Air Pollution and Avoidance Behavior: Evidence from Daily Activities in the U.S.*

Xinhui Sun†

January, 2024

Abstract

Individuals take action to avoid costly air pollution exposure, yet empirical evidence is limited. I investigate how people modify their daily activities to mitigate the adverse health effects of air pollution. Using phone-location based data from Safegraph, I conduct a large-scale analysis to examine the causal effect of air pollution on visitation rates to leisure facilities across the United States. By 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 0.65% decrease in visitation rates (or a loss of 6595 visits) per day nationwide. This reduction is more pronounced in counties with higher incomes and larger elderly populations, suggesting better awareness of the elevated air pollution among wealthier or more vulnerable groups.

*I am grateful to Tatyana Deryugina, Julian Reif, and seminar participants at the Applied Micro Research Lunch for useful feedback. All errors are my own.

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

1 Introduction

Air pollution is considered as the world's biggest environmental health threat[1]. Existing evidence shows that 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 hours worked (Hanna and Oliva, 2015; Aragón et al., 2017). While the negative effects of air pollution have been discussed a lot, previous literature has predominantly focused on health and labor outcomes. However, individual take various avoidance strategies in order to mitigate the negative health impact of air pollution. On the one hand, they increase spending on defensive expenditures such as face masks (Zhang and Mu, 2018) and air purifiers (Ito and Zhang, 2020). On the other hand, they reduce or cancel their activities to reduce exposure to pollutants. If individuals take action to reduce their exposure to air pollution, the estimated effects of air pollution that ignore these actions are severely biased (Neidell, 2009).

I investigate the relationship between air pollution and leisure activities by analyzing visitation rates to various leisure facilities in the United States. Using anonymized mobile phone location data from the company SafeGraph, I extract patterns of daily visits to these facilities. A primary identification concern when estimating the effect of air pollution is that air pollution is endogenous. For example, existing estimates may be biased due to reverse causality, as visitation-related traffic could increase local air pollution. To resolve the concern, I use change in wind direction as an instrumental variable (IV) for air pollution to derive the causal effect of air pollution on visit rates. I find that, on average, a 1 $\mu g/m^3$ (about 10 percent of the mean) increase in PM2.5 concentration leads to a 0.65% decrease in visit rates, and this negative impact is significant across different types of locations (i.e., zoos, nature parks, amusement parks, golf courses, marinas, museums, casinos, restaurants and supermarkets) and different demographic groups. This translates to a visitation loss of 6595 visits per day, or an annual welfare loss of over \$ 200 million nationwide[2]. The reduction in visitation rates is more pronounced in counties with higher incomes and larger proportions of elderly residents, potentially indicating that wealthier or more vulnerable groups have a better awareness of elevated air pollution levels. In addition, I also show that

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

[2] Give that the recreational use value per person per day is \$93.89, as referenced from the Recreational Use Values Database, the annual welfare loss is calculated as 6595 visits × \$93.89 × 365 days.

air quality advisories, commonly displayed on weather applications or websites when PM2.5 exceeds a certain level, serve as a potential channel through which pollution reduces leisure visit rates.

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. For example, visitors to tourist attractions might not respond to air pollution in the same manner as those exercising in urban parks. 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. Consequently, my findings suggest a slightly smaller effect than what has been previously reported.

Moreover, this paper adds to the relatively understudied literature on avoidance behavior from a different perspective by investigating whether individuals respond to daily pollution fluctuations. A few recent studies show that providing information on air pollution, such as air quality alerts, prompts avoidance behavior. Neidell (2009) uses aregression discontinuity design to estimate the causal effect of smog alerts on visitation to the Los Angeles Zoo and Griffith Park Observatory. Altindag et al. (2017) investigate the impact of avoidance behavior triggered by pollution alerts on various birth outcomes, providing evidence for the effectiveness of pollution alerts in promoting public health. However, air quality warnings are rare and are triggered when Air Quality Index exceeds a certain level, but the negative effects of air pollution continue to increase before and above this threshold (Zivin and Neidell, 2009). Therefore, how individuals respond to these alerts does not necessarily correspond to how individuals respond to air quality itself. Without an air quality alert, individuals might not be aware of the elevated pollution levels. Thus, it remains an open question as to whether people adjust their behavior in response to day-to-day pollution fluctuations. Furthermore, with the development of information technologies making real-time air pollution information readily available nowadays (Yoo, 2021), studying how people respond to this information is becoming increasingly important.

Finally, this paper reinforces the importance of characterizing avoidance behavior when quantifying the externalities of air pollution. The results in this paper are consistent with the hypothesis that people reduce leisure activities on polluted days, which implies existing literature that assumes no avoidance behavior may underestimate the costs of air pollution. Properly accounting for these avoidance behaviors is essential for accurately measuring the externalities of air pollution. In addition, 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. In my case, a lack of leisure activities may induce depression, which impose additional costs on society. Accurately quantifying these effects is important for understanding the costs of air pollution and determining optimal policies.

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, estimates the effect of pollution alerts, 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 SafeGraph[3], a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places, via the SafeGraph Community. The dataset includes information collected from over 45 million smart mobile devices and provides over 3.6 million Points of Interest (POI) covering the entire United States. To enhance privacy, SafeGraph excludes census block group information if fewer than two devices visited an establishment in a month from a given census block group.

In this paper, I focus on leisure facilities and extract these locations from the POI data using their North American Industry Classification System (NAICS) code.

[3] https://www.safegraph.com/. Accessed Sep 12, 2023.

Specifically, I focus on POIs corresponding to these ten categories: Zoos and Botanical Gardens (712130), Nature Parks and Other Similar Institutions (712190), Amusement and Theme Parks (713110), Golf Course and Country Clubs (713910), Marinas (713930), Museums (712110), Casinos (713210), Bowling Centers (713950), Full-Service Restaurants (722511), and Supermarkets and Other Grocery Retailers (445110). In total, I obtain over 600 million observations for these ten categories from January 1, 2018, to December 30, 2021, across the United States.

The POI location data from SafeGraph were collected from over 45 million mobile devices, accounting for over 10% of the US population. SafeGraph conducted 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, I match each location to its county based on latitude and longitude, and then aggregate visits at the county level. After aggregating all visits (including all visits to zoos and botanical gardens, nature parks, amusement and theme parks, golf courses, marinas, museums, casinos, bowling centers, restaurants and supermarkets) at the county level, I have 4,463,557 county-day observations. Additionally, I aggregate the visits to different categories at the county level separately, and the number of county-day observations for each category can be found in Table 1[4].

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 the population in each county varies a lot and more populated counties tend to have a larger number of visits, I use the visitation rates rather than the visitation numbers as the dependent variable in this paper. To obtain the county-level visitation rates, I first aggregate the total number of visits in each county and then divide the total number of visits by the total population of this county.

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

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

the monitors collect data on a 1-in-3 day schedule or 1-in-6-day schedule[5], 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[6]. 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 speed are constructed using the East-West wind vector (u-wind) and the North-South wind vector (v-wind) provided in the database[7]. Wind direction is defined as the direction the wind is blowing from.

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

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

[7]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.

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 4,463,557 county-day observations. The average daily visit rate[8] to all leisure facilities within a county is 7.28 per 1000 people[9]. The average daily concentration of PM2.5 is 11.50 $\mu g/m^3$[10], with a standard deviation of 16.13.

[8]The summary statistics of the number of visits are reported in Table A1

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

[10]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 Facilities	7.28	4.97	4,463,557
Outdoor Facilities			
Zoos and Botanical Gardens	0.09	0.23	927,381
Nature Parks	1.27	1.27	3,552,409
Amusement Parks	0.19	0.47	1,952,944
Golf Courses	0.48	0.70	2,984,282
Marinas	0.13	0.35	1,145,578
Other Facilities			
Museums	0.30	0.72	2,723,077
Casinos	0.77	1.90	398,944
Bowling Centers	0.28	0.49	1,857,387
Restaurants	4.48	3.56	4,415,096
Supermarkets	1.31	1.38	4,139,044
Pollution			
PM2.5 ($\mu g/m^3$)	11.50	16.13	4,463,557
Weather			
Temperature (°C)	13.64	10.69	4,463,557
Total Precipitation (mm)	0.31	0.69	4,463,557
Wind Direction (degrees)	193.46	94.60	4,463,557
Wind Speed (m/s)	2.71	1.54	4,463,557
Demographic			
Population	106,194	38,243	4,463,557
Median Income	49,509	12,902	4,463,557

3 Empirical Strategy

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

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

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}, and month-by-year fixed effect θ_{my}. 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. 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:

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

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[11]. 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).

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 102.2, 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 0.65% decrease in visit rates on average. In addition, 2 shows that warmer temperatures increase visit rates, whereas precipitation and strong wind reduce visit rates.

[11]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.

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 Outdoor Recreation Visit Rates

	(1) OLS log(visit rates) ×100	(2) IV log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.02***	-0.65***
	(0.00)	(0.03)
Temperature (°C)	0.08***	0.22***
	(0.01)	(0.01)
Precipitation (mm)	-2.30***	-1.74***
	(0.04)	(0.14)
Wind speed (m/s)	-0.55***	-0.94***
	(0.01)	(0.02)
First-stage F stat		102.2
Dependent Variable Mean	0.73	0.73
Fixed Effects	Yes	Yes
R^2	0.82	0.80
Observations	4,457,413	4,457,413

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 outdoor recreational facilities. 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 and year-by-month FE. Standard errors are clustered at the county level.

I further investigate if the decrease in visit rates is offset by an increase in visit rates on subsequent days. This assessment considers the possibility of more dynamic behavioral responses to air pollution, such as people adjusting the timing of leisure activities rather than reducing them altogether. 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 the effect of the lagged PM2.5 is not significant (Table 3), suggesting that individuals do not reschedule their activities. Thus, the reduced visit rates in the main result are not compensated by increased visit rates on the following day.

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 visits to leisure facilities. 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 0.73% in my sample, this translates into a 0.005% decrease in visit rates per day per county, or a 6595[12] 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

	(1) log(visit rate) ×100	(2) log(visit rate) × 100
PM2.5, contemporaneous($\mu g/m^3$)	-0.65***	-0.63***
	(0.03)	(0.03)
PM2.5, 1 day lag ($\mu g/m^3$)		-0.09
		(0.06)
PM2.5, 2 day lag ($\mu g/m^3$)		0.03
		(0.09)
First-stage F stat	102.2	100.6
Dependent Variable Mean	0.73	0.73
Fixed Effects	Yes	Yes
R^2	0.80	0.80
Observations	4,457,413	4,454,341

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 outdoor recreational facilities. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Fixed effects include county-by-year, county-by-month 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.

4.2 Heterogeneity

Heterogeneous across Groups of Individuals

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 in-

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

come 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 Column (1) 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 leisure activities[14]. Furthermore, if I focus on counties with income higher than the 3^{rd} 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.

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

	(1) log(visit rates) ×100	(2) log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.53***	-0.50***
	(0.04)	(0.04)
PM2.5 $\times \mathbf{1}\{income > median\}$	-0.20***	
	(0.06)	
PM2.5 $\times \mathbf{1}\{income > 3^{rd}\ quartile\}$		-0.45***
		(0.08)
First-stage F stat	102.1	102.1
Dependent Variable Mean	0.73	0.73
Fixed Effects	Yes	Yes
R^2	0.80	0.80
Observations	4,455,959	4,455,959

Notes: This table reports the effect of daily PM2.5 on outdoor reaction visit rates 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 $\mathbf{1}\{income > median\} = 1$ ($\mathbf{1}\{income > 3^{rd}\ quartile\} = 1$) if personal income in county c is higher than the median (third quartile). Dependent variable mean is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, and month-by-year FE. Standard errors are clustered at the county level.

Furthermore, I investigate the potential influence of a county's age composition on pollution avoidance behavior by using the proportion of residents over 65 years.

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

[14]It is important to note that higher-income counties register a higher average visit rate initially: 0.79% versus 0.67% in lower-income counties. The decrease is proportional to the mean visit rates.

Similar to the income categorization, counties are grouped based on their elderly population percentage: fewer elderly (below median) and more elderly (above median). Their interaction with PM2.5 levels is included in the regression. In Column (1) and (2) of Table 5, the estimated coefficient for the interaction term is negative and statistically significant. As elderly are more vulnerable to air pollution exposure (Schlenker and Walker, 2016; Deschenes et al., 2017), this suggests the vulnerable group would respond more to air pollution.

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

	(1) log(visit rates) ×100	(2) log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.62***	-0.63***
	(0.03)	(0.03)
PM2.5 $\times \mathbf{1}\{age > median\}$	-0.32***	
	(0.10)	
PM2.5 $\times \mathbf{1}\{age > 3^{rd}\ quartile\}$		-0.35**
		(0.17)
First-stage F stat	102.1	102.1
Dependent Variable Mean	0.73	0.73
Fixed Effects	Yes	Yes
R^2	0.80	0.80
Observations	4,455,959	4,455,959

Notes: This table reports the effect of daily PM2.5 on leisure facilities visit rates for different age 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 $\mathbf{1}\{age > median\} = 1$ ($\mathbf{1}\{age > 3^{rd}\ quartile\} = 1$) if proportion of people over 65 years old in county c is exceeds the median (third quartile). Dependent variable mean is the average visit rate in percentage terms. Fixed effects include county-by-year, county-by-month, and month-by-year FE. Standard errors are clustered at the county level.

Heterogeneous Effect across Types of Locations

The effect estimated in the main results might mask variations across different types of location. To provide a comprehensive view, I utilize the broad coverage of the SafeGraph dataset to examine the effect of air pollution on various types of locations. These include outdoor facilities such as nature parks, zoos, marinas, golf courses, and amusement parks, as well as indoor establishments like museums, bowling centers, casinos, restaurants, and supermarkets. Using the same IV model as the main analysis, I separately estimate the effects for each location type.

As shown in Figure 1, visit rates to all outdoor facilities are negatively affected by air pollution. Among them, golf courses have 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.

Furthermore, Figure 2 indicates that visit rates to most indoor facilities are also negatively affected by air pollution. This finding is consistent with He et al. (2022), and they suggest that the negative effect is mainly due to pollution exposure during transportation to the destination. Moreover, the magnitude of this negative impact is generally smaller compared to outdoor facilities, likely because engaging in outdoor activities intensifies the negative health effects of air pollution due to increased respiration and exposure.

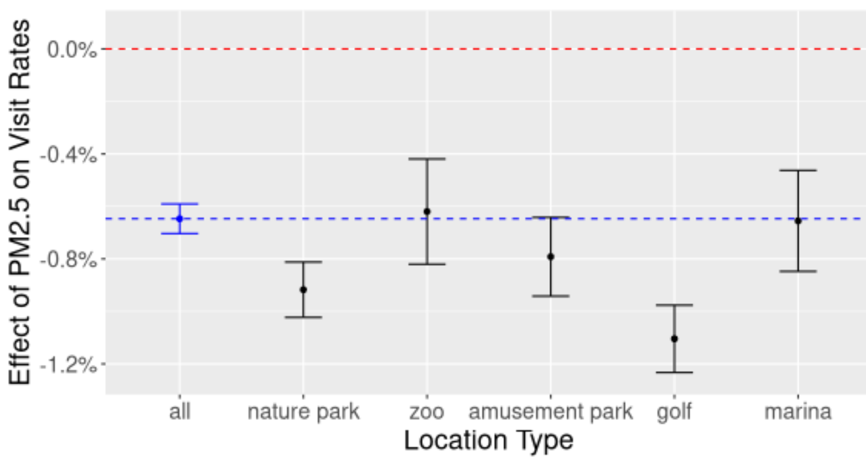


Figure 1. Heterogeneity by Outdoor Location Types

Notes: This figure displays the heterogeneous treatment effect of air pollution on visit rates at outdoor facilities, including nature parks, zoos, marinas, golf courses, and amusement parks. Points represent the estimates, and vertical lines represent the 95% confidence intervals. The blue dashed line represents the estimate for all facilities as presented in the main result.

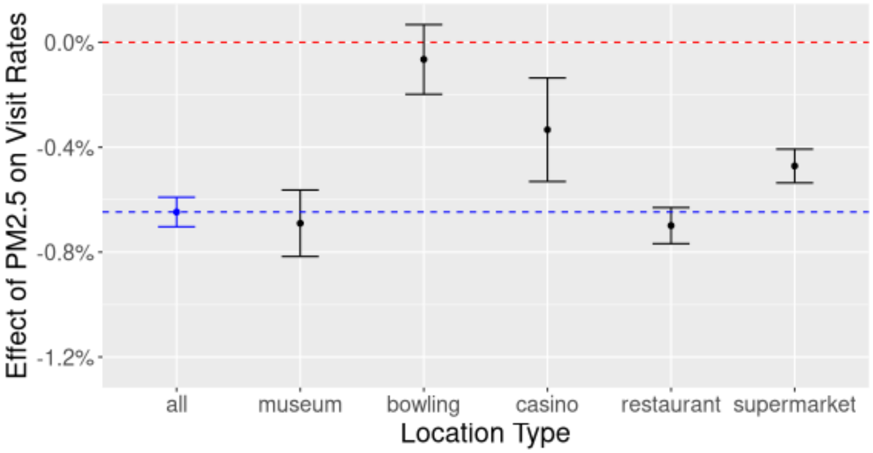


Figure 2. Heterogeneity by Indoor Location Types

Notes: This figure displays the heterogeneous treatment effect of air pollution on visit rates at indoor facilities, including museums, bowling centers, casinos, restaurants, and supermarkets. Points represent the estimates, and vertical lines represent the 95% confidence intervals. The blue dashed line represents the estimate for all facilities as presented in the main result.

4.3 Air Quality Advisories

In this section, I show that Air Quality Index (AQI) Advisory is a possible channel through which pollution reduces leisure visit rates. AQI is an index that spans from 0 to 500, created by the EPA for telling 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[15]. Table A2 displays the behavioral guidelines and PM2.5 concentrations 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). In my sample, 85% of the days are good (code green) and 14% of the days are moderate (code yellow). Individuals actively searching for air quality information will typically find a green category, making an orange day a noticeable change. Therefore, I create a dummy variable indicating whether the real-time PM2.5 is above the orange category (35.5 $\mu g/m^3$) to estimate the effects of air quality advisories. Additionally, since AQI advisories are based on data from EPA's outdoor monitors, I switch to monitor-based

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

PM2.5 data for the following estimation[16].

A number of studies have shown that providing information on air pollution, such as smog alerts, prompts avoidance behavior (Neidell, 2009; Zivin and Neidell, 2009). However, my study is different from theirs in two aspects. First, they are focusing on smog alerts, which are based on ozone and are more severe events[17]. In contrast, given that different states have varying criteria for issuing air quality alerts or warnings, I focus on air quality advisories based on the unified AQI category. Second, they study forecast alerts, whereas I focus on real-time information. A study more closely related to mine is Yoo (2021), which investigates the impact of real-time air quality advisories on attendance at baseball games in South Korea.

Since air quality advisories are displayed on mobile applications and websites when PM2.5 levels exceed 35.5 $\mu g/m^3$, I use a regression discontinuity (RD) Design to estimate the causal effect of real-time advisories. 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. Following Neidell (2009), I estimate the following equation:

$$log(\frac{Y_{ct}+1}{Pop_c}) = \alpha_1 \cdot Advisory_{ct} + g(PM2.5_{ct}, \alpha_2) + \mathbf{X}'_{ct}\beta_3 + \sigma_{cy} + \eta_{cm} + \theta_{my} + \epsilon_{ct}, \tag{3}$$

where $Advisory_{c,t}$ is a dummy variable indicating whether PM2.5 is above the orange category in county c at date t. If $\alpha_1 < 0$, this implies individuals respond to air quality advisories and reduce their leisure activities accordingly. PM2.5 is included as a linear function, quadratic function, or as 10 $\mu g/^3$ interval dummies. Other control variables X_{ct} and fixed effects are defined as in equation (1).

[16]Since some counties do not have monitors, there are 951,465 observations from January 1, 2018 to December 30, 2021.

[17]California state law mandates an air quality episode be declared when ozone is forecasted to surpass 0.2 ppm, equivalent to an AQI of 300.

Table 6. Effect of real-time Air Quality Advisories on Visit Rates

	(1)	(2)	(3)	(4)
Advisory	-0.68***	-0.53**	-0.51**	-0.67**
	(0.26)	(0.25)	(0.23)	(0.32)
PM2.5 ($\mu g/m^3$)	-0.01***	-0.02***	-0.02***	
	(0.00)	(0.01)	(0.01)	
Fixed Effects	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.73	0.73	0.73	0.73
Function Form of PM2.5	Linear	Linear, Different Slope	Quadratic	Interval Dummies
R^2	0.81	0.81	0.81	0.81
Observations	1,660,881	1,660,881	1,660,881	1,660,881

Notes: This table presents RD estimates using equation (3). $Advisory_{ct}$ indicates that PM2.5 concentration is above the orange category in county c at time t. The dependent variable is the log of visit rates to outdoor recreational facilities. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Fixed effects include county-by-year, county-by-month, and month-by-year FE. Dependent variable mean is the average visit rate in percentage terms. Standard errors clustered at the county level are reported in parentheses.

As shown in Table 6, real-time air pollution advisories effectively reduce visit rates to leisure facilities. This result is significant across various functional forms of PM2.5 and is consistent with findings in the literature. Additionally, the coefficient of PM2.5 is also significant, suggesting people react to both the continuous PM2.5 level and the simpler information provided in the advisories.

I conduct additional robustness checks for this model, shown in Table A6. I first show the robustness of the results to different model specifications. Column (1) estimates the regression with the outcome variable expressed in the IHS of visits rather than the log of visit rates. Column (2) displays the results when only county-by-month fixed effects are included. Furthermore, I provide a falsification test by testing for a discontinuity where one should not exist. The results in Column (3) and (4) show that the effects of the artificial discontinuities are not significant, providing further support for the validity of the model.

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 $\mu g/m^3$ increase in PM2.5 decreases visit rates to leisure facilities by 0.65% on average. In this study, the average visit rate is 0.73%. This translates into a 0.005% decrease in visit rates per day per county, or a decline of 6595 visits[18] per day across the United

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

States due to a 1 $\mu g/m^3$ 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)[19]. 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 $\mu g/m^3$ increase in PM2.5 on leisure visitation is approximately 226 million[20] in 2022 USD.

The estimated welfare impact of air pollution on outdoor recreational 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. 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 outdoor recreational activities represents only a tiny fraction of this overall cost. A more meaningful comparison would be a recent paper (Fan et al., 2020), which estimates that the cost of heavy pollution day on park visitation in China is \$20.8 million in 2020 USD per day in northern China, which is \$22.6 in 2022 USD. There are two reasons why my estimates are smaller. First, Fan et al. (2020) find only citizens in northern cities in winter respond to air pollution and estimate the welfare loss base on this subsample. The welfare loss might be smaller if all regions and seasons were considered. Second, they focus on the effect of a heavily polluted day (days with PM2.5 $> 150\mu g/m^3$), while my analysis examines the effect of a 1 $\mu g/m^3$ increase in PM2.5.

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

[19] The database offers 21 primary activity types. However, it is predominantly focused on outdoor recreation activities. However, as both indoor and outdoor activities are forms of leisure, I use these estimates as a rough proxy for welfare valuation.

[20] 6595 visits reduction per day $\times$ 93.89 welfare loss per day $\times$ 365 days

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 A3, the estimates remain significant.

Table 7. Robustness of IV Estimates to Instrument Choices

	(1)	(2)	(3)	(4)	(5)	(6)
PM2.5 ($\mu g/m^3$)	-0.65***	-0.78***	-0.64***	-0.67***	-0.54***	-0.54***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(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	0.80	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	

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 outdoor recreational facilities. 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 3, 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.

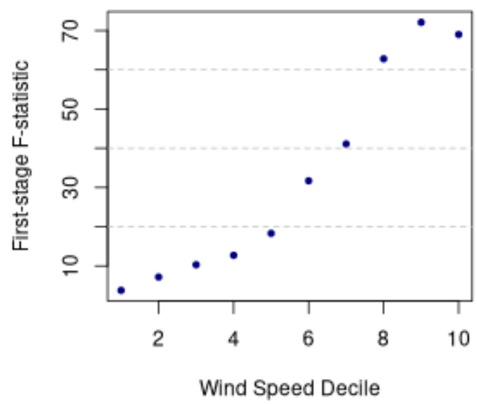


Figure 3. Relationship Between the First-stage F Statistics and Wind Speed

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) 1 lead and 1 lag	(2) 1 lag	(3) 2 lags	(4) 3 lags	(5) 4 lags	(6) 5 lags
PM2.5 ($\mu g/m^3$)	-0.65***	-0.64***	-0.62***	-0.62***	-0.64***	-0.64***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
R^2	0.80	0.80	0.80	0.80	0.80	0.80
Observations	4,457,413	4,457,413	4,454,341	4,451,269	4,448,197	4,445,125
F-statistic	102.2	123.1	107.0	118.8	103.2	99.0
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 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 outdoor recreational facilities. 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)
PM2.5 ($\mu g/m^3$)	-0.65***	-0.57***	-0.62***	-0.71***	-0.45***
	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)
form of weather controls	linear	linear	linear	quadratic	quadratic
county-by-year fixed effects	✓	✓	✓	✓	✓
county-by-month fixed effects	✓	✓			
year-by-month fixed effects	✓		✓	✓	
state-by-month fixed effects			✓		✓
state-by-year fixed effects				✓	
R^2	0.80	0.77	0.78	0.77	0.76
Observations	4,457,413	4,457,413	4,457,413	4,457,413	4,457,413
F-statistic	102.2	105.1	87.0	65.4	83.0
Dependent Variable Mean	0.73	0.73	0.73	0.73	0.73

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) IV IHS(visits) × 100
PM2.5 ($\mu g/m^3$)	-0.67***
	(0.04)
Temperature (°C)	0.23***
	(0.00)
Precipitation (mm)	-1.75***
	(0.04)
Wind speed (m/s)	-0.98***
	(0.03)
First-stage F stat	94.2
Dependent Variable Mean	0.73
Fixed Effects	Yes
R^2	0.97
Observations	4457,413

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 outdoor recreational facilities. 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 pandemic[21]. As shown in Table A4, the estimates are very similar, indicating the negative impact is not driven by pandemic-related restrictions. In addition, A5 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 visit rates at leisure facilities across the United States. I use an instrumental variable approach to address the endogeneity of air pollution, and find a significant negative effect of

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

PM2.5 on visit rates. My preferred model implies that a 1 $\mu g/m^3$ increase in PM2.5 leads to a 0.65 % decrease in visit rates on average, which is significant across different income groups and location types. This translates a reduction of 6595 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 leisure 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 leisure 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 leisure activities that they could potentially have enjoyed on a cleaner day. Consequently, the lost leisure 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.

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.

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.

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.

Fan, Yichun et al., "Air pollution, avoidance behaviors, and neglected social costs: Evidence from outdoor leisure and commuting behaviors." PhD dissertation, Massachusetts Institute of Technology 2020.

Gellman, Jacob, Margaret Walls, and Matthew Wibbenmeyer, "Wildfire, smoke, and outdoor recreation in the western United States," *Forest Policy and Economics*, 2022, *134*, 102619.

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.

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.

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.

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.

Schlenker, Wolfram and W Reed Walker, "Airports, air pollution, and contemporaneous health," *The Review of Economic Studies*, 2016, *83* (2), 768–809.

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

Yoo, Geunsik, "Real-time information on air pollution and avoidance behavior: evidence from South Korea," *Population and Environment*, 2021, *42* (3), 406–424.

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.

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. Summary Statistics - Number of Visits

Variables	Mean	SD	N
All Facilities	818.47	2915.74	4,463,557
Outdoor Facilities			
Zoos and Botanical Gardens	7.80	16.30	927,381
Nature Parks	207.73	681.24	3,552,409
Amusement Parks	14.30	32.45	1,952,944
Golf Courses	34.75	73.41	2,984,282
Marinas	10.07	25.83	1,145,578
Other Facilities			
Museums	13.98	37.78	2,723,077
Casinos	26.30	75.03	398,944
Bowling Centers	17.10	29.91	1,857,387
Restaurants	518.45	1953.93	4,415,096
Supermarkets	95.58	341.19	4,139,044

Appendix Table A2. 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 [22].

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

Appendix Table A3. 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 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 outdoor recreational facilities. 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.

Appendix Table A4. 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 ($\mu g/m^3$)	-0.40***	-0.60***	-0.52***
	(0.03)	(0.06)	(0.05)
First-stage F stat	136.3	38.2	26.1
Dependent Variable Mean	0.68	0.70	0.85
Fixed Effects	Yes	Yes	Yes
R^2	0.83	0.80	0.80
Observations	2,443,150	830,890	1,183,373

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. Dependent variable mean is the average visit rate in percentage terms. Before COVID denotes observations before March 11, 2020, when many U.S. states began implementing stay-at-home orders. During COVID denotes observations from March 15, 2020, to December 11, 2020, when the FDA issued an emergency use authorization for the COVID-19 vaccine. After COVID denotes observations after December 11, 2020.

Appendix Table A5. Robustness to Excluding Counties without Satellite Data

	(1) IV log(visit rates) × 100
PM2.5 ($\mu g/m^3$)	-0.59***
	(0.05)
Temperature (°C)	0.66***
	(0.00)
Precipitation (mm)	-5.94***
	(0.15)
Wind speed (m/s)	-1.40***
	(0.06)
First-stage F stat	24.0
Dependent Variable Mean	0.74
Fixed Effects	Yes
R^2	0.78
Observations	1,658,591

Notes: This table reports the effect of daily PM2.5 on outdoor reaction visit rates 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 and county-by-month FE. Standard errors are clustered at the county level.

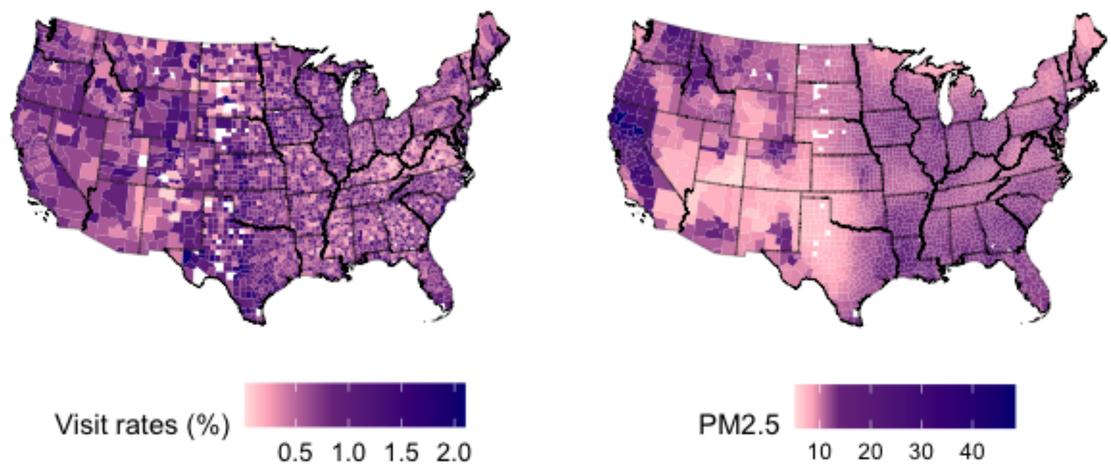
Appendix Table A6. Robustness Check for the Effect of Air Quality Advisories

	(1) IHS(visit)	(2) log(visit rates)	(3) log(visit rates)	(4) log(visit rates)
Advisory	-0.65**	1.06***	-0.08	-0.11
	(0.39)	(0.31)	(0.06)	(0.07)
PM2.5 ($\mu g/m^3$)	-0.01***	-0.02***	-0.01***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Cutoff	35.5 $\mu g/m^3$	35.5 $\mu g/m^3$	10 $\mu g/m^3$	15 $\mu g/m^3$
county-by-month FE	✓	✓	✓	✓
county-by-year FE	✓		✓	✓
year-by-month FE	✓	✓	✓	✓
R^2	0.97	0.78	0.81	0.81
Dependent Variable Mean	0.73	0.73	0.73	0.73
Observations	1,660,881	1,660,881	1,660,881	1,660,881

Notes: This table presents RD estimates using equation (3). $Advisory_{ct}$ indicates that PM2.5 concentration is above the orange category in county c at time t. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. Fixed effects include county-by-year, county-by-month, and month-by-year FE. Dependent variable mean is the average visit rate in percentage terms. The cutoff for the running variable is 35.5 $\mu g/m^3$. Column (1) shows the results when the outcome is IHS of visits. Column (2) displays the results when county-by-year FE is excluded. Column (3) and (4) present the results for a false cutoff value (10 $\mu g/m^3$ or 15 $\mu g/m^3$) for the running variable. Standard errors clustered at the county level are reported in parentheses.

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

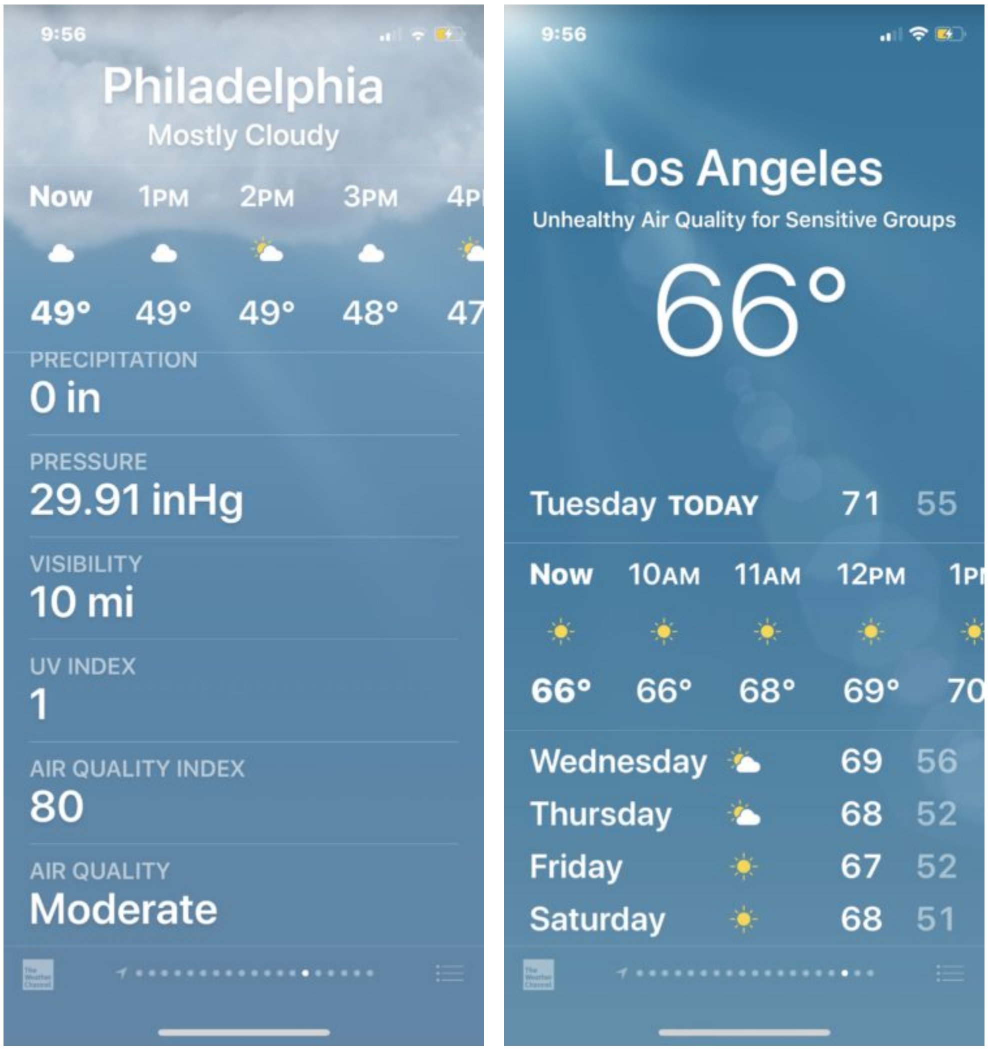
(a) County-level Visit Rates (b) County-level PM2.5 Concentration



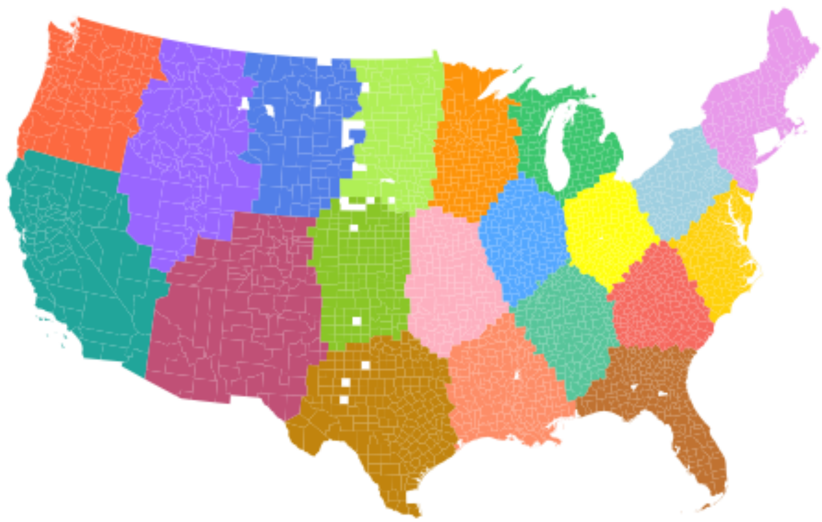
Notes: This figure displays average daily county means for the number of visitations (left panel) and PM2.5 concentration (right panel) from January 1, 2018, to December 30, 2021. As a few counties do not have any visitation data, there are some missing values in the figure.

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

(a) when AQI $\leq$ 100 (b) when AQI $>$ 100

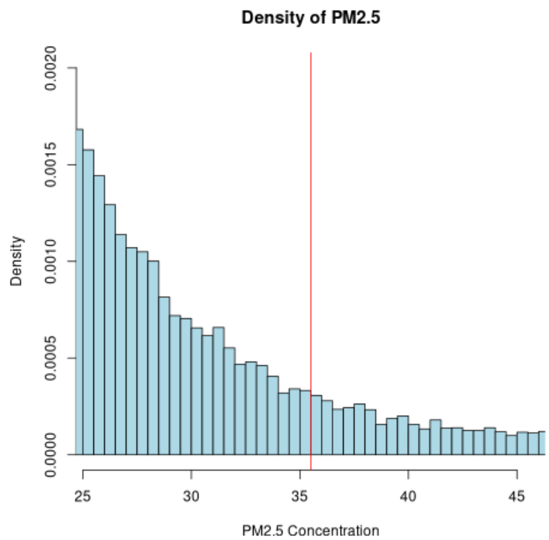


Notes: This figure displays the weather application interface when AQI $\leq$ 100 (left panel) and AQI $>$ 100 (right panel). Specifically, if there is no air quality concern (AQI $\leq$ 100), then there is no message at the top of the Weather overview. Users have to scroll down to find the air quality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