ISSN: 2641-6336

Annals of
Biostatistics & Biometric Applications

DOI: 10.33552/ABBA.2025.06.000646

Iris Publishers

Review Article

Risk Factors for COVID-19 in Ohio: A Retrospective Study at the County Level

Young Jun Park[1], Jeffrey B. Travers[2] and Trevor Bihl[3]*

[1]*Department of Pharmacology & Toxicology, Wright State University, Dayton, OH, USA*

[2]*Department of Pharmacology & Toxicology, Wright State University, Dayton, OH, USA*

[3]*Russ College of Engineering and Technology, Ohio University, Athens, OH, USA*

***Corresponding author:** Trevor Bihl, Russ College of Engineering and Technology, Ohio University, Athens, OH, USA

Received Date: **August 24, 2025**

Published Date: **September 16, 2025**

Abstract

The COVID-19 pandemic has changed how we view public health and its importance. Through the pandemic, understanding the epidemiology of COVID-19 is critical to help mitigate the damage future pandemics might have. This study has attempted to identify factors that can explain differences in COVID-19 infection rates and death rates between the counties of Ohio. Results showed that there are regional differences that explain the number of COVID-19 cases and a multitude of factors together that can explain COVID-19 death rates. Furthermore, as the number of ICU beds increase, the number of COVID-19 death rates decrease. This study is a start but further research examining other factors needs to be done to understand disparate COVID-19 health impact on a county level.

Keywords: COVID-19; appalachia; ohio; epidemiology

Introduction

COVID-19, also known as SARS-CoV-2, is an acute respiratory syndrome that has caused a global pandemic since 2020 [1]. Milder symptoms of the condition include fever, cough, difficulty breathing and in worst case, it can lead to death [2]. The virus originated in Wuhan, China where there was a pneumonia-like outbreak of people who were at the local food market near the Wuhan Institute of Virology [3]. What started in the local market rapidly progressed in the hospital and close-contact transmission among family and friends. The main transmission of COVID-19 is through droplet transmission, but aerosol is a potential alternative route. Since the first discovery in November 2019, COVID-19 has killed 1 million people in the US [4]. This alarming rate led the U.S. Department of Health and Human Services to declare a public health emergency on January 31st, 2020. Since then, major societal changes occurred with the transition to virtual learning, drop in the U.S. GDP, and transitioning to working from home. After almost 3 years of the pandemic, and a couple years of lockdown, the public health emergency declaration finally ended on May 11th, 2023 [5]. By April 26th, 2023, there were more than 104 million U.S. COVID-19 cases. Considerable unknowns emerged due to COVID-19 in how to respond to the situation and how to handle the public health emergency [6].

Despite COVID-19 cases significantly reducing since the onset of the pandemic, learning from this pandemic will be important for when the next pandemic occurs. Notably, different communities were associated with different results in the COVID-19 pandemic, but these differences in outcomes were not always obvious. Visibly, this is apparent in Figure 1, which shows the overall COVID-19 death rate by Ohio county for 2020-2023. Notably, the major cities of Cincinnati (bottom left), Columbus (middle), and Cleveland (top right) have an overall death rate (per 100,000) less than many rural counties and those along the Ohio River. The high-level observation about COVID-19 death rate by county implies a regional or cultural difference in the impact that COVID-19 had on the state of Ohio. This seems apparent when further looking at the Hospital and ICU bed rate by county, Figure 2, where many counties that had high COVID-19 death rates are visibly seen to have no hospital beds. This paper thus presents a retrospective study to understand various county level risk factors and their potential relationship to COVID-19 results. In general, this paper explains the differing results in COVID-19 outcomes across the counties in Ohio.

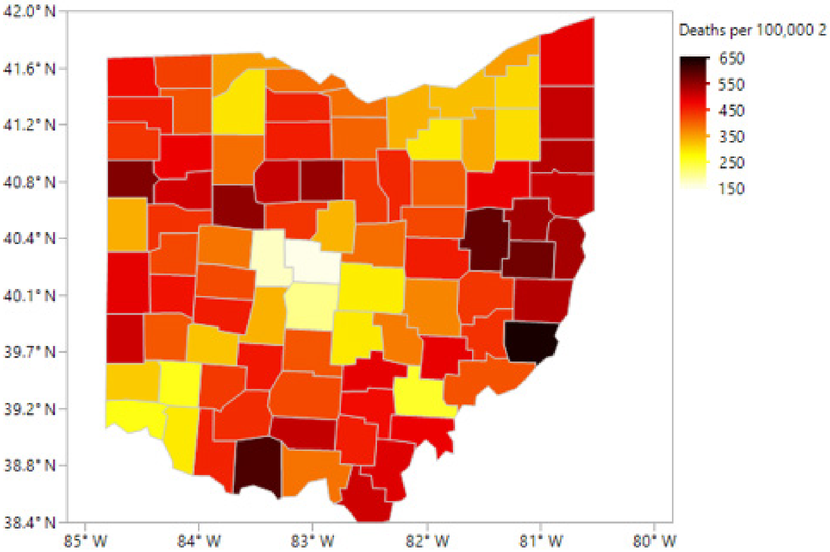


Figure 1: Overall COVID-19 Death Rate (per 100,000) by Ohio County.

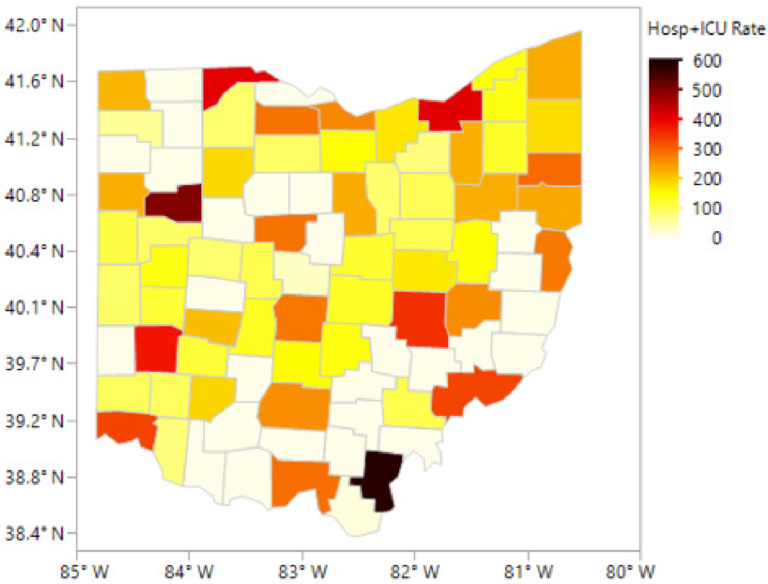


Figure 2: Hospital and ICU bed rate (per 100,000) in each county.

Background

Recent studies have explored various population level statistics and their relationship to COVID-19 results. One paper written by Khuda Bukhsh et al. [7] tried to create a model that can predict how many COVID-19 cases in Ohio to understand the hospital burden with the result that county level COVID-19 outcomes appeared related to political party affiliations in the county. One study done by Lhila et al. observed that 25% of additional COVID-19 deaths were explained by the differences between the Democratic and Republican party and their support President Trump [8]. Another study aimed at identifying if income inequality differences by Zip-codes could explain differences in COVID-19 infection and mortality outcomes in Chicago [9]. They found that Zip-codes with lower income had higher infection rates than those with higher income. Medical conditions like obesity have placed obese people at triple the risk of hospitalization due to COVID-19 infection [10].

In a study observe COVID-19 outcomes of children diagnosed with obesity, they were 3.07 times as likely to be hospitalized and 1.42 times higher to be put in an ICU or on an invasive mechanical ventilator [11]. Besides the COVID-19 pandemic, another public health issue that Ohio faces is drug overdose. In 2007, drug poisoning was the leading cause of injury death in Ohio and has only increased with 5,017 deaths in 2020 [12]. A study showed that non-fatal overdosing has been linked with a greater risk of developing severe complications due to COVID-19 [13]. Limited access to health insurance has also increased individuals to a 1.9 times higher risk of hospitalization [14]. With COVID-19 being a respiratory infection, air quality has the potential to influence the severity and mortality of COVID-19. A thorough literature review showed limited studies done to understand which risk factors were effective in predicting the severity of COVID-19 cases and deaths in different Ohio counties.

Methodology

For this study, data was collected from multiple sources. This is summarized in Table 1. This data was then analyzed through various statistical methods to understand the relationship between the dependent and hypothesized independent variables.

Table 1: Collected country level data and sources.

Variable	Role	Year of data	Source
Average Income per capita (USD)	I		[16]
Average age-adjusted rate of unintentional drug overdose deaths (per 100,000 population)	I	2015-2020	[12]
Percent of obese adults	I		[17]
Uninsured rate (per 100,000 population)	I		[18]
Air pollution (PM2.5)	I		[19]
Feb 2020 Unemployment	I	2020	[20]
Geographic Region	I	N/A	[21]
Appalachian County	I	2008+	[22,23]
Number of hospital beds	I	2020	[24]
Number of ICU beds	I	2020	[24]
Overall Covid-19 rate	D	2020-2023	[25]
Overall COVID-19 death rate	D	2020-2023	[25]

+An emergency stay at home order was issued on March 23, 2020 [26]. Thus, this variable possibly explains the economic conditions of a given county going into the pandemic.

Data Collection

For this study, data was collected from multiple sources. This is summarized in Table 1. Dependent variables of interest included both the COVID-19 cases per 100,000 per county in Ohio and the COVID-19 deaths per 100,000 per county in Ohio. The dependent variables were both collected for the overall pandemic, i.e., 2020-2023 numbers, and were scaled by the population of each county to compute the rate. The number of COVID-19 cases and deaths were acquired from The New York Times Ohio COVID-19 dashboard that gets their data from the Ohio Department of Health [15-20]. Independent variables were collected as factors which could explain these dependent variables and thus the authors aimed to collect statistical variables that described Ohio counties as they were going into the pandemic, i.e., 2020 numbers, or overall numbers for the county. These variables are described in Table 1 and include an "I" to denote independent variable. An example of the collection logic for exploring different years is as such, the unemployment rate for February 2020 was collected since this was the last full month before the Ohio

The 88 counties of Ohio were categorized into five different geographical regions based on the Ohio Self Determination Association [21-26]. Table 2 shows the 88 Ohio counties which are served by the Appalachian Regional Commission (ARC) as expanded to include counties in Ohio by 2008, most of these Appalachian counties are in the unglaciated region of Ohio, along the Ohio River valley, and were the first settled counties in Ohio. Many of these counties are rural and have depressed economies The World Population Review was used to obtain the 2023 population and population density (per square mile) [27]. To normalize the data, COVID-19 cases and deaths, drug overdose death, and number of uninsured people converted to rates (per 100,000 population). To obtain the number of staff hospital beds and ICU beds per county, the American Hospital Directory and an article published in Data Central was used [28,29]. Other risk factors and their source are included in Table 1.

Table 2: Ohio Counties by Geographic Regions, Appalachian Counties are bold and underlined.

Region	Ohio Counties
Northwest	Allen, Crawford, Defiance, Erie, Fulton, Hancock, Hardin, Henry, Huron, Lucas, Morrow, Ottawa, Paulding, Putman, Sandusky, Seneca, Williams, Wood, Wyandot, Marion, and Van Wert
Northeast	Ashland, **Ashtabula**, **Columbiana**, Cuyahoga, Geauga, Lake, Lorain, **Mahoning**, Medina, Portage, Richland, Stark, Summit, **Trumbull**, and Wayne
Southwest	**Adams**, Auglaize, **Brown**, Butler, Champaign, Clark, **Clermont**, Clinton, Drake, Fayette, Greene, Hamilton, **Highland**, Logan, Mercer, Miami, Montgomery, Preble, Shelby, and Warren
Southeast	**Athens**, **Belmont**, **Carroll**, **Coshocton**, **Gallia**, **Guernsey**, **Harrison**, **Hocking**, **Holmes**, **Jackson**, **Jefferson**, Knox, **Lawerence**, **Meigs**, **Monroe**, **Morgan**, **Muskingum**, **Noble**, **Perry**, Pickaway, **Pike**, **Ross**, **Scioto**, **Tuscarawas**, **Vinton**, **and Washington**
Central	Delaware, Fairfield, Franklin, Licking, Madison, and Union

All the data was organized by Ohio counties and no other data editing was done. Our objective to understand the risk factors for the differences in COVID-19 cases and deaths can be further divided into more detailed questions. The first question is to understand if there is a statistical difference in COVID-19 cases between regions. Understanding this question can help us see if COVID-19 infection can be attributed to geographical locations. The second question is to understand a possible relationship between the number of ICU beds per county and COVID-19 deaths. Understanding this information would help identify counties of attention during another viral pandemic to provide adequate numbers of ICU beds. The third question is to identify statistically significant risk factors that can predict which counties will have a higher COVID-19 infection rate. The last question is to identify statistically significant risk factors that can predict which counties will have higher COVID-19 mortality.

Statistical Analysis

Statistical analysis considered one-way analysis of variance (ANOVA) and linear regression for continuous responses were considered with F-statistics and p-values presented to show statistical significance relative to hypothesis tests that the model of selected variables does not have a relationship with the dependent variable. From these determined statistically significant relationships, stepwise regression, consistent with [30], was used to develop parsimonious multiple linear regression models wherein variables were removed sequentially in a backward manner based on minimum Bayesian Information Criterion (BIC). Statistical calculations were performed using JMP 14 (SAS, Cary, NC) consistent with [31]. In statistical tests, $p < 0.05$ was considered statistically significant, however raw p-values are presented from the underlying statistical tests since many p-values were considerably smaller than 5%.

Results

Regional Differences in COVID-19 Responses are Apparent

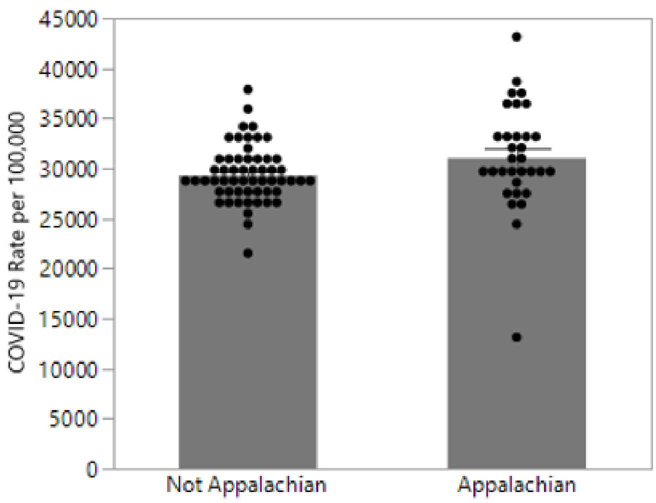


Figure 3: COVID-19 infection rate by county for Appalachian vs Not Appalachian counties (T=2.02, p=0.047).

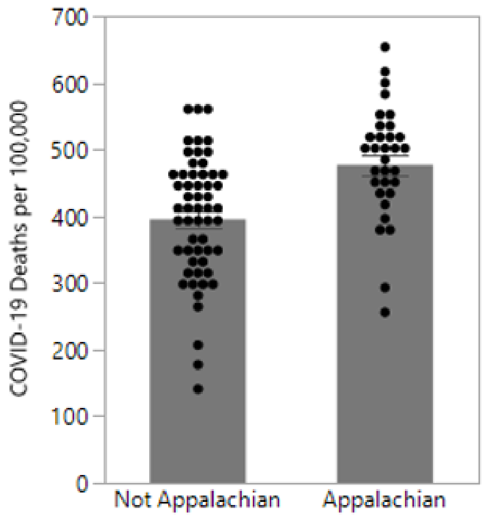


Figure 4: COVID-19 Death rate by county for Appalachian vs Not Appalachian counties (T=4.106, p<0.0001).

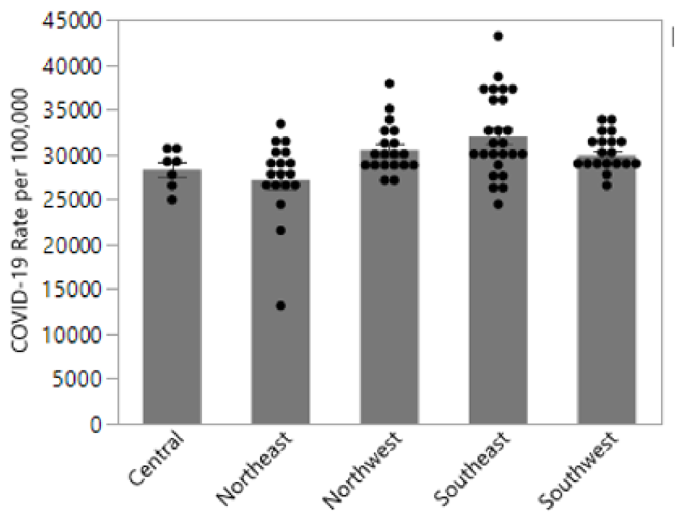


Figure 5: COVID-19 infection rate by county for regions counties (R2=0.20, F=5.203, p=0.0009).

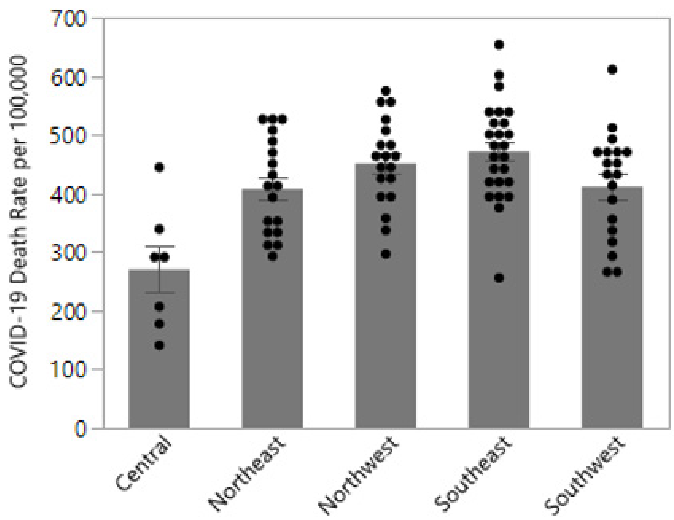
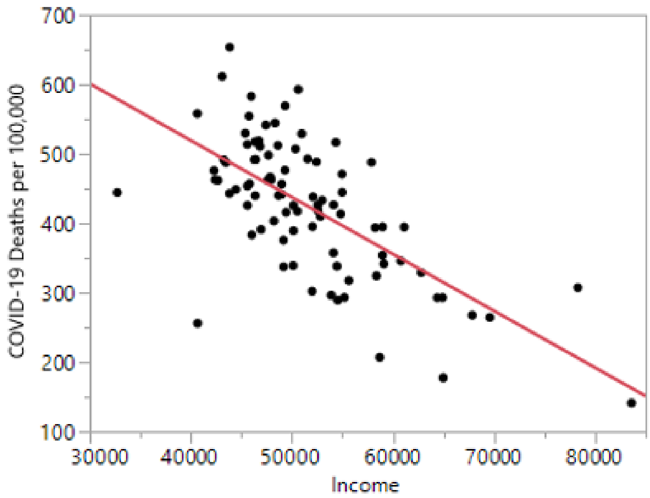


Figure 6: COVID-19 death rate by county for regions counties (R2=0.29, F=8.622, p<0.0001).

First of all, the difference between Non-Appalachian and Appalachian counties were measured for COVID-19 rates by county, Figure 3, and COVID-19 death rates by county, Figure 4. Pooled t-tests were used, ANOVA was used and depicted in Figure 3. As shown in Figures 3&4, a statistically significant difference is apparent in response where Appalachian counties had a high overall COVID-19 infection rate (T=2.02, p=0.047) and death rate (T=4.106, p<0.0001). To identify if there is a statistical difference of COVID-19 cases between regions, ANOVA was used and showed in Figure 4 (for overall COVID-19 rate) and Figure 5 (for COVID-19 death rates). In both cases statistically significant relationships exist, indicating a regionality aspect of the overall response to the COVID-19 pandemic in Ohio. Results suggest that there are statistically significant differences between the Southeast and Northeast, Southeast and Central, Northwest and Northeast, and Southwest and Northeast Figure 6.

Overall Differences



$$Deaths\ per\ 100,000 = 846,273 - 0.008 * Income\ R^2 = 0.44$$

Figure 7: Linear regression relationship between income and COVID-19 death rate.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	p
Model	6	768367886	128061314	18.0602	<0.0001
Error	81	574356004	7090814.9		
Total	87	1342723890			

Summary of Fit

R^2	0.572
R^2 Adjusted	0.540

Parameter Estimates

Variable	Estimate	p
Intercept	46437.477	<0.0001*
Region(Northeast&Central)	-746.23	0.0506†
Region(Northeast)	-1453.51	0.0278*
Drug Overdose Death Rate	93.37	<0.0001*
Income	-0.11	0.0170*
Uninsured rate	-0.68	<0.0001*
Air Pollution	-1048.777	0.0027*

Figure 8: Multiple regression relationship with COVID-19 infection rate as dependent variable.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	p
Model	6	473587.93	78931.3	17.9833	<0.0001
Error	81	355521.04	4389.1		
Total	87	829108.97			

Summary of Fit	
R^2	0.571
R^2 Adjusted	0.539

Parameter Estimates		
Variable	Estimate	p
Intercept	446.32813	0.0003*
Municipal Ratio	108.37085	0.0903†
Region{Central}	44.79815	0.0021*
February 2020 Unemployment Rate	14.186224	0.0248*
Population density	-0.057416	0.0077*
Uninsured rate	0.0067296	0.0880†
Income	-0.004204	0.0016*

Figure 9: Multiple regression relationship with COVID-19 death rate as dependent variable.

An interesting and statistically significant relationship was seen between county average (mean) income and the COVID-19 death rate. This is seen in Figure 7. As seen in this figure, as incomes increase the death rates generally decrease (R^2 = 0.44, p < 0.0001). For this model, the residuals were approximately normal but had a heteroscedasticity issue, possibly due to an outlier with Athens County. Notably, Athens has an overall low average income but also low death rate; this is possibly due to the county being predominantly rural and poor but also having with a major R1 university (Ohio University). Notably, while there were statistically significant linear relationships between the overall hospital bed (hospital + ICU) rates by county and both COVID-19 overall infection rate (p=0.047) and COVID-19 Death rate (p=0.03). Both were of minimally predictive value (R^2 =0.06 and R^2 =0.07, respectively). In both cases, counties without hospital or ICU beds were excluded to avoid a bias, and thus this is limited in its results. Similarly, many other bivariate relationships were not statistically or practically significant.

Complex Relationships

Due to the apparent complex relationship in the data, a multiple regression was considered for both COVID-19 rate and COVID-19 death rate as separate dependent variables. In both cases, all collected variables were considered as independent variables using a stepwise approach, with a p-value threshold of 0.25 to enter and 0.1 to leave. This was iterative and performed separately for each dependent variable. First the model in Figure 8 was produced. This model has a statistically significant multiple linear regression predictive relationship between these variables and COVID-19 rate (p < 0.0001, R^2 = 0.57). When a similar stepwise regression approach was used with the COVID-19 Death Rate as a dependent variable and other all collective data variables as independent variables using a stepwise approach, the model in Figure 8 was produced. This model has a statistically significant multiple linear regression predictive relationship between the variables in the parameter estimate table and the COVID-19 death rate (p < 0.0001, R^2 = 0.57).

Discussion

Determining if there were significant differences of COVID-19 responses (infections and deaths) between counties and regions of Ohio based on external and lifestyle variables was approached through available resources and analyzed. The results firstly showed that Appalachian counties in Ohio, which are mostly rural and poor, had considerably higher incidents of both COVID-19 infections and deaths than the rest of Ohio. Further differences were seen across other geographical regions of Ohio, suggesting that there are geographical differences between regions that allow for the differences in COVID-19 infection rates. Further research needs to be done to identify what geographical or demographics factors create these differences. When considering income as a predictor of COVID-19 deaths, a negative relationship was seen where counties with higher income saw lower death rates. The implication is that counties with higher income rates have better healthcare or healthcare outcomes. When considering this result with the observations of Appalachian counties is that the relatively lower income of Appalachian counties could be a predictor for pandemic health impacts and thus a tool for use in future healthcare emergencies.

When considering multiple variables through stepwise regression means, two different models were seen for COVID-19 infection rates and death rates. While different variables were seen in both; some commonality was found where both included regional data about the counties, implying a cultural aspect of the health outcomes, as well as both income and the uninsured rate, further illustrating disparate health outcomes based on income and availability of medical care. Interestingly, COVID-19 death rates were further seen to have a predictive relationship with the unemployment level of the county going into the 2020 pandemic while the overall COVID-19 incident rate was seen to have a predictive relationship with the drug overdose rate of the counties. More understanding of these relationships could provide vectors to understanding and improving healthcare results. Finally, this study was limited by data availability (additional factors such as public health spending rate per county were unavailable) and the small sample size (limited to only Ohio).

Conclusions

This study found predictive relationships between county level COVID-19 outcomes in Ohio and various demographic variables. Chief among these was if the county was considered part of Appalachia. Additionally, a link between COVID-19 outcomes and income and financial security was discovered; which could provide an indicator for intervention for future pandemics and medical emergencies.

References

1. M Ciotti, M Ciccozzi, A Terrinoni, WC Jiang, CB Wang, et al, (2020) The COVID-19 pandemic. Critical Reviews in Clinical Laboratory Sciences 57(6): 365-388.
2. O Albitar, R Ballouze, JP Ooi, SMS Ghadzi (2020) Risk factors for mortality among COVID-19 patients. Diabetes Research and Clinical Practice PP: 166.
3. J Sun, WT He, L Wang, A Lai, X Ji, et al. (2020) COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends in Molecular Medicine 26(5): 483-495.
4. JP Loannidis (2020) Global perspective of COVID-19 epidemiology for a full-cycle pandemic. European Journal of Clinical Investigation 50(12): e13423.
5. BJ Silk, HM Scobie, WM Duck, T Palmer, BF Ahmad, et al. (2023) COVID-19 Surveillance After Expiration of the Public Health Emergency Declaration- United States. MMWR Morb Mortal Wkly Rep 72(19): 523-528.
6. L Dial, BL Hammers (2020) Navigating the unknowns of COVID-19. Marshall Journal of Medicine 6: (2).
7. WR KhudaBuksh, CD Bastian, M Wascher, C Klaus, SY Sahai (2023) Projecting COVID-19 cases and hospital burden in Ohio. Journal of Theoretical Biology PP: 561.
8. A Lhila, F. Alghanem (2023) Along party Lines: Examining the gubernational party difference in COVID-19 mortality rates in U.S. Counties. Prev Med Rep.
9. SD Ramos, L Kannout, H Khan, L Klasko-Foster, BN Chronister, et al. (2023) A Neighborhood-level analysis of mental health distress and income inequality as quasi-longitudinal risk of reported COVID-19 infection and mortality outcomes in Chicago. Dialogues Health.
10. Obesity and COVID-19 (2022) Centers for Disease Control and Prevention.
11. L Kompaniyets, NT Agathis, JM Nelson, LE Preston, JY Ko, et al. (2021) Underlying Medical Conditions Associated with Severe COVID-19 Illness Among Children. JAMA 4(6): e2111182.
12. Ohio Department of Health (2020) Drug Overdose. Ohio Department of Health.
13. AK Slaunwhite, WQ Gan, C Xavier, B Zhao, JA Buxton (2020) Overdose and risk factors for coronavirus disease in 2019. Drug Alcohol Depend.
14. KH McConnell, A Hajat, C Sack, SJ Mooney, CM Khosropour (2023) Associations Between Insurance, Race and Ethnicity, and COVID-19 Hospitalization, Beyond Underlying Health Conditions: A Retrospective Cohort Study. AJPM Focus.
15. Ohio coronavirus map and case count. (2023) The New York Times.
16. St. Louis Fed (2020) Per Capita Personal Income by County, Annual: Ohio. FRED.
17. County Health Rankings (2020) Adult Obesity. County Health Rankings & Roadmaps.
18. United States Census Bureau (2020) Small Area Health Insurance Estimates (SAHIE). U.S. Census Bureau.
19. County Health Ratings (2020) Air Pollution- Particulate Matter.
20. Ohio Department of Job and Family Services (2020) Ohio Unemployment Rates.
21. Find My Region. Ohio Self Determination Association.
22. L Peaceful (1996) A Geography of Ohio, Kent, OH: The Kent State University Press.
23. S Scales, E Satterwhite, A August (2018) Mapping Appalachia's boundaries: Historiographic overview and digital collection. Journal of Appalachian Studies 24(1): 89-100.
24. R Exner (2020) How many hospital beds are near you? Details by Ohio county.
25. New York Times (2023) Tracking Coronavirus in Ohio: Latest Map and Case Count.
26. Mike DeWine Governor of Ohio (2020) Ohio Issues "Stay at Home" Order; New Restrictions Placed on Day Cares for Children.
27. Population of Counties in Ohio. (2023) World Population Review.
28. Individual Hospital Statistics for Ohio. (2023) American Hospital Directory.
29. R Exner (2020) How many hospital beds are near you? Details by Ohio county. Data Central.
30. M Kutner, C Nachtsheim, J Neter, W Li (2025) Applied linear statistical models. McGraw-Hill.
31. T Bihl (2017) Biostatistics using JMP: A practical guide. SAS Institute.