Air Pollution, Student Health, and School Absences: Evidence from China*

Siyu Chen†, Chongshan Guo‡, Xinfei Huang§

Abstract: Little is known about the impact of air pollution on school children in developing countries. This paper aims to fill this gap by quantifying the causal effect of air pollution on the health status and the school attendance of the Chinese students. We relate the arguably exogenous daily variation in air pollution-instrumented by the occurrence of temperature inversion-with student illnesses and absences from more than 3,000 schools in Guangzhou City. We find a sizable deleterious effect of air pollution on school attendance through the health channel. The impact persists for at least four days and displays a monotonically increasing pattern. Notably, this harmful effect is non-negligible even when pollution levels are below the official standard for air quality in China, suggesting that the current ambient air quality standards in China are not low enough to protect students.

Keywords: Air pollution, Student health, School absences, China, Air quality standards

JEL classification: I18, I24, Q51, Q53

---

*We are deeply indebted to Jessica Pan, Wenlan Qian, Sumit Agarwal, Alice Chen, Olivier Deschenes, Sebastian Galiani, John Ham, Rema Hanan, Aloysius Siow, Hongjia Zhu, Klaus Zimmermann, and seminar participants at National University of Singapore. We would also like to give special thanks to the editor and two anonymous reviewers for their constructive comments. Chen also thanks the Musim Mas Dissertation Award for financial support. Any remaining errors are our own.

†**Chen:** Department of Economics and Center for Population and Family Research, National University of Singapore, 1 Arts Link, Singapore, 117570. Email: chensynus1220@gmail.com

‡**Guo:** Guangzhou Center for Disease Control and Prevention, Dexing 1st Cross Street, Guangzhou, China, 510000.

§**Huang:** International School of Business and Finance, Sun Yat-sen University, 135 Xingang West Road, Guangzhou, China, 510275. Huang is the corresponding author. E-mail: huangxf3@mail.sysu.edu.cn

1. Introduction

Many developing countries face severe air pollution problems. For instance, the levels of air pollution in India and China are six to eight times as high as those in the U.S., as documented by Greenstone and Hanna (2014). Ample evidence shows that air pollution poses a significant threat to human respiratory and cardiovascular systems.[1] In 2012, approximately 3.7 million lives lost were attributed to air pollution globally, and 88% of this death burden is borne by developing countries (WHO, 2014).

Apart from the direct health consequences, a growing body of literature documents the social costs of air pollution. One strand of the literature concerns school children, which represent one of the high-risk groups for pollution exposure.[2] Currie et al. (2009) examined the impact of air pollution on the school absences in Texas (U.S.) and find that a high level of carbon monoxide (CO) significantly increases the absence incidences. Liu and Salvo (2017) document that fine particulate matter ($PM_{2.5}$) reduces school attendance of international students who come from rich nations and temporarily reside in China.[3] Ham et al. (2011) and Lavy et al. (2016) investigate the relationship between air pollution and the educational outcomes of students. They suggest that poor air quality negatively influences the performance of students on tests. These findings highlight the additional benefits that can be gained from air pollution abatement.

However, most of the aforementioned studies are derived from developed countries, such as the U.S. and Israel, rather than from developing countries, where pollution levels are relatively higher. The estimates based on developed countries have limited

---

[1]Some examples that study this association include Schwartz et al. (1994), Levy et al. (2001) and Neidell (2009). These adverse health effects raise the infant mortality rates (Chay and Greenstone, 2003ab; Currie and Neidell, 2005; Greenstone and Hanna, 2014; Tanaka, 2015; Arceo et al., 2016), increase the respiratory-related hospital visits (Neidell, 2004; Moretti and Neidell, 2011; Jans et al., 2014), and even cause premature deaths (Chen et al., 2013; He et al., 2016; Ebenstein et al., 2017). Besides the physical problems, air pollution can also affect mental health (Zhang et al., 2017; Chen et al., 2018).

[2]School children are particularly vulnerable to air pollution, because their bodies are at the stage of development, and their metabolic rates are high. The higher the metabolic rate, the larger the amount of air breathed. Hence, more air pollutants are inhaled into the body (Currie et al., 2009; WHO, 2005).

[3]They mainly focus on foreign children from high-income expatriate families temporarily living in China, a group of students who are supposed to be quite different from local students in terms of avoidance behavior and the preference for absenteeism.

external validity for developing ones, given the potential nonlinear effect of air pollution on health and the different avoidance behaviors in the developing world (Arceo et al., 2016). Governments in developing countries have long realized that air pollution constitutes one of the key factors that worsen the national health and constrain sustainable development. Nevertheless, the lack of policy guidance and the fear that the economic costs associated with environmental regulations may outweigh the benefits gained, deter these governments from undertaking large measures (Ebenstein et al., 2015). Thus, evidence that offers insight into both the tangible and intangible costs of air pollution is important for policy-makers in order to efficiently plan the future air quality management.

In this paper, we examine the causal effects of air pollution on the school attendance of students in China, as well as their health status. This question is highly relevant because an extensive body of research is concerned with child health and school participation as vital components of human capital attainment (Currie, 2009; Currie et al., 2009). If school children are more prone to get sick due to high levels of air pollution, and hence become less engaged in school activities due to frequent absences, it is plausible that their educational attainment will be negatively affected. Given the strong relationship between human capital accumulation and future earnings, they are more likely to be at a disadvantage for well-paid jobs. This is particularly the case in developing countries, where higher education is regarded as one of the most important ways in which people acquire high labor incomes.

Related studies for developing countries in the economics literature are scarce, at least because of data limitation. The access to disaggregated data on student attendance is challenging in developing countries, and such data are often absent or suffer from poor quality. This is because many governments and schools do not maintain these data in permanent archives, leading to short or discontinuous observation periods with less variation to explore. Even when long-term data are available, they may only cover a sample of schools with substantial selections, making the analysis less representative.

To study this relationship, we explore a unique administrative dataset drawn from the Guangzhou Center for Disease Control and Prevention (CDC). The dataset not only provides the temporal granularity, the daily records of student illnesses and absences from 2013 to 2015, but also covers almost all the schools in Guangzhou City, China, including the kindergartens, elementary, middle, and high schools. Furthermore, it contains information on symptoms, hospital diagnoses and reasons for absences. The detailed information of the data permits fine-grained analysis of how air pollution influences student absences on a daily basis, as well as enable us to explore the mechanisms that drive this adverse effect. One caveat of this detest is that the CDC hides the individual identities (names and IDs of the students) when communicating the data for research to ensure the privacy of the students. Therefore, we aggregate the data into the school-grade-date level in the analysis.

We measure air pollution as the levels of Air Quality Index, a composite measure that captures the binding pollutant on any certain date.[4] We also provide suggestive evidence on the effects of fine particulate matter ($PM_{2.5}$), ozone, sulfur dioxide (SO_2), and carbon monoxide (CO), which have been extensively studied in the literature showing strong associations with health.[5] However, we are cautious in interpreting these results, as our research design is not able to perfectly isolate the effect of one pollutant from the others.[6] The estimates of individual pollutants work in the same

[4]Different countries have their own algorithms to calculate the air quality indices. In Mainland China, AQI is based on six atmospheric pollutants, namely, sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), carbon monoxide (CO), and ozone. The maximum AQI across pollutants is reported to the public. As summarized in Table A.I, the correlations between AQI and the other pollutants are quite high, often exceeding 0.6, indicating that AQI is a good proxy variable for air pollution. For the purpose of robustness, we also construct a single-pollution index to measure air pollution levels using the principal components method. Our regression results are similar by using the two measurements.

[5]The other two primary pollutants, PM_{10} and NO_2, are not separately examined in our analysis. The reasons are as follows: 1) The correlation between $PM_{2.5}$ and PM_{10} reaches 0.94 during the sample period, as shown in Table A.I. Hence, it is empirically challenging to split the effects of these two pollutants from one another. Besides, numerous new studies parade that $PM_{2.5}$ is more harmful than PM_{10}, especially in urbanized areas (Kappos et al., 2004; Peters et al., 2006). Thus, we focus on $PM_{2.5}$ in our paper instead of PM_{10}; 2) NO_2 pollutant constitutes a part of the NO_x family of pollutants, which forms the particulate matter (PM) and also the ozone. Thus, it is difficult to isolate the effects of NO_2 from those of $PM_{2.5}$ and ozone.

[6]Research in this area is faced with the challenge of disentangling the effect of one particular pollution from the pollutant mixture. Some settings allow for a clean route around this problem. Lavaine and Neidell (2017) exploits oil refinery strikes, which generate an exogenous variation in SO_2 concentrations, while the other pollutants are unchanged. Ransom and Pope (2013) take advantage of the mill closure, which is the main source of particulate matter in Utah Valley, to isolate the effect of PM_{10} from CO. Otherwise, studies typically focus on a single pollutant (a subset of pollutants) of relevance (Currie et al., 2009; Arceo et al., 2016; Schlenker and Walker., 2016), or use a

direction as the results for AQI. Therefore, we use AQI to measure the combined effect of all the air pollutants.

In the empirical analysis, we first employ a fixed-effect technique, controlling for a series of meteorological factors in a flexible fashion and a rich set of fixed effects netting out any unobserved determinants of student absences (health) that also covary with ambient pollution levels. The corresponding results indicate a positive relationship between air pollution and the incidences of respiratory illnesses and absences among students. However, the magnitude of the coefficients tends to be moderate. This is possibly because our OLS estimates suffer from attenuation bias pertaining to classical measurement error in the pollution variable (Griliches and Hausman, 1986).[7] Additionally, there may exist omitted-variables bias despite the comprehensive controls, as discussed in detail below.

Therefore, to reinforce our causal interpretation, we adopt an instrumental-variables approach, which helps to alleviate the omitted-variables bias and the attenuation bias from classical measurement error. Specifically, we instrument the air pollution by using the occurrence of temperature inversion. This occurs when air temperature abnormally increases with height, resulting in a layer of cool air overlain by warmer air. As a meteorological phenomenon, it does not affect absenteeism (health) directly, but it may trigger the accumulation of air pollution by reducing the upward movement of air from the layers below (Jacobson, 2002; Secretaría del Medio Ambiente, 2005). It has been proven to have a strong impact on pollution levels in many studies (Ransom and Pope, 2013; Jans et al., 2014; Arceo et al., 2016; Hicks et al., 2016; Sager, 2016; Fu et al., 2017; Liu and Salvo, 2017; Chen et al., 2018). Following Arceo et al. (2016), we define the inversion dummy as the temperature of the second layer (320 m) being higher than

composite measure of air pollution (Chang et al., 2018).

[7]We use ambient pollutant concentrations from nearby stations to construct pollution levels for each school. There are 11 public pollution monitors across 12 districts in Guangzhou City during our sample period. Pollution levels may be measured with errors due to sporadic monitoring sites. The inclusion of fixed effects may exacerbate the classical measurement error, making our OLS estimates biased downward. Moretti and Neidell (2011) provide evidence and insightful discussion on the problems concerning measurement error in pollution when studying the link between ozone and health.

that of the first layer (110 m). Consistent with the existing literature, it indicates a strong relevance with pollution levels.

In our preferred IV estimations, we find that air pollution statistically significantly worsens student health and further increases school absences. In terms of the magnitude, a 10-unit increase in AQI shifts up the daily respiratory illness rate among students by approximately 0.00582 percentage points, equivalent to 3.31% of the mean level. Apart from the direct health consequences, we observe that a 10-unit increase in AQI raises the total absence rate by about 0.00417 percentage points, which amounts to 2.31% of the daily mean.[8] Notably, these adverse effects are predominately driven by respiratory-related absences, implying that air pollution influences absenteeism mainly through the health channel rather than by inducing avoidance behaviors.

The impact of air pollution is also found to be cumulative. This study reveals that air pollution on previous days considerably influences respiratory illnesses and school absences on the current period, by gradually adding lag terms of pollution measurements into the regression model. Furthermore, the impact is shown to persist for at least four days. We additionally advance evidence that this relationship is nonlinear and displays a monotonically increasing pattern. The higher the pollution levels, the larger the adverse effect is. More importantly, air pollutants exert toxic effects at levels lower than the official regulatory thresholds. This suggests that the current standards for air quality in China may not be low enough to protect students. We also conduct a heterogeneity analysis and find that the impact of air pollution varies across school grades, but not across school quality.

This paper makes three contributions to the literature. First, to the best of our knowledge, this is the first estimate on the causal effect of air pollution on absenteeism

[8]Equivalently, a one standard deviation of daily AQI levels increases the respiratory rate by 10.04% and the total absence rate by 7.01%.

for students from a developing country.[9] In the economics literature, one innovative work by Currie et al. (2009) rigorously examined this linkage in the City of Texas (U.S.), but whether air pollution influences student attendance in developing countries, and to what extent, remains an open question.[10] They show that high levels of CO lead to a significant decrease in school attendance, but no observable consequences of PM_{10} and ozone are documented. However, our study shows that both $PM_{2.5}$ and ozone exert pronounced adverse impacts on school participation, providing a contrast to the previous findings. Moreover, Currie et al. (2009) explore the cumulative effect of air pollution on absenteeism over a six-week attendance period due to limited data. In this study, we provide supplementary evidence. In addition to the impact of cumulative pollution levels over several days, we also investigate the linkage between day-to-day variation in air pollution and school absences and point out unappreciated costs induced by short-term pollution exposure. Furthermore, we document the mechanism that drives the adverse effect on school attendance. Most absences are attributable to pollution-related illnesses, suggesting that air pollution affects attendance mainly through the health channel.

Second, this paper adds to the ample literature linking air pollution and health by providing the dose-response function for the population of students at higher concentrations of pollutants.[11] School children represent one of the most at-risk groups for pollution exposure (WHO, 2005). The disturbance of their health conditions has been linked to negative academic, socio-emotional and socio-economic outcomes (Currie et al., 2009; Gottfried, 2014). Thus, their health is a central question that

---

[9]Many existing studies for developing countries are in epidemiological literature. They mainly exploit differences in pollution over time while controls for weather and time fixed effects. We would caution against interpreting them as causal analyses since their estimates may be confounded by local economic conditions and unobserved social-demographic determinants of pollution and attendance, such as residential sorting whereby poor families stay in communities with bad air quality while their children are more prone to be absent from schools.

[10]Other nice studies include Ransom and Pope (2013), and Liu and Salvo (2017). Ransom and Pope (2013) exploit the exogenous variation in pollution levels generated by mill closure in Utah Valley (U.S.) in 1986 and 1987 to study the impact of PM_{10} and CO on school absences for primary school pupils. They find a significantly positive effect of PM_{10} but their preferred IV estimates of the effect of CO are negative and implausible. Liu and Salvo (2017) study foreign children from high-income expatriate families temporarily living in China and find that $PM_{2.5}$ increases their absences.

[11]Our sample includes students aged from 2 to19 years old.

concerns all the countries in the world. As stated above, most existing studies are related to developed countries, where pollution levels are relatively low. Guangzhou, as a major manufacturing city in China, continues to be exposed to high levels of air pollution with large variations in these levels. This useful setting serves as an opportunity for examining the nonlinear effect of air pollution on health. Moreover, this study employs a more sensitive proxy variable for student health than the measurements commonly used in other studies, such as hospitalization and emergency room visits (Neidell 2004; Neidell 2009). For example, some illnesses, such as runny noses and sore throats, are not severe enough for children to be hospitalized or to be absent from school. The sample in the present study accounts for approximately 21% of sick students that did not visit hospitals. Thus, if the proportion of ill students who consult doctors is positively correlated with pollution levels, using hospital visits may overstate the impact to be evaluated.

Third, this paper delivers strong policy implications. Our assessment suggests that pollution reductions will significantly improve student health and school participation. It demonstrates the potential benefits that will be achieved by adopting air quality regulations in developing countries. Besides, our results in nonlinear models reveal an underlying problem regarding the current standards for air quality in China. Precisely, school children are not exempted from harm even on days when pollution levels are below official standards. The Government should be urged to take this finding into consideration when setting future air quality guideline values. Moreover, our suggestive evidence shows that apart from $PM_{2.5}$, which is commonly known by Chinese people due to wide media coverage, other less detectable air pollutants, such as zone, also have pronounced harmful effects.[12] This can serve as a warning to the public, the media, and the government to take these conspicuous gases in the atmosphere pollutants more seriously.

[12]Panels (a) and (b) in Figure 1 show the weekly search frequency for $PM_{2.5}$, ozone, SO_2, and CO in 2015. These panels manifest that people in China, not limited to Guangzhou City, search far less for air pollutants other than $PM_{2.5}$ on the internet.

The rest of this paper is organized as follows: Section 2 introduces the background. Section 3 describes the data sources and summary statistics. Section 4 discusses our empirical model and identification strategy. We present our results in Section 5, along with the placebo tests, robustness check, and heterogeneous analyses. Section 6 discusses the implications of our results and concludes the paper.

1. Background on Air Pollution and Health

China reports daily AQI for most cities. It is a composite measure of air pollution that ranks air quality based on the associated health risks as a means to facilitate comprehensibility by the public. The higher the values, the poorer the air quality. According to the Chinese Ministry of Environmental Protection (CMEP, hereafter), AQI is divided into six groups: 0 to 50 (Level I, excellent), 51-100 (Level II, good), 101-150 (Level III, lightly polluted), 151-200 (Level IV, moderately polluted), 201-300 (Level V, heavily polluted), and above 300 (Level VI severely polluted). Level I and II have no health concern to the public, suitable for outdoor activities. Level III and Level IV increase the risk of health problems, especially for sensitive individuals. Outdoor exercise should be reduced. Level V and VI may induce noticeably respiratory symptoms even for healthy people. The elderly and the sick should avoid being outside. The algorithm to compute AQI in China is similar to that developed by the U.S. Environmental Protection Agency (EPA, 2006). It is based on six atmospheric pollutants, namely, sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), carbon monoxide (CO), and ozone. The pollutant that has the highest index is referred to as the primary pollutant and determines the AQI of the day.

Next, we briefly summarize how different air pollutants affect human health.

Fine particulate matter ($PM_{2.5}$). $PM_{2.5}$ is a type of particulate matter with diameters equal to or smaller than 2.5 micrometers. It is mainly generated by power

plants, industry, and automobiles (WHO, 2006). $PM_{2.5}$ can penetrate deep into the lungs and even enter the bloodstream due to its small size. Toxicological and epidemiological evidence shows that $PM_{2.5}$ poses a great threat to the respiratory and cardiovascular systems (Seaton et al., 1995; Delfino, 2002; Ling and van Eeden, 2009). The short-term exposure to $PM_{2.5}$ at high concentrations can aggravate asthma, cause lung diseases, and exacerbate existing heart diseases. Long-term exposure is associated with reduced pulmonary function and even premature death (Pope et al., 2000).

Ozone (O_3). Ground-level ozone is formed by the chemical reactions between volatile organic compounds (VOC) and nitrogen oxides (NO_x) in the presence of sunlight at high temperature (Moretti and Neidell, 2011). Therefore, ozone concentration is generally higher in metropolitan areas with large amounts of motor vehicle exhausts and industrial emissions that contain VOC and (NO_x). Additionally, the ozone concentration usually peaks in the summertime, when the temperature is high and the sunlight is at its strongest level. Breathing ozone deeply can trigger a variety of health problems such as shortness of breath, coughing, and chest pain. It can also aggravate asthma, emphysema, chronic bronchitis, and other respiratory diseases.

Sulfur dioxide (SO_2). The main sources of SO_2 pollution are the emissions from fossil-fuel combustion in power plants and other industrial facilities. It generally coexists with other air toxics or can react with other compounds to form new particles ($PM_{2.5}$). WHO (2006) reports that SO_2can trigger many health problems (mainly respiratory), even after brief exposure only. For perspective, SO_2 causes difficulty in breathing by narrowing the airways, exacerbates asthma symptoms, and increases the risk of other respiratory morbidities. Children, the elderly, and people with asthma are the most vulnerable to SO_2, especially in the course of exercising or fast and deep breathing (Linn et al., 1987).

Carbon monoxide (CO). CO is a colorless, odorless, and dangerous gas pollutant. The reaction between CO and hemoglobin reduces the amount of oxygen transported

in the blood. As a consequence, the functions of critical organs such as the brain and heart may be impaired due to tissue hypoxia. CO can also poison the neurological system at high concentrations, impairing the coordination and cognition temporarily or permanently (WHO, 2000). CO is generated mainly by the incomplete combustion of carbonaceous materials. Therefore, the largest sources of CO in outdoor space are petroleum-derived emissions, either from vehicle exhausts or industrial processes.

2. Data Sources and Descriptive Statistics

This paper focuses on Guangzhou, one of the most developed cities in China, for two main reasons that make it suitable and desirable for this study.[13] First, Guangzhou, as a large manufacturing city, is exposed to high levels of air pollution that are notably above the international standards.[14] For instance, the annual $PM_{2.5}$ mean in the city is 45.9 μg/m^3 during our sample period, which is around four and a half times as large as the regulatory threshold set by the WHO (10 μg/m^3). Moreover, its AQI ranges from 3 (excellent) to 226 (heavily polluted) during our test period. Given the high and variable levels of air pollution, Guangzhou is an ideal setting for studying the relationship of our interest, which may potentially be nonlinear. Second, the data on student illnesses and absences cover almost all the schools in the entire city rather than a particular sample of schools. Our data contain 3,139 schools across 12 districts in Guangzhou City, including 1,671 kindergartens (53.2%), 1,075 elementary (34.2%), 439 middle (14.0%), and 127 high schools (4.0%).[15] Exploiting such a large dataset helps to avoid the sample selection problem and renders our analysis more representative.

2.1 Data Sources

Student illness and absence records. Our data are drawn from the Student Health

[13]Guangzhou is City ranked the third in 2014 in the gross domestic product (GDP) among 290 cities in China, after Shanghai and Beijing, according to China City Statistical Yearbook 2015.
[14]The three pillar industries of Guangzhou City are auto, electronic products, and petroleum chemical manufacturing (China City Statistical Yearbook, 2015). In 2015, it ranked at the bottom in air quality among 22 cities in Guangdong Province (Chinese Ministry of Environmental Protection, 2015).
[15]The sum of the four components is larger than 100% because some schools operate two or three types of schools at the same time.

Monitoring System operated by the Guangzhou Center for Disease Control and Prevention (CDC).[16] The system collects illness and absence records from schools for each school day, based on which the CDC monitors the incidence and prevalence of infectious diseases among students in Guangzhou.[17] It was implemented on a pilot basis in 2010, but the vast participation of schools in the program and the realization of their daily records were achieved at the beginning of 2013, with the exception of local universities and colleges. Thus, our sample starts from January 2013.

Here, we summarize how these records are collected. The school nurses conduct daily morning checks according to roll books with the assistance of head teachers. Figure A.2 presents a sample of the paper form to be filled out by the school nurses. Sick children are reported to their parents and, when necessary, delivered to hospitals for diagnoses and treatments. Parents of absent children are contacted by telephone to inquire whether the absence is related to illness, and the encountered symptoms if it is the case. Sick students, who are absent, are required to submit the hospital diagnoses upon returning to school. All the records are entered into the system by the school nurses on the same day.[18] The paper forms are stored by the schools and are checked frequently by the CDC on a random basis to ensure the quality of reports.[19] Figure A.3 displays an electronic record sample containing all the information collected from the morning checks. Each record includes the demographic characteristics of the sick or absent student, including age, gender, class number, grade, school address, and dormitory number in case of residence in the campus, as well as information on the illness and absence, such as date, symptoms, hospital diagnosis, and reason of the potential request for absence.

We treat respiratory diseases as the most potential outcome influenced by air

[16]Figure A.1 shows the log-in page of the system (http://xsjk.gzcdc.org.cn/gzjkjc/).
[17]Student illness and absence records are collected by the CDC for monitoring and research purposes. These data are not publicized or used to assess the school quality or to affect the educational resource allocation. Thus, there is little incentive for schools to misreport illnesses or absences.
[18]Hospital diagnoses are updated online when the absent students return to school.
[19]If any intentional misreporting behaviors are detected, the corresponding school will receive a warning along with posting a public notice on the incidence.

pollution, following the convention in the pollution-health literature. We define an illness as a respiratory disease if it satisfies at least one of the following criteria: (1) the hospital diagnosis clearly states pneumonia, tonsillitis, influenza, emphysema, asthma, bronchitis, or other respiratory infections (Deschenes et al., 2017);[20] (2) it includes at least one of the following symptoms: runny nose/sneezing, sore throat, wet or dry coughing, wheezing, asthma attack, and so forth (Schwartz et al., 1994). We also identify gastrointestinal illnesses, which are likely to be unaffected by air pollution, and use them to test the validity of our specification.

Air pollution. Ambient air pollution data are obtained from the Chinese Ministry of Environmental Protection.[21] We mainly focus on AQI, $PM_{2.5}$, ozone, SO_2, and CO. The hourly readings for AQI and the four air pollutants are recorded at 11 outdoor monitoring sites across 12 districts in Guangzhou City. Only a few missing and invalid readings in our sample period exist. In particular, $PM_{2.5}$ levels are not available for less than 3% of the sample days and the levels of the other pollutants are lacking for around 7%. These missing values are uniformly spread across dates and across monitors for all pollutants, indicating that the monitors are operated reasonably well.

Figure 2 and 3 plots the hourly and monthly patterns of AQI and the four pollutants. As shown in Panel (a), AQI rises from the morning rush hours, tends to be slightly higher during the evening, and considerably higher in colder months than in warmer months. In Panel (b), $PM_{2.5}$ shares a similar pattern with AQI, suggesting that $PM_{2.5}$ is the most relevant pollutant to Guangzhou City. Turning to Panel (c), ozone drastically rises with the temperature, peaks at noon, and falls from the late afternoon till night and remains at a low level in the early morning. It is higher during the warmer months from

[20]The classification is based on the International Classification of Diseases and Related Health Problem, 10th Revision ((ICD-10)-WHO Version for 2016). ICD is the diagnostic classification standard for all the clinical and research purposes. It defines the universe of diseases, disorders, injuries, and other related health conditions and has been widely used in clinical care and research (Seaton et al., 1995; Schlenker and Walker, 2016).

[21]There has been a concern that Chinese authorities manipulated published pollution readings around key thresholds (Chen et al., 2012; Ghanem and Zhang, 2014). Stoerk (2016) compares the official data to the U.S. Embassy data and finds that the Chinese data fits Benford's Law from 2013 onwards. In our sample period, the correlation coefficient between 24-hour $PM_{2.5}$ mean across CMEP sites and the average at the U.S. embassy is 0.91.

May to October. This pattern is consistent with ozone formation which requires the presence of heat and sunlight. As can be seen in Panel (d), SO_2 starts to rise in the late afternoon and continues during the night until the next morning. It peaks in January and December, but it is relatively steady during the other months. We observe in Panel (e) that CO levels are elevated during the morning and evening rush hours, consistent with the fact that CO is closely related to vehicle exhaust. Regarding its monthly pattern, CO is found to be typically higher in cold months. This is possibly because the control systems of car emissions operate less effectively during cold winter days. Conclusively, the patterns of the four air pollutants are consistent with the reported documents in the previous studies (Davis, 2008; Neidell, 2009).

We match the pollution data to the absence data as follows. First, we convert hourly pollution readings into 24-hour means. Notably, to capture the impact of the concurrent and the most recent pollution exposure, we define the treatment period as the 24 hours prior to classes ending of the day in consideration (i.e., from 6 pm of the previous day till same-day 6 pm). We measure ozone levels with the maximum 8-hour moving averages that culminate with every clock hour within the 24-hour period and measure AQI, $PM_{2.5}$, SO_2, and CO levels using the 24-hour means. To properly account for the missing observations in the pollution data, we keep track of the number of observations per day and only use monitors with at least 6 out of 8 zone concentration readings or 20 out of 24 mass concentration readings for the other pollutants.[22] Second, we match each school with monitors within a 40 km radius.[23] We then create daily school pollution measures by taking the average monitor reading of all qualified monitors, weighting by the inverse distance between the monitor and the school.[24] Hence, the closer monitors are valued more heavily, reducing the measurement error.[25]

[22]This approach is also adopted by the CMEP as well as Schlenker and Walker (2016). If a monitor has not a single reading for a day, we exclude this monitor on that day.
[23]Under the restriction of 40 km, each school is matched with nine monitors on the average. In Table A.II, we present the robustness check using various restrictions of the maximum distance (20 and 30 km) between a school and a pollution monitor. Our results are not sensitive to alternative distance thresholds.
[24]We require that a school-date cell should have at least two eligible monitors (within 40 km) with valid readings on that day. Otherwise, we treated the school-date cell as a missing observation.
[25]The inverse-distance weighting method has been adopted to impute pollution in many studies (Currie and Neidell, 2005; Deschênes and Greenstone, 2007; Schlenker and Walker, 2016).

Weather. Our 3-hour readings of the meteorological data for Guangzhou City come from the National Oceanic and Atmospheric Administration (NOAA), which is a comprehensive source of publicly available global historical weather and climate data.[26] We mainly use the information on cumulative precipitation, temperature, wind speed, dew point (a measure of humidity), and atmospheric pressure, since these weather conditions may influence the student absence (health). We aggregate the 3-hour weather readings into 24-hour means. Considering the potential nonlinear dose-response relationship between weather and absences (health), we include all weather variables in non-parametric forms. Following Deschenes et al. (2017), we classify five categories for each weather variable: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of its distribution during the sample period. Each category equals to one if the weather variable falls into the category it belongs to and equals to zero if otherwise.

The temperature inversion data for Guangzhou City are available from the National Aeronautics and Space Administration (NASA). The data record air temperatures every six hours at 42 vertical layers ranging from 110 meters to 36 thousand meters within 50 by 60-kilometer grids (0.5×0.625-degree grid).[27] We follow the approach applied in creating pollution measures to construct our temperature inversion measures at the schools. Specifically, for each school, we average temperatures across all observation points within a 40 km radius using the inverse-distance weighting scheme for every six hours and for each layer. Following Arceo et al. (2016), we define a temperature inversion at a school if the temperature of the second layer (320 m) is higher than that of the first layer (110 m) for any six-hour period during the 24 hours from 6 pm of the previous day till 6 pm of the day in consideration.[28]

---

[26]The NOAA only has one weather station in Guangzhou. Therefore, we are not able to precisely measure the weather conditions across various locations within the city. Consequently, our OLS estimates of the weather conditions may suffer from attenuation bias due to classical measurement error.

[27]There are twelve observation points across Guangzhou City. They are located at (22.5°N, 112.5°E), (22.5°N, 113.125°E), (22.5°N, 113.75°E), (22.5°N, 114.375°E), (23°N, 112.5°E), (23°N, 113.125°E), (23°N, 113.75°E), (23°N, 114.375°E), (23.5°N, 112.5°E), (23.5°N, 113.125°E), (23.5°N, 113.75°E), and (23.5°N, 114.375°E), respectively.

[28]Our main results are robust when we use layers at 110 and 550 meters to define temperature inversions.

3.2 Descriptive statistics

In our analysis, we aggregate the student illness and absence data into the school-grade-date level and merge them with the 24-hour period, school-level pollution measurements, and meteorological variables. We exclude observations during summer and winter vacations, public holidays, and weekends. We also remove observations for children in care classes (grade one in kindergartens), whose ages are between 0 and 2 years, since their school times are relatively flexible.[29] Finally, we have 12,238 school-grade pairs during school terms between January 2013 and November 2015 for a total of 3,122,724 observations.

Table I presents the summary statistics for the baseline regression sample. Each observation for health and absence outcomes represents a school-grade by school day pair (school-grade-date for simplicity). The major health measurement is the daily respiratory illness rate (%), defined as the proportion of students with respiratory diseases over the total number of students in a given school-grade. The outcome of interest for absenteeism is the daily total absence rate (%), defined as the proportion of absent students over the total number of students in a given school-grade. Other health and absence measurements are computed in a similar way.

Two points are worth mentioning in Table I. First, student illnesses are highly related to respiratory diseases. On average, 0.22% of students are sick per day.[30] Among them, 79.64% have respiratory illnesses, 6.79% have gastrointestinal problems, and the remaining students are sick for other reasons. Second, most school absences are related to illnesses, and particularly to the respiratory ones.[31] Notably, the illness-related absence rate is less than the total illness rate, which indicates that some students

[29]Attendance in care classes is relatively random, and often depends on the interest of family members. Observations from care classes account for approximately 3% of our full sample. Full results that include this part of observations are quantitatively similar and available upon request.

[30]Taken literally, there are approximately 2.27 million students in Guangzhou City in 2015. Thus, around 2.27 million×0.22%=4,994 students are reported every day to the CDC as feeling unwell.

[31]The illness-induced absence rate in Guangzhou is 0.17 %, which is comparable to those in Beijing (0.54%), Shanghai (0.12%), Chengdu (0.12%), and Jiangyin (0.37%) (Gao and Ma, 2007; Dong et al., 2014; Wang et al., 2015; Miao et al., 2015).

in our sample prefer not to be absent although they are not feeling well. This is possibly because the opportunity cost of not attending school in certain periods (e.g., the week before final exams) is relatively high.

In Figure 4, we further summarize how these outcome variables vary over time and across grades. Panel (a) displays the distributions of respiratory illness rates and total absence rates over school days. As can be seen, they distribute similarly, with absence rates being slightly higher, reinforcing that school absences are mostly due to respiratory diseases. In terms of the dispersion, the median day exhibits an absence rate of 0.21% (0.14% for the respiratory illness rate), and the days in the 10th and 90th percentiles have absence rates of 0.14% and 0.31%, respectively (0.11% and 0.25% for the respiratory illness rate). These day-to-day variations provide the opportunity to explore the relationship of our interest. Panel (b) demonstrates their monthly patterns, with lower rates at the beginning of an academic year (September) and higher rates in the last months of school terms (June for the spring term and January for the fall term, respectively).[32] The respiratory illness rate significantly surpasses the absence rate in June and January. This is possibly due to the high opportunity cost when approaching the final examinations, so a larger number of sick students choose to attend school instead of resting at home for recovery. Panel (c) illustrates the day-of-week patterns. The total absence rates are higher on Monday and Friday than during the middle of the week, while there is no much variation in the respiratory illness rate. Finally, we plot the average rates by grade in Panel (d), with "K", "P", "M", and "H" denoting kindergartens, primary, middle, and high schools, respectively. Both the respiratory illness rate and total absence rate display monotonic decreasing patterns. Higher grades, lower the probabilities of getting respiratory illnesses and being absent from school. This is consistent with the epidemiological evidence, i.e., the immune systems of the elder students are stronger than those of the younger ones, and thus the absences of the elder students tend to be less (WHO, 2005). Another possible explanation is that high-

[32]Spring term is from March to June. Fall term is from September till the next January. Summer vacation starts from July and ends in August, while winter vacation is in February.

grade students spend much less time on outdoor activities due to the heavy study load, and therefore they are less likely to be exposed to air pollution.

3. Empirical Strategy

Herein, we first introduce a fixed-effects model for estimating the contemporaneous effects of air pollution on student health and school attendance. We then discuss several challenges to the validity of OLS estimation. Finally, we describe our favored instrumental-variable strategy to solve the potential endogeneity problem and attenuation bias from classical measurement error.

4.1 The Fixed-effects Model

The primary econometric framework is as follows:

$$Y_{st} = \beta_i Pollutant_{ist} + W'_t\gamma + H'_t\delta + school + year + week + dow + school \times year + year \times week + school \times week + week \times dow + school \times year \times week + \varepsilon_{st}, \quad (1)$$

where Y_{st} refers to the respiratory illness rate or the total absence rate for school-grade *s* on date *t*. $Pollutant_{ist}$ is the average mass concentrations of air pollutant *i* (AQI, $PM_{2.5}$, ozone, SO_2, or CO, measured in a 10-unit basis) for the 24-hour period prior to classes ending on date *t* (i.e., from prior-day 6 pm to same-day 6 pm). The weather conditions jointly impact both pollution and health (Deschenes et al., 2009), and therefore we include a vector of meteorological factors, W_t, consisting of daily temperature, wind speed, dew point (a measure of humidity), atmospheric pressure, and cumulative precipitation. They enter as indicators to accommodate possible nonlinear effects. Specifically, we classify five categories for each of these variables, namely, 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of its distribution during the sample period (Deschenes et al., 2017). Each category takes the value of one if the weather variable falls into the category it belongs to, and the value of zero if

otherwise.[33]

$school$ represents school-grade fixed effects designed to capture any time-invariant school-grade characteristics. $year$ and $week$ represent year and calendar week fixed effects, respectively. They are included to account for the common patterns of air pollution and health over years and weeks. To address the concern that absences (health) may be different on holidays from non-holidays, and weekdays from weekends, we include dummy variables indicating one day before or one day after public holidays (H_t) and day of the week (dow).

Considering many time-varying confounding factors, we add three pairwise interactions between school-grade, year and week. Specifically, $school \times year$ captures the school-grade characteristics that vary by year, such as the improvement in sanitation conditions. $year \times week$ captures the year-specific seasonal patterns, such as the outbreak of contagious diseases in certain periods, which high pollution levels coincide with. $school \times week$ controls for any school-grade specific seasonal patterns. For perspective, students of the 10th-grade are seldom absent from schools in late June, several weeks before the senior high entrance examination, although ozone levels are relatively high during these hot, sunny days. If we were to directly link the low absence rates with the high pollution levels without controlling for $\text{school} \times week$, we would underestimate the impact of air pollution. Since the seasonal effects are expected to have an important influence, we further include day-of-week by week dummy variables to flexibly capture seasonality in air quality and absences (health). To be conservative, we include a three-way interaction term, $\text{school} \times year \times week$, into Equation (1) to capture any school-grade specific shock in a given year-week.[34] Finally, ε_{st} denotes the idiosyncratic error term. The standard errors are clustered at the community level to adjust for potentially within-community serial correlations over

[33]We also use quadratic estimation, an alternative approach to nonlinearity in weather, and the results are undisturbed and are available upon request.

[34]Robustness of the OLS estimates to different sets of fixed effects can be found in Tables A.III and A.IV. The results look almost identical if we exclude the three-way interaction term, so we decide to include the more restrictive set of fixed effects.

time in pollution.[35]

The parameter of interest is β_i, which captures the effect of air pollutant i. It can be interpreted as a β-increase in the respiratory illness rate or the total absence rate due to additional 10 units of pollutant. Our identifying assumption for the fixed-effects model is that after partialling out idiosyncratic factors affecting specific school-grades in certain year-weeks, as well as the systematic day-of-week patterns by week, the remaining daily pollutant variations are exogenous to the daily incidence of student illnesses, as well as school absences.[36]

There are several challenges to the validity of our OLS estimation. First, endogeneity concerns may still remain given these comprehensive controls. An example that could violate our identifying assumption is a random disaster, such as a chemical plant explosion, which may raise the air pollution levels while increasing the school absences. However, to the best of our knowledge, this type of incidences is not reported during our sample period. Second, pollution levels may be measured with errors due to sporadic monitoring sites.[37] The inclusion of fixed effects may exacerbate the classical measurement error, making our OLS estimates biased downward.

Figure 5 emphasizes the relationships between air pollution and our outcome variables, partialling out the meteorological factors and the full set of fixed effects. As the figure illiterates, both the respiratory illness rate and the total absence rate are positively correlated with daily AQI levels.[38]

---

[35]We did not cluster at the district or monitor level, which may deliver too few clusters as suggested by Angrist and Pischke (2008). Small cluster numbers can bias downwards cluster-robust standard errors, leading to an overstated statistical significance of results (Bertrand et al., 2004). We currently have 215 clusters. We also clustered on both community and day to account for the potential spatial correlation across communities on a given day. Our estimates are robust to this two-way cluster-robust standard errors and are available upon request.

[36]Note that as we exploit day-to-day variation within school-grade pairs in this setting, sorting across different school-grade pairs due to differential pollution should not be a concern.

[37]We have 11 outdoor pollution monitors across 12 districts in Guangzhou City during our sample period.

[38]We also describe the raw correlations between air pollution and the outcome variables in Figure A.4, where we plot the weekly fluctuations of average daily respiratory illness and total absence rates against average daily AQI levels. Our outcome variables comove with AQI in most cases, suggesting a positive correlation between air pollution and student respiratory illnesses and school absences.

4.2 The Instrumental-variable Approach

Next, we adopt an instrumental-variable approach to alleviate the potential bias from omitted variables and classical measurement error. Specifically, we exploit the temperature inversion, a meteorological phenomenon that occurs when air temperature abnormally increases with height. The formation of a thermal inversion mainly relies on meteorological and geographical factors, and thus, it does not affect absenteeism (health) directly (Arceo et al., 2016). However, it traps air pollutants near the ground by reducing the upward movement of air from the layers below (Jacobson, 2002; Secretaría del Medio Ambiente, 2005). Hence, we use the occurrence of temperature inversion at a school as a random shock and link the induced exogenous variation in pollution levels to the variation in school absences (student illnesses).

We propose the two-stage-least-square model as follows:

$$Pollutant_{ist} = \alpha TI_{st} + W'_t\delta + H'_t\varphi + school + year + week + dow + school \times year + year \times week + school \times week + week \times do + school \times year \times week + \varepsilon_{ist}, \quad (2)$$

$$Y_{st} = \beta_i \widehat{Pollutant}_{ist} + W'_t\theta + H'_t\rho + school + year + week + dow + school \times year + year \times week + school \times week + week \times do + school \times year \times week + \sigma_{st}, \quad (3)$$

where TI_{st} represents a temperature inversion dummy, which varies at the school-date level. In this specification, we include the same set of fixed effects and meteorological factors as controlled for in Equation (1).[39] Our 2SLS estimates are based on the exclusive restriction assumption that, conditional on the weather and the set of fixed effects, the occurrence of temperature inversion affects absenteeism (health) only through the accumulation of air pollution.

Finally, we replace the linear pollution measurements with two dummy variables

[39]Temperature inversions covary with meteorological patterns at ground-level (Chen et al., 2017; Arceo et al., 2016), and therefore it is crucial to control for weather in the specifications.

that separate pollution levels into three categories. This replacement is applied to account for the nonlinear effect of air pollution and to further investigate the relationship between the CMEP standards for air quality and absenteeism (health). Specifically, we classify AQI levels into three groups: 0-50 (the base group), 50-100, and 100 above.[40] Each category takes the value of one if AQI falls into it, and the value of zero if otherwise. By doing this, we are enabled to test the effects of pollution categories below 100% of the CMEP thresholds to examine the effectiveness of the current standards for air quality in public health protection. We estimate the nonlinear models using 2SLS regressions. We exploit the fact that the impact of temperature inversion on pollution concentrations may differ in the thickness of the inversion, and thus, we construct two instruments for the AQI categories (50-100 and 100 above), namely the occurrence of temperature inversion and the thickness of the inversion.

4. Results

5.1 The Relationship between Temperature Inversion and Pollution

We begin by showing the raw correlations between the occurrence of temperature inversion and pollution levels. As illustrated by Panel (a) in Figure 6, the concentrations of AQI as well as the other four air pollutions (especially for $PM_{2.5}$ and ozone) increase with temperature inversion. Additionally, in Panel (b) we compare outcome variables on days with temperature inversion versus on days without. It is evident that both the respiratory illness and the total absence rate are higher when temperature inversion occurs.

We formally test the first stage by estimating Equation (2) in Table II. We find that the occurrence of temperature inversion increases AQI levels by 12.38%.[41] The estimates are statistically significant at the 1% level and pass the weak-instrument test. The results highlight the strength of our instrument.

[40]According to the CMEP, AQI levels below 100 are considered to have no health concerns to the public, suitable for outdoor activities.
[41]The impact of temperature inversion on AQI levels is calculated by 0.83334/6.727*100≈12.38%.

5.2 The contemporaneous effects of air pollution

We summarize the direct health consequences caused by air pollution in Table III. Column (1) presents the OLS estimates based on Equation (1) controlling for weather conditions and the full set of fixed effects. We find a statistically significant positive effect of AQI on the respiratory illness rate. Specifically, a 10-unit rise in daily AQI levels increases the respiratory illness rate by 0.00126 percentage points, equivalent to 0.72% of the sample mean.[42]

In contrast, the corresponding 2SLS estimate is presented in column (2). Likewise, we control for the same meteorological factors and the set of fixed effects as reported in column (1). The coefficient in column (2) becomes larger in magnitude while remaining statistically significant at the 1% level as compared to the OLS estimates. Specifically, a 10-unit increase in the daily AQI shifts up the respiratory illness rate by approximately 0.00582 percentage points, which amounts to 3.31% of the mean level. Converting to standard deviations, we find that a one-standard-deviation increase in daily AQI raises the probability of having respiratory illness for students by 10.04%.[43]

To validate our specification, in columns (3) and (4), we replace the dependent variable with the rate of gastrointestinal illnesses, which is less likely correlated with air pollution. This can be viewed as a placebo test. If we had isolated the effect of pollution from other unobserved socio-demographic determents of student health, we should find no effect of air pollution on the gastrointestinal illness rate. As expected, both the OLS and 2SLS estimates are statistically insignificant, and more importantly, they are an order of magnitude much smaller than the effect size of the respiratory illness rate. The fact that we find that air pollution mainly affects respiratory-related illness confirms the validity of our main specification.

[42]The impact of AQI on the respiratory illness rate is calculated by 0.00126/0.17582*100≈0.72%.
[43]The standard deviation for AQI is 3.034 units. The calculation is 0.00582/0.17582*3.034*100≈10.04%.

Next, we examine the robustness of our 2SLS specification. In column (1) of Table A.V., we start by only including weather, dummies indicating one day preceding or following public holidays, school-grade FE, year FE, week FE, and day-of-week FE.[44] The estimate is statistically significant at the 1% level and suggests that a 10-unit rise in daily AQI increases the respiratory illness rate by 0.00483 percentage points. From column (2) to (3), we additionally include week by school-grade and week by year fixed effects. Their inclusion increases the coefficient size, suggesting that the seasonality is a key to our specification. From column (4) to (5), we add year by school-grade and day-of-week by week fixed effects. Our 2SLS estimates change modestly and remain at the 1% significance level. In the last column, we conservatively include a three-way interaction term. The fact that its inclusion has a little additional effect further supports our approach.

We then proceed to investigate the relationship between air pollution and absenteeism. Table IV reports the corresponding results with the daily total absence rate as the dependent variable. Our favored 2SLS estimates (column (2)) are substantially larger in magnitude as compared to the OLS estimates in column (1). The 2SLS estimate indicates that a 10-unit rise in daily AQI increases the total absence rate by 0.00417 percentage points, equivalent to 2.31% of the sample mean. Converting to standard deviations, we find that a one-standard-deviation increase in AQI raises the probability of being absent by 6.99%.[45]

To understand the mechanism through which air pollution influences absenteeism, we compare the effects of air pollution on various types of absences. We expressly classify the total absences into three groups: respiratory-related, gastrointestinal-related, and non-illness-related.[46] Columns (3) to (8) present the corresponding results.

[44] Weather controls ensure that air pollution is the only channel through which temperature inversion affects respiratory illnesses.

[45] Our 2SLS estimates on the effect of air pollution on absenteeism are robust to a different set of fixed effects (see Table A.VI).

[46] Respiratory-related, gastrointestinal-related, and non-illness-related absences account for 77.35%, 6.83%, and 6.17% of the total absences, respectively. The other absence reasons include chronic diseases, trauma, and so on. They are not affected by air pollution. Results are available upon request.

Column (4) indicates that the respiratory-related absence rate is significantly positively correlated with AQI levels. More importantly, the coefficients account for the majority of those for the total absence rate (see column (2)), indicating that air pollution affects absenteeism mainly through the health channel. In column (6), the effect of air pollution on gastrointestinal-related absence rate is found to be statistically insignificant, thus validating our specification. Results in column (8) demonstrate that the non-illness-related absence rate is positively associated with air pollution.[47] However, such effects are relatively small in magnitude and statistically insignificant. In general, our results indicate that health effect is the dominant channel through which air pollution increases school absences.

It is important to check whether our specification fully controls for seasonality. If it is not the case, our results may be driven by some omitted seasonal differences in climate or the prevalence of flu, which may have similar timing with the occurrence of temperature inversion. Hence, we add an interaction term between the warm season and AQI levels in the regressions for testing. We define the summer period as one from May to October, where the monthly temperature is above 77°F (25°C). As demonstrated in columns (1) and (2) of Table V, the interaction terms are statistically insignificant, indicating that the effects on student illnesses and absences do not appear to be significantly different across seasons. These findings further suggest that: conditional on our meteorological factors and the set of fixed effects, the seasonality in thermal inversions is not driving our main results.

Next, to check whether AQI captures the combined effects of $PM_{2.5}$, ozone, SO_2, and CO, we construct another single-pollution index by using the principal components method (Arceo et al., 2016). The single endogenous variable is based on fluctuations of the four air pollutants. Table A.VII presents the corresponding 2SLS estimates. The units of the index can be interpreted in terms of standard deviations. A one-standard-

[47]This could be possible because air pollution triggers some avoidance behaviors.

deviation rise in the index increases the respiratory illness rate and the total absence rate by 10.84% and 7.55%, respectively. The effects are similar in magnitude when the dependent variable is AQI. These results indicate that in our setting, AQI reasonably measures the combined effect of air pollution.

Finally, we supplement the analysis with the results of individual air pollutants ($PM_{2.5}$, ozone, SO_2, and CO), as summarized in Table A.VIII (for the health outcome) and Table A.IX (for the absence outcome). We start by estimating single pollutant models, where we separately examine the impact of one pollutant at a time without controlling for co-pollutants. Column (1) of Tables A.VIII and A.IX emphasize the OLS estimates based on Equation (1). As can be seen, $PM_{2.5}$, ozone, and SO_2 display statistically significant positive relationships with both the respiratory illness and the total absence rate. Although the coefficient for CO is positive, it is imprecisely estimated. Column (2) of Tables A.VIII and A.IX report the corresponding 2SLS estimates from Equation (3).[48] As compared to the OLS estimates, all coefficients become larger in magnitude, while remaining statistically significant at the 1% level.

However, one concern is that the estimates on individual pollutants may capture the effects of other pollutants that covary with the one in consideration. We therefore conduct joint estimation in columns (3) and (4) of Tables A.VIII and A.IX. For column (3), we simultaneously include $PM_{2.5}$, ozone, SO_2 and CO into one specification and estimate it by OLS. The coefficients based on the multiple-pollutant model become much smaller in magnitude than the separate estimates in column (1). The effect of SO_2 and CO is not statistically significant.

In column (4) of Tables A.VIII and A.IX, we use an instrumental-variable approach to estimate the multiple-pollutant model. Following Moretti and Neidell (2011), we include the mass concentrations of the other pollutants as linear controls in both the

[48]We present the first stage estimates based on Equation (2) in Table A.X. All the estimates are statistically significant at the 1% level and pass the weak-instrument tests.

first and second stage regressions when one pollutant is examined.[49] In Table A.XI, we present the first-stage estimates. As can be seen, the first stage exercise works well for $PM_{2.5}$ and ozone, but not for SO_2 and CO. Therefore, we only report the 2SLS estimates on the effects of $PM_{2.5}$ and ozone in column (4). This approach restores the significance on $PM_{2.5}$ to 1% level and suggests that both $PM_{2.5}$ and ozone statistically significantly affect student health as well as absenteeism.[50]

5.3 Cumulative effects

In the previous subsection, we document the contemporaneous effects of air pollution. We now examine whether and to what extent previous pollution levels affect health as well as absenteeism in the current period using a lag-structure regression model. This model gives insights on the cumulative effects of air pollution. We commence with the main specification, then consecutively add one additional lag term of pollution into the model until the total number of lags is five.

We initially exploit the reduced form to examine the prolonged effects of the occurrence of temperature inversion. The estimates, shown in Table VI, are monotonically increasing from zero lag to three lags, after which they level out. This pattern suggests that the effect of temperature inversion through the accumulation of air pollution persists for four days.

We then present the cumulative effects of AQI on the respiratory illness rate and the total absence rate in columns (1) and (2) of Table VII, respectively, with pollution levels instrumented by the occurrence of temperature inversion and its lag terms.[51] We note that the estimates are monotonically increasing from zero lag to three lags, and

[49]In principle, we could also instrument each air pollutant using various instrumental variables, and estimate the model simultaneously including all instrumented air pollutants. But in the current setting, we have difficulty in creating multiple instruments that can provide independent variations.
[50]We are cautious in interpreting these results and wary about attributing estimated effects to a particular pollutant, as our research design is not able to perfectly isolate the effect of one pollutant from the others.
[51]Specifically, we instrument each lag term of air pollutant with the corresponding lag term of temperature inversion occurrence.

they level out after that. This pattern is similar to that of temperature inversion observed in Table VI. This summary suggests that the adverse effects of air pollution on both health and absenteeism persist for 4 days.

5.4 Nonlinear effects

In this section, we explore the nonlinear effects of air pollution and further assess the effectiveness of the official standards for air quality in China. We replace the linear air pollution measurements in Equation (3) with two dummy variables that divide AQI levels into three groups: 0-50 (the omitted group), 50-100, and 100 above.[52] Each category takes the value of one if AQI falls into it, and the value of zero if otherwise. By doing this, we are enabled to test the effects of pollution categories below 100% of the CMEP thresholds to examine the effectiveness of the current standards for air quality in public health protection.

We estimate the nonlinear models using 2SLS regressions. We exploit the fact that the impact of temperature inversion on pollution concentrations may differ in the thickness of the inversion, and thus, we construct two instruments for the AQI categories (50-100 and 100 above), namely the occurrence of temperature inversion and the thickness of the inversion. As illustrated in Table A.XII, both instruments are statistically significantly correlated with AQI levels. The pollution levels are higher when the temperature inversion layer is thicker. The F-statistic for each of the two first stages is above the Stock-Yogo 10% critical value for a single endogenous regressor, proving the strong relevance of our instruments.

Columns (1) and (2) of Table VIII demonstrate that the two AQI categories are significantly positively correlated with the daily respiratory illness rate as well as the total absence rate. Furthermore, the effect displays a monotonically increasing pattern.

[52]According to the CMEP, AQI below 100 is considered safe to the public health. In contrast, AQI above 100 has health concerns.

The coefficients of the two AQI categories are statistically significantly different from each other at the traditional level. An additional day with AQI exceeding the official threshold (100 units) statistically significantly increases the daily respiratory illness and the total absence rate by 0.00164 and 0.00089 percentage points, respectively. Remarkably, air pollution affects student health as well as school attendance even at pollution levels that are lower than the regulatory standard. This suggests that the current official standard for air quality may not be low enough to protect students.

5.5 Heterogeneous analyses

Effects across school types. We first study the heterogeneous effects of air pollution across school types.[53] Table IX statistically summarizes the corresponding 2SLS estimates on the effects of daily AQI levels on student health (column (1)) as well as on absenteeism (column (2)).

The impact on children in kindergarten is first examined in Panel (a). All coefficients are statistically significant at 1% level. A 10-unit increase in the daily AQI levels raises the respiratory illness rate by approximately 0.00564 percentage points, which amounts to 2.36% relative to the sample mean of kindergartens. In terms of the impact on absenteeism, an additional 10-unit of daily AQI leads to an increase in the probability of being absent by 0.00356 percentage points, equivalent to 1.64% of the mean level.

Panel (b) reveals that the adverse effects of air pollution on primary-school students are relatively large. Specifically, an additional 10-unit of AQI increases the respiratory illness rate by 0.00677, which amounts to 4.44% of the sample mean. For absenteeism, the number is and 0.00532, equivalent to 2.95% of the mean level. We find similar effects in magnitude for middle-school students in Panel (c). A 10-unit rise in the daily

[53]Students in kindergarten, primary, middle, and high schools are generally 2-6, 7-12, 13-15, and 16-19 years old, respectively. Note that the observations from care classes with students aged 0 to 2 years are dropped.

AQI levels increases the risk of respiratory illness rate by 4.26% and the probability of being absent by 2.66%.

Panel (d) demonstrates the results for high-school students. As can be found, they are less influenced by air pollution than the other age groups, especially in school attendance as illustrated in column (2). A 10-unit increase in the daily AQI levels only augments the absence rate by approximately 0.60%. This could be explained by several reasons. First, they have relatively stronger immune systems and thus they are less susceptible to air pollution. Second, they spend much less time in outdoor activities due to the heavy study load, and therefore they are less likely to be exposed to air pollution. If these two explanations work, we would expect a smaller effect of air pollution on the respiratory illness rate for high-school students. In column (1), we indeed find that the impact of air pollution on health is the smallest for students from high schools, speaking of the biological effect channel.

Additionally, we argue that senior students may be less willing to ask for leave even when they get sick, as their opportunity cost of being absent is relatively high. This point is supported by the comparison between the coefficient for absenteeism and that for health. The absence to health ratio for kindergartens, primary, middle and high schools are 0.63, 0.79, 0.65 and 0.42, respectively.[54] This implies that high-school students are less likely to be absent conditional on health.

Effects across school quality. We finally examine whether the effects of air pollution vary with different family backgrounds. Hypothetically, students from rich families are more likely to adopt avoidance behaviors to air pollution, such as wearing particulate-filtering face masks (Zhang and Mu, 2017), and therefore, they might be less affected by air pollution.

[54] Take the high-school students as an example, the absence to health ratio for high-school students is 0.00040/0.00096≈0.42.

Lack of information on student family background, we here use school quality as a proxy to test the hypothesis. School quality is measured by a dummy variable, the provincial-level key school. During our sample period, Guangzhou City has 83, 112, 69, and 67 provincial-level key kindergartens, primary, middle, and high schools of Guangdong province, respectively. To test whether the adverse effects of air pollution differ for provincial-level key versus non-key schools, we include the interaction between AQI levels and the key-school indicator in the regression. Table X summarizes the results. As can be seen, the interaction terms are negative for both the health and the absence outcomes, implying that key-school student tends to be less affected by air pollution. However, they are not statistically significant at the conventional level. This suggests that air pollution impacts students from various families indifferently, possibly because the willingness to avoid air pollution is low considering students from rich families.

5. Conclusion

This paper estimates the effects of air pollution on student health and school absences in a developing country context. We exploit a novel administrative data containing daily student illness and absence records from more than 3,000 schools across Guangzhou City, China. We isolate the contemporaneous effect of air pollution by instrumenting daily pollution levels with the occurrence of temperature inversion, a meteorological phenomenon that does not present a risk for absenteeism (health) but may trigger the accumulation of air pollutants by reducing the upward movement of air.

We find that air pollution significantly increases the incidences of respiratory illnesses and the probability of being absent among students. We demonstrate that air pollution affects school absences mainly through the health channel, and this adverse effect persists for at least four days. We also explore the nonlinear dose-response function of air pollution and find a monotonically increasing pattern. More importantly, we reveal that air pollution exerts non-ignorant negative effects on student health as

well as absenteeism at levels below the Chinese official regulatory thresholds for air quality. Finally, we show that the effects of air pollution are relatively larger for younger students.

This paper contributes to the existing literature by providing the first estimate on the causal effect of air pollution on absenteeism for students from developing countries. It also shows that air-pollution abatement does indeed matter for developing countries, by revealing the additional costs caused by air pollution in terms of reduced school participation and the direct health consequences. Since health and school participation are central parts of human capital accumulation and their disturbance has linked to a set of negative academic and social-economic outcomes, the government should take swift actions to regulate air pollution and guide the public to protect themselves from the harmful impact of air pollution.

We conclude this work by conducting a back-of-the-envelope calculation for China. In 2015, China has approximately 0.21 billion students (including kindergartens, primary, middle, and high schools).[55] Based on our estimates, a 10-unit rise in the daily AQI levels results in approximately 87.57 thousand more absent student across the country on a daily basis.[56] The estimated government education expenditure per student per day in 2015 is around 100 RMB (15 USD), according to the Ministry of Education. Therefore, a 10-unit rise in the daily AQI levels yields a sizable daily cost of 1.31 million USD.

[55]This number is derived from China Statistical Yearbook 2016.
[56]We acknowledge that these estimates rely on a strong assumption, i.e., the estimated effects of air pollution could be applied to other areas in China.

References

1. Angrist, J. D., and Pischke, J. S. (2008), Mostly Harmless Econometrics: An Empiricist's Companion. (Princeton, NJ: Princeton University Press).
2. Arceo, E., Hanna, R., and Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico City. Economic Journal, 126(591):257–280.
3. Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics, 119(1), 249-275.
4. Chang, T., Zivin, J. G., Gross, T., & Neidell, M. (2016). The effect of pollution on worker productivity: Evidence from call-center workers in China (No. w22328). National Bureau of Economic Research.
5. Chay, K. Y. and Greenstone, M. (2003a). Air quality, infant mortality, and the Clean Air Act of 1970. NBER Working Paper.
6. Chay, K. Y. and Greenstone, M. (2003b). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. The Quarterly Journal of Economics, pages 1121–1167.
7. Chen, S., Oliva, P., and Zhang, P. (2018). Air Pollution and Mental Health: Evidence from China (No. w24686). NBER Working Paper.
8. Chen, Y., Ebenstein, A., Greenstone, M., and Li, H. (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. Proceedings of the National Academy of Sciences, 110(32):12936–12941.
9. Chen, Y., Jin, G. Z., Kumar, N., and Shi, G. (2012). Gaming in air pollution data? Lessons from China. The BE Journal of Economic Analysis and Policy, 12(3).
10. Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. Journal of Economic Literature, 47(1):87–122.
11. Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? The Review of Economics and Statistics, 91(4):682–694.
12. Currie, J. and Neidell, M. (2005). Air pollution and infant health: What can we learn from California's recent experience? The Quarterly Journal of Economics, 120(3):1003–1030.
13. Davis, L. W. (2008). The effect of driving restrictions on air quality in Mexico City. Journal of Political Economy, 116(1):38–81.
14. Delfino, R. J. (2002). Epidemiologic evidence for asthma and exposure to air toxics: linkages between occupational, indoor, and community air pollution research. Environmental health perspectives, 110(Suppl 4), 573.
15. Deschenes, O., Greenstone, M., and Shapiro, J. S. (2017). Defensive investments and the demand for air quality: Evidence from the NOx Budget Program and ozone

reductions. The American Economic Review, 107(10):2958–89.
16. Dong, Y., Zhang, Q., Sun, S., and Yuan, H. (2014). An analysis of school absences in Shanghai. Shanghai Journal of Preventive Medicine, 26(12):675–677 (in Chinese).
17. Ebenstein, A., Fan, M., Greenstone, M., He, G., Yin, P., and Zhou, M. (2015). Growth, pollution, and life expectancy: China from 1991–2012. The American Economic Review, 105(5):226–231.
18. Ebenstein, A., Fan, M., Greenstone, M., He, G., and Zhou, M. (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proceedings of the National Academy of Sciences, 201616784.
19. Fu, S., Viard, V. B., and Zhang, P. (2017). Air quality and manufacturing firm productivity: comprehensive evidence from China. SSRN Working Paper.
20. Gao, X. and Ma, H. (2007). An analysis of school absences due to illness among elementary and middle school students in Beijing. Chinese Journal of School Health, 28(11):1037–1038 (in Chinese).
21. Ghanem, D., and Zhang, J. (2014). 'Effortless Perfection': Do Chinese cities manipulate air pollution data?. Journal of Environmental Economics and Management, 68(2), 203-225.
22. Greenstone, M. and Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in India. The American Economic Review, 104(10):3038–3072.
23. Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. Journal of Education for Students Placed at Risk (JESPAR), 19(2), 53-75.
24. Griliches, Z., and Hausman, J. A. (1986). Errors in variables in panel data. Journal of econometrics, 31(1), 93-118.
25. Ham, J. C., Zweig, J. S., and Avol., E. (2011). Pollution, test scores and the distribution of academic achievement: Evidence from California schools 2002-2008. IZA Working Paper.
26. He, G., Fan, M., & Zhou, M. (2016). The effect of air pollution on mortality in China: evidence from the 2008 Beijing Olympic Games. Journal of Environmental Economics and Management, 79, 18-39.
27. Hicks, D., Marsh, P., and Oliva, P. (2016). Air pollution and procyclical mortality: Causal evidence from thermal inversions. working paper.
28. Jacobson, M. (2002). Atmospheric Pollution. History, Science, and Regulation, Cambridge: Cambridge University Press.
29. Jans, J., Johansson, P., and Nilsson, P. (2014). Economic status, air quality, and child health: Evidence from inversion episodes. IZA Working Paper.
30. Kappos, A.D., Bruckmann, P., Eikmann, T., Englert, N., Heinrich, U., Hoppe, P., Koch, E., Krause, G.H., Kreyling, W.G., Rauchfuss, K., Rombout, P., Schulz-Klemp, V., Thiel, W.R. and Wichmann, H.E., Health effects of particles in ambient

air. International journal of hygiene and environmental health. 207(4) 2004 399–407.

31. Lavaine, E., and Neidell, M. (2017). Energy production and health externalities: Evidence from oil refinery strikes in France. Journal of the Association of Environmental and Resource Economists, 4(2), 447-477.
32. Lavy, V., Ebenstein, A., and Roth, S. (2016). The long run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. American Economic Journal: Applied Economics, 8(4):36-65.
33. Levy, J., Timothy C., Jouni T., James H., and John E. (2001). Assessing the Public Health Benefits of Reduced Ozone Concentrations. Environmental Health Perspectives, 109(12):9–20.
34. Ling, S. H., & van Eeden, S. F. (2009). Particulate matter air pollution exposure: role in the development and exacerbation of chronic obstructive pulmonary disease. International journal of chronic obstructive pulmonary disease, 4, 233.
35. Linn, W. S., Avol, E. L., Peng, R.-C., Shamoo, D. A., and Hackney, J. D. (1987). Replicated dose-response study of sulfur dioxide effects in normal, atopic, and asthmatic volunteers. American Review of Respiratory Disease, 136(5):1127–1135.
36. Liu, Haoming and Alberto Salvo, 2017. "Severe Air Pollution and School Absences: Knowledge for International Knowledge Workers". IZA DP No. 11134.
37. Miao, G., Yuan, H., and Xu, C. (2015). Monitoring situation analysis for the students absent from class because of illness from 2012 to 2013 academic year in Jiangyin City. Journal of Medical Pest Control, (12):1342–1345 (in Chinse).
38. Moretti, E. and Neidell, M. (2011). Pollution, health, and avoidance behavior evidence from the ports of Los Angeles. Journal of Human Resources, 46(1):154–175.
39. Neidell, M. (2004). Air pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma. Journal of Health Economics, 23(6):1209–1236.
40. Neidell, M. (2009). Information, avoidance behavior, and health: The effect of ozone on asthma hospitalizations. Journal of Human Resources, 44(2):450–478.
41. Peters, A., Veronesi, B., Calderon-Garciduenas, L., Gehr, P., Chen, L.C., Geiser, M., Reed, W., Rothen-Rutishauser, B., Schurch, S. and Schulz, H., Translocation and potential neurological effects of fine and ultrafine particles, Particle and Fibre Toxicology, 3(13) 2006.
42. Pope 3rd, C. (2000). Epidemiology of fine particulate air pollution and human health: Biologic mechanisms and who's at risk? Environmental Health Perspectives, 108(Suppl 4):713.
43. Ransom, M. and Pope, C. A. (2013). Air pollution and school absenteeism: Results from a natural experiment. Working Paper.
44. Sager, L. (2016). Estimating the Effect of Air Pollution on Road Safety using Atmospheric Temperature Inversions. Grantham Research Institute on Climate

Change and the Environment working paper 251.
45. Schlenker, W., and Walker, W. R. (2015). Airports, air pollution, and contemporaneous health. The Review of Economic Studies, 83(2), 768-809.
46. Schwartz, J., Dockery, D. W., Neas, L. M., Wypij, D., Ware, J. H., Spengler, J. D., Koutrakis, P., Speizer, F. E., and Ferris Jr, B. G. (1994). Acute effects of summer air pollution on respiratory symptom reporting in children. American Journal of Respiratory and Critical Care Medicine, 150(5):1234–1242.
47. Seaton, A., Godden, D., MacNee, W., and Donaldson, K. (1995). Particulate air pollution and acute health effects. The Lancet, 345(8943):176–178.
48. Secretaria del Medio Ambiente (2005). Informe Climatologico Ambiental del Valle de Mexico. Mexico: Gobierno del Distrito Federal.
49. Stoerk, T. (2016). Statistical corruption in Beijing's air quality data has likely ended in 2012. Atmospheric Environment, 127, 365-371.
50. Tanaka, S. (2015). Environmental regulations on air pollution in China and their impact on infant mortality. Journal of Health Economics, 42:90–103.
51. Wang, K., Du, C., and Liu, Y. (2015). Monitoring of illness-induced school absence among primary and secondary school students in Chengdu. Journal of Preventive Medicine Information, 31(9):719–722 (in Chinese).
52. World Health Organization (2000). Air quality guidelines-second edition: Chapter 5.5 Carbon Monoxide.
53. World Health Organization (2005). Effects of air pollution on children's health and development: A review of the evidence.
54. World Health Organization (2006). Air quality guidelines: global update 2005: Particulate matter, ozone, nitrogen dioxide, and sulfur dioxide.
55. World Health Organization (2014). Burden of disease from household air pollution for 2012.
56. Zhang, J., and Mu, Q. (2017). Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks. Journal of Environmental Economics and Management.
57. Zhang, X., Zhang, X., and Chen, X. (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being?. Journal of environmental economics and management, 85, 81-94.

Table I: Sample Statistics

Variable	N	Mean	Std. Dev.	Min.	Max.
Panel (a): Health outcomes					
Total illness rate (%)	3,122,724	0.221	0.731	0.000	20.000
Respiratory illness rate (%)	3,122,724	0.176	0.642	0.000	20.000
Gastrointestinal illness rate (%)	3,122,724	0.015	0.171	0.000	20.000
Panel (b): Absence outcomes					
Total absence rate (%)	3,122,724	0.181	0.655	0.000	20.000
Respiratory-related absence rate (%)	3,122,724	0.140	0.565	0.000	20.000
Gastrointestinal-related absence rate (%)	3,122,724	0.012	0.149	0.000	20.000
Panel (c): Pollution variables					
Air Quality Index (AQI, 10 units)	3,122,724	6.727	3.034	0.258	22.639
$PM_{2.5}$ (10μg/m^3)	3,122,724	4.592	2.520	0.300	18.146
SO_2 (10μg/m^3)	3,122,724	1.674	0.870	0.021	6.986
CO (10mg/m^3)	3,122,724	0.100	0.033	0.002	0.658
Ozone (10μg/m^3)	3,122,724	8.204	4.538	0.080	33.200
Panel (d): Weather variables					
Temperature (°F)	3,122,724	73.888	10.188	49.292	90.396
Dew point (°F)	3,122,724	64.097	13.240	15.729	80.146
Sea level pressure (mb)	3,122,724	1,013.591	6.239	994.862	1,028.863
Wind speed (mph)	3,122,724	5.474	2.579	1.125	16.500
Precipitation (inches)	3,122,724	0.228	0.616	0.000	6.655
Temperature inversion (yes=1)	3,122,724	0.297	0.457	0.000	1.000

Notes: This table provides descriptive statistics for our regression sample, consisting of 12,238 school-grade units in Guangzhou city from January 2013 through November 2015. Each observation represents a school-grade by school day pair (school-grade-date for simplicity). Panel (a) provides information on the illness rates, defined as the proportion of sick students over total students in a given school-grade. Panel (b) provides information on the absence rates, defined as the proportion of absent students over total students in a given school-grade. Panel (c) reports the inverse-distance weighted averages of the 24-hour means of AQI Index, $PM_{2.5}$, SO_2, and CO, and the maximum 8-hour mean of ozone in 10 units. Panel (d) reports the 24-hour means of temperature (°F), dew point temperature (a measure of humidity, °F), sea level pressure (mb), and wind speed (mph), and the 24-hour cumulative precipitation (inches). Temperature inversion is a binary variable, equaling to 1 if the inverse-distance weighted average temperature of the second layer (320 m) is higher than that of the first layer (110 m) during the 24-hour period. Health and absence data were provided by the Guangzhou Center for Disease Control and Prevention. Pollution data come from the Chinese Ministry of Environmental Protection. Weather data are from the National Oceanic and Atmospheric Administration.

Table II: The Effect of Temperature Inversion Occurrence on Air Pollution (The First Stage)

VARIABLES	(1) AQI (10 units)
Temperature inversion	0.83334***
	(0.02714)
Angrist-Pischke F-statistic	943.00
Stock-Yogo crtical value	16.38
Mean of dept.var	6.72725
Number of schools	3,139
Observations	3,122,724
R-squared	0.751
Weather controls	Yes
School-grade FE	Yes
Year FE	Yes
Week FE	Yes
Before/after holiday	Yes
Day-of-week FE	Yes
School-grade×week FE	Yes
Year×week FE	Yes
School-grade×year FE	Yes
Week×day-of-week FE	Yes
School-grade×year×week FE	Yes

Notes: This table presents the OLS estimate on how the occurrence of temperature inversion affects daily AQI levels. Each observation represents a school-grade-date. The specification includes school-grade FE, year FE, week FE, day-of-week FE, dummies indicating one day preceding or following public holidays, school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE. Weather controls contain cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of the overall distribution during the sample period. The Angrist-Piscke F-statistic is reported to assess the strength of the instrument variable. The statistic should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias caused by weak instruments. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table III: The Effect of Air Pollution on Student Health, Estimated by OLS and 2SLS

VARIABLES	Respiratory (%)		Gastrointestinal (%)	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
AQI (10 units)	0.00126***	0.00582***	0.00007	0.00008
	(0.00025)	(0.00169)	(0.00005)	(0.00031)
Mean of dept. var	0.17582	0.17582	0.01527	0.01527
Number of schools	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes

Notes: This table compares the OLS and 2SLS estimates on the effect of air pollution on student health. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variables are the respiratory illness rate (%) for columns (1) and (2), and the gastrointestinal illness rate (%) for columns (3) and (4). All specifications include school-grade FE, year FE, week FE, day-of-week FE, dummies indicating one day preceding or following public holidays, school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE. Weather controls contain cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of the overall distribution during the sample period. In columns (2) and (4), daily AQI levels are instrumented by the occurrence of temperature inversion of the day. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, *at 10%.

Table IV: The Effect of Air Pollution on Student absences, Estimated by OLS and 2SLS

VARIABLES	Total abs (%)		Respiratory abs (%)		Gastrointestinal abs (%)		Non-illness-related abs (%)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	8)
AQI (10 units)	0.00120***	0.00417***	0.00110***	0.00416***	0.00003	-0.00021	0.00003	0.00009
	(0.00026)	(0.00179)	(0.00023)	(0.00151)	(0.00004)	(0.00024)	(0.00012)	(0.00087)
Mean of dept. var	0.18090	0.18090	0.13980	0.13980	0.01236	0.01236	0.01116	0.01116
Number of schools	3,139	3,139	3,139	3,139	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table compares the OLS and 2SLS estimates on the effect of air pollution on student absences. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variables are the total absence rate (%) for columns (1) and (2), the respiratory absence rate (%) for columns (3) and (4), the gastrointestinal absence rate (%) for columns (5) and (6), and the non-illness-related absence rate (%) for columns (7) and (8), respectively. All specifications include the same set of fixed effects and weather controls as reported in the main specification in Table III. In 2SLS regressions, daily AQI levels are instrumented by the occurrence of temperature inversion of the day. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, *at 10%.

Table V: The Effects of Air Pollution on Student Health and Absences, **By Season**, Estimated by 2SLS

VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
AQI (10 units)	0.00638***	0.00487***
	(0.00144)	(0.00158)
Summer*AQI	-0.00308	-0.00385
	(0.00396)	(0.00404)
Angrist-Pischke F-statistic, first stage1	957.55	957.55
Angrist-Pischke F-statistic, first stage2	231.70	231.70
Stock-Yogo critical value	19.93	19.93
Mean of dept.var	0.17582	0.18090
Number of schools	3,139	3,139
Observations	3,122,724	3,122,724
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table examines the heterogeneous effects of air pollution by season. Each regression is estimated by 2SLS and controls for the same set of fixed effects and weather controls as reported in the main specification in Table III. Each observation represents a school-grade-date. The dependent variables are the respiratory illness rate (%) for column (1) and the total absence rate (%) for column (2). We define the summer period from May to October, where the monthly temperature is above 77°F (25°C). To test whether the results differ in warm versus cold periods, we include the interaction between AQI levels and the summer period indicator as a right-hand-side variable. Our instrument set consists of the occurrence of temperature inversion and the interaction between it and the summer period indicator. To assess the strength of the instruments, we report the Angrist-Piscke F-statistic for each of the two first stages of the model. These statistics should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias due to weak instruments. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10\%.

Table VI: The Cumulative Effects of Temperature Inversion on Student Health and Absences, **Reduced Form Analysis**

VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
Zero Lag (β_t)	0.00485***	0.00347**
	(0.00136)	(0.00146)
One Lag ($\beta_t + \beta_{t-1}$)	0.00700**	0.00673**
	(0.00273)	(0.00275)
Two Lags ($\beta_t + \beta_{t-1} + \beta_{t-2}$)	0.00869***	0.00679**
	(0.00280)	(0.00307)
Three Lags ($\sum_{n=0}^{3} \beta_{t-n}$)	0.01832***	0.01631***
	(0.00418)	(0.00420)
Four Lags ($\sum_{n=0}^{4} \beta_{t-n}$)	0.01422***	0.01492***
	(0.00428)	(0.00440)
Five Lags ($\sum_{n=0}^{5} \beta_{t-n}$)	0.01434***	0.01400***
	(0.00438)	(0.00460)
Mean of dept.var	0.17582	0.18090
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table presents the reduced form results for the cumulative effects of the temperature inversion on student health and school absences under a model with distributed lag structures increasing from 0 to 5 days prior to illness and absence incidences. Each cell represents a separate regression. Each observation represents a school-grade-date. The dependent variables are respiratory illness rate (%) in column (1) and total absence rate (%) in column (2). All specifications control for the same weather controls and fixed effects as reported in the main specification in Table III. Each coefficient denotes the sum of the current period (β_t) and lag terms (β_{t-n}). Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, * at 10%.

Table VII: The Cumulative Effect of Air Pollution on Student Health and Absences, Estimated by 2SLS

VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
AQI (10 units):		
Zero Lag (β_t)	0.00582***	0.00417**
	(0.00169)	(0.00179)
One Lag ($\beta_t + \beta_{t-1}$)	0.00706***	0.00778**
	(0.00257)	(0.00348)
Two Lags ($\beta_t + \beta_{t-1} + \beta_{t-2}$)	0.00956***	0.01020**
	(0.00358)	(0.00491)
Three Lags ($\sum_{n=0}^{3} \beta_{t-n}$)	0.02526***	0.02329***
	(0.00514)	(0.00602)
Four Lags ($\sum_{n=0}^{4} \beta_{t-n}$)	0.02248***	0.01938***
	(0.00621)	(0.00671)
Five Lags ($\sum_{n=0}^{5} \beta_{t-n}$)	0.01631***	0.01352**
	(0.00526)	(0.00662)
Mean of dept.var	0.17582	0.18090
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table presents the 2SLS estimates on the cumulative effect of air pollution on student health as well as absences using a model with distributed lag structures increasing from 0 to 5 days prior to illness (absence) incidences. Each cell represents a separate regression. Each observation represents a school-grade-date. The dependent variables are the respiratory illness rate (%) for column (1) and the total absence rate (%) for column (2). The instrument for each lag term of AQI is the corresponding lag term of temperature inversion occurrence. All specifications control for the same set of fixed effects and weather controls as reported in the main specification in Table III. The coefficient is the sum of the current period (β_t) and lag terms (β_{t-n}). Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, * at 10%.

Table VIII: The Nonlinear Effect of Air Pollution on Student Health and Absences, Estimated by 2SLS

VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
AQI:		
50-100% CMEP threshold	0.00125***	0.00089***
	(0.00026)	(0.00025)
Above 100% CMEP threshold	0.00164***	0.00158***
	(0.00012)	(0.00013)
Equal Slopes F-statistic	2.75	4.49
p-value	0.10	0.03
Mean of dept.var	0.17582	0.18090
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table displays the nonlinear effects of air pollution on student health and total absence rate in columns (1) and (2), respectively. Each column represents a separate regression, estimated by 2SLS and controlling for the same set of fixed effects and weather controls as reported in the main specification in Table III. We have two instruments for the two pollution categories, i.e., the occurrence of temperature inversion of the day and the thickness of the inversion. For ease of interpretation, we include a test for equal slopes to examine whether the effect of air pollution on the outcomes differs at high versus low pollution levels. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, * at 10%.

Table IX: The Effect of Air Pollution on Student Health and Absences, **By School Type**, Estimated by 2SLS

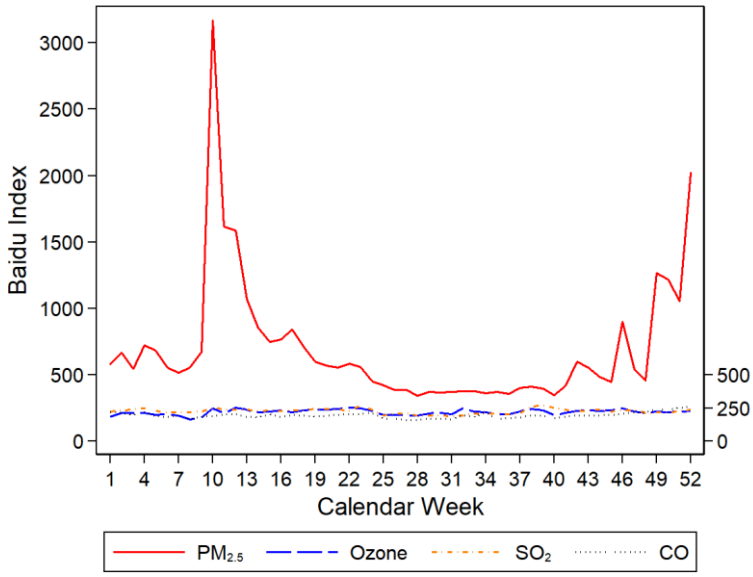
VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
Panel (a): Kindergarten		
AQI (10 units)	0.00564***	0.00356***
	(0.00186)	(0.00136)
Mean of dependent var.	0.23865	0.21700
Number of schools	1,671	1,671
Observations	1,184,119	1,184,119
Panel (b): Primary School		
AQI (10 units)	0.00677***	0.00532***
	(0.00243)	(0.00201)
Mean of dependent var.	0.15258	0.18013
Number of schools	1,075	1,075
Observations	1,483,544	1,483,544
Panel (c): Middle School		
AQI (10 units)	0.00401***	0.00260**
	(0.00147)	(0.00122)
Mean of dependent var.	0.09422	0.0978
Number of schools	439	439
Observations	332,880	332,880
Panel (d): High School		
AQI (10 units)	0.00096**	0.00040**
	(0.00044)	(0.00019)
Mean of dependent var.	0.07158	0.06684
Number of schools	127	127
Observations	122,181	122,181
Weather controls	Yes	Yes
All fixed effects	Yes	Yes

Notes: This table examines whether the effects of air pollution on student health and school absences vary by school type. Each coefficient represents a separate regression, estimated by 2SLS and controlling for the same set of fixed effects and weather controls as reported in the main specification in Table III. The dependent variables in column (1) and (2) are the respiratory illness rate (%) and total absence rate (%), respectively. The instrument for AQI levels is the occurrence of temperature inversion of the day. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

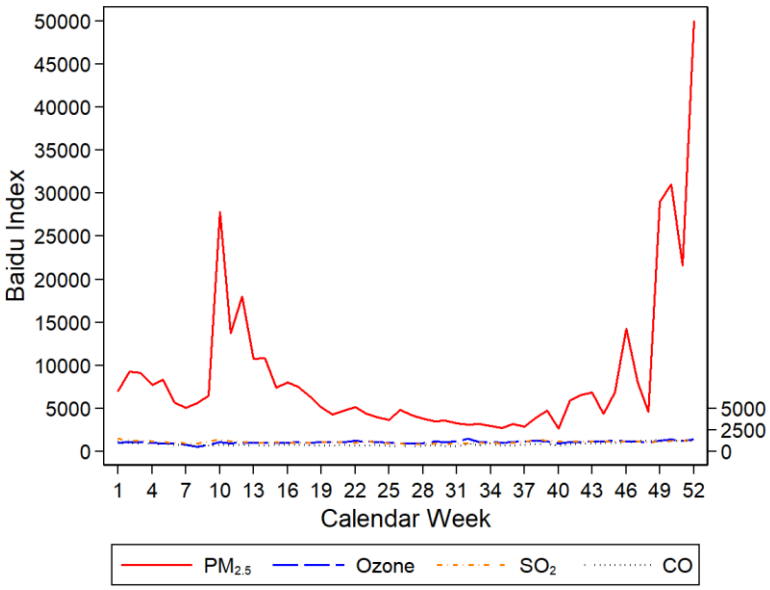
Table X: The Effects of Air Pollution on Student Health and Absences, **By School Quality**, Estimated by 2SLS

VARIABLES	(1) Respiratory illness rate (%)	(2) Total absence rate (%)
AQI (10 units)	0.00604***	0.00445***
	(0.00183)	(0.00197)
Key School*AQI	-0.0014	-0.00177
	(0.00172)	(0.00182)
Angrist-Pischke F-statistic, first stage 1	502.95	502.95
Angrist-Pischke F-statistic, first stage2	328.28	328.28
Stock-Yogo critical value	8.68	8.68
Mean of dept.var	0.17582	0.18090
Number of schools	3,139	3,139
Observations	3,122,724	3,122,724
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table explores whether the effects of air pollution on student health and school absences differ by school quality. The Key School dummy denotes provincial-level key schools of Guangdong province. Each regression is estimated by 2SLS and controls for the same weather controls and fixed effects as reported in the main specification in Table III. The dependent variables are the respiratory illness rate (%) for column (1) and the total absence rate (%) for column (2). To test whether results differ for key versus non-key schools, we include the interaction between pollutant concentrations and the key-school indicator as a right-hand side variable. Our instrument set consists of temperature inversion occurrence and the interaction between it and the key-school indicator. To assess the strength of the instruments, we report Angrist-Piscke F-statistic for each of the two first stages of the model. These statistics should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias due to weak instruments. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

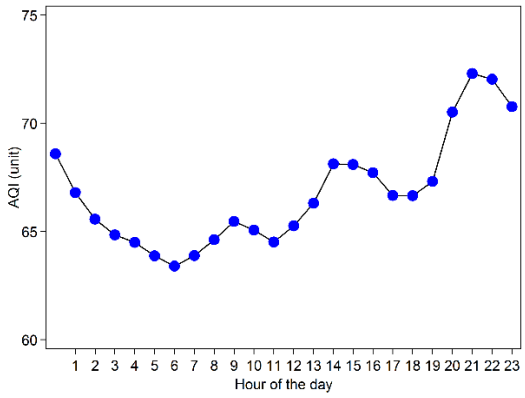


(a) Guangzhou City

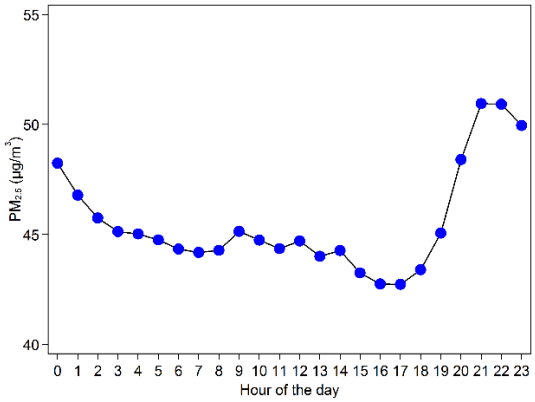


(b) China

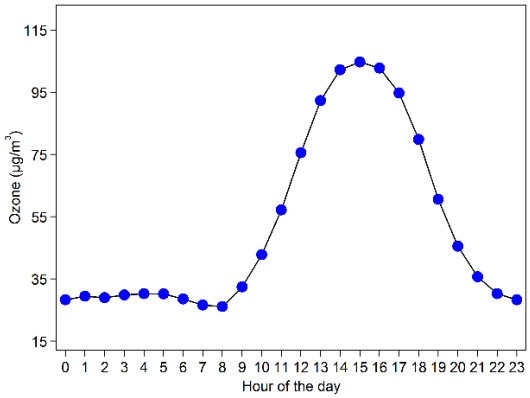
Figure 1: This figure demonstrates the weekly Baidu Index for $PM_{2.5}$, ozone, SO_2, and CO in Guangzhou City (Panel (a)) and in China (Panel (b)) in 2015. Baidu Index, which is similar to Google Trends, is a keyword-analysis tool launched by Baidu, the largest search-engine company in China. It reflects the search frequency of certain keywords on Baidu website.



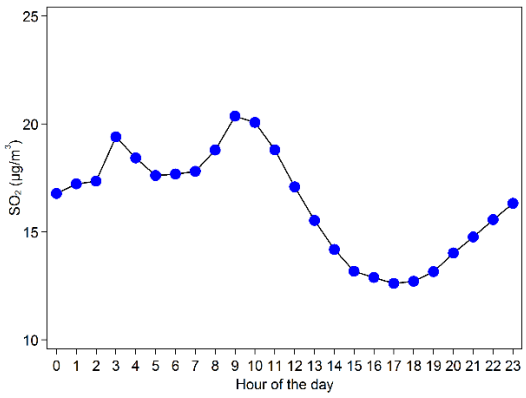
(a) Air Quality Index (AQI)



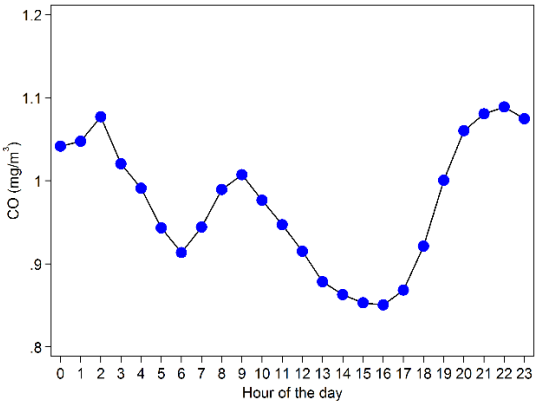
(b) $PM_{2.5}$



(c) Ozone

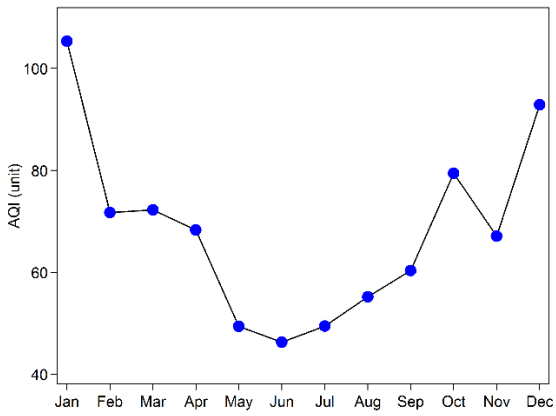


(d) SO_2

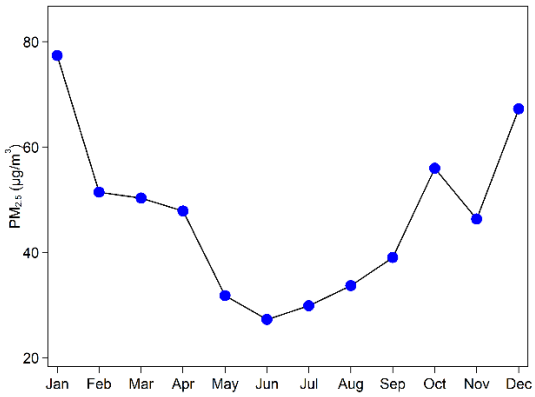


(d) CO

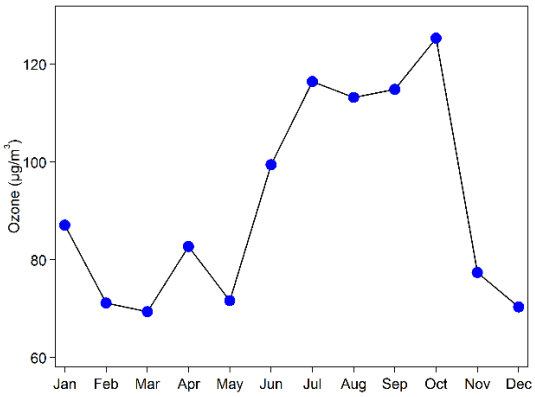
Figure 2: This figure displays the average hourly concentrations of AQI, $PM_{2.5}$, ozone, SO_2, and CO, respectively, in Guangzhou City from January 2013 to November 2015.



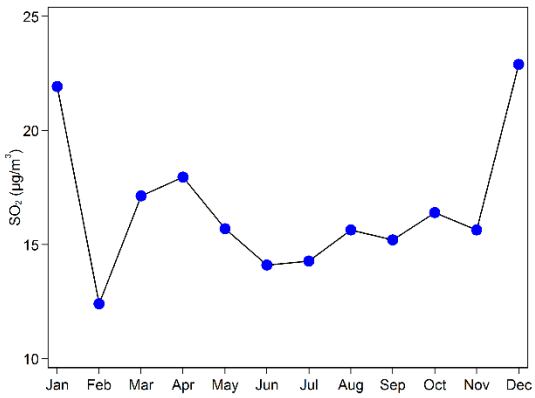
(a) Air Quality Index (AQI)



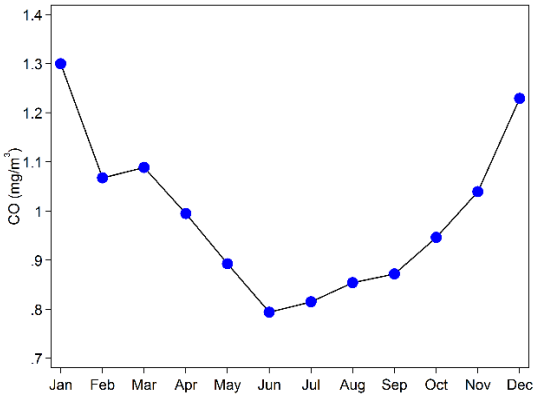
(b) $PM_{2.5}$



(c) Ozone

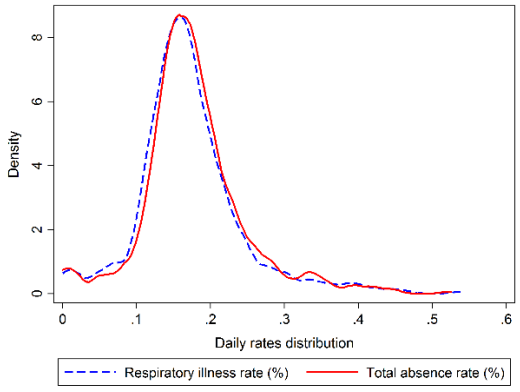


(d) SO_2

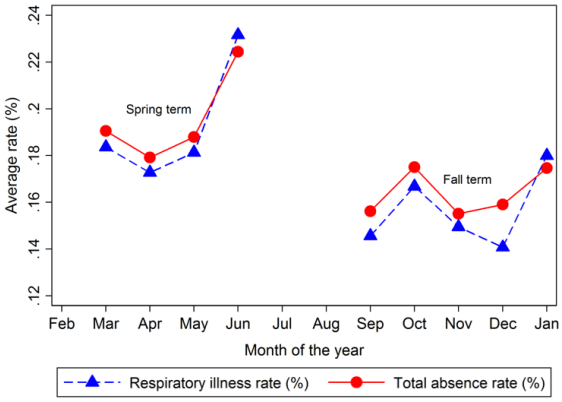


(d) CO

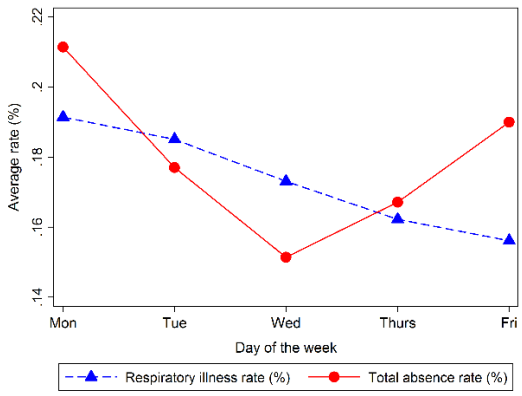
Figure 3: This figure displays the average monthly concentrations of AQI, $PM_{2.5}$, ozone, SO_2, and CO, respectively, in Guangzhou City from January 2013 to November 2015.



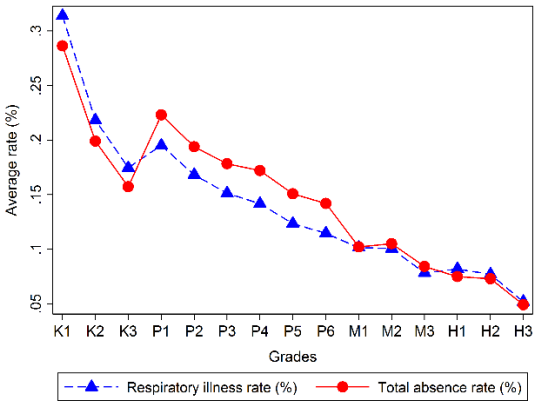
(a) Distribution over School Days



(b) By Month

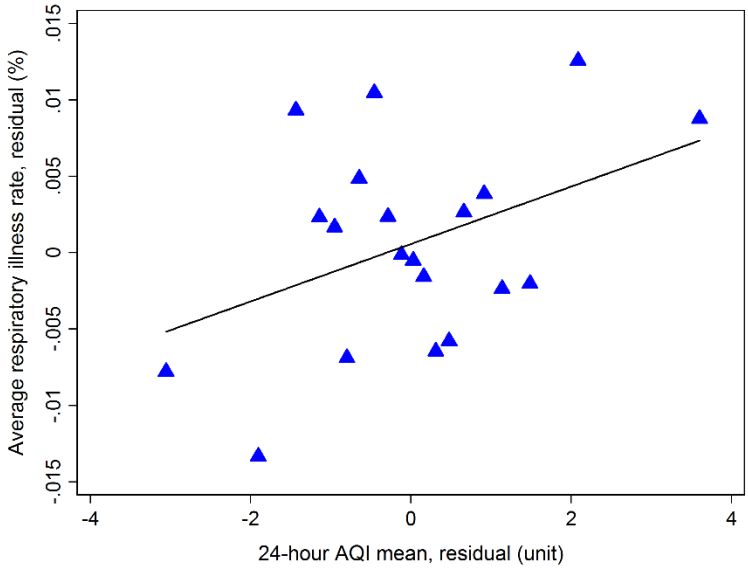


(c) By Day-of-week

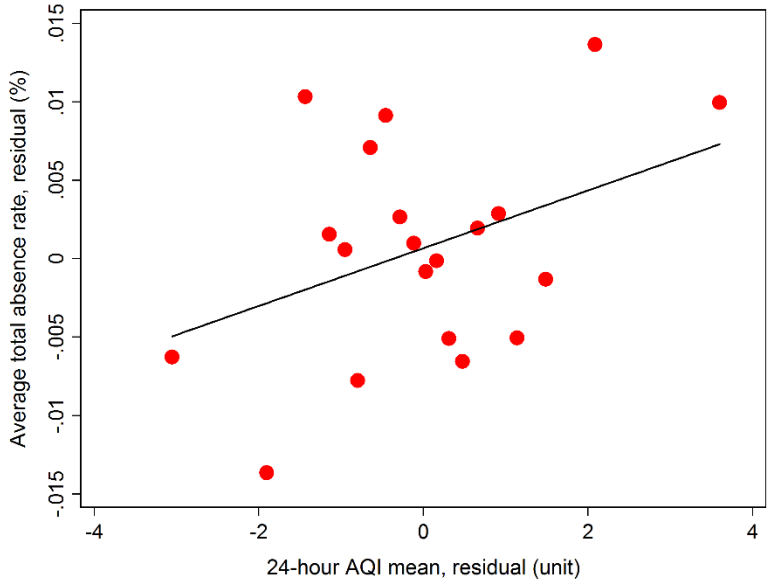


(d) By Grade

Figure 4: This figure summarizes how the respiratory illness rate and total absence rate vary over time and across grades. Panel (a) presents their distributions over school days; Panels (b) to (d) present their variations by month, day-of-the-week, and grade, respectively.

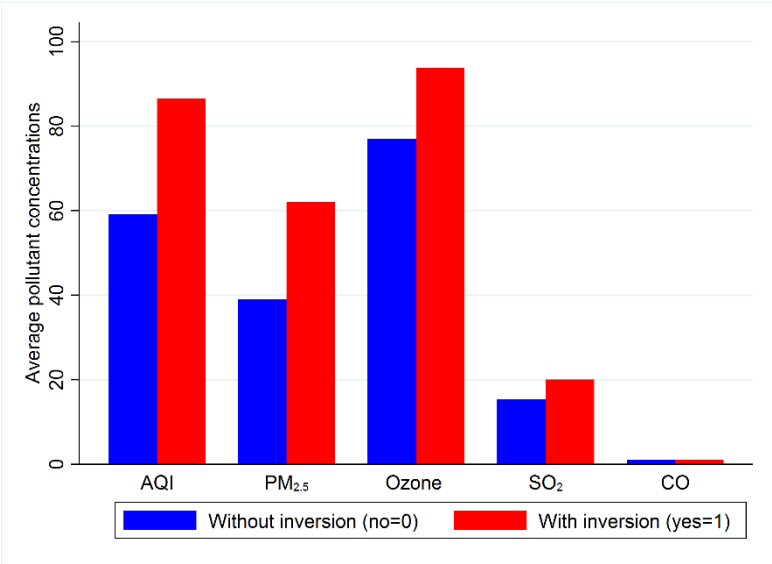


(a) Respiratory Illness Rate v.s. AQ

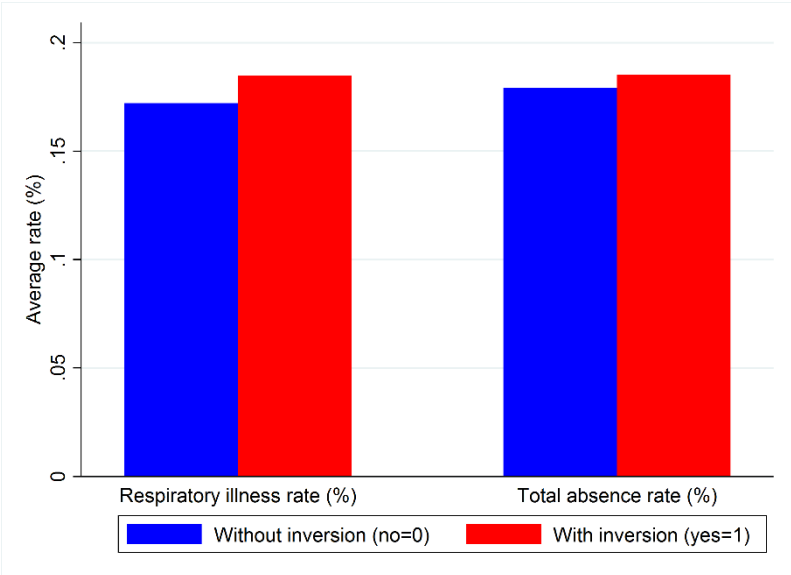


(b) Total Absence Rate v.s. AQI

Figure 5: This figure plots residual respiratory illness rate against residual AQI levels in Panel (a) and residual total absence rate against residual AQI levels in Panel (b). Each dot represents an average school-grade-date within a bin. We first regress the outcome variables and AQI levels on the meteorological factors and the set of fixed effect in the main specification, respectively. We then use the residuals to plot these patterns in binned scatters.



(a) Air Pollutants v.s. Temperature Inversion



(b) Outcomes v.s.Temperature Inversion

Figure 6: Panel (a) plots the average concentrations of air pollutants by the occurrence of temperature inversion. The magnitudes for the blue bars from AQI to CO are 59.12, 39.11, 77.02, 15.35, and 0.96, respectively. The magnitudes for the red bars from AQI to CO are 86.56, 62.05, 99.89, 20.03, and 1.12, respectively. Panel (b) plots the average outcome variables by the occurrence of temperature inversion. When temperature inversion occurs, the average respiratory illness rate increases from 0.172% to 0.184% and the average total absence rate rises from 0.179% to 0.185%.

Table A.I: Correlation Matrix

VARIABLES	(1) AQI	(2) $PM_{2.5}$	(3) Ozone	(4) SO_2	(5) CO	(6) PM_{10}	(7) NO_2
AQI	1.0000						
$PM_{2.5}$	0.9858*	1.0000					
Ozone	0.2952*	0.2464*	1.0000				
SO_2	0.6478*	0.6430*	0.2735*	1.0000			
CO	0.5849*	0.5970*	-0.1258*	0.3557*	1.0000		
PM_{10}	0.9472*	0.9360*	0.2662*	0.6225*	0.5866*	1.0000	
NO_2	0.6797*	0.6819*	0.0287*	0.5036*	0.5358*	0.6972*	1.0000

Notes: This table presents the pair-wise correlation coefficients between AQI, $PM_{2.5}$, ozone, SO_2, $PM_{10,}$ and NO_2 in the main regression sample. Star (*) is specified at the 1% significance level.

Table A.II: Robustness to Alternative Thresholds for Distance of Furthest Pollution Monitor, Estimated by OLS or 2SLS

VARIABLES	Respiratory illness rate (%)				Total absence rate (%)			
	30 km radius		20 km radius		30 km radius		20 km radius	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AQI (10 units)	0.00128***	0.00536***	0.00121***	0.00514***	0.00124***	0.00384**	0.00116***	0.00416**
	(0.00025)	(0.00165)	(0.00024)	(0.00160)	(0.00026)	(0.00173)	(0.00026)	(0.00166)
Mean of dept.var	0.17526	0.17526	0.17440	0.17440	0.18025	0.18025	0.17959	0.17959
Number of schools	3,091	3,091	2,894	2,894	3,091	3,091	2,894	2,894
Observations	3,050,332	3,050,332	2,841,724	2,841,724	3,050,332	3,050,332	2,841,724	2,841,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table tests the robustness of the OLS and 2SLS estimates using various restrictions of the maximum distance from schools to pollution monitors. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variables are the respiratory illness rate (%) for columns (1) to (4) and the total absence rate (%) for columns (5) to (8), respectively. All specifications include the same weather controls and fixed effects as reported in the main specification in Table III. In the 2SLS regressions, AQI levels are instrumented by the occurrence of temperature inversion. The more restrictive distance threshold leads to a smaller sample size as school-grade pairs without an available pollution monitor within a 30 or 20 km radius are dropped. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, *at 10%.

Table A.III: Robustness to Alternative Fixed Effects - The Effect of Air Pollution on Student Health, Estimated by OLS

	Respiratory illness rate (%)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
AQI (10 units)	0.00295***	0.00276***	0.00166***	0.00183***	0.00109***	0.00126***
	(0.00026)	(0.00027)	(0.00029)	(0.00025)	(0.00026)	(0.00025)
Mean of dept. var	0.17582	0.17582	0.17582	0.17582	0.17582	0.17582
Number of schools	3,139	3,139	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×week FE	No	Yes	Yes	Yes	Yes	Yes
Year×week FE	No	No	Yes	Yes	Yes	Yes
School-grade×year FE	No	No	No	Yes	Yes	Yes
Week×day-of-week FE	No	No	No	No	Yes	Yes
School-grade×year×week FE	No	No	No	No	No	Yes

Notes: This table tests the robustness of the OLS estimates on the effect of air pollution on student health across specifications. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variable is the respiratory illness rate (%). Column (1) controls for weather, school-grade FE, year FE, week FE, day-of-week FE, and dummies indicating one day preceding or following public holidays. From columns (2) to (6), we add school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE one by one (in sequence). Weather controls include cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of the overall distribution during the sample period. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, * at 10%.

Table A.IV: Robustness to Alternative Fixed Effects - The Effect of Air Pollution on School Absences, Estimated by OLS

VARIABLES	Total absence rate (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
AQI (10 units)	0.00225***	0.00219***	0.00157***	0.00160***	0.00104***	0.00120***
	(0.00025)	(0.00026)	(0.00028)	(0.00025)	(0.00027)	(0.00026)
Mean of dept. var	0.18090	0.18090	0.18090	0.18090	0.18090	0.18090
Number of schools	3,139	3,139	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×week FE	No	Yes	Yes	Yes	Yes	Yes
Year×week FE	No	No	Yes	Yes	Yes	Yes
School-grade×year FE	No	No	No	Yes	Yes	Yes
Week×day-of-week FE	No	No	No	No	Yes	Yes
School-grade×year×week FE	No	No	No	No	No	Yes

Notes: This table tests the robustness of the OLS estimates of the effect of air pollution on student absences across specifications. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variable is the total absence rate (%). Column (1) controls for weather, school-grade FE, year FE, week FE, day-of-week FE, and dummies indicating one day preceding or following public holidays. From columns (2) to (6), we add school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE one by one (in sequence). Weather controls include cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-5%, 50-75%, and 75-100% of the overall distribution during the sample period. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, **at 5%, * at 10%.

Table A.V: Robustness to Alternative Fixed Effects - The Effect of Air Pollution on Student Health, Estimated by 2SLS

VARIABLES	Respiratory illness rate (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
AQI (10 units)	0.00483***	0.00518***	0.00693***	0.00680***	0.00571***	0.00582***
	(0.00132)	(0.00113)	(0.00155)	(0.00156)	(0.00171)	(0.00169)
Mean of dept. var	0.17582	0.17582	0.17582	0.17582	0.17582	0.01527
Number of schools	3,139	3,139	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×week FE	No	Yes	Yes	Yes	Yes	Yes
Year×week FE	No	No	Yes	Yes	Yes	Yes
School-grade×year FE	No	No	No	Yes	Yes	Yes
Week×day-of-week FE	No	No	No	No	Yes	Yes
School-grade×year×week FE	No	No	No	No	No	Yes

Notes: This table tests the robustness of the 2SLS estimates on the effect of air pollution on student health across specifications. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variable is the respiratory illness rate (%). The instrument for daily AQI levels is the occurrence of temperature inversion of the day. Column (1) controls for weather, school-grade FE, year FE, week FE, day-of-week FE, and dummies indicating one day preceding or following public holidays. From columns (2) to (6), we add school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE one by one (in sequence). Weather controls include cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of the overall distribution during the sample period. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table A.VI: Robustness to Alternative Fixed Effects - The Effect of Air Pollution on Student Absences, Estimated by 2SLS

VARIABLES	Total absence rate (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
AQI (10 units)	0.00366***	0.00398***	0.00551***	0.00570***	0.00393**	0.00417**
	(0.00133)	(0.00118)	(0.00165)	(0.00163)	(0.00180)	(0.00179)
Mean of dept. var	0.18090	0.18090	0.18090	0.18090	0.18090	0.18090
Number of schools	3,139	3,139	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724	3,122,724
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes	Yes	Yes
School-grade×week FE	No	Yes	Yes	Yes	Yes	Yes
Year×week FE	No	No	Yes	Yes	Yes	Yes
School-grade×year FE	No	No	No	Yes	Yes	Yes
Week×day-of-week FE	No	No	No	No	Yes	Yes
School-grade×year×week FE	No	No	No	No	No	Yes

Notes: This table tests the robustness of the 2SLS estimates on the effect of air pollution on student absences across specifications. Each coefficient represents a separate regression. Each observation represents a school-grade-date. The dependent variable is the total absence rate (%). The instrument for daily AQI levels is the occurrence of temperature inversion of the day. Column (1) controls for weather, school-grade FE, year FE, week FE, day-of-week FE, and dummies indicating one day preceding or following public holidays. From columns (2) to (6), we add school-grade×week FE, year×week FE, school-grade×year FE, week×day-of-week FE, and school-grade×year×week FE one by one (in sequence). Weather controls include cumulative precipitation, temperature, dew point temperature (a measure of humidity), sea level pressure, and wind speed in non-parametric forms: 0-25% (the omitted group), 25-50%, 50-75%, and 75-100% of the overall distribution during the sample period. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table A.VII: The Effects of Air Pollution on Student Health and Absences, Based on Single-pollution-index Model, Estimated by 2SLS

	(1)	(2)
VARIABLES	Respiratory illness rate (%)	Total absence rate (%)
Single Pollution Index	0.01305***	0.00935**
	(0.00373)	(0.00399)
Angrist-Pischke F-statistic, first stage	1,369.64	1,369.64
Stock-Yogo critical value	16.38	8.68
Mean of dept.var	0.17582	0.18090
Number of schools	3,139	3,139
Observations	3,122,724	3,122,724
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: In this table, we use the principal components method to construct an index of air pollution. This index summarizes information on all the four air pollutants. We use the occurrence of temperature inversion to instrument for this single endogenous variable. Each column denotes a separate regression, controlling for the same weather controls and fixed effects as reported in the main specification in Table III. The dependent variables are the respiratory illness rate (%) for column (1) and total absence rate (%) for column (2), respectively. Standard errors, in parentheses, are clustered at the community level. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table A.VIII: The Effects of Each Air Pollutant on Student Health, Estimated by OLS and 2SLS

VARIABLES	Single estimation		Joint estimation	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
$PM_{2.5}$ (10 μg/m^3)	0.00138***	0.00645***	0.00078*	0.00958***
	(0.00030)	(0.00187)	(0.00044)	(0.00322)
Ozone (10 μg/m^3)	0.00161***	0.01388***	0.00149***	0.02977***
	(0.00015)	(0.00399)	(0.00017)	(0.00510)
SO_2 (10 μg/m^3)	0.00237***	0.03535***	0.00025	-
	(0.00082)	(0.00969)	(0.00101)	-
CO (10 mg/m^3)	0.00012	0.88982***	-0.01997	-
	(0.01926)	(0.26771)	(0.02366)	-
Mean of dept. var	0.17582	0.17582	0.17582	0.17582
Number of schools	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724
Co-pollutants	**No**	**No**	**Yes**	**Yes**
Weather controls	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes

Notes: This table reports the OLS and 2SLS estimates on the effect of $PM_{2.5}$, ozone, SO_2, and CO on student health. The dependent variable is the respiratory illness rate (%). Each observation represents a school-grade-date. Each coefficient represents a separate regression except for the estimates in column (3), where a single regression is reported. Columns (1) and (2) summarize estimates based on single-pollutant models, with each pollutant in turn being examined without co-pollutants. Column (3) reports the OLS estimates from the multiple-pollutant model, in which we simultaneously include $PM_{2.5}$, ozone, SO_2, and CO into one specification. Column (4) reports the 2SLS estimates, with each pollutant in turn being the one that is instrumented conditional on co-pollutants. To be clear, we include the concentrations of the other pollutants as linear controls in both the first and the second stage regressions of one particular pollutant. Each observation represents a school-grade-date. The instrument for pollution levels is the occurrence of temperature inversion during the same period. All specifications include the same set of fixed effects and weather controls as reported in main specification as reported in Table III. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, *at 10%.

Table A.IX: The Effects of Each Air Pollutant on Student Absences, Estimated by OLS and 2SLS

VARIABLES	Single estimation		Joint estimation	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
$PM_{2.5}$ (10 μg/m^3)	0.00136***	0.00462**	0.00069*	0.00636*
	(0.00031)	(0.00199)	(0.00040)	(0.00340)
Ozone (10 μg/m^3)	0.00159***	0.00995**	0.00149***	0.02678***
	(0.00016)	(0.00432)	(0.00017)	(0.00512)
SO_2 (10 μg/m^3)	0.00235***	0.02532**	0.00023	-
	(0.00087)	(0.01057)	(0.00106)	-
CO (10 mg/m^3)	0.00644	0.63750**	-0.01043	-
	(0.01968)	(0.27719)	(0.02377)	-
Mean of dept. var	0.18090	0.18090	0.18090	0.18090
Number of schools	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724
Co-pollutants	**No**	**No**	**Yes**	**Yes**
Weather controls	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes

Notes: This table reports the OLS and 2SLS estimates on the effect of $PM_{2.5}$, ozone, SO_2, and CO on student absences. The dependent variable is the total absence rate (%). Each observation represents a school-grade-date. Each coefficient represents a separate regression except for the estimates in column (3), where a single regression is reported. Columns (1) and (2) summarize estimates based on single-pollutant models, with each pollutant in turn being examined without co-pollutants. Column (3) reports the OLS estimates from the multiple-pollutant model, in which we simultaneously include $PM_{2.5}$, ozone, SO_2, and CO into one specification. Column (4) reports the 2SLS estimates, with each pollutant in turn being the one that is instrumented conditional on co-pollutants. To be clear, we include the concentrations of the other pollutants as linear controls in both the first and the second stage regressions of one particular pollutant. Each observation represents a school-grade-date. The instrument for pollution levels is the occurrence of temperature inversion during the same period. All specifications include the same set of fixed effects and weather controls as reported in main specification as reported in Table III. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, *at 10%.

Table A.X: The Effect of Temperature Inversion Occurrence on Air Pollution (The First Stage)

	(1)	(2)	(1)	(2)
VARIABLES	$PM_{2.5}$ (10 μg/m^3)	Ozone (10 μg/m^3)	SO_2 (10 μg/m^3)	CO (10 mg/m^3)
Temperature inversion	0.75104***	0.34908***	0.13713***	0.34908***
	(0.02243)	(0.02228)	(0.00535)	(0.00036)
Angrist-Pischke F-statistic	1121.47	1110.24	656.96	231.86
Stock Yogo critical value	16.38	16.38	16.38	16.38
Mean of dept.var	4.59249	8.95822	1.67434	0.10037
Number of schools	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724
R-squared	0.760	0.728	0.704	0.684
Co-pollutants	**No**	**No**	**No**	**No**
Weather controls	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes

Notes: This table summarizes the OLS estimates on the effects of temperature inversion occurrence on the concentrations of $PM_{2.5}$, ozone, SO_2, and CO. Each column represents a separate regression. Each observation represents a school-grade-date. All specifications include the same weather controls and a full set of fixed effects as reported in the main specification in Table III. The Angrist-Piscke F-statistics are reported to assess the strength of the instrument variable. These statistics should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias caused by weak instruments. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table A.XI: The Effect of Temperature Inversion Occurrence on Air Pollution (The First Stage)

	(1)	(2)	(1)	(2)
VARIABLES	$PM_{2.5}$ (10 μg/m^3)	Ozone (10 μg/m^3)	SO_2 (10 μg/m^3)	CO (10 mg/m^3)
Temperature inversion	0.43861***	0.29839***	-0.01150	0.00016
	(0.01710)	(0.02292)	(0.00786)	(0.00015)
Angrist-Pischke F-statistic	657.98	169.45	2.14	6.67
Stock-Yogo critical value	16.38	16.38	16.38	16.38
Mean of dept.var	4.59249	8.95822	1.67434	0.10037
Number of schools	3,139	3,139	3,139	3,139
Observations	3,122,724	3,122,724	3,122,724	3,122,724
R-squared	0.868	0.758	0.786	0.781
Co-pollutants	**Yes**	**Yes**	**Yes**	**Yes**
Weather controls	Yes	Yes	Yes	Yes
School-grade FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Before/after holiday	Yes	Yes	Yes	Yes
Day-of-week FE	Yes	Yes	Yes	Yes
School-grade×week FE	Yes	Yes	Yes	Yes
Year×week FE	Yes	Yes	Yes	Yes
School-grade×year FE	Yes	Yes	Yes	Yes
Week×day-of-week FE	Yes	Yes	Yes	Yes
School-grade×year×week FE	Yes	Yes	Yes	Yes

Notes: This table summarizes the OLS estimates on the effects of temperature inversion occurrence on the concentrations of $PM_{2.5}$, ozone, SO_2, and CO. Each observation represents a school-grade-date. All specifications control for co-pollutants as well as a full set of fixed effect and weather variables as reported in Table II. To be clear, when estimating the link from temperature inversion to $PM_{2.5}$, we include ozone, SO_2, and CO concentrations as linear controls. We repeat the exercise for each case from column (2) to (4). The Angrist-Piscke F-statistics are reported to assess the strength of the instrument variable. These statistics should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias caused by weak instruments. Standard errors, in parentheses, are clustered at the community level. ***Significant at 1%, ** at 5%, * at 10%.

Table A.XII: The Effect of Temperature Inversion Occurrence and The Effect of The Thickness of Inversion on Air Pollution (The First Stage)

VARIABLES	AQI 50-100	AQI >100
	(1)	(2)
Single Pollution Index	0.00232***	0.03766***
	(0.00076)	(0.00037)
Thickness	-0.18498***	0.00932***
	(0.00543)	(0.00262))
Angrist-Pischke F-statistic, first stage1	1,885.69	1,885.69
Angrist-Pischke F-statistic, first stage2	28,368.94	28,368.94
Stock-Yogo critical value	7.03	7.03
Mean of dept.var	0.60408	0.06201
Number of schools	3,139	3,139
Observations	3,122,724	3,122,724
R-squared	0.5346	0.5567
Weather controls	Yes	Yes
School-grade FE	Yes	Yes
Year FE	Yes	Yes
Week FE	Yes	Yes
Before/after holiday	Yes	Yes
Day-of-week FE	Yes	Yes
School-grade×week FE	Yes	Yes
Year×week FE	Yes	Yes
School-grade×year FE	Yes	Yes
Week×day-of-week FE	Yes	Yes
School-grade×year×week FE	Yes	Yes

Notes: This table presents the OLS estimates on how the occurrence of temperature inversion and the thickness of temperature inversion affects each of AQI categories (50-100 and 100 above). Each observation represents a school-grade-date. All specifications include the full suite of fixed effects and weather controls as reported in the main specification in Table III. To assess the strength of the instruments, we report the Angrist-Piscke F-statistic for each of the two first stages of the model. These statistics should be above the Stock-Yogo critical value for a single endogenous regressor listed below in order to rule out more than 10% bias due to weak instruments. ***Significant at 1%, ** at 5%, * at 10%.

Figure A.1: The log-in page of the Guangzhou Student Health Monitoring System. Visit the page at http://xsjk.gzcdc.org.cn/gzjkjc/.

附表：

The "Morning Check" Record Table

_____学校托幼机构学生健康监测日报表

班级 Class	姓名 Name	性别 Gender	年龄 Age	是否缺勤		缺勤原因			是否出现如下症状										就诊情况		是否住宿	
				是	否	因病	因事	其它	发热（温度）	咳嗽	咽痛	腹泻	呕吐	眼结膜充血	皮疹	黄疸	腮腺肿大	伤害	其它	是（诊断结果）	否	是（房号）	否

Absenteeism? Reasons? Symptom Description Hospital Diagnosis Boarder?

填报人：
Reporter:

填报时间： 年 月 日
Date: (YY/MM/DD)

Figure A.2: This figure presents the morning check table to be filled out by school nurses. They are required to fill in the demographic characteristics of the sick or the absent student, including age, gender, class number, grade, school address, and dormitory number in case of residence in the campus, as well as information on illness and absence, such as date, symptoms, hospital diagnosis, reason of the potential request for absence.

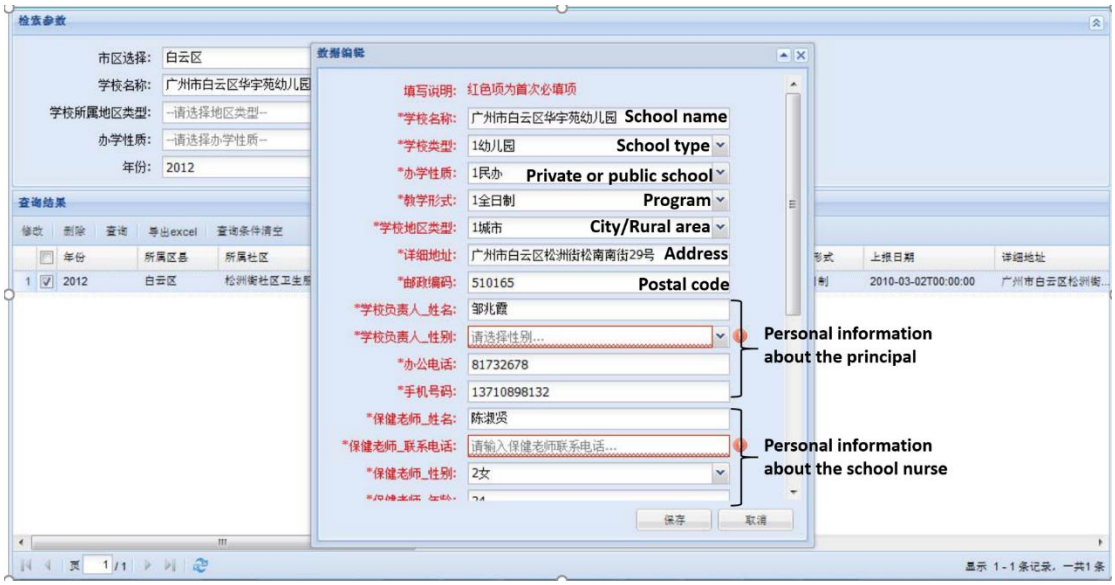


Figure A.3: This figure presents an electronic record sample containing all the information collected from the morning checks.

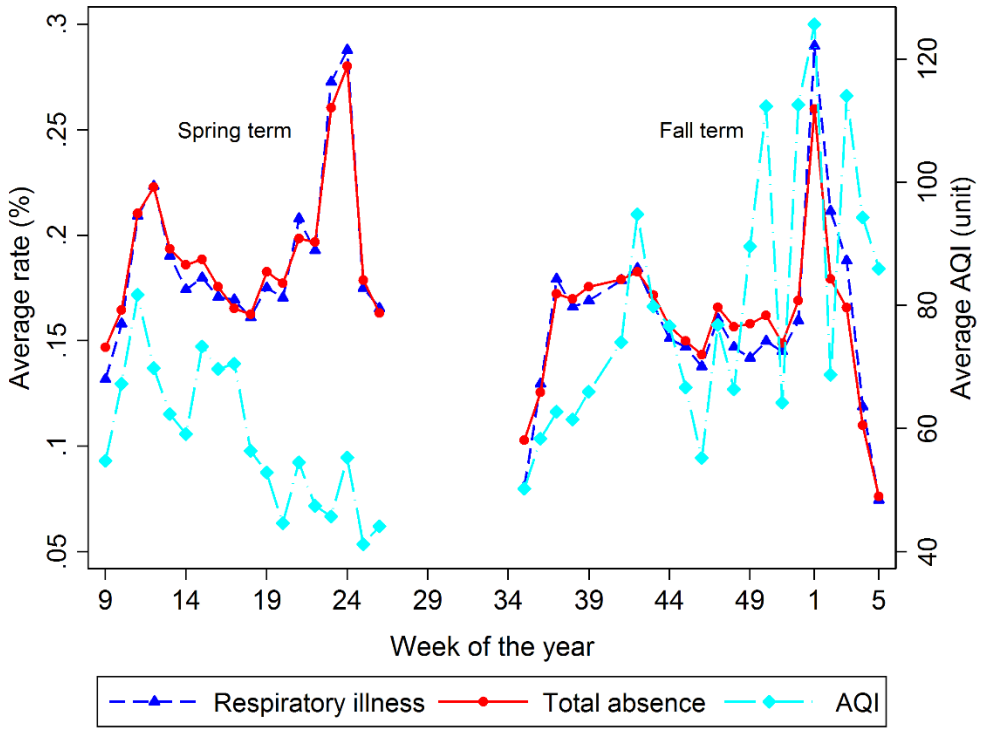


Figure A.4: This figure plots the weekly fluctuations of average daily respiratory illness and total absence rates among students against average daily AQI levels. We arrange the weeks according to the school calendar in Guangzhou City. Spring term starts from week 9 to week 26 of the year, equivalent to from March to June. Fall term starts from week 35 to week 5 of the next year, equivalent to from September till the next January.