# Air Pollution and School Absences in New York City

Daniel Guth

November 10, 2022

**Abstract**

In this paper, I analyze the effects of changes in day-to-day air pollution levels on daily absences for New York City schools from 2006 to 2019. I combine EPA air quality data with absences for more than 1600 schools. To alleviate endogeneity concerns I use wind as an instrument for transport of air pollution. I estimate that an additional 1 $\mu g/m^3$ of PM2.5 pollution increases absences across all schools by 0.044%, and an extra part-per-billion (PPB) of Ozone increases it by 0.029%. PM2.5 pollution has the largest effects on elementary and middle schools, and on schools with more impoverished students. Examining trends across 14 years of pollution and absences, my results suggest that the decrease in average daily PM2.5 pollution of 5 $\mu g/m^3$ from 2006 to 2019 led to at least 0.2% fewer absences across NYC schools every day. In contrast, Ozone concentrations did not decline over time, and I find that Ozone affects high school absences more than it affects elementary or middle school absences. This work shows the improvements over time in air quality in New York City but also highlights the disparate impacts of air pollution.

## 1 Introduction

Air pollution has darkened American skies for a century. It is well understood on the individual level that air pollution is a risk factor for a variety of respiratory and cardiovascular diseases. However, measuring the effects of air pollution on human health is difficult because the consequences unfold over time and many other factors influence health outcomes. In this paper, I estimate the consequences of exposure to air pollution in day-to-day life by analyzing its effects on school absences for millions of students in New York City. I find strong evidence that increased levels of PM2.5 and Ozone cause additional students to be absent over the next few days, and this effect persists even when pollution levels are below federal limits.

Much of the research on air pollution in the US was done when when carbon monoxide (CO) levels were frequently above the federal limit and the smallest particulate matter that could be measured was 10 microns (PM10). More recent work demonstrates that small particle pollutants (PM2.5) are particularly harmful to human health because their smaller size allows them to penetrate deeper into the lungs. Deryugina et al., 2019 found large increases in mortality and medical spending among Medicare recipients due to PM2.5 pollution, as well as annual benefits of more than $24 billion from national reductions in PM2.5 emissions from 1999 to 2013. I contribute to this literature by providing evidence for the negative effects of PM2.5 and Ozone on NYC children while also showing in the same context that CO and other pollutants do not cause similar harms. I analyze school absences from 2006 to 2019 and find suggestive evidence that the reduction in average PM2.5 pollution over that time contributed to a citywide decline in student absences. My data and methodology also allow me to separately identify the effects of air pollution on each school, and I find novel results for high school absences caused by Ozone.

Previous work on pollution and school absences took place in different pollution environments and with less data. Currie, Hanushek, et al., 2009 analyzed school absences aggregated by six week attendance blocks for Texas in the late 1990s and found that carbon monoxide (CO) was responsible for large increases in absences, PM10 and Ozone had a small effect on absences, and they could not evaluate PM2.5 because atmospheric measurements did not exist at the time. I analyze absences from 2006 to 2019 in NYC at pollution levels much lower than in Currie, Hanushek, et al., 2009, but I find a statistically significant number of absences caused by PM2.5 and Ozone pollution. I use absences at the daily level and develop a framework that allows me to estimate the effects of increased air pollution on different days while controlling for seasonal variation. I further leverage modern climate modeling to use exogenous variation in daily wind direction to show that wind-carried PM2.5 and Ozone directly causes absences. NYC has PM2.5 pollution comparable to the median county in the United States but Ozone pollution higher than the median county, so school districts across the country may have similar amounts of absences caused by air pollution. My estimates represent a lower bound for the harms experienced by other countries, though, most of which have much higher PM2.5 or Ozone concentrations.

School absences represent a child missing class for whatever reason. In papers that track individual students and further break down the cause of absences, respiratory-related absences are the most responsive to pollution. Gilliland et al., 2001 found that Ozone led to increases in both upper and lower respiratory illnesses. Gilliland et al. note that these kinds of respiratory issues often do not land children in the hospital, and so "School absences caused by respiratory illnesses may usefully represent the first tier of adverse effects that are far more common than severe adverse effects." S. Chen, Guo, and Huang, 2018 studied the effects of pollutants in China from 2013 to 2015, finding that PM2.5 and Ozone particularly increased respiratory-illness related absences. I do not observe reasons for school absences in my NYC data, so for each school I control for the average number of absences on less-polluted days and seasonal effects using multiple years of data. After netting out those effects, the remaining absences caused by pollution are likely due to respiratory issues.

There are two advantages to my approach. First, by analyzing school absences at the daily level I can look at the same school cohort before and after a high-pollution event. This analysis allows each school to act as both a control and treatment group on different days, and I can test multiple lag structures to identify the time frame at which pollution causes absences. Specifically, I find differences between the time it takes PM2.5 pollution to cause absences (largest same-day effects) and the time takes Ozone pollution to cause absences (one or two days). Second, by using wind as an instrument, pollution shocks (after accounting for snow) are uncorrelated with school absences except through wind-carried pollutant transport. This allows me to get causal estimates of pollution's effects on school absences using years of variation in air quality levels.

I use two kinds of regression models to estimate the effects of air pollution on absences at the school-day level. In the first set of models, I use the pollutant concentration (for PM2.5 in terms of $\mu g/m^3$, and Ozone in parts-per-billion) and directly regress school absences on air pollution through ordinary least squares (OLS). The second set of regressions uses wind direction as an instrumental variable (IV) in a two-stage least squares procedure (2SLS). In the first stage, daily PM2.5 or Ozone pollution is regressed on wind direction, and then the wind-fitted pollution is used in the second stage absence regression. All models include a panel of fixed effects for day of week, school, school times school year, and school times month to control for unobserved school and seasonal characteristics. To the extent that wind satisfies the exclusion restriction (that is, wind being uncorrelated with school absences except through pollution-transport), this IV regression produces causal estimates of PM2.5 and Ozone on absences. I also show that snowfall in NYC has large effects on absences and is correlated with wind direction, so the main regressions are run on months without snow to avoid model misspecification and violating this exclusion principle. In the robustness section I include OLS regressions run on all months and directly controlling for snow, but measurement error makes it difficult to separate absences caused by snowfall from absences caused by air pollution.

Before previewing the results, it's worth considering these effects relative to the overall burden of pollution. First, we should expect the coefficient on pollution for absences to be relatively small, because there are many factors that cause sickness or for students to otherwise miss school. For instance, snowfall in NYC sometimes cancels school but short of that, icy roads often cause some students to be absent. Other non-illness related absences can be caused by students travelling for holidays or otherwise being unable to get transportation to school. Illnesses can also be caused by seasonal flu or infectious respiratory diseases, although it is generally understood that air pollution impairs the body's immune system and ability to fight off disease. At the same time, air pollution negatively affects everyone and in many more ways than are captured in school absences. Elevated levels of air pollution have been shown to be associated with reductions in test scores, increases in mortality and hospital spending, as well as increase in crime rates.[1] This paper provides a precise

[1] Papers on air pollution and test scores include Xin Zhang, X. Chen, and Xiaobo Zhang, 2018 and Ebenstein, Lavy, and Roth, 2016. Mortality effects from Medicare populations are analyzed in Deryugina et al., 2019 and Di

estimation of one of the short-run causal effects of air pollution, but these estimates are lower bounds on the societal costs of pollution.

Analyzing school absences allows me to test multiple lag structures without issues of displacement or 'harvesting' that arise when analyzing mortality or hospitalization data. In those settings, the effects of pollution might be strongest among people who were going to die or require hospitalization soon regardless of exposure, so researchers often average pollution measurements or mortality over many days. Absences are high-frequency indicators of student health where I can separately identify the effects of pollution on day-of absences from its effects on absences one or more days later. I can identify these changes because I have absence data on the same group of students across many days and wind direction provides day-specific pollution shocks. My extended results show that PM2.5-caused absences are largest on the day-of pollution exposure, with somewhat smaller effects one or two days afterwards, and the effect become approximately zero three or more days later. I find that Ozone-caused absences are also approximately zero after three days, but I find the largest effect on absences two days later. These findings are significant because they represent illnesses that are generally too minor to appear in hospital data but are common in the everyday life of children. In this paper I am able to quantify both the effect size as well as timescale at which pollution causes absences.

I find that elevated levels of PM2.5 and Ozone each lead to absences across all schools and in a variety of regression specifications. My preferred specification, the IV regression with full set of fixed effects, finds that every additional 1 $\mu g/m^3$ of PM2.5 leads to 0.0443% more absences and every additional part-per-billion (PPB) of Ozone leads to 0.0298% more absences. The average school day in NYC has an absence rate of approximately 8%, and going from the median PM2.5 concentration of 8.05 $\mu g/m^3$ to the top decile of most polluted days at 16.98$\mu g/m^3$ times that coefficient implies an additional 0.393% absences that day. Doing the same comparison for Ozone, going from the median pollution day of 32.6 PPB to the top decile of 60.8 PPB implies an additional 0.84% absences two days later. I document a decline in both school absences and daily PM2.5 concentration over this period, and multiplying the decrease of 5 $\mu g/m^3$ from 2006 to 2019 suggests that there are 0.2% fewer absences across NYC schools every day because air quality improved. In contrast to the PM2.5 trends, I find that Ozone concentration slightly increased over this period and continues to cause school absences through the present day. These are lower bounds for the effects of air pollution, though, and cumulative exposure may cause harms not captured in my analysis.

Finally, to understand the treatment heterogeneity I separately analyze each school's absences and compare regression coefficients across schools. I find that elementary and middle schools had much larger PM2.5-induced absences compared to high schools. This difference in effect size is consistent with medical literature on these kinds of respiratory issues, because children's immune systems are developing and they are known to be more susceptible to a variety of illness. Ozone,

et al., 2017. Crime papers include Bondy, Roth, and Sager, 2020 and Herrnstadt et al., 2021.

in contrast, has largest effect sizes for high schools, followed by elementary schools, and smallest effect sizes for middle schools. I am unaware of research that would suggest high school students are more (and middle school students less) susceptible to Ozone-related health issues, but high school students who participate in after school sports programs might spend hours outside during hours when Ozone concentration is highest. If that is the case, high school students might have the most exposure to Ozone pollution and further research should investigate the relationship between Ozone and other health outcomes. Breaking down the effect size for schools by measures of student poverty, I find the elasticity of PM2.5-related absences is higher for schools with more students classified as economically disadvantage. This means that the reduction in PM2.5 pollution from 2006 to 2019 might have had largest benefits for poorer students. Repeating that exercise for Ozone, I find that economic status has less of an effect on pollution elasticity.

The rest of the paper is as follows. Section 2 provides background on air quality regulations and previous papers on school absences. Section 3 outlines the absence, air pollution, and weather data. Section 4 describes the methodology and instrumental variables (IV) approach. Section 5 provides results for OLS, IV, and individual school regressions. Section 6 includes extensions of the model to compare lagged days of pollution and robustness results run separately on each school year as well as OLS results including winter months. Section 7 concludes and discusses the implications for future pollution regulations.

# 2 Background

This section gives background on air pollution and related literature of its effects on humans. Most of the research on air pollution has been done using observational studies which can identify correlations but are less effective at estimating the effects from marginal changes in pollutant concentrations. More recently, environmental economists have developed methods for exploiting quasi-random pollution shocks or changes in regulation that allows for causal identification and estimation of counterfactuals for different pollution levels. I continue this line of research by analyzing the effects of changes in air pollution on school absences, which allows for comparing the same set of students on days of high and low pollution.

## 2.1 Effects of Air Pollution

Air pollution leads to many health problems, but mortality is the most studied because it is the worst possible outcome and governments have kept detailed death records for decades. Anderson, 2009 describes the history of research on pollution and mortality, which have improved over time due to better pollution measurements and computational power. The landmark six-cities study of Dockery et al., 1993 found large increases in mortality from PM2.5 over a fourteen year observational period. Follow-up studies, such as Lepeule et al., 2012, found that average PM2.5 concentrations declined but were still associated with higher mortality. Deryugina et al., 2019 analyzed mortality among Medicare populations using daily changes in pollution from wind and found significant effects from PM2.5 across the US. The other population that is often studied in this context is newborn infants, who have developing lungs and are especially affected by pollution. Currie and Neidell, 2005 finds that reductions in California of CO and PM10 pollution in the 1990s likely prevented more than one thousand infant deaths. Several papers study pollution from car exhaust, Knittel, Miller, and Sanders, 2016 finds similar effects of particulate matter on infant mortality. Currie and R. Walker, 2011 finds that the introduction of E-ZPass reduced congestion and resulted in nearly 10% fewer premature births near toll plazas. School-aged children are underrepresented in these kinds of studies because they are rarely hospitalized for respiratory issues, so absences are a useful kind of high-frequency data to analyze as health outcomes.

A separate strand of literature analyzes the educational effects of pollution on test scores. Ebenstein, Lavy, and Roth, 2016 analyzed students retaking exams in Israel and found that increased PM2.5 pollution on the day of the exam reduced student test scores, which had reductions in average earnings and university completion later in life. Marcotte, 2017 finds that students score between 1-2% lower on tests on days with high pollen or PM2.5 concentration. Xin Zhang, X. Chen, and Xiaobo Zhang, 2018 showed negative effects of both transient and cumulative air pollution on verbal and math test scores. Heissel, Persico, and Simon, 2022 finds that students who transitioned to schools downwind of highways had lower test scores and more absences.

## 2.2 Regulations on Air Pollution

There are costs associated with reducing air pollutant emissions, and a better understanding of the harms caused by air pollution informs cost-benefits analysis for regulation. W. R. Walker, 2013 examines industrial plants newly affected by the 1990 Clean Air Act Amendments and finds they led to more than $5 billion in lost wages, but notes that this is approximately two orders of magnitude less than the health benefits from pollution reduction. Currie and R. Walker, 2019 reviews the economic literature on the Clean Air Act, distinguishing between papers using casual identification on short-run changes in pollution versus longitudinal changes in yearly pollution. The EPA updated the PM2.5 regulations in 1997[2], but PM2.5 and Ozone pollution below the federal limits still has negative effects. Counties in the US that are above the annual limits are designated as 'non-attainment', but in this paper I am able to identify costs from pollution at levels that are well below these limits. I show large reductions of PM2.5 concentration in NYC from 2006 to 2019 [3], but Ozone concentration increased over this time and causes continues to cause absences.

[2] The updated thresholds were in large part due to Dockery et al., 1993 which found large increases in mortality from PM2.5 even below the previous threshold.

[3] This link, shows the reductions in pollution by neighborhood using NYCCAS measurements from approximately 100 locations.

## 2.3 Literature on School Absences

Ransom and Pope III, 1992 was one of the first papers to study the association between PM10 and school absences for schools in Utah. Hales et al., 2016 continues that Utah analysis into the 2010s using PM2.5 and uses a 'control' district exposed to lower levels of pollution to explain residual absences caused by pollution. Gilliland et al., 2001 identified a longitudinal sample of fourth graders and tracked the cause of their absences over six months, finding increases in Ozone concentrations were associated with more respiratory-illness related absences. L. Chen et al., 2000 similarly found absences caused by CO and Ozone for Nevada in 1996-1998, but found a negative association from PM10. The most recent paper on school absences is S. Chen, Guo, and Huang, 2018, which analyzed school absences and their causes for two years in Guangzhou, and found that air pollution increased respiratory illness-related absences. The mean daily absence rate in that paper was 0.22%, though, with worse air quality than in NYC so comparatively each of those absences represents more severe illness. Currie, Hanushek, et al., 2009 is the most comprehensive school absence study in the US, but it was limited to analysis of 6-week absence periods and PM10 instead of PM2.5 concentrations. I update analysis of school absences in the United States for years 2006 - 2019, which allows me to analyze Ozone and PM2.5 to identify causal effects of absences at the individual day level. Daily absence data across years allows me to identify the timescale at which pollution causes absences as well as document a decline in absences over time due to reductions in PM2.5 pollution.

In addition to the negative health effects, missing school causes a child to also miss opportunities to learn. Research that uses individual student-level data is able to study the effects of absences themselves on other educational outcomes. Most of the absence research in this area analyzes the educational effects of chronic absenteeism, such as Chang and Romero, 2008 or Allen, Diamond-Myrsten, and Rollins, 2018, because that has large effects on graduation rates. Less attention is focused on students who are infrequently absent, they but still experience learning loss. Goodman, 2014 finds that school absences had more effect than school closures on test performance, suggesting that teachers are less able to accommodate students who miss school at different times. Liu, Lee, and Gershenson, 2021 connects school absences in specific class periods and shows that they can reduce the probability of graduating. I do not directly test for those educational outcomes in my analysis, but my research provides a method and preliminary evidence showing that PM2.5 and Ozone in NYC over this period caused school absences. Further research is warranted to understand the effects of these pollution-induced absences on educational outcomes.

| School Year | Average Daily PM2.5 (mcg/m3) | Average Daily Ozone (PPB) | Number of Schools | Average Absences (%) | Average Absences Elementary (%) | Total Enrollment |
|---|---|---|---|---|---|---|
| 2006-2007 | 12.8 | 32.73 | 1313 | 9.88 | 7.85 | 917876 |
| 2007-2008 | 13.06 | 34.36 | 1369 | 9.58 | 7.42 | 924888 |
| 2008-2009 | 10.54 | 31.32 | 1428 | 9.47 | 7.56 | 931849 |
| 2009-2010 | 9.85 | 35.27 | 1491 | 8.8 | 6.88 | 957476 |
| 2010-2011 | 10.36 | 36.25 | 1525 | 9.01 | 7.03 | 967880 |
| 2011-2012 | 9.08 | 34.99 | 1551 | 8.37 | 6.36 | 969586 |
| 2012-2013 | 9.08 | 35.03 | 1577 | 8.53 | 6.72 | 976174 |
| 2013-2014 | 8.92 | 34.25 | 1607 | 8.58 | 6.92 | 980377 |
| 2014-2015 | 8.22 | 34.48 | 1615 | 8.21 | 6.63 | 983312 |
| 2015-2016 | 7.83 | 36.49 | 1608 | 7.97 | 6.37 | 977592 |
| 2016-2017 | 7.22 | 35.07 | 1594 | 8.19 | 6.69 | 972781 |
| 2017-2018 | 7.69 | 35.44 | 1594 | 8.41 | 6.97 | 964161 |
| 2018-2019 | 7.35 | 35.16 | 1565 | 8.38 | 6.97 | 947389 |

Table 1: NYC daily absences and pollution concentration by school year, with PM2.5 and Ozone calculated starting in August. Average absences are expressed as a daily percent weighted by school size. Average elementary absences are also weighted by school size and calculated for the approximately 1182 schools that are either elementary or middle schools. Total enrollment is calculated yearly using the fraction of schools each year (96-100%) with demographic information.

# 3 Data

This section describes the different data sources combined to analyze school absences. Table 1 provides school and pollution summary statistics for each school year. School absence data is described first, then air pollution data, and then finally weather data comprising wind and snow.

## 3.1 School Absence Data

School absence data for NYC is publicly available from 2006 to 2021. This paper ends at the 2018-2019 school year to avoid Covid-related school absences starting in 2020.[4] The absence rates

[4]School absence data is available at this link. NYC School locations were obtained using location data from 2012 to 2013 school data and a few dozen school locations from later years.

are reported at the school level, so all analysis will be done at the school level. I analyze a total of 1669 different schools in NYC from 2006 to 2019, representing more than three million school-day absence records and twelve million student-years. Figure 1 shows the number of students as well as the average absences for all schools and for elementary-to-middle schools. The data includes a total of 1182 elementary schools, and previous research has suggested that younger children are most susceptible to pollution, but I also analyze high schools. High schools also have higher absence rates because older students may be absent for work, sports, or other reasons. One advantage of my analysis is that NYC schools are all in the same district and thus have the same regulations and absences procedures.[5] For methodological reasons, I do not directly include school demographics as fixed effects in the absence regressions.[6] In one specification I run each school separately and group the results by type (elementary, middle, or high) and socioeconomic information (fraction of students classified as living in poverty).

### 3.2 Air Pollution Data

Air pollution data comes from the EPA's Air Quality System (AQS) that maintains a network of outdoor monitors across the United States.[7] Using this data, pollution concentrations are measured multiple times throughout the day and then aggregated to form daily readings. The EPA regulates and measures six principal pollutants: particulate matter (PM2.5)[8], Ozone, NO2, SO2, CO, and lead. AQS data on airborne lead concentrations is sparse, so lead is not analyzed in this paper. Outdoor concentrations of CO never went above 50% of the daily limit or near harmful levels from Currie, Hanushek, et al., 2009, so it is analyzed in Appendix Section 8.1 but has no significant effect.[9] OLS results for NO2 and SO2 are also reported in the appendix, with no significant results. Thus, the main pollutants of interest are PM2.5 and Ozone.[10]

For every school and every day over this period, I assigned each school the pollution measurement from the closest PM2.5 or Ozone monitoring site. As described in Zou, 2021, many PM2.5 monitoring sites are active on 1-in-3 or 1-in-6 day schedules, so individual schools are assigned

[5] In NYC when a student is suspended (Suspension Link), they receive alternative instruction and are not marked absent unless they miss those activities. Alternative Reference.

[6] Including yearly school student demographic variables in the model corresponds to trying to estimate and control for the average daily absence rate by race. Across all of NYC and 14 years of data, I do not think it is meaningful to try and estimate those quantities using ecological inference at the school level, so I instead use school fixed effects. Including school by year fixed effects similarly controls for changes in the student body by year without assuming race has a constant absence elasticity across the entire sample.

[7] This link allows you to download daily data by state and year.

[8] The AQS stopped measuring PM10 concentrations for New York in approximately 2006, so PM10 is not analyzed in this paper.

[9] Different pollutants are aggregated on different time-scales: PM2.5 is reported as daily mean concentration; Ozone is reported as daily maximum 8-hour concentration; SO2 is reported as daily maximum 1-hour concentration; CO is reported at daily maximum 8-hour concentration; NO2 is reported at maximum 1-hour concentration

[10] Pollution is used here in direct atmospheric concentrations, whereas some other papers use the scaled Air Quality Index (AQI) with Technical Documentation here.

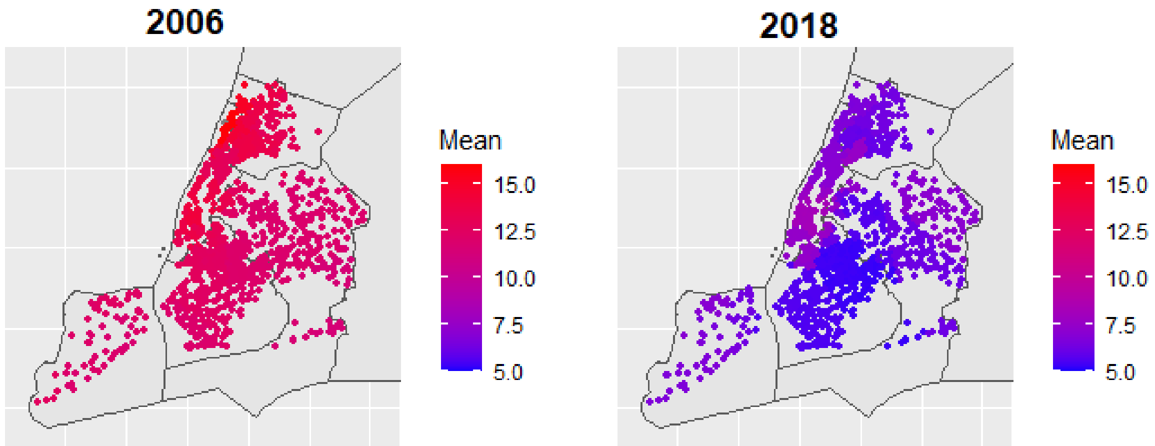


Figure 1: NYC average PM2.5 by school using a common scale for 2006-2007 and 2018-2019 school years. A total of 1312 schools are plotted in 2006 and 1551 schools in 2018. Borough-wide PM2.5 averages for (Bronx, Brooklyn, Manhattan, Queens, Staten Island) school days in 2006 were (13.49, 11.73, 13.44, 11.83, 11.77) and in 2018 were (6.50, 5.71, 7.40, 6.36, 6.61)

multiple monitoring stations in a single year. The median distance between schools and PM2.5 monitoring station across all days was 1.91 miles, 93% of all school day-monitor distances were below 5 miles, and 99.5% of all distances were less than 10 miles. There are fewer Ozone monitoring stations, so the median distance between schools and Ozone monitoring station across all days was 4.14 miles, 97.2% of all school-day monitor distances were less than 10 miles, and 99.6% of all distances were less than 12 miles. Figure 1 shows the average PM2.5 measurement for each school in school years 2006-2007 and 2018-2019 across all school days. PM2.5 pollution in 2006 was highest in the Bronx and Manhattan whereas in 2018 Manhattan is most polluted, but PM2.5 air quality improved everywhere in NYC over this time. NYC is dense and so the absolute difference in pollution between boroughs on the same day is not large, but there is significant variation in average pollution across days and years. Wind is discussed in the next section and acts as a daily pollution shock common to all schools in NYC.

## 3.3 Weather Data

### 3.3.1 Wind Data

As shown in Deryugina et al., 2019, wind-carried transport is responsible for significant variation in daily PM2.5 pollution across the US. In this paper, I leverage changes in wind direction using an instrumental variable analysis that exploits this quasi-random pollution shock. Wind data comes from the North American Regional Reanalysis (NARR) from 2006 to 2019[11], with implementation

[11]Link to data at NARR Monolevel Data Catalog, data used is near-suface (10m) uwnd and vwnd. Data description can be found at NARR homepage.

details borrowed from Deryugina et al., 2019. NARR combines multiple data sets to produce consistent and longitudinal atmospheric data at a 32 km by 32 km resolution. Wind conditions are reported as an east-speed (u-component) and a north-south speed (v-component). Simple trigonometric functions allow for the combination of u- and v-wind into a wind direction.[12] Specifically, the wind angle is calculated as $\theta = \frac{180}{\pi}\text{Arctan}(\frac{|v|}{|u|})$ and then converted from 0-360 degrees depending on the signs of $u$ and $v$ following conventions in Deryugina et al., 2019 and elsewhere:

$$WINDDIR = \begin{matrix} 270 - \theta \text{ if } u > 0 \text{ and } v > 0 \\ 270 + \theta \text{ if } u > 0 \text{ and } v < 0 \\ 90 + \theta \text{ if } u < 0 \text{ and } v > 0 \\ 90 - \theta \text{ if } u < 0 \text{ and } v < 0 \end{matrix}$$

In this form, $WINDDIR$ of zero corresponds to wind blowing from the north into the south, and increasing angle moves clockwise with 90 degrees corresponding to east-west, 180 degrees south-north, and 270 degrees west-east. New York is covered by a single 32km by 32km grid[13], and all school-monitor pairs are assigned the same daily wind direction.

Figure 2 shows a regression of average daily PM2.5 in NYC from 2006 to 2019 against daily wind directions. The median daily PM2.5 concentration across this period was 8.05$\mu$g/$m^3$ and so wind coming out of the south-to-southwest adding an additional 6-7$\mu$g/$m^3$ is a nearly 75% increase relative to wind from the north. This relationship between PM2.5 and wind in NYC is very similar to a figure in the appendix of Deryugina et al., 2019 for King's County, New York from 1999 to 2013. Figure 3 shows the same regression of average daily Ozone on wind direction, and the effect size is significant but smaller (relative to the mean) than for PM2.5 pollution. Against a median daily average of 32.6 PPB, an additional 10 PPB from South-West originating wind compared to East-originating wind adds 30% more Ozone pollution.

#### 3.3.2 Snow Data

NYC gets snow multiple times a year, and snow can lead to school delays or closings as well as icy roads that cause individual students to be unable to get to school. Controlling for snow is important in analyzing winter school absences, and snow data from 2006 to 2019 is taken from NOAA's Global Historical Climatology Network (GHCN) of snowfall observations. Unfortunately, there are only a few snow surface stations compared the pollution monitoring stations, and so most schools are 10 miles away from where their nearest snow measurements were taken. Even at that distance measured snowfall has a large effect on each school's daily absences, but it represents significant measurement error from the amount of snow on the roads near each school. I create

[12] The current model does not use wind speed, which is calculated as $\sqrt{u^2 + v^2}$.

[13] Using NARR's grid, the row coordinate is 259 and the columnn coordinate is 130, centered at latitude and longitude of (40.656, -73.816), near JFK airport.

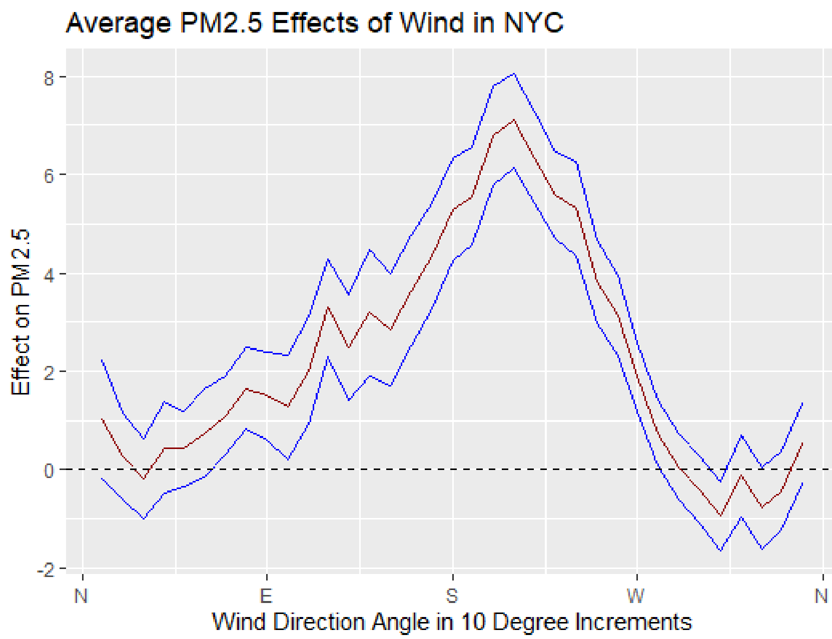


Figure 2: Regression of average daily PM2.5 measurements against 10-degree binned wind direction from 2006 to 2019 with Month by Year fixed effects. Regression coefficients are in red with blue lines representing 95% confidence intervals using robust standard errors. Omitted comparison angle is zero degrees.

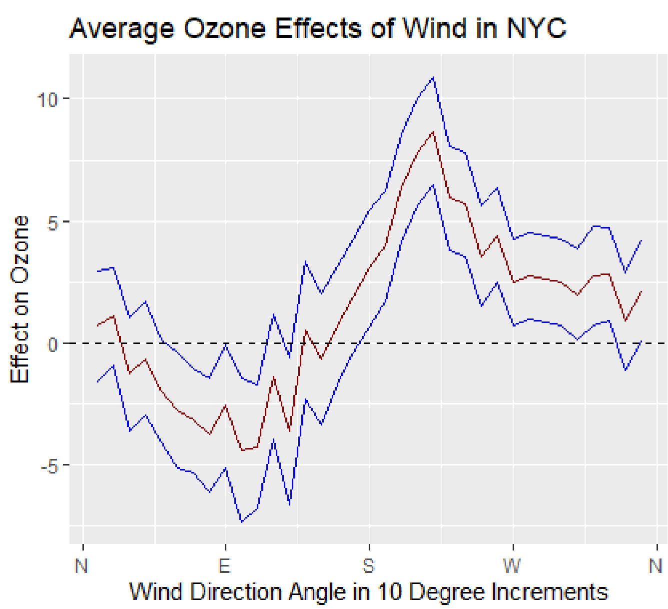


Figure 3: Regression of average daily Ozone measurements against 10-degree binned wind direction from 2006 to 2019 with Month by Year fixed effects. Regression coefficients are in red with blue lines representing 95% confidence intervals using robust standard errors. Omitted comparison angle is zero degrees.

indicator variables for whether any snowfall was measured on the day of absences or the day before, as well as additional indicator variables if cumulative snowfall over the last four days was more than 30cm or 50cm. These snow variables are included in an OLS regression for the robustness section, but the main regressions are restricted to the months of April through November to avoid snow's effect on absences.

# 4 Methodology

This section considers how to analyze the effects of air pollution on school absences. To inform policy, we want estimates of the harms of pollution that are either unbiased or biased downwards. Daily records of absences and pollution allow for testing multiple lag or dose-response functions, but require controlling for seasonal and non-pollution related absences. I first describe the OLS specification of regressing daily school absences on air pollution. The next subsection describes threats to identification, particularly omitted variables that affect both pollution and absences. I then propose using wind in an instrumental variable regression, because daily wind direction is quasi-random and wind transports air pollution. This section concludes by discussing snowfall, which is measured with error and violates the exclusion restriction.

## 4.1 OLS Regression Specification

The first regression specification directly regresses daily absences on daily pollutant concentration.

$$Y_i^t = \beta_c[\sum_j^n \text{Pollution}_i^{t-j}] + dayofweek + school + school \times year + school \times month + \epsilon_i^t \qquad (1)$$

$Y_i^t$ is the percent of students absent for school $i$ at day $t$, the coefficient of interest is $\beta_c$ for each different pollutant. I test several lag structures using pollution from previous days (shown in Section 6.1) and find different time-lags based on pollutant. The main results for PM2.5 uses the same day of pollution (n = 0) on absences, but Ozone has delayed effect on absences and so I use two-days lagged (n = 2). Results are presented in the main section for PM2.5 and Ozone, while CO, NO2, and SO2 have non-significant results and are relegated to the appendix.

Following papers such as Hales et al., 2016 and S. Chen, Guo, and Huang, 2018, I use day of the week fixed effects because absences are statistically more likely on Mondays and Fridays compared to the middle of the week. To further control for unobserved school and seasonal characteristics, I also include school, school times month, and school times year fixed effects. The remaining error term $\epsilon_i^t$ is also at the school $i$ and day $t$ level, with results weighted by school population and the coefficients calculated using HC2 standard errors.

### 4.2 Threats to Identification

Consider the above OLS regression of daily air pollution on school absences. The first problem we might encounter is one of selection, where students who live areas with low pollution might have different probabilities of being absent than students who live in areas with high pollution. I can control for differences in average absence rates by using school and school times year fixed effects, so that kind of selection is unlikely to be an issue. Differences in elasticity or response to pollution by school, however, would lead to biased estimates. For this reason I also have a specification where I directly analyze each school's absences separately and then compare results across schools. I find that type of school (elementary, middle, or high) changes the elasticity for absences caused by Ozone or PM2.5, and school poverty has effects on absences caused by PM2.5 pollution.

The next possible threat to identification involves potential changes across years in school absences or the way they are calculated. I am unaware of any such rule changes, but over a period of 13 calendar years there might also be changes in the availability of school busing or prevalence of mental health related absences. Papers like Twenge et al., 2019 and Bitsko et al., 2018 find increases of anxiety and depression among students over this sample period, which could translate into more absences. Empirically, however, I document a significant decline in NYC average absence rates from 2006 to 2019, combined with a decline in PM2.5 pollution and a slight increase in Ozone pollution. My analysis uses day-to-day variation in pollution that is present across all years and is thus an unbiased estimate of effects across time.

The final threat to identification in the school absence context is omitted variables or seasonality that affects both pollution and absences. For PM2.5, concentrations are highest in winter months which have higher absences due to holiday breaks as well as seasonal influenza. For Ozone, concentrations are highest in summer months which have higher absences potentially due to end-of-year effects. Comparing school days from different months could lead to coefficients biased either upwards or downwards depending on the size of these seasonal correlations. For this reason I add month fixed effects, which makes the comparison between high pollution and low pollution days while attempting to keep seasonal effects constant by comparing days within the same month. To analyze pollution variation that is uncorrelated with these kinds of seasonal effects, I use wind direction in an instrumental variables specification. This IV specification also prevents any possible issues of reverse causality, because wind transports air pollution that causes absences but school absences have zero effect on wind direction.

### 4.3 IV Regression Specification

Using wind direction as an instrument on daily pollution, this regression attempts to determine the causal effects of air pollution. Following Deryugina et al., 2019 I use wind direction as an instrument for pollution in 2SLS. Wind is binned into eight 45-degree indicator variables. The first-stage specification for PM2.5 is

$$PM2.5_i^t = \alpha_c[\sum_{k=0}^{3} WindDirectionBin_i^{t-j-k}] + \text{Fixed Effects} + \epsilon_i^t \tag{2}$$

Pollution is estimated for each day $t$ and school $i$. Deryugina et al., 2019 includes two days lagged wind measurements, and I include three lags because absences are calculated in the morning. In the second stage I use this estimated $\widehat{PM2}.5$ based on each school's closest pollution monitor to get the daily measurements.

$$Y_i^t = \beta_c[\sum_{j}^{n} \widehat{PM2.5}_i^{t-j}] + dayofweek + school + school \times year + school \times month + \epsilon_i^t \tag{3}$$

Fixed effects of day of the week, school, school times year, and school times month are used in both equations. Figure 2 shows a simplified version of the first-stage regression with large variations in daily PM2.5 pollution based on wind direction. The above equations are written in terms of PM2.5, but I repeat the 2SLS procedure for Ozone using two-days lagged pollution and the corresponding lagged wind. Figure 3 shows a version of the first-stage regression for Ozone, with significant variation based on wind direction but smaller effects (relative to the mean concentration) compared to the regression of PM2.5 on wind direction. The remaining error term $\epsilon_i^t$ is also at the school $i$ and day $t$ level, with results weighted by school population and the coefficients calculated using HC2 standard errors.

#### 4.3.1 Individual School Regressions

To directly compare absences between students at the same school as well as examine treatment heterogeneity across schools, I also run the OLS and IV regressions on absences for each school individually. I analyze both PM2.5 and Ozone and report coefficients for individual schools, as well as schools grouped by type (elementary, middle, or high) and by fraction of students that are classified as economically disadvantaged. The OLS and IV regressions are set up in the same way as in equations 1, 2, and 3 except without school fixed effects (because these regressions include only one school at a time); instead, the fixed effects are "$dayofweek + year + month$" for all individual school results.

#### 4.3.2 Snow Problems

Snow poses two challenges in analyzing school absences in New York City. The first problem is that snow has large effects on school attendance but is measured with error. As described in the

Data section on snow, most schools are 10 miles away from where their nearest snow measurements were taken. Even at that distance snow has a large effect on each school's daily absences, but it represents significant measurement error from the amount of snow on the roads near each school. Snow might melt at different rates across NYC depending on plow timing and road salt usage, so the effects on absences for following days might also vary. For this reason the main OLS and IV regressions in the results section are run on months that do not have snow, but I include OLS results using all months in Section 6.3.

The second issue with snow is that it is correlated with wind direction. For the IV wind regression to satisfy the exclusion restriction, there must be no channel by which daily wind affects school absences except through wind-carried pollution. In a meteorological analysis Blechman, 2002 finds that "New York City exhibited preferences for snow with east and northeast winds with few westerly wind events". Figure 4 updates the analysis of wind direction on snowy days for NYC from 2006 to 2019, and shows there remains a strong correlation between wind direction and snowfall. Specifically, snow most often occurs on days that have low PM2.5 pollution due to winds from the north and north-east and rarely occurs on high PM2.5 days with winds coming from the south and west. Snowfall has a strong effect on school absences, so the IV exclusion restriction fails on those days and the coefficient is biased downwards. The main IV results are thus run on months without snow, April though November. More localized snow data or improved functional forms for the snow-absence relationship could allow for analysis of winter months. I am unaware of other omitted variables or correlations that would violate the exclusion restriction for my IV regressions, but if there were, my OLS estimates still provide evidence of the same magnitude and timing of absences caused by PM2.5 and Ozone pollution.

# 5 Results

## 5.1 OLS Results

This section provides OLS regression results for PM2.5 and Ozone across all years from 2006 to 2019. As described in Section 4.3.2, snow is measured with error and causes large amounts of absences unrelated to pollution, so the main OLS regressions omit snow-containing months December through March. OLS estimates for all months and directly controlling for snow are presented in Section 6.3.

Table 2 shows the PM2.5 OLS results of same-day pollution on absences. Against an average of approximately 8% students absent per day across NYC during this period, the coefficient represents absences caused by an additional 1 $\mu g/m^3$ of daily PM2.5 pollution. The coefficients are positive in all specifications, although adding more fixed effects reduces the magnitude. The preferred specification is Model 5 with all fixed effects, showing an extra 0.0251% more absences per $\mu g/m^3$.

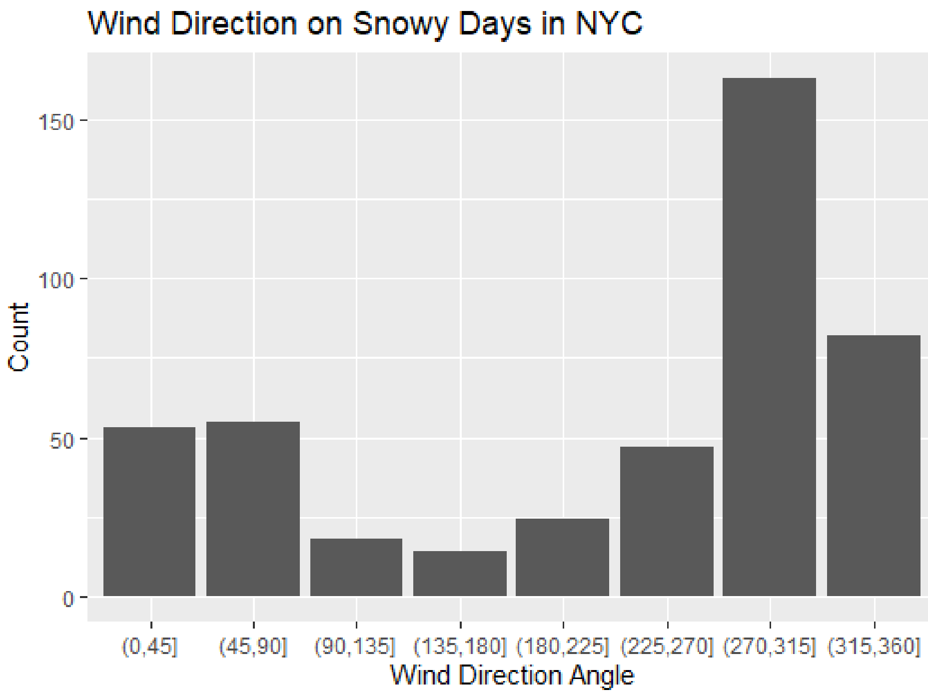


Figure 4: Wind direction of NYC days that had non-zero snowfall from 2006 to 2019. With a total of 456 days with snow accumulation over that range, 245 (53.7%) had wind coming from north-to-northwest (270 to 360 degrees).

Going from the median NYC day of PM2.5 pollution of $8.05\mu g/m^3$ to the ninetieth percentile day of $16.98\mu g/m^3$ times that coefficient implies an additional 0.224% more absences that day.

Table 3 shows the Ozone OLS results of two-days lagged pollution on absences. Against an average of approximately 8% students absent per day across NYC during this period, the coefficient represents absences caused by an additional part-per-billion (PPB) of daily Ozone pollution two days earlier. As shown in later Section 6.1, Ozone has largest effects on absences one or two days following exposure. The coefficients are positive in all specifications, and school times month fixed effect especially reduces the coefficient magnitude. This is because Ozone is highest in hot summer months, and comparing absences between months may attribute seasonal effects to differences in average Ozone, so it is important to include fixed effects to control for monthly variation. The preferred specification is Model 5 with all fixed effects, showing an extra 0.0209% more absences per PPB of Ozone. Going from the median NYC day of Ozone pollution of 32.6 PPB to the ninetieth percentile day of 60.8 PPB times that coefficient implies an additional 0.589% more absences two days later.

| PM2.5 OLS Summer Regression | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration.x | 0.083662*** | 0.060179*** | 0.039141*** | 0.048445*** | 0.025061*** |
| | (0.001326) | (0.000995) | (0.000893) | (0.001028) | (0.000918) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.000 | 0.536 | 0.616 | 0.538 | 0.618 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 2: OLS Regression of same-day PM2.5 on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

| Ozone OLS Regression of Summer Months | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day2 | 0.057419*** | 0.076309*** | 0.016713*** | 0.079816*** | 0.020912*** |
| | (0.000520) | (0.000380) | (0.000420) | (0.000379) | (0.000420) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.011 | 0.554 | 0.616 | 0.557 | 0.619 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 3: OLS Regression of two-days lagged Ozone on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

| | No FE | School FE | School + Month | School + Year | School + Month + Year |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration.x | 0.050818*** | 0.050530*** | 0.049211*** | 0.057193*** | 0.044293*** |
| | (0.002287) | (0.001590) | (0.001436) | (0.001627) | (0.001458) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.013 | 0.508 | 0.597 | 0.511 | 0.600 |
| R2 Adj. | 0.013 | 0.508 | 0.596 | 0.510 | 0.600 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

PM2.5 IV Regression of Summer Months

Table 4: IV Regression of same-day PM2.5 on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

### 5.2 Wind IV Regression Estimates

This section provides IV regression results for PM2.5 and Ozone across all years from 2006 to 2019. As described in Section 4.3.2, snow is measured with error and causes large amounts of absences unrelated to pollution, so the main IV regressions omit snow-containing months December through March. Figures 2 and 3 show the first stage regressions of daily PM2.5 or two-days lagged Ozone on school absences.

Table 4 shows the PM2.5 IV results of same-day pollution on absences. Against an average of approximately 8% students absent per day across NYC during this period, the coefficient represents absences caused by an additional 1 $\mu g/m^3$ of daily PM2.5 pollution. The coefficients are positive in all specifications, although adding more fixed effects reduces the magnitude. The preferred specification is Model 5 with all fixed effects, showing an extra 0.0442% more absences per $\mu g/m^3$ of PM2.5 pollution. Going from the median NYC day of PM2.5 pollution of $8.05 \mu g/m^3$ to the ninetieth percentile day of $16.98 \mu g/m^3$ times that coefficient implies an additional 0.393% more absences that day. Comparing the PM2.5 IV coefficient magnitude to the OLS results, with no fixed effects the OLS coefficient is approximately sixty percent larger than the IV coefficient, but in Model 5 the IV coefficient is seventy-five percent larger than the OLS coefficient.

Table 5 shows the Ozone IV results of two-days lagged pollution on absences. Against an average of approximately 8% students absent per day across NYC during this period, the coefficient represents absences caused by an additional part-per-billion (PPB) of daily Ozone pollution two days earlier. As shown in later Section 6.1, Ozone has largest effects on absences one or two days

| Ozone IV Regression of Summer Months | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day2 | 0.079053*** | 0.087996*** | 0.030160*** | 0.091381*** | 0.029830*** |
| | (0.001476) | (0.001067) | (0.001037) | (0.001042) | (0.001029) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.020 | 0.526 | 0.596 | 0.531 | 0.601 |
| R2 Adj. | 0.020 | 0.525 | 0.596 | 0.531 | 0.600 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 5: IV Regression of two-days lagged Ozone on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

following exposure. The coefficients are positive in all specifications and, similarly to Table 3, the school times month fixed effect especially reduces the coefficient magnitude. The preferred specification is Model 5 with all fixed effects, showing an extra 0.0298% more absences per PPB of Ozone. Going from the median NYC day of Ozone pollution of 32.6 PPB to the ninetieth percentile day of 60.8 PPB times that coefficient implies an additional 0.84% more absences two days later. Comparing the Ozone IV coefficient magnitude to the OLS results, the IV coefficient is consistently larger than the OLS coefficients regardless of fixed effects, and it is approximately fifty percent larger in Model 5.

### 5.3 Individual School IV Analysis

This section presents PM2.5 and Ozone IV results from individual schools, with the corresponding OLS results relegated to the appendix. Results are shown first for all schools in histogram form, then grouped by school type, and finally grouped by percentage of students considered economically disadvantaged.

Figure 5 shows the PM2.5 IV results run on each individual school. The vertical red line marks zero effects on absences, and 176 out of 1623 (10.8%) schools have negative PM2.5 coefficients. Running each school individually increases the standard errors, though, and so only 319 out of 1623 (19.7%) of schools have IV coefficients that are significantly above zero. Comparing across all schools, the median IV coefficient is 0.0494 and the mean IV coefficient is 0.0522, both of which are slightly larger than the full regression IV coefficient of 0.0443. This distribution of effect size

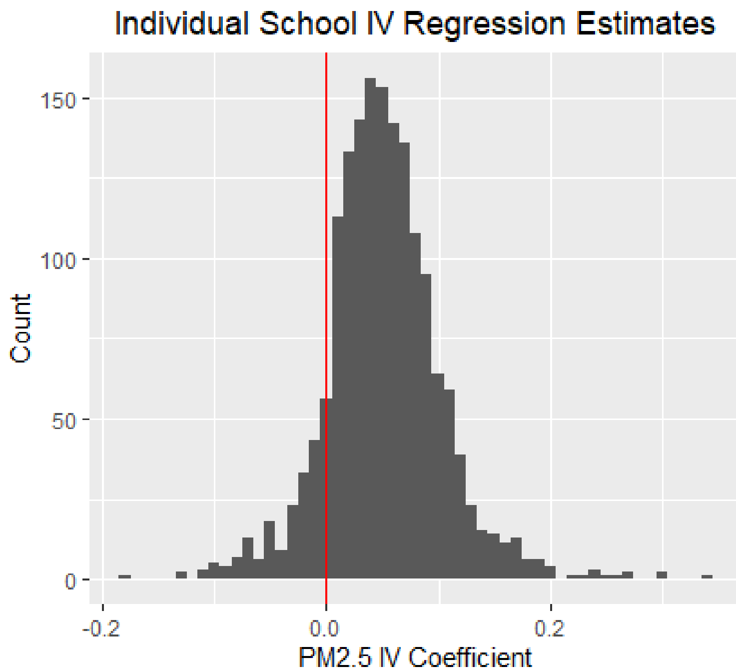


Figure 5: PM2.5 IV regression with every school run individually. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

by school is consistent with PM2.5 causing absences combined with some noise by school, but it is possible some schools do not have absences caused by PM2.5 pollution.

Figure 6 shows the Ozone IV results run on each individual school. The vertical red line marks zero effects on absences, and 188 out of 1623 (11.6%) schools have negative coefficients. Running each school individually increases the standard errors, though, and so only 321 out of 1623 (19.8%) of schools have IV coefficients that are significantly above zero. Comparing across all schools, the median IV coefficient is 0.0322 and the mean IV coefficient is 0.0386, both of which are slightly larger than the full regression IV coefficient of 0.0298. The distribution of effect size by school is consistent with Ozone causing absences combined with some noise by school, but it is also possible some schools do not have absences caused by Ozone.

Figure 7 presents a box plot of PM2.5 IV coefficients grouped into elementary, middle, and high schools. The median PM2.5 coefficients by type are 0.539 for elementary schools, 0.0556 for middle schools, and 0.0353 for high schools. The difference between the average elementary and middle school coefficient is not statistically significant ($p = 0.12$), but both are statistically larger than the average coefficient among high schools; in fact, 25% of all high schools have a negative PM2.5 coefficient. Previous work has generally found that PM2.5 causes absences for elementary or middle schools, but high schools absences are often not analyzed. I include all types of schools in the main analysis, but this difference by school type fits with the medical understanding that PM2.5 harms younger children more than high school aged children.

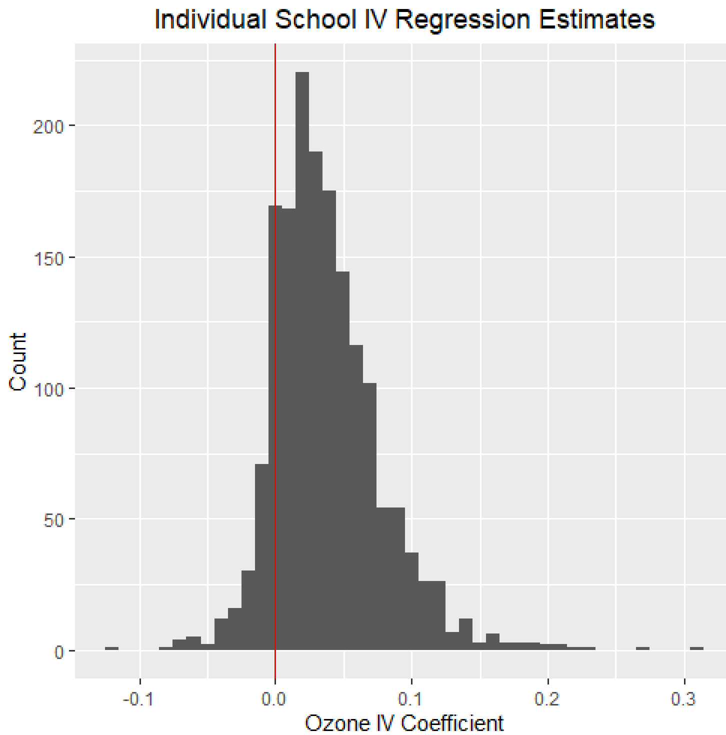


Figure 6: Ozone IV regression with every school run individually. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

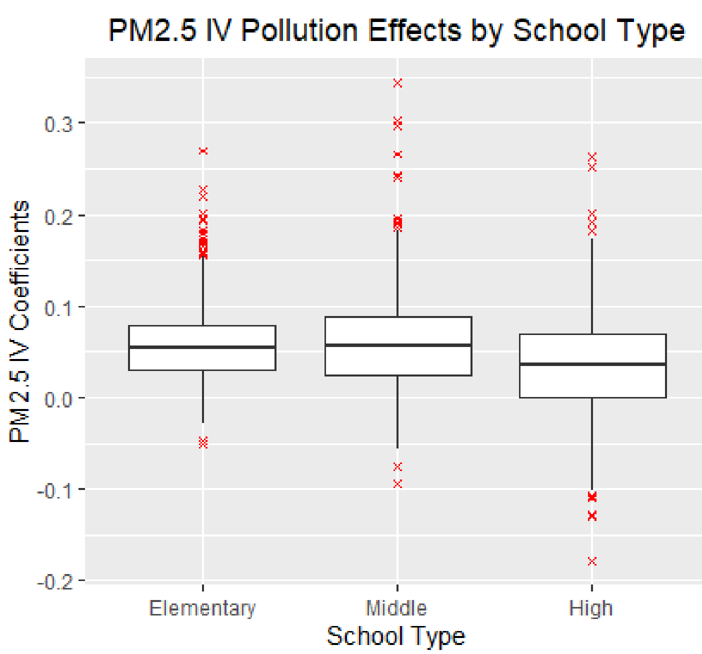


Figure 7: PM2.5 IV regression with every school run individually and grouped into school type. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

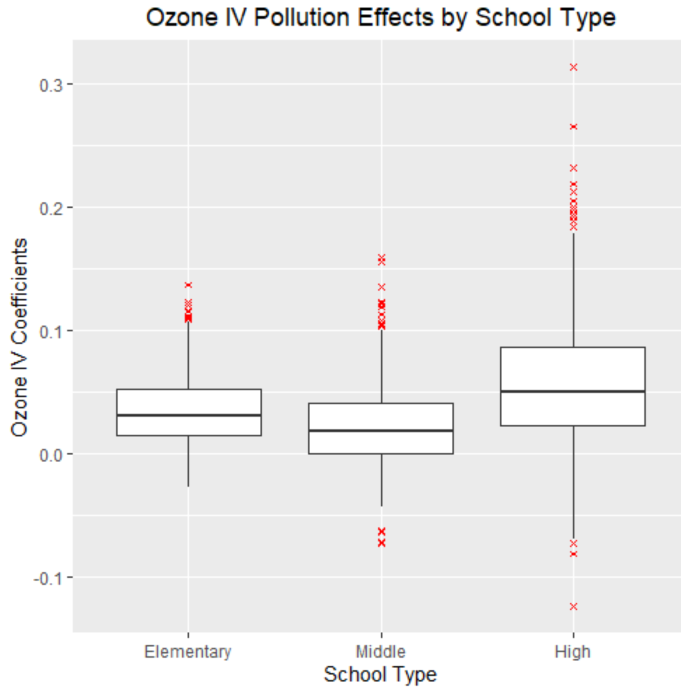


Figure 8: Ozone IV regression with every school run individually and grouped into school type. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

Figure 8 presents a box plot of Ozone IV coefficients grouped into elementary, middle, and high schools. The median Ozone coefficients by type are 0.0305 for elementary schools, 0.0178 for middle schools, and 0.0497 for high schools. Pairwise differences between school types are all statistically significant, with high schools having statistically larger effects than elementary schools which have larger effects than on middle schools. Most of the medical literature suggests that younger children are generally more at risk because of their developing immune systems, so this large Ozone effect on high school absences (and small effect on middle school absences) is surprising. I conjecture that this difference might be due to high school students doing after-school sports, which would have them exercising outdoors during peak Ozone hours. Previous analysis of school absences often focused only on elementary or middle school, so these Ozone-caused absences among high school students warrant further investigation.

Figure 9 presents a box plot of Ozone IV coefficients grouped by a school's percentage of students considered economically disadvantaged. Most of NYC public school students are classified as poor, with poverty quartiles by school of Q4 = [0, 76.07], Q3 = (76.07, 87.92], Q2 = (87.92, 93.08], and Q1 = (93.08, 100], with Q1 being the richest schools and Q4 being the poorest schools. The median coefficients by poverty quartile are 0.0381 for Q1, 0.0465 for Q2, 0.0552 for Q3, and 0.0689 for Q4. Average effect sizes are increasing in poverty quartile, with statistically significant increases from Q4 to Q3 ($p = 0.016$), from Q3 to Q2 ($p = 0.021$), and from Q2 to Q1 ($p = 0.0001$). Previous

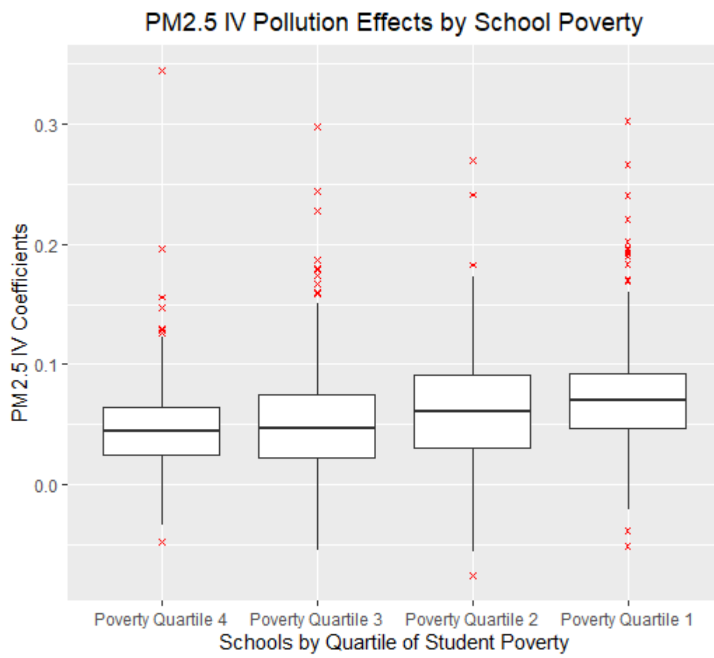


Figure 9: PM2.5 IV regression with every school run individually and grouped by percentage of students considered economically disadvantaged, where Quartile 4 is richest and Quartile 1 is poorest. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

research has shown that socioeconomic status affects both the frequency and severity of asthma, and higher asthma rates at schools with more students in poverty could explain these findings. Alternatively, students from families that are more well-off may have smaller cumulative exposure from pollution or more ways to mitigate the harms of pollution through air filtration or increased healthcare spending.

Figure 10 presents a box plot of Ozone IV coefficients grouped by a school's percentage of students considered economically disadvantaged, using the same poverty quartile bins as in the Figure 9 with Q1 being the richest schools and Q4 being the poorest schools. The median coefficients by poverty quartile are 0.0220 for Q1, 0.0440 for Q2, 0.0427 for Q3, and 0.0456 for Q4. In contrast to PM2.5, average effect sizes are not increasing in poverty quartile, with statistically significant increases from Q4 to Q3 ($p = 0.0000$) but not from Q3 to Q2 ($p = 0.652$) or Q2 to Q1 ($p = 0.248$). This analysis suggests that socioeconomic status has less of an intermediary effect on absences caused by Ozone than on absences caused by PM2.5 pollution.

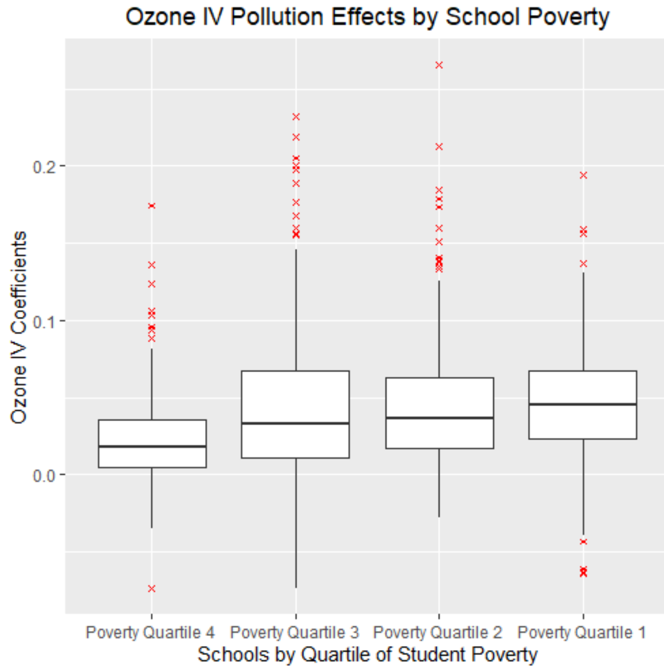


Figure 10: Ozone IV regression with every school run individually and grouped by percentage of students considered economically disadvantaged, where Quartile 4 is richest and Quartile 1 is poorest. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

# 6 Extensions and Robustness

This section contains extensions of the main IV regressions and robustness tests. The first subsection compares the effects of pollution on different days, which motivates the choice of same-day PM2.5 pollution and two-days lagged Ozone. The next subsection runs IV regressions on individual years to check for changes across time, and the final subsection runs OLS regressions on all data (including winter months) and directly controlling for snow.

## 6.1 Effects of Lagged Pollution

This section analyzes different pollution lag structures for Ozone and PM2.5 pollution. These two chemicals are measured differently (maximum 8-hour concentration for Ozone, average 24-hour concentration for PM2.5) and the process of Ozone formation is heavily influenced by heat, so Ozone is highest in the afternoon after school absences are determined by morning attendance. For this reason, we might expect different time-responses for absences caused by PM2.5 versus Ozone pollution. PM2.5 measurements are averaged over the entire day, so it also includes after-school hours, but ambient concentrations do not change as much throughout the day as Ozone concentrations do.

| PM2dot5 IV-OLS Regression of Summer Months Multiple Days Comparison | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | IV 0 Day Lag | IV 1 Day Lag | IV 2 Day Lag | IV 3 Day Lag | OLS 0 Day Lag | OLS 1 Day Lag | OLS 2 Day Lag | OLS 3 Day Lag |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
| Pollution_Concentration_Day0 | 0.044293*** | | | | 0.025061*** | | | |
| | (0.001458) | | | | (0.000918) | | | |
| Pollution_Concentration_Day1 | | 0.032376*** | | | | 0.021202*** | | |
| | | (0.001620) | | | | (0.000897) | | |
| Pollution_Concentration_Day2 | | | 0.025409*** | | | | 0.013863*** | |
| | | | (0.001679) | | | | (0.000912) | |
| Pollution_Concentration_Day3 | | | | -0.018194*** | | | | -0.000938 |
| | | | | (0.001607) | | | | (0.000901) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.600 | 0.600 | 0.600 | 0.600 | 0.618 | 0.618 | 0.618 | 0.618 |
| R2 Adj. | 0.600 | 0.600 | 0.600 | 0.599 | | | | |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | | | | |

Table 6: PM2.5 IV and OLS regression comparing each day of lagged pollution separately. Models 1-4 are IV and 5-8 are OL5, and all correspond to the full set of Day of Week, School, School × Month, and School × Year fixed effects.

Table 6 runs the full fixed effects model of PM2.5 separately for each day of pollution. For the IV models, in the first stage each day of pollution is regressed on wind from that day and three additional days of lagged wind. In both the IV and OLS models days 0 through 2 all have positive and significant coefficients, but day 3 has a negative and/or insignificant coefficient. This regression is evidence of the timescale at which pollution causes respiratory issues that lead to absences, and it appears that the short-run increase in absences caused by PM2.5 subsides after 2 days. Average PM2.5 over the past week or month is likely to affect school absences, but my econometric design can only test for short-run changes in pollution. Effects are largest on the same-day that PM2.5 increases, but there are (smaller) statistically significant absence effects on the following day and two days later. Based on this analysis of the effects by day, I run the main regressions on same-day PM2.5 concentrations, but results are similar if I include all pollution from days 0 through 2 as explanatory variables.

Table 7 runs the full fixed effects model of Ozone separately for each day of pollution. For the IV models, in the first stage each day of pollution is regressed on wind from that day and three additional days of lagged wind. In both the IV and OLS models days 0 through 2 all have positive and significant coefficients, but day 3 has a insignificant IV coefficient and smaller OLS coefficient. Same-day Ozone pollution also has much smaller effects than lagged Ozone; the same-day coefficient magnitude is three times smaller compared to two-days IV lagged coefficient and twelve times smaller than the comparable OLS coefficient. This regression is evidence of the

| Ozone IV-OLS Regression of Summer Months Multiple Days Comparison | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | IV 0 Day Lag | IV 1 Day Lag | IV 2 Day Lag | IV 3 Day Lag | OLS 0 Day Lag | OLS 1 Day Lag | OLS 2 Day Lag | OLS 3 Day Lag |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
| Pollution_Concentration_Day0 | 0.009369*** | | | | 0.001670*** | | | |
| | (0.000847) | | | | (0.000403) | | | |
| Pollution_Concentration_Day1 | | 0.026109*** | | | | 0.019799*** | | |
| | | (0.000913) | | | | (0.000394) | | |
| Pollution_Concentration_Day2 | | | 0.029830*** | | | | 0.020912*** | |
| | | | (0.001029) | | | | (0.000420) | |
| Pollution_Concentration_Day3 | | | | -0.001567 | | | | 0.011649*** |
| | | | | (0.000990) | | | | (0.000415) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.600 | 0.601 | 0.601 | 0.600 | 0.618 | 0.619 | 0.619 | 0.618 |
| R2 Adj. | 0.599 | 0.600 | 0.600 | 0.599 | | | | |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | | | | |

Table 7: Ozone IV and OLS regression comparing each day of lagged pollution separately. Models 1-4 are IV and 5-8 are OL5, and all correspond to the full set of Day of Week, School, School $\times$ Month, and School $\times$ Year fixed effects.

timescale at which Ozone causes minor respiratory issues that lead to absences, and it appears that the short-run increase in absences caused is strongest one or two days later, but then this increase subsides. As with PM2.5, average Ozone concentration over the past week or month is likely to affect school absences, but my econometric design can only test short-run changes in pollution. Based on this analysis of the effects by day, I run the main regressions on two-days lagged Ozone concentrations, but results are similar if I include all pollution from days 0 through 2 as explanatory variables.

## 6.2 Regression Results By Year

This section reports results for each school year regressed separately against PM2.5 and Ozone pollution. The models include school and school times month fixed effects but the coefficient is estimated separately by year instead of including year fixed effects. These yearly regressions are more susceptible to one-off events that affect absences, such as Hurricane Sandy in October of 2012, but they allow for coefficient comparisons across years.

Figure 11 shows the coefficients for PM2.5 estimated by year and Figure 12 does the same for Ozone. The PM2.5 estimate is generally positive but negative for 2017 to 2018, and after examining data from that year could be due to several inches of snow that NYC had on April

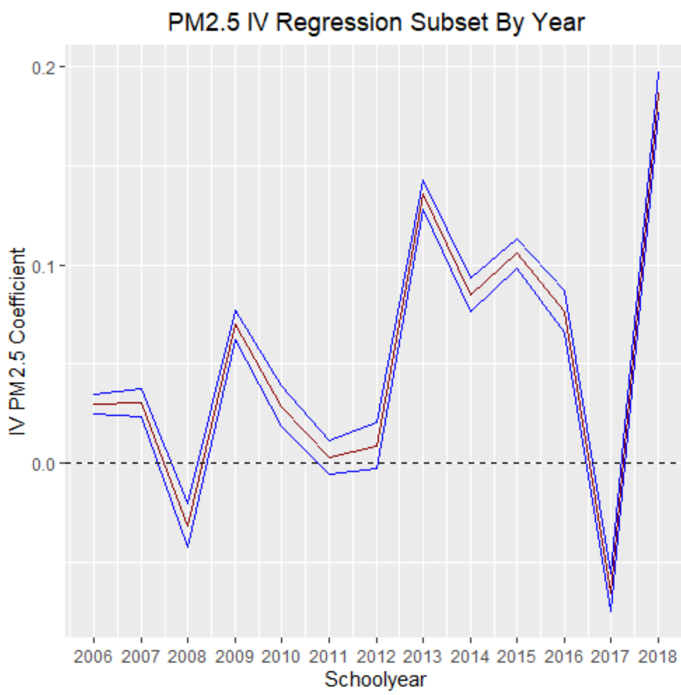


Figure 11: PM2.5 IV regression with every school year run separately using same-day PM2.5 concentration. These models include school and school times month fixed effects, and the (dark red) coefficient is reported for every year with (blue) confidence intervals.

2nd, 3rd, and 6th of 2018 - April normally does not have snowfall and this specification does not control for snow, so the large (and unrelated to pollution) increase in daily absences could bias the coefficient downwards. Similarly, the negative coefficient on Ozone in 2012 to 2013 could be related to Hurricane Sandy hitting NYC in October of 2012. Year-to-year changes in individual coefficient estimates represent yearly absence shocks that are unrelated to air pollution, which is why using multiple years of data is useful in controlling for seasonal and yearly variation.

### 6.3 OLS Results Including Snow

This subsection runs OLS regressions on the full school absence data set and directly controls for snow. Snow is included as binary indicators for {Any snow at $t$ or $t-1$, between 0.1 and 0.5 meters of snow from $t$ to $t-3$, More than 0.5 meters of snow from $t$ to $t-3$}. Table 8 shows the panel of fixed effects and coefficients on the snow variables. The PM2.5 OLS coefficients in Model 1 and 2 are similar to the corresponding coefficients in the main results, but adding school times year and school times month fixed effects reduces the coefficient and makes it negative in Model 5. The issue is that snow is measured with error and has much larger effects on absences than pollution. Using coefficients from Model 5, any snow that day or the day before causes 1.86% more absences, which is added to either 2.53% or 6.32% if there is more than 0.1 or 0.5 meters of accumulated snow.

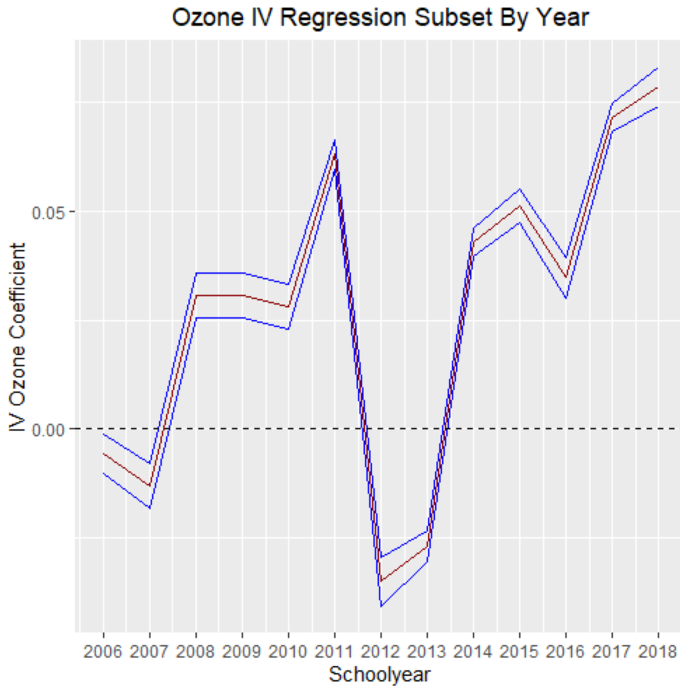


Figure 12: Ozone IV regression with every school year run separately, using two days lagged Ozone concentration. These models include school and school times month fixed effects, and the coefficient (dark red) is reported for every year with (blue) confidence intervals.

A day of heavy snowfall can outweigh the average absence effects of a month of high-pollution, so month and year fixed effects end up fitting average snowfall instead of average pollution. This is an econometric and measurement issue, though, and pollution still causes absences during winter months.

Table 9 shows the Ozone OLS regression using all months of data and directly controlling for snow. In contrast to PM2.5, the Ozone results are remarkably similar to the main Ozone OLS results run on months without snow. Specifically, all specifications for Ozone with snow are positive and significant and the Model 5 coefficient is only twenty percent larger in the comparable regression of non-winter months. Ozone formation depends on heat and is lowest during winter months, so it is not too surprising that this regression is able to distinguish snow absences from Ozone-caused absences.

# 7 Conclusion

This paper estimates the causal effects of air pollution encountered in day-to-day life in the United States. Previous papers analyzing the absences in the US took place either at smaller scale or using less fine-grained data. I analyze 1669 schools in NYC from 2006 to 2019 and using daily

| PM2.5 OLS Full Snow Regression | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration.x | 0.043250*** | 0.017510*** | 0.001683** | 0.002716*** | -0.016305*** |
| | (0.000906) | (0.000644) | (0.000643) | (0.000664) | (0.000667) |
| Snow_Bin_nonzero | 1.732690*** | 1.894971*** | 1.887067*** | 1.899375*** | 1.861915*** |
| | (0.021147) | (0.015784) | (0.015681) | (0.015587) | (0.015503) |
| Snow_Bin_above100 | 2.180877*** | 2.310145*** | 2.398036*** | 2.447013*** | 2.538757*** |
| | (0.038192) | (0.031962) | (0.031970) | (0.032061) | (0.032045) |
| Snow_Bin_above500 | 5.097064*** | 6.105872*** | 5.962005*** | 6.437452*** | 6.321915*** |
| | (0.242387) | (0.221243) | (0.220250) | (0.221847) | (0.220852) |
| Num.Obs. | 3499057 | 3499057 | 3499057 | 3499057 | 3499057 |
| R2 | 0.007 | 0.554 | 0.604 | 0.555 | 0.605 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 8: PM2.5 OLS regression using all months of absence data and directly controlling for snow.

| Ozone OLS Full Snow Regression | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day2 | 0.032202*** | 0.049033*** | 0.012475*** | 0.052335*** | 0.017302*** |
| | (0.000404) | (0.000292) | (0.000343) | (0.000293) | (0.000344) |
| Snow_Bin_nonzero | 1.972587*** | 2.204839*** | 1.903836*** | 2.220451*** | 1.894420*** |
| | (0.021068) | (0.015576) | (0.015543) | (0.015376) | (0.015383) |
| Snow_Bin_above100 | 2.331500*** | 2.489993*** | 2.461915*** | 2.642160*** | 2.602623*** |
| | (0.038067) | (0.031809) | (0.031937) | (0.031923) | (0.032028) |
| Snow_Bin_above500 | 5.239541*** | 6.203282*** | 6.091221*** | 6.543811*** | 6.401542*** |
| | (0.240796) | (0.219629) | (0.220057) | (0.220297) | (0.220562) |
| Num.Obs. | 3490118 | 3490118 | 3490118 | 3490118 | 3490118 |
| R2 | 0.024 | 0.532 | 0.580 | 0.536 | 0.584 |
| R2 Adj. | 0.024 | 0.531 | 0.580 | 0.535 | 0.583 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 9: Ozone OLS regression using all months of absence data and directly controlling for snow.

attendance data from millions of students find absences caused by PM2.5 as well as Ozone. Because I can analyze responses to wind-induced pollution shocks across different days, I show that PM2.5 pollution causes absences the same day it increases but Ozone has largest effects on absences one or two days later. I separately analyze schools by type and find, consistent with the literature, that elementary and middle schools experience larger effects from PM2.5 pollution than do high schools. For Ozone, by contrast, the largest absence effect is on high schools, followed by elementary and then middle schools.

Although the magnitudes of the estimated absences caused by PM2.5 and Ozone are not especially large, they are robust to multiple specifications, are estimated for all NYC students, and can be interpreted as causal. This means that, across NYC with an average yearly enrollment of 960000, an additional $\mu$g/$m^3$ of PM2.5 pollution causes 423 students to stay home sick that day and an extra PPB of Ozone causes 286 students to be sick from school two days later. Going from the median PM2.5 day to the ninetieth percentile day causes 2150 students to be absent that day and going from the median Ozone day to the ninetieth percentile day causes 5654 students to be absent two days later. The decline in average daily PM2.5 pollution of 5 $\mu$g/$m^3$ from 2006 to 2019 caused a reduction in school absences of 2116 students per day, or 381024 total absences over the entire school year. The state of New York funds individual schools on the basis of average daily attendance, with a school losing around \$50 for every day that a student is absent, so this reduction in school absences increased annual NYC school funding by approximately \$19 million dollars in 2019. My identification strategy compares school days in the same month to minimize bias from seasonal trends, but if some of the difference in absences between months is due to changes in pollution across months, then these are likely to be lower bounds. For example, if school absences are high in summer months not because of end-of-year reasons but only because Ozone concentrations are much higher, then including School $\times$ Month fixed effects biases the results towards zero. Assuming that all differences in absences across months are due to pollution corresponds to Model 4 in my results, and the Ozone coefficients are approximately 3 times larger than the corresponding Model 5 coefficients.

These pollutants were measured in NYC from 2006 to 2019 with concentrations below the respective federal limits on approximately 99.3% of days for PM2.5 and 98.3% of days for Ozone. The measured harms were larger for days above those thresholds, but there were absences caused by pollution below those thresholds as well. I was able to identify effects of pollution at these levels because of daily school absence data representing more than twelve million student years. Reporting absences by school means that the absence data is not sensitive and is publicly available in multiple settings. Student-level data allows for the tracking of students exposed to high pollution and the long-run effects, but that data is harder to access and typically can't be released for replication. In contrast, the NYC school-level data is available online and anyone can replicate this paper with different econometric specifications. Other variables of interest to connect to absences in future work include influenza rates, asthma rates, or seasonal allergies.

These absence results also have implications for the rest of the country. PM2.5 concentrations declined significantly over this period, but Ozone remained high and these pollutant concentrations are common to the US and other countries in the modern era. A back-of-the-envelope calculation suggests that: counties containing more than 140 million people have higher yearly average PM2.5 pollution than NYC; counties containing more than 103 million people have higher 98th percentile daily PM2.5 compared to NYC; counties containing more than 76 million people have higher Ozone concentrations.[14] While those other counties might have different distributions of polluted days, it suggests that schools in those counties would have similar numbers of absences due to pollution.

## References


[1] Claudia W Allen, Sharon Diamond-Myrsten, and Lisa K Rollins. "School absenteeism in children and adolescents". In: *American family physician* 98.12 (2018), pp. 738–744.

[2] H Ross Anderson. "Air pollution and mortality: A history". In: *Atmospheric Environment* 43.1 (2009), pp. 142–152.

[3] Rebecca H Bitsko et al. "Epidemiology and impact of health care provider–diagnosed anxiety and depression among US children". In: *Journal of developmental and behavioral pediatrics: JDBP* 39.5 (2018), p. 395.

[4] Jerome B Blechman. "Surface wind directions associated with snowfall in upstate New York". In: *National Weather Digest* 25.3/4 (2002), pp. 19–27.

[5] Malvina Bondy, Sefi Roth, and Lutz Sager. "Crime is in the air: The contemporaneous relationship between air pollution and crime". In: *Journal of the Association of Environmental and Resource Economists* 7.3 (2020), pp. 555–585.

[6] Hedy N Chang and Mariajosé Romero. "Present, Engaged, and Accounted for: The Critical Importance of Addressing Chronic Absence in the Early Grades. Report." In: *National Center for Children in Poverty* (2008).

[7] Lei Chen et al. "Elementary school absenteeism and air pollution". In: *Inhalation Toxicology* 12.11 (2000), pp. 997–1016.

[8] Siyu Chen, Chongshan Guo, and Xinfei Huang. "Air pollution, student health, and school absences: Evidence from China". In: *Journal of Environmental Economics and Management* 92 (2018), pp. 465–497.

[9] Janet Currie, Eric A Hanushek, et al. "Does pollution increase school absences?" In: *The Review of Economics and Statistics* 91.4 (2009), pp. 682–694.


[14] Calculation done using the EPA's Air Quality Statistics by County data for 2021, Link Here, with 2010 Population. 'Wtd AM' is weighted annual mean concentration, (24-hr) is the 98th percentile 24-hour PM2.5 concentration, and 'O3' is the fourth daily maximum 8-hour concentration. NYC is calculated as the mean of New York County (Manhattan), Kings County (Brooklyn), Bronx County (The Bronx), Richmond County (Staten Island), and Queens County (Queens).

[10] Janet Currie and Matthew Neidell. "Air pollution and infant health: what can we learn from California's recent experience?" In: *The Quarterly Journal of Economics* 120.3 (2005), pp. 1003–1030.

[11] Janet Currie and Reed Walker. "Traffic congestion and infant health: Evidence from E-ZPass". In: *American Economic Journal: Applied Economics* 3.1 (2011), pp. 65–90.

[12] Janet Currie and Reed Walker. "What do economists have to say about the Clean Air Act 50 years after the establishment of the Environmental Protection Agency?" In: *Journal of Economic Perspectives* 33.4 (2019), pp. 3–26.

[13] Tatyana Deryugina et al. "The mortality and medical costs of air pollution: Evidence from changes in wind direction". In: *American Economic Review* 109.12 (2019), pp. 4178–4219.

[14] Qian Di et al. "Air pollution and mortality in the Medicare population". In: *New England Journal of Medicine* 376.26 (2017), pp. 2513–2522.

[15] Douglas W Dockery et al. "An association between air pollution and mortality in six US cities". In: *New England journal of medicine* 329.24 (1993), pp. 1753–1759.

[16] Avraham Ebenstein, Victor Lavy, and Sefi Roth. "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution". In: *American Economic Journal: Applied Economics* 8.4 (2016), pp. 36–65.

[17] Frank D Gilliland et al. "The effects of ambient air pollution on school absenteeism due to respiratory illnesses". In: *Epidemiology* (2001), pp. 43–54.

[18] Joshua Goodman. *Flaking out: Student absences and snow days as disruptions of instructional time*. Tech. rep. National Bureau of Economic Research, 2014.

[19] Nicholas M Hales et al. "A quasi-experimental analysis of elementary school absences and fine particulate air pollution". In: *Medicine* 95.9 (2016).

[20] Jennifer A Heissel, Claudia Persico, and David Simon. "Does pollution drive achievement? The effect of traffic pollution on academic performance". In: *Journal of Human Resources* 57.3 (2022), pp. 747–776.

[21] Evan Herrnstadt et al. "Air pollution and criminal activity: Microgeographic evidence from Chicago". In: *American Economic Journal: Applied Economics* 13.4 (2021), pp. 70–100.

[22] Christopher R Knittel, Douglas L Miller, and Nicholas J Sanders. "Caution, drivers! Children present: Traffic, pollution, and infant health". In: *Review of Economics and Statistics* 98.2 (2016), pp. 350–366.

[23] Johanna Lepeule et al. "Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009". In: *Environmental health perspectives* 120.7 (2012), pp. 965–970.

[24] Jing Liu, Monica Lee, and Seth Gershenson. "The short-and long-run impacts of secondary school absences". In: *Journal of Public Economics* 199 (2021), p. 104441.

[25] Dave E Marcotte. “Something in the air? Air quality and children’s educational outcomes”. In: *Economics of Education Review* 56 (2017), pp. 141–151.

[26] Michael R Ransom and C Arden Pope III. “Elementary school absences and PM10 pollution in Utah Valley”. In: *Environmental research* 58.1-2 (1992), pp. 204–219.

[27] Jean M Twenge et al. “Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017.” In: *Journal of abnormal psychology* 128.3 (2019), p. 185.

[28] W Reed Walker. “The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce”. In: *The Quarterly journal of economics* 128.4 (2013), pp. 1787–1835.

[29] Xin Zhang, Xi Chen, and Xiaobo Zhang. “The impact of exposure to air pollution on cognitive performance”. In: *Proceedings of the National Academy of Sciences* 115.37 (2018), pp. 9193–9197.

[30] Eric Yongchen Zou. “Unwatched pollution: The effect of intermittent monitoring on air quality”. In: *American Economic Review* 111.7 (2021), pp. 2101–26.

# 8 Appendix

## 8.1 OLS Results Other Pollutants

CO OLS Regression of Summer Months

| | No FE | School FE | School + Month | School + Year | School + Month + Year |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day0 | 0.138121*** | -0.425931*** | 0.746628*** | -1.154102*** | 0.279166*** |
| | (0.021638) | (0.016154) | (0.015274) | (0.017718) | (0.016588) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.010 | 0.506 | 0.596 | 0.511 | 0.600 |
| R2 Adj. | 0.010 | 0.506 | 0.596 | 0.511 | 0.600 |

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 10: OLS Regression of same-day CO on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School $\times$ Month; Model 4 adds School and School $\times$ Year; Model 5 corresponds to the full set of School, School $\times$ Month, and School $\times$ Year fixed effects.

| SO2 OLS Regression of Summer Months | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day0 | 0.050014*** | -0.002443** | 0.035953*** | -0.056939*** | 0.001350 |
| | (0.001172) | (0.000777) | (0.000761) | (0.001007) | (0.000984) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.012 | 0.506 | 0.597 | 0.511 | 0.600 |
| R2 Adj. | 0.012 | 0.505 | 0.596 | 0.511 | 0.600 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 11: OLS Regression of same-day SO2 on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

| NO2 OLS Regression of Summer Months | | | | | |
|---|---|---|---|---|---|
| | No FE | School FE | School + Month | School + Year | School + Month + Year |
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Pollution_Concentration_Day0 | 0.009084*** | -0.001563*** | 0.005677*** | -0.010471*** | -0.002705*** |
| | (0.000483) | (0.000321) | (0.000301) | (0.000329) | (0.000308) |
| Num.Obs. | 2103540 | 2103540 | 2103540 | 2103540 | 2103540 |
| R2 | 0.010 | 0.506 | 0.596 | 0.510 | 0.600 |
| R2 Adj. | 0.010 | 0.505 | 0.595 | 0.510 | 0.600 |
| + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001 | | | | | |

Table 12: OLS Regression of same-day NO2 on daily school absences. All models have day of week fixed effects; Model 1 has no other fixed effects; Model 2 adds School fixed effects; Model 3 adds School and School × Month; Model 4 adds School and School × Year; Model 5 corresponds to the full set of School, School × Month, and School × Year fixed effects.

### 8.2 Additional PM2.5 OLS Regressions

### 8.3 Additional Ozone OLS Regressions

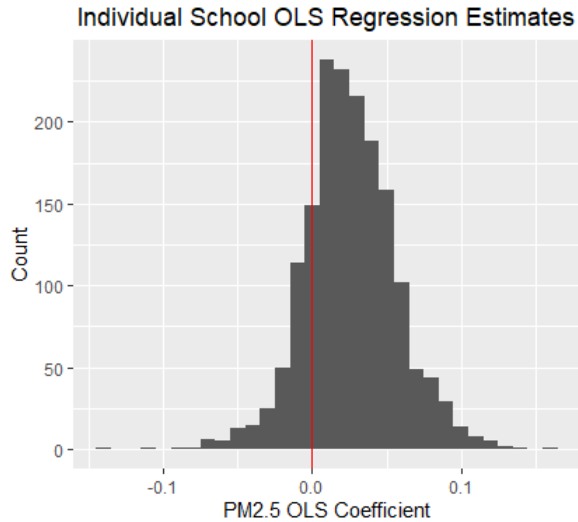


Figure 13: PM2.5 OLS regression with every school run individually. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

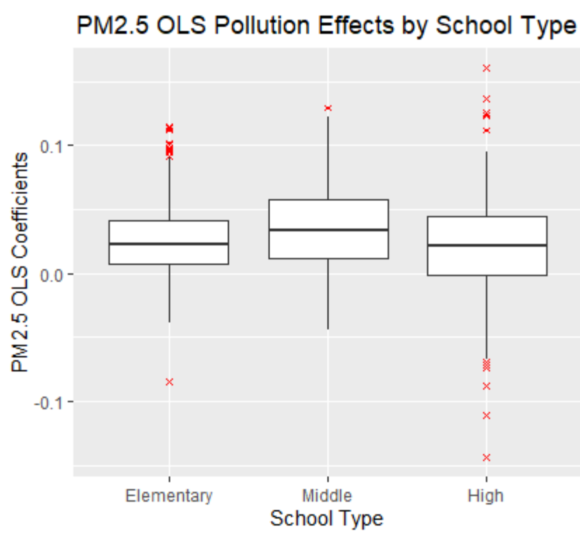


Figure 14: PM2.5 OLS regression with every school run individually and grouped into school type. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

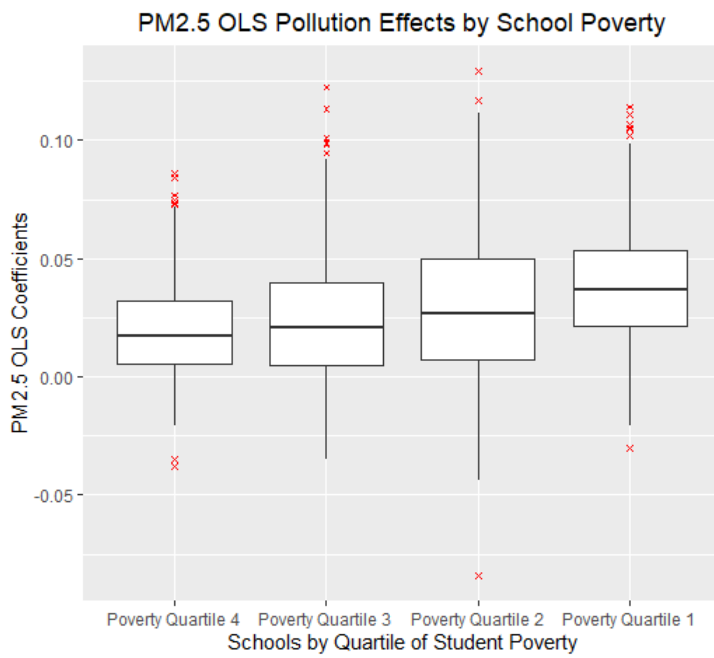


Figure 15: PM2.5 OLS regression with every school run individually and grouped by percentage of students considered economically disadvantaged, where Quartile 4 is richest and Quartile 1 is poorest. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

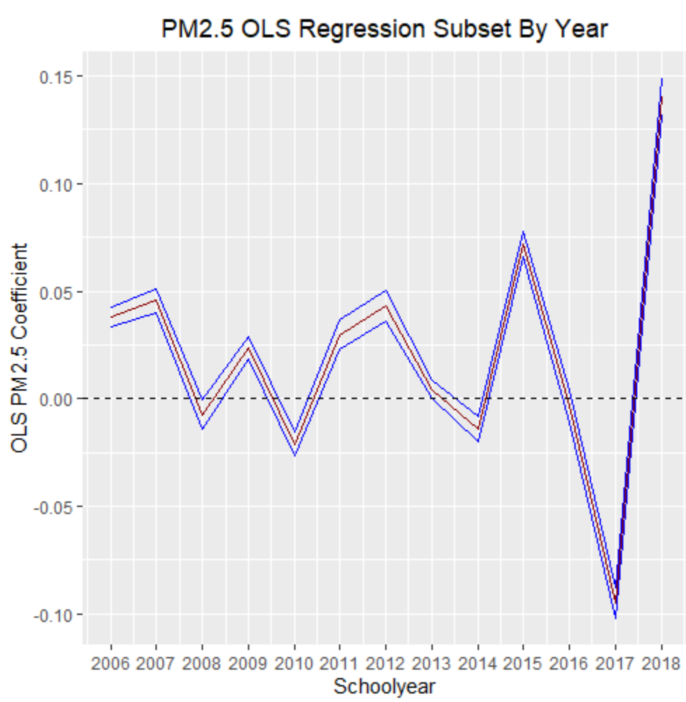


Figure 16: PM2.5 OLS regression with every school year run separately using same-day PM2.5 concentration. These models include school and school times month fixed effects, and the (dark red) coefficient is reported for every year with (blue) confidence intervals.

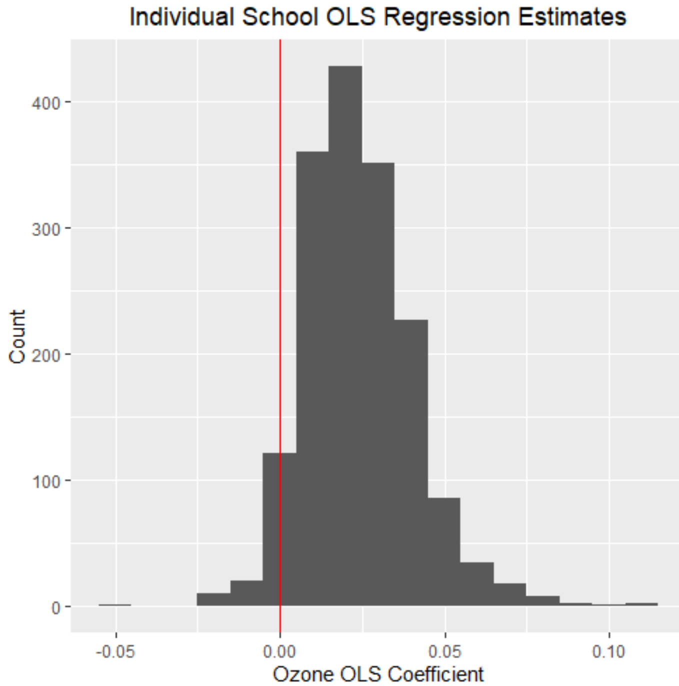


Figure 17: Ozone OLS regression with every school run individually. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

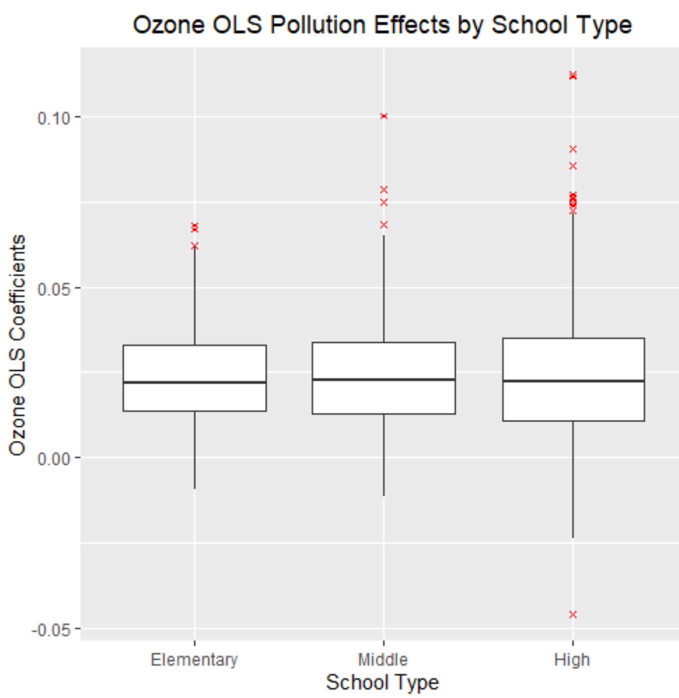


Figure 18: Ozone OLS regression with every school run individually and grouped into school type. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

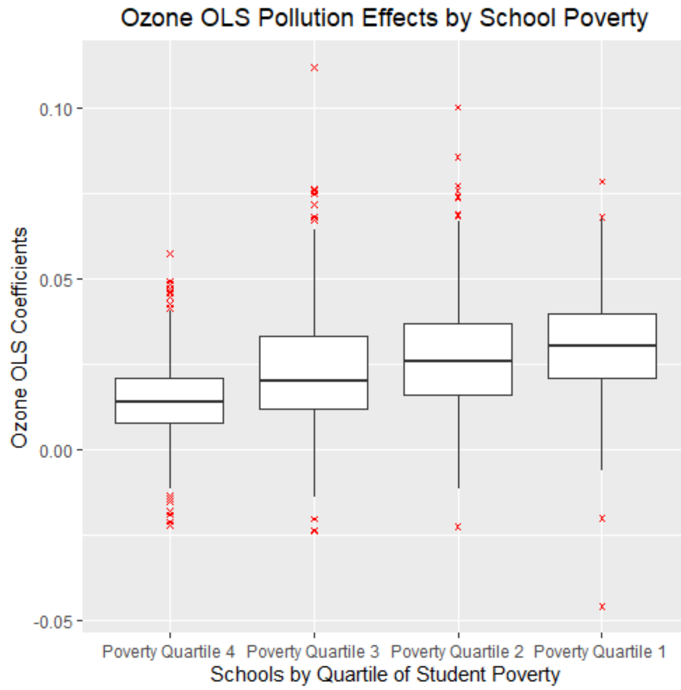


Figure 19: Ozone OLS regression with every school run individually and grouped by percentage of students considered economically disadvantaged, where Quartile 4 is richest and Quartile 1 is poorest. These models include day of week, month, and year fixed effects, and the coefficient is reported in bins for every school.

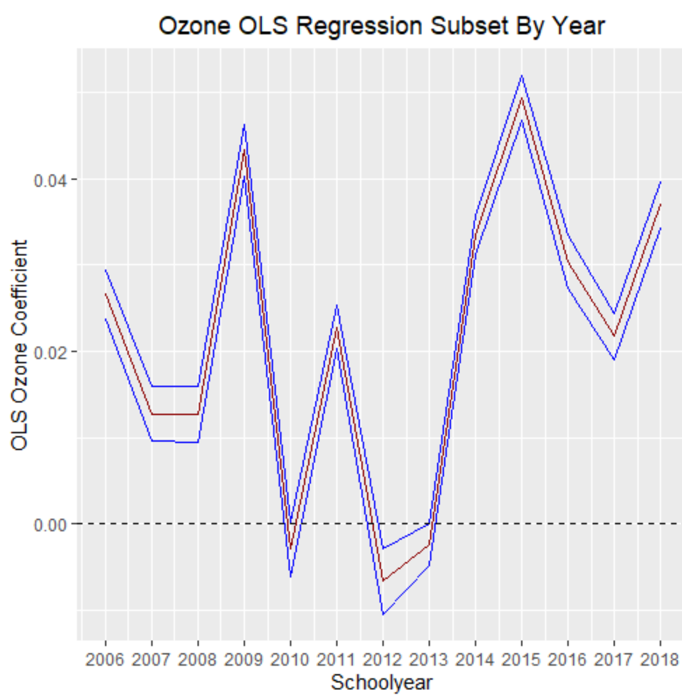


Figure 20: Ozone OLS regression with every school year run separately using two-day lagged Ozone concentration. These models include school and school times month fixed effects, and the (dark red) coefficient is reported for every year with (blue) confidence intervals.