Exploration of Potential Risk Factors For Texas County COVID-19 Infection Rates

Kahsai, Keren; Machiorlatti, Michael
Department of Population Health and Biostatistics, University of Texas Rio Grande Valley School of Medicine, Edinburg, Texas, USA

Acknowledgements
I would like to express my deep gratitude for Dr. Michael Machiorlatti for providing valuable advice, suggestions, and guidance throughout the entirety of this project.

Abstract

There have been over 2 million COVID-19 cases within the United States. When considering where infections take place, there appears to be variability between states and even between counties (Gardner 2020). To understand the reasons underlying this phenomenon, we used Texas county data on Covid-19 infection rates and utilized auxiliary data such as age, race, gender, diabetes and obesity rates, temperature, humidity, median household income, metropolitan or rural designation, and poverty rates to see what confers greater risk for higher total infection at a county level. Our study found a positive relationship between diabetes and obesity rates and COVID-19 infection rates, and a negative relationship between the rates of lower educational status and COVID-19 infection rates. Further studies should investigate the underlying mechanisms regarding why those with diabetes, obesity, or a High School Diploma or less are more susceptible to COVID-19 infection as well as the role that population density and different stay-at-home policies affected infection rates in each county.

Intro

Identified in Wuhan, China near the end of 2019, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the virus responsible for Coronavirus Disease 2019 (COVID-19). As an RNA, beta coronavirus within the same subgenus of viruses responsible for Severe Acute Respiratory Syndrome (SARS) and Middle East respiratory syndrome (MERS), SARS-Cov-2 binds to the angiotensin-converting enzyme 2 (ACE2) and primarily causes respiratory illnesses. Currently, SARS-CoV-2 is transmitted person-to-person through direct contact, respiratory droplets, and potentially airborne (McIntosh, 2020). As a result, globally, there have been more than nine million confirmed cases, causing close to five hundred deaths, with the United States contributing 2.43 million cases and 124,000 deaths (Gardner, 2020). Within Texas, there has been variability regarding infection rates within the counties, potentially due to differences in the following risk factors: age, race, median household income, poverty rates, temperature, humidity, gender breakdown, underlying health conditions, metropolitan or rural designation, and education level. The goal of this study is to quantify what variables specifically confer greater risk for higher county-level infection rates.

Methods & Materials

To analyze whether Texas county risk factors affected the infection rates, county level data and the number of COVID-19 cases were collected from a number of online, public databases. COVID-19 cases were found from usafacts.org, a private company that retrieved their information from the Texas Department of State Health Services (Texas Department of State Health Services, 2020). Race breakdown was found on the United States Census Bureau website from the 2010 Decennial Census (United States census Bureau, 2010). Race was classified as White, Black, Asian, American Indian, Pacific, two or more races, or other. Income and poverty levels per Texas county was also retrieved from the United States Census Bureau through the Small Area Income and Poverty Estimate program in 2018 (United States census Bureau, 2018).

Gender breakdown was found through the American Community Survey on the United States Census Bureau in 2018 (United States Census Bureau, 2018). Temperature data was found on usa.com from the American Community Survey from 2010-2014 (Usa.com, 2020). Humidity data was found on usa.com from the American Community Survey from 2010-2014 (Usa.com, 2020). The metropolitan/rural designation was found from Texasagriculture.gov from the Texas state office of rural health, office of rural affairs, and Texas department of agriculture in 2012 (Texas Department of Agriculture, 2012). The Diabetes and Obesity was found from the Centers of Disease Control and Prevention in 2016 within the United States Diabetes Surveillance System (Centers for Disease Control and prevention, 2016).

Education levels for each Texas county was found from the American Community Survey 2014-2018 through the United States Department of Agriculture Economic Research Service (United States Department of Agriculture Economic Research Service, 2018). Education levels were classified as less than high school, High school or less, and Bachelors or higher Age breakdown was found from http://healthdata.dshs.texas.gov/ which gathered its information from the Texas Department of State Health Services, Center for Health Statistics, Agency Analytics Unit in 2015 (Texas Department of State Health Services, 2015). Age was categorized by those under five years of age, 5-14 years, 15-44 years, 25-64 years, and over 65 years.

Minitab version 18 software was used to calculate the infection rates and perform the analyses. All data was at the county level. The race data was further refined by looking at the percentage of the population that was White versus non-white (WhitePercent). Poverty levels were further defined by identifying the percent of the population in the county that lived under the poverty line (poverty percent). Gender data was further defined by identifying the male percentage of the population (malepercent). Diabetes and obesity burden on the county was further refined by diabetes and obesity percent (diabetespercent and obesitypercent). The age breakdown was refined to an older percent (olderpercent), referring to those above 65, and younger percent (youngerpercent), or those younger than 5. Education was further broken down to the percent of those with a High school diploma or less (HighSchoolorless).

The mean, standard deviation, minimum, maximum, and median were calculated for the following continuous variables: PercentWhite, poverty percent, Income level of the county, malepercent, temperature and humidity, diabetespercent and obesitypercent, HighSchoolorless, and olderpercent (Table 1). A bar graph was completed for the categorical variable of whether a county was metro or rural (Figure 1). The Pearson Correlation coefficient was calculated for all of the continuous variables to measure the linear association with infection rates (Table 2). Finally, multiple linear regression was used to measure the association between each covariate and infection rate (Table 3). To create a final model, a selection criterion of p-value < 0.25 was utilized (Table 4). From the set of variables that met this inclusion criteria, interaction terms and quadratic terms where then explored and final model was then created (Table 5).

Results

Table 1-Descriptive Statistics for Continuous Covariates

Variable	Min	Max	Mean	Median	SD
PercentWhite	0.52	0.96	0.82	0.83	0.083
povertypercent	0.02	1.64	0.16	0.15	0.105
Income	61.00	102858.00	50918.00	48542.00	12956.000
malepercent	0.06	3.85	0.51	0.50	0.216
Temperature	48.40	73.60	60.39	60.10	5.182
Humidity	0.73	0.82	0.77	0.77	0.020
diabetespercent	0.00	8.77	0.13	0.08	0.600
obesitypercent	0.02	20.32	0.34	0.23	1.313
olderpercent	0.02	1.63	0.19	0.17	0.124
Fatality Rate	0.00	72.73	2.23	0.00	6.845
Infection Rate	0.00	5.28	0.02	0.00	0.333
HighSchoolorless	0.03	28.69	0.50	0.35	1.834
olderpercent	0.02	1.63	0.19	0.17	0.124

Table 1 displays the descriptive statistics of all of the continuous variables. On average, the counties were mostly white; most counties held lower levels of diabetes and obesity rates. Overall, the counties demonstrated low percentages of residents over the age of 65. While the humidity levels were very similar, there was a wide range of average temperatures among the counties. Additionally, there also appeared a wide range within the education category amongst the counties. The gender makeup of each county was about equal.

Figure 1-Bar graph of Texas counties classified as metro or rural

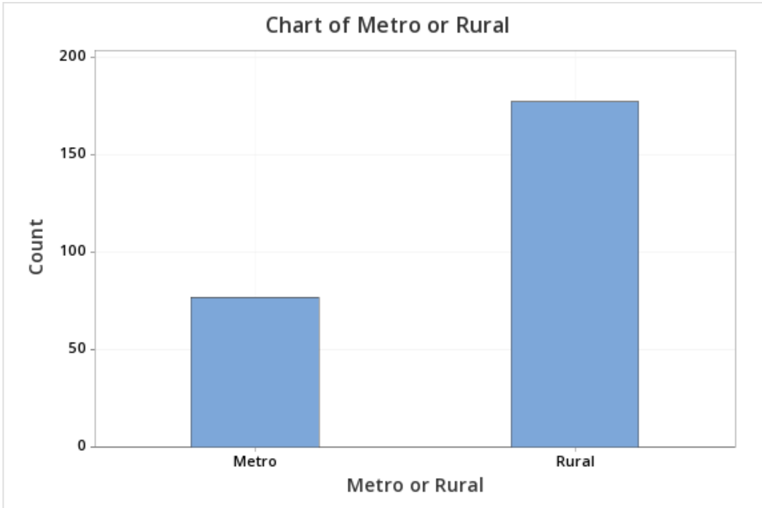

Figure 1 displays the number of counties listed as a metropolitan or rural county, with 177 (70%) counties are rural. Whereas 77 (30%) counties are metropolitan.

Table 2-Pearson Correlation Coefficient - Continuous Covariates and Infection Rate

Variables	Infection Rate
PercentWhite	-0.009
Income	-0.025
povertypercent	-0.013
malepercent	-0.007
Temperature	-0.047
Humidity	-0.014
diabetespercent	0.075
obesitypercent	0.071
HighSchoolorless	0.065
olderpercent	-0.002
youngerpercent	-0.002

To analyze the relationship between infection rate and the other continuous variables, the Pearson correlation coefficient was completed, seen in Table 2. Due to the low Pearson Correlation Coefficients, there did not appear to be a strong relationship among any of the covariates and infection rate.

Table 3-Regression Results – All Covariates with Infection Rate

Term	Coef	T-Value	P-Value
Constant	0.07	0.07	0.942
Income	-0.00	-0.46	0.643

povertypercent	-0.22	-0.34	0.733
malepercent	0.04	0.08	0.940
Temperature	-0.00	-0.48	0.635
Humidity	0.23	0.20	0.840
diabetespercent	1.04	1.97	0.050
obesitypercent	0.35	1.39	0.165
HighSchoolorless	-0.55	-2.35	0.019
olderpercent	-0.00	0.00	0.996
PercentWhite	-0.03	-0.11	0.912
Metrorural	0.05	0.93	0.356

These results were used to screen for the variables that had a p-value of less than 0.25, leaving percentDiabetes, percentObesity, and HighSchoolorLess. Additionally, the R^2-Adj value was 3.57%.

Table 4- Regression Results – Covariates Remaining With Cutoff of Screened With p-value < 0.25

Term	Coef	T-Value	P-Value
Constant	0.04	1.65	0.100
diabetespercent	0.87	1.80	0.073
obesitypercent	0.19	0.90	0.369
HighSchoolorless	-0.39	-2.04	0.042

The linear regression was run again, only including the 3 variables that produced a p-value less than 0.25 and exploring interactions and quadratic terms.

Table 5- Final Regression Results – Considering Interactions and Quadratic Terms

Term	Coef	T-Value	P-Value
Constant	-0.46	-20.86	0.000
diabetespercent	-0.68	-1.57	0.117
obesitypercent	1.10	4.43	0.000
HSorless	0.97	6.24	0.000
diabetespercent*diabetespercent	-18.37	-3.08	0.002
obesitypercent*obesitypercent	-4.17	-2.92	0.004
HighSchoolorless*HighSchoolorless	0.42	0.99	0.322
diabetespercent*obesitypercent	31.71	6.29	0.000
diabetespercent*HSorless	-7.00	-2.38	0.018
obesitypercent*HSorless	-1.96	-1.38	0.170

Table 5 demonstrates interactions between the variables in regards to the infection rate. The R^2-Adj value increased to 80.27%.

Discussion

From the initial regression, the R^2-Adj value was 3.57%, demonstrating that only 3.57% of the infection rate could be explained by all of the variables. However, there were three variables that had a p-value of less than 0.25: percentdiabetes, percentobesity, and HSorLess. From these results, another ANOVA was run, only incorporating the three significant variables, in which the R squared value increased to 80.27%, indicating that 80.27% of the infection rate could be explained by those three significant variables. Additionally, the p-values for those three variables was less than 0.05, indicating a relationship between those three variables and infection rate. Also looking at the interactions between those variables, there appears to be a relationship between having diabetes and obesity and increasing infection rates as well as having diabetes and less than a High school diploma and increases in infection rate. The results found in this study aligns with the current literature as older men with comorbidities tend to have more severe cases of COVID causing higher fatality rates (Jian, 2020). However, the current literature also supports a negative relationship between temperature and humidity and infection and fatality rates (Bhardwaj, 2020), negative between lower income countries and infection rates (Reeves 2020), and higher cases among minorities (*Racial and Ethnic Minority Groups, 2020).* This data does not align with other conclusions and hypotheses found in the literature.

One limitation of the study is that only race was taken into consideration and not ethnicity. Therefore, those who are Hispanic and identify as White were lumped into the White category. This could affect the overall results. Additionally, one aspect that this study does not take into consideration is that each county began its stay-at-home orders at different times, affecting infection rates. It should also be noted that population density could also affect infection rates, which this study did not take into consideration. Further studies should consider the effects of these variables on fatality rates as well as the underlying mechanisms implicating diabetes and obesity and lower education levels in higher infection rates. This information could be valuable when the second COVID-19 wave hits, identifying vulnerable populations ahead of time and ensuring that those populations are protected.

References

Bhardwaj, Rajneesh. (8 June 2020). Likelihood of survival of coronavirus in a respiratory droplet deposited on a solid surface. AIP *Physics of Fluids, 32. https://doi.org/10.1063/5.0012009*

Centers for Disease Control and Prevention. (2016). Diagnosed Diabetes. Retrieved from https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html#

Gardner, L. (2020, February 19). *COVID-19 United States Cases by County*. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/us-map

Jian-Min, Jin. (29 April 2020). Gender Differences in patients with COVID-19: Focus on Severity and Mortality. *Frontiers in Public Health, 8. https://doi.org/10.3389/fpubh.2020.00152*

McIntosh, K. (2020). Coronavirus disease 2019 (COVID-19):ble from Epidemiology, virology, and prevention. In: T.W. Post, M. Hirsch, A. Bloom (Eds). *UpToDate.* Available from https://www-uptodate-com.ezhost.utrgv.edu/contents/coronavirus-disease-2019-covid-19-epidemiology-virology-and-prevention?search=covid-19&topicRef=128323&source=see_link

(25 June 2020*). Race and Ethnic Minority Groups.* Coronavirus Disease 2019. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html

Reeves, Richard. (27 March 2020). *Class and Covid: How the less affluent face double risks.* Brookings. https://www.brookings.edu/blog/up-front/2020/03/27/class-and-covid-how-the-less-affluent-face-double-risks/

(2020). Texas Average Humidity County Rank. USA.com. http://www.usa.com/rank/texas-state--average-humidity--county-rank.htm

(2020). Texas average max. temperature county rank. Usa.com. http://www.usa.com/rank/texas-state--average-max-temperature--county-rank.htm?yr=9000&dis=&wist=&plow=&phigh=

Texas Department of Agriculture. (2012, April). Texas County Designations. Retrieved from https://www.texasagriculture.gov/Portals/0/forms/ER/Rural-Metro%20Counties.pdf

Texas Department of State Health Services. (2020, June 8). *Cases Over Time by County*. Retrieved from https://www.dshs.state.tx.us/coronavirus/additionaldata/

Texas Department of State Health Services. (2015). Health Facts Profile. Retrieved from http://healthdata.dshs.texas.gov/dashboard/surveys-and-profiles/health-facts-profile-2014-2015

Texas Department of State Health Services. (2020, June 8). *Fatalities Over Time by County*. Retrieved from https://www.dshs.state.tx.us/coronavirus/additionaldata/

United States Census Bureau. (2010, April 1). Summary of Modified Race and Census 2010 Race Distributions for the United States. Retrieved from https://www.census.gov/data/datasets/2010/demo/popest/modified-race-data-2010.html

United States Census Bureau. (2019, December). SAIPE State and County Estimates for 2018**. Retrieved from** https://www.census.gov/data/datasets/2018/demo/saipe/2018-state-and-county.html

United States Census Bureau. (2018) Age and Sex. (**Retrieved from** https://data.census.gov/cedsci/table?q=S01&d=ACS%201-Year%20Estimates%20Subject%20Tables&tid=ACSST1Y2018.S0101

United States Department of Agriculture Economic Research Service. (2018). Completion rates, 2014-2018. Retrieved from https://www.ers.usda.gov/data-products/county-level-data-sets/