Spatial Analyses of State Level COVID-19 Mortality and Chronic Disease Covariates in the United States

Avery H. Richards

School of Public Health, University of California Berkeley

PHW272C: Applied Spatial Data Science for Public Health

Dr. Hugh Sturrock

May 2nd, 2021

Background:

In the year 2020, United States experienced a great deal of excess mortality attributable to the SARS-COV-2 virus and COVID-19 (US CDC, 2021). The rapid onset and growth of COVID-19 case mortality over this period prompted a frenzy of research and inquiry into the etiology of COVID-19 mortality cases. Meta-analyses of COVID-19 disease severity outline social determinants of health impacting outcomes (Burstrom & Tao, 2020), as well as how "*preexisting chronic diseases were strongly correlated with an increased admittance to ICU*." (Liu et al. 2020). Along with social determinants being implicated, co-morbidity with pre-existing health condition and chronic disease were associated as risk factors in the COVID-19 related mortality (Harrison et al, 2020).

In this paper I walk through a computationally intensive analysis that seeks to compare crude state level COVID-19 mortality counts with the prevalence of two chronic disease covariates: state level prevalence ratio values of diabetes and asthma. I hypothesized an autocorrelative spatial relationship exists between one (or both) of these disease prevalence variables as they are related to the state-level aggregate counts of COVID-19 mortality cases. Once I established a global measure of spatial autocorrelation with crude mortality data via a global Moran's I test, I ran two General Linear Models (GLM) to explore statistical significance between the mortality counts and disease prevalence coefficients without a spatial interaction term. After isolating a significant covariate in the GLM analyses, I tested the model's deviance residuals for evidence of global spatial autocorrelation. I then attempted to adjust for spatial autocorrelation by implementing a Conditional Autoregressive Model (CAR) on statistically significant covariates, or "covariate" in this case. I ran a global Moran's I again, using the standard deviation residuals of the best fit model and evaluated the outputs to determine if autocorrelation could be statistically detectable in the CAR model's fitted values post-hoc. All computational resources were accessible online via The Comprehensive R Archive Network (CRAN, 2021).

Methods:

Initial data was sourced from the John Hopkins Center for Systems Science and Engineering COVID-19 Data Repository (John Hopkins, 2021). The dataset included confirmed, active, and recovered cases, test result counts, and state population. Each variable was aggregated to state-level observations with corresponding centroid spatial coordinates, with a sample size relating to each state in the contiguous United States, ($n = 49$). After importing this dataset into a local computing environment, I ran calculations determine a crude COVID-19 mortality count per 10,000 persons. The denominator was scaled to allow for frequency counts with an appropriate general linear model (GLM) formula.

Most recent state level chronic disease prevalence ratios from were collected from public facing US government websites. Diabetes prevalence data was collected from the State of Childhood Obesity website (SCO, 2021), and asthma prevalence data was collected from the US Center for Disease Control website (CDC, 2021). These two variables were joined to the initial John Hopkins dataset through matching by state names on each dataset. Once a final dataset was assembled with the included covariates, I conducted an initial evaluation of global clustering in

the state-level with the COVID-19 mortality counts alone. To do this, I created a "queen" contiguity neighbor matrix of all 49 areal state polygons, as seen below in figure 1.

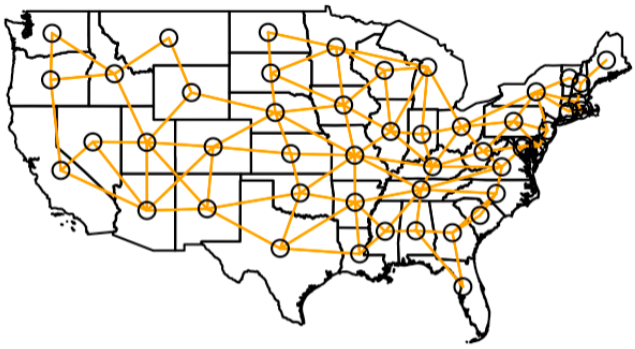

Figure 1*: Queen Neighbor Contiguity Structure, United States*

I then extracted coordinate values from each state polygon and created a binary spatial weights coding scheme to run a Moran's I test, which passed the adjusted mortality counts per state polygon within the weighting structure. Results from this test indicated a failure to reject the alternative hypothesis of positive global autocorrelation present with a p-value of Pr(>|z|) = 5.644e-05.

After global spatial autocorrelation was established in the present data, I sought to explore relationships between chronic disease prevalence variables and population adjusted mortality counts with multivariate regression models. To do this, I assembled a Poisson family General Linear Model (GLM) function to compare normalized mortality counts with diabetes and asthma prevalence ratios. Due to adjustment for population apriori, an offset term was not included in the models.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.374959	0.3756395	6.322442	0.0000000
diabetes_ratio	7.808997	1.9100683	4.088334	0.0000434
asthma_ratio	-5.797244	3.1189057	-1.858743	0.0630636

Table 1: *Coefficient values from initial regression model.*

Returns from an initial GLM model indicated a statistically significant relationship between diabetes prevalence ratios and population adjusted COVID-19 mortality, with a p-value of Pr(>|z|) = 4.34e-05. The asthma prevalence variable returned marginal significance at best, with a Pr(>|z|) = .0631. Asthma's marginally significant return aligns with findings from preliminary research conducted in New York City where it was reported that, "*no statistically significant association between asthma status and mortality among patients with COVID-19*." (Lieberman-Cribbin et al, 2020). Due to Pr(>|z|) = .05 thresholds of significance not being met, I ran a second GLM model excluding the asthma ratio variable. While excluding asthma prevalence

from the second model, diabetes prevalence values retained statistical significance, with a p-value of Pr(>|z|) = 2.05e-05 (see Table 2).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.802495	0.2123782	8.487193	0.0000000
diabetes_ratio	8.008521	1.8801039	4.259616	0.0000205

Table 2: *Coefficient values from an updated regression model.*

I then extracted the deviance residual values from this second GLM and ran global Moran's I test using the residuals from each state level object and the binary weight structure outlined in figure 1. The test failed to reject the null hypothesis of spatial autocorrelation with a significant p-value of Pr(>|z|) = 4.97e-06. Since a statistical relationship was determined between state level diabetes prevalence and COVID-19 mortality counts, and evidence of spatial autocorrelation was determined within the model I constructed, I continued by attempting to adjust for spatial autocorrelation with a spatial regression model. To do this, I used a Conditional Autoregressive Model (CAR) with an adjacency function to account for the spatial relationship of the state polygons together (*see figure 1*).

Returns from this model did not include p-value measurements in the summary object. I accessed coefficients from the model's summary and was able to generate a 95% Confidence Interval (CI) from those data. The fitted output of the model produced a lower bound of CI = 4.05 and upper bound of CI = 14.62, indicating statistical significance of the model's output. When plotting the fitted values of the model against the mortality counts, I evaluated the strength of fit on visual terms, as seen a scatterplot below in figure 2, and comparative choropleth maps in figure 3 below.

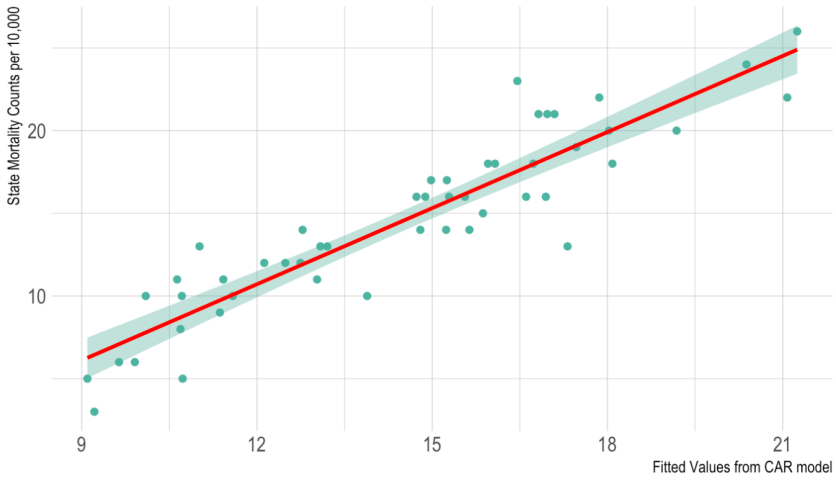

Figure 2: *Scatterplot of observed mortality counts plotted again fitted values from CAR model.*

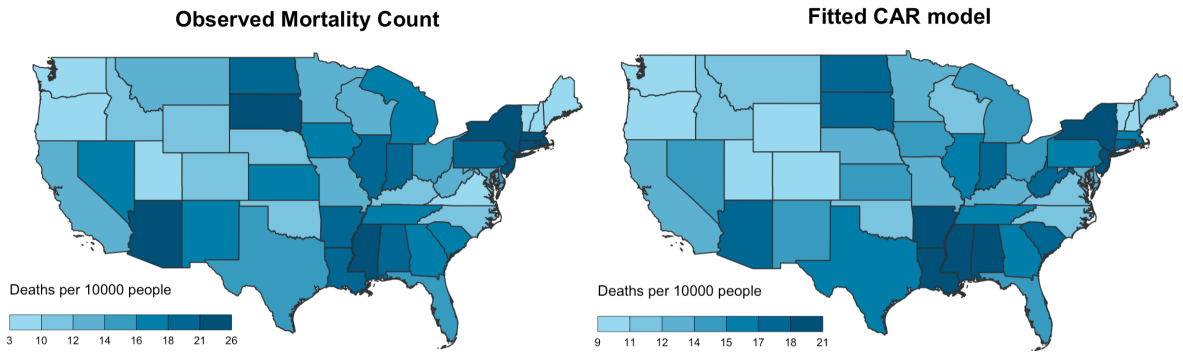

Figure 3*: Comparative maps illustrating observed values (*left*) and fitted CAR model values (*right*).*

Visual analysis of the CAR model's output illustrated a reasonable fit between observed COVID-19 mortality counts and state-level diabetes prevalence ratio values while accounting for spatial autocorrelation between states. I went on to explore Marginal Akaike Information Criterion (AIC) scores to evaluate a step wise fit of three model configurations associated with this study's analyses. Three comparisons were made: a). CAR model with spatial covariate, sans the disease prevalence; b) CAR model with disease prevalence and spatial covariate, and c) CAR model with disease prevalence, sans spatial covariate. While measuring AIC values on an identical computational package and functional framework, it was determined that the CAR model with disease prevalence and spatial interaction term included demonstrated a best fit amongst the three combinations tested, even if divergence was relatively minimal in comparison (*see Figure 4*).

	Marginal AIC
a). CAR sans Disease Covariate	308.6696
b). CAR with Disease Covariate	299.4750
c). CAR sans Spatial Covariate	305.1189

Figure 4: *Comparing AIC values of three regression models.*

In the final steps of this study's analysis, I explored residual values extracted from the "best fit" of CAR models outlined above. I first plotted a histogram to observe a distribution of standard deviations of residual values, (*see figure 5*). I then used those values to test for persisting global spatial autocorrelation with a post-hoc global Moran's I test. To do this, I used an identical weighting structure and adjacency matrix to the first Moran's I test, (*see figure 1*). Results from this test failed to reject the alternative hypothesis of spatial autocorrelation with p-value Pr(>|z|) = .0147. These results indicated a continued presence of spatial autocorrelation within the fitted

model, suggesting that further adjustment, approaches, and/or inclusion of covariates into the model may better explain the spatial relationship of mortality outcomes between states.

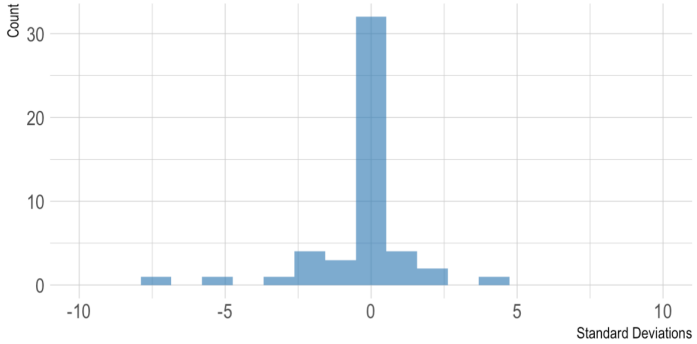

Figure 5*: Histogram plotting counts of residual standard deviations from CAR model.*

Limitations:

Areal objects at this scale pose unique challenges to geospatial inference. The raw data of this analysis could be considered sparse, with the selected variables using only few measurements while hoping to shed light on vastly complex phenomena and interconnected relationships. As variance exists within spaces, the state-level scale of measurements may not reflect the impact of the selected variables on smaller units of measurement such as county, census tract, or community.

The arithmetic methods used in this study are crude, and do not account for variance of COVID-19 testing between states. During the interval of time this study's data draws from, state-level areas experienced unique approaches and participation with COVID-19 testing. "Confirmed" cases relative to population between states, as well as administrative lag around mortality statistics, is a likely effect modifier informing differences in the numerator of the "Death per 10000 people" measurement used in this study's analysis.

Discussion:

As hypothesized, evidence of spatial autocorrelation within a chronic disease variable (diabetes), and crude COVID-19 mortality counts appears present. The extent of that relationship does not explain spatial variation within these data alone. The methods this study employ highlight thoughtful approaches to multivariate regression analyses of spatial data aggregated within an areal structure. More than an end to itself, this work reflects initial explorations into how one may approach understanding regression model fit of covariate data and inference around larger spatial relationships alive in our world.

Bibliography:

Burstrom, Bo. Tao, Wenjing. "Social Determinants of Health and Inequalities in COVID-19". European Journal of Public Health. Vol 30, Issue 4, Aug. 2020 DOI: https://doi.org/10.1093/eurpub/ckaa095

Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. 2020. "Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis." PLoS Med 17(9): e1003321. https://doi.org/10.1371/journal.pmed.1003321

Lieberman-Cribbin, Wil. Rapp, Joseph. Alpert, Naomi. Tuminello, Stephanie. Taioli, Emanuela. "The Impact of Asthma on Mortality in Patients With COVID-19". Asthma Research Letter, CHEST. 2020. American College of Chest Physicians. DOI: https://doi.org/10.1016/j.chest.2020.05.575

Liu, Hong. Chen, Shiyan. Liu, Min. Nie, Hao. Lu, Hongyun. "Comorbid Chronic Diseases are Strongly Correlated with Disease Severity among COVID-19 Patients: A Systematic Review and Meta-Analysis". June 2020, Aging and Disease. Volume 11, Number 3; 668-678. DOI: http://dx.doi.org/10.14336/AD.2020.0502

State of Childhood Obesity, "Diabetes in the United States". 2021. Web Resource. Retrieved from: https://stateofchildhoodobesity.org/diabetes/

The Comprehensive R Archive Network. "Contributed Packages". 2021. Web Resource. Retrieved from: https://cran.r-project.org/

United States Center for Disease Control. "Excess Deaths Associated with COVID-19", 2021. Web resource. Retrieved from: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm

United States Center for Disease Control. "Most Recent Asthma State or Territory Data", 2021. Web resource. Retrieved from: https://www.cdc.gov/asthma/most_recent_data_states.htm

Project code and data repository:
https://github.com/Averysaurus/spatial_regression_covid19