

Air Pollution and Avoidance Behavior: Evidence from Daily Activities in the US*

Xinhui Sun[†]

September, 2024

Abstract

Air pollution is costly, yet empirical evidence on how individuals take action to reduce pollution exposure is limited. I investigate how people modify their daily activities in response to air pollution. Using over 600 million phone-location based foot traffic data from SafeGraph, I conduct a large-scale analysis to examine the causal effect of air pollution on daily activities across the United States. Using changes in local wind direction as an instrumental variable (IV) for air pollution, I find that a $1 \mu\text{g}/\text{m}^3$ increase in PM2.5 concentration leads to a 0.50% decrease in daily activities, resulting in a loss of 117,642 visits per day nationwide. Recreational activities are the most affected, likely due to their flexibility compared to other daily activities. This reduction is more pronounced in counties with higher incomes and larger proportions of children, suggesting better awareness among wealthier or more vulnerable groups.

*I am grateful to Tatyana Deryugina, David Molitor, and Julian Reif for invaluable guidance and support. This paper has also benefited from comments and suggestions by Alex Bartik, George Deltas, Eliza Forsythe, Russell Weinstein, Ben Marx, Adam Osman, and seminar participants at the Applied Micro Research Lunch. All errors are my own.

[†]Department of Economics, University of Illinois at Urbana-Champaign. xinhuis2@illinois.edu

1 Introduction

Air pollution is considered the world’s biggest environmental health threat¹. Accurately quantifying these effects is crucial for understanding the costs of air pollution and determining optimal environmental policies. Exposure to air pollution can induce premature mortality (Currie and Neidell, 2005; He et al., 2016; Deryugina et al., 2019), reduce labor productivity (Graff Zivin and Neidell, 2012; Borgschulte et al., 2018; He et al., 2019), and decrease hours worked (Hanna and Oliva, 2015; Aragón et al., 2017). While the negative effects of air pollution have been widely discussed, previous literature has predominantly focused on health and labor outcomes. However, its impacts on other aspects, particularly those without explicit market values, are often neglected. For example, individuals might take actions to mitigate the negative health impact of air pollution. Previous studies on avoidance behavior are typically confined to narrow geographic areas or specific activity types, raising questions of external validity. Such studies also lack the power to detect heterogeneity of avoidance behavior across different industries, as well as the distributional effects across different demographic groups.

I conduct the first large-scale estimation of the causal effect of day-to-day air pollution fluctuations on daily activities in the United States. I overcome the challenges described above by utilizing foot traffic patterns from SafeGraph, which aggregates de-identified geospatial data from millions of US smartphones. Specifically, I assemble a dataset by combining data from over 600 million trips, covering a broad range of industries, with satellite-based air pollution data for the continental United States from 2018 to 2021. A primary identification concern when estimating the effect of air pollution is endogeneity. For instance, existing estimates may be biased due to reverse causality, as visitation-related traffic could increase local air pollution. To address this concern, I use changes in wind direction as an instrumental variable (IV) for air pollution to derive the causal effect on daily activities.

I find that, on average, a 1 $\mu\text{g}/\text{m}^3$ (about 10 percent of the mean) increase in PM2.5 concentration leads to a 0.50% decrease in daily activities. This translates to a nationwide loss of 117,642 visits per day. This estimate remains robust across various specifications and is significant across different industries. Among all sectors, the entertainment and recreation industries experience the most substantial decline, likely due to the greater flexibility in recreational activities compared to other daily

¹<https://unece.org/air-pollution-and-health>. Accessed Jan 2, 2022.

activities. Additionally, the reduction in visits is generally more pronounced for outdoor facilities than for indoor ones, presumably because outdoor activities intensify the negative health effects of pollution due to increased respiration and exposure.

Moreover, I examine the heterogeneity across different income levels and demographics. The reduction in daily activities is more pronounced in counties with higher incomes and larger proportions of children, suggesting that wealthier or more vulnerable groups may have a greater awareness of elevated air pollution levels. Furthermore, I find that the Black population shows a smaller response to air pollution, which might partially explain the greater negative health impact of air pollution on Black people compared to White people (Alexander and Currie, 2017; Gillingham and Huang, 2021). These findings highlight the endogeneity of avoidance behavior, which might exacerbate environmental justice issues. Specifically, low-income and minority groups respond less to alleviating pollution levels; given these groups are more likely to live in more polluted areas (Mikati et al., 2018; Heblich et al., 2021; Tessum et al., 2021), they are in need of targeted policy interventions to mitigate the unequal burden of air pollution.

I also explore two potential channels for the negative relationship between air pollution and daily activities. First, using monitor-based PM2.5 data, I find suggestive evidence that air quality advisories reduce daily activities, although the effect is weak and statistically insignificant. Additionally, I present evidence that visibility serves as a potential channel through which pollution reduces daily visits: specifically PM2.5 reduce visibility and then visibility reduce daily visits. Specifically, increased PM2.5 levels reduce visibility, which in turn decreases daily visits. This suggests that people might perceive poor air quality through reduced visibility and, as a result, choose to spend more time at home rather than going out.

This paper provides the first large-scale estimation of the causal effect of air pollution on daily activities in the United States. Existing literature studying the effect of air pollution on daily activities is generally based on a limited sample from a specific region (Bresnahan et al., 1997; Zivin and Neidell, 2009), or a specific activity type, such as visits to national parks (Keiser et al., 2018), camping (Gellman et al., 2022) or movie watching (He et al., 2022), which makes the generalizability of these estimates worth questioning. In contrast, my analysis uses nationwide phone-location based data and focuses on visits to a broader range of facilities, making it more representative than previous studies. My findings suggest a slightly smaller effect than what has been previously reported.

Moreover, this paper reinforces the importance of characterizing avoidance behavior when quantifying the externalities of air pollution. The results show that individuals, especially vulnerable groups, are actively engaging in self-protective behaviors to reduce their exposure to air pollution, such as spending more time at home. Consequently, estimates that ignore these behaviors are likely biased downward, as the health impact would be greater without such avoidance actions. Given the debates about air quality regulations, accounting for these behaviors is crucial to accurately measure the cost of pollution. From a policy perspective, researchers often view avoidance behavior as the primary policy target and are intensively investigating ways to promote voluntary self-protection (Lee et al., 2020). However, avoidance behavior itself can be costly, either in terms of increased expenditures or utility losses. For example, when individuals choose to stay at home and limit their daily activities, it could lead to depression, resulting in additional societal costs. Furthermore, there could be a decrease in GDP stemming from affected industries. Accurately quantifying these effects is important for understanding the costs of air pollution and determining optimal policies.

Finally, this paper investigates whether individuals respond to day-to-day pollution fluctuations, adding to the relatively understudied topic on avoidance behavior. Previous studies show that air quality alerts prompt avoidance behavior (Neidell, 2009; Altindag et al., 2017). However, air quality warnings are rare and triggered only when the Air Quality Index exceeds a certain level, while negative effects of air pollution increase even before this threshold (Zivin and Neidell, 2009). Therefore, how individuals respond to these alerts does not necessarily correspond to how they respond to air quality itself. Without an air quality alert, individuals might not be aware of the elevated pollution levels. I find that people respond to daily pollution fluctuations, suggesting that providing timely, accessible real-time pollution information could effectively safeguard public health.

The rest of the paper is organized as follows. Section 2 describes the data and provides summary statistics. Section 3 introduces the empirical strategy in detail. Section 4 presents the main results, discusses the heterogeneity, explores potential channels, and provides back-of-the-envelope calculation. Section 5 presents the robustness checks. Section 6 concludes.

2 Data

The data used in the paper come from three main sources: mobile phone-based visitation data from SafeGraph, satellite-based air pollution from CAMS global reanalysis (EAC4), and satellite-based weather data from EAC4 and ECMWF Reanalysis (ERA5). The linkages and further details are described below.

2.1 Visitation Data

I obtain the visitation data from SafeGraph². The dataset includes information collected from over 45 million smart mobile devices and covers over 3.6 million Points of Interest (POI) across the United States. In total, I obtain over 600 million trips from January 1, 2018, to December 30, 2021, across the United States.

SafeGraph conducts a data quality evaluation by comparing its demographic data with the American Community Survey (ACS) data from the US Census and found that their data are statistically representative of the population at the county level and above (Squire, 2019; Chang et al., 2022). Therefore, for empirical analysis, I match each location to its county based on latitude and longitude, and then aggregate visits at the county level. After aggregating all visits at the county level, I have 4,561,337 county-day observations. Additionally, since the dataset includes industry categories based on the North American Industry Classification System (NAICS) code, I was able to analyze foot traffic to different categories separately. The number of county-day observations for each category is shown in Table 1³.

Because this dataset does not contain socioeconomic and demographic information about mobile device users for privacy protection reasons, I obtain county-level population and income data from the United States Census Bureau. Since county populations vary widely and more populated counties tend to have more visits, I use visitation rates rather than raw visitation numbers as the dependent variable. To calculate county-level visitation rates, I aggregated the total number of visits in each county and then divided by the county’s total population.

²<https://www.safegraph.com/>. Accessed Sep 12, 2023.

³Note that the number of observations varies across categories because some counties may not have facilities of certain types.

2.2 Air Pollution Data

Although the United States Environmental Protection Agency (EPA) has been reporting air quality and other atmospheric data since 1970 and the number of pollutant monitors has increased over the years, there are still limitations: more than half of the monitors collect data on a 1-in-3 day schedule or 1-in-6-day schedule⁴, resulting in a lack of data on certain days. Interpolating the missing data on these days can lead to bias, as air quality on unmonitored days is found to be worse than on monitored days due to strategic responses (Zou, 2021).

Therefore, rather than using monitor-based data, I use satellite-based air pollution data from the EAC4 reanalysis database⁵. EAC4 reports PM2.5 and other atmospheric data every 3 hours with a 0.75×0.75 ($\approx 81km \times 81km$) resolution, which is derived from the combination of satellite observation and computer simulation of the atmosphere. I construct the county-level daily PM2.5 level in the following manner: for counties that have multiple satellite data, I average the gridded values overlapping each county; for counties that do not have satellite data, I interpolate their PM2.5 levels using inverse distance weighting (IDW) base on their latitude and longitude. Then, I match the visitation data with the air pollution data using county code and date. Figure A1 shows the average county-level visit rates and PM2.5 levels from January 1, 2018, to December 30, 2021. As a few counties do not have any visitation data for leisure facilities from Safegraph, there are some missing values in the figure.

2.3 Weather Data

The analysis in this paper contains a flexible set of control variables for weather, including temperature, wind speed, and precipitation. Additionally, wind direction is used as the instrument of PM2.5 concentrations.

Daily temperature, wind direction, and wind speed data are also obtained from the EAC4 reanalysis database. I average the daily measures across all grid points in a particular county to obtain the county-level daily measure. For counties without satellite data, I interpolate their temperature, wind direction, and wind speed using IDW based on their latitude and longitude. Specifically, wind direction and wind

⁴See EPA’s Sampling Schedule Calendar: <https://www.epa.gov/amtic/sampling-schedule-calendar>. Accessed October 10, 2022

⁵See <https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=overview>. Accessed September 19, 2022.

speed are constructed using the East-West wind vector (u-wind) and the North-South wind vector (v-wind) provided in the database⁶. Wind direction is defined as the direction the wind is blowing from.

In addition, I obtain precipitation data from the Copernicus ERA5 reanalysis hourly databases. Precipitation data are reported on a 0.25×0.25 degrees grid ($\approx 27km \times 27km$). I constructed the county-level daily precipitation by averaging the hourly data on a given day with grid points within a particular county. For counties without satellite data, I interpolate their precipitation using IDW based on their latitude and longitude.

2.4 Summary Statistics

Table 1 displays the summary statistics for the main estimation sample, which consists of 4,561,337 county-day observations. The average daily visit rates⁷ to all POI within a county is 69.54 per 1000 people⁸. The Retail Trade sector has the highest mean visit rate of 15.46 per 1,000 people. The average daily concentration of PM2.5 is $11.50 \mu g/m^3$ ⁹, with a standard deviation of 16.13.

⁶Note that wind directions and speed are vectors, so they cannot be averaged or interpolated numerically. Therefore, when averaging or interpolating, I first take the average of the two vectors and then calculate the average wind direction and wind speed using the average vectors.

⁷The raw number of visits before aggregating to the county level is reported in Table A1

⁸Note that the data from Safegraph were collected from over 45 million mobile devices, which is around 14% of the US population.

⁹This is slightly higher than the average PM2.5 concentration calculated using EPA Ground monitors. One possible explanation for this discrepancy is strategic monitor placement, as discussed in Graginer et al. (2018).

Table 1. Summary Statistics

Variables	Mean	SD	N
Visit Rates (per 1,000 people)			
All POIs	69.93	41.53	4,513,600
44-45: Retail Trade	22.50	14.28	4,506,558
72: Accommodation and Food Services	14.58	12.39	4,491,452
61: Educational Services	8.03	8.84	4,466,749
71: Arts, Entertainment, and Recreation	6.48	9.13	4,322,150
53: Real Estate and Rental and Leasing	5.73	7.44	3,858,106
62: Health Care and Social Assistance	4.88	3.86	4,449,158
81: Other Services	4.54	6.01	4,495,995
92: Public Administration	1.11	1.55	4,468,576
52: Finance and Insurance	0.73	0.72	4,282,607
48-49: Transportation and Warehousing	1.33	2.62	4,482,408
51: Information	0.46	0.78	3,944,703
42: Wholesale Trade	0.45	0.76	3,677,317
31-33: Manufacturing	0.87	2.43	3,875,683
54: Professional, Scientific, and Technical Services	0.41	0.48	4,040,757
22: Utilities	0.24	0.63	2,228,604
23: Construction	0.31	0.50	3,614,912
55: Management of Companies and Enterprises	0.31	1.70	1,383,221
21: Mining, Quarrying, and Oil and Gas Extraction	0.17	0.20	8,750
56: Administrative and Support and Waste Services	0.18	0.34	3,009,650
11: Agriculture, Forestry, Fishing and Hunting	0.11	0.27	658,854
Pollution			
PM2.5 ($\mu g/m^3$)	11.48	16.05	4,513,600
Weather			
Temperature ($^{\circ}C$)	13.61	10.71	4,513,600
Total Precipitation (mm)	0.30	0.69	4,513,600
Wind Direction (degrees)	193.49	94.62	4,513,600
Wind Speed (m/s)	2.72	1.55	4,513,600
Visibility (km)	17.52	3.93	4,513,600
Demographic			
Population	105,050	336,537	4,513,600
Age under 5 (%)	5.84	1.19	4,512,144
Black (%)	4.56	7.29	4,512,144
Per Capita Income	26,011	6,214	4,512,144

3 Empirical Strategy

To investigate the impacts of air pollution on daily activities, I fit a fixed-effect Ordinary Least Squares (OLS) using the equation:

$$\log\left(\frac{Y_{ct} + 1}{Pop_c}\right) = \alpha \times PM2.5_{ct} + \mathbf{X}_{ct}'\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \quad (1)$$

where c indexes county, t indexes time which is at the daily level, y indexes year, and m indexes month. The outcome variable is the visit rate, which is calculated using the visits to all leisure facilities in county c on date t (Y_{ct}) divided by the county population in the year 2020 (Pop_c). Since the outcome variable is highly right-skewed, I perform a log transformation (Y_{ct} is added by 1 to avoid numerical errors when taking the log transformation). The coefficient of interest is $PM2.5_{ct}$, which is the average PM2.5 level in county c on date t . Control variable \mathbf{X}_{it} includes other weather variables, such as temperature, precipitation, and wind speed. To minimize concerns about autocorrelation, I include one lead and one lag of the weather controls, as well as PM2.5 (OLS) or the instruments (IV). My results are robust to different forms of weather controls.

In addition, I include a rich set of fixed effects, including county-by-year fixed effect σ_{cy} , county-by-month fixed effect η_{cm} , month-by-year fixed effect θ_{my} and day-of-week fixed effect γ_w . Specifically, county-by-year fixed effects σ_{cy} pick up within-year variations in county-level factors that determine visits but are not captured by the control variables, such as demographic characteristics and economic conditions. County-by-month fixed effects η_{cm} control for seasonal unobservables across counties, such as different peak seasons due to different geographic features. Day-of-week fixed effect γ_w pick up cyclical visit patterns within week. Lastly, month-by-year fixed effect θ_{my} captures the time-varying shocks that are common in each month, such as economic recessions and pandemic outbreaks. I also examine the robustness of the results by including different fixed effects. The standard errors are clustered at the county level.

The coefficient α captures the impact of air pollution on leisure activities. The identification assumption is that, conditional on control variables and fixed effects included in equation (1), unobserved determinants of visit rates (ϵ_{ct}) are independent of variation in PM2.5. However, such an assumption can be violated since local air pollution is endogenous to local activities. For instance, as more people drive to the parks, emissions around the park will increase, thereby biasing the estimate of α .

To address this concern, I leverage the pollution variation due to changes in wind patterns to identify pollution impacts. Specifically, since wind directions are random, I use the changes in wind direction as an instrumental variable for air pollution to derive the causal relationship (Deryugina et al., 2019). The assumption of this approach is that after controlling for covariates and fixed effects, changes in wind direction only affect people’s leisure activities through their effects on air pollution. The specification for the first stage is:

$$PM2.5_{ct} = \sum_{g \in G} \sum_{b=0}^3 \gamma_b^g \mathbf{1}[G_c = g] \times WindDir_{ct}^{90b} + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \quad (2)$$

In equation (2), the instrument variable is constructed in the following manner. $WindDir_{ct}^{90b}$ equals 1 if wind direction in county c falls in the 90-degree interval $[90b, 90b+90)$ and 0 otherwise. To allow the effect of the wind instruments on PM2.5, denoted as γ_b^g , to vary across geographic regions, I use the K-means clustering algorithm to classify counties into 20 spatial groups based on their latitude and longitude. The clustering result is shown in Figure A3. $\mathbf{1}[G_c = g]$ equals 1 if county c is classified into monitor group g and 0 otherwise. Other control variables X_{ct} and fixed effects are defined as in equation (1).

Equation (2) restricts the effect of wind direction on pollution to be the same for all counties within each geographic cluster. Intuitively, non-local sources located outside of the cluster are more likely to have similar effects on pollution levels in all (or most) counties in the cluster group. As a result, Equation (2) is more likely to capture the pollution variation driven by non-local sources. This is advantageous because pollution driven by local sources may not affect all individuals residing within the area in the same way, leading to measurement error¹⁰. In section 5, I provide evidence that the pollution variation I employed is primarily driven by non-local sources. Therefore, the effect of wind direction on pollution should be similar for all counties in the same geographic group. I employ 4 bins and 20 clusters for computational ease. The results are robust to varying the number of wind direction bins and geographic clusters (Table 7).

¹⁰Consider a local pollution source located in the center of a cluster. When the wind blows from the west, counties to the west of this source will record low pollution levels, and counties to the east will record high pollution levels. On net, a researcher who uses such variation may conclude that short-term pollution fluctuations have no effect on visit rates to leisure facilities.

4 Results

4.1 Main Effect

I find a significant negative relationship between air pollution and visit rates to leisure facilities. Table 2 displays the results from both fixed-effects models and instrumental variable models. For the IV strategy, I use daily changes in county-level wind direction as an instrument for daily changes in county-level PM2.5 concentrations. The first-stage F-statistic in Column 2 is 118.3, which implies the issue of the weak instrument is not a problem in this approach. Since PM2.5 is endogenous, I rely on the IV approach as the preferred empirical strategy. The estimate in Column 2 implies that a 1 $\mu\text{g}/\text{m}^3$ increase in PM2.5 leads to a 0.50% decrease in visit rates on average. In addition, Table 2 shows that warmer temperatures increase visit rates, whereas precipitation and strong wind reduce visit rates.

The IV estimate in Column 2 is substantially larger than the OLS estimate in Column 1, suggesting that OLS estimation suffers from significant bias. This downward bias is common in quasi-experimental studies on air pollution and is generally thought to be, at least in part, due to measurement errors in pollution exposure ([Deryugina et al., 2019](#); [Alexander and Schwandt, 2022](#)).

Table 2. Effect of PM2.5 on Daily Visit Rates

	(1) OLS log(visit rate) \times 100	(2) IV log(visit rate) \times 100
PM2.5 ($\mu g/m^3$)	-0.01*** (0.00)	-0.50*** (0.02)
Temperature ($^{\circ}C$)	0.16*** (0.00)	0.25*** (0.01)
Precipitation (mm)	-2.40*** (0.03)	-1.98*** (0.03)
Wind speed (m/s)	-0.50*** (0.01)	-0.82*** (0.02)
First-stage F stat		123.7
Dependent Variable Mean	6.99	6.99
Fixed Effects	Yes	Yes
R ²	0.89	0.88
Observations	4,507,400	4,507,400

Notes: This table reports the OLS and IV estimates using equation (1) and (2). The dependent variable is the log of visit rates at all POIs. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. All regressions control for temperature, precipitation, and wind speed; one lead and one lag of these weather controls. OLS (IV) estimates also include one lag and one lead of PM2.5 (instruments). Dependent variable mean is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, day-of-week, and year-by-month FE. Standard errors are clustered at the county level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

I further investigate the possibility of more dynamic behavioral responses to air pollution by considering whether the decrease in visit rates is offset by an increase on subsequent days or worsened by a further decrease. For example, people might adjust the timing of their activities rather than reducing them altogether, or they might exhibit inertia, staying home on subsequent days even after pollution levels decline. To evaluate this temporal impact, I include lagged PM2.5 values in the main specification and estimate the effects on successive days.

I find that that same-day PM2.5 levels have the largest impact, slightly lower than the impact captured by the main specification. The effect of a 1-day lag is about 20% of the contemporaneous effect, and the coefficient becomes insignificant for a 2-day lag (Table 3). This suggests that instead of compensating for lost activities on days with high pollution, people tend to stay home even as pollution levels decline, which further magnifies the negative impacts of pollution on daily activities.

The observed negative relationship between air pollution and visit rates indicates

that individuals actively adjust their behavior in response to daily fluctuations in air pollution. Specifically, when air quality deteriorates, people tend to reduce their daily activities. This avoidance behavior decreases their exposure to air pollution, thereby mitigating its negative impacts. As a result, previous studies that did not consider this avoidance behavior may have underestimated the costs of air pollution. Given that the average visit rate to leisure facilities is 6.99% in my sample, this translates into a 0.035% decrease in visit rates per day per county, or a 117,642¹¹ decrease in the number of visits per day across the United States due to a 1 $\mu g/m^3$ increase in PM2.5.

Table 3. Effect of lagged PM2.5

	log(visit rate) \times 100		
	(1)	(2)	(3)
PM2.5, contemporaneous($\mu g/m^3$)	-0.50*** (0.02)	-0.35*** (0.02)	-0.31*** (0.02)
PM2.5, 1 day lag ($\mu g/m^3$)		-0.08*** (0.01)	-0.05*** (0.01)
PM2.5, 2 day lag ($\mu g/m^3$)			0.01 (0.01)
First-stage F stat	123.7	134.0	132.1
Dependent Variable Mean	6.99	6.99	6.99
Fixed Effects	Yes	Yes	Yes
R ²	0.8	0.89	0.89
Observation	4,507,400	4,504,300	4,501,200

Notes: This table reports the IV estimates using equation (1) and equation (2) when including lagged PM2.5. The dependent variable is the log of visit rates at all POIs. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Fixed effects include county-by-year, county-by-month, day-of-week, and year-by-month FE. Dependent variable mean is the average visit rate in percentage terms. Standard errors clustered at the county level are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

4.2 Heterogeneity

Heterogeneity across Industries The effect estimated in the main results might conceal variations across different industries. To provide a comprehensive view, I use the extensive coverage of the SafeGraph dataset to examine the impact of air pollution on various industries (defined by 2-digit NAICS codes). Using the same IV

¹¹The World Bank reported a total population of 331.9 million in the United States in 2021.

model as in the main analysis, I separately estimate the effects for each industry. The raw number of visits and percentage of each industry in my sample can be found in Table A1.

As shown in Figure 1, activities in most industries are negatively affected by air pollution. Among them, recreation activities (NAICS code 71) experience the largest decline in visits. This is likely because recreational activities are more flexible compared to work commitments. Therefore, when pollution levels are high, people can easily cancel their recreational plans and stay at home. The second most affected industry is educational services, which involves primary schools. This suggests that vulnerable groups, such as young children, may be more responsive to air pollution.

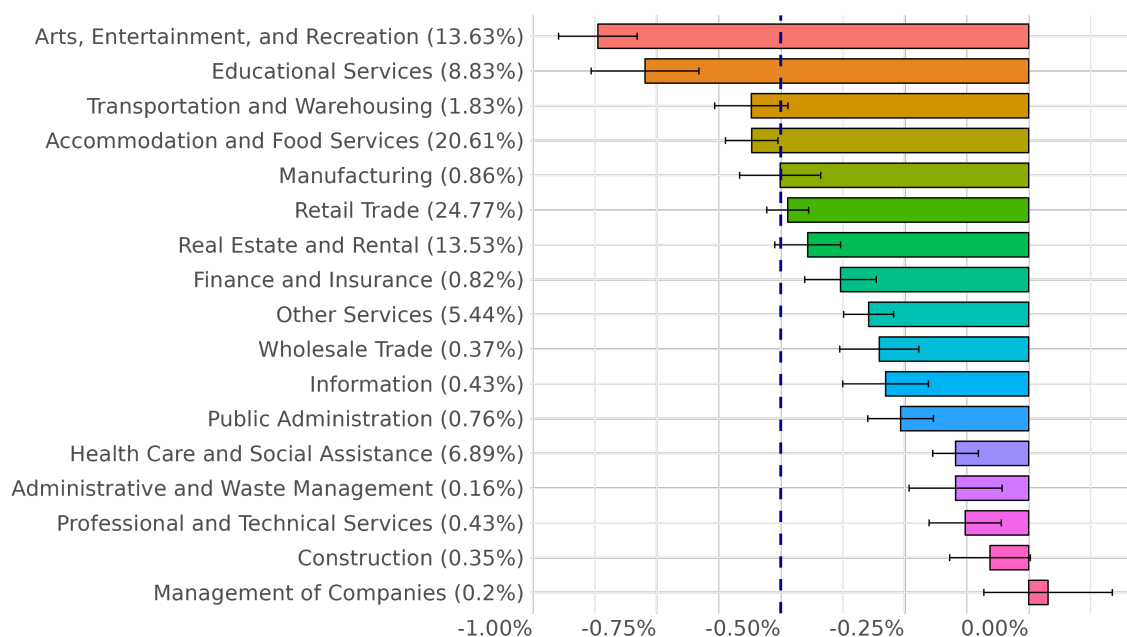


Figure 1. Heterogeneity by 2-digit NAICS Codes

Notes: This figure displays the heterogeneous treatment effects of air pollution on daily activities across various industries, categorized by their 2-digit NAICS codes. The percentage following each industry name indicates the share of raw visits that the industry represents in the sample. Points represent the estimates, and horizontal lines represent the 95% confidence intervals. The vertical blue dashed line represents the average effect in the main results for all POIs. Industries with raw visits accounting for less than 0.1% are omitted from the figure due to large variance.

In addition, for the most affected industry—Arts, Entertainment, and Recreation—I separately estimate the effects for each facility type (defined by 6-digit

NAICS codes). The percentage of each facility type in my sample can be found in A2. As shown in Figure 2, air pollution reduces activities in all of the outdoor facilities and most of the indoor facilities¹², with the exception of a slight increase in bowling centers. Moreover, the reduction in outdoor facilities is generally greater than in indoor facilities. This is likely because engaging in outdoor activities intensifies the negative health effects of air pollution due to increased respiration and exposure. Lastly, golf courses experience the largest decline in visits. This might stem from golf’s popularity among the urban rich, further suggesting that higher-income groups could be more responsive to air pollution.

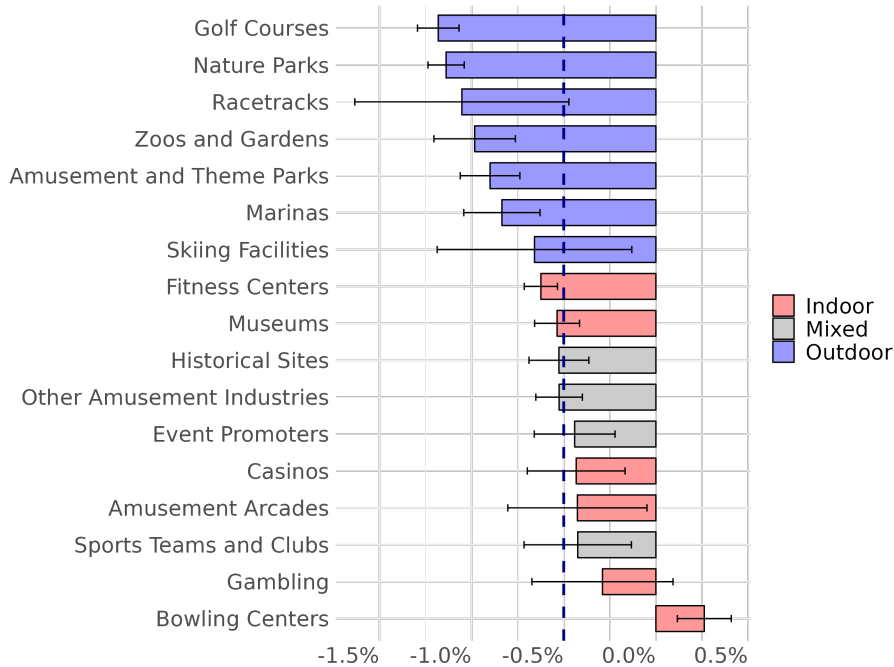


Figure 2. Heterogeneity in Recreational Facilities

Notes: The figure displays the heterogeneous treatment effect of air pollution on visit rates at recreational facilities, including outdoor ones (in blue, such as golf courses, nature parks, and racetracks) and indoor ones (in red, such as fitness centers, casinos, and bowling centers). Points represent the estimates, and horizontal lines represent the 95% confidence intervals. Establishments with raw visits less than 0.1% are omitted from the figure due to large variance.

I also estimate the effect for the Health Care and Social Assistance sector. As shown in Figure 3, the effect of air pollution on all health industries is smaller, approx-

¹²He et al. (2022) suggest that the negative effect on indoor facilities is mainly due to pollution exposure during transportation to the destination.

imately one-fifth of the reduction observed for recreation facilities. This is contrary to my initial assumption that hospital visits would increase with worsening pollution. This could be because there is no separate category specifically for respiratory-related visits; thus, any increase in visits to respiratory or related emergency centers may be offset by a decrease in visits for other types of services, such as social assistance and child day care. The increase in visits by vulnerable groups might be "obscured" by an overall decline in visits for unrelated healthcare services. Similarly, the increase in visits by vulnerable groups might be obscured by an overall decline in visits for unrelated healthcare services. To further test the effect in vulnerable groups, I then limit my sample to counties with an elderly population larger than the third quartile. As shown in Figure ??, the results indicate that air pollution does increase visits to outpatient care centers and other individual and family services.

I also estimate the effect of air pollution on the Health Care and Social Assistance sector. As shown in Figure 3, the impact on all health-related industries is smaller, at roughly one-fifth of the reduction observed for recreation facilities. This finding contradicts my initial assumption that hospital visits would increase with worsening pollution. One possible reason for this result is the absence of a separate NAICS category specifically for respiratory-related visits; therefore, any increase in respiratory or related emergency visits may be masked by a decrease in visits for other types of medical services. Likewise, the increase in visits by vulnerable groups could be obscured by an overall decline in visits for unrelated healthcare services. To further examine the effect on vulnerable groups, I limit my sample to counties with an elderly population above the third quartile. As shown in Figure 4, after this adjustment, the estimates generally shift from negative to less negative or positive. Additionally, there is a significant increase in visits to outpatient care centers and other individual and family services.

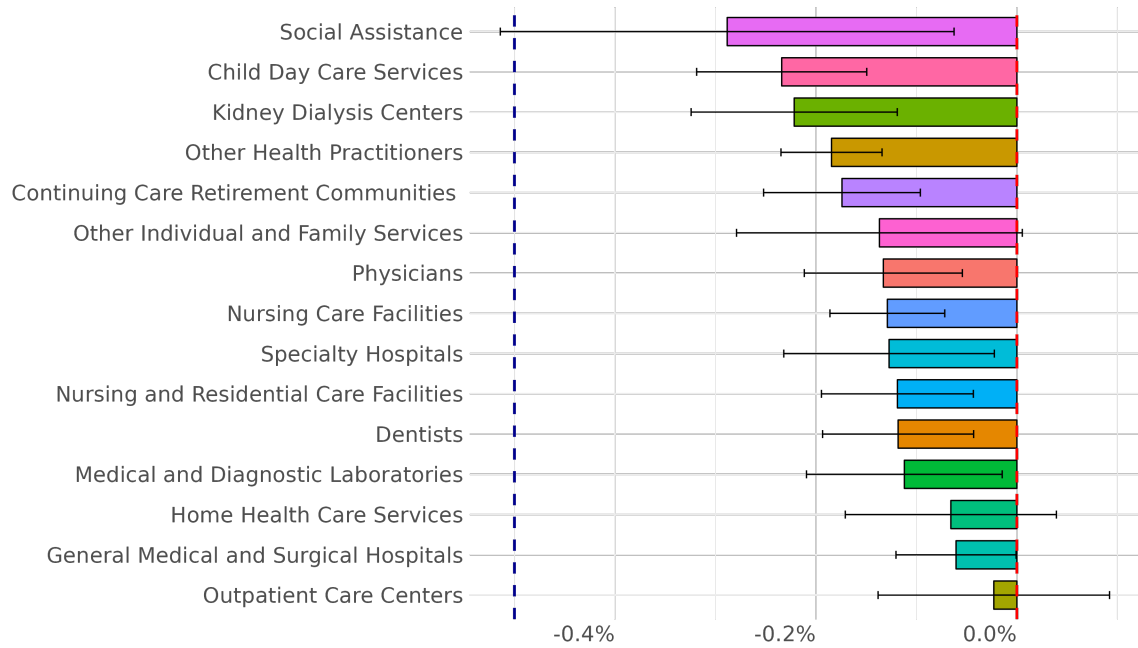


Figure 3. Heterogeneity in Health Care Facilities

Notes: The figure displays the heterogeneous treatment effect of air pollution on visit rates at health facilities. Points represent the estimates, and horizontal lines represent the 95% confidence intervals. The vertical blue dashed line represents the estimates for overall visits, and the red vertical line indicates zero. Establishments with raw visits less than 0.1% are omitted from the figure due to large variance.

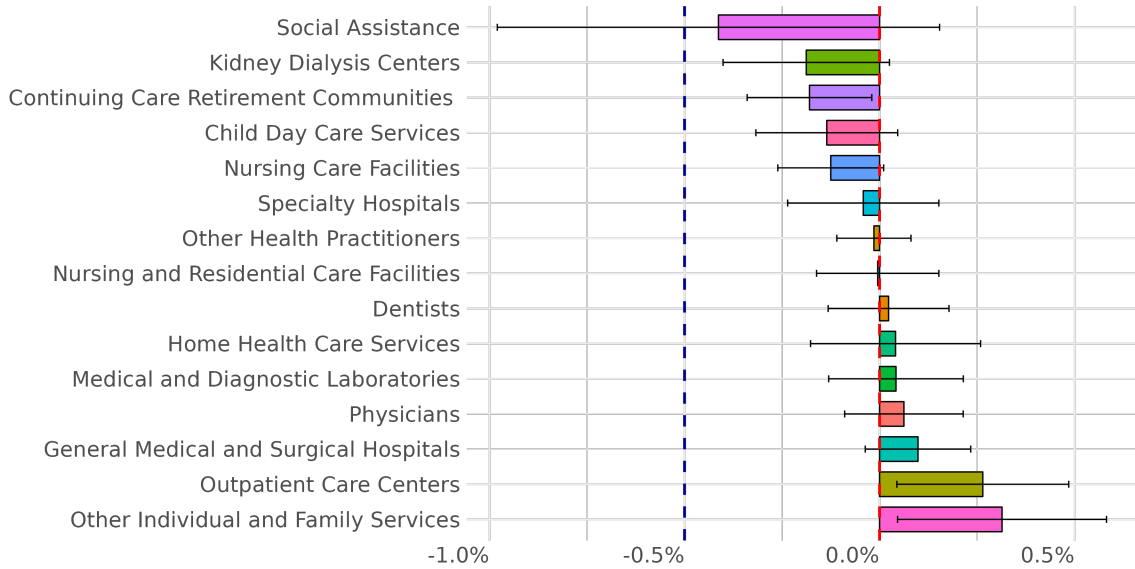


Figure 4. Heterogeneity in Health Care Facilities for Counties with Elderly Population Above the Third Quartile

Notes: This figure shows the heterogeneous treatment effect of air pollution on visit rates at health care facilities in counties where the elderly population is above the third quartile. The points represent the estimated effects, while the horizontal lines show the 95% confidence intervals. The vertical blue dashed line represents the estimates for overall visits, and the red vertical line indicates zero. Facilities with raw visit rates below 0.1% are excluded due to high variance.

Heterogeneity across Demographic Groups A growing literature shows that exposure to air pollution and other environmental risks is unequally distributed across different groups of individuals (Mohai et al., 2009; Hsiang et al., 2019). To examine whether the effects of air pollution differ across income groups, I categorize counties into two income groups: low income (below the national median) and high income (above the national median), and include their interaction with PM2.5 levels¹³. As shown in Panel A of Table 4, the estimated coefficient for the interaction is negative and statistically significant at the 99% level. This indicates that high-income counties have greater sensitivity to air pollution in their activities¹⁴. Furthermore, if I focus

¹³Since PM2.5 is endogenous, the interaction of PM2.5 and the income group dummy is instrumented using wind directions.

¹⁴It is important to note that higher-income counties have a slightly higher average visit rate initially: 7.08% versus 6.84% in lower-income counties. The decrease is proportional to the mean

on counties with income higher than the third quartile, the magnitude of the estimate becomes larger. This result suggests that as income increases, the avoidance response to air pollution also increases. One possible interpretation is that individuals in high-income counties have a better awareness of air pollution.

Children and infants are among the most susceptible to air pollution ([Aragón et al., 2017](#); [Jayachandran, 2009](#)) because lung development continues throughout adolescence, making developing lungs particularly at risk from exposure to toxins ([Dietert et al., 2000](#)). I investigate the potential influence of a county’s age composition on pollution avoidance behavior by using the proportion of children below 5 years old. Similar to the income categorization, counties are grouped based on their percentage of children below 5: fewer children (below median) and more children (above median). Their interaction with PM2.5 levels is included in the regression. In Panel B of Table 4, the estimated coefficient for the interaction term is negative and statistically significant, indicating that vulnerable groups, such as young children, are more responsive to air pollution.

Lastly, I investigate whether different racial groups respond differently to air pollution. Similar to the income and age categorization, counties are grouped based on their percentage of Black population: fewer Black people (below median) and more Black people (above median). Their interaction with PM2.5 levels is included in the regression. As shown in Panel C of Table 4, counties with a higher percentage of Black people (above the third quartile) respond less to air pollution. Literature shows that air pollution has larger health effects on Black people than on White people ([Currie and Walker, 2011](#); [Chay and Greenstone, 2003](#)). This disparity may be partly due to the different magnitude of avoidance behavior, where Black people respond less to air pollution, leading to greater exposure and consequently larger negative health impacts.

visit rates.

Table 4. Heterogeneous Effect of PM2.5 across Different Groups

	(1) log(visit rate) \times 100	(2) log(visit rate) \times 100
Panel A: Different Income Group		
PM2.5 ($\mu g/m^3$)	-0.16*** (0.03)	-0.28*** (0.02)
PM2.5 \times 1{Inc. > Median}	-0.61*** (0.04)	
PM2.5 \times 1{Inc. > 3rd Quartile}		-0.71*** (0.07)
Panel B: Different Age Group		
PM2.5 ($\mu g/m^3$)	-0.37*** (0.04)	-0.40*** (0.03)
PM2.5 \times 1{Pct. Children > Median}	-0.34*** (0.10)	
PM2.5 \times 1{Pct. Children > 3rd Quartile}		-0.60*** (0.13)
Panel C: Different Race Group		
PM2.5 ($\mu g/m^3$)	-0.56*** (0.04)	-0.55*** (0.03)
PM2.5 \times 1{Pct. Black > Median}	0.10 (0.04)	
PM2.5 \times 1{Pct. Black > 3rd Quartile}		0.18*** (0.04)
First-stage F stat	113.5	113.5
Dependent Variable Mean	6.99	6.99
Fixed Effects	Yes	Yes
R ²	0.88	0.88
Observations	4,505,946	4,505,946

Notes: This table reports the effect of daily PM2.5 on daily activities for different income groups using equation (1) and equation (2). The dependent variable is the log of the visit rates in county c at date t . All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. The dummy variable $\mathbf{1}\{\text{Income} > \text{Median}\} = 1$ ($\mathbf{1}\{\text{Income} > \text{3rd Quartile}\} = 1$) if personal income in county c is higher than the median (third Quartile). Similarly, $\mathbf{1}\{\text{Pct. Children} > \text{Median}\}$ and $\mathbf{1}\{\text{Pct. Children} > \text{3rd Quartile}\}$ are dummies for the percentage of children, and $\mathbf{1}\{\text{Pct. Black} > \text{Median}\}$ and $\mathbf{1}\{\text{Pct. Black} > \text{3rd Quartile}\}$ are dummies for the percentage of the Black population. The dependent variable mean is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, day-of-week, and month-by-year FE. Standard errors are clustered at the county level.

4.3 Mechanisms

In this section, I explore two potential mechanisms through which air pollution reduce daily activities. First, individuals may respond to Air Quality Index (AQI) warnings. A second potential mechanism is that individuals spend more time at home on days with poor visibility.

AQI Advisories AQI is an index that spans from 0 to 500, created by the EPA to tell the public how polluted the air is. Real-time AQI information is disseminated to the public through various channels, such as website portals (www.airnow.gov) and mobile applications. An AQI value of 100 generally corresponds to the national air quality standard for the pollutant, which is the level EPA has set to protect public health¹⁵. Table A4 displays the behavioral guidelines and PM2.5 concentrations associated with each category.

When PM2.5 concentrations exceed $35.5 \mu\text{g}/\text{m}^3$, air quality is considered unhealthy (code orange), and information about unhealthy air quality appears on most weather applications or websites (see Figure A2 for an example). Therefore, I use a regression discontinuity (RD) Design to estimate the causal effect of AQI advisories: the real-time PM2.5 levels serve as the running variable, and the data are examined on either side of the $35.5 \mu\text{g}/\text{m}^3$ cutoff. One main assumption is that there is no manipulation at the cutoff, which is reasonable since PM2.5 values are automatically recorded by monitors. Figure A4 further supports the validity of this assumption. Additionally, since AQI advisories are based on data from EPA’s outdoor monitors, I switch to monitor-based PM2.5 data for the following estimation¹⁶.

Table 5 summarizes the results across different bandwidth and kernel selections. I find that there is a negative effect of air quality advisories on visit rates, but this effect is imprecisely estimated and not significant. Similarly, in Figure 5, there does not appear to be a significant discontinuity at the cutoff point. One potential explanation for this result is the small number of observations around the threshold; in my sample, only 0.34% of the data points are above the $35.5 \mu\text{g}/\text{m}^3$ (equivalent to AQI 100) threshold. Another explanation is that individuals respond more only at higher alert thresholds. For example, Neidell (2009) and Zivin and Neidell (2009) show that smog alerts (issued when ozone is forecasted above 0.2 ppm, equivalent to AQI 300)

¹⁵source: <https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information>. Accessed March 30, 2023

¹⁶Since some counties do not have monitors, there are 601,894 observations from January 1, 2018 to December 30, 2021.

in California prompt avoidance behavior. However, there is even less data around higher alert thresholds. Lastly, although AQI categories and health concerns are defined uniformly across the US, different states have different criteria for issuing air quality alerts or advisories, which could make the estimate less precise.

Table 5. Estimation Results from RD of AQI Advisories

	(1)	(2)	(3)	(4)	(5)	(6)
AQI Advisories	-4.3 (3.9)	-4.9 (2.7)	-3.0 (1.9)	-4.6 (3.8)	-5.1 (2.6)	-2.7 (1.8)
Kernel	Triangular	Triangular	Triangular	Uniform	Uniform	Uniform
Bandwidth	5	10	20	5	10	20
Dependent Variable Mean	7.60	7.60	7.60	7.60	7.60	7.60
Observations	695,494	695,494	695,494	695,494	695,494	695,494
Effective Observations	2,990	7,500	45,135	2,990	7,500	45,135

Notes: This table presents RD estimates with PM2.5 as the running variable and a cutoff at $35.5 \mu\text{g}/\text{m}^3$ (equivalent to AQI 100). The dependent variable is the log of visit rates to all POIs. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. The dependent variable mean is the average visit rate in percentage terms.

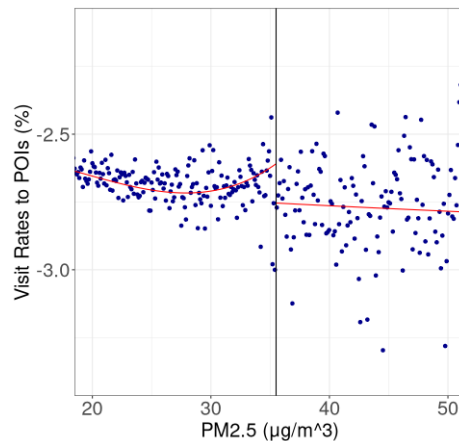


Figure 5. Visit Rates by PM2.5 levels

Visibility The same pollutants that contribute to PM2.5 can also reduce visibility: pollution particles affect visibility by altering the way light is absorbed and scattered in the atmosphere, reducing the clarity and color of what we see¹⁷. When visibility is low, people might perceive that the air quality is poor and choose to spend more time at home instead of going out. To test this channel, I use satellited-based visibility

¹⁷Source: https://www.epa.gov/sites/default/files/2015-05/documents/haze_brochure_20060426.pdf. Accessed June 10, 2024.

data from NCEP North American Regional Reanalysis database (NARR)¹⁸ and apply a 2SLS-style analysis. In the 2SLS-style analysis, I use PM2.5 level as an instrument of visibility and then regress visit rates on the predicted visibility at the second stage. Specifically, I use the following two stage specification:

$$Visibility_{ct} = PM2.5_{ct} + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \quad (3)$$

$$\log\left(\frac{Y_{ct} + 1}{Pop_c}\right) = \alpha \times Visibility_{ct} + \mathbf{X}'_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct}. \quad (4)$$

The results are reported in Table 6. Column 1 shows the positive and significant correlation in the first stage. The F-stage is far larger than 10, indicating that there is no weak instrument problem. Columns 2 report the results of the second stage: a 1 km decrease in visibility leads to 5.37% decrease in visit rates. These results are consistent with Keiser et al. (2018), who find that visitors decrease visitation to national parks on days with poor visibility.

Table 6. Effect of Visibility on Visit Rates

	(1) First Stage Visibility	(2) Second Stage log(visit rate) $\times 100$
Visibility (<i>km</i>)		5.15*** (0.81)
PM2.5 ($\mu g/m^3$)	-0.24*** (0.03)	
First-stage F stat	675.3	
Dependent Variable Mean	17.52	6.99
Fixed Effects	Yes	Yes
R ²	0.48	0.84
Observations	4,507,400	4,507,400

Notes: This table reports the results of effect of visibility on visit rates using equation (3) and (4). All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. All regressions control for temperature, precipitation, and wind speed, including one lead and one lag of these weather controls. The dependent variable mean in column (1) is the average visibility, and the dependent variable mean in column (2) is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, day-of-week, and year-by-month fixed effects. Standard errors are clustered at the county level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

¹⁸See <https://psl.noaa.gov/data/gridded/data.narr.html>. Accessed July 18, 2024.

4.4 Welfare Analysis

In this section, I put the results into context by generating some back-of-the-envelope estimates of the cost of the lost leisure. In the previous section, I show that a $1 \mu\text{g}/\text{m}^3$ increase in PM2.5 decreases visit rates to leisure facilities by 0.50% on average. In this study, the average visit rate is 6.95%. This translates into a 0.035% decrease in visit rates per day per county, or a decline of 115,335 visits¹⁹ per day across the United States due to a $1 \mu\text{g}/\text{m}^3$ increase in PM2.5.

To understand the economic implications, I source my valuation data from the Recreational Use Values Database (Rosenberger, 2016). The values estimated in this dataset represent measures of net willingness-to-pay or consumer surplus derived from various recreation activities (e.g., sightseeing, hiking)²⁰. Based on Recreational Use Values Database, the recreational use value per person per day is \$77 (in 2016 USD), which equates to \$93.89 in 2022 USD. Therefore, the estimated annual cost in the United States of a $1 \mu\text{g}/\text{m}^3$ increase in PM2.5 on leisure visitation is approximately 3.95 billion²¹ in 2022 USD.

The estimated welfare impact of air pollution on daily activities is relatively minor compared to other air pollution cost estimates. For example, the World Bank estimated that the welfare cost of air pollution in the United States was around \$886.5 billion in 2016 (\$1,080.96 billion in 2022 USD). However, these two estimates are not directly comparable. The World Bank’s estimate assesses the aggregated cost of air pollution on human health and the environment, while the decrease in daily activities represents only a tiny fraction of this overall cost. As people become more educated about the impacts of pollution, the costs associated with the reduction in daily activities are expected to increase. However, the cost of pollution related to mortality and morbidity might decrease due to more widespread adoption of avoidance behaviors.

5 Robustness Checks

In this section, I first test the validity of the IV. IV estimates can be interpreted as the local average treatment effect (LATE) when the monotonicity assumption holds

¹⁹The World Bank reported a total population of 331.9 million in the United States in 2021.

²⁰The database offers 21 primary activity types. However, it is predominantly focused on recreation activities. However, the vast majority of POIs in my sample are non-work related POIs such as restaurants, retail stores, entertainment and recreation (Table A1), I use these estimates as a rough proxy for welfare valuation.

²¹ $115,335 \text{ visits reduction per day} \times 93.89 \text{ welfare loss per day} \times 365 \text{ days}$

(Angrist and Imbens,1995). In this paper, this assumption will be satisfied if every county within a geographic cluster group experienced a change in pollution in the same direction when the wind blows from a 90-degree direction bin, and will be violated if some counties experience changes in different directions with others counties within the same cluster group. One way Deryugina et al. (2019) assess the validity of this assumption is by varying the number of geographic clusters and the sizes of the wind direction bins. I follow a similar approach by changing the number of geographical clusters from 20 to 10 and 30 and reducing the size of wind angle bins from 90 degrees to 60 degrees. In Table 7, Column 1 is the original specification, Column 2 and 3 change the number of geographical clusters, and Column 4, 5, and 6 decrease the size of wind angle bins. In all cases, the IV estimates are similar to the main specification, supporting the robustness of the main result to different instrument choices (Table 7). In addition, I address potential spatial autocorrelation by clustering the standard errors at the county-day level. As shown in Table ??, the estimates remain significant.

Table 7. Robustness of IV Estimates to Instrument Choices

	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 ($\mu g/m^3$)	-0.50*** (0.02)	-0.56*** (0.02)	-0.48*** (0.02)	-0.54*** (0.02)	-0.49*** (0.02)	-0.46*** (0.02)
Number of geographic clusters	20	10	30	10	20	30
Size of wind angle bins (degrees)	90	90	90	60	60	60
R ²	0.88	0.88	0.88	0.88	0.89	0.88
Observations	4,507,400	4,507,400	4,507,400	4,507,400	4,507,400	4,507,400
F-statistic	123.7	207.6	83.4	151.5	86.1	62.0
Dependent Variable Mean	6.99	6.99	6.99	6.99	6.99	6.99

Notes: This table reports the IV estimates using equation (1) and equation (2) when varying the instrument choices. The baseline model (shown in column (1)) aggregates location into 20 clusters and wind direction into 90-degree intervals. The dependent variable is the log of visit rates at all POIs. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Standard errors clustered at the county level are reported in parentheses. Dependent variable mean is the average visit rate in percentage terms.

Another underlying assumption of this IV approach is that the variation comes primarily from the pollution that is transported by wind rather than generated locally. If this underlying assumption holds, then the first stage should be generally weak on days with low wind speeds and vice versa. To further examine the validity of this IV approach, I calculate the first-stage F-statistics separately by quintiles of daily wind speed. As shown in Figure 6, the strength of the first stage increases as wind speed increases. This implies the pollution variation is mainly due to non-local transport

by wind, which assesses the validity of my approach.

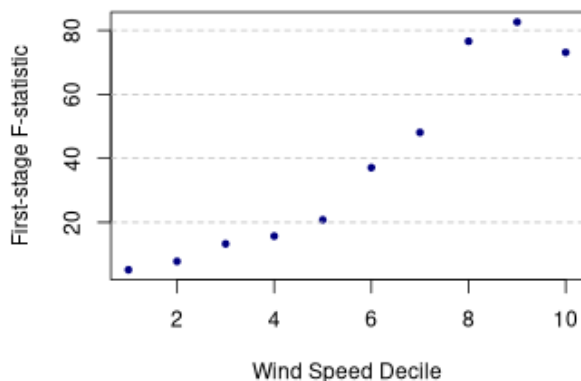


Figure 6. Relationship Between the First-stage F Statistics and Wind Speed

Notes: This figure displays the First-stage F Statistics for five subsamples that each include days that fall within a particular wind speed quintile. The first-stage F-statistics are generally smaller on days with low wind speeds and bigger on days with high wind speeds.

In addition, Table 8 indicates that the main specification is robust to variations in the number of instrument lags included. This demonstrates that the main estimates are not driven by lagged effects from PM2.5 on previous days, and therefore, can be properly interpreted as the impact of a one-unit increase in daily PM2.5 levels.

Table 8. Robustness of IV Estimates to Including Different Instrument Lags

	(1)	(2)	(3)	(4)	(5)	(6)
	1 lead and 1 lag	1 lag	2 lags	3 lags	4 lags	5 lags
PM2.5 ($\mu\text{g}/\text{m}^3$)	-0.50*** (0.02)	-0.44*** (0.02)	-0.44*** (0.0002)	-0.43*** (0.02)	-0.45*** (0.02)	-0.45*** (0.02)
R ²	0.89	0.89	0.89	0.89	0.89	0.89
Observations	4,507,400	4,507,400	4,504,300	4,501,200	4,498,100	4,495,000
Dependent Variable Mean	6.99	6.99	6.99	6.99	6.99	6.99

Notes: This table reports the IV estimates using equation (1) and equation (2) when including different number of instrument lags. The baseline model (shown in column (1)) does not include any lags. The dependent variable is the log of visit rates at all POIs. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Standard errors clustered at the county level are reported in parentheses. Dependent variable mean is the average visit rate in percentage terms.

Then, I check the robustness of the model specification. As a first test, I estimate the regressions with different sets of fixed effects and weather controls. As shown in Table 9, the main result is robust to including different combinations of fixed

effects and weather control, which implies the estimate in the main specification is not driven by seasonal or regional patterns. Additionally, I estimate the regressions with the dependent variable in the form of inverse hyperbolic sine of visit levels rather than the log of visit rates. The estimates in Table 10 are comparable to the main specification, which implies the main result is insensitive to the form of the outcome variable.

Table 9. Robustness of IV Estimates to Including Different Forms of Weather Controls and Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 ($\mu g/m^3$)	-0.50*** (0.02)	-0.60*** (0.04)	-0.53*** (0.08)	-0.51*** (0.10)	-0.43*** (0.08)	-0.63*** (0.03)
Form of weather controls	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
Day-of-week FE	✓	✓	✓	✓	✓	
County-by-year FE	✓	✓				✓
Year-by-month FE	✓	✓		✓		
County-by-month FE	✓					✓
State-by-year FE			✓	✓	✓	
State-by-Month FE			✓		✓	
R ²	0.88116	0.84591	0.30827	0.34691	0.31506	0.78988
Observations	4,507,400	4,507,400	4,507,400	4,507,400	4,507,400	4,507,400
F-statistic	123.7	83.2	96.0	64.2	85.6	101.1
Dependent Variable Mean	6.99	6.99	6.99	6.99	6.99	6.99

Notes: This table reports the IV estimates using equation (1) and equation (2) when including different combinations of fixed effects and weather controls. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Dependent variable mean is the average visit rate in percentage terms. Standard errors clustered at the county level are reported in parentheses.

Table 10. Robustness to Including Different Forms of Outcome

	(1)	(2)
	$\log(\text{visit rate}) \times 100$	$\text{IHS}(\text{visits}) \times 100$
PM2.5 ($\mu\text{g}/\text{m}^3$)	-0.50*** (0.02)	-0.51*** (0.02)
Temperature ($^{\circ}\text{C}$)	0.25*** (0.00)	0.25*** (0.00)
Precipitation (mm)	-1.95*** (0.03)	-1.96*** (0.03)
Wind speed (m/s)	-0.81*** (0.02)	-0.82*** (0.02)
First-stage F stat	123.7	123.7
Dependent Variable Mean	6.99	6.99
Fixed Effects	Yes	Yes
R ²	0.88	0.99
Observations	4,507,400	4,507,400

Notes: This table reports the OLS and IV estimates using equation (1) and equation (2) when the dependent variable is the inverse hyperbolic sine transformation (IHS) of visits. In the main specification, the dependent variable is the log of visit rates at all POIs. Standard errors clustered at the county level are reported in parentheses. Dependent variable mean is the average visit rate in percentage terms.

For a final robustness check, I estimate the regression using different sub-samples. The sample period ranges from January 1, 2018, to December 31, 2021, which includes the COVID-19 pandemic that dramatically affects individuals' mobility patterns. To ensure that the results are not influenced by public health guidance on leisure activities, I estimate the effect separately before and after the break out of the COVID-19 pandemic²². As shown in Table A5, the estimates are very similar, indicating the negative impact is not driven by pandemic-related restrictions. In addition, A6 shows the robustness of the main specification when excluding counties without satellite data.

6 Conclusion

This paper presents a large-scale analysis of the impacts of PM2.5 on daily activities across the United States. I use an instrumental variable approach to address the endogeneity of air pollution, and find a significant negative effect of PM2.5 on visit rates. My preferred model implies that a 1 $\mu\text{g}/\text{m}^3$ increase in PM2.5 leads to a 0.50 %

²²The cutoff date for dividing the sample is March 11, 2020, which is when the World Health Organization (WHO) declared COVID-19 a pandemic.

decrease in visit rates on average, which is significant across different income groups and location types. This translates a reduction of 117,642 visits or an annual economic cost of over 200 million nationwide. Overall, the results in this paper indicate the presence of behavioral adjustments in response to air pollution fluctuations, which underscore the importance of characterizing avoidance behavior when analyzing the impacts of air pollution.

This study is not without limitations. First, while I focus on the effects of air pollution on daily leisure activities, I am unable to study the potential long-term adjustments individuals make. Second, even though the phone location-based visitation data is shown to be representative of the population at the county level and above, aggregating data to the county level ignores individual variation and causes information loss. For example, county-level data does not account for individual-specific variables, such as age, race, education level, and health condition. Third, the reduced-form approach can only identify the combined effect of air pollution on leisure activities, which makes it hard to disentangle the pure avoidance effect. For example, people who fall ill on a polluted day may not engage in leisure activities.

Despite these limitations, this paper makes several contributions. First, it provides a large-scale estimation of the causal effect of air pollution on daily activities, which is more representative than previous studies. Second, this paper has important policy implications. It provides evidence that people take avoidance behavior and reduce daily activities on polluted days, which suggests studies ignoring avoidance behavior when estimating the cost of air pollution may suffer from bias. In addition, avoidance behavior itself is costly. When individuals choose to stay home to avoid pollution, they forgo the daily activities that they could potentially have enjoyed on a cleaner day. Consequently, the lost activities caused by air pollution should also be considered as part of the pollution cost.

References

- Alexander, Diane and Hannes Schwandt**, “The impact of car pollution on infant and child health: Evidence from emissions cheating,” *The Review of Economic Studies*, 2022, 89 (6), 2872–2910.
- **and Janet Currie**, “Is it who you are or where you live? Residential segregation and racial gaps in childhood asthma,” *Journal of health economics*, 2017, 55, 186–200.
- Altindag, Duha T, Deokrye Baek, and Naci Mocan**, “Chinese yellow dust and Korean infant health,” *Social Science & Medicine*, 2017, 186, 78–86.
- Angrist, Joshua and Guido Imbens**, “Identification and estimation of local average treatment effects,” 1995.
- Aragón, Fernando M., Juan Jose Miranda, and Paulina Oliva**, “Particulate matter and labor supply: The role of caregiving and non-linearities,” *Journal of Environmental Economics and Management*, 2017, 86, 295–309.
- Borgschulte, Mark, David Molitor, and Eric Zou**, “Air pollution and the labor market: Evidence from wildfire smoke,” *Rev Econ Stat*, 2018.
- Bresnahan, Brian W, Mark Dickie, and Shelby Gerking**, “Averting behavior and urban air pollution,” *Land Economics*, 1997, pp. 340–357.
- Chang, Ting, Yingjie Hu, Dane Taylor, and Brian M Quigley**, “The role of alcohol outlet visits derived from mobile phone location data in enhancing domestic violence prediction at the neighborhood level,” *Health & Place*, 2022, 73, 102736.
- Chay, Kenneth Y and Michael Greenstone**, “The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession,” *The quarterly journal of economics*, 2003, 118 (3), 1121–1167.
- Currie, Janet and Matthew Neidell**, “Air pollution and infant health: what can we learn from California’s recent experience?,” *The Quarterly Journal of Economics*, 2005, 120 (3), 1003–1030.
- **and Reed Walker**, “Traffic congestion and infant health: Evidence from E-ZPass,” *American Economic Journal: Applied Economics*, 2011, 3 (1), 65–90.
- Deryugina, Tatyana, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif**, “The mortality and medical costs of air pollution: Evidence from changes in wind direction,” *American Economic Review*, 2019, 109 (12), 4178–4219.

- Dietert, Rodney R, Ruth A Etzel, David Chen, Marilyn Halonen, Steven D Holladay, Annie M Jarabek, Kenneth Landreth, David B Peden, Kent Pinkerton, Ralph J Smialowicz et al.**, “Workshop to identify critical windows of exposure for children’s health: immune and respiratory systems work group summary.” *Environmental health perspectives*, 2000, 108 (suppl 3), 483–490.
- Gellman, Jacob, Margaret Walls, and Matthew Wibbenmeyer**, “Wildfire, smoke, and outdoor recreation in the western United States,” *Forest Policy and Economics*, 2022, 134, 102619.
- Gillingham, Kenneth and Pei Huang**, “Racial disparities in the health effects from air pollution: Evidence from ports,” Technical Report, National Bureau of Economic Research 2021.
- Graginer, Corbett, Andrew Schreiber, and Wonjun Chang**, “Do regulators strategically avoid pollution hotspots when siting monitors,” *Evidence from remote sensing of air pollution. University of Madison–Wisconsin Working Paper*, 2018.
- Hanna, Rema and Paulina Oliva**, “The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City,” *Journal of Public Economics*, 2015, 122, 68–79.
- He, Guojun, Maoyong Fan, and Maigeng Zhou**, “The effect of air pollution on mortality in China: Evidence from the 2008 Beijing Olympic Games,” *Journal of Environmental Economics and Management*, 2016, 79, 18–39.
- He, Jiaxiu, Haoming Liu, and Alberto Salvo**, “Severe air pollution and labor productivity: Evidence from industrial towns in China,” *American Economic Journal: Applied Economics*, 2019, 11 (1), 173–201.
- He, Xiaobo, Zijun Luo, and Junjie Zhang**, “The impact of air pollution on movie theater admissions,” *Journal of Environmental Economics and Management*, 2022, 112, 102626.
- Heblich, Stephan, Alex Trew, and Yanos Zylberberg**, “East-side story: Historical pollution and persistent neighborhood sorting,” *Journal of Political Economy*, 2021, 129 (5), 1508–1552.
- Hsiang, Solomon, Paulina Oliva, and Reed Walker**, “The distribution of environmental damages,” *Review of Environmental Economics and Policy*, 2019.
- Jayachandran, Seema**, “Air quality and early-life mortality: Evidence from Indonesia’s wildfires,” *Journal of Human resources*, 2009, 44 (4), 916–954.
- Keiser, David, Gabriel Lade, and Ivan Rudik**, “Air pollution and visitation at U.S. national parks,” *Science advances*, 2018, 4 (7), eaat1613.

- Lee, Ki-Kwang, YoungKi Park, Sang-Pil Han, and Hyun Cheol Kim**, “The alerting effect from rising public awareness of air quality on the outdoor activities of megacity residents,” *Sustainability*, 2020, *12* (3), 820.
- Mikati, Ihab, Adam F Benson, Thomas J Luben, Jason D Sacks, and Jennifer Richmond-Bryant**, “Disparities in distribution of particulate matter emission sources by race and poverty status,” *American journal of public health*, 2018, *108* (4), 480–485.
- Mohai, Paul, David Pellow, and J Timmons Roberts**, “Environmental justice,” *Annual review of environment and resources*, 2009, *34*, 405–430.
- Neidell, Matthew**, “Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations,” *Journal of Human resources*, 2009, *44* (2), 450–478.
- Rosenberger, Randall S**, “Recreation Use Values Database–Summary,” 2016.
- Squire**, “What about bias in the SafeGraph dataset?,” 2019.
- Tessum, Christopher W, David A Paoella, Sarah E Chambliss, Joshua S Apte, Jason D Hill, and Julian D Marshall**, “PM_{2.5} polluters disproportionately and systemically affect people of color in the United States,” *Science advances*, 2021, *7* (18), eabf4491.
- Zivin, Joshua Graff and Matthew Neidell**, “Days of haze: Environmental information disclosure and intertemporal avoidance behavior,” *Journal of Environmental Economics and Management*, 2009, *58* (2), 119–128.
- Zivin, Joshua Graff and Matthew Neidell**, “The impact of pollution on worker productivity,” *American Economic Review*, 2012, *102* (7), 3652–73.
- Zou, Eric Yongchen**, “Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality,” *American Economic Review*, July 2021, *111* (7), 2101–26.

A Appendix Tables and Figures

Appendix Table A1. Percentage of Raw Visits by 2-digit NAICS Codes

NAICS Codes	Description	Raw Visit	Percentage
44-45	Retail Trade	9,146,929,016	24.77
72	Accommodation and Food Services	7,610,742,956	20.61
71	Arts, Entertainment, and Recreation	5,034,071,398	13.63
53	Real Estate and Rental and Leasing	4,996,089,773	13.53
61	Educational Services	3,261,648,196	8.83
62	Health Care and Social Assistance	2,543,964,413	6.89
81	Other Services	2,010,628,936	5.44
48-49	Transportation and Warehousing	676,689,537	1.83
31-33	Manufacturing	317,021,279	0.86
52	Finance and Insurance	303,659,378	0.82
92	Public Administration	279,409,983	0.76
51	Information	160,409,689	0.43
54	Professional, Scientific, and Technical Services	158,697,298	0.43
42	Wholesale Trade	136,210,405	0.37
23	Construction	130,696,514	0.35
55	Management of Companies and Enterprises	72,815,213	0.20
56	Administrative and Support and Waste Management	59,912,819	0.16
22	Utilities	28,890,736	0.08
11	Agriculture, Forestry, Fishing and Hunting	2,093,013	0.01
21	Mining, Quarrying, and Oil and Gas Extraction	185,256	0.00

Appendix Table A2. Percentage of Raw Visits in Recreation Industry

NAICS Codes	Description	Raw Visit	Percentage
712190	Nature Parks and Other Similar Institutions	2,610,937,377	51.97
713940	Fitness and Recreational Sports Centers	908,175,772	18.08
713910	Golf Courses and Country Clubs	525,323,866	10.46
713990	All Other Amusement and Recreation Industries	187,873,033	3.74
713110	Amusement and Theme Parks	184,452,834	3.67
711211	Sports Teams and Clubs	126,947,155	2.53
712110	Museums	93,538,708	1.86
711310	Promoters of Performing Arts and Events	91,428,501	1.82
713210	Casinos (except Casino Hotels)	87,878,401	1.75
712120	Historical Sites	65,689,209	1.31
713950	Bowling Centers	56,054,357	1.12
713920	Skiing Facilities	26,600,467	0.53
712130	Zoos and Botanical Gardens	24,056,950	0.48
713930	Marinas	17,077,159	0.34
713120	Amusement Arcades	6,400,510	0.13
713290	Other Gambling Industries	4,858,269	0.10
711212	Racetracks	4,849,757	0.10
711219	Other Spectator Sports	915,133	0.02
711130	Musical Groups and Artists	528,696	0.01
711190	Other Performing Arts Companies	364,952	0.01
711510	Independent Artists, Writers, and Performers	70,944	0.00
711110	Theater Companies and Dinner Theaters	33,414	0.00
711410	Agents and Managers for Artists	30,905	0.00

Appendix Table A3. Percentage of Raw Visits in Healthcare Industry

NAICS Codes	Description	Raw Visit	Percentage
6221	General Medical and Surgical Hospitals	602,393,987	24.73
6213	Offices of Other Health Practitioners	391,980,708	16.09
6244	Child Day Care Services	349,496,920	14.35
6211	Offices of Physicians	317,171,391	13.02
6212	Offices of Dentists	226,457,069	9.30
6231	Nursing Care Facilities (Skilled Nursing Facilities)	156,062,572	6.41
621492	Kidney Dialysis Centers	61,429,935	2.52
6223	Specialty (except Psychiatric and Substance Abuse) Hospitals	49,666,480	2.04
6233	Continuing Care Retirement Communities and Assisted Living Facilities for the Elderly	43,721,328	1.80
624190	Other Individual and Family Services	39,871,875	1.64
6214	Outpatient Care Centers	39,677,848	1.63
6216	Home Health Care Services	37,830,065	1.55
6215	Medical and Diagnostic Laboratories	36,241,509	1.49
624110	Child and Youth Services	21,330,836	0.88
6242	Community Food and Housing, and Emergency and Other Relief Services	16,862,249	0.69
6219	Other Ambulatory Health Care Services	13,277,127	0.55
621498	All Other Outpatient Care Centers	13,132,241	0.54
621493	Freestanding Ambulatory Surgical and Emergency Centers	8,481,763	0.35
624120	Services for the Elderly and Persons with Disabilities	6,366,018	0.26
6222	Psychiatric and Substance Abuse Hospitals	3,974,312	0.16
6241	Individual and Family Services	161,226	0.01
621420	Outpatient Mental Health and Substance Abuse Centers	139,374	0.01

Appendix Table A4. AQI Categories corresponding to PM2.5 Concentrations

Category	Designated Color	AQI Index	PM2.5 Concentration ($\mu g/m^3$)
Good	Green	0-50	0.0-12.0
Moderate	Yellow	51-100	12.1-35.4
Unhealthy for Sensitive Groups	Orange	101-150	35.5-55.4
Unhealthy	Red	151-200	55.5-150.4
Very Unhealthy	Purple	201-300	150.5-250.4
Hazardous	Maroon	301-500	250.5-500

Source: National Ambient Air Quality Standards for Particle Pollution ²³.

Appendix Table A5. Robustness to COVID-19 Pandemic

	(1) Before COVID log(visit rates) $\times 100$	(2) During COVID log(visit rates) $\times 100$	(3) After COVID log(visit rates) $\times 100$
PM2.5 ($\mu g/m^3$)	-0.27*** (0.02)	-0.39*** (0.06)	-0.46*** (0.03)
First-stage F stat	158.2	34.2	19.1
Dependent Variable Mean	6.51	4.93	8.76
Fixed Effects	Yes	Yes	Yes
R ²	0.90	0.93	0.89
Observations	2,467,600	49,600	855,600

Notes: This table presents IV estimates using equation (1) and equation (2) for different time periods. Standard errors are reported in parentheses and are clustered at the county level. The dependent variable mean is the average visit rate in percentage terms. Before COVID denotes observations before March 15, 2020, when many U.S. states began implementing stay-at-home orders. During COVID denotes observations from March 15, 2020, to April 2021, when the FDA issued an emergency use authorization for the COVID-19 vaccine. After COVID denotes observations after April 2021.

²³https://www.epa.gov/sites/default/files/2016-04/documents/2012_aqi_factsheet.pdf. Accessed Feb 25, 2023

Appendix Table A6. Robustness to Excluding Counties without Satellite Data

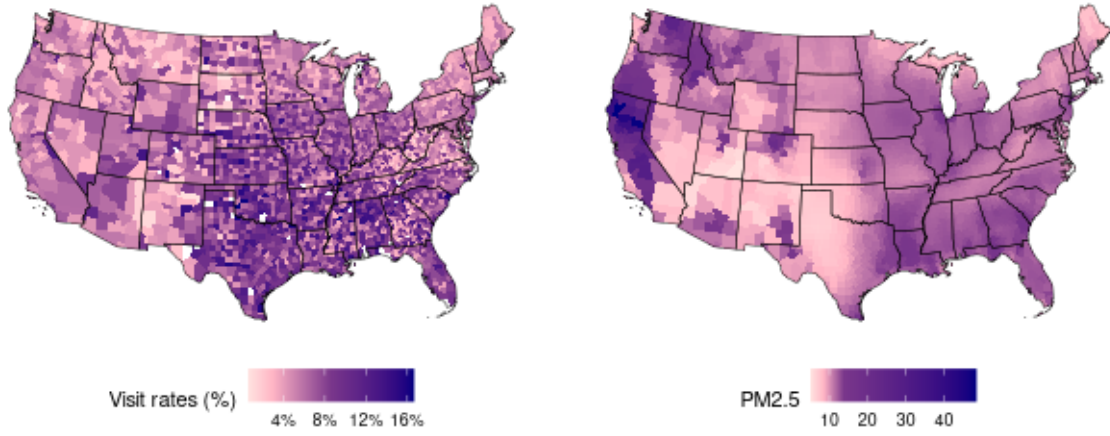
	(1) With IDW log(visit rates) \times 100	(2) Without IDW log(visit rates) \times 100
PM2.5 ($\mu g/m^3$)	-0.50*** (0.02)	-0.39*** (0.03)
Temperature ($^{\circ}C$)	0.25*** (0.00)	0.22*** (0.00)
Precipitation (mm)	-1.98*** (0.03)	-1.98*** (0.05)
Wind speed (m/s)	-0.82*** (0.02)	-0.67*** (0.03)
First-stage F stat	123.7	24.6
Dependent Variable Mean	6.99	6.73
Fixed Effects	Yes	Yes
R ²	0.88	0.88
Observations	4,507,400	1,674,454

Notes: This table reports the effect of daily PM2.5 on daily activities using only counties with satellite data. In the main specification, counties without satellite data are interpolated using IDW. Dependent variable mean is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, year-by-month and day-of-week FE. Standard errors are clustered at the county level.

Appendix Figure A1. County-level visit rates and PM2.5 concentration

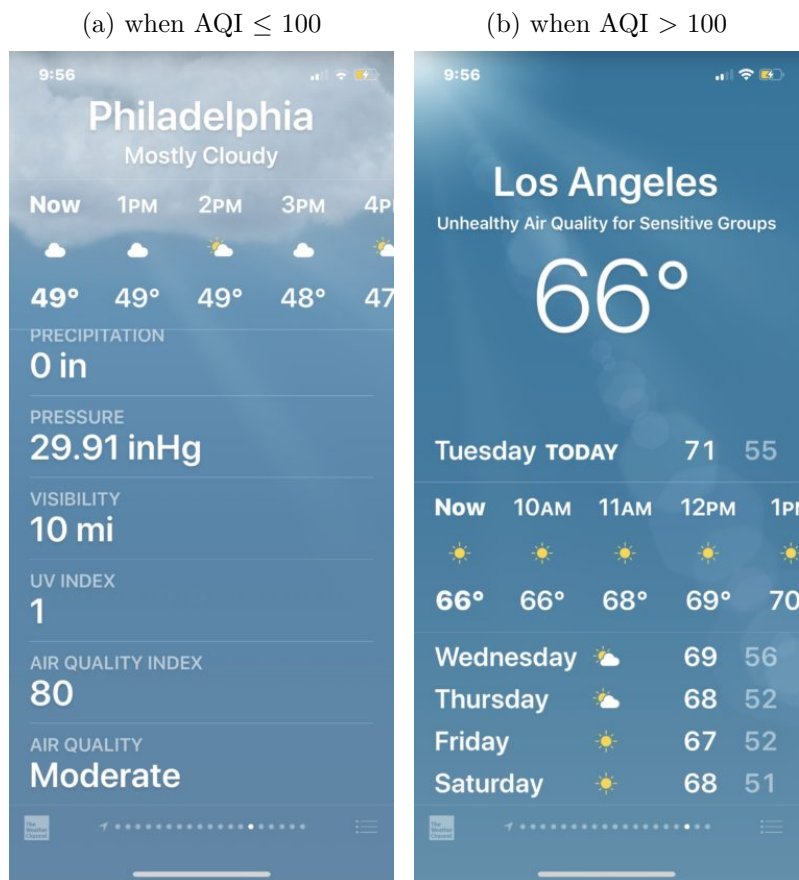
(a) County-level Visit Rates

(b) County-level PM2.5 Concentration



Notes: This figure displays average daily county means for the number of visitations (left panel) and PM2.5 concentration (right panel) from January 1, 2018, to December 30, 2021. As a few counties do not have any visitation data, there are some missing values in the figure.

Appendix Figure A2. Example: Air Quality Information on iPhone with Weather



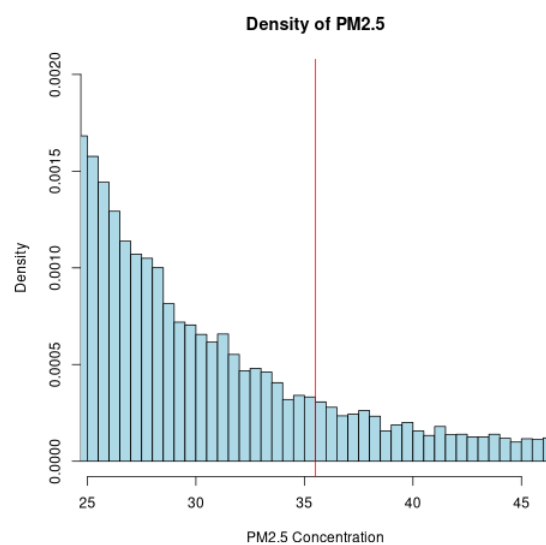
Notes: This figure displays the weather application interface when $AQI \leq 100$ (left panel) and $AQI > 100$ (right panel). Specifically, if there is no air quality concern ($AQI \leq 100$), then there is no message at the top of the Weather overview. Users have to scroll down to find the air quality information. However, if the air quality is unhealthy ($AQI > 100$), then the app prominently displays information regarding unhealthy air quality at the top of the interface.

Source: <https://osxdaily.com/2018/11/20/get-air-quality-info-iphone-weather/>



Appendix Figure A3. K-means Clustering Result

Notes: This figure displays the K-means clustering result based on latitude and longitude. As a few counties do not have any visitation data for leisure facilities from Safegraph, there are some missing values in the figure. There are 20 spatial groups in total, and each of them is represented by a different color. After clustering, γ_b^g in equation (2) can vary across geographic regions.



Appendix Figure A4. No Manipulation at the Threshold

Notes: This figure displays the density of the PM2.5 concentrations and indicates there is no discontinuity in density at the threshold. The red line is at $35.5 \mu\text{g}/\text{m}^3$, which is the level EPA set to protect public health.