

# Leading Causes of Death in Texas Counties in 2021

Elena Bleicher, Valerie Routzong, Jennifer Loreda  
 Faculty Sponsor: Dr. Rebecca J. Davis  
 Stephen F. Austin State University



This Photo by Unknown author is licensed under CC BY-SA.

This Photo by Unknown author is licensed under CC BY-SA.

## Introduction:

After being given data from the County Health Rankings for Texas in 2021, we wanted to determine the leading causes of death in Texas counties for the given year by utilizing the data and other variables found to estimate multiple linear regression models that would help form a conclusion. Statistical research and literature have shown that there are several leading causes of death in the United States: heart disease, cancer, COVID-19, and unintentional accidents being the top four. After comparing our data to current literature, we found that our results only somewhat aligned with one another.

## Methodology:

Based on the data on the County Health Rankings website (<https://www.countyhealthrankings.org/app/texas/2021/downloads>), we estimate a multiple linear regression model in Excel to identify which variables have the most impact on the death rate by county in Texas for 2021. The original data includes 28 variables associated with medical, environmental, and socioeconomic causes of death. Out of those 28 variables in the data, we find that the variables most closely related to deaths in Texas counties are the following: percent of fair or poor health, percent of smokers, percent of excessive drinking, percent with access to exercise opportunities, percent of uninsured, violent crime rate, average daily PM2.5 intake, and percent of long commute who drives alone.

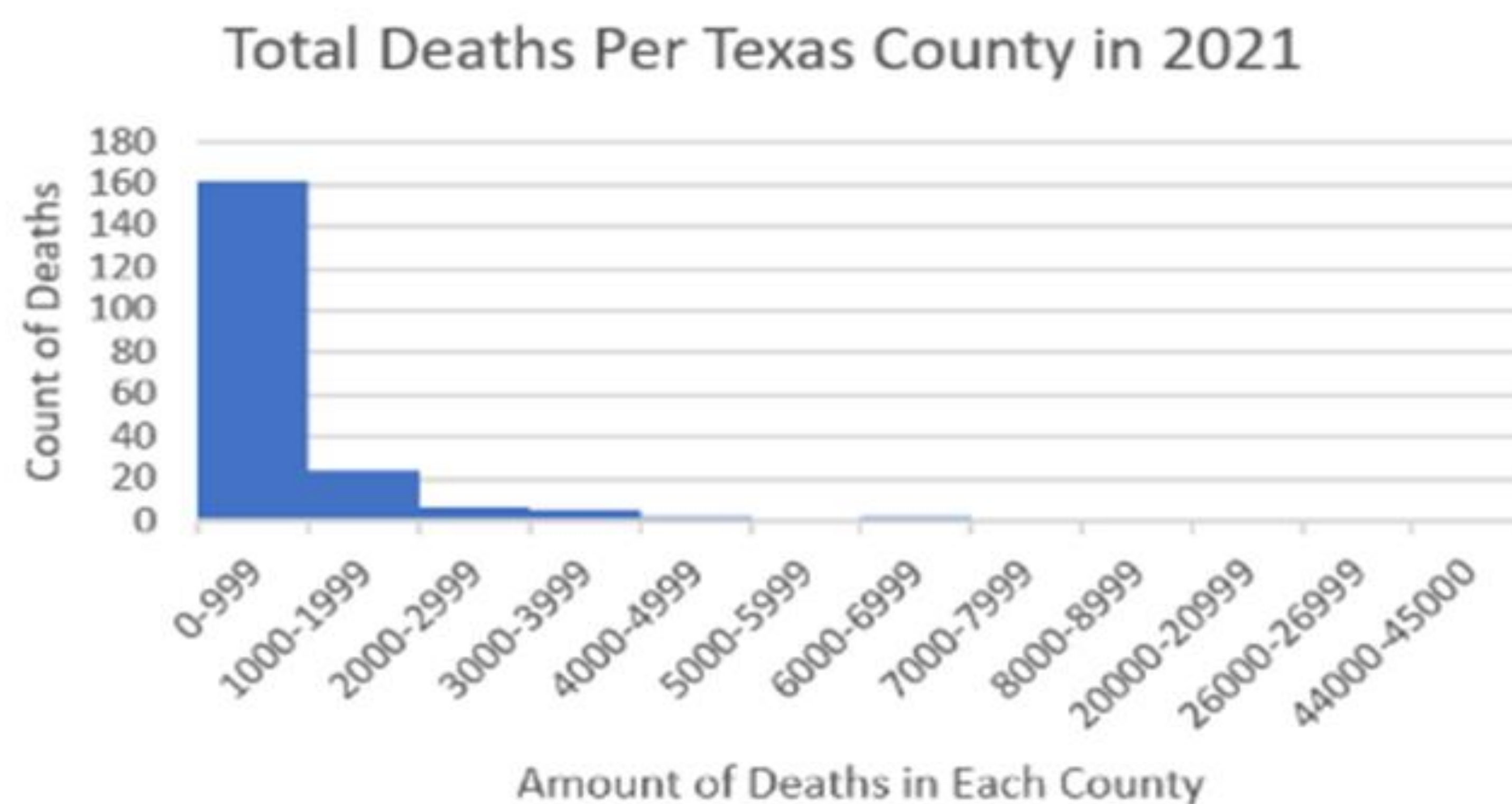
## Descriptive Statistics

Variables	Mean	Median	Standard Deviation	Minimum	Maximum	Description
Deaths	1413.89	392.00	4179.37	38.00	44591.00	n=213 right skewed
%Fair or poor health	23.47	22.87	4.87	13.54	41.00	n=213 right skewed
%Smokers	19.05	19.20	2.37	11.54	23.69	n=213 left skewed
%With access to exercise opportunities	58.86	60.14	21.31	0.00	97.57	n=213 left skewed
%Excessive drinking	18.78	18.91	1.54	13.78	22.29	n=213 left skewed
%Uninsured	20.82	20.62	3.76	12.38	32.23	n=213 right skewed
Violent Crime Rate	281.98	252.29	159.17	19.09	863.40	n=212 right skewed
Average Daily PM2.5	7.62	7.70	1.12	5.40	10.40	n=213 left skewed
% Long Commute – Drives Alone	31.91	30.10	13.50	7.60	64.90	n=213 right skewed

## Results Table:

Variables	Coefficients	Standard Errors
%Fair or poor health	-281.96**	106.69
%Smokers	-517.25***	116.69
%With access to exercise opportunities	28.96**	12.74
%Excessive drinking	-930.11***	302.19
%Uninsured	243.13***	81.75
Violent Crime Rate	6.91***	1.63
Average Daily PM2.5	868.85***	278.72
% Long Commute – Drives Alone	43.67*	22.72

## Deaths per Texas County 2021:



The death variable corresponds to the number of deaths by county. There was an average of 1,414 deaths per county in Texas for 2021 with a maximum of 44,591 deaths per county and a minimum of 38 deaths per county. The graph above represents the total deaths per Texas county in 2021.

## Equation:

$$deathsi = \beta_0 + \beta_1 \% \text{ fair or poor health}_i + \beta_2 \% \text{ smokers}_i + \beta_3 \% \text{ with access to exercise opportunities}_i + \beta_4 \% \text{ excessive drinking}_i + \beta_5 \% \text{ uninsured}_i + \beta_6 \text{ violent crime rate}_i + \beta_7 \text{ average daily PM2.5}_i + \beta_8 \% \text{ long commute - drives alone}_i + e_i$$

## Results Discussion:

Our results indicate that a higher percent fair or poor health, percent smokers, and percent excessive drinking per county are associated with lower deaths per county in Texas. Alternatively, higher percent uninsured, violent crime rate, average daily PM2.5, and percent long commute – drives alone increase deaths per county in Texas. The variables described on the left are included in our preferred model, which has the highest adjusted  $r^2$  ( $r^2=0.34973$ ) compared to other models that included the different subsets of the variables from the original dataset. The adjusted  $r^2$  shows that 34.97% of the variation in deaths per county in Texas for 2021 is explained by the preferred model. Also, all p-values from the variables in our preferred model are less than the 1%, 5%, or 10% level of significance. After analyzing and comparing the regression model results to articles included in the literature review, we determined that our findings only somewhat align with the previous literature, since the top four leading causes of death in the U.S.— heart disease, cancer, COVID-19, and unintentional accidents— were not variables included in our preferred model even though statistics show that these causes significantly affect the death rate.

## Limitations:

The original dataset from the County Health Rankings website included 254 Texas counties. However, after sorting our data, we narrowed our list to 211 Texas counties due to missing data for several variables for certain counties. Thus, we are missing 17% of the total counties in Texas, which could affect our results. If there were no missing data, the results would be a more accurate representation of Texas in 2021.