Article

Predictors of Death Rate During the COVID-19 Pandemic

Ian Feinhandler 1, Benjamin Cilento 2, Brad Beauvais 3, Lawrence Fulton 4,*

1 Department of Geography and Political Sciences, Front Range Community College, Longmont, CO, ian.feinhandler@gmail.com
2 Memorial Hermann Hospital, Houston, TX, ben.cilento@me.com
3 School of Health Administration, Texas State University, San Marcos, TX, bmb230@txstate.edu
4 School of Health Administration, Texas State University, San Marcos, TX, lf25@txstate.edu

Received: date; Accepted: date; Published: date

Abstract: COVID-19 is a potentially fatal viral infection. This study investigates geography, demography, socioeconomics, health conditions, hospital characteristics, and politics as potential explanatory variables for death rates at the state and county levels. Data from the Centers for Disease Control and Prevention, the Census Bureau, and other sources were used to evaluate spatial regression models. Yearly pneumonia and flu death rates (state level, 2014-2018) were evaluated as a function of the governors’ political party using repeated measures analysis. Spatial regression at the county level discovered a statistically significant model that included only geography, racial and ethnic status along with a political factor. State level analysis was consistent with this finding. The political factor did not, however, appear in a subsequent analysis of 2014-2018 pneumonia and flu death rates. This study suggests racial/ethnic composition and geographic relationships with the outbreak are important considerations but do not fully explain death rates without inclusion of political factors. The pathogenesis of COVID-19 has greater and disproportionate effect within racial and ethnic minority groups. While population density was not found to be significant, political influence on the reporting of COVID-19 mortality was a significant finding.

Keywords: COVID-19, Geospatial Regression, Health Disparities, Public Health

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiologic agent of the coronavirus (COVID-19) pandemic. As of August 7, 2020, the associated death toll in the United States is reported to have surpassed 150,000 [1], the highest of any country in raw numbers but equivalent to many other developed countries when adjusted for population [2]. The proper recognition and remediation of the disease are pressing concerns and each will likely be subject to debate in the months prior to the 2020 presidential election [3,4]. However, there is some concern surrounding the veracity of the data and factors contributing to COVID-19 deaths. Media outlets provide daily updates on the number of cases and deaths but draw this information from data collection and reporting agencies that have adjusted their methods over time [5]. The resulting inconsistencies have led to charges of under-reporting [6,7] and over-reporting [8,9], and have contributed to the politicization of the pandemic.

COVID-19 data inconsistencies and potential political bias in data reporting can have significant implications. If the data that politicians rely on is faulty, subsequent policies may harm public
health, the economy, and other aspects of society. Thus, given the novel nature of the virus and its progression, and the known inconsistencies in the reported data, we sought to gain a deeper understanding of the factors that contribute to COVID-19 deaths.

1.1. Research Questions

We investigated three research questions. First, what attributes of geography, demography, population density, the economy, state health conditions, hospital characteristics, and politics might explain deaths per 100,000 (death rate) at the county-level as of August 7, 2020? Second, did COVID-19 death rates at the state level differ based upon Governor party affiliation after accounting for other relevant variables? As a control for our second line of inquiry, we also examined whether variation existed in previous flu/pneumonia death rates (2014-2018) based upon the governor’s party affiliation.

1.2. Significance and Motivation

To our knowledge, this research is the first to evaluate SARS2-COVID using combined data from multiple areas covering demographic, socio-economic, health system, population health, and political factors using spatial regression. It is also the first study to evaluate the effects of state and county political affiliation on COVID-19 death rates. The motivation behind this study is to address the media promulgation of explanatory factors that may or may not be scientifically verifiable (e.g., population density and political factors), particularly when placed in the context of other known factors established at the individual unit of analysis (e.g., race).

2. Methods

2.1. Sample Sizes, Data Sets, Variables

Sample sizes for the research questions were 3,116 (county), 51 (state), and 250 (50 states by 5 years). The dependent variable was the death rate. Cumulative COVID-19 deaths were obtained from USAfacts.org [1] for August 7, 2020. Flu data were from the Centers for Disease Control and Prevention – CDC (2014-2018)[10]. Definitive Healthcare data provided descriptive hospital-related information [11]. Population and demographic data were from the Census Bureau [12]. The Centers for Medicare and Medicaid provided the source for relevant patient morbidity proportions by state and county [13]. Geographic variables in the analysis included the shapefiles from the Census Bureau’s state and county Tiger Files [14]. Race and ethnicity variables included the proportion of African American, Native American, and Hispanic. Population density (population per square kilometer), and the proportion 65 and older served as additional control variables. Economic variables included median household income and unemployment. Population health status variables included the population proportions with chronic obstructive pulmonary disorder (COPD), heart failure, diabetes, obesity, and cancer; all of which have been identified as risk factors at the individual level [15]. Hospital characteristics included the average case-mix index (CMI) and aggregate acute beds. The number of physicians and patient census by geographic
unit were highly collinear and thus omitted. At the county level, the winning party in the 2016 presidential election was a proxy for possible reporting differences. At the state level, we looked at each governor’s affiliated party in 2020. See Appendices 1 and 2 for a description of all variables.

2.2. Models

2.2.1. County-Level Analysis

Lasso, elastic net, and ordinary least square regression techniques [16] with 10-fold cross validation (CV) evaluated deaths per 100,000 to identify predictive variables (all variables scaled). Those variables were then included in an OLS and a subset of significant variables (α = .05 level) were used for residual analysis. Moran’s I, a global test for spatial correlation, informed the need for spatial analysis. Lagrangian multiplier diagnostic tests (non-robust and robust) informed model selection. A refined spatial model composed only of statistically significant variables was then estimated.

2.2.2. State-Level COVID-19 and Influenza Analysis

Based on county-level results, we conducted a state-level analysis in which we examined the relationship between death rates, race, ethnicity, and party of the governor in 2020. We ran spatial regression to estimate those effects and account for geographic relationships.

We also investigated reporting differences that might exist for flu and pneumonia deaths. Using repeated measures analysis, we modeled the logarithm of flu and pneumonia deaths as a function of year and governor party. All analyses were performed in R Statistical Software [17].

3. Results

3.1. Descriptive Statistics

Table 1 presents a county-level summary of the association between 2016 presidential election results, population density, and deaths from COVID-19. Population density is higher for counties that voted Democratic (116.2 versus 23.5), as are death rates (64.3 versus 29.6).

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Counties Won</th>
<th>Mean Population Density</th>
<th>Deaths</th>
<th>Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinton</td>
<td>491</td>
<td>116.2</td>
<td>114,644</td>
<td>64.30</td>
</tr>
<tr>
<td>Trump</td>
<td>2625</td>
<td>23.5</td>
<td>44,227</td>
<td>29.60</td>
</tr>
<tr>
<td>Total</td>
<td>3116</td>
<td>41.5</td>
<td>158,871</td>
<td>48.40</td>
</tr>
</tbody>
</table>

Table 2 summarizes the descriptive statistics study variables. The COVID-19 death rate averages by state and county are 40.25 and 26.04, respectively. The flu and pneumonia death rate, 2014-2018, is 15.10. The mean county population was 9% African American, 2% Native American, 9%
Hispanic, and 20% aged 65 and over. Population density, income, and unemployment averages were 106.45 per square kilometer, $58,000 per county-person, and 3.6% per county, respectively. The largest comorbidity proportion average was obesity (32.85%), and the median number of acute beds was 35 with an average of 215. Average CMI was 1.06 with a median of 1.17. Sixteen percent of counties voted Democratic in 2016.

### Table 2. Descriptive statistics (as of 7 August 2020)

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>State COVID-19 Death Rate</td>
<td>51</td>
<td>40.25</td>
<td>39.46</td>
<td>28.94</td>
<td>2.26</td>
<td>178</td>
</tr>
<tr>
<td>County COVID-19 Death Rate</td>
<td>3,116</td>
<td>26.04</td>
<td>40.81</td>
<td>11.63</td>
<td>-</td>
<td>413.86</td>
</tr>
<tr>
<td>Flu State Death Rate</td>
<td>250</td>
<td>15.10</td>
<td>3.76</td>
<td>14.65</td>
<td>7.00</td>
<td>29.60</td>
</tr>
<tr>
<td>African American</td>
<td>3,116</td>
<td>0.09</td>
<td>0.15</td>
<td>0.02</td>
<td>-</td>
<td>0.87</td>
</tr>
<tr>
<td>Native</td>
<td>3,116</td>
<td>0.02</td>
<td>0.06</td>
<td>-</td>
<td>-</td>
<td>0.90</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3,116</td>
<td>0.09</td>
<td>0.14</td>
<td>0.04</td>
<td>-</td>
<td>0.99</td>
</tr>
<tr>
<td>% 65+</td>
<td>3,116</td>
<td>0.20</td>
<td>0.05</td>
<td>0.19</td>
<td>0.05</td>
<td>0.58</td>
</tr>
<tr>
<td>Population Density</td>
<td>3,116</td>
<td>106.45</td>
<td>697.38</td>
<td>17.52</td>
<td>0.09</td>
<td>27,755.49</td>
</tr>
<tr>
<td>Unemployment %</td>
<td>3,116</td>
<td>3.96</td>
<td>1.39</td>
<td>3.70</td>
<td>0.70</td>
<td>18.30</td>
</tr>
<tr>
<td>Income</td>
<td>3,116</td>
<td>52,686</td>
<td>13,838</td>
<td>50,530</td>
<td>25,385</td>
<td>140,382</td>
</tr>
<tr>
<td>Cancer</td>
<td>3,116</td>
<td>7.41</td>
<td>1.40</td>
<td>33.10</td>
<td>12.40</td>
<td>57.70</td>
</tr>
<tr>
<td>Obesity</td>
<td>3,116</td>
<td>32.85</td>
<td>5.43</td>
<td>33.10</td>
<td>12.40</td>
<td>57.70</td>
</tr>
<tr>
<td>COPD</td>
<td>3,116</td>
<td>12.82</td>
<td>3.76</td>
<td>12.45</td>
<td>-</td>
<td>32.15</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3,116</td>
<td>26.94</td>
<td>5.08</td>
<td>27.13</td>
<td>-</td>
<td>49.62</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>3,116</td>
<td>14.40</td>
<td>3.27</td>
<td>14.15</td>
<td>-</td>
<td>33.75</td>
</tr>
<tr>
<td>Acute Beds</td>
<td>3,116</td>
<td>215</td>
<td>721</td>
<td>35</td>
<td>-</td>
<td>19,274</td>
</tr>
<tr>
<td>Case Mix Index</td>
<td>3,116</td>
<td>1.06</td>
<td>0.59</td>
<td>1.17</td>
<td>-</td>
<td>2.71</td>
</tr>
<tr>
<td>Winning Party</td>
<td>3,116</td>
<td>0.16</td>
<td>0.36</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### 3.2. COVID-19 Death Analysis, County

Elastic net, lasso and OLS models performed similarly based on Root Mean Squared Error (0.831, 0.830, 0.830, respectively), and on adjusted $R^2$ (0.307, 0.307, 0.307, respectively). Model variable selection was identical when rounding coefficients to two significant digits. A final OLS model included only statistically significant variables. The residuals of the error terms from this model were mapped (Figure 1), indicating a pattern of positive spatial autocorrelation or “clustering.” Moran’s I confirmed this visual analysis, (I=0.356, p<.001). Lagrangian multiplier diagnostics were statistically significant for both error and lag (p<0.001 in all cases).
Figure 1. Choropleth map, deaths per 100,000 residuals from regression

3.2.1. Generalized Spatial Two-Stage Least Squares Model (GS2SLS)

A generalized spatial two-stage least squares model (GS2SLS) [18] using the final OLS model resulted in even a smaller subset of significant variables due to geospatial correlation effects. The final subset of variables included the geographic relationship weights (\(\rho\)), African American, Native American, Hispanic, and winning party status (Table 3). The estimated effect size is moderate, \(R^2 = 0.459\).

| Variable             | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| Rho                  | 0.593    | 0.047      | 12.750  | <0.001   |
| (Intercept)          | 0.541    | 0.985      | 0.549   | 0.583    |
| Native               | 46.295   | 8.404      | 5.509   | <0.001   |
| Hispanic             | 27.473   | 4.301      | 6.387   | 0.003    |
| African American     | 63.338   | 6.398      | 9.900   | <0.001   |
| Winning Party        | 6.162    | 1.674      | 3.682   | <0.001   |

Native American, Hispanic, and or African American proportions are associated with a 46.295, 27.473, and 63.338 increase in deaths per 100,000 individuals, respectively. County political leaning based on the 2016 presidential election is associated with an increase of 6.162 deaths per 100,000 individuals. Moran’s I is not significant (I = -0.0.035, p=0.961).

After conducting spatial regression analysis, we investigated the effect of party voting on death rates marginally using a descriptive table. These results also supported the inclusion of the political factor. An important result is that while we evaluated population density, its effect disappears when other factors are considered. This county-level analysis is congruent with Pew Research
findings that death rates are higher in Democratic-led counties [19]. This study suggests that racial/ethnic composition and geographic relationships with the outbreak are important considerations but do not fully explain death rates without inclusion of the political factor.

3.3. **COVID-19 Death Analysis, State**

Given the results of the political analysis at the county level we further evaluated political leadership at the state level, replicating the analysis previously conducted by examining race, ethnicity, and governor’s party variables. Death rates were mapped, and states in the Northeast (New York, New Jersey, Massachusetts, Connecticut, and Rhode Island) had higher death rates than other areas of the country.

Based on residual diagnostics, we estimated a spatial regression model which included the two remaining statistically significant estimators from the non-spatial analysis: African American and governor’s party. The coefficient estimates for the spatial regression are shown in Table 4. There is strong evidence that increases in the African American proportion in a state are associated with higher death rates. Moran’s I was not significant after regression (I=-0.099, p=0.902).

### Table 4. Spatial regression coefficient table, state level

| Variable           | Estimate | Std. Error | t value | Pr(|t|) |
|--------------------|----------|------------|---------|--------|
| Rho                | 0.789    | 0.225      | 3.509   | <0.001 |
| (Intercept)        | -9.640   | 8.856      | -1.089  | 0.276  |
| Governor’s Party   | 13.900   | 7.579      | 1.834   | 0.067  |
| African American   | 98.022   | 34.806     | 2.816   | 0.005  |

Using this analysis, we estimated how removing the effect of the governor’s party would impact the death rate estimates (results presented in supplementary material online). The presence of Democratic leadership is associated with an increase of 13.9 deaths per 100,000 individuals after accounting for other factors.

3.4. **Flu Death Analysis, State**

The results of the state-level analysis prompted an examination of the changes in death rate between past influenza outbreaks and COVID-19, to determine whether the change is consistent between states controlled by Democratic governors when compared to states controlled by Republican governors. To investigate, we ran a repeated measures (by state) analysis of variance on the log-transformed death rate for 2014-2018. The model identified no effects associated with the governor party affiliation ($F_{1,244}=1.5311$, $p=0.2165$), only reporting year ($F_{4, 244}=2.382$, $p=0.040$).

4. Discussion

4.1. **Native American, Hispanic, and African American Population Effects**
Our study confirms the findings of numerous researchers pertaining to health care–disparities in the United States, particularly with respect to Native American, Hispanic, and African American populations [20-22]. We found an increase in the percentage of these populations to be associated with an increase in mortality from COVID-19. McLaren (2020) attributes this difference to disparities in education, occupation, and commuting patterns [22]. Although we did not include these factors in our analysis, we did find the mortality disparities do not appear to be attributable to differences in unemployment rates or household income. Our findings affirm that resources are ineffectively distributed in the United States, but may also be indicative of a pathogenesis of COVID-19 that has greater and disproportionate effect within these three racial groups [23, 24].

4.2. Population Density Effects

Population density has been identified as a predictive factor in disease progression[25, 26]. A superficial examination of county-level data indicates that a relationship might exist between population density and death rate from COVID-19 (see Table 1). Consistent with prior analysis [27, 28], Table 1 also shows urban areas tended to vote Democrat in the 2016 presidential election. Due to these associations, media outlets have presented the urban-rural divide as a viable explanation for the difference in death rates between counties that voted Democrat in 2016, and those that voted Republican [29, 30]. This divide has also provided an explanation for the divergent response to the disease based on party affiliation. For example, Democrats are more concerned about COVID-19 than Republicans, and are more likely to wear a facemask and practice other forms of social distancing [31-33]. However, population density falls out of the model when other factors are considered. The failure of population density to provide a statistically significant explanation for deaths from COVID-19 has been one of the most surprising results from our analysis.

4.3. Political Party Effect

The influence of politics on the reporting of COVID-19 mortality was a significant finding in our analysis. County level Democratic affiliation was highly and significantly associated with increased COVID deaths, even after controlling for factors such as population density, etc. To the best of our knowledge, this is the first time that population density and urbanization are ruled out as an explanation for the disparity in death rates between Democratic and Republican states.

5. Conclusions

During our analysis, we evaluated the data that pointed toward political interference in the reporting of COVID-related deaths. As of August 7, 2020, it is clear that the national death rate from COVID-19 is higher than from other flu pandemics, but the increase in the death rate in states with Democratic governors has been far greater than the increase in states with Republican governors. The explanation for this – in part – may be related to how the CDC collects and analyzes data.

In past years, the CDC retrospectively tabulated the number of flu-associated illnesses, hospitalizations, and deaths – a process that takes up to two years to generate an estimate. The
process relies on estimation modeling in and out of hospitals based on behavioral algorithms[34]. The CDC never relies solely on death certificate data because it recognizes that there is never large-scale testing and that the clinicians do not routinely list influenza data on death certificates if the patient died of pneumonia, heart failure, or deteriorating lung disease. According to the CDC, this leads to significant underreporting of deaths due to flu every year [34].

On February 20, 2020, the CDC published guidelines for the diagnosis and mandatory reporting of COVID-19 for any patients evaluated with “COVID related” illnesses. This applied to all health care practitioners and included a comprehensive set of instructions and codes to document any relationship to COVID-19 on the death certificates [35]. This represents a significant change in reporting of the disease and consequently the inclusion on the death certificate. Three separate additional guidelines put out in March and April affirmed these measures. In addition, the new CDC guidance stated that: “In cases where a definite diagnosis of COVID–19 cannot be made, but it is suspected or likely, it is acceptable to report COVID–19 on a death certificate as ‘probable’ or ‘presumed’”[35]. This change introduced significant potential variation in the tabulation of COVID-19 death tolls.

At approximately the same time, the Centers for Medicare and Medicaid Services (CMS) authorized an additional 20% reimbursement for patients carrying a diagnosis of COVID-19 pursuant to Sections 3710 and 3711 of the CARES Act [36]. These changes created a financial incentive for hospitals to classify patients as positive for COVID-19. Importantly, at the time these measures were introduced, the dominant model used by policy-makers – based on Ferguson et al. [37] – predicted an exceptionally high mortality rate [38]. By late March more accurate estimates predicted a mortality rate well below original expectations [39]. This should have triggered a policy reversal from the CDC and CMS, but no changes were noted. In short, in the politically charged landscape of 2020, the CDC’s new way of collecting data, combined with CMS’ monetary incentives, may have resulted in the over-reporting of COVID-19 deaths. The introduction of these two new sources of reporting bias makes historical comparisons unreliable at best. Without reliable data, it is difficult to effectively fight a pandemic. This conundrum associated with the reliability of data on COVID-related deaths highlights the need for objective and uniform standards for case-identification and data-collection.
Supplementary Materials: Variables and sources posted at mdpi.com


Funding: This research received no external funding.

Acknowledgments: None.

Conflicts of Interest: The authors declare no conflict of interest.

References


