

A Multiple Linear Regression Analysis of Rural-Urban COVID-19 Risk Disparities in Texas

Amber K. Luo^{1*}, Sophia Zhong², Charles Sun³, Jasmine Wang⁴, and Alexander White^{5*}

¹ *Ward Melville High School, East Setauket, New York, USA*

² *Westwood High School, Austin, Texas, USA*

³ *Westlake High School, Austin, Texas, USA*

⁴ *Seven Lakes High School, Katy, Texas, USA*

⁵ *Department of Mathematics, Texas State University, San Marcos, Texas, USA*

*Corresponding author

A Multiple Linear Regression Analysis of Rural-Urban COVID-19 Risk Disparities in Texas

Abstract

As the number of COVID-19 cases in the U.S. rises, the differential impact of the pandemic in urban and rural regions becomes more pronounced, and the major factors relating to this difference remain unclear. Using the 254 counties of Texas as units of analysis, we utilized multiple linear regression to investigate the influence of 83 county-level predictor variables including race demographics, age demographics, healthcare and financial status, and prevalence of and mortality rate from COVID-19 risk factors on the incidence rate and case fatality rate from COVID-19 in Texas on September 15, 2020. Here, we report that urban counties experience, on average, 41.1% higher incidence rates from COVID-19 than rural counties and 34.7% lower case fatality rates. Through comparisons between our models, we found that this difference was largely attributable to four major predictor variables: namely, the proportion of elderly residents, African American residents, and Hispanic residents, and the presence of large nursing homes. According to our models, counties with high incidence rates of COVID-19 are predicted to have high proportions of African American residents and Hispanic residents coupled with low proportions of elderly residents. Furthermore, we found that counties with the highest case fatality rates are predicted to have high proportions of elderly residents, obese residents, and Hispanic residents, coupled with low proportions of residents ages 20-39 and residents who report smoking cigarettes. In our study, major variables and their effects on COVID-19 risk are quantified, highlighting the most vulnerable populations and regions of Texas.

Keywords: COVID-19, multiple linear regression, incidence rate, case fatality rate, rural-urban health disparities

Introduction

The COVID-19 (Coronavirus Disease 2019) pandemic continues to exert major effects on healthcare systems and on the well-being of residents in both urban and rural regions. Texas is one of the hardest hit states by COVID-19, experiencing alarming surges in confirmed cases, hospitalizations, and deaths starting in mid June of 2020. As of September 15, 2020, there have been 15,785 deaths and 668,746 confirmed cases of COVID-19 across the 254 counties in Texas [1].

Previous studies have shown that urban counties in Texas experience significantly higher

incidence rates and lower case fatality rates from COVID-19 when compared to rural counties [2]. Rural communities experience lower testing rates and higher positive test rates for COVID-19 [3], along with fewer financial resources [4], limited access to healthcare [5], and higher all-cause mortality rates in comparison to urban communities [6]. Furthermore, many hospitals in rural communities lack the ICU beds and primary care physicians necessary to deal with a local COVID-19 outbreak [7]. Rural counties also have higher prevalences of CDC (Centers for Disease Control and Prevention)-defined COVID-19 risk factors [8], including cardiovascular disease, chronic respiratory disease, obesity, and smoking, as well as higher rates of poverty and lower levels of physical activity [9]. Of those hospitalized with COVID-19, 75% have some underlying medical condition, regardless of age, such as diabetes, chronic respiratory diseases, and cardiovascular diseases [10]. Furthermore, rural and urban counties have many demographic differences, including a higher proportion of elderly and black residents in rural counties, populations that are thought to be especially vulnerable to COVID-19 [11]. These health and demographic disparities, among others, hold the key to understanding the differential impact of COVID-19 on urban and rural counties. Identifying the variables responsible for these differences and quantifying their effects on measures of COVID-19 risk will allow policy makers and healthcare professionals to determine the most suitable plans for COVID-19 based on the predicted vulnerability of a region.

Currently, few studies have investigated the relation of demographic differences in rural and urban counties to COVID-19 risk. Our study is the first to use multiple linear regression to definitively quantify the relationship of population characteristics to urban-rural COVID-19 disparities by creating models that explain the effect of various characteristics (predictor variables) on a single measure of COVID-19 risk (response variable). These multiple linear regression models were run for on two response variables: incidence rate from COVID-19 and case fatality rate from COVID-19. The county-level predictor variables we analyzed include, but are not limited to, prevalence of and mortality rate from CDC-defined COVID-19 risk factors, age demographics, race demographics, financial and living conditions, political leaning, and potential super-spreading sites. Our study highlights the most significant county-level variables in predicting COVID-19 incidence rate and case fatality rate and identifies the major underlying variables responsible for urban-rural disparities in COVID-19 risk.

Data and Methods

Data Sources

The units of analysis for this study are $N = 254$ counties in Texas. The COVID-19 Data Hub 2.2.0, accessed through the COVID19 Package in R, provided cumulative county-level reports of COVID-19 confirmed cases and deaths. Data was filtered to a 7 day window around September 15 (9/12 - 9/19), and the 7 day rolling averages of COVID-19 incidence rate and case fatality rate were calculated from this window. The 83 predictor variables and their sources can be grouped into the following four categories:

1. Population demographics: County-level data on percent population by five-year age strata from 0-85 years of age, race, education status, health insurance status, sex, unemployment status, and poverty status, as well as data on median income, population density per square mile, and the percentage of healthcare workers per county were obtained from the 2018 American Community Survey and the 2020 County Health Rankings [12, 13].
2. Health-related risk factors: County-level data on the prevalence of diabetes, cigarette smoking, obesity, and hypertension, as well as the mortality rate per 100,000 population from chronic respiratory disease, cardiovascular disease, alcohol use disorders, and drug use disorders were obtained from the IHME (Institute for Health Metrics and Evaluation) GHDx (Global Health Data Exchange) US Health Datasets. Data on survey responses on mask use, where participants from all counties in the US were asked to self-report how often they wear masks in public from 7/2/2020 - 7/14/2020, was obtained from the New York Times [14, 15].
3. Potential superspreading sites: Three discrete variables were created to examine the true effect of sites generally thought to have a higher incidence of COVID-19 on COVID-19 risk. These variables represented the presence of one or more meatpacking facilities, the presence of one or more prisons, and the presence of one or more large nursing homes (population >100) by county from Niche Meat Processing, Texas Almanac, and Texas Health and Human Services, respectively [16, 17, 18].
4. Political leaning: Due to the heavily politicized climate surrounding the COVID-19 pandemic, several measures of political leaning were collected. Data on the percentage of voters for Trump/Clinton in the 2016 presidential election and the percentage of voters for O'Rourke/Cruz in the 2018 Texas Senate election was obtained from the New York Times to estimate the true political leanings of the counties in Texas [19, 20].

All statistical analyses, models, and mapping of data were done using R version 1.3.959. A complete list of the predictor and response variables analyzed and their sources and dates can be found in Table 1.

Statistical Analysis

Prior to model creation, counties were grouped into quintiles (1-5) by measures of risk: COVID-19 incidence rate per 100,000 population and case fatality rate, both calculated from 7-day rolling averages. Medians and IQRs were calculated for select predictor variables by quintile, and their significance in relation to COVID-19 risk was determined using Kruskal-Wallis tests, with a p -value of < 0.05 considered significant. These tables of COVID-19 risk quintiles (Table 2, 4) can be compared with a study by Khose, Moore, and Wang [2] using data from April 8, 2020.

Models

Multiple Linear Regression (MLR) was used to model the effect of different subsets of predictor variables on each of the response variables. To investigate the difference between urban and rural counties, all models include the *Urban* variable, which denotes the urban-rural status of the county.

$$y_i = \beta_0 + \beta_1 Urban_i + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i \quad (1)$$

y_i represents either incidence rate or case fatality rate from COVID-19 for the i th county, and x_{ij} represents the value of the j th predictor variable for the i th county. The errors, $\epsilon_i \sim N(0, \sigma^2)$, are assumed independent.

The best subset of predictors for each response variable was chosen in two phases. In phase one, using the LEAPS package in R, all possible subsets of our 83 predictor variables were compared using the Bayesian information criteria. The optimal Box-Cox transformation was then fit to the best four models. In each case, the 95% confidence interval for the optimal power included 0, which suggested the natural log transformation. In phase two, the LEAPS package in R was used once more with the transformed response variable. The best four models for each response were then compared for strength of prediction, as determined by adjusted R^2 and F -statistic; validity of prediction, as determined by normally distributed residual plots; and significance, as determined by the p -values of the variables included in the model and ANOVA tables. To assess possible multicollinearity, VIF (variance of inflation) factors were calculated for

all predictors in our MLR models using the `car` package in R. In each case, the VIF never exceeded 2.744, indicating a lack of variable redundancy (threshold: $VIF=5$).

Results

Incidence Rate Analysis

To elucidate the basic nature of the predictor variables in relation to COVID-19 risk and to provide for a deeper understanding of the coefficients in our MLR models, we analyzed the medians and IQRs of certain predictor variables by COVID-19 risk quintiles. Table 2 presents the distribution of county-level demographics and health outcomes by quintiles of COVID-19 incidence rates in Texas on September 15, 2020. Counties with higher incidence rates had generally higher median percentages of residents with diabetes (p -value: $< .001$), African American residents (p -value: $.002$), Hispanic residents (p -value: $< .001$), obese residents (p -value: $< .001$), residents 20-39 years of age (p -value: $< .001$), unemployed residents (p -value: $.078$) and generally higher population density per square mile (p -value: $< .001$). These counties also had generally lower median percentages of Caucasian American residents (p -value: $.002$) and elderly residents (p -value: $< .001$, > 70 years of age). In particular, the fifth quintile of COVID-19 incidence rate (> 2816.3 per 100,000 population) had the highest median percentage of residents beneath the poverty line (median: 16.6%; p -value: $.002$) and the lowest percentage of residents with health insurance (median: 97.7%; p -value: $.012$).

These data identify significant variables in relation to COVID-19 incidence rate and illustrate the trend of these variables as the COVID-19 incidence rate increases. To further analyze these trends, multiple linear regression models were run. For each response variable (incidence rate — IR; case fatality rate — CFR), a base model was run: a simple linear regression of the form

$$\ln(\hat{y}_i) = \beta_0 + Urban * x_1 \quad (2)$$

where the discrete variable *Urban* is 1 for urban counties and 0 for rural counties, in which the urban/rural status of counties is defined by the US Office of Management and Budget [21]. Table 3 presents the base, full, and rural-only incidence rate models. Three counties (Fig. 1, left map; Loving: left; Borden: center; King: right) with no confirmed cases were removed to allow for the natural log transformations. The base model regresses $\ln(IR)$ on *Urban*, providing a control to mathematically establish both the significance and nature of a county's urban-rural status when *Urban* is the only predictor in consideration. Changes in the p -value of *Urban* as a result of adding more predictor variables to the base model highlight the importance

of certain predictor variables in determining urban-rural COVID-19 risk disparities.

The IR base model predicts an incidence rate of 0.0151 in a rural county and 0.0213 in an urban county, not accounting for any other variables: a 41.1% increase. Of note, *Urban* is a highly significant variable in the base model (p -value: $< .001$).

According to our full incidence rate model, counties with high percentages of Hispanic and/or African American residents are expected to experience significantly higher incidence rates of COVID-19, as are counties with one or more large nursing homes (> 100 residents). *Hispanic* is the most significant variable in the model, with a very small p -value (p -value: $< .001$) and the largest impact on the R^2 upon removal (Table 9; -19.3%). Counties with one or more large nursing homes are particularly at risk for COVID-19; specifically, a county with one or more large nursing homes ($Nursing\ Homes = 1$) is expected to experience a 40.6% higher incidence rate from COVID-19 than a similar county without any large nursing homes. These data underscore the importance of outbreaks in nursing homes, establishing the presence of large nursing homes as a major factor affecting the incidence rate of COVID-19 in a particular county. *Elderly* is the only predictor variable in the model with a negative coefficient, indicating that counties with a higher percentage of elderly (> 70 years of age) residents are expected to have lower COVID-19 incidence rates. *Black* is the variable with the largest positive coefficient; however, the standard deviation of the percentage of black residents (Table 7; 0.066) is lower than the standard deviation of Hispanic residents (Table 7; 0.233), so that *Black* has a lesser impact on $\ln(\hat{IR})$ than *Hispanic*.

The objective of these MLR models is to analyze urban-rural differences, hence the presence of the variable *Urban* in both the base and full models. The p -value of *Urban* was significant in the base model and insignificant in the full model; adding the variables *Elderly*, *Nursing Homes*, *Hispanic* and *Black* increased the p -value of *Urban* from $< .001$ to 0.957, implying that the increased incidence rate of COVID-19 in urban counties is largely attributable to differences in the percentages of elderly, Hispanic, and black residents, and by the presence of large nursing homes. Adding the four aforementioned variables also hugely improved the adjusted R^2 (0.052 to 0.448) and F statistic (14.59 to 41.61).

Full models were filtered to include only rural counties to observe any unusual shifts in coefficients or p -values of the variables in the full model. As expected, little difference was observed between the full and rural IR models. *Elderly* is slightly less significant in the rural model (.020 vs. .046) and *Black* has a considerably larger coefficient (2.553 vs. 3.261). No variables changed signs and there were no major changes in significance, as is consistent with the insignificance of *Urban* as a predictor variable in the full model (p -value: .957; Table 9; $-.001\%$).

Case Fatality Rate Analysis

Table 4 presents the distribution of county-level demographics and health outcomes by quintiles of COVID-19 case fatality rate in Texas on September 15, 2020. The median percentages of obese residents, unemployed residents, and diabetic residents follow generally positive trends, with medians increasing as they approach the third quintile of case fatality rate (1.9% – 2.9%) and deviating as they approach the fourth and fifth quintiles (> 2.9%). The percentage of elderly residents follows a strictly positive trend from the second quintile to the fifth quintile, with the first quintile deviating from this pattern. The fifth quintile of case fatality rate (> 4.2%) has the highest median percentage of elderly residents (median: 14.3%; p -value: < .001), lowest median percentage of Hispanic residents (median: 23.5%; p -value: .079), and lowest median percentage of residents ages 20-39 (median: 21.2%; p -value: < .001), in contrast with the low median percentage of elderly residents (median: 9.3%; p -value: < .001), high median percentage of Hispanic residents (median: 50.2; p -value: < .001), and high median percentage of residents ages 20-39 (median: 27.2; p -value: < .001) in the fifth quintile of COVID-19 incidence rate.

Table 5 presents the base, full, and rural-only CFR models. 28 counties with no deaths as of September 15, 2020 were not included in the CFR models to allow for a natural log transformation on the response variable. The immediate takeaway from the CFR base model is that rural counties experience higher case fatality rates from COVID-19 as opposed to lower incidence rates from COVID-19. The base model predicts a case fatality rate of 1.94% in a urban county and 2.97% in an rural county: a 53.1% increase. Once again, *Urban* is highly significant in the base model (p -value: < .001).

From our full CFR model, counties with a high percentage of elderly residents are predicted to experience a higher case fatality rate from COVID-19, as are counties with a high percentage of obese residents. Counties with a high percentage of residents ages 20-39 and, interestingly, counties with a high percentage of smokers are predicted to have lower case fatality rates. The coefficient of *Nursing Homes* is negative (-0.210), indicating that a county with one or more large nursing homes is expected to experience a 18.9% lower case fatality rate from COVID-19 than a similar county without any large nursing homes.

To provide a basis for direct comparison between case fatality rate and incidence rate of COVID-19, a model was created that regressed $\ln(C\hat{F}R)$ on the predictor variables from the full incidence rate model (*Urban*, *Elderly*, *Nursing Homes*, *Hispanic*, *Black*; Table 6).

The following changes were observed: 1) The coefficient of *Elderly* reversed signs, increased considerably in magnitude (IR: -2.324 to CFR: 8.079), and had a much larger impact on the R^2 upon

removal (Table 9; IR: -1.21% vs. CFR: -11.7%), 2) The coefficient of *Nursing Homes* reversed signs (IR: 0.343 vs. CFR: -0.251), 3) *Hispanic* became a much less significant variable (Table 9; IR: -19.3% vs. CFR: -3.49%), and 4) *Black* was not significant in the case fatality rate model (p -value: $.527$). Of note, the coefficient of *Urban* remained high (IR: 0.957 vs. CFR: 0.945), indicating that both the difference in incidence rate and the difference in case fatality rate between urban and rural counties is largely attributable to these four variables.

To visualize model performance and identify region-specific variations in prediction, we created residual maps. Fig. 1 presents the residuals of the 251 counties in the full incidence rate model (left) and the 226 counties in the full case fatality rate model (right). Of note, the full IR model generally overestimates the incidence rate of COVID-19 in the urbanized regions of the Coastal Plains, while the full CFR model underestimates the case fatality rate in these same regions. The opposite is true for counties in the Mountains and Basins region, for which the full IR model underestimates the incidence rate and the full CFR model overestimates the case fatality rate.

The full CFR model was filtered to include only rural counties to observe any unexpected differences. Similar to the rural-only incidence rate model, no variables changed signs or shifted considerably in magnitude, but *Elderly* (p -value: $.069$) and *Nursing Homes* (p -value: $.058$) were no longer significant when including only rural counties in the CFR model.

Discussion

Our study highlights a concerning disparity between urban and rural counties, consistent with several other studies, wherein urban counties experience higher incidence rates, yet rural counties experience higher case fatality rates [2, 22]. Moreover, our study is the first to identify Hispanic & African American populations, elderly populations, and large nursing homes as major contributing factors to this urban-rural disparity in Texas. Knowing these important determinants of county health and their quantitative effect on COVID-19 incidence rate and case fatality rate makes it possible to predict which counties are at the highest risk for serious effects from COVID-19. As the counties with the highest levels of variables that positively impact the predicted COVID-19 case fatality rate are also those with lesser health resources and higher prevalence of health-related COVID-19 risk factors, it is imperative that traditionally vulnerable populations in such counties, including the elderly and immunocompromised, receive adequate essential resources from local providers and health agencies. Furthermore, the alarming positive effect of elderly populations on case fatality rate stresses the need to insulate and closely monitor

counties with high percentages of elderly people. Equally concerning is the observation that rural counties, which are predicted to experience higher case fatality rates due to their high percentage of elderly residents and obese residents, lack the health resources and infrastructure of urban counties, exacerbating these effects. Streamlining hospital transportation, providing at-home medical services, and stockpiling resources for rural hospitals can lessen COVID-19 mortality in rural regions. Urban counties face a different problem, having higher incidence rates and lower case fatality rates. While urban counties have more widely available healthcare services, they also have denser populations. These denser populations, coupled with higher predicted incidence rates, have the potential to strain hospital capacities, as demonstrated in mid-July. High incidence rates in urban counties are linked to lower percentages of elderly residents, higher percentages of black residents, and presence of large nursing homes. Of these, the variable that can most reasonably be addressed is nursing homes. The significance of COVID-19 outbreaks in nursing homes is widely known, and such outbreaks are often a result of a lack of testing and close contact with healthcare providers. A plausible measure could be to conduct consistent facility-wide testing, which has been shown to prevent outbreaks in nursing homes [23].

In addition to analyzing COVID-19 incidence rate and case fatality rate, our study uncovered interesting behavior in the predictor variables analyzed. *Black* is not significant in the case fatality rate model and has a very low R^2 upon removal, in contrast with its significance in the incidence rate model. This indicates that while counties with a high percentage of black residents have higher incidence rates, they do not necessarily have higher case fatality rates, despite these counties having higher rates of poverty and lesser access to healthcare. Counties with many Hispanic residents experience both higher incidence rates and higher case fatality rates, which can be partially attributed to high levels of obesity and diabetes coupled with low rates of health insurance coverage and higher representation in "superspreader" sites such as prisons or homeless shelters [24, 25]. Furthermore, Hispanic residents face language barriers and immigration status, which can complicate the process of obtaining medical aid [26, 27]. This highlights the concerning vulnerability of Hispanic populations along with the importance of adequate communication and provision of medical aid during this time. The negative coefficient of *Smoking* in the case fatality rate model was unexpected, as smoking is a CDC-defined risk factor for COVID-19 and prevalence of smoking in a county is positively correlated with elderly populations and with the incidence of respiratory disease, hypertension, obesity, and cardiovascular disease in Texas. Our study is not the first to report this relationship: Norden et al. [28] at Stanford University and the University of Washington found a similar negative relationship between smoking and COVID-19 case fatality rate across multiple countries, controlling for sex ratio, obesity, temperature, and elderly populations. In their study, several biological

explanations for a protective effect of smoking on COVID-19 fatality were proposed. While the possibility of a true protective effect of smoking on COVID-19 can not be ruled out, it is likely that a confounding variable, a variable not accounted for in our model, the effects of county-level data aggregation, or an interaction between another predictor variable and smoking is responsible for the negative coefficient. Regardless, the magnitude of the negative effect of smoking on COVID-19 case fatality rate prompts further exploration to clarify the role of smoking in COVID-19 fatality, both biologically and statistically.

The variables included in our full models are those that are the statistically strongest in predicting COVID-19 risk. In creating our models, we were surprised to find that health access (health insurance, primary care physicians), financial status (median income, poverty levels), mortality rate from respiratory disease, and population density were not among the best predictors, as these variables are thought to have significant effects on COVID-19 risk. However, several variables (*Elderly, Hispanic, Nursing Homes*) made consistent appearances, decreasing the likelihood that these variables were merely acting as proxies for more significant variables and indicating that there is some intrinsic aspect of populations of elderly/Hispanic residents, apart from correlations with other variables, that affects COVID-19 risk. This emphasizes the need to ensure that these populations are properly insulated against COVID-19 outbreaks and establishes them as important factors in population heterogeneity in relation to COVID-19 risk. Moreover, the residuals of our models contain important information about which counties are performing better or worse than expected. Comparing counties with negative residuals for common preventative measures could reveal which prevention strategies are most effective, which could then be implemented in counties with high residuals.

There exist a number of limitations in our study and opportunities for further work. Several counties with a value of 0 for the response variable were removed from our models to allow for a log transformation on the response variable. Excluding counties with no cases or deaths causes our study to underestimate the impact of variables that may be particularly significant in those counties, potentially leading to the omission of important variables in determining incidence rate or case fatality rate from COVID-19 and lowering the accuracy of prediction in counties with no deaths and/or confirmed cases. As the pandemic progresses, this limitation may fade. Another limitation was the designation of urban/rural and variables for super spreading contexts as discrete variables. In reality, the urban/rural designation is not black or white. Many counties cannot be classified as definitively urban or rural, with some sections being urban and others rural. A simple discrete variable does not capture this variability, and a more proper designation may be large metropolitan/midsize metro/small metro/micropolitan/semirural/rural. Regarding super spreading contexts, a discrete variable does not adequately convey the extent to which the

specific super spreading context impacts the county, and a variable that captures the proportion of people in the population who are employed or reside in a super spreading context may be a more robust predictor. Future research can extend the premise of our models to investigate more response variables, including R_t , testing rates, and possibly hospital capacity. By identifying at-risk counties and predicting the relative magnitude of county-level COVID-19 incidence rate or case fatality rate based on quantifiable population characteristics, resources can be more appropriately allocated to decrease the disproportionate impact of COVID-19 in Texas and a groundwork can be laid for future population studies to explore the relationships we discovered in depth.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Availability of data and material

The codes and data used to generate our results are available from the corresponding authors (email: ambermath99@gmail.com, aw22@txstate.edu) upon reasonable request.

Author contributions

AW designed the study. AKL, SZ, CS, and JW were responsible for data curation and model creation. AKL conducted statistical analysis and produced tables and figures. All authors contributed towards interpretation of study findings. AKL wrote the first draft, and all authors critically reviewed and approved the final contents of the paper.

Acknowledgements

A big thanks to Texas State University and the Mathworks Honors Summer Math Camp for providing a platform for us to collaborate virtually to undertake this research. We'd also like to thank Dr. Lauren Ancel-Meyers and the UT COVID-19 Modeling Consortium for sharing their knowledge and advice on a daily basis, and for providing inspiration and data leads for this project.

References

- [1] Texas Department of State Health Services. *DSHS COVID-19 Dashboard*. Aug. 11, 2020. URL: txdshs.maps.arcgis.com/apps/opstdashboard.
- [2] Swapnil Khose, Justin Xavier Moore, and Henry E Wang. “Epidemiology of the 2020 Pandemic of COVID-19 in the State of Texas: The First Month of Community Spread”. In: *Journal of Community Health* 45.4 (2020), pp. 696–701.
- [3] Jacob M Souch and Jeralynn S Cossman. “A Commentary on Rural-Urban Disparities in COVID-19 Testing Rates per 100,000 and Risk Factors”. In: *The Journal of Rural Health* (2020), pp. 1–3.
- [4] John Pender et al. *Rural America at a Glance*. Tech. rep. U.S. Department of Agriculture, 2019.
- [5] The Cecil G. Sheps Center for Health Services Research. *128 rural hospital closures: January 2010 - present*. Sept. 21, 2020. URL: <http://www.shepscenter.unc.edu/programs-projects/rural-health/rural-hospital-closures/>.
- [6] Arthur G Cosby et al. “Growth and persistence of place-based mortality in the United States: the rural mortality penalty”. In: *American Journal of Public Health* 109.1 (2019), pp. 155–162.
- [7] Brystana G Kaufman et al. “Half of Rural Residents at High Risk of Serious Illness Due to COVID-19, Creating Stress on Rural Hospitals”. In: *The Journal of Rural Health* (2020).
- [8] Centers for Disease Control and Prevention. *People who are at higher risk for severe illness*. Sept. 11, 2020. URL: <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html>.
- [9] Laura Richman et al. “Addressing health inequalities in diverse, rural communities: An unmet need”. In: *SSM-Population Health* 7 (2019), p. 100398.
- [10] Nancy et al. “Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019—United States, February 12–March 28, 2020”. In: *Morbidity and Mortality Weekly Report* 69.13 (2020), p. 382.
- [11] Carrie Henning-Smith. “The Unique Impact of COVID-19 on Older Adults in Rural Areas”. In: *Journal of Aging & Social Policy* (2020), pp. 1–7.
- [12] United States Census Bureau. *American Community Survey*. Aug. 13, 2020. URL: <https://www.census.gov/programs-surveys/acs>.
- [13] County Health Rankings & Roadmaps. *Rankings Data & Documentation*. Aug. 12, 2020. URL: <https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation>.

- [14] Global Health Data Exchange. *US Data*. Aug. 12, 2020. URL: <http://ghdx.healthdata.org/us-data>.
- [15] New York Times. *A Detailed Map of Who Is Wearing Masks in the U.S.* July 17, 2020. URL: <https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html>.
- [16] Niche Meat Processing. *Texas Meat Processing Plants*. Sept. 2014. URL: <https://www.nichemeatprocessing.org/wp-content/uploads/2014/09/Texas-Meat-Processing-Plants.pdf>.
- [17] Texas Almanac. *Texas Prisons*. Aug. 31, 2014. URL: <https://texasalmanac.com/topics/government/texas-prisons>.
- [18] Texas Health and Human Services. *Nursing Facilities*. Oct. 29, 2020. URL: <https://hhs.texas.gov/doing-business-hhs/provider-portals/long-term-care-providers/nursing-facilities-nf>.
- [19] New York Times. *An Extremely Detailed Map of the 2016 Election*. July 25, 2018. URL: <https://www.nytimes.com/interactive/2018/upshot/election-2016-voting-precinct-maps.html>.
- [20] New York Times. *Texas Senate Election Results: Beto O'Rourke vs. Ted Cruz*. Jan. 28, 2019. URL: <https://www.nytimes.com/elections/results/texas-senate>.
- [21] Texas Commission on the Arts. *Rural Texas Counties*. Aug. 13, 2020. URL: <https://www.arts.texas.gov/initiatives/rural-initiatives/rural-texas-counties/>.
- [22] Rashid Ahmed et al. "United States County-level COVID-19 Death Rates and Case Fatality Rates Vary by Region and Urban Status". In: *Healthcare*. Vol. 8. 3. Multidisciplinary Digital Publishing Institute. 2020, p. 330.
- [23] Carson T Telford et al. "Preventing COVID-19 Outbreaks in Long-Term Care Facilities Through Preemptive Testing of Residents and Staff Members—Fulton County, Georgia, March–May 2020". In: *Morbidity and Mortality Weekly Report* 69.37 (2020), pp. 1296–1299.
- [24] Raul Macias Gil et al. "COVID-19 Pandemic: Disparate Health Impact on the Hispanic/Latinx Population in the United States". In: *The Journal of Infectious Diseases* (2020).
- [25] Yiling J Cheng et al. "Prevalence of diabetes by race and ethnicity in the United States, 2011-2016". In: *Journal of the American Medical Association* 322.24 (2019), pp. 2389–2398.
- [26] Benjamin D Sommers. "Targeting in Medicaid: The costs and enrollment effects of Medicaid's citizenship documentation requirement". In: *Journal of Public Economics* 94.1-2 (2010), pp. 174–182.

- [27] Office of Minority Health. *Profile: Hispanic/Latino Americans*. Aug. 22, 2019. URL: <https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3>.
- [28] Michael J Norden et al. “National smoking rates correlate inversely with COVID-19 mortality”. In: *medRxiv* (2020).

Appendices

Table 1

Data Descriptions and Sources

Variable ^a	Description & Year of Data Collection	Source & R package used ^b
<i>IR</i>	7-day rolling average of the incidence rate of COVID-19 from the week of 9/12 - 9/19 2020	COVID19 Data Hub COVID19
<i>CFR</i>	7-day rolling average of the case fatality rate of COVID-19 from the week of 9/12 - 9/19 2020	COVID19 Data Hub COVID19
<i>Urban</i>	Metropolitan classification: 1 for urban and 0 for rural 2020	Texas Commission on the Arts –
<i>Black</i>	Percentage of population that is African American 2018	American Community Survey Tidycensus
<i>Hispanic</i>	Percentage of population that is Hispanic 2018	American Community Survey Tidycensus
<i>Nursinghomes</i>	Presence of a nursing home with > 100 residents 2020	Licensed Nursing Facilities –
<i>Elderly</i>	Percentage of people ages 70+ 2018	American Community Survey Tidycensus
<i>Twentieththirties</i>	Percentage of people ages 20-39 2018	American Community Survey Tidycensus
<i>Smoking</i>	Percentage of people surveyed who report that they smoke tobacco cigarettes (regularly or irregularly) 2012	Global Health Data Exchange –
<i>Obesity</i>	Prevalence of obesity (BMI > 30)	Global Health Data Exchange

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	2011	–
NAME	Name of the Texas county (_ County, Texas)	COVID19 Data Hub
	2020	COVID19
meatpacking	Presence of a meatpacking plant	Niche Meat Processing
	2014	–
prisons	Presence of a correctional institution	Texas Almanac
	2018	–
population	Total population	COVID19 Data Hub
	2020	COVID19
poverty	Percentage of households below the poverty threshold	American Community Survey
	2018	Tidycensus
medincome	Median gross income in 2018 inflation-adjusted dollars of employed residents ages 15+	American Community Survey
	2018	Tidycensus
zero_4	Percentage of population ages 0-4	American Community Survey
	2018	Tidycensus
five_9	Percentage of population ages 5-9	American Community Survey
	2018	Tidycensus
ten_14	Percentage of population ages 10-14	American Community Survey
	2018	Tidycensus
fifteen_19	Percentage of population ages 15-19	American Community Survey
	2018	Tidycensus
twenty_24	Percentage of population ages 20-24	American Community Survey
	2018	Tidycensus
twentyfive_29	Percentage of population ages 25-29	American Community Survey
	2018	Tidycensus
thirty_34	Percentage of population ages 30-34	American Community Survey
	2018	Tidycensus
thirtyfive_39	Percentage of population ages 35-39	American Community Survey
	2018	Tidycensus

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
fourty_44	Percentage of population ages 40-44 2018	American Community Survey Tidycensus
fourtyfive_49	Percentage of population ages 45-49 2018	American Community Survey Tidycensus
fifty_54	Percentage of population ages 50-54 2018	American Community Survey Tidycensus
fiftyfive_59	Percentage of population ages 55-59 2018	American Community Survey Tidycensus
sixty_64	Percentage of population ages 60-64 2018	American Community Survey Tidycensus
sixtyfive_69	Percentage of population ages 65-69 2018	American Community Survey Tidycensus
seventy_74	Percentage of population ages 70-74 2018	American Community Survey Tidycensus
seventyfive_79	Percentage of population ages 75-79 2018	American Community Survey Tidycensus
eighty_84	Percentage of population ages 80-84 2018	American Community Survey Tidycensus
eightyfive_above	Percentage of population ages 85+ 2018	American Community Survey Tidycensus
young	Percentage of population ages 19 and below 2018	American Community Survey Tidycensus
asian	Percentage of population that is Asian American 2018	American Community Survey Tidycensus
white	Percentage of popluation that is Caucasian American 2018	American Community Survey Tidycensus
native	Percentage of population that is American Indian or Alaska Native 2018	American Community Survey Tidycensus
female	Percentage of population that is female	American Community Survey

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	2018	Tidycensus
diabetes	Prevalence of diagnosed and undiagnosed diabetes	Global Health Data Exchange
	2012	–
hypertension	Prevalence of chronic hypertension in residents ages 30 and above	Global Health Data Exchange
	2009	–
respiratory	Age-standardized mortality rate from chronic respiratory disease per 100,000 population	Global Health Data Exchange
	2014	–
cardiovascular	Age-standardized mortality rate from cardiovascular disease per 100,000 population	Global Health Data Exchange
	2014	–
asthma	Prevalence of asthma	Global Health Data Exchange
	2014	–
alcohol	Age-standardized mortality rate from alcohol use disorders per 100,000 population	Global Health Data Exchange
	2014	–
violence	Age-standardized mortality rate from interpersonal violence per 100,000 population	Global Health Data Exchange
	2014	–
drug_use	Age-standardized mortality rate from drug use disorders per 100,000 population	Global Health Data Exchange
	2014	–
Density	County population divided by county area in square miles	Texas Association of Counties
	2020	–
CruzR	Percentage of all voters that voted for Ted Cruz in the 2018 TX Senate Election	New York Times
	2019	–
BetoD	Percentage of all voters that voted for Beto O'Rourke in the 2018 TX Senate Election	New York Times

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	2019	–
TrumpR	Percentage of all voters that voted for Donald Trump in the 2016 presidential election	New York Times
	2018	–
ClintonD	Percentage of all voters that voted for Hillary Clinton in the 2016 presidential election	New York Times
	2018	–
bachelors	Percentage of the population with a bachelor’s degree	American Community Survey
	2018	Tidycensus
masters	Percentage of the population with a master’s degree	American Community Survey
	2018	Tidycensus
doctorate	Percentage of the population with a doctorate degreee	American Community Survey
	2018	Tidycensus
collegedegree	Percentage of the population with a college degree	American Community Survey
	2018	Tidycensus
healthinsurance	Percentage of the population with health insurance	American Community Survey
	2018	Tidycensus
NEVER	Percentage of survey respondents who report never wearing a mask in public places	New York Times
	2020	–
RARELY	Percentage of survey respondents who report rarely wearing a mask in public places	New York Times
	2020	–
SOMETIMES	Percentage of survey respondents who report sometimes wearing a mask in public places	New York Times
	2020	–
FREQUENTLY	Percentage of survey respondents who report frequently wearing a mask in public places	New York Times
	2020	–
ALWAYS	Percentage of survey respondents who report	New York Times

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	always wearing a mask in public places	
	2020	–
YPLL	Age-adjusted years of potential life lost per 100,000 population	County Health Rankings
	2020	–
poorhealth	Age-adjusted percentage of adults reporting fair or poor health	County Health Rankings
	2020	–
fei	Index of factors that contribute to a healthy food environment, from 0 (worst) to 10 (best)	County Health Rankings
	2020	–
physically.inactive	Percentage of adults age 20 and over reporting no leisure-time physical activity.	County Health Rankings
	2020	–
flu.vaccination	Percentage of fee-for-service (FFS) Medicare enrollees that had an annual flu vaccination.	County Health Rankings
	2020	–
unemployed	Percentage of population ages 16 and older unemployed but seeking work.	County Health Rankings
	2020	–
childpoverty	Percentage of children under age 18 living in poverty	County Health Rankings
	2020	–
income.80	80th percentile of household income	County Health Rankings
	2020	–
income.20	20th percentile of household income	County Health Rankings
	2020	–
income.ratio	Ratio of household income at the 80th percentile to income at the 20th percentile	County Health Rankings
	2020	–
single.parent	Percentage of children that live in a household headed	County Health Rankings

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	by single parent	
	2020	–
social.association	Number of membership associations per 10,000 population	County Health Rankings
	2020	–
violent.crime	Number of reported violent crime offenses per 100,000 population	County Health Rankings
	2020	–
overcrowding	Percentage of households with overcrowding	County Health Rankings
	2020	–
facilities.inadequate	Percentage of households with lack of kitchen or plumbing facilities	County Health Rankings
	2020	–
life.expectancy	Average number of years a person can expect to live	County Health Rankings
	2016-2018	–
physical.distress	Percentage of adults reporting 14 or more days of poor physical health per month	County Health Rankings
	2017	–
mental.distress	Percentage of adults reporting 14 or more days of poor mental health per month	County Health Rankings
	2017	–
hiv	Number of people aged 13 years and older living with a diagnosis of human immunodeficiency virus (HIV) infection per 100,000 population	County Health Rankings
	2016	–
food.insecurity	Percentage of population who lack adequate access to food	County Health Rankings
	2017	–
sleep.insufficient	Percentage of adults who report fewer than 7 hours of sleep on average	County Health Rankings
	2016	–
free.lunch	Percentage of children enrolled in public schools that are	County Health Rankings

Continued on next page

Table 1 cont.

Variable	Description & Year of Data Collection	Source & R package
	eligible for free or reduced price lunch	
	2017-2018	–
suicide	Number of deaths due to suicide per 100,000 population	County Health Rankings
	2014-2018	–

Information about each variable is presented in two rows. The first row contains a brief description of the variable on the left and the source of the variable on the right. The second row contains the year of data collection on the left and the R package used to obtain the data on the right. If the variable description overflows from the first row, the information in the second row is shifted down to make room for a continued description of the variable. Variables that were included in MLR models are capitalized, italicized, and presented at the top.

^aName of the variable as it appears in our dataset.

^bR package used to retrieve the county-level dataset for the indicated variable, if applicable.

Table 2

Texas County-Level Variable Comparisons by Quintiles of COVID-19 Incidence Rate through September 15, 2020^a

Incidence Rate	0.00-964.7	964.7-1426.7	1426.7-1923.6	1923.6-2816.3	>2816.3	<i>p</i> -value ^b
	Data presented as median (IQR)					
Counties	51	51	50	51	51	–
<i>Demographics</i>						
Black	1.8 (3.8)	4.6 (8.0)	4.9 (10.3)	6.3 (8.7)	4.8 (8.6)	0.002
White	90.2 (9.2)	86.7 (12.1)	85.8 (11.5)	81.8 (15.1)	83.4 (14.8)	0.002
Hispanic	18.5 (18.3)	19.3 (24.6)	23.5 (17.2)	36.8 (33.9)	50.2 (38.6)	<.001
Elderly	14.4 (5.6)	12.7 (5.9)	12.5 (4.7)	9.8 (4.1)	9.3 (2.3)	<.001
Age 20-39	20.8 (4.7)	23.1 (6.1)	23.1 (4.2)	25.9 (5.2)	27.2 (5.6)	<.001
Female	49.9 (3.3)	50.3 (1.8)	50.4 (2.0)	49.7 (2.0)	49.4 (3.3)	0.097
<i>Health Status</i>						
Obesity	37.6 (3.1)	38.9 (3.5)	39.6 (3.5)	39.3 (3.3)	40.3 (3.3)	<.001
Smoking	21.2 (4.4)	22.4 (3.7)	21.4 (4.3)	20.8 (5.4)	20.6 (3.2)	0.108
Diabetes	14.4 (1.2)	15.6 (2.2)	15.4 (2.2)	16.2 (2.2)	17.0 (3.8)	<.001
Health Insurance	98.8 (1.8)	98.7 (1.7)	98.8 (1.2)	98.8 (2.1)	97.7 (6.2)	0.012
<i>Living Conditions</i>						
Unemployed	3.4 (0.6)	3.5 (1.0)	3.7 (0.9)	3.6 (1.4)	3.8 (1.7)	0.078
Poverty	13.5 (5.8)	14.1 (5.5)	15 (5.0)	14.4 (6.4)	16.6 (4.9)	0.002
Density	5.5 (19.7)	27.4 (63.3)	19.5 (42.2)	40.1 (169.4)	36.2 (63.7)	<.001

^aCounty-level incidence rate calculated as the 7-day rolling average of confirmed COVID-19 cases divided by population. Expressed per 100,000 population on September 15, 2020.

^b*p*-values determined using the Kruskal-Wallis test. *p*-value < 0.05 is considered significant.

Table 3
COVID-19 Incidence Rate Models

	Base ^a	Full ^b	Rural ^c
Data presented as β_n (<i>p</i> -value)			
<i>Variables</i>			
Intercept	-4.191 (<.001)	-4.725 (<.001)	-4.781 (<.001)
Urban	0.341 (<.001)	0.005 (.957)	–
Elderly	–	-2.324 (.020)	-2.321 (.046)
Nursing Homes	–	0.343 (<.001)	0.344 (<.001)
Hispanic	–	1.531 (<.001)	1.581 (<.001)
Black	–	2.553 (<.001)	3.261 (<.001)
<i>Model Strength</i>			
R^2	0.052	0.448	0.418
F	14.59	41.61	34.52

–: Variable is not included in the model.

^aRegresses COVID-19 incidence rate on urban only; control.

^bStatistically strongest model by BIC and adjusted R^2 .

^cFull model filtered to include only rural counties.

Table 4

Texas County-Level Variable Comparisons by Quintiles of COVID-19 Case-Fatality Rate through September 15, 2020^a

Case Fatality Rate	0%-1.3%	1.3%-1.9%	1.9%-2.9%	2.9%-4.2%	>4.2%	<i>p</i> -value ^b
Data presented as median (IQR)						
Counties	50	50	50	50	51	–
<i>Demographics</i>						
Black	3.8 (8.6)	7.2 (10.4)	5.3 (5.8)	2.5 (5.8)	3.0 (6.8)	0.004
White	85.6 (10.3)	81.7 (16.2)	87 (10.4)	88.3 (12.1)	88.0 (13.3)	0.142
Hispanic	26.9 (20.3)	23.9 (9.8)	40.3 (39.7)	39.3 (39.5)	23.5 (40.9)	0.079
Elderly	11.3 (6.4)	9.9 (4.4)	10.5 (4.4)	11.4 (3.3)	14.3 (6.1)	< .001
Age 20-39	24.6 (9.0)	25.8 (5.2)	25.2 (5.7)	24.6 (4.1)	21.2 (5.5)	< .001
Female	49.4 (4.0)	50.3 (1.9)	50.2 (2.2)	50.4 (2.9)	50.4 (1.9)	0.155
<i>Health Status</i>						
Obesity	37.7 (3.3)	38.5 (3.8)	39.9 (2.3)	39.6 (2.7)	39.7 (3.1)	< .001
Smoking	20.9 (3.5)	21.3 (3.5)	22.1 (4.2)	21.5 (5.1)	20.7 (3.7)	0.455
Diabetes	15.0 (1.8)	15.1 (2.4)	16.4 (2.3)	16.1 (3.1)	16.1 (2.6)	< .001
Health Insurance	98.5 (6.8)	98.8 (3.3)	98.8 (1.2)	98.4 (3.9)	98.6 (1.8)	0.645
<i>Living Conditions</i>						
Unemployed	3.3 (0.7)	3.4 (1.0)	3.7 (1.2)	3.7 (1.3)	4.1 (1.8)	< .001
Poverty	12.5 (4.6)	13.5 (5.3)	15.5 (3.5)	15.8 (5.9)	16.8 (7.7)	< .001
Density	9.4 (25.9)	48.8 (170.1)	39.5 (69)	26.1 (43.1)	11.5 (27.2)	< .001

^aCounty-level case fatality rate calculated as the 7-day rolling average of COVID-19 deaths divided by the 7-day rolling average of confirmed COVID-19 cases.

^b*p*-values determined using the Kruskal-Wallis test. *p*-value < 0.05 is considered significant.

Table 5

COVID-19 Case Fatality Rate Models

	Base ^a	Full ^b	Rural ^c
Data presented as β_n (<i>p</i> -value)			
<i>Variables</i>			
Intercept	-3.516 (<.001)	-4.113 (<.001)	-4.113 (<.001)
Urban	-0.426 (<.001)	0.030 (.773)	–
Elderly	–	3.876 (.015)	3.545 (.069)
Nursing Homes	–	-0.210 (.013)	-0.186 (.058)
Age 20-39	–	-4.347 (<.001)	-4.633 (.003)
Obesity	–	6.354 (<.001)	6.485 (.001)
Smoking	–	-4.881 (<.001)	-5.752 (.002)
<i>Model Strength</i>			
<i>R</i> ²	0.086	0.332	0.238
F	21.06	19.62	11.10

–: Variable is not included in the model.

^aRegresses case fatality rate on urban only; control.

^bModel with the lowest BIC (Bayesian Information Criteria).

^cFull model filtered to include only rural counties.

Table 6

Case Fatality Rate Model with Full Incidence Rate Model

Predictors

	IR	CFR
<i>Variables</i>		
Intercept	-4.725 (<.001)	-4.698 (<.001)
Urban	0.005 (.957)	0.007 (.945)
Elderly	-2.324 (.020)	8.079 (<.001)
Nursing Homes	0.343 (<.001)	-0.251 (.006)
Hispanic	1.531 (<.001)	0.685 (.002)
Black	2.553 (<.001)	0.435 (.527)
<i>Model Strength</i>		
R^2	0.463	0.235
F	43.81	14.79

CFR is regressed on the predictor variables from the full IR model: *Urban*, *Elderly*, *Nursing Homes*, *Hispanic*, and *Black*. Full IR model is provided for comparison.

Table 7

Summary Statistics for Incidence

Rate Models

	Mean	SD ^a
Incidence Rate		
Urban	0.024	0.012
Rural	0.019	0.012
<i>Overall</i> ^b	0.020	0.012
Black		
Urban	0.098	0.077
Rural	0.053	0.058
<i>Overall</i>	0.064	0.066
Hispanic		
Urban	0.325	0.234
Rural	0.359	0.233
<i>Overall</i>	0.351	0.233
Elderly		
Urban	0.090	0.026
Rural	0.129	0.038
<i>Overall</i>	0.119	0.039
Nursing Homes		
Urban	0.968	0.177
Rural	0.436	0.498
<i>Overall</i>	0.570	0.496

^aStandard deviation.

^bIncludes $N = 251$ counties in the full IR model.

Table 8

Summary Statistics for Case Fatality Rate Models

	Mean	SD ^a		Mean	SD
CFR			Black		
Urban	0.022	0.011		0.098	0.076
Rural	0.038	0.033		0.057	0.060
Overall ^b	0.033	0.029		0.068	0.067
Hispanic			Nursing Homes		
Urban	0.325	0.234		0.968	0.177
Rural	0.372	0.242		0.491	0.501
Overall	0.359	0.240		0.624	0.485
Elderly			Age 20-39		
Urban	0.090	0.026		0.276	0.041
Rural	0.127	0.036		0.238	0.045
Overall	0.116	0.037		0.249	0.047
Obesity			Smoking		
Urban	0.382	0.037		0.203	0.034
Rural	0.399	0.026		0.217	0.028
Overall	0.394	0.030		0.213	0.030

^aStandard deviation.

^bIncludes $N = 226$ counties in the full CFR model.

Values for *Hispanic*, *Black*, *Nursing Homes*, and *Elderly* differ from Table 7 because of the 28 counties removed in the CFR model, as opposed to 3 in the IR model.

Table 9

Impact of Each Predictor Variable on the R^2 Upon Removal

	Incidence Rate	CFR - IR Predictors	CFR - Full
Urban	-.001%	-.003%	-.024%
Elderly	-1.21%	-11.7%	-1.79%
Hispanic	-19.3%	-3.49%	-
Nursing Homes	-5.25%	-2.70%	-1.85%
Black	-4.85%	-.142%	-
Smoking	-	-	-4.18%
Obesity	-	-	-7.23%
Age 20-39	-	-	-4.20%

Each cell is the negative impact that the respective variable has on the R^2 of the indicated model upon removal. Percentages are differences from the R^2 , and are not percentages of the R^2 .

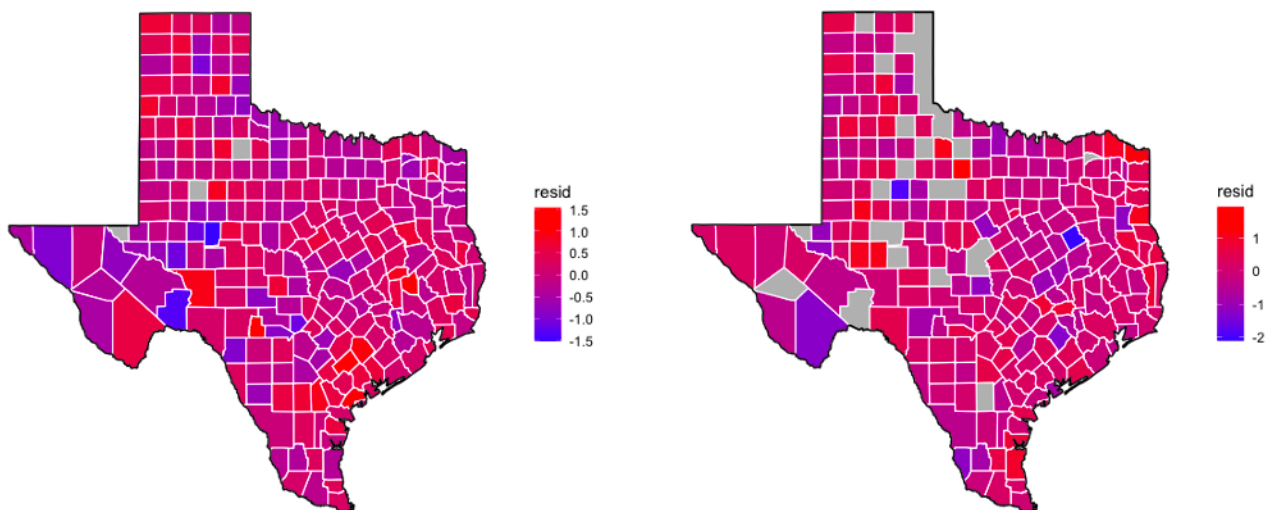


Figure 1. Residuals of the full incidence rate model (left) and full case fatality rate model (right). Counties with no confirmed cases are displayed in gray on the left, and counties with no deaths are displayed in gray on the right. The scales for the map on the left and the map on the right refer to $\ln(IR)$ and $\ln(CFR)$, respectively.