**, 725–734 (2011).
29. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. R. Stat. Soc. B* **71**, 319–392 (2009).
30. Carroll, R. fillmap: Create maps with SpatialPolygons objects. (2016).
31. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2015).
32. Carroll, R. *et al.* Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spat. Spat. Epidemiol.* **14–15**, 45–54 (2015).

Acknowledgements

This research was supported in part by the University of North Carolina Wilmington's Center for Social Impact with special thanks to Cape Fear Collective.

Author contributions

R.C. performed the statistical analysis. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93433-z>.

Correspondence and requests for materials should be addressed to R.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021