Accumulated COVID-19 Death-Rate - A County-level comparison

May, 2021

Zhizun Li, Yuval Shpira, Yao Tong, Linan Zhou

Abstract

This thesis seeks to analyze the determinants of county-level accumulated death rate caused by COVID-19 (measured in deaths per thousand people). We accumulated county-level death cases from January 21 2020, when the first reported coronavirus case took place in Washington, to April 12 2021. We used the death rate per thousand people to measure the severity of the pandemic in each county. Demographic and climate related features were included to picture the difference between counties and detect how they might affect the fatality rates. Using our data, we showed that higher population density, larger proportion of elder population, lower median household income and lower education level were associated with higher death rate. Proportion of minority races: Black, Asian and Hispanic, contributed to the difference of fatality rates as well, though a higher proportion of Asian population was correlated with a lower death rate whereas blacks and Hispanics were higher. The thesis also provided evidence that scientific interventions can effectively decrease the overall death rate: vaccinations and high propensity of wearing masks contributed to a decrease in the death rate. Warmer weather climate is also shown to have a negative effect on the death toll, while rainfall (proportion of rainy days in a year) and wind speed show the opposite effect.

1.Introduction

COVID-19 hit the world at the beginning of 2020, including the US. According to WHO, 30,772,857 confirmed cases of COVID-19 with 555,712 deaths were reported, from 3 January 2020 to 2:37pm CEST[1]. Different counties took different actions while facing the same pandemic, resulting in differences in death rates caused by COVID-19. Various demographic and political features across the country might also contribute to differentiating the results.

In our analysis, we tried to build causal relationships between economic indicators and death rate in geographic regions[2] in the United States, then quantify the influence of each determinant. More specifically, 48 contiguous states[3] plus the District of Columbia are included here.

The first section summarizes the data used. The second section presents the OLS model with results. The last section discusses the interpretation of our models as well as deviations between our expected value and the actual regression results.

[1] World Health Organization data of United States: https://covid19.who.int/region/amro/country/us
[2] The geographic regions in this paper are divided according to county-level FIPS code and state-level FIPS code. Detailed information can be found at https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt.
[3] Continuous United States: https://en.wikipedia.org/wiki/Contiguous_United_States

2.Data

In this section, we calculated the statistics for our variables and summarized them in Table 1. Explanations for key variables are shown below.

Accumulated death rate

The accumulated death rate describes the accumulated death cases caused by COVID-19 to the total population in each county. In this paper we use death rate per thousand people to measure the accumulated death rate. This dataset includes the accumulated death cases of each county as of April 12, 2021 published by the The New York Times[4], and the total population of each county published at the ERS[5] website. About 20% of the counties listed have accumulated death rates of lower than 1 per thousand people, about 15% of the counties listed have accumulated death rates of greater than 3 per thousand people, and the average death rate of the listed counties is 1.92 per thousand people.

Vaccination

In this dataset, we included vaccination data generated at 21:58am ET on April 12, 2021, published at the official website of CDC[6]. For counties included here, about 90% of them had fully vaccinated rates between 18-28% for populations aged more than 18. Almost all regions managed to fully-vaccinated more than half of their elder population (aged more than 65), except for Alabama with 49.3% which was also quite close to 50%.

Poverty rate

This is the estimated percentage of people of all ages in poverty in 2019 in the United States, published by the U.S. Department of Commerce, Bureau of the Census, Small Area Income and Poverty Estimates (SAIPE) Program[7]. The poverty rate of 71% counties listed in the dataset was lower than 17%. The poverty rate of 0.71% counties was larger than 35%. Ziebach, Todd and Madison were the only three counties which had a poverty rate larger than 41%.

Net-migration rate

This data describes the net migration (both including domestic and international migration) in the United States from July 1, 2018 to June 30, 2019. 49.5% of the counties had positive

---

[4] The geographic regions and county-level FIPS code and state-level FIPS code: New York Times data of COVID 19: https://github.com/nytimes/covid-19-data

[5] County level data: https://www.ers.usda.gov/data-products/county-level-data-sets/download-data.aspx

[6] Centers for Disease Control and Prevention of the U.S. (CDC) updates vaccination data across the country with the information received from each state. Data on doses of vaccine distributed and administered include data received by CDC as of 6:00 am ET on the day of reporting, then CDC would update vaccination data on COVID Data Track between 1:30 pm and 8:00 pm ET each day.

[7] County level data: https://www.ers.usda.gov/data-products/county-level-data-sets/

net-migration rates, and 49.98% of the counties had negative net-migration rates. 73% of the counties had the net-migration rates ranging from -10% to 10%. Gulf was the smallest negative outlier and Loving was the largest positive outlier in the data.

Median household income

This data describes the estimated median income of each county in 2019. 90% of counties had the median income ranging from $35,232 to $77,232. Loudoun had the largest median income at $151,806, which was much higher than the other counties.

Habit of wearing masks

The habit of wearing masks was measured by the percentage of the population that always wear masks in each county. The original dataset came from a survey conducted by The New York Times and Dynata in July 2020[8]. The habit of wearing masks was nearly normal-distributed, with an average value of 51% among all the listed counties. Only 8% of the listed counties reported that less than 30% people do not always wear masks, while half of the listed counties reported over 50% people always wore masks.

Propensity to wear a mask

Initially there were five distinct variables of self-reported mask wearing (Always, Frequently, Sometimes, Rarely, Never), we decided to perform a linear transformation to include all of them into one variable that will incorporate all of them. The formula chosen was:
Always + 0.7 * Frequently + 0.4 * Sometimes + 0.1 * Rarely (+ 0 * Never).
By doing so we accounted for self over reporting, and didn't disregard any of the reports.

Average temperature

The average temperature in the county, gathered in the year 2020. It is measured in Fahrenheit (°F). The variable shows an almost normal distribution, with very little skewness (Mean=Median).

Wind Speed

The average wind speed in the county, gathered in the year 2020. We encountered 193 missing observations. We attempted to face this problem in a number of ways, as can be seen in section 4 models.

[8] Josh Katz, Margot Sanger-Katz & Kevin Quealy (2020). A Detailed Map of Who Is Wearing Masks in the U.S, New York Times,
https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html

Rain

The proportion of days it rained in the county over the year 2020. While the amount of precipitation is an important variable and might "scale" better, due to the varying size of the counties, we surmised that the chances of a rainy day would keep the population indoors and provide a better outlook than if it rains in copious amounts for a few selected days.

Uninsurance

This indicator describes the county-level health insurance conditions. It is measured by the percentage of the population without health insurance.

Race

We collected percentages of Black, Asian and Hispanic population in each county as part of our demographic indicators.

Percentage of female

Percentage of female population per county is included in the dataset

Table 1: Statistics for numeric variables

Variable	Mean	Median	Std. Dev	Min	Max
Accumulated Death rate per thousand people due to COVID-19	1.92	1.77	1.12	0	8.66
Percentage of population fully-vaccinated	21.78	21.70	2.99	15.00	29.40
Percentage of population age 65+ fully-vaccinated	62.45	62.40	6.50	49.30	77.10
Estimated percent of people of all ages in poverty 2019	14.47	13.4	5.79	2.7	47.7
Death rate in period 7/1/2018 to 6/30/2019	10.45	10.5	2.65	0	22.1
Natural increase rate in period 7/1/2018 to 6/30/2019	0.53	0.3	3.91	-13.5	24.4
Net migration rate in period 7/1/2018 to	0.39	0	12.31	-165.4	126.2

6/30/2019					
Percentage of female	49.91	50.32	2.23	26.84	56.87
Percentage of uninsured population	11.42	10.52	5.11	2.26	33.75
Population density per square mile	214.81	45.1	791.82	0.11	18131.89
Percentage of population 65+	17.64	17.3	4.45	3.9	53.1
Mean temperature (°F)	57.40	57.12	7.89	28.94	79.58
Rain (% of rainy days)	0.3040	0.3381	0.1584	0	0.7049
Habit of wearing a mask (Percentage of population always wear a mask)	0.51	0.50	0.15	0.12	0.89
Propensity to wear a mask	0.7095	0.7110	0.1143	0.3137	0.9547
Percentage of Black	9.06	2.27	14.36	0	85.41
Percentage of Asian	1.46	0.73	2.45	0	38.31
Percentage of Hispanic	9.66	4.38	13.88	0.61	96.36

3.Model

Though COVID-19 is still threatening the health of the public around the world with hundreds of thousands of on-going studies to unveil its origin and causes, a few existing papers have already measured the relationships between the mortality rate of COVID-19 patients and other features.

Bhadra et al.[9] investigated the influence of population density on COVID-19 spread and related mortality in the context of India and found moderate association between COVID-19 spread and population density. Cahill et al.[10] accessed the New York Times GitHub repository COVID-19 data and 2018 United States Census data for all United States Counties and suggested that COVID-19 fatality rates be related to advanced population age. Roser et al.[11] analyzed country-by-country data on mortality risk of the COVID-19 pandemic via visualization. They explained that if infected with this virus, the elderly were at the greatest risk of dying because they were more likely to have health conditions such as cardiovascular diseases, respiratory diseases or diabetes. Sandhu et al.[12] uniquely examined relationships between in-hospital COVID-19 mortality risk of African Americans in urban cities and census tract social vulnerability characteristics and concluded that there is an elevated mortality risk for Black people living in areas flagged for extreme socioeconomic vulnerability. Their study supported the possibility that socioeconomic vulnerability (i.e., poverty, unemployment) explains the higher death risk in their sample.

We started building our model by including only the demographic factors.

Based on the former study, we included population density per square mile, natural increase rate in population and net migration rate to represent the living environment at country level concerning the population density. All of these 3 variables were correlated with death rate at 1% significance level.

As for economic factors, we collected unemployment rate, poverty rate and median household income for each recorded county. A preliminary study showed all of them were statistically significantly correlated with death rate at 5% level, but they also had strong relationships with one another. Including them all into the model would definitely cause a serious endogeneity problem. In

---

[9] Bhadra, A., Mukherjee, A. & Sarkar, K. Impact of population density on COVID-19 infected and mortality rate in India. Model. Earth Syst. Environ. 7, 623–629 (2021). https://doi.org/10.1007/s40808-020-00984-7

[10] Gina Cahill, Carleigh Kutac, Nicholas L. Rider, Visualizing and assessing US county-level COVID19 vulnerability, American Journal of Infection Control, 2020, ISSN 0196-6553.

[11] Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus'

[12] Avnish Sandhu, Steven J. Korzeniewski, et al., Elevated COVID-19 mortality risk in Detroit area hospitals among patients from census tracts with extreme socioeconomic vulnerability, EClinicalMedicine, Volume 34, 2021, 100814, ISSN 2589-5370.

this case, we selected only median household income as our economic indicator as it directly determined the poverty level for a family and was hugely impacted by unemployment.

For socioeconomic vulnerability indicators other than income, we included sex (percentage of female) and education (percentage of population with no high school diploma). Preliminary analysis suggested that sex had a slightly positive relationship with death rate (0.011), whereas education was statistically significantly correlated with it (0.361).

Model 1: death rate ~ log(population density) + old + female + log(income) + education + net migration rate
Model 1.a: Model 1 + race factors

Model 1 showed that the higher population density, the higher the death rate in that county, though its coefficient is not statistically significant here. Also, counties with larger proportions of population aged more than 65 were associated with higher mortality rates. Lower education level, as expected as well, increased the death rate per county level.

Then, we added race factors to Model 1 and summarized the result in Model 1.a. It seemed that each percent increase in Black and Hispanic population would cause an increase in mortality rate in that county, while the increase of Asian population lowered the death rate. This move also improved both the R-squared and adjusted R-squared ratios and decreased the residual standard error, so for now, we took it as a better model.

Regression results for Model 1 and Model 1.a are shown in Table 2.

Table 2: Regression results for Model 1 and Model 1.a

	Dependent variable: Death rate per thousand	
	Model 1	Model 1.a
log(Population Density)	0.003 (0.014)	-0.003 (0.015)
Percentage of the Old	0.032*** (0.005)	0.041*** (0.005)
Percentage of Female	0.024*** (0.009)	0.017* (0.009)
log(Income)	-0.502*** (0.109)	-0.274** (0.123)
Percentage of people without High School Diploma	0.051*** (0.004)	0.040*** (0.004)
Net Migration Rate	-0.015*** (0.002)	-0.016*** (0.002)
Percentage of Asian		-0.028*** (0.009)
Percentage of Black		0.012*** (0.002)
Percentage of Hispanic		0.006*** (0.002)
Observations	3,103	3,103
R2	0.197	0.216
Adjusted R2	0.196	0.214
Residual Std. Error	1.008 (df = 3096)	0.997 (df = 3093)
F Statistic	126.844*** (df = 6; 3096)	94.867*** (df = 9; 3093)
Note:		*p<0.1; **p<0.05; ***p<0.01

Model 2: death rate ~ log(population density) + old + female + log(income) + education + net migration rate +race factors + percentage of population fully-vaccinated + percentage of uninsured population

Model 2a: Model 2 + percentage of population aged over 65 fully-vaccinated (better)

Based on model 1, we added vaccination and health care factors of each county in Model 2. We introduced the percentage of fully-vaccinated population to measure vaccination and the percentage of the uninsured population to measure health care conditions. The correlation between percentage of the uninsured population and death rate was positive as expected and statistically significant, which means higher uninsurance rates were associated with higher death rates of COVID-19. The correlation between vaccination factors and death rate, however, was not statistically significant in Model 2.

Considering that counties with larger proportions of population aged over 65 were associated with higher mortality rates, we introduced a new variable percentage of fully-vaccinated population aged over 65 to explain vaccination factors in Model 2a. The new model worked better in explaining the relationship between vaccination and death rate, it turned out that the correlation between vaccination rate and death rate was negative and was statistically significant. The new Model 2a outperformed Model 2 with higher R-squared and adjusted R-squared value, smaller residual standard error and smaller AIC and BIC.

Regression results for Model 2 and Model 2a are shown in Table 3.

Table 3: Regression results for Model 2 and Model 2a

	Dependent variable: Death rate per thousand	
	Model 2	Model 2a
log(Population Density)	0.017	0.020
	(0.016)	(0.016)
Percentage of the Old	0.041***	0.041***
	(0.005)	(0.005)
Percentage of Female	0.012	0.015*
	(0.009)	(0.009)
log(Income)	-0.264**	-0.338***
	(0.122)	(0.123)
Percentage of People without High School Diploma	0.033***	0.036***
	(0.005)	(0.005)
Net Migration Rate	-0.017***	-0.017***
	(0.002)	(0.002)
Percentage of Asian	-0.028***	-0.024***
	(0.009)	(0.009)
Percentage of Black	0.011***	0.011***
	(0.002)	(0.002)
Percentage of Hispanic	0.004**	0.003*
	(0.002)	(0.002)
Percentage of Fully-Vaccinated Population	-0.006	-0.033***
	(0.008)	(0.009)
Percentage of Fully-Vaccinated Population (over 65)		0.019***
		(0.004)
Percentage of Uninsured People in the Population	0.020***	0.012**
	(0.005)	(0.005)
Observations	3,103	3,103
AIC	8778.5082	8756.8582
BIC	8857.0298	8841.4199
R2	0.222	0.228
Adjusted R2	0.219	0.225
Residual Std. Error	0.993 (df=3091)	0.990 (df=3090)
F Statistic	80.114*** (df=11; 3091)	75.946*** (df=12; 3090)
Note:		*p<0.1; **p<0.05; ***p<0.01

Model 3a: death rate ~ log(population density) + old + female + log(income) + education + net migration rate + race factors + percent vaccinated + percent vaccinated over age 65 + percent of uninsured + always wearing a mask

Model 3b: Model 3a + (always wearing a mask) + propensity to wear a mask

Health Affair[13] compared the COVID-19 growth rate before and after mask mandates in 15 states and the District of Columbia. It turned out that mask mandates led to a slowdown in daily COVID-19 growth rate, which became more apparent over time.

Based on Model 1 and Model 2, we added the variable of wearing a mask in Model 3a. We have the data of the frequency of wearing a mask including: never wearing, rarely, sometimes, frequently, and always. Since the data is self-reported, and the best way to prevent the spread is by always wearing a face covering, we used the data of always wearing masks. Just like what we thought, always wearing a mask could efficiently decrease the death rate.

In Model 3b, we removed the variable of the percentage of people reported 'always wear a mask', as this variable was included in the variable we created that measures the propensity of the population to wear a mask. Because we thought that summarizing the propensity to wear a mask would be better than only taking the variable of wearing a mask. R-squared, adjusted R-squared and residual standard error did not change in two models, but from AIC, we found that the value of AIC decreased. Hence, we took Model 3b as the better one.

Regression results for Model 3.a and Model 3.b are shown in Table 4.

[13] Wei Lyu, George L. Wehby, Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US, doi: 10.1377/hlthaff.2020.00818, 10.1377/hlthaff.2020.00818

Table 4: Regression results for Model 3.a and Model 3.b

	Dependent variable: Death rate per thousand	
	Model 3.a	Model 3.b
log(Population Density)	0.056***	0.055***
	(0.016)	(0.016)
Percentage of the Old	0.050***	0.049***
	(0.005)	(0.005)
Percentage of Female	0.013	0.012
	(0.009)	(0.009)
log(Income)	-0.208*	-0.197
	(0.123)	(0.123)
Percentage of People without High School Diploma	0.037***	0.036***
	(0.005)	(0.005)
Net Migration Rate	-0.016***	-0.016***
	(0.002)	(0.002)
Percentage of Asian	-0.021**	-0.023**
	(0.009)	(0.009)
Percentage of Black	0.014***	0.014***
	(0.002)	(0.002)
Percentage of Hispanic	0.009***	0.009***
	(0.002)	(0.002)
Percentage of Fully-Vaccinated Population	-0.023**	-0.026***
	(0.009)	(0.009)
Percentage of Fully-Vaccinated Population (over 65)	0.013***	0.014***
	(0.004)	(0.004)
Percentage of Uninsured People in the Population	0.010*	0.009*
	(0.005)	(0.005)
Percentage of People Always Wearing a Mask	-1.106***	
	(0.156)	
Percentage of Mask Wearers (Calculation)		-1.462***
		(0.204)
Observations	3,103	3,103
AIC	8708.8792	8707.8208
BIC	8799.4811	8798.4227
R2	0.240	0.240
Adjusted R2	0.237	0.237
Residual Std. Error	0.982 (df = 3088)	0.982 (df = 3088)
F Statistic	75.077***(df = 14; 3088)	75.184***(df = 13; 3089)
Note:		*p<0.1; **p<0.05; ***p<0.01

Model(s) 4: death rate ~ log(population density) + old + female + log(income) + education + net migration rate + race factors + percent vaccinated + percent vaccinated over age 65 + percent of uninsured + (always wearing a mask) + propensity to wear a mask + rain + (wind speed) + average temperature

With all the families of variables (demographic, vaccination and health care, mask usage and climate), we ran over half a dozen models to determine the best one. We encountered a problem with wind speed, as we had 129 missing observations. We had four models (named "lean"), identical apart from our handling of wind speed; we either removed the missing observations (lean), imputed the missing observations with the country's average (lean 1), imputed the missing observations with the average of the relevant state the county is in (lean 2), and removed the variable altogether (lean 0). We could see that wind speed is a significant factor that increases the number of people dying from COVID-19. We believed that this is tied to the fact that the "Wind Belt", the middle corridor of the US, has been known to take very little action to combat the coming pandemic[14]. Another climate factor is average temperature, which also had (across all models) a positive effect on deaths. This is, in our opinion, due to the fact that counties enjoying warmer climates are more likely to have people outside their homes, not adhering to restrictions. The last of the climate variables is the proportion of rainy days in a year, a variable that also proved to be significant in many of our models (but not all). Again, we were led to believe that the more rainy days, the more people stay in and are inclined to stay in. However, there is a large number of counties with no rainfall (desert counties), where people also might avoid the outdoors, therefore making it less significant in our model.

In all of our models we reached an R-Squared of no less than .244 and no more than .253, meaning the bulk of the variable is still unexplained. There was no clear-cut winner between our models, as all were fairly significant in R-Squared, AIC and BIC. The climate data, especially temperature, proves to be telling, although we suspect there is a mediator variable that is not presented in our models (for instance, the fact that warmer climate counties in the US are wealthier and provide better level healthcare). We decided that the best model was Model 4.lean2, which used the wind speed with a mean state imputation, as it had the lowest AIC and BIC out of the "full" observations models (while the models that removed the counties also enjoyed a high R-Square, and obviously a lower AIC due to fewer observations, we opted to use the model that could predict ALL counties and had the highest R-Square and lowest AIC).

Regression results for Models 4a, 4a.1, 4a.2 and 4b are shown in Table 5.

Regression results for Models 4.lean0, 4.lean, 4.lean1 and 4.lean2 are shown in Table 6.

[14] Madani, Doha(2020). Sturgis Rally May Have Caused More Than 250,000 New Coronavirus Cases, Study Finds, NBC News.
https://www.nbcnews.com/news/us-news/sturgis-rally-may-have-caused-250-000-new-coronavirus-cases-n1239577

Table 5: Regression results for Model 4.a, Model 4.a1, Model 4.a2 and Model 4.b

	Dependent variable: Death rate per thousand			
	Model 4.a	Model 4.a1	Model 4.a2	Model 4.b
log(Population density)	0.071*** (0.018)	0.061*** (0.017)	0.064*** (0.017)	
Population Density per Square Mile				0.0001*** (0.00003)
Percentage of the Old	0.052*** (0.005)	0.049*** (0.005)	0.049*** (0.005)	0.044*** (0.005)
Percentage of Female	0.010 (0.009)	0.013 (0.009)	0.014 (0.009)	0.019** (0.008)
log(Income)	-0.205* (0.124)	-0.252** (0.123)	-0.258** (0.123)	-0.159 (0.124)
Percentage of People without High School Diploma	0.042*** (0.005)	0.042*** (0.005)	0.042*** (0.005)	0.043*** (0.005)
Net Migration Rate	-0.015*** (0.002)	-0.015*** (0.002)	-0.015*** (0.002)	-0.014*** (0.002)
Percentage of Asian	-0.022** (0.009)	-0.021** (0.009)	-0.022** (0.009)	-0.021** (0.010)
Percentage of Black	0.013*** (0.002)	0.013*** (0.002)	0.013*** (0.002)	0.012*** (0.002)
Percentage of Hispanic	0.003 (0.002)	0.004* (0.002)	0.004* (0.002)	0.002 (0.002)
Percentage of Fully-Vaccinated Population	-0.024** (0.010)	-0.021** (0.010)	-0.023** (0.010)	-0.027*** (0.010)
Percentage of Fully-Vaccinated Population (over 65)	0.013*** (0.004)	0.012*** (0.004)	0.012*** (0.004)	0.015*** (0.004)
Percentage of Uninsured People in the Population	0.002 (0.006)	0.003 (0.005)	0.002 (0.005)	-0.005 (0.005)
Percentage of People Always Wearing a Mask	-0.120 (0.433)	-0.977*** (0.159)	-0.951*** (0.159)	-0.032 (0.433)
Percentage of Mask Wearers (Calculation)	-1.175** (0.566)			-1.119** (0.566)
Rainy Days 2020	-0.254* (0.136)	-0.192 (0.124)	-0.199 (0.124)	-0.137 (0.133)
Average Wind Speed 2020 (missing data not replaced)	0.061*** (0.011)			0.055*** (0.011)
Average Wind Speed 2020 (Country mean Imputation)		0.059*** (0.011)		
Average Wind Speed 2020 (State mean Imputation)			0.064*** (0.011)	

Average Temperature	0.013***	0.012***	0.012***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	2,974	3,103	3,103	2,974
AIC	8242.57	8673.28	8669.21	8252.11
BIC	8356.52	8782.00	8777.94	8366.06
R2	0.252	0.250	0.251	0.250
Adjusted R2	0.248	0.246	0.247	0.245

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Regression results for Model 4.lean0, Model 4.lean, Model 4.lean1 and Model 4.lean2

	Dependent variable: Death rate per thousand			
	Model 4.lean0	Model 4.lean	Model 4.lean1	Model 4.lean2
log(Population density)	0.050***	0.071***	0.061***	0.064***
	(0.017)	(0.018)	(0.017)	(0.017)
Percentage of the Old	0.047***	0.052***	0.048***	0.048***
	(0.005)	(0.005)	(0.005)	(0.005)
Percentage of Female	0.011	0.009	0.012	0.013
	(0.009)	(0.009)	(0.009)	(0.009)
log(Income)	-0.211*	-0.206*	-0.241*	-0.247**
	(0.123)	(0.124)	(0.123)	(0.123)
Percentage of People without High School Diploma	0.035***	0.042***	0.041***	0.042***
	(0.005)	(0.005)	(0.005)	(0.005)
Net Migration Rate	-0.016***	-0.015***	-0.015***	-0.015***
	(0.002)	(0.002)	(0.002)	(0.002)
Percentage of Asian	-0.024***	-0.022**	-0.023**	-0.023**
	(0.009)	(0.009)	(0.009)	(0.009)
Percentage of Black	0.013***	0.013***	0.013***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)
Percentage of Hispanic	0.007***	0.003	0.004*	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
Percentage of Fully-Vaccinated Population	-0.017*	-0.024**	-0.024**	-0.025**
	(0.010)	(0.010)	(0.010)	(0.010)
Percentage of Fully-Vaccinated Population (over 65)	0.015***	0.013***	0.013***	0.013***
	(0.004)	(0.004)	(0.004)	(0.004)
Percentage of Uninsured People in the Population	0.004	0.002	0.002	0.002
	(0.005)	(0.006)	(0.005)	(0.005)
Percentage of Mask Wearers (Calculation)	-1.512***	-1.320***	-1.304***	-1.272***
	(0.205)	(0.212)	(0.208)	(0.208)
Rainy Days 2020	-0.236*	-0.256*	-0.199	-0.205*
	(0.125)	(0.136)	(0.124)	(0.124)
Average Wind Speed 2020 (missing data not replaced)		0.061***		
		(0.011)		
Average Wind Speed 2020 (Country mean Imputation)			0.058***	
			(0.011)	
Average Wind Speed 2020 (State mean Imputation)				0.064***
				(0.011)
Average Temperature	0.012***	0.013***	0.012***	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	3,103	2,974	3,103	3,103
AIC	8697.82	8240.64	8671.73	8667.61

BIC	8800.51	8348.60	8780.45	8776.33
R2	0.244	0.252	0.251	0.252
Adjusted R2	0.240	0.248	0.247	0.248

==

Note: *p<0.1; **p<0.05; ***p<0.01

4.Conclusion

According to our model, holding everything else unchanged, one percent increase in population density per square mile would lead to a 6.4% increase in accumulated death rate. Each percentage increase in population age over 65 was associated with a 4.8% increase in fatality rate. Statistically, the impact of household income was significant, with a 24.7% decrease in death rate for each percent increase in median household income in that county. For minority races, a higher proportion of Asian population was correlated with lower death rates, while higher blacks and Hispanics proportion were higher. This is an interesting finding and may worth further analysis for potential reasons. Another finding is the correlation between education level and fatality rate. In our model, one percent increase in population without a high school diploma contributed to a 4.2% increase in accumulated death rates. Explanation behind this phenomenon could be worth further study. In terms of the effects of vaccination, one percent increase in the fully-vaccinated population would result in a 2.5% decrease in the overall death rate caused by COVID-19, providing statistical evidence that vaccinations are effective in mitigating the pandemic. The impact of wearing masks was remarkable and statistically significant. According to our model, one percent increase in mask wearers contributed to 127% decrease in mortality rate.

As we can see in our final models, while we achieved relative success with it, and the rest of our models (R-Squared≈0.25, across all final models), there is a lot more in the sense of unexplained variance. Therefore, we believe that the following steps must be taken to improve this model:

First and foremost, we believe that using aggregated data is not the best course of action. By constructing a time series model, we can truly observe the microchanges on a daily basis. It would also help us understand the full effect of vaccination, as our data is both a bit "early" (before vaccination became widespread) and fails to quantify the full staggering effects and benefits that come along with it.

Recommended by experts around the world, wearing a face cover (preferably a mask) might be one the most effective to stop the spread of virus. However, to accurately estimate the proportion of the population protected by this move is a difficult and daunting task. We sought to quantify this by including a linear transformation of questionnaire results of the frequency of people to wear a mask (on a ranked scale) as the likelihood of people to wear a mask. Although CDC published the guidance for wearing masks and recommended it as a life-saving gesture, many individuals choose not to wear masks in public as a means of protecting themselves and others from COVID-19[15]. In this case, it is definitely difficult to get a precise number to picture this propensity or frequency for the public to wear a mask in each county.

[15] Gillespie, Claire (2021). Why Do Some People Refuse to Wear a Face Mask in Public? Health, https://www.health.com/condition/infectious-diseases/coronavirus/face-mask-refuse-to-wear-one-but-why

Furthermore, it is our opinion that transportation plays a key role in the spread of COVID-19 and its death toll. The virus can be spread among the population via public transportation. According to the CDC, masks are required on planes, buses, trains, and all other forms of public transportation traveling into, within, or out of the United States, and in U.S. transportation hubs such as airports and stations, ever since February 2, 2021[16]. Due to data limitations, we did not include public transportation factors into our model. Further analysis could consider measuring these impacts by adding regressors like number of flights, length of highways, degree of crowdedness in trains and subways, if possible.

Finally, we are aware that collinearity problems are likely to exist in our model. We attempted to combat this with the limitation of mask usage to one variable, but there is reason to suspect that the median household income in each county can also be affected by the other features in that county. For example, Card[17] conducted a convincing analysis of the causal link between years of schooling and earning. Bayard et al.[18] found that approximately one-half of the sex gap in wages remains attributable to the individual's sex. Bielby[19] measured workplace race bias and concluded that blacks and Hispanics earn, respectively, 22% and 32% less than white. We are still searching for better instrumental variables to decrease correlations between regressors and solve the omitted variable problem.

[16] Protect Yourself When Using Transportation, 2021 Published online at Center for Disease Control and Prevention. Retrieved from:
https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/using-transportation.html
[17] Card, David. (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. https://doi.org/10.3386/w4483
[18] Kimberly Bayard, Judith Hellerstein, David Neumark, & Kenneth Troske. (2003). New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee‐Employer Data. Journal of Labor Economics, 21(4), 887–922. https://doi.org/10.1086/377026
[19] William T. Bielby. (2000). Minimizing Workplace Gender and Racial Bias. Contemporary Sociology (Washington), 29(1), 120–129. https://doi.org/10.2307/2654937

References

- Avnish Sandhu, Steven J. Korzeniewski, et al., Elevated COVID19 mortality risk in Detroit area hospitals among patients from census tracts with extreme socioeconomic vulnerability, EClinicalMedicine, Volume 34, 2021, 100814, ISSN 2589-5370.

- Bhadra, A., Mukherjee, A. & Sarkar, K, Impact of population density on Covid-19 infected and mortality rate in India. Model. Earth Syst. Environ. 7, 623–629 (2021)

- Card, David, Using Geographic Variation in College Proximity to Estimate the Return to Schooling, 1993

- County level data

- County-level FIPS code and state-level FIPS code, Detailed information

- Gillespie, Claire, Why Do Some People Refuse to Wear a Face Mask in Public? Health, 2021

- Gina Cahill, Carleigh Kutac, & Nicholas L. Rider, Visualizing and assessing US county-level COVID19 vulnerability, American Journal of Infection Control, 2020, ISSN 0196-6553.

- Josh Katz, Margot Sanger-Katz & Kevin Quealy, A Detailed Map of Who Is Wearing Masks in the U.S, New York Times, 2020

- Kimberly Bayard, Judith Hellerstein, David Neumark, & Kenneth Troske, New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data, 2003 Journal of Labor Economics, 21(4), 887–922.

- Madani, Doha, Sturgis Rally May Have Caused More Than 250,000 New Coronavirus Cases, Study Finds, NBC News, 2020

- Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina and Joe Hasell, Coronavirus Pandemic (COVID-19), 2020 Published online at OurWorldInData.org. Retrieved from 'our world in data'

- Protect Yourself When Using Transportation, 2021 Published online at Center for Disease Control and Prevention. Retrieved from the CDC website

- The geographic regions and county-level FIPS code and state-level FIPS code: New York Times data of COVID 19

- Wei Lyu, George L. Wehby, Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US, 2020, doi: 10.1377/hlthaff.2020.00818, 10.1377/hlthaff.2020.00818

- William T. Bielby, Minimizing Workplace Gender and Racial Bia, 2000, Contemporary Sociology (Washington), 29(1), 120–129.

- World Health Organization data of United States