Air Pollution and Economic Activity: Evidence from Foot Traffic Patterns*

Xinhui Sun†

January, 2025

Abstract

I investigate how air pollution affects economic activity. Using over 600 million phone-location-based foot traffic data points from SafeGraph, I conduct a large-scale analysis to examine the causal effect of air pollution on activity patterns across the US. Using changes in local wind direction as an instrumental variable (IV) for air pollution, I find that a 1 $\mu g/m^3$ increase in PM2.5 concentration leads to a 0.50% decrease in economic activity, resulting in a nationwide reduction of 43 million trips annually. The reductions are widespread across different economic sectors, with recreational activities experiencing the largest decline. The effect is more pronounced in higher-income counties and areas with a larger share of children, suggesting greater awareness among wealthier or more vulnerable populations.

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

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

1 Introduction

Air pollution imposes significant costs on human well-being. Its negative impacts on morbidity and mortality are well-documented (Currie and Neidell, 2005; He et al., 2016; Deschenes et al., 2017; Deryugina et al., 2019). Air pollution can also influence daily behaviors, affecting decisions about where people go and what they do. However, the extent to which air pollution affects economic activities remains largely unexplored. Prior studies typically focus on specific types of activities, such as visits to zoos, national parks, or movie theaters (Neidell, 2009; Keiser et al., 2018; He et al., 2022). Yet, air pollution may have much broader effects, affecting other important activites such as shopping and restaurant dining[1]. These widespread behavioral changes can carry significant economic implications, as reductions in economic activity due to air pollution can also lower overall economic output. Quantifying these broader effects of air pollution on daily activities is important for understanding its impact on human welfare and for designing optimal environmental policies.

In this paper, I conduct the first large-scale estimation of the causal effect of daily air pollution fluctuations on economic activity in the United States. Conventional datasets on daily activities often focus on specific regions or sectors (e.g., hiking, cycling, or zoo visits), making it difficult to assess the broader effects of air pollution on economic activities. I address this issue using foot traffic data from SafeGraph, which aggregates de-identified geospatial data from millions of US smartphones. My dataset combines over 600 million trips across 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 estimate its causal effect on daily activities.

I find that air pollution leads to statistically and economically significant reductions in daily activities. On average, a 1 $\mu g/m^3$ (about 10 percent of the mean) increase in PM2.5 concentration results in a 0.50% decrease in daily activities, resulting in a nationwide reduction of 43 million trips annually. A back of the envelope calculation suggest that this corresponds to an annual welfare loss of $1.63 billion. This reduction is significant across many industries, indicating that the effect of air

[1] In my sample, retail trade, accommodation and food services, and entertainment sectors are the three largest sectors, accounting for more than 60% of trips.

pollution on daily activities is more widespread than previously thought. Entertainment and recreation activities experience the most substantial declines (a 1 $\mu g/m^3$ increase in PM2.5 concentration results in a 0.87% decrease in recreational activities), likely due to the greater flexibility in recreational activities compared to other daily routines. Furthermore, the reduction in visits is generally more pronounced for outdoor facilities than for indoor ones. This is presumably because outdoor activities amplify the negative health effects of pollution due to increased respiration and exposure.

I also examine the heterogeneity of responses to air pollution across different income levels and demographic groups. The reduction in daily activities is more pronounced in counties with higher incomes and larger proportions of children, suggesting that wealthier or more vulnerable populations may have greater awareness of elevated air pollution levels. I find that the Black population exhibits a smaller behavioral response to air pollution, which may partially explain the greater negative health impacts of air pollution on Black individuals compared to White individuals (Alexander and Currie, 2017; Gillingham and Huang, 2021). While previous literature on environmental injustice has primarily focused on the unequal distribution of pollution (Banzhaf et al., 2019; Jbaily et al., 2022), my findings emphasize the role of avoidance behavior in exacerbating these disparities. Even when exposed to the same levels of air pollution, low-income and minority groups are less likely to adjust their behavior to mitigate exposure, which may further deepen environmental justice issues. Given that these groups are disproportionately likely to live in more polluted areas (Mikati et al., 2018; Heblich et al., 2021; Tessum et al., 2021), addressing this unequal burden may require targeted policy interventions.

This paper has three main contributions. First, it 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 unclear. In contrast, my analysis uses nationwide phone-location data and examines a broader range of activities, making it more representative than previous studies. My findings suggest that air pollution leads to reductions in activities across most industries, indicating that the behavioral response to air pollution is more widespread than previously understood.

Second, this paper reinforces the importance of characterizing avoidance behavior when quantifying the externalities of air pollution. The results show that individuals, especially vulnerable groups, actively engage in self-protective behaviors to reduce their exposure, such as spending more time at home. As a result, estimates that ignore these behaviors are likely biased downward, since health impacts would be greater without such avoidance actions. These avoidance responses also suggest that the marginal disutility of pollution exposure outweighs the marginal utility people derive from daily activities. Additionally, avoidance behavior itself can be costly, either through increased expenditures (Ito and Zhang, 2020; Zhang and Mu, 2018) or utility losses. For example, staying home and limiting daily activities can reduce physical activity, leading to health issues such as obesity (Hankinson et al., 2010), depression, and anxiety (Paluska and Schwenk, 2000), which impose further societal costs. Lastly, behavioral changes can negatively impact GDP by disrupting key industries. Given that avoidance behavior is widespread, accurately quantifying these effects is important for understanding the true costs of air pollution and determining optimal policies.

Third, 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. My results are not solely driven by air quality alerts, indicating that people adjust their activities in response to more common, everyday pollution fluctuations. This finding suggests that avoidance behavior is more widespread than previously recognized, which in turn implies that the costs of pollution are underestimated.

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 economic activity 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 Economic Activity Data

I obtain the economic activity data from SafeGraph[2]. 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 conductes 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 1[3]. As shown in Table A1, the vast majority of trips documented in the SafeGraph data are business-related, including visits to retail stores, hotels, restaurants, and entertainment facilities, which account for 57% of the total raw visits.

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.

[3] Note 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 schedule[4], 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 database[5]. 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 gird 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

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

[5]See 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 database[6]. 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 $\times$ 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 rates[7] to all POI within a county is 69.93 per 1000 people[8]. The Retail Trade sector has the highest mean visit rate at 22.50 per 1,000 people, followed by the Accommodation and Food Services sector, which has a mean visit rate of 14.58 per 1,000 people. The average daily concentration of PM2.5 is 11.48 $\mu g/m^3$[9], with a standard deviation of 16.95.

[6]Note 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.

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

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

[9]This 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 (°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:

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

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. 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 economic 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. Although high-frequency air pollution is relatively more random than long-term trends, some sources of endogeneity could still bias the estimate of α. First, there could be omitted variables that correlate with both ambient air pollution and economic activity. For example, a local event in a community could increase local PM2.5 levels by raising traffic, while simultaneously affecting economic activity by altering individuals' time allocation. Additionally, although the reverse causality between economic activities and air pollution is arguably weak, it cannot be completely ruled out. For instance, if people choose to stay home instead of going out, the resulting decrease in economic activities could also reduce local air pollution.

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 economic activity 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 error[10]. 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 A5).

[10]Consider 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 economic activities. 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 $\mu g/m^3$ 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) × 100	(2) IV log(visit rate) × 100
PM2.5 ($\mu g/m^3$)	-0.01***	-0.50***
	(0.00)	(0.02)
Temperature (°C)	0.16***	0.25***
	(0.00)	(0.01)
Precipitation (mm)	-2.40***	-1.98***
	(0.03)	(0.03)
Wind speed (m/s)	-0.50***	-0.82***
	(0.01)	(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. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

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 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.

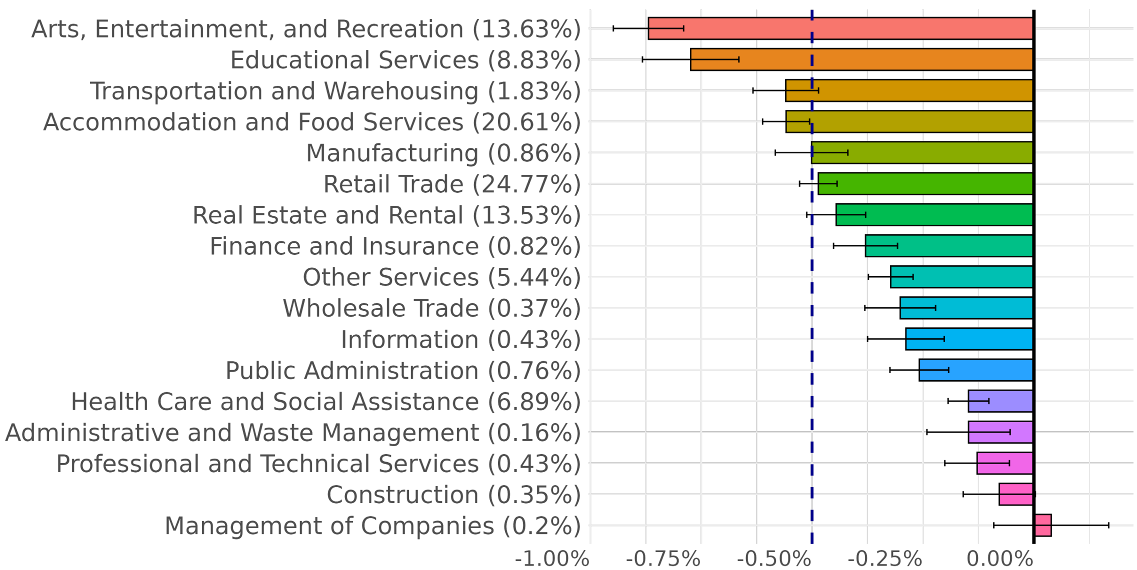


Figure 1. Heterogeneity by 2-digit NAICS Codes

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[11]. and most of the indoor facilities[12], with the exception of a slight increase in bowling centers. Moreover, the reduction in outdoor facilities is generally greater than

[11]Roughly 67% of recreational facilities in my dataset are outdoors, largely because nature parks constitute a significant portion of the data. I divided only the recreational sector into outdoor and indoor categories because (1) it exhibits the largest declines, and (2) other sectors, such as food services and retail trade, are predominantly indoor or a mix of indoor and outdoor settings.

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

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.

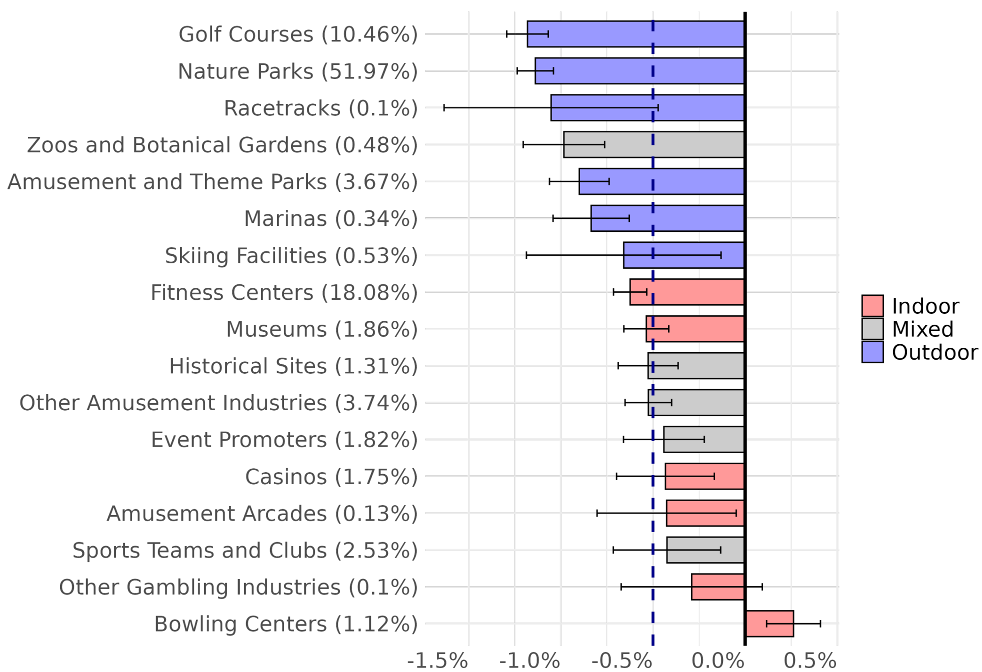


Figure 2. Heterogeneity in Recreational Facilities

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 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.

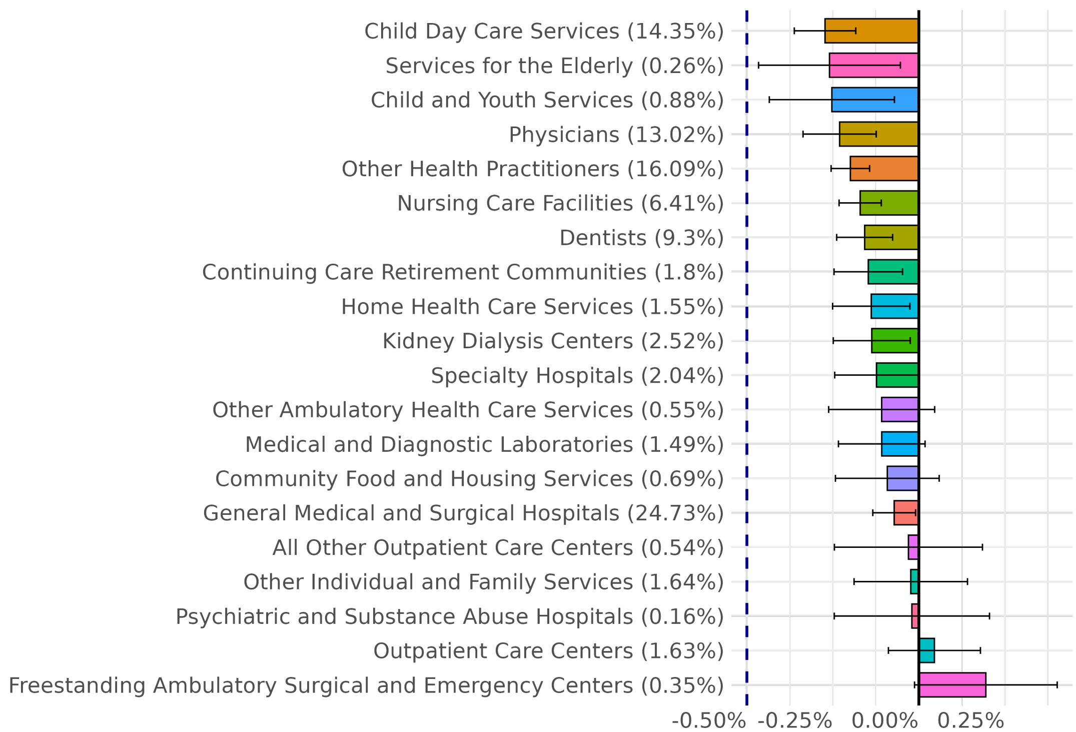


Figure 3. Heterogeneity in Health Care Facilities

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.

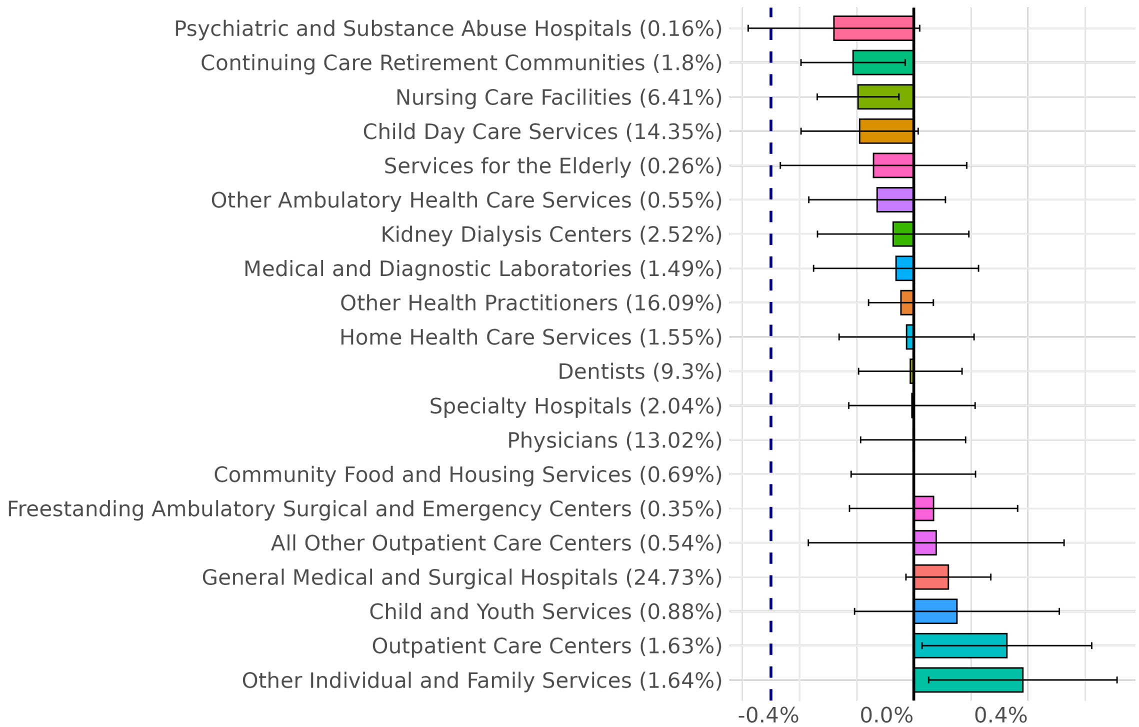


Figure 4. Heterogeneity in Health Care Facilities for Counties with Elderly Population Above the Third Quartile

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 levels[13]. As shown in Panel A of Table 3, 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 activities[14]. Furthermore, if I focus

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

[14]It 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 the negative impact 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 3, 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 3, 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 3. Heterogeneous Effect of PM2.5 across Different Groups

	(1) log(visit rate) × 100	(2) log(visit rate) × 100
Panel A: Different Income Group		
PM2.5 ($\mu g/m^3$)	-0.16***	-0.28***
	(0.03)	(0.02)
PM2.5 × 1{Inc. > Median}	-0.61***	
	(0.04)	
PM2.5 × 1{Inc. > 3rd Quartile}		-0.71***
		(0.07)
Panel B: Different Age Group		
PM2.5 ($\mu g/m^3$)	-0.37***	-0.40***
	(0.04)	(0.03)
PM2.5 × 1{Pct. Children > Median}	-0.34***	
	(0.10)	
PM2.5 × 1{Pct. Children > 3rd Quartile}		-0.60***
		(0.13)
Panel C: Different Race Group		
PM2.5 ($\mu g/m^3$)	-0.56***	-0.55***
	(0.04)	(0.03)
PM2.5 × 1{Pct. Black > Median}	0.10	
	(0.04)	
PM2.5 × 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 **1**{Income > Median} = 1 (**1**{Income > 3rd Quartile} = 1) if personal income in county c is higher than the median (third Quartile). Similarly, **1**{Pct. Children > Median} and **1**{Pct. Children > 3rd Quartile} are dummies for the percentage of children, and **1**{Pct. Black > Median} and **1**{Pct. Black > 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 Intertemporal Substitution

I further investigate the possibility of more dynamic behavioral responses to air pollution by examining whether the decrease in economic activity is offset by an increase on subsequent days or exacerbated by a further decline. For instance, 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 have improved. I use the following distributed lag model to capture this relationship:

$$\log\left(\frac{Y_{ct}}{Pop_c}\right) = \sum_{\tau=0}^{k} \alpha_\tau PM2.5_{c,t-\tau} + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \tag{3}$$

where Y_{ct} is the number of visits in county c on date t, and $PM2.5_{c,t-\tau}$ represents either contemporaneous pollution exposure ($\tau = 0$) or lagged pollution exposure ($\tau \geq 1$). The weather controls and fixed effects are defined the same as in Equation (1).

However, since pollution levels on consecutive days are highly correlated, simply including lagged variables in the regression can cause severe multicollinearity problems. To address this issue, a quadratic distributed lag model is applied to estimate the temporal lagged effect of pollution[15]. The model assumes that the effect over time follows a smooth quadratic function, which is a relatively benign assumption. Specifically, for a quadratic lag function, the lag coefficients are defined as:

$$\alpha_\tau = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 \tag{4}$$

where τ represents the number of lags in the model, and $\gamma_0, \gamma_1, \gamma_2$ describe the lag weights. In my case, $\tau = 5$. Substituting Equation (4) into Equation (3), we obtain:

$$\begin{aligned}
\log\left(\frac{Y_{ct}}{Pop_c}\right) =& \sum_{\tau=0}^{k}(\gamma_0 + \gamma_1\tau + \gamma_2\tau^2) PM2.5_{c,t-\tau} + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \\
=& \gamma_0 \sum_{\tau=0}^{k} PM2.5_{c,t-\tau} + \gamma_1 \sum_{\tau=0}^{k} \tau PM2.5_{c,t-\tau} + \gamma_2 \sum_{\tau=0}^{k} \tau^2 PM2.5_{c,t-\tau} \\
& + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct}
\end{aligned} \tag{5}$$

or:

$$\log\left(\frac{Y_{ct}}{Pop_c}\right) = \gamma_0 z_t^0 + \gamma_1 z_t^1 + \gamma_2 z_t^2 + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \tag{6}$$

[15] A similar estimation strategy was adopted by Barwick et al. (2024), Burkhardt et al. (2019), and Fan (2024).

where:

$$z_t^0 = \sum_{\tau=0}^{k} PM2.5_{c,t-\tau}, \quad z_t^1 = \sum_{\tau=0}^{k} \tau PM2.5_{c,t-\tau}, \quad z_t^2 = \sum_{\tau=0}^{k} \tau^2 PM2.5_{c,t-\tau} \qquad (7)$$

The coefficients on the lagged variables can be recovered from Equation (4), and standard errors are computed using the delta method. The instruments for these endogenous pollution variables $\{z_{ct}\}$ are constructed analogously to Equation (2), except that the lagged endogenous pollution variables are replaced with the corresponding lagged vector of exogenous instrumental variables.

As shown in Figure 5, the contemporaneous effects of air pollution are the largest, followed by a 50% negative effect on the following day. The negative coefficient remains significant up to two-day lags. Instead of compensating for the loss in economic activity during pollution episodes, individuals exhibit inertia, staying home even when pollution levels decline. This behavior further amplifies the negative economic impact of pollution.

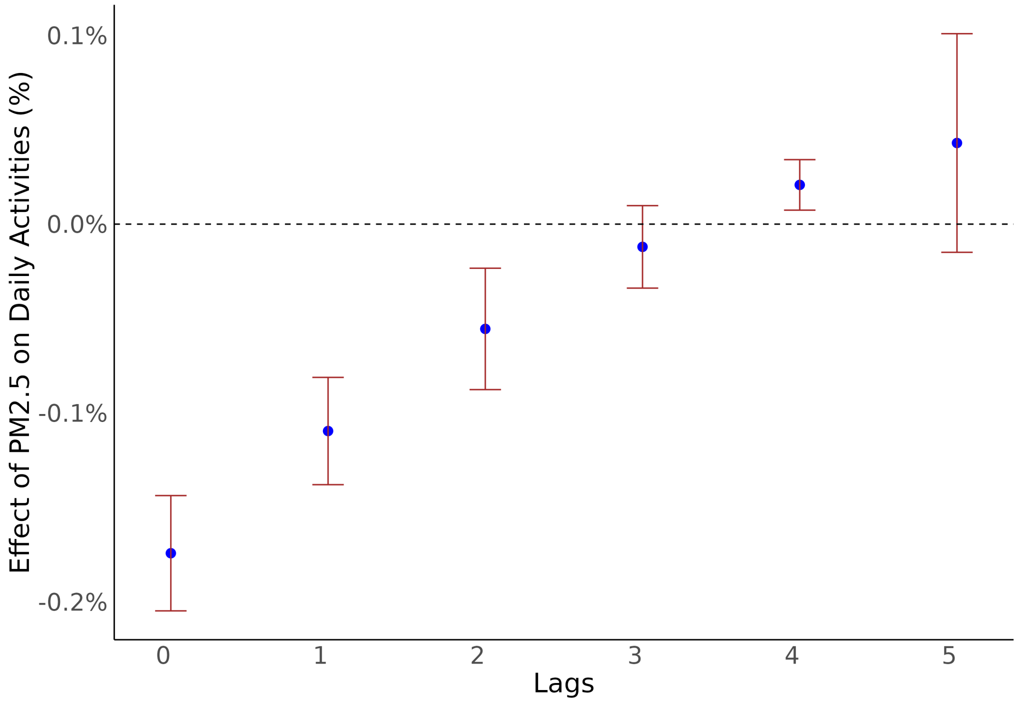


Figure 5. Intertemporal Dynamics of the effect of PM2.5 on Daily Activities

Notes: This figure displays the intertemporal dynamics of the effect of 1 $\mu g/m^3$ increase in PM2.5 on economic activities. Quardratic spline is used to constrain the relationship of inter-temporal effects. Point estiamtes are shown in blue. horizontal bar shows the 95% confidence intervals.

4.4 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 health[16]. Table A4 displays the behavioral guidelines and PM2.5 concentrations

[16]source: https://www.epa.gov/outdoor-air-quality-data/air-data-basic-information. Accessed March 30, 2023

associated with each category.

When PM2.5 concentrations exceed 35.5 $\mu g/m^3$, 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 $\mu g/m^3$ 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 estimation[17].

Table 4 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 6, 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 $\mu g/m^3$ (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) 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.

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

Table 4. Estimation Results from RD of AQI Advisories

	(1)	(2)	(3)	(4)	(5)	(6)
AQI Advisories	-4.3	-4.9	-3.0	-4.6	-5.1	-2.7
	(3.9)	(2.7)	(1.9)	(3.8)	(2.6)	(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 $\mu g/m^3$ (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.

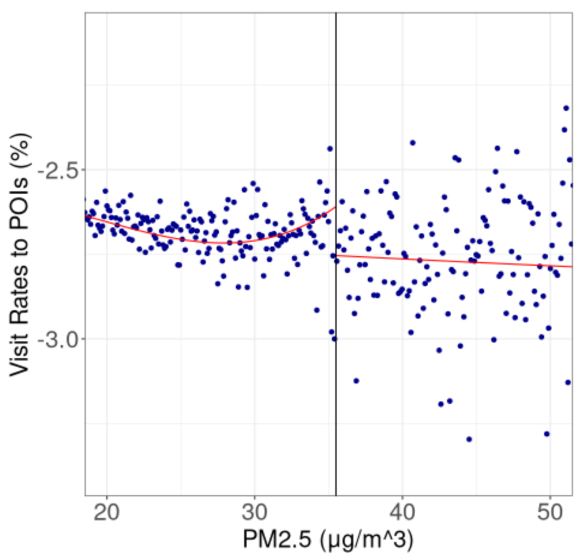


Figure 6. Visit Rates by PM2.5 levels

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 see[18]. 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 data from NCEP North American Regional Reanalysis database (NARR)[19] and apply a 2SLS-style analysis. In the 2SLS-style analysis, I use PM2.5 level as an instrument

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

[19]See https://psl.noaa.gov/data/gridded/data.narr.html. Accessed July 18, 2024.

of visibility and then regress visit rates on the predicted visibility at the second stage. Specifically, I use the following two stage specification:

$$Visiblity_{ct} = PM2.5_{ct} + \mathbf{X}_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct} \tag{8}$$

$$log(\frac{Y_{ct}}{Pop_c}) = \alpha \times Visiblity_{ct} + \mathbf{X}'_{ct}\beta + \sigma_{cy} + \eta_{cm} + \gamma_w + \theta_{my} + \epsilon_{ct}. \tag{9}$$

The results are reported in Table 5. 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 5. Effect of Visibility on Visit Rates

	(1) First Stage Visibility	(2) Second Stage log(visit rate) ×100
Visibility (km)		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 (8) and (9). 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. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

4.5 Welfare Analysis

In this study, I present novel evidence on individuals' daily adjustments in response to air pollution. A one $\mu g/m^3$ increase in PM2.5 leads to a 0.5% reduction in economic activities. The behavioral responses I observe are relatively smaller compared to existing literature on avoidance behaviors. For example, Zivin and Neidell (2009) find that ozone alerts reduce visits to zoos and observatories in Los Angeles by 5–8%, and He et al. (2016) find that an API above 100 reduces movie theater admissions by 2.26%.

This finding is not surprising for two reasons. First, existing literature typically focuses on recreational activities, such as visiting zoos and watching movies—activities more likely to experience larger declines—whereas my study examines a broader range of daily activities, where effects are expected to be smaller. Second, previous studies often focus on air quality alerts, which tend to elicit stronger responses, while my study captures behavioral changes in more common, day-to-day settings. My results are most comparable to Fan (2024), who find that a 10 $\mu g/m^3$ increase in PM2.5 leads to a 1.4% reduction in outdoor exercise.

This study reveals an important social cost of air pollution that has been overlooked in previous research. Air pollution is costly not only due to direct health consequence, but also because the economic consequence resulting from behavioral changes in a daily setting. In this section, I put the results into context by generating some back-of-the-envelope estimates of the economic costs associated with the decline in daily activities due to air pollution. To understand the economic implications, I estimate the following equation

$$log(Income_{cy}) = \beta \cdot log(\frac{Y_{ct}}{Pop_c}) + \sigma_c + \eta_y + \epsilon_{cy} \tag{10}$$

where $Income_{cy}$ is the personal income per capita in county c in year t, and $VisitRates_{cy}$ is the average visit rates in county c in year t. The coefficient of interest, β, represents the association between county daily activities and county income per capita. The estimated coefficient is 0.018 with a standard deviation of 0.004, indicating that a one percent decrease in daily activities is associated with a one percent decrease in personal income per capita.

In the previous section, I show that a 1 $\mu g/m^3$ increase in PM2.5 decreases daily activities by 0.50% on average. Therefore, this is associated with a 0.009%[20] decrease in personal income per year. The average personal income per capita is \$64,665 in the US, which translates to an income loss of \$5.80 per person per year. This implies a total loss of approximately \$1.93 billion[21] annually across the entire US. Since daily

[20] $0.5\% \times 0.018 = 0.009\%$

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

activities also contribute to physical (e.g., reducing obesity) and psychological well-being (e.g., reducing depression), this estimate is likely a lower bound.

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.

A more meaningful comparison would be with two recent papers on avoidance behaviors. Zhang and Mu (2018) estimates the cost of defensive expenditures on face masks during heavily polluted periods to be $187 million USD, and Fan (2024) estimates the cost of physical inactivity, as it relates to major non-communicable diseases, to be $0.55 billion USD in 2017. There are two reasons why my estimated costs are larger, despite smaller declines in daily activities. First, previous studies focus on the costs of specific avoidance behaviors, such as purchasing face masks or reducing outdoor exercise, whereas my estimate captures a more general cost. Second, previous analyses examine the costs associated with heavily polluted days, which occur only a few times a year, while my analysis considers behavioral adjustments to everyday pollution, which occurs consistently throughout the year.

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 (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 6, 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. In addition, I address potential spatial autocorrelation by clustering the standard errors at the county-day level. As shown in Table A5, the estimates remain significant.

Table 6. Robustness of IV Estimates to Instrument Choices

	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 ($\mu g/m^3$)	-0.50***	-0.56***	-0.48***	-0.54***	-0.49***	-0.46***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(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 7, 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.

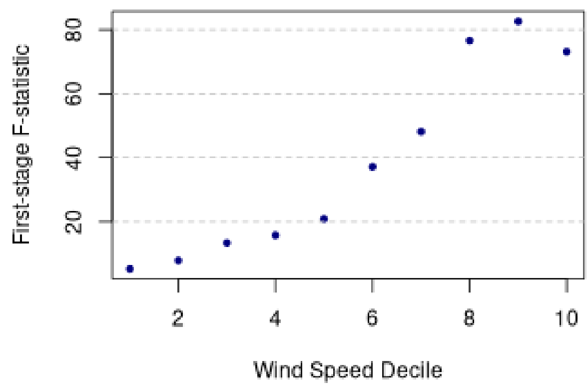


Figure 7. Relationship Between the First-stage F Statistics and Wind Speed

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 7 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 7. Robustness of IV Estimates to Including Different Instrument Lags

	(1) 1 lead and 1 lag	(2) 1 lag	(3) 2 lags	(4) 3 lags	(5) 4 lags	(6) 5 lags
PM2.5 ($\mu g/m^3$)	-0.50***	-0.44***	-0.44***	-0.43***	-0.45***	-0.45***
	(0.02)	(0.02)	(0.0002)	(0.02)	(0.02)	(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 8, 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 9 are comparable to the main specification, which implies the main result is insensitive to the form of the outcome variable.

Table 8. 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***	-0.60***	-0.53***	-0.51***	-0.43***	-0.63***
	(0.02)	(0.04)	(0.08)	(0.10)	(0.08)	(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 9. Robustness to Including Different Forms of Outcome

	(1) log(visit rate) × 100	(2) IHS(visits) × 100
PM2.5 ($\mu g/m^3$)	-0.50***	-0.51***
	(0.02)	(0.02)
Temperature (°C)	0.25***	0.25***
	(0.00)	(0.00)
Precipitation (mm)	-1.95***	-1.96***
	(0.03)	(0.03)
Wind speed (m/s)	-0.81***	-0.82***
	(0.02)	(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 economic activity, I estimate the effect separately before and after the break out of the COVID-19 pandemic[22]. As shown in Table A6, the estimates are very similar, indicating the negative impact is not driven by pandemic-related restrictions. In addition, A7 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

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

rates. My preferred model implies that a 1 $\mu g/m^3$ increase in PM2.5 leads to a 0.50 % 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 economic activity, I am unable to examine the potential long-term adjustments individuals may make. Second, although the phone location-based visitation data is representative of the population at the county level and above, aggregating data to the county level introduces information loss and ignores individual variation. For instance, county-level data does not account for individual-specific factors such as age, race, education level, or health condition. Third, the reduced-form approach identifies only the combined effect of air pollution on economic activity, making it difficult to disentangle the pure avoidance effect. For example, people who fall ill on a polluted day may spend more time at home because they are physically unable to engage in activities, rather than actively choosing to avoid pollution.

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 economics activity 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.

Banzhaf, Spencer, Lala Ma, and Christopher Timmins, "Environmental justice: The economics of race, place, and pollution," *Journal of Economic Perspectives*, 2019, *33* (1), 185–208.

Barwick, Panle Jia, Shanjun Li, Deyu Rao, and Nahim Bin Zahur, "The healthcare cost of air pollution: Evidence from the world's largest payment network," *Review of Economics and Statistics*, 2024, pp. 1–52.

Bresnahan, Brian W, Mark Dickie, and Shelby Gerking, "Averting behavior and urban air pollution," *Land Economics*, 1997, pp. 340–357.

Burkhardt, Jesse, Jude Bayham, Ander Wilson, Ellison Carter, Jesse D Berman, Katelyn O'Dell, Bonne Ford, Emily V Fischer, and Jeffrey R Pierce, "The effect of pollution on crime: Evidence from data on particulate matter and ozone," *Journal of Environmental Economics and Management*, 2019, *98*, 102267.

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.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro, "Defensive investments and the demand for air quality: Evidence from the NOx budget program," *American Economic Review*, 2017, *107* (10), 2958–2989.

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.

Fan, Yichun, "Social cost of lifestyle adaptation: Air pollution and outdoor physical exercise," *Journal of Environmental Economics and Management*, 2024, *127*, 103042.

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.

Hankinson, Arlene L, Martha L Daviglus, Claude Bouchard, Mercedes Carnethon, Cora E Lewis, Pamela J Schreiner, Kiang Liu, and Stephen Sidney, "Maintaining a high physical activity level over 20 years and weight gain," *JAMA*, 2010, *304* (23), 2603–2610.

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, 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.

Ito, Koichiro and Shuang Zhang, "Willingness to pay for clean air: Evidence from air purifier markets in China," *Journal of Political Economy*, 2020, *128* (5), 1627–1672.

Jayachandran, Seema, "Air quality and early-life mortality: Evidence from Indonesia's wildfires," *Journal of Human resources*, 2009, *44* (4), 916–954.

Jbaily, Abdulrahman, Xiaodan Zhou, Jie Liu, Ting-Hwan Lee, Leila Kamareddine, Stéphane Verguet, and Francesca Dominici, "Air pollution exposure disparities across US population and income groups," *Nature*, 2022, *601* (7892), 228–233.

Keiser, David, Gabriel Lade, and Ivan Rudik, "Air pollution and visitation at U.S. national parks," *Science advances*, 2018, *4* (7), eaat1613.

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.

Paluska, Scott A and Thomas L Schwenk, "Physical activity and mental health: current concepts," *Sports medicine*, 2000, *29*, 167–180.

Squire, "What about bias in the SafeGraph dataset?," 2019.

Tessum, Christopher W, David A Paolella, Sarah E Chambliss, Joshua S Apte, Jason D Hill, and Julian D Marshall, "PM2. 5 polluters disproportionately and systemically affect people of color in the United States," *Science advances*, 2021, *7* (18), eabf4491.

Zhang, Junjie and Quan Mu, "Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks," *Journal of Environmental Economics and Management*, 2018, *92*, 517–536.

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.

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	279409983	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	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	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 Clustering Level

	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 ($\mu g/m^3$)	-0.65***	-0.78***	-0.64***	-0.67***	-0.54**	-0.54**
	(0.22)	(0.23)	(0.20)	(0.22)	(0.21)	(0.20)
Number of geographic clusters	20	10	30	10	20	30
Size of wind angle bins (degrees)	90	90	90	60	60	60
R^2	0.80	0.80	0.80	0.80	0.80	0.80
Observations	4,457,413	4,457,413	4,457,413	4,457,413	4,457,413	4,457,413
F-statistic	102.2	191.9	71.3	136.8	76.9	55.0
Dependent Variable Mean	0.73	0.73	0.73	0.73	0.73	0.73

Notes: This table reports the IV estimates using equation (1) and equation (2) when the standard errors are clustered at the county-day level. 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. Dependent variable mean is the average visit rate in percentage terms.

Appendix Table A6. Robustness to COVID-19 Pandemic

	(1) Before COVID log(visit rates) ×100	(2) During COVID log(visit rates) × 100	(3) After COVID log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.27***	-0.39***	-0.46***
	(0.02)	(0.06)	(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.

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

Appendix Table A7. Robustness to Excluding Counties without Satellite Data

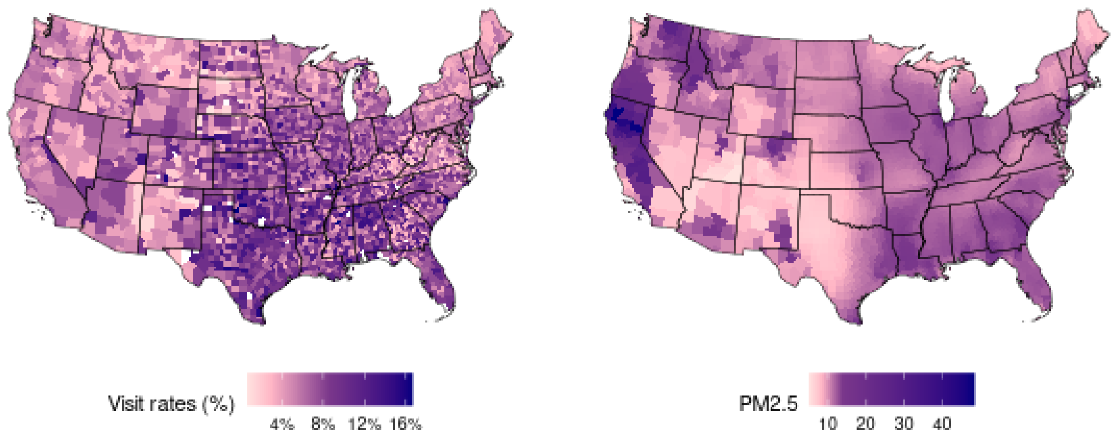
	(1) With IDW log(visit rates) × 100	(2) Without IDW log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.50***	-0.39***
	(0.02)	(0.03)
Temperature (°C)	0.25***	0.22***
	(0.00)	(0.00)
Precipitation (mm)	-1.98***	-1.98***
	(0.03)	(0.05)
Wind speed (m/s)	-0.82***	-0.67***
	(0.02)	(0.03)
First-stage F stat	123.7	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



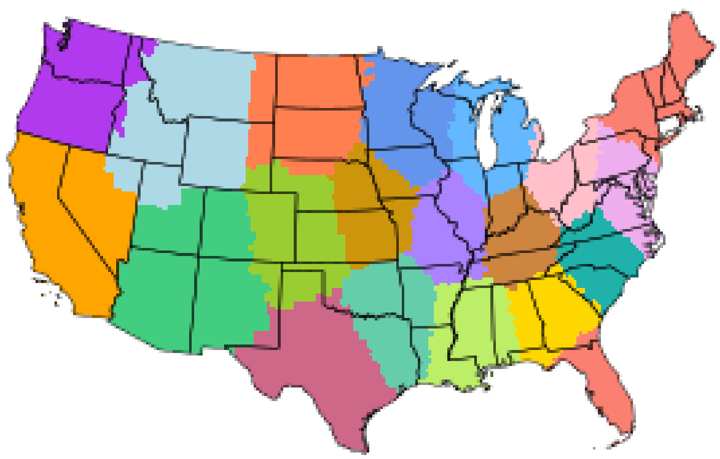
Notes: This figure displays average daily county means for the number of visitations (left panel) and PM2.5 concentration (right panel) from January 1, 2018, to December 30, 2021. As a few counties do not have any visitation data, there are some missing values in the figure.

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

(a) when AQI $\leq$ 100

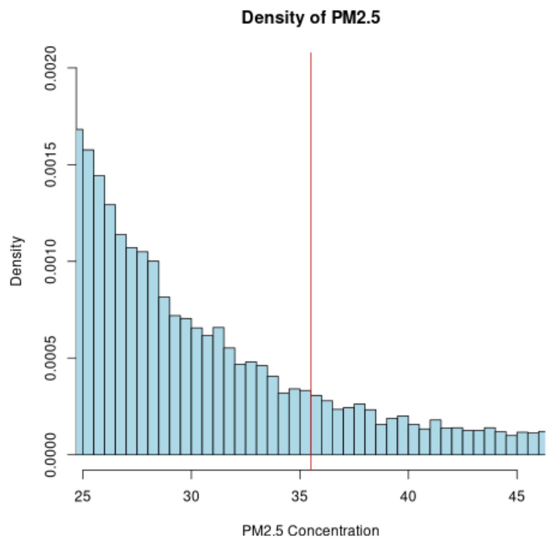
(b) when AQI > 100

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 $\mu g/m^3$, which is the level EPA set to protect public health.