The Influence of Air Pollution on Health Expenditures by US Counties*

Does Local Air Pollution Answer for the Variations in Health Care Spending Across US Counties

Joshua Kwon

December 10, 2025

Abstract here.

Introduction

Health care spending in the U.S. reached $14,570 per person in 2023, growing 7.3 percent over the previous year according to the National Health Expenditure Accounts (NHEA). Health care costs now account for 17.3 percent of the nation's GDP, far exceeding spending levels in other high-income nations, many of which spend roughly half as much per person. As public and private expenditures continue to rise, identifying cost drivers, especially preventable ones, has become increasingly urgent.

Air pollution is one such driver with profound health implications, contributing largely to noncommunicable diseases. It is also a major contributor to respiratory and cardiovascular disease and responsible for an estimated 6.7 million deaths annually, according to the World Health Organization (WHO). The WHO also reports that in 2019, 99 percent of the world's population lived in areas exceeding recommended air quality guideline levels. Because of this, air pollution will continue to exacerbate respiratory illnesses, hospitalizations, and demands for medical treatment, placing additional burden on people's and government's health care spending.

*Project repository available at: https://github.com/peteragao/MATH261A-project-template.

Understanding how air pollution translates into health care spending is not only a public health priority but also an economic one. As health care costs rise and public budgets become strained, research clarifying these connections can help guide policy interventions aimed at both improving air quality and reducing medical expenditures. Recent studies explore the effect poor air quality has on health care expenditures. My review of the current literature summarizes the evidence linking air pollution to increased health care spending. This review forms the basis for my research plan, which will analyze the relationship between air pollution levels and county-level health care expenditures within the United States.

Literature Review

A growing body of literature has examined the economic and social costs of air pollution, particularly its effect on healthcare expenditures. Empirical studies consistently show that increased pollution levels are associated with higher household or per capita health spending even across differing countries and pollutants observed.

In China, several studies highlight the direct relationship between air quality and household healthcare expenses. Zhou, Zhong, and Yang (2022) analyze the impact of air pollution, measured by the Air Quality Index (AQI), on household medical and healthcare expenditures. Using panel data from the China Family Panel Studies (CFPS) for 2014 and 2016, they match city-level AQI data with household-level consumption data. Their analysis demonstrates that poorer air quality significantly increases household consumption, including medical expenditures, with every additional AQI unit raising consumption by 0.275 yuan. They identify two mechanisms driving this effect: households spend more on anti-pollution products and seek better living conditions in response to pollution incidents, and they incur higher medical costs to mitigate health impacts. The study further finds non-linear effects of income, consistent with precautionary savings theory, and confirms robustness using six main pollutants, CO, NO2, SO2, O3, PM2.5, and PM10.

Similarly, Anwar et al. (2022) examine developing countries and focus on ambient PM2.5 pollution as a determinant of per capita healthcare expenditures. Using World Bank data from 2000 to 2018, their regression models consider government, private, and total health expenditures, controlling for GDP per capita, education, population density, and temperature. Their findings indicate that higher PM2.5 levels increase healthcare spending, with the effect stronger for government expenditures, suggesting that public healthcare systems bear a larger burden of the pollution induced health costs.

Yang and Zhang (2018) also focus on China, estimating the marginal effect of ambient PM2.5 exposure on household healthcare expenditures using the China Urban Household Survey (UHS) and satellite-based pollution estimates. Their study accounts for both demographic and environmental confounders, including temperature. Employing both OLS and instrumental variable (IV) strategies to address endogeneity, they find that a 1 percent increase in annual PM2.5 exposure increases healthcare expenditure by 0.536 percent under OLS, rising to 2.942 percent under IV estimation. These findings underscore the substantial causal impact of air pollution on healthcare costs.

Chen and Chen (2021) contributes to these studies by noting methodological challenges in quantifying the health costs of air pollution, such as economic development, sample selection bias in hospital records, and measurement error. Using the China Health and Nutrition Survey (CHNS), they match individual-level health expenditures with city-level AQI data, controlling for individual characteristics and weather. Their analysis confirms a significant and positive effect of pollution exposure on health spending, especially for cumulative exposure over multiple weeks, with PM2.5 and PM10 exerting the largest effects.

Research in the United States and other developed countries also documents significant effects of environmental quality on healthcare spending. Apergis et al. (2018) examine the impact of per capita CO2 emissions on real per capita healthcare expenditures across all 50 U.S. states over 1966–2009. They find that the effect of emissions is stronger in states with higher healthcare spending. Since the relationship between CO2 emissions and health care expenses varies considerably across U.S. states, they find that the net benefit of reducing a unit of carbon dioxide emissions also varies significantly.

Narayan and Narayan (2008) extend the literature to selected OECD countries, and thereby wealthier countries, analyzing sulfur oxide (SOx), nitrogen oxide (NOx), and carbon monoxide (CO) emissions as determinants of healthcare spending from 1980–1999. They report significant long-run effects of SOx and CO emissions on per capita health expenditures, though the effect of NOx is statistically insignificant. Short-run effects are smaller, suggesting that the impact of pollution accumulates over time.

Across developing and developed countries, the evidence consistently shows that air pollution, whether measured via specific pollutants or composite indices, like AQI, exerts a positive effect on health expenditures. While studies differ in their data sources, measurement approaches, and estimation techniques, a common finding is that fine particulate matter, PM2.5 and PM10, and carbon monoxide are particularly impactful. Moreover, endogeneity remains an important consideration, as unobserved economic, environmental, or behavioral factors can confound estimates. This points to the need to use beyond just OLS.

Despite this extensive literature, gaps remain. The interaction between air pollution, household income, and healthcare spending remains underexplored, particularly in the U.S. context.

Data

My dataset contains 665 observations of key air pollution measurements, healthcare expenditures, and median household income across 113 counties for the years 2010-2019. I selected six air pollutants, based on their use in the United States Air Quality Index (AQI), to reflect the air pollution in the observed counties. This lets me observe the effect of each pollutant independent of each other. These pollutants are carbon monoxide, CO; nitrogen dioxide, NO2; ozone, O3; sulfur dioxide, SO2; and fine particulate matters PM2.5 and PM10. These are reported in units of ppm, ppb, ppm, ppb, ug/m3 and ug/m3 respectively. Measurements for these pollutants come from the U.S. Environmental Protection Agency (2025) Air Quality Statistics Report and are contained in the `AQSR2010.csv` through `AQSR2019.csv` datasets. They report the 2nd highest non-overlapping 8-hour average in the year for CO; the annual mean of all the 1-hour measurements in the year for NO2; the 4th highest daily max 8-hour average in the year for O3; the annual mean of all the 1-hour measurements in the year for

SO2; and the Weighted Annual Mean (mean weighted by calendar quarter) for the year for PM2.5 and PM10.

Estimates for the county-level healthcare expenditures are provided by the Institute for Health Metrics and Evaluation (2025) and contained in the `IHME_estimates.csv` dataset. They estimate the per capita healthcare expenditures of 3110 counties across the years 2010-2019, adjusting their estimates to 2019 prices. The authors of this data calculated their estimates from the over 40 billion insurance claims and nearly 1 billion facility records they accessed across Medicare, Medicaid, private insurance, and out-of-pocket payments.

I collected county-level income data from the `SAIPE_income.csv` dataset provided by the U.S. Census Bureau (2025)'s Small Area Income and Poverty Estimates (SAPIE) Program which reports on the median household income for every US county. I adjusted these values for inflation to 2019 prices to better match my healthcare expenditure data.

County-level health expenditure data was only available for the years 2010-2019, so I limited my air pollution and income data to those same years and matched the datasets together by year and county. Due to inconsistent monitoring of air pollution, data for all pollutants in every county for each year was not available. Observations that did not have measurements for all six pollutants were dropped. Observations for counties without data for all 10 years were included however, and as such, this dataset is unbalanced. Additionally, states can flag data that may have been influenced by an exceptional event such as high winds or wildfire. These data points were excluded for this analysis.

State Summary Table		
Metric	State	Count
Number of unique states	-	47
Most common state	CA	105
Least common state	ID	1
Average counties per state	-	14

Table 1: Frequency statistics of states included in the dataset.

Monitoring sites and their capabilities are not uniformly distributed across the US. As seen in Table 1, not every state is included and California is over represented. The trends in California will have an amplified effect on my analysis and bias my results.

In 2010, the Patient Protection and Affordable Care Act (ACA) was passed. Estimates for healthcare expenditures all fall within this period, so I do not expect this to have largely affected my data or results.

In 2004, the SAIPE Program changed from using the Current Population Survey to the American Community Survey; however, no such changed occurred between the years 2010-2019 and I do not expect this to have influenced the data.

Summary Statistics					
variable	mean	sd	median	min	max
CO	1.54	0.86	1.40	0.10	5.80
NO2	11.88	5.74	11.00	2.00	32.00
SO2	1.02	1.21	1.00	0.00	7.00
O3	0.07	0.01	0.07	0.04	0.12
PM2.5	9.48	2.50	9.20	2.20	18.40
PM10	22.18	9.83	20.00	5.00	57.00
Healthcare Expenditures	6763.45	1156.25	6759.35	3740.64	11977.94
Median Household Income	61955.42	14037.83	59369.48	35726.18	132444.00

Table 2: Summary statistics for key variables, including air pollutants, per-capita healthcare expenditures, and median household income. Year and County identifiers are excluded.

As seen in Table 2, the max values of many pollutants greatly exceed the median values. Even after removing exceptional events from the dataset, maximum values for some pollutants, notably CO, NO2, and PM10, greatly exceed the median. These point to several outliers in pollutant measurements that I expect will affect my analysis. Outliers in healthcare expenditures and median household income are visualized in Figure 1. Notably, the county with the highest median household income outpaces the average increase in other counties. These will also affect my analysis.

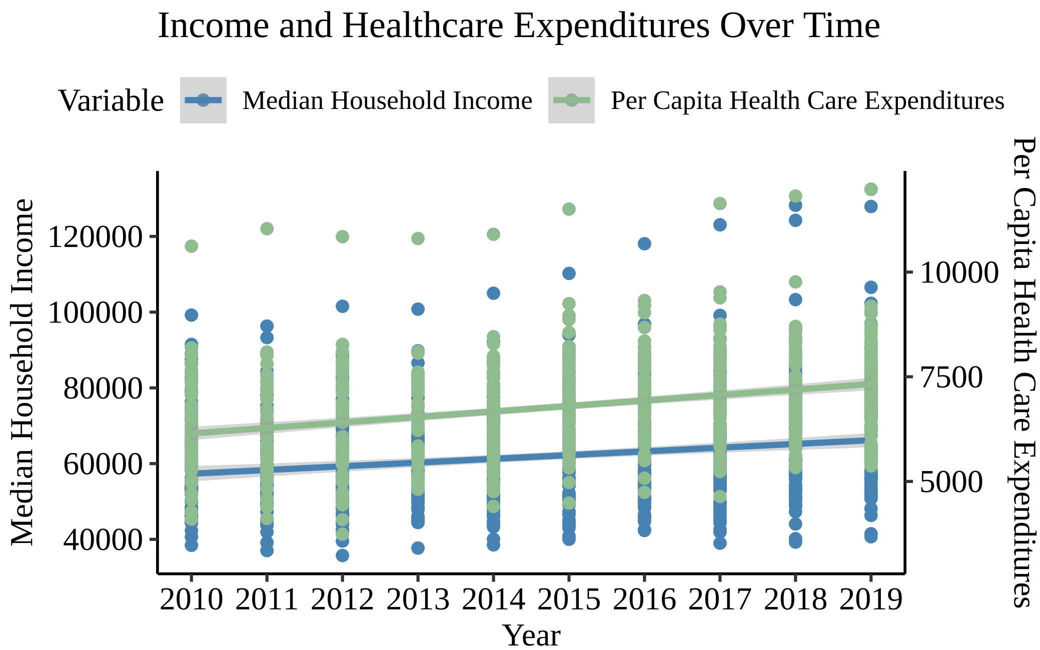

Figure 1: Scatter plot of median household income and healthcare spending in US dollars across the years 2010-2019 for the 665 observations in the dataset. Regression line is included to better illustrate the change in median household income and healthcare spending.

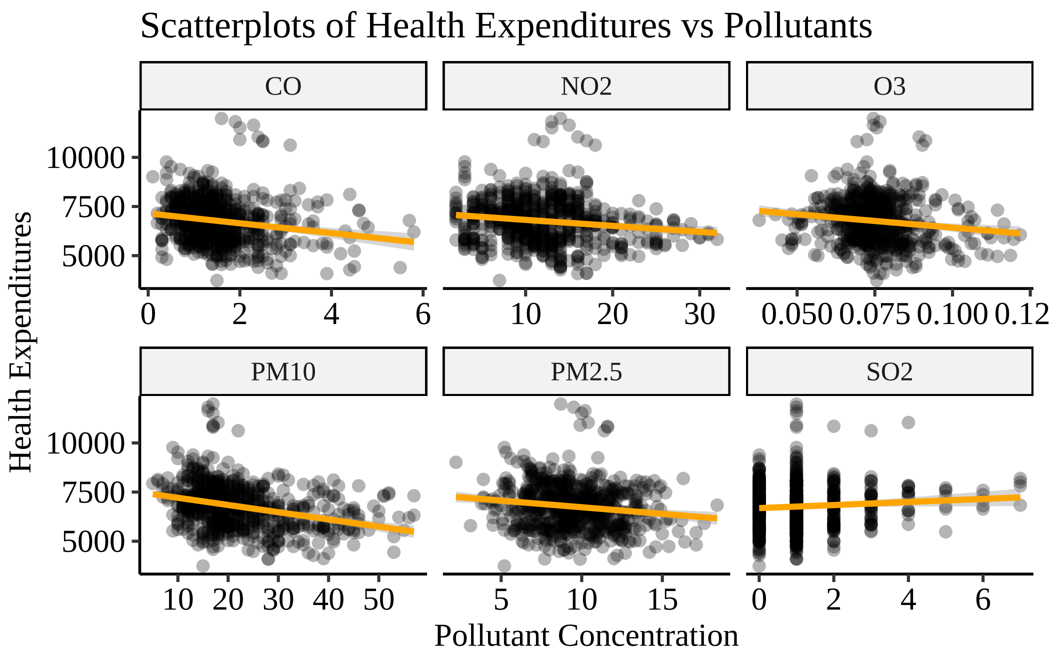

Figure 2: Scatter plot of healthcare spending versus each of the air pollutants specified. Regression line is included to better illustrate the change in healthcare expenditures.

Many of the trends in Figure 2 indicate that there is a negative relationship between air pollution and healthcare expenditure, against the findings of previous studies. I expect this is due to the limitations in my data.

Methods

To investigate the relationship between air pollution and per capita healthcare expenditures across U.S. counties, I use a multiple linear regression model. The outcome variable is the county-level per capita healthcare spending. The key predictor variables are the concentrations of six air pollutants, carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and fine particulate matter (PM2.5 and PM10). Because socioeconomic conditions are known to influence healthcare spending, median household income (HI) is included as a control variable. The model also includes county and year fixed effects. County fixed effects account for time-invariant demographic and economic characteristics that differ across counties. Year fixed effects capture nationwide trends that affect all counties in a given year. The model is expressed as

$$Y_{it} = \beta_0 + \beta_1 CO_{it} + \beta_2 NO2_{it} + \beta_3 SO2_{it} + \beta_4 O3_{it} + \beta_5 PM2.5_{it} + \beta_6 PM10_{it} + \beta_7 HI_{it} + \gamma_i + \delta_t + \varepsilon_{it}$$

where Y_{it} denotes per capita healthcare expenditure in county i during year t. The intercept term β_0 represents the expected per capita healthcare expenditure when all predictors are equal to zero for the reference county and reference year. Coefficients β_1 through β_7 represent the expected change in per capita healthcare expenditures associated with a one unit increase in the respective predictor, holding all other variables constant and controlling for county and year effects. The terms $ □_i$ and δ_t represent county fixed effects and year fixed effects, respectively.

We assume the error term, ε_{it} is random with mean 0 and finite variance σ^2, conditional on the fixed effects. The validity of the OLS estimates relies on standard assumptions, including linearity, independence of errors, homoscedasticity, normality of the errors, and no perfect multicollinearity among predictors.

The parameters are estimated using the method of least squares which minimizes the sum of the squared residuals

$$Q = \sum_{i,t} (Y_{it} - \beta_0 - \sum_{j=1}^{7} \beta_j X_{ijt} - \gamma_i - \delta_t)^2$$

producing fitted values $\hat{Y_i}t = \hat{\beta}_0 + \sum_{j=1}^{7} \hat{\beta}_j X_{ijt} + \hat{\gamma}_i + \hat{\delta}_t$ that best predict healthcare expenditures based on pollutant concentrations and income after accounting for county and year fixed effects.

We perform hypothesis tests to determine whether each predictors has a statistically significant association with healthcare spending. For each predictor, we define the null hypothesis as $H_0 : \beta_j = 0$, indicating no linear relationship with healthcare expenditures, holding all other variables and fixed effects constant. We define the alternative hypothesis as $H_A : \beta_j \neq 0$, indicating that a linear relationship exists. All tests are evaluated at the 5% significance level.

Because the sources of air pollutants are often highly correlated, we expect to see evidence of multicollinearity which can destabilize OLS estimates. To detect multicollinearity among predictors, I use variance inflation factors (VIF). These factors measure how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related.

To address this, we use a Ridge regression, which applies L2 regularization to shrink coefficient estimates toward zero without eliminating predictors, reducing variance and improving the stability of estimated effects. Ridge regression modifies the method of least squares to allow biased estimators of the regression coefficients. Ridge regression estimates are obtained by the method of penalized least squares. The penalized least squares criterion, given by

$$Q = \sum_{i,t}(Y_{it} - \beta_0 - \sum_{j=1}^{7} \beta_j X_{ijt} - \gamma_i - \delta_t)^2 + \lambda \sum_{j=1}^{7} \beta_j^2$$

combines the usual sum of squared errors with a penalty for large regression coefficients. The penalty parameter for Ridge regression is selected via cross-validation to minimize prediction error.

One limitation with our use of VIFs is that they cannot distinguish between several simultaneous multicollinearities, but they are still valuable as a method to detect multicollinearity. A major limitation of ridge regression is that ordinary inference procedures are not applicable and exact distributional properties are not known. Bootstrapping can be employed to evaluate the precision of ridge regression coefficients, but this was not performed in this paper. Another limitation of ridge regression is that the choice of the biasing constant, λ, is a judgmental one.

Both OLS and Ridge regression models are implemented in the R programming language (R Core Team 2023).

Results

Table 3 shows the results of the multiple linear regression. Using a 5% significance level, only SO2 and PM10 were positively and significantly associated with per capita healthcare spending. Specifically, a one-unit increase in SO2 measurements was associated with an increase of 54.9 dollars in per capita healthcare expenditures, and a one-unit increase in PM10 was associated with an increase of 5.77 dollars

Regression Table		
Predictor	Estimate	Pr(>|t|)
Intercept	4642.8331	0.0000
CO	-35.8485	0.0518
NO2	-7.0213	0.1033
SO2	54.9081	0.0002
O3	3590.1495	0.0997
PM2.5	-1.0095	0.9041
PM10	5.7674	0.0239
Median Household Income	0.0098	0.0084

Table 3: Coefficients from OLS regression of per capita healthcare expenditures on air pollutants and household income, including county and year fixed effects.

in per capita healthcare expenditures. These results provide statistical evidence supporting a positive relationship between these pollutants and spending.

CO and O3 showed weaker associations with spending, but were not statistically significant at the 5% level. In contrast, NO2 and PM2.5 showed no evidence of an association with healthcare spending.

Some of our regression results are counter intuitive. We would not expect increases in any pollution concentration to decrease healthcare expenditures, and many of these pollutants have established negative health effects, so non significance is unexpected. These results reaffirm our suspicions of multicollinearity, so I use variance inflation factors to measure this.

The variance inflation factor (VIF) results in Table 4 indicate moderate multicollinearity, and we would expect OLS to produce unstable coefficients. Several pollutants, notably NO2, O3, and PM10, have VIF greater than three and median household income has a VIF greater than six. To address this, we supplement our OLS with a ridge regression.

We examined the associations between air pollutants, median household income, and per capita healthcare spending using ridge regression. Predictors were standardized for clearer comparisons of the effects, and the results are shown in Table 5. Median household income had the largest positive effect on spending, with a one standard deviation increase associated with an increase of $179 in per capita

VIF Table			
Variable	GVIF	Df	GVIF_Scaled
CO	3.91	1	1.98
NO2	9.64	1	3.10
SO2	4.82	1	2.20
O3	9.08	1	3.01
PM2.5	6.92	1	2.63
PM10	9.89	1	3.15
Median.Household.Income	43.01	1	6.56

Table 4: Variance Inflation Factors (VIF) for predictors in the regression of per capita healthcare expenditures on air pollutants and median household income. County and year fixed effects are included but omitted from this table.

Ridge Table	
Variable	Coefficient
CO	-50.3414
NO2	-46.6607
SO2	54.3656
O3	0.7389
PM2.5	-5.0785
PM10	-26.2765
Median.Household.Income	179.4630

Table 5: Ridge regression coefficients for per capita healthcare expenditures on air pollutants and median household income. County and year fixed effects included but not shown.

healthcare expenditures. Among air pollutants, SO2 showed the greatest positive association with each standard deviation increase associated with a $54 increase in spending. Surprisingly, CO and NO2 continued to exhibit large negative associations. PM10 and PM2.5 also had negative effects, but to a lesser extent. O3 had a negligible effect on healthcare expenditures.

Discussion

References

Anwar, Asim, Shabir Hyder, Russell Bennett, and Mustafa Younis. 2022. "Impact of Environmental Quality on Healthcare Expenditures in Developing Countries: A Panel Data Approach." *Healthcare* 10 (9). https://doi.org/10.3390/healthcare10091608.

Apergis, Nicholas, Rangan Gupta, Chi Keung Marco Lau, and Zinnia Mukherjee. 2018. "U.s. State-Level Carbon Dioxide Emissions: Does It Affect Health Care Expenditure?" *Renewable and Sustainable Energy Reviews* 91: 521–30. https://doi.org/https://doi.org/10.1016/j.rser.2018.03.035.

Chen, Fanglin, and Zhongfei Chen. 2021. "Cost of Economic Growth: Air Pollution and Health Expenditure." *Science of The Total Environment* 755: 142543. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.142543.

Institute for Health Metrics and Evaluation. 2025. "United States Health Care Spending by Health Condition and County 2010-2019." Institute for Health Metrics; Evaluation.

Narayan, Paresh Kumar, and Seema Narayan. 2008. "Does Environmental Quality Influence Health Expenditures? Empirical Evidence from a Panel of Selected OECD Countries." *Ecological Economics* 65 (2): 367–74. https://doi.org/https://doi.org/10.1016/j.ecolecon.2007.07.005.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

U.S. Census Bureau. 2025. "Small Area Income and Poverty Estimates (SAIPE) Interactive Data Tool." https://www.census.gov/data-tools/demo/saipe/#/.

U.S. Environmental Protection Agency. 2025. "Air Quality Statistics Report." https://www.epa.gov/outdoor-air-quality-data/air-quality-statistics-report.

Yang, Jing, and Bing Zhang. 2018. "Air Pollution and Healthcare Expenditure: Implication for the Benefit of Air Pollution Control in China." *Environment International* 120: 443–55. https://doi.org/https://doi.org/10.1016/j.envint.2018.08.011.

Zhou, Li, Qian Zhong, and Jingjing Yang. 2022. "Air Pollution and Household Medical Expenses: Evidence from China." *Frontiers in Public Health* Volume 9 - 2021. https://doi.org/10.3389/fpubh.2021.798780.