Review of Economic Studies (2016) **83**, 768–809 doi:10.1093/restud/rdv043

Advance access publication 20 October 2015

Airports, Air Pollution, and Contemporaneous Health

WOLFRAM SCHLENKER
Columbia University and NBER

and

W. REED WALKER
University of California at Berkeley and NBER

First version received November 2012*; final version accepted July* 2015 (*Eds.*)

We link daily air pollution exposure to measures of contemporaneous health for communities surrounding the twelve largest airports in California. These airports are some of the largest sources of air pollution in the US, and they experience large changes in daily air pollution emissions depending on the amount of time planes spend idling on the tarmac. Excess airplane idling, measured as residual daily taxi time, is due to network delays originating in the Eastern US. This idiosyncratic variation in daily airplane taxi time significantly impacts the health of local residents, largely driven by increased levels of carbon monoxide (CO) exposure. We use this variation in daily airport congestion to estimate the population dose-response of health outcomes to daily CO exposure, examining hospitalization rates for asthma, respiratory, and heart-related emergency room admissions. A one standard deviation increase in daily pollution levels leads to an additional $540 thousand in hospitalization costs for respiratory and heart-related admissions for the 6 million individuals living within 10 km (6.2 miles) of the airports in California. These health effects occur at levels of CO exposure far below existing Environmental Protection Agency mandates, and our results suggest there may be sizable morbidity benefits from lowering the existing CO standard.

Key words: Health effects of pollution, Airport congestion, Network delays, Instrumental variables.

JEL Codes: Q53, J1, C26

The effect of pollution on health remains a highly debated topic. The US Clean Air Act (CAA) requires the Environmental Protection Agency (EPA) to develop and enforce regulations to protect the general public from exposure to airborne contaminants that are known to be hazardous to human health. In January 2011, the EPA decided against lowering the existing CAA carbon monoxide standard due to insufficient evidence that relatively low carbon monoxide (CO) levels adversely affect human health. In order to assess the benefits of lowering the standard, accurate estimates are needed that link contemporaneous air pollution exposure to observable health outcomes at levels of pollution currently faced by local populations. However, these estimates are hard to come by as pollution is rarely randomly assigned across individuals, and individuals who live in areas of high pollution may be in worse health for reasons unrelated to pollution. Preferences for clean air may covary with unobservable determinants of health (*e.g.* exercise), which can lead to various forms of omitted variable bias in regression analysis. Moreover, heterogeneity

across individuals in either preference for, or health responses to, ambient air pollution implies that individuals may self-select into locations on the basis of these unobserved differences. In both cases, estimates of the health effects of ambient air pollution may reflect the response of various subpopulations and/or spurious correlations pertaining to omitted variables. While recent research attempts to address the issue of non-random assignment using various econometric tools such as fixed effects or instrumental variables, these studies often focus on infant health over longer periods of time (Chay and Greenstone, 2003; Currie and Neidell, 2005). Much less is known about short-term, daily effects of ambient air pollution on the health of the more general population, such as the non-elderly, non-child, adult population.[1]

We develop a framework for estimating the contemporaneous effect of air pollution on health using variation in local air pollution driven by airport runway congestion. Airports are one of the largest sources of air pollution in the US with Los Angeles International Airport (LAX) being the largest source of carbon monoxide in the state of California (Environmental Protection Agency, 2005). A large fraction of airport emissions come from airplanes, with the largest aggregate channel of emissions stemming from airplane idling (Transportation Research Board, 2008). We show that airport runway congestion, as measured by the total time planes spent taxiing between the gate and the runway, is a significant predictor of local pollution levels. Since local runway congestion may be correlated with other determinants of pollution such as weather, we exploit the fact that California airport congestion is driven by network delays that began in large airports outside of California.[2] A recent article in the New York Times (New York Times 27 January, 2012) provides a useful motivation:

> [Airplane] delays ripple across the country. A third of all delays around the nation each year are caused, in some way, by the New York airports, according to the F.A.A. Or, as Paul McGraw, an operations expert with Airlines for America, the industry trade group, put it, 'When New York sneezes, the rest of the national airspace catches a cold'.

Our analysis hence links health outcomes of residents living near California airports to changes in air pollution driven by runway congestion at airports on the East Coast. The identifying variation in California pollution is caused by events several thousand miles away (*e.g.* weather in Atlanta), which is unlikely to be correlated with determinants of health in California.

The goal of this article is to identify the ways in which short run, daily variation in air pollution affects population health. In doing so, this article makes four primary contributions to the existing literature in this area. First, while most existing literature focuses on the health impacts of infants or elderly, we are able to examine the health responses of the entire population. We find that

1. There is a larger literature in epidemiology which focuses on daily responses to air pollution (see *e.g.* Ito *et al.* (2007), Linn *et al.* (1987), Peel *et al.* (2005), Schildcrout *et al.* (2006), Schwartz *et al.* (1996)). The work in our article complements the existing epidemiological literature by focusing on issues pertaining to measurement error, avoidance behaviour, and self-selection bias in the context of susceptibility to pollution exposure. Each of these issues is critically important to providing unbiased estimates of the causal relationship between pollution and health. The instrumental variables approach in this article exploits arguably exogenous pollution shocks that are unlikely to be known by local residents, allowing us to simultaneously address issues of measurement error and avoidance behaviour. Recent work in economics and environmental health, discussed in more detail below, suggests that short run variation in pollution exposure may be significant predictors of mortality and morbidity (Moretti and Neidell, 2011; Knittel *et al.*, 2011).

2. This relationship is well known within the transportation literature (Welman *et al.*, 2010). Optimal airplane scheduling incorporates anticipated ripple effect. For example, Pyrgiotisa *et al.* (2013) use queuing theory to simulate how delays propagate through the system. They quote a study that found a multiplier effect of seven, *i.e.*, each 1-hour delay of a particular airplane leads to a combined 7 hours delay for the airline.

infants as well as the elderly are most sensitive to ambient air pollution. At the same time, a one-unit increase in pollution has much larger aggregate effects for adults aged 20–64 years, given their large share of the overall population. Studies that focus on infants or the elderly significantly underestimate overall health effects.

The second contribution of this article is to estimate the contemporaneous effect of multiple pollutants simultaneously. It has traditionally been difficult to decipher which pollutant is responsible for adverse health outcomes since short-term fluctuations among ambient air pollutants are highly correlated. Our solution to this identification problem is to rely on the fact that wind speed and wind direction transport individual pollutants in different ways. By using interactions between taxi time, wind speed, and wind angle from airports, we can pin down the direct effect of each pollutant, while holding the others constant. We use over-identified models to instrument for several pollutants simultaneously, an approach that was simultaneously developed in related work by Knittel *et al*. (2011). We find that CO is responsible for the majority of the observed increase in hospital admissions, although we cannot rule out that this may be driven by other unobserved pollutants that are correlated with airplane-driven CO emissions. This finding has direct policy implications. The EPA recently decided to maintain the current CO pollution standard, citing a lack of evidence that reducing CO below current ambient levels would improve population health outcomes.

We believe there are two additional features that set this article apart from existing work in both economics and epidemiology. Our article is most closely related to the recent work of Moretti and Neidell (2011) and Knittel *et al*. (2011) who also instrument daily pollution in health regressions with variation in local transportation conditions (*i.e*. container shipping in Long Beach, CA, and automobile congestion in Central and Southern CA, respectively). Relative to these article and the existing literature, we believe this paper is the first to use the network structure of transportation to generate local variation in congestion that is driven by events that occur several thousand miles away. This matters because one of the key drivers of transportation congestion is local weather, and local weather is also likely to affect ambient pollution, violating the identifying assumptions of the model. By way of example, we show that instrumenting local airport congestion with network delays that are not correlated with local weather doubles our point estimates, relative to the baseline case. We also explicitly model the spatial dispersion of air pollution emissions, as it varies with wind speed, wind direction, and distance from the airport. Pollutant transport is very locally heterogeneous, and failing to account for this spatial heterogeneity leads to bias when estimating the population dose-response function.

The fourth contribution of this study is the use of newly available Emergency Discharge Data to better capture the morbidity impacts of air pollution. Previous research has predominantly focused on the effects of pollution on mortality or morbidity as measured in Inpatient Discharge Records. Inpatient Discharge data consist only of observations for patients that stayed overnight in a hospital, and thus exclude a large fraction of respiratory-related emergency room admissions that do not require overnight hospital visits. We show that estimates using the more commonly used Inpatient Discharge data substantially underestimate the morbidity impacts of air pollution, relative to estimates from the combined Emergency Discharge and Inpatient Discharge datasets.[3]

In summary, our approach combines newly available data with arguably exogenous daily changes in air pollution that originate several thousand miles away and are unknown to the

3. Knittel *et al*. (2011) focus on infant mortality, while Moretti and Neidell (2011) examine morbidity outcomes, but only for individuals of Emergency Room visits that eventually get admitted to an overnight stay as the authors rely on the Inpatient Discharge data.

local population. The instrumental variables setting allows us to simultaneously address issues pertaining to both avoidance behaviour and classical forms of measurement error, each of which lead to significant downward bias in conventional dose-response estimates. The primary estimation framework examines how zip code level emergency room admissions covary with these quasi-experimental increases in air pollution stemming from airports.

We find that a one standard deviation increase in daily pollution explains roughly one third of average daily admissions for asthma problems. It leads to an additional $540 thousand per day in hospitalization costs for respiratory and heart related admissions of individuals within 10 km of one of the twelve largest airports in California. This is likely a significant lower bound of the social costs as the willingness to pay to avoid a sickness might be significantly larger than the medical reimbursement cost (Grossman, 1972). Our baseline IV estimates are an order of magnitude larger than uninstrumented fixed effects estimates, highlighting the importance of accounting for measurement error and/or avoidance behaviour in conventional estimators. We find no evidence that airport runway congestion affects diagnoses unrelated to air pollution such as bone fractures, stroke, or appendicitis. We also present a variety of evidence in favour of a non-linear dose-response function. As pollution levels increase the marginal effect of a 1 unit increase in pollution increases but at a decreasing rate. This is consistent with thresholds at which the health effect of air pollution levels off (*i.e.* the dose-response function is not convex over the levels of pollution we observe), and along the lines of what research in epidemiology has observed (Pope *et al*., 2009; Pope III *et al*., 2011).

We present several sensitivity checks of results that do not alter our conclusions. For example, we focus on morning airport congestion in the East since it is possible that California airport delays impact airports on the East Coast, which then feedback to California airports. Due to the difference in time zones, very few flights from California reach East Coast airports before 12 pm. Estimates remain similar to our baseline estimates. A distributed lag model finds no evidence for delayed impacts or forward displacement, *i.e*., that individuals on the brink of an asthma or heart attack may experience an episode that would have otherwise occurred in the next few days anyway. A Poisson model linking sickness counts to pollution levels gives comparable estimates to our baseline linear probability model, which does not account for the truncation of daily sickness rates at zero. Finally, we find little evidence of treatment effect heterogeneity that would raise concerns pertaining to forms of self-selection bias and/or the external validity of the underlying dose-response estimates.

The findings in this article suggest that daily variation in ambient air pollution has economically significant health effects at levels below current EPA mandates, at least for the population that comprises our study. We believe this is particularly important due to the fact that in January 2011, the EPA decided against lowering the existing CAA CO standard due to insufficient evidence that relatively low CO levels adversely affect human health. The maximum hourly CO concentration in our data is 7.5 ppm (see Supplementary Table A2), which is below the ambient air quality standard of 35 ppm for any 1-hour reading or 9 ppm for any 8-hour average, *i.e*., air quality levels were always within the limit.[4] Yet, fluctuations in pollution levels significantly below the standard still have sizable health consequences. While a full-fledged benefit–cost analysis would have to balance the cost of reducing CO against the benefits, EPA indicated there were no appreciable benefits from lowering the standard to begin with, which we find not to be the case.

4. The same is not true for NO_2. The maximum 1-hour reading in our data is 136 ppb, which is above the 1-hour standard of 100 ppb.

1. BACKGROUND: AIRPORTS, AIRPLANES, AND AIR POLLUTION

Regulators have long been aware of the pollution generated by cars, trucks, and public transit. There have been countless legislative policies designed to curtail harmful emissions from these sources (Auffhammer and Kellogg, 2011). However, aircraft and airport emissions have only recently become the subject of regulatory scrutiny, although little has been done to reduce or manage emissions generated by airports and air travel. While there has been some effort to curtail the substantial CO_2 emissions generated by aircraft,[5] there has been relatively little effort to control or contain some of the more pernicious air pollutants generated by jet engines. This lack of regulatory scrutiny can be traced back to the way in which pollutants are regulated in the US under the Clean Air Act. Current Federal law preempts all federal, state, and local agencies except the Federal Aviation Administration from establishing measures to reduce emissions from aircraft due to potential interstate and international commerce conflicts that might arise from other decentralized regulations.[6]

Aircraft jet engines, like many other mobile sources, produce carbon dioxide (CO_2), nitrogen oxides (NO_x), carbon monoxide (CO), oxides of sulphur (SO_x), unburned or partially combusted hydrocarbons (also known as volatile organic compounds, or VOCs), particulates, and other trace compounds (Federal Aviation Administration, 2005). Each of these pollutants is emitted at different rates during various phases of operation, such as idling, taxing, takeoff, climbing, and landing. NO_x emissions are higher during high power operations like takeoff when combustor temperatures are high. On the other hand, CO emissions are higher during low power operations like taxiing when combustor temperatures are low and the engine is less efficient (Federal Aviation Administration, 2005).[7] Even though the aircraft engine is often idling during taxi-out, the per minute CO and NO_x emissions factors are higher than at any other stage of a flight (Environmental Protection Agency, 1992). Combining this with the long duration of taxi-out times during peak periods of the day, total taxiing over the course of a day can add up to a substantial amount. Consistent with these facts, Los Angeles International airport is estimated to be the largest point source of CO emissions in the state of California, the second largest of NO_x, the twenty-ninth largest of SO_2, and the 2,763 and 2,782 largest of PM_{10} and $PM_{2.5}$, respectively (Environmental Protection Agency, 2005).

Airports provide a particularly compelling setting through which to estimate the contemporaneous relationship between air pollution and health. Not only are airports some of the largest polluters of ambient air pollution in the US but they also have extraordinarily rich data on daily operating activity, detailing for each domestic flight the length of time spent taxiing to and from the gate before takeoff and after landing. This allows for a precise understanding of the aggregate amount of daily runway congestion at airports. Moreover, daily runway congestion at airports exhibits a great degree of residual variation even after controlling for normal scheduling patterns. Much of the variation in runway congestion is driven by network delays propagating from major airport hub delays thousands of miles away. Network delays at distant airports serve as an ideal instrumental variable for local pollution; the effect of a snow storm in Chicago on congestion at LAX should be orthogonal to any other confounding influences of air pollution in the Los Angeles

5. The European Union has recently approved greenhouse gas measures, which oblige airlines, regardless of nationality, that land or take off from an airport in the European Union to join the emissions trading system starting on 1 January, 2012.

6. Currently, the EPA has an agreement with the FAA to voluntarily regulate ground support equipment at participating airports known as the Voluntary Airport Low Emission (VALE) program (United States Environmental Protection Agency, 2004).

7. As a result, reducing engine power for a given operation like takeoff or climb out generally increases the rate of CO emissions and reduces the rate of NO_x emissions.

area. In addition, local residents are likely unaware of increases in taxi time and hence cannot engage in self-protective behaviour. Finally, every airport has detailed weather data, allowing researchers to exploit the spatial distribution of airport-generated pollution. We can therefore estimate how areas downwind of an airport on a given day are disproportionately affected by runway congestion relative to areas upwind. Understanding this spatial variation in pollutant transport improves the efficiency of our estimates, while also providing important tests of the validity of our research design.

2. DATA

This project uses the most comprehensive data currently available on airport traffic, air pollution, weather, and daily measures of health in California. These data are rich in both temporal and spatial dimension, allowing for fine-grained analysis of how daily airport congestion impacts areas downwind of an airport on a given day. The various datasets and linkages are described in more detail below.

2.1. *Airport traffic data*

A useful feature of a study involving airports is the detailed nature of daily flight data. The Bureau of Transportation Statistics (BTS) Airline On-Time Performance Database contains flight-level information by all certified US air carriers that account for at least 1% of domestic passenger revenues. It has a wealth of information on individual flights: flight number, the origin and departure airport, scheduled departure and arrival times, actual departure and arrival times, the time the aircraft left the runway and when it touches down. We construct a daily congestion measure for each of the twelve major airports in California by aggregating the combined taxi time of all airplanes at an airport. This measure consists of (1) the time airplanes spend between leaving the gateway and taking off from the runway and (2) the time between landing and reaching the gate. An interesting feature of aggregate daily taxi time is the large amount of residual variation remaining after controlling for daily airport scheduling, weather, and holidays. We relate this variation to local measures of pollution and health in our econometric analysis. One caveat of the BTS data is that it only includes information for major domestic airline passenger travel.[8] As long as international flights are not treated differently in the queuing system and are hence colinear to the taxi time of domestic flights, congestion of national flights should be a good proxy for overall congestion.

We limit our analysis to the twelve largest airports in California by passenger count. These airports are in alphabetical order (including airport call sign in brackets): Burbank (BUR), Los Angeles International (LAX), Long Beach (LGB), Oakland International (OAK), Ontario International (ONT), Palm Springs (PSP), San Diego International (SAN), San Francisco International (SFO), San Jose International (SJC), Sacramento International (SMF), Santa Barbara (SBA), and Santa Ana / Orange County (SNA). The locations of these airports are shown as dots in Figure 1. Average flight statistics at each of these airports are reported in Supplementary Table A1. There is significant variation in daily ground congestion at airports: the standard deviation of daily taxi time at the largest airport (LAX) is 1,852 minutes. Once we account for year, month, weekday and holiday fixed effects as well as local weather, the remaining variation is still 891 minutes. Most of the airports are close to urban areas as they serve the travel needs of these populations.

8. In January 2005, international departures (both cargo and passenger) accounted for 8.5% of total departures, whereas cargo (both international and domestic) accounted for 5.9% of all US airport departures (Department of Transportation, 2009).

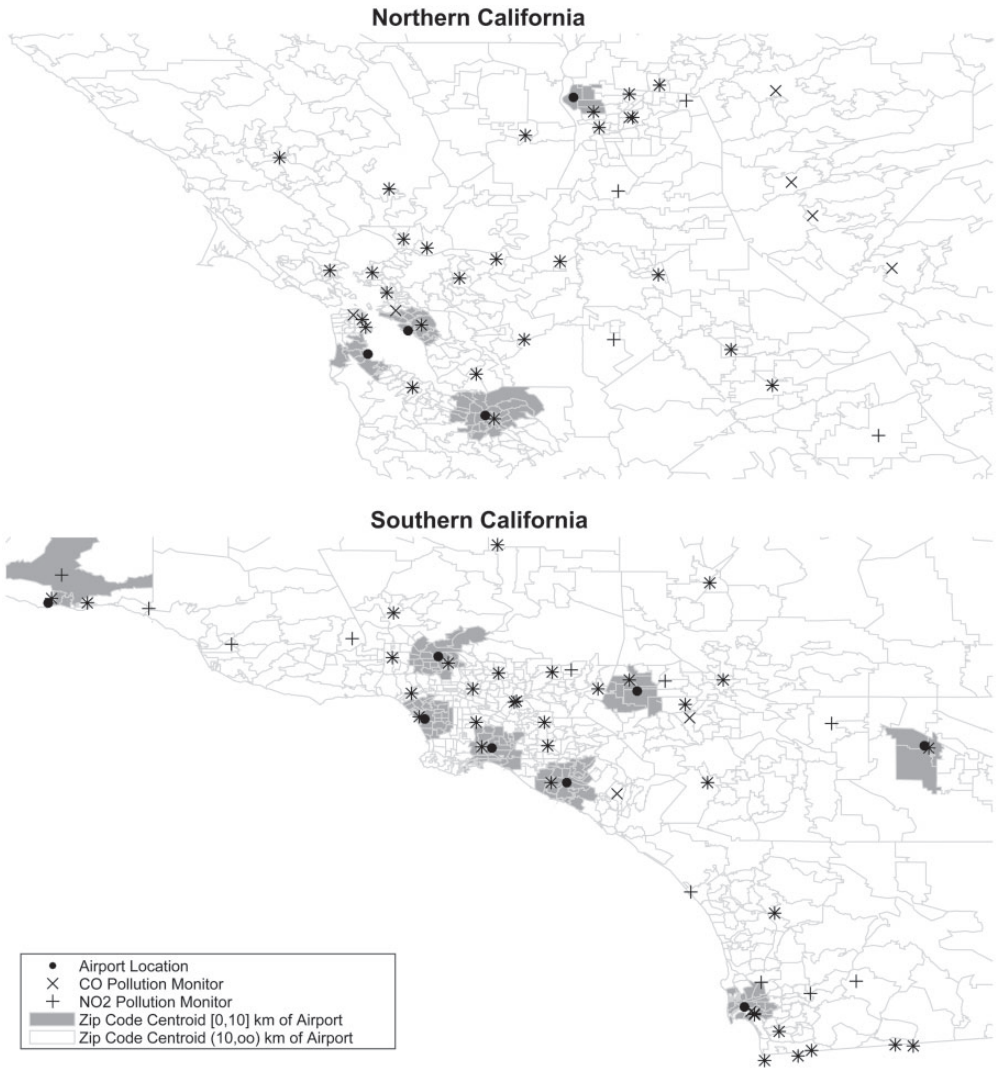

FIGURE 1

Location of airports, pollution monitors, and zip codes.

Notes: The 12 largest airports in California are shown as dots. The location of CO pollution monitors in the California Air Resource Board (CARB) data base are shown as **x**, the location of NO_2 monitors as **+**. Zip code boundaries are shown in grey. They are shaded if the centroid is within 10 km (6.2 miles) of an airport.

Seven airports in California rank among the top fifty busiest airports in the nation according to passenger enplanement (Federal Aviation Administration, 2009).

A potential concern when linking daily airport activity to daily ambient air pollution levels is that runway congestion in California airports may be highest in the late afternoon and evening. This would lead us to erroneously misclassify some of the daily airport effects to the wrong day. Supplementary Figure A2 plots the distribution of aggregate taxi time within a day. Most ground activity at airports is skewed towards the beginning of the day. We will address the sensitivity

of our estimates towards these issues of misclassification or across-day spillovers in subsequent sections.

2.2. *Pollution data*

We construct daily measures of air pollution surrounding airports using the monitoring network maintained by the California Air Resource Board (CARB). This database combines pollution readings for all pollution monitors administered by CARB, including information on the exact location of the monitor. Data includes both daily and hourly pollution readings. We concentrate on the set of monitors with hourly emission readings for CO, NO_2, and O_3 in the years 2005–7.[9] The locations of all CO and NO_2 monitors in relation to airports are shown in Figure 1.

A unique feature of pollution data is the significant number of missing observations in the database. We therefore use the following algorithm when we aggregate the hourly data to daily pollution readings: Our measure of the daily maximum pollution reading is simply the maximum of all hourly pollution readings. The daily mean is the duration-weighted average of all hourly pollution readings. We define the duration as the number of hours until the next reading.[10] We prefer this approach to simply taking the arithmetic average of all hourly readings on a day since hourly pollution data exhibit great temporal dependence. A missing hourly observation is better approximated by the previous non-missing value than the daily average. We also keep track of the number of observations per day. In a sensitivity check (not reported) we rerun the analysis using only monitors with at least 20 or 12 readings per day.[11]

We create daily zip code pollution measures by taking the average monitor reading of all monitors within 15 km of a zip code centroid, weighting by the inverse distance between the monitor and the zip code centroid.[12] Summary statistics are given in Panel A of Supplementary Table A2. Since we have both the longitude and latitude of all airports and zip code centroids, we are able to derive (1) the distance between the airport and a zip code, and (2) the angle at which the zip code is located relative to the airport. In order to leverage the spatial features of our data, we normalize the angle between a zip code centroid and an airport to 0 if the zip code is lying to the north of the airport. Degrees are measured in clockwise fashion, *e.g.* a zip code that is directly east of an airport will have an angle of 90°. The angle between an airport and a zip code allows us to explore the link between airport emissions and pollution downwind of airports using the weather data described next.

9. While data exist for other pollutants in California, we limit our analysis to using CO, NO_2 as they are directly emitted by airplanes and have better coverage than PM10. O_3 forms from VOC and NO_x. In a sensitivity check we do not find that O_3 pollution levels are impacted by airport congestion. Nevertheless, we present sensitivity analyses that include O_3 and PM10 as controls with little effect on our results. While monitor data exists as far back as 1993, portions of our hospital data, described further in this section, exists only from 2005 onwards.

10. Readings occur on the hour of each day ranging from midnight to 11 pm. If readings at the beginning of a day (midnight, 1 am, etc.) are missing, we adjust the duration of the first reading from midnight to the second reading. For example, if readings occur on 3 am, 5 am, and 8 am, the 3 am reading would be assigned a duration of 5 hours and the 5 am reading would be assigned a duration of 3 hours. By the same token, if the last reading of a day is not 11pm, the duration of that last reading is from the time of the reading until midnight.

11. If a monitor has not a single reading for a day, we approximate its value in a three step procedure: (1) we derive the cumulative density function (cdf) at each monitor; (2) take the inverse-distance weighted average of the cdf for a given day at all monitors with non-missing data; (3) we fill the missing observation with the same percentile of the station's cdf. For example, if surrounding monitors with non-missing data on average have pollution levels that correspond to the 80th percentile of their respective distributions, we fill the missing value of a station with the 80th percentile of its own distribution of pollution readings. This procedure gives us a balanced panel.

12. Inverse distance weighting pollution measures has been used to impute pollution in previous research. See for example, Currie and Neidell (2005).

2.3. *Weather data*

We use temperature, precipitation, and wind data in our analysis to both control for the direct effects of weather on health (Deschênes *et al*., 2009) and also to leverage the quasi-experimental features of wind direction and wind speed in distributing airport pollution from airports. Our weather data comes from Schlenker and Roberts (2009), which provides minimum and maximum temperature as well as total precipitation at a daily frequency on a 2.5×2.5 mile grid for the entire US.[13] To assign daily weather observations to an airport or zip code, we use the grid cell in which the zip code centroid is located. Summary statistics for the zip-code level data are given in Panel B of Supplementary Table A2.

Average wind speed and wind direction come from the National Climatic Data by the National Oceanic and Atmospheric Administration's (NOAA) hourly weather stations. Most airports have weather stations with hourly readings. We construct wind direction, which is normalized to equal zero if the wind is 'blowing' northward and counted in clockwise fashion. If the angles of the zip code and the wind direction are identical, the zip code is hence exactly downwind from the airport. An angle of 180° implies that the zip code is upwind from the airport. The hourly wind speed and wind direction is aggregated to the daily level by calculating the duration-weighted average between readings comparable to the pollution data above. The distribution of wind directions is shown in Figure 2. Airports at the ocean predominantly have winds coming from the direction of the ocean. For example, Santa Barbara, located on the only portion of the California coast that runs east–west has winds blowing northward. Note again that we are measuring the direction in which the wind is blowing, not from which it is coming. In our empirical analysis, we use this daily variation in wind speed and wind direction to predict how pollution from airports disproportionately impacts some zip codes more than others on a given day.

2.4. *Hospital discharge and emergency room data*

Health effects are measured by overnight hospital admission and emergency room visits to any hospital in the state of California. We use the California Emergency Department & Ambulatory Surgery data set for the years 2005–7.[14] The dataset gives the exact admission date, the zip code of the patient's residence (as well as the hospital), the age of the patient, as well as the primary and up to twenty-four secondary diagnosis codes. An important limitation of the Emergency Department data is that any person who visits an ER and is subsequently admitted to an overnight stay drops out of the dataset. This is done to prevent double counting in California's hospital admissions records, as overnight hospital stays are logged in California's Inpatient Discharge data. Therefore, we also obtained Inpatient Discharge data for all individuals who stayed overnight in a hospital in the years 2005–7. In our baseline model, we focus on the sum of emergency room visits and overnight stays in a zip code-day to avoid non-random attrition in the ER data. Focusing only on emergency room admittance would suffer from selection bias as higher pollution levels (and more severe health outcomes) could result in more overnight stays, yet the emergency room numbers would actually appear smaller.

We count the daily admissions of all people in a zip code who had a diagnosis code pertaining to three respiratory illnesses: asthma, acute respiratory, and all respiratory. Note that each category

13. There is one exception: in a set of regression models where we estimate the effect of airport weather on taxi time we use the closest non-missing daily weather station data from NOAA's COOP station data set for each airport. This is because Schlenker and Roberts (2009) use a spatial interpolation procedure that might result in artificial correlation between weather data at airports due to the spatial interpolation technique.

14. The Emergency Room data was not collected prior to 2005.

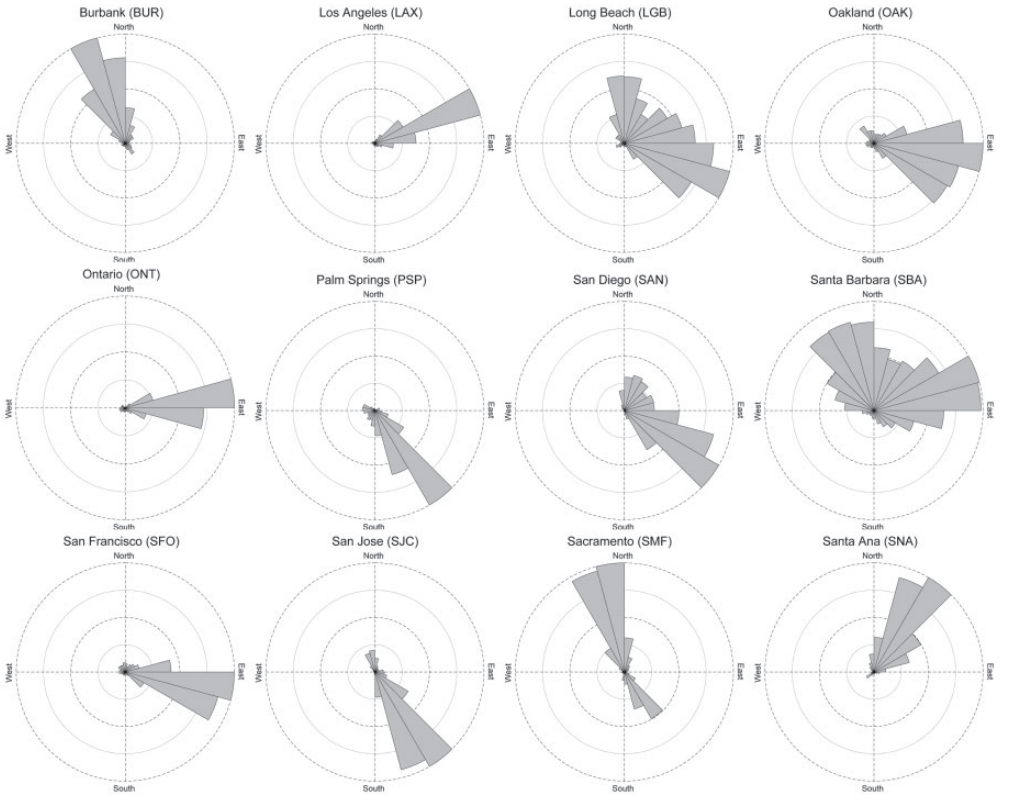

FIGURE 2

Histogram of daily wind direction at airports.

Notes: Histogram of the distribution of daily directions in which the wind is blowing (2005–7). Plot is normalized to the most frequent category. The four circles indicate the quartile range. Airport locations are shown in Figure 1.

adds additional sickness counts but includes the previous. For example, asthma attacks are also counted in all respiratory problems. We also count heart-related problems, which Peters *et al*. (2001) have shown to be correlated with pollution. Finally, we include three placebos: stroke, bone fractures, and appendicitis.[15] In our baseline model, we count a patient as suffering from a sickness if either the primary or one of the secondary diagnosis codes lists the illness in question.

We merge the zip code level hospital data with age-specific population counts in each zip code obtained from both the 2000 and 2010 Censuses. We use the weighted average between the 2000 (weight 0.4) and 2010 (weight 0.6) counts, as the midpoint of our data is 2006. We limit our analysis to the 164 zip codes whose centroid lies within 10 km of an airport and which have at least 10,000 inhabitants.[16] The total population of these 164 zip codes is around 6 million people, or roughly one sixth of the overall population of California at the time. Summary statistics for the zip codes in the study are given in Panel C of Supplementary Table A2. We use

15. The exact ICD-9 codes are: asthma: [493, 494); acute respiratory: [460,479), [493,495), [500,509), [514,515), [516,520); all respiratory: [460, 520); heart problems: [410, 430); stroke [430, 439); bone fractures [800, 830); appendicitis: [540, 544).

16. The latter sample restriction excludes 0.8% of the total population that lives in a zip code whose centroid is within 10 km of an airport but has less than 10,000 inhabitants.

these age-specific population counts to construct daily hospitalization rates for each zip code. Supplementary Table A3 provides sickness rates per 10 million inhabitants for both the entire population as well as population subgroups of those over 65 years of age and under 5 years of age.

2.5. *External validity — populations close to airports*

Our analysis focuses on areas within 10 km of airports. This raises the broader question as to how our estimated results generalize to populations outside of the 10 km airport radius. Table A4 investigates this question by examining zip code characteristics from the 2000 Census. We present three comparisons: First, we look at zip codes that are in our sample in columns (1a)–(1c) but divide them into zip codes whose centroids are within [0,5] km and (5,10] km of an airport. Second, we compare zip codes within 10 km of an airport versus neighbouring zip codes that are between 10 km and 20 km of an airport in columns (2a)–(2c). Third, we compare zip codes within 10 km of an airport to all other zip code in California in columns (3a)–(3c).

For the first two sets of comparisons, few comparison tests are significant, roughly at a rate that should happen due to randomness. In other words, areas [0,5] km from an airport are comparable to areas (5,10] km or (10,20] km.[17] On the other hand, the third set of comparisons shows that areas within 10 km are not comparable to the rest of the state of California, which includes more rural areas. Zip codes closer to airports are on average more urban, more populated, wealthier, and have higher housing prices. Therefore, we would caution against interpreting the estimated dose–response relationship as representative for the entire population at large. From the standpoint of airport externalities, the population close to airports is the population of interest. Moreover, much of the air pollution regulation in the US is spatially targeted towards urban areas (*i.e.* those areas with higher degrees of ambient air pollution), and in that case, these estimates may be more appropriate for regulatory analysis than a dose response function averaged over individuals in both urban and rural locations.

3. EMPIRICAL METHODOLOGY

We are estimating the link between ground level airport congestion, local pollution levels, and contemporaneous hospitalization rates for major airports in the state of California. To begin, we consider the effects of increased levels of airport traffic congestion on local measures of pollution.

3.1. *Aggregate daily taxi time and local pollution levels*

Ambient air pollution is a function of the distance between a point source and the receptor location, as well as many other atmospheric variables including, but not limited to, wind speed, wind direction, humidity, temperature, and precipitation. To model the effects of increases in aggregate airport taxi time on pollution levels, we adopt the following additive linear regression model

$$\textbf{Model 1:}\quad p_{zat} = \alpha_1 T_{at} + \underbrace{\mathbf{W}_{zt}\boldsymbol{\Phi} + weekday_t + month_t + year_t + holiday_t}_{\mathbf{Z}_{zt}\Gamma} + v_{za} + e_{zat}, \qquad (1)$$

17. In all, 47% of Californians live in a zip code within 20 km of an airport.

where pollution p_{zat} in zip code z that is paired with airport a on day t is specified as a function of taxi time T_{at} and a vector of zip code level controls $\mathbf{Z}_{zt}$ that include weather controls $\mathbf{W}_{zt}$.[18] Our baseline regressions include seventeen weather controls: a quadratic in minimum and maximum temperature, precipitation, and wind speed (eight terms) as well as nine terms for wind direction that are included in equation (3) below.[19] To model this relationship formally, we define wind direction by the cosine of the difference between the wind direction and the direction in which the zip code is located. The variable will be equal to 1 in the case that the angle in which the wind is blowing equals the direction in which the zip code is located, and the variable will be equal to zero when they are at a right angle (the difference is 90°). The vector $\mathbf{W}_{zt}$ includes all possible time-varying interactions between distance, wind speed and angle (up and downwind) to control for pollution formation not directly influenced by taxi time. We also control for temporal variation in pollution by including weekday fixed effects ($weekday_t$), month fixed effects ($month_t$), and year fixed effects ($year_t$) as well holiday fixed effects ($holiday_t$) to limit the influence of airport congestion outliers.[20] In a sensitivity check (available upon request), we instead include day fixed effects, *i.e.*, one for each of the 1,095 days, and the results remain robust. Since there may be time-invariant unobserved determinants of pollution for any given zip code, all regressions include zip code fixed effects, ν_{za}. The parameter of interest is α_1, which tells us the effect of a 1,000 minute increase in aggregate daily ground congestion on local ambient air pollution levels. Increased airplane taxiing leads to an increase in airplane emissions and presumably increases in ambient air pollution. Hence, we would expect this coefficient to be positive.

We also estimate models similar to equation (1), where we interact taxi time (or instrumented taxi time) with the distance between an airport and the monitor. The idea would be to allow the marginal effect of taxi time to differ based on monitors that were closer relative to further from the airport. This results in the following equation:

$$\textbf{Model 2:} \quad p_{zat} = \alpha_1 T_{at} + \alpha_2 T_{at} d_{za} + \mathbf{Z}_{zt}\Gamma + \nu_{za} + e_{zat}. \tag{2}$$

The additional coefficient is α_2. The effect of taxi time on pollution should fade out with distance, and we would hence expect this coefficient to be negative. The marginal effect of taxi time in model 2 is $\alpha_1 + \alpha_2 d_{za}$.

In a third step we also include interactions with wind direction and wind speed. The intuition is that both wind direction and speed transport airport emissions across space. Thus, holding speed constant, areas downwind should be relatively more affected by aggregate daily taxi time relative to areas upwind. To model this relationship formally, we let v_{at} be the wind speed and c_{zat} the cosine of the difference between the wind direction and the direction in which the zip code is located, which can differ upwind $c_{zat} > 0$ and downwind $c_{zat} < 0$.[21] Allowing for all possible

18. In principle, a zip code z could be paired with more that one airport a. In practice, our baseline model uses zip codes whose centroid is within 10 km of an airport. Each zip code is assigned to exactly one airport as none is within 10 km of two airports.

19. Specifically, our weather controls include the terms corresponding to α_3, α_4 and $\alpha_6 - \alpha_{12}$ in equation (3) without the interaction with taxi time T_{at}. Results are robust to different functional forms of weather control variables. Additionally, we have estimated models that exclude all weather controls, and the coefficients for our primary pollutant of interest (CO see below) are not significantly affected (although the standard errors increase).

20. We include fixed effects for New Year, Memorial Day, July 4th, Labor Day, Thanksgiving, and Christmas, as well as the three days preceding and following the holiday.

21. The cosine is 0 if the angle is 90°, *i.e.*, the separately estimated effect is different upwind and downwind.

time-varying interactions we get:[22]

$$\begin{aligned}\textbf{Model 3:}\quad p_{zat} = {} & \alpha_1 T_{at} + \alpha_2 T_{at} d_{za} + \alpha_3 T_{at} c_{zat} I_{[c_{zat}>0]} + \alpha_4 T_{at} c_{zat} I_{[c_{zat}<0]} \\ & + \alpha_5 T_{at} v_{at} + \alpha_6 T_{at} d_{za} c_{zat} I_{[c_{zat}>0]} + \alpha_7 T_{at} d_{za} c_{zat} I_{[c_{zat}<0]} \\ & + \alpha_8 T_{at} d_{za} v_{at} + \alpha_9 T_{at} c_{zat} I_{[c_{zat}>0]} v_{at} + \alpha_{10} T_{at} c_{zat} I_{[c_{zat}<0]} v_{at} \\ & + \alpha_{11} T_{at} d_{za} c_{zat} I_{[c_{zat}>0]} v_{at} + \alpha_{12} T_{at} d_{za} c_{zat} I_{[c_{zat}<0]} v_{at} \\ & + \mathbf{Z}_{zt}\Gamma + \nu_{za} + e_{zat}. \end{aligned} \tag{3}$$

The new coefficients are α_3 through α_{12}. The predicted signs of these coefficients are less intuitive. While higher wind speeds can clear the air they may also carry greater amounts of the pollutant further distances.[23] Moreover, downwind areas should have higher pollution levels relative to those areas upwind, but aircrafts usually start against the wind. To better interpret the combination of all of these interactions, we plot the marginal effects of this particular regression model using contour plots in subsequent sections. These contour plots provide strong visual evidence of the relationship between daily aggregate airport taxi time, wind speed, wind direction, and local pollution levels.

One potential cause for concern in equations (1)–(3) are any omitted transitory determinants of local pollution levels that may also covary with ground congestion. If such omitted variables exist, then least squares estimates of the coefficients on taxi time (*e.g.* α_1) will be biased. This could occur, for example, if weather adversely affected airport activity while also affecting local pollution levels. To address this potential source of bias, we need an instrumental variable that is correlated with changes in ground congestion at an airport but is unrelated to local levels of pollution. A natural instrument comes from delays at major airport hubs outside California, which propagate through the air network as connecting flights are delayed, leading to more ground congestion at airports in California. The basic logic is that instead of smoothing out scheduling over the course of the day, planes now arrive in more distinct blocks of time, leading to more waiting/taxiing by those planes taking off as the runway space is shared. Specifically, we instrument taxi time at each California airport with taxi time at major airports outside of California (Atlanta (ATL), Chicago O'Hare (ORD), and New York John F. Kennedy (JFK)), in the following first stage regression:[24]

$$T_{at} = \alpha_{a0} + \sum_{k=1}^{3}\sum_{a=1}^{12} \alpha_{ak} T_{kt} I_a + \mathbf{Z}_{at}\Theta + \omega_{at}. \tag{4}$$

Supplementary Figure A1 shows the location of those airports in relation to the California airports. We allow the coefficients α_{ak} in equation (4) to vary by airport a by interacting taxi time with an airport indicator I_a. These interactions allow for heterogeneity in the impact of delays from major airports outside of California T_{kt} on each of the California airports T_{at}. This is important

22. The exact dispersion of pollution depends on additional factors like acceleration and height of emissions. Benson (1984) presents a formal model of pollution dispersion around roads that includes many variables we do not observe. The standard in the literature has hence been to estimate reduced-form relationship with wind direction and speed (Batterman *et al*., 2010).

23. Recall that we are already controlling for overall wind speed in $\mathbf{W}_{zt}$, but it has so far not been interacted with taxi time or any other weather measure.

24. These airports were chosen because they are among the largest airports in the country, they serve different regions, and they are subject to different weather systems. The results are robust to different airport specifications.

as the impact of delays in Atlanta on California airports is likely to differ across airports. Our baseline model utilizes thirty-six instruments (three airports outside California interacted with each of the twelve airports in California).[25] We use two-way cluster robust standard errors for inference, clustering on both zip code and day. The two-way cluster robust variance–covariance estimator implicitly adjusts standard errors to properly account for both spatial correlation across zip codes on a given day, which are all due to the same network delays, as well as within-zip code serial correlation in air pollution over time.[26]

The standard conditions for consistent estimation of α_1 in the context of our 2SLS estimator are that $\alpha_{ak} \neq 0$ in equation (4) and $\mathbb{E}[T_{kt} \cdot e_{azt} \mid \mathbf{Z}_{zt}, \nu_{za}] = 0$. Subsequent sections will show that the first condition clearly holds; taxi time at airports on the East Coast leads to large increases in taxi time at California airports. The second condition requires that the error term in the instrumental variable regression be uncorrelated with taxi time at major airports outside of California, T_{kt}. This condition would be violated if ground congestion in Chicago somehow co-varied with pollution levels in California through reasons unrelated to California airport congestion due to network delays.

While the second condition is not explicitly testable, our data and research design permit several indirect tests. First, we show evidence that taxi time in California is predicted by weather fluctuations at airports inside and outside of California, but the reverse is not true: weather at the major airports in California has no significant effect on taxi time at Eastern airports. Second, we show that network delays propagate East to West rather than West to East. Taxi time in Atlanta is not higher due to increased taxi time in Los Angeles.[27] Further sensitivity checks show that using only taxi time before noon at Eastern Airports or directly instrumenting with observed weather variables at airports in the Eastern US has little impact on our baseline estimates. In the following sections we use the variation in California airport taxi time, and the spatial distribution of emissions from an airport, as a predictor of local air pollution measures in order to better understand contemporaneous relationships between elevated levels of air pollution and hospital admissions.

3.2. *Aggregate daily taxi time, local pollution, and health*

To estimate the pollution–health association in our data we begin by assuming that the relationship between health and ambient air pollution can be summarized by the following linear model:

$$y_{zat} = \beta p_{zat} + \mathbf{Z}_{zt}\Pi + \eta_{za} + \epsilon_{zat}, \tag{5}$$

where the dependent variable y_{zat} is our observable measure of health in zip code z when paired with airport a on day t.[28] The remaining notation is consistent with the previous models, $\mathbf{Z}_{zt}$ are the same weather and time controls and η_{za} is a zip code fixed effect.

25. Model 2 instruments both T_{at} and $T_{at}d_{az}$ with the taxi time outside California T_{kt} and $T_{kt}d_{az}$, and thus uses seventy-two instruments. Similarly, model 3 instruments all twelve interaction of taxi time T_{at} at the twelve airports by the taxi time at the three largest airports outside California T_{kt}, which results in $12 \times 12 \times 3 = 432$ instruments.

26. Standard errors clustering on both airport and day tend to be smaller than those using zip code and day. We choose the latter when conducting inference, as they tend to be the more conservative of the two. Results with airport and day clustering are available upon request.

27. This issue is largely addressed by the difference in time zones between our instrumental variable airports and California. Airplane traffic in the US generally starts around 6 am in the morning and slows down in the evening. Due to the change in time zones, a flight that leaves at LAX in the morning to go to one of the airports does not reach of the three airports outside California before noon. On the other hand, a flight that leaves at 6 am on the East Coast will reach California by 9 am.

28. Our analysis implicitly assumes that we can summarize health responses and behaviour at the zip code level and that the effect of interest, β, is stable over time and across airports.

We focus primarily on respiratory-related hospital admissions as defined by International Statistical Classification of Diseases and Related Health Problems ICD-9 (Friedman *et al*., 2001; Seaton *et al*., 1995). The dependent variable y_{zat} is the number of admissions to either the emergency room or an overnight hospital stay where either the primary or one of the secondary diagnosis code fell in one of the following admission categories: asthma, acute respiratory, all respiratory, or heart-related diagnoses. These daily zip code counts are scaled by zip code population so that the dependent variable represents hospitalization rates per 10 million zip code residents. We also estimate models for diagnoses unrelated to pollution: strokes, bone fractures, and appendicitis. These outcomes are meant to serve as an important test for the internal validity of our research design. Since these health outcomes are unrelated to pollution exposure, they should not be significantly related to changes in pollution.

The coefficient of interest in this model is β which provides an estimate of the effect of a one unit increase in pollution levels on daily hospitalization rates in zip code z and time t. Consistent estimation of β requires $\mathbb{E}[p_{zat} \cdot \epsilon_{zat} \mid \mathbf{Z}_{zt}, \eta_{za}] = 0$. The inclusion of a zip code fixed effect implicitly controls for any time invariant determinants of local health that also covary with average pollution levels. For example, if relatively disadvantaged households live in more polluted areas and have poorer health for reasons unrelated to air pollution, then the zip code fixed effect will control for this time-invariant unobserved heterogeneity. However, least squares estimation of β will be biased if there are time-varying influences of both health and pollution (*e.g.* weather), and/or if there is measurement error in p_{zat}. Since we are proxying for pollution exposure using the average level of pollution in a zip code on a given day, measurement error might be substantial (*i.e.* people's actual exposure to ambient air pollution might differ significantly from that which is reported by a monitor).

Instrumental variables provide a convenient solution to the bias from omitted variables as well as the bias introduced from measurement error in the independent variable.[29] We use airport ground congestion as an instrumental variable for local pollution levels in the following first stage regression equation:

$$\textbf{First Stage (Model 1):} \quad p_{zat} = \alpha_1 \widehat{T_{at}} + \mathbf{Z}_{zt}\Gamma + \nu_{za} + e_{zat}. \tag{6}$$

The first stage regression, equation (6), estimates the degree to which instrumented airport taxi time $\widehat{T_{at}}$ predicts local pollution levels in areas surrounding airports.[30] The second stage equation uses the predicted values from the first stage to estimate the impact of local pollution variation on health. We also estimate versions of equation (6) using models that interact $\widehat{T_{at}}$ with distance, wind speed, and wind direction as in equations (2) and (3), models 2 and 3, respectively.

Aside from the relationship between pollution and health, we are also explore "reduced form" relationships between health outcomes and taxi time. These "reduced form" estimates are directly policy relevant; namely, how does aggregate daily taxi time impact the health of nearby residents? Understanding the degree to which variation in airport runway congestion directly impacts health

29. Instrumental variables only solves the bias from measurement error in the independent variable when the measurement error is classical, namely mean zero and i.i.d. (Griliches and Hausman, 1986).

30. We are using predicted aggregate taxi time $\widehat{T_{at}}$ as an instrumental variable in these regression models. In standard OLS regression, inference using generated regressors should be corrected for first-stage sampling variance (*e.g.* Murphy and Topel (2002)). When the generated regressor is used as an instrumental variable this is no longer the case. (Wooldridge, 2002, p. 117) presents a weak set of assumptions for which the standard errors of 2SLS regressions using generated instruments are unbiased. The key assumption turns on strict exogeneity between the error term in the structural model and the covariates used to generate the instrument in the auxiliary regression. See Dahl and Lochner (2012) for a similar approach, using a predicted variable as an instrumental variable in a 2SLS setting. These issues are also discussed tangentially in Wooldridge (1997) and Wooldridge (2003).

has implications for both managing congestion through either demand pricing mechanisms (*e.g.* a congestion tax) or a more efficient runway queuing system.

3.3. *Health outcomes: alternative models*

We supplement our baseline health regressions with several alternative models, exploring model specification and model dynamics in more detail. These various regression models are described in more detail below.

3.3.1. Health Outcomes: Non-linearities and Threshold Effects. There is reason to suspect that the relationship between pollution and health outcomes is non-linear in the level of pollution. Do highly polluted days matter more for predicting negative health outcomes than moderately polluted days? We test these hypotheses in two different ways. First, we examine heterogeneity in the dose–response relationship between seasons of the year as pollution levels of CO and NO_2 are higher in the winter months as shown in Supplementary Figure A3. We interact all variables in all regressions (first and second stage) with a dummy for summer (April–September), thereby allowing the effect to be different for two subsets of the year. Marginal changes at higher baseline levels of pollution (*i.e.* winter) should be larger than marginal changes at lower levels of baseline pollution (*i.e.* summer) if the dose–response function was in fact non-linear. There may be other important differences in health outcomes across seasons that could explain these seasonal disparities. For example, pollen levels might be higher in the winter as most precipitation occurs in the winter and hence flowering occurs in early spring. A body that is weakened by the elevated pollen levels might be more (or less) susceptible to pollution shocks. The non-linearity we are measuring might be an interaction effect with other substances and not directly related to the average pollution level.

Second, we use the over-identified model 3 to instrument higher order polynomials of average daily pollution levels and plot the responding dose–response function. Pollution spreads non-linearly in wind direction and wind speed, and our overidentified models allow us to identify higher-order polynomials.

3.3.2. Health outcomes: dynamic effects and forward displacement. By looking at the daily response of health outcomes to contemporaneous pollution shocks, we may be neglecting important dynamic effects of pollution and health. For example, contemporaneous exposure to air pollution may have lagged effects on health, leading people to seek care one or two days after the initial pollution episode. Our contemporaneous regression models might miss these important lagged impacts. Alternatively, health estimates may be driven by various forms of forward displacement. Short-term spikes in pollution might lead individuals on the brink of an asthma or heart attack to experience an episode that would have otherwise occurred in the next few days anyway. Such behaviour would overestimate the dose–response function as an increase in hospitalization rates is followed by a decrease once pollution levels subside. We explore the dynamic effects of pollution on health by estimating the following distributed lag model:

$$y_{zat} = \sum_{k=0}^{3} \beta_k p_{za(t-k)} + \mathbf{Z}_{zt}\Pi + \eta_{za} + \epsilon_{zat}. \tag{7}$$

Instrumented pollution p_{zat} is again obtained using either model 1, 2, or 3 from previous sections. In the case of forward displacement, the spike in hospital admissions should be followed by a

decrease in admissions, and hence $\sum_{k=0}^{3}\beta_k < \beta$, where the latter β comes from the baseline, contemporaneous regression. In a sensitivity check (available upon request) we include six lags and three leads.

3.3.3. Health outcomes: heterogeneity and self-selection. Our baseline models rely upon the relatively unattractive assumption that the relationship between pollution and health is the same for everyone in the population. If there is heterogeneity in a person's relative susceptibility to pollution (or in how people respond to adverse health outcomes), then people may sort themselves into locations based on these observed or unobserved differences. This heterogeneity may manifest itself through access to medical care or through biological differences in the pollution–health relationship among certain segments of the population. Previous research (*e.g.* Chay and Greenstone (2003)) and results presented in subsequent sections of this article suggest that health effects differ by observable characteristics of the population. If people sort themselves based on this underlying heterogeneity, then our estimates may identify the average effect of pollution on health for a non-random subpopulation in the data (Willis and Rosen, 1979; Garen, 1984; Wooldridge, 1997; Heckman and Vytlacil, 1998).

We address these issues in various ways. In a sensitivity check, we limit our estimates to people 65 years and older who have guaranteed health insurance in the form of Medicare. Thus, any heterogeneity in hospitalization should no longer be driven by access to health insurance. Another concern is that the severity of the particular health shock determines whether a person will seek emergency care. We therefore also include heart problems as a category, which are severe enough that patients will seek medical help independent of their insurance or financial situation. There may also exist significant heterogeneity based on unobservable characteristics. Previous research suggests that individuals engage in avoidance behaviour on days where pollution is predicted to be high (Neidell, 2009), which implies there is likely heterogeneity in β as well as correlation between β and p_{zat}. In a previous version of this article, we developed a framework to test whether selection on unobserved heterogeneity leads to bias in our estimates (Schlenker and Walker, 2011), but did not find this to be the case. The lack of self-selection bias may be in part driven by our research design, where airport-driven pollution is relatively stochastic and unforecastable, making it difficult to select on.

3.3.4. Health outcomes: poisson model. Since our dependent variable is measured as hospital visits in a given zip code day (before we convert it to a sickness rate), we also estimate regression models that account for the non-negative and discrete nature of the data. Specifically, we use a conditional (fixed effects) quasi-maximum likelihood Poisson model (Hausman *et al*., 1984; Wooldridge, 1999).[31] To account for the endogeneity of pollution exposure, we generalize the standard conditional Poisson model into an instrumental variables setting. To do this, we adopt a control-function approach to the conditional Poisson model (see *e.g.* Wooldridge (1997) and Wooldridge (2002)), whereby we include the residual ($\widehat{e_{zat}}$) from our first-stage regression (*i.e*. the effect of taxi time on pollution) in our regression equation of interest:

$$\mathbb{E}[s_{zat}|p_{zat}, T_{at}, \mathbf{Z}_{zt}, \eta_{za}] = \eta_{za}\exp(\beta p_{zat} + \gamma_1\widehat{e_{zat}} + \mathbf{Z_{zt}}\Pi), \tag{8}$$

where s_{zat} are sickness counts (no longer rates), p_{zat} is the observed pollution level in a county, and $\widehat{e_{zat}}$ is the residual from one of the first-stage regression of pollution on taxi time using model

31. The Poisson model is generally preferred to alternative count data models, such as the negative binomial model, because the Poisson model is more robust to distributional misspecification provided that the conditional mean is specified correctly (Cameron and Trivedi, 1998; Wooldridge, 2002).

1, 2, or 3. The fixed effect model allows the marginal effect of pollution to differ by zip code. The model accounts for the fact that zip codes have different number of residents through the fixed effects η_{za}.

While including the first-stage error purges the estimates of the various selection biases outlined above (Wooldridge, 2002, p. 663), the standard errors need to be corrected for the variation coming from the first stage estimation. To account for the first-stage sampling error in the e_{zat}, we bootstrap the regression using a block-bootstrap procedure where we randomly draw the entire history of a zip code with replacement.

4. EMPIRICAL RESULTS

4.1. *Aggregate daily taxi time and local pollution levels*

We start by examining the effect of airport congestion on pollution levels in surrounding areas. Table A5 gives the first-stage results when taxi time is instrumented using runway congestion at the three major airports outside of California. There is one noteworthy result: For major hubs in California, an increase in taxi time at East Coast airports increases taxi time as delays propagate through the system. On the other hand, the sign reverses for smaller airports: an increase in taxi time at East Coast airports decreases local taxi time. As Pyrgiotisa *et al*. (2013) point out, propagation through the system can have "counter-intuitive results". If planes bunch up at one hub, the effects on close-by commuter airports can be the opposite as the connectors now arrive more evenly spread, or because flights are canceled.[32]

Table 1 presents regression estimates using the specifications outlined in equations (1), (2), and (3), presented in columns a, b, and c, respectively. Each column represents a different regression, where the dependent variable in the columns (1a)–(1c) is the daily mean CO measured in parts per billion (ppb). Columns (2a)–(2c) report regression estimates for daily mean NO_2, while columns (3a)–(3c) report estimates for ozone O_3.[33] Taxi time is reported in thousands so that the coefficients in Table 1 report the marginal effect of a 1,000 minute increase in taxi time on local pollution levels. All regressions report robust standard errors, clustering on both zip code and day.[34]

Column (1a) suggests that a 1,000 minute increase in taxi time increases ambient CO concentrations in zip codes within 10 km of an airport by 45 ppb (an 8% increase relative to the mean, or 12% of the day-to-day standard deviation). Since the standard deviation of taxi time at LAX in Supplementary Table A1 is 1,852, a one-standard deviation increase in taxi time leads

32. For example, flights out of Santa Barbara frequently get canceled if Los Angeles is backed up to reduce the queue of incoming airplanes into Los Angeles.

33. OLS estimates are presented in Supplementary Table A6.

34. The heavily over-identified models from equation (3) impose significant computational burdens when estimating IV models containing two-way, cluster-robust standard errors. To circumvent this issue, we report the results from running the first stage and then using the predicted values in the second stage without accounting for the fact that we are using generated regressors in the second stage. Plugging in the predicted regressors is computationally much easier because we do not cluster all the first-stage regressions, instead we simply recover the point estimates from each regression. Two-way cluster robust routines require estimating three variance–covariance matrices, one corresponding to the first cluster group, one corresponding to the second cluster group, and one corresponding to the two-way expansion of the two groups. Since we have more than a 100 instruments in model 3 (12 variables times 12 airports times 3 east coast airports = 432 first stage regressions), this imposes a significant computational burden. To understand the likely magnitude of this bias, Supplementary Table A7 reports two sets of standard errors for equations (1) and (2): (i) the IV results; and (ii) running the first stage and using the predicted values in the second stage with two-way clustered errors but no other adjustments. The results suggest that the standard errors from the IV are quite similar to those from manual 2SLS.

TABLE 1
Pollution regressed on instrumented taxi time

	CO pollution			NO_2 pollution			O_3 pollution		
Variable	**(1a)**	**(1b)**	**(1c)**	**(2a)**	**(2b)**	**(2c)**	**(3a)**	**(3b)**	**(3c)**
Taxi Time	44.78***	56.26***	52.56***	0.57***	0.67***	0.67***	−0.00	0.08	0.16
	(5.04)	(9.48)	(10.49)	(0.09)	(0.15)	(0.22)	(0.09)	(0.11)	(0.20)
Taxi x Distance		−1.62	−2.13		−0.01	−0.02		−0.01	−0.03
		(1.22)	(1.37)		(0.02)	(0.03)		(0.01)	(0.02)
Taxi x $Angle_u$			13.16*			0.31			−0.50***
			(7.78)			(0.22)			(0.18)
Taxi x $Angle_d$			5.48			0.05			0.05
			(6.97)			(0.18)			(0.12)
Taxi x Speed			−2.05			−0.08*			0.04
			(1.89)			(0.04)			(0.05)
Taxi x Distance x $Angle_u$			−0.60			−0.02			0.05**
			(1.10)			(0.03)			(0.02)
Taxi x Distance x $Angle_d$			0.16			−0.01			−0.01
			(0.89)			(0.03)			(0.02)
Taxi x Distance x Speed			0.55**			0.01*			−0.00
			(0.25)			(0.01)			(0.01)
Taxi x $Angle_d$ x Speed			1.70			0.10*			−0.07
			(2.66)			(0.05)			(0.06)
Taxi x $Angle_u$ x Speed			−10.41***			−0.19**			0.26***
			(3.74)			(0.10)			(0.09)
Taxi x Dist. x $Angle_u$ x Speed			1.50***			0.03**			−0.03**
			(0.50)			(0.01)			(0.01)
Taxi x Dist. x $Angle_d$ x Speed			−0.63*			−0.01			0.01
			(0.35)			(0.01)			(0.01)
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip Codes	164	164	164	164	164	164	164	164	164
Days	1,095	1,095	1,095	1,095	1,095	1,095	1,095	1,095	1,095
F-stat(joint sig.)	78.48	42.24	14.11	39.67	19.80	4.88	0.00	0.76	1.26
p-value (joint sig.)	1.33e-15	1.66e-15	7.89e-20	2.68e-09	2.00e-08	7.48e-07	.9773	.4705	.2452

Notes: Table regresses zip code level pollution measures on airport congestion (total taxi time in 1,000 min) in 2005–7. Taxi time at the local airport is instrumented with the taxi time at three major airports in the Eastern US. All regressions include weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction), temporal controls (year, month, weekday, and holiday fixed effects), and zip code fixed effects. Regressions are weighted by the total population in a zip code. Errors are two-way clustered by zip code and day. Significance levels are indicated by *** 1%, ** 5%, * 10%.

to 0.23 standard deviation increase in CO pollution of the zip codes around LAX. Column (1b) of Table 1 includes an interaction of taxi time with distance to the airport. The non-interacted taxi time coefficient now reports the effect of airplane idling on pollution levels directly at the airport. The point estimate implies that a one standard deviation increase in taxi time at LAX leads to 0.28 standard deviation increase in CO levels in areas adjacent to LAX. The interaction term shows how this effect decays linearly with distance.

Finally, column (1c) reports the coefficients from the estimated version of equation (3) that interacts taxi time with wind speed and wind angle from an airport. The F-test for the joint significance of these coefficients is given in the last two rows of the table and shows that they are highly significant. Since individual coefficients are difficult to interpret, we plot the marginal effect of an extra 1,000 minutes of taxi time for four wind speeds in the first row of Figure 3. Wind speeds increase from left to right. The colour indicates the marginal impact ranging from low (blue) to high (red). If a zip code is directly downwind, it is on the positive x-axis, while

areas upwind are on the negative x-axis.[35] Figure 3 makes clear that there is significant spatial heterogeneity in the marginal effect of taxi time, and this heterogeneity depends on distance from an airport, wind speed, and wind direction. As such, equation (3) (*i.e.* model 3) is best able to capture this heterogeneity.

Columns (2a)–(2c) of Table 1 give estimates pertaining to the effect of taxi time on NO_2 levels. The results are comparable to those from CO, although the linear decrease in distance from the airport is not significant. A one standard deviation increase in taxi time at LAX increases NO_2 concentrations by roughly 1 ppb, or 10% of the day-to-day standard deviation. The second row of Figure 3 shows again that downwind areas are much more impacted than upwind areas. Both Table 1 and Figure 3 show that the relative impact of NO_2 is different than CO: the range of marginal impacts for CO in Figure 3 is between -71% and $+43\%$ relative to the average impact from column (1a) in Table 1. In contrast, the marginal effect of taxi time on NO_2 varies between -60% and $+33\%$ relative to the average effect from column (2a) of Table 1. The spatial pattern is also somewhat different. In subsequent sections, we use these relative differences in pollutant dispersion to jointly estimate the effect of both CO and NO_2. Recall from Section 1 that CO emissions are higher during low power operation, while NO_2 is higher during high power operation. Larger wind speeds require more thrust during takeoff and hence change the mix of CO and NO_2 emissions.

Finally, columns (3a)–(3c) replicate the same analysis for ozone (O_3), a pollutant that is not directly emitted from airplanes.[36] The results in Table 1 suggest that airport taxi time has little significant impact on ozone levels, although some of the interaction terms are significant, the joint F-test is never significant. In the remainder of the analysis we focus on CO and NO_2, the two criteria air pollutants for which airplanes are large emitters, while acknowledging that we may be picking up the health effects of other pollutants that are correlated with airplane emissions.

Our baseline pollution estimates presented above come from models in which airport taxi time is instrumented with taxi time at large airports outside of California. We instrument taxi time because delays and runway congestion might be correlated with local weather, which in turn might impact pollution levels. In addition, there is likely measurement error in our taxi time variable as it only includes domestic, commercial flight activity. While we control for weather in our regressions, there might be unobserved weather (or other) variables that jointly impact both pollution and taxi time. Supplementary Table A6 replicates the baseline IV analysis of Table 1 using local taxi time at California airports, which is not instrumented. The estimated effect is generally half as big for CO and NO_2. The smaller OLS estimates are consistent with adverse weather (*e.g.* precipitation) causing both airport delays and at the same time reducing ambient air pollution. Alternatively, these results could be driven by the well-known attenuation bias stemming from measurement error in fixed effects models. In the remainder of the article, we rely on instrumented taxi time stemming from network delays.

35. Areas downwind are more affected by taxi time than areas upwind. For the very highest wind speeds, the largest marginal impact of taxi time can be found just upwind from the centroid of the airport (although the average marginal impact remains highest downwind). This is possibly due to the fact that airplanes start against the wind and mostly line up in the opposite direction, *i.e.*, the direction from which the wind is blowing. Local wind is highly predictive of congestion. When local wind is strong and the average local taxi time is high and the queue is long, an additional unit of congestion due to network delays will hence "add" an additional plane that is idling upwind from the airport centroid. For example, the four runways of LAX are between 2.7 km and 3.7 km long, which is significant as we are examining monitors within 10 km of the airport centroid.

36. Ozone is formed through a complicated chemical reaction between both nitrogen dioxides and VOCs in the presence of sunlight. As Auffhammer and Kellogg (2011) have shown, increasing VOC in VOC-rich environments can have no effect on ozone or slightly decrease it, while it will increase ozone if VOCs are limited compared to NO_x. This poses a challenge for the monotonicity assumption behind IV regressions.

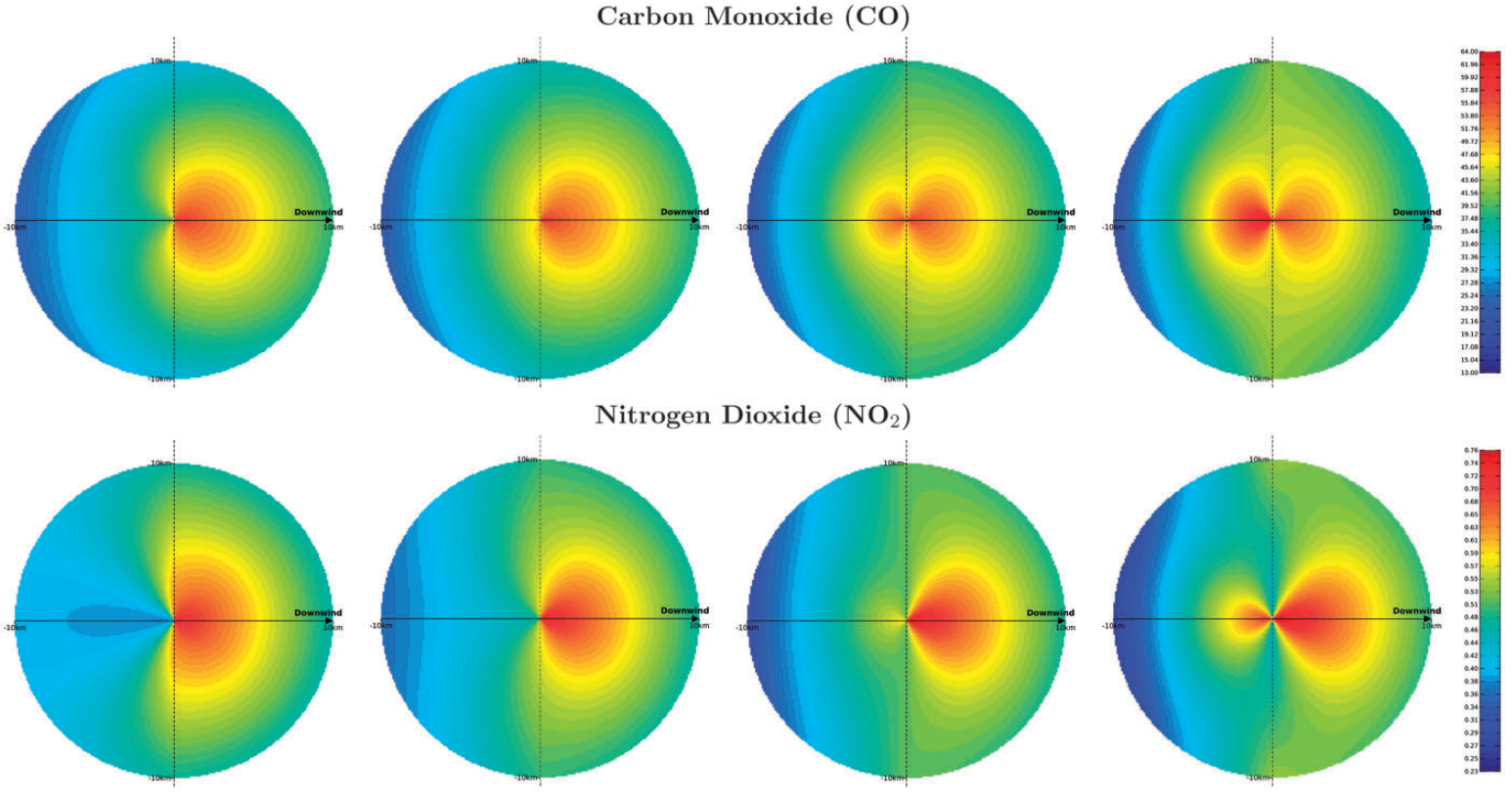

FIGURE 3

Contour maps: marginal impact of taxi time on pollution levels.

Notes: Graphs display the marginal impact of taxi time (ppb per 1000 minute of taxi time, *i.e.*, kmin) on pollution levels across space for different wind speeds. The x-axis shows the direction in which the wind is blowing: positive x-values imply the location is downwind, negative value imply they are upwind. Points on the y-axis are at a right angle to the wind direction. The wind speeds in columns 1–4 are 0.1 m/s, 1 m/s, 2 m/s, and 3 m/s corresponding to the 0.1, 10.6, 34.5, and 66.5 percentiles of the distribution of wind speeds in 2005–7 at the twelve airports in our study (see Figure 1).

We use taxi time at three major airports in our baseline regressions: Atlanta (ATL), Chicago (ORD), and New York (JFK). Supplementary Table A7 presents first-stage F-statistics if we instrument taxi time at California on up to four airports outside of California. Recall that we allow the coefficients to vary by airport, as network congestion will have different absolute effects on California airports. Irrespective of whether we use 1, 2, 3, or 4 airports outside of California, the F-statistic is well above 10. In our baseline model, we use three airports that cover weather patterns in three regions of the Eastern US: Southeast (Atlanta), Midwest (Chicago), and Northeast (New York JFK), and the first-stage F-stat is 50. The fourth large airport outside of California that we include in columns (d) is Dallas/Fort Worth (DFW). While results are not particularly sensitive to including DFW, we exclude it from our baseline specifications as it is significantly closer to California airports and thus may be more endogenous than the other three airports (*i.e.* Dallas/Fort Worth may be delayed because California airports are delayed).

Reverse causality is less of a concern for the other three airports: a flight that leaves a California airport at 6 am will not reach Atlanta, Chicago, or New York until roughly noon due to the change in time zones. Supplementary Table A8 tests for reverse causality directly by regressing taxi time at an airport on eight weather measures we generally include as controls: a quadratic in minimum and maximum temperature, precipitation, as well as wind speed.[37] The column heading gives the airport at which the congestion is measured while the row indicates the airport at which the weather variables are measured.[38] The table reports p-values of a hypothesis test pertaining to the joint significance of the weather variables. The diagonal is highly significant as local weather measures impact airport taxi time. While weather at the eastern airports (ATL, ORD, or JFK) sometimes impacts taxi time at the two largest airports in California (LAX and SFO), the reverse is not true. This is consistent with weather at Eastern airports causing local network delays that propagate through the airspace and impact taxi time in California. The reverse direction does not hold. California airports do not affect East Coast airports on the same day. This result is not simply an arteifact of there being less weather variation in California, as weather at LAX significantly impacts taxi time at SFO.

We have also run two sensitivity checks to further rule out endogeneity through reverse causality, the results of which are reported at the end of the subsequent section on health effects and shown in Supplementary Table A13. First, we only utilize the combined taxi time between 5 am and noon at the three major Eastern airports to rule out California feedback effects. This reduces the F-stat in model 1 from 50 to 35.5, but the results remain similar to baseline estimates. Second, instead of using taxi time at the three major Eastern airports, we use the eight weather variables at each of these airports. Since this effectively increases the number of instruments by a factor of eight, we no longer estimate model 3 (which had 432 instruments to begin with). The F-statistic for the weather-instrumented regression is 5,436. Again, results remain similar to our baseline estimates but the standard errors in the second-stage increase. Going forward we instrument using the overall daily taxi time, as it has a higher F-statistic than focusing only on the mornings yet is more tractable than using weather measures, which would result in 3,456 instruments in model 3.

We conduct two last robustness checks. First, since the variation in pollution due to delays outside of California should be uncorrelated with weather in California, we have estimated models (not reported) that exclude California weather controls altogether. Reassuringly, our baseline

37. Weather measures in our baseline regression also include the direction in which the wind is blowing relative to the direction in which the zip code is located. Since the dependent variable in the current regression is at the airport level and not the zip code level, these variables are not well defined and hence dropped.

38. If we pair airport taxi time with weather from another airport, we also include the local weather measure as control. The local weather measures are not included in the joint test of significance.

estimates for the most important pollutant (CO, see below) are similar whether we include or exclude California weather controls, but the error terms increase. Second, there may of course be some omitted variable that affects congestion outside of California and health outcomes in California. This hypothesis is not directly testable, but we have estimated models (available upon request) which include taxi time at other CA airports as a control variable in our baseline reduced form regressions, and the results remain very similar.

To put the magnitude of these effects into perspective, it is useful to consider the current ambient air standards in place for CO as regulated by the EPA under the Clean Air Act. The current one hour carbon monoxide standard specifies that pollution may not exceed 35 ppm (or 35,000 ppb) more than once per year. California has their own CO standard which is 20 ppm. A one standard deviation increase in LAX airplane idling (1,852 minutes) translates into an 83 ppb increase (44.78×1.852) in carbon monoxide levels for areas within 10 km of LAX using estimates from column (1a) of Table 1. Adding this number to the average daily maximum CO level at zip codes from Panel A of Supplementary Table A2 (1234 ppb), the estimated increase in pollution concentrations is far below the current EPA standard. Similarly, for NO_2, the current EPA 1-hour standard is 100 ppb. Using estimates from column (2a) of Table 1, a standard deviation increase in LAX taxi time would lead to a 1 ppb increase in NO_2 levels. Evaluated relative to the average daily maximum NO_2 levels of 35.5 ppb, these are again well below the ambient criteria standard. Note that the maximum of the maximum daily NO_2 levels is above the standard as some areas are out of attainment. The remaining sections estimate the social costs of these congestion-related increases in ambient air concentrations by focusing on heath outcomes of the populations most affected by these emissions.

4.2. *Effects of taxi time on local measures of health*

We begin by investigating the "reduced form" health effects of airports, relating aggregate daily taxi time to local measures of health. Namely, how does variation in airport congestion predict local health outcomes? Table 2 presents the results from a regression relating daily measures of airport taxi time to local hospital admissions for the overall population as well as two susceptible subgroups: individuals below 5 years of age and individuals aged 65 years and above. The dependent variable is measured as the daily sum of hospital and emergency room visits for persons living in a particular zip code scaled by the population (per 10 million individuals) in that particular zip code. The regressions are weighted by zip code population size, and taxi time is instrumented using taxi time at three major airports in the East. The estimated coefficient on the taxi time variable corresponds to the increased rate of hospitalizations per 10 million individuals in a zip code for an extra 1,000 minutes of taxi time. Using various diagnosis codes, we examine the impact of taxi time on asthma, respiratory, and heart-related admissions separately. As a falsification exercise, we also estimate the incidence of taxi time on strokes, bone fractures, and appendicitis rates. The reported standard errors are clustered on both zip code and day.

For the overall population (Panel A), all respiratory sickness rates as well as heart problems are significantly impacted by taxi time, while the placebo effects for stroke, bone fractures, and appendicitis are not significantly affected. Results become larger in magnitude for the at-risk age groups. For the population 65 years and above, the incidence of stroke and bone fractures is marginally significant at the 10% level. This may be do to statistical chance or may be explained by the fact that senior citizens may also be more susceptible to sicknesses that covary with one another (*e.g.* a respiratory problem might make them fall and break a bone). Additionally, Medicare provides doctors implicit incentives to add additional diagnosis codes to receive higher reimbursement rates. Consistent with this explanation, models for which the dependent variable

TABLE 2
Sickness rates regressed on instrumented taxi time

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	All Heart (2)	Stroke (3)	Bone fractures (4)	Appendicitis (5)
Panel A: all ages							
Taxi Time	13.84***	24.77***	33.89***	19.35***	2.55	−1.33	0.26
	(2.72)	(7.73)	(9.82)	(5.24)	(1.71)	(2.87)	(0.68)
Panel B: ages below 5 years							
Taxi Time	24.46**	84.28	116.12*	6.63*	0.80	2.16	−0.29
	(11.21)	(51.35)	(62.46)	(3.47)	(0.94)	(5.84)	(1.38)
Panel C: age 65 years and above							
Taxi Time	36.89***	63.80***	100.53***	156.54***	22.87*	19.13*	0.75
	(11.39)	(16.43)	(25.25)	(36.98)	(12.95)	(9.95)	(1.21)
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip codes	164	164	164	164	164	164	164
Days	1,095	1,095	1,095	1,095	1,095	1,095	1,095

Notes: Table regresses zip-code level sickness rates (counts for primary and secondary diagnosis codes per 10 million people) on daily congestion (taxi time in 1000 min) that is caused by network delays (taxi time at three major airports in the Eastern US). All regressions include weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction), temporal controls (year, month, weekday, and holiday fixed effects), and zip code fixed effects. Regressions are weighted by the total population in a zip code. Errors are two-way clustered by zip code and day. Significance levels are indicated by *** 1%, ** 5%, * 10%.

is measured only using the primary diagnosis code, the placebo effects for 65 years and older are no longer significant.

4.3. *Hospital admissions and instrumented pollution exposure*

Results thus far have shown that aggregate airplane taxi time generates variation in pollution levels of nearby communities. We exploit this variation to examine the relationship between pollution and health explicitly. Table 3 summarizes regression results for various pollutants and illnesses using a variety of traditional econometric specifications. Each entry corresponds to a different regression, where the dependent variable is measured as hospital admission rates, and the independent variable is the daily mean ambient pollution concentration in a particular zip code. As before, regression estimates are weighted by zip code population and standard errors are clustered on both zip code and day.[39]

The first row within each panel presents estimates from a pooled OLS version of equation (5) without any controls $\mathbf{Z}_{zt}$, which suggests that increased ambient air concentrations lead to adverse health outcomes for respiratory and heart problems. Since various pollutants are often correlated with one another, these estimates should be interpreted with caution, as the pollutant of interest will proxy for other correlated air pollutants. Each consecutive row adds more controls. The second row uses time controls (year, month, weekday, and holiday fixed effects), and the third row additionally adds weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction). To control for unobserved,

39. Unweighted regressions yield similar results and are available upon request.

TABLE 3
Sickness rates regressed on pollution

	Asthma (1a)	Acute Respiratory (1b)	All Respiratory (1c)	All Heart (2)	Stroke (3)	Bone fractures (4)	Appendicitis (5)
Panel A: CO pollution — all ages							
No controls	0.070***	0.265***	0.353***	0.035	−0.002	−0.022***	−0.001
	(0.017)	(0.041)	(0.053)	(0.028)	(0.006)	(0.007)	(0.001)
Time controls	0.030	0.058	0.070	−0.022	−0.014*	−0.008	0.001
	(0.024)	(0.057)	(0.075)	(0.040)	(0.008)	(0.010)	(0.001)
Time + Weather	0.056**	0.047	0.071	0.002	−0.005	−0.012	−0.001
	(0.028)	(0.068)	(0.091)	(0.052)	(0.010)	(0.012)	(0.001)
Time + Weather + Zip Code FE	0.011	0.049***	0.077***	0.027***	−0.001	−0.007*	0.002
	(0.007)	(0.018)	(0.022)	(0.008)	(0.003)	(0.004)	(0.001)
Panel B: NO_2 pollution — all ages							
No controls	3.1***	10.7***	14.6***	4.3***	0.6***	−0.3	0.1**
	(0.5)	(1.3)	(1.7)	(1.1)	(0.2)	(0.2)	(0.0)
Time controls	1.7**	6.0***	7.9***	1.0	−0.1	0.6*	0.1**
	(0.7)	(1.5)	(2.1)	(1.4)	(0.3)	(0.3)	(0.0)
Time + Weather	4.2***	8.3***	11.5***	3.0	0.8*	0.7	−0.0
	(1.0)	(2.6)	(3.6)	(2.4)	(0.5)	(0.5)	(0.1)
Time + Weather + Zip code FE	0.1	1.2*	2.5***	0.9***	0.1	−0.0	0.1*
	(0.2)	(0.6)	(0.8)	(0.3)	(0.1)	(0.2)	(0.0)
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip codes	164	164	164	164	164	164	164
Days	1,095	1,095	1,095	1,095	1,095	1,095	1,095

Notes: Table regresses zip code level sickness rates (based on primary and secondary diagnosis codes) on daily pollution (ppb) in 2005–7. Each entry is a separate regression. Columns use sickness rates (counts per 10 million people) for different diseases, while rows use different controls. The first specification (row) in each panel has no controls, while the second adds time controls (year, month, weekday as well as holiday fixed effects), the third adds weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction), and the fourth adds zip code fixed effects. All regressions are weighted by the total population in a zip code. Errors are two-way clustered by zip code and day. Significance levels are indicated by *** 1%, ** 5%, * 10%.

time-invariant determinants of health, the fourth row of each panel in Table 3 reports regression estimates from a model using zip code fixed effects. The model is identified by examining how within zip code changes in pollution are related to hospitalization rates of that particular zip code. Again, pollution is often strongly correlated with health, although the estimates in the fourth row are usually smaller than those in the first three. These smaller point estimates are consistent with time-invariant omitted variables introducing bias into the estimates from rows one through three. Alternatively, classical measurement error in the pollution variable may lead to significant attenuation bias in fixed effects models (Griliches and Hausman, 1986), and this may be responsible for the smaller point estimates in the last row.

Aside from attenuation bias, fixed effects models may also suffer from biases introduced by any unobserved, time-varying determinants of both pollution and health (*e.g.* weather). To explore this issue further, Table 4 presents instrumental variable estimates of the pollution–health relationship, using instrumented aggregate airport taxi time as an instrumental variable for daily mean pollution. Table 4 presents results for both the overall population in Panel A as well as children below 5 year in Panel B and people aged 65 years and above in Panel C.[40] The three

40. Results for the two remaining groups: children ages 5–19 years and adults ages 19–64 years are given in Supplementary Table A9. Children between 5 and 19 years of age show no sensitivity to pollution shocks. Conversely,

TABLE 4
Sickness rates regressed on instrumented pollution

	Asthma (1a)	Acute respiratory (1b)	All respiratory (1c)	Heart problems (2)	Stroke (3)	Bone fractures (4)	Appendicitis (5)
Panel A: all ages							
Model 1: CO	0.311***	0.556***	0.761***	0.434***	0.057	−0.030	0.006
	(0.065)	(0.162)	(0.207)	(0.134)	(0.039)	(0.063)	(0.015)
Model 2: CO	0.307***	0.550***	0.755***	0.419***	0.050	−0.030	0.003
	(0.062)	(0.163)	(0.210)	(0.128)	(0.038)	(0.064)	(0.015)
Model 3: CO	0.194***	0.396***	0.515***	0.226***	0.020	−0.039	0.002
	(0.047)	(0.125)	(0.165)	(0.079)	(0.030)	(0.040)	(0.011)
Model 1: NO_2	24.5***	43.8***	59.9***	34.2***	4.5	−2.4	0.5
	(6.2)	(16.2)	(20.5)	(10.5)	(3.1)	(5.2)	(1.2)
Model 2: NO_2	24.3***	43.6***	59.8***	33.5***	4.2	−2.4	0.3
	(6.1)	(16.3)	(20.8)	(10.4)	(3.1)	(5.2)	(1.2)
Model 3: NO_2	12.4***	18.9*	24.2*	17.1**	0.7	−1.0	0.3
	(4.0)	(11.0)	(14.2)	(7.1)	(2.2)	(3.0)	(0.9)
Panel B: ages below 5 years							
Model 1: CO	0.565**	1.948*	2.683**	0.153*	0.018	0.050	−0.007
	(0.240)	(1.124)	(1.353)	(0.081)	(0.021)	(0.136)	(0.032)
Model 2: CO	0.579**	1.930*	2.624*	0.127	0.020	0.064	−0.013
	(0.235)	(1.111)	(1.356)	(0.078)	(0.023)	(0.132)	(0.034)
Model 3: CO	0.669***	2.166***	2.493**	0.075	0.023	−0.012	−0.009
	(0.170)	(0.796)	(0.980)	(0.057)	(0.015)	(0.122)	(0.022)
Model 1: NO_2	42.2**	145.3	200.2*	11.4*	1.4	3.7	−0.5
	(20.5)	(95.2)	(117.3)	(6.4)	(1.6)	(10.0)	(2.4)
Model 2: NO_2	43.2**	144.1	195.9*	9.5	1.5	4.8	−0.9
	(20.2)	(94.3)	(117.6)	(6.2)	(1.7)	(9.7)	(2.5)
Model 3: NO_2	43.6***	122.2*	140.8*	4.5	2.9**	3.7	0.6
	(14.9)	(67.7)	(82.4)	(4.6)	(1.3)	(9.3)	(2.0)
Panel C: ages 65 years and older							
Model 1: CO	0.849***	1.469***	2.314***	3.604***	0.526*	0.440*	0.017
	(0.312)	(0.440)	(0.642)	(1.001)	(0.297)	(0.242)	(0.028)
Model 2: CO	0.815***	1.413***	2.275***	3.529***	0.502*	0.409*	0.016
	(0.288)	(0.422)	(0.637)	(0.971)	(0.302)	(0.241)	(0.028)
Model 3: CO	0.493**	0.696**	1.424***	1.937***	0.198	0.187	−0.025
	(0.204)	(0.309)	(0.511)	(0.620)	(0.249)	(0.161)	(0.026)
Model 1: NO_2	66.5***	114.9***	181.1***	282.0***	41.2*	34.5*	1.4
	(22.8)	(34.3)	(52.3)	(76.1)	(24.2)	(18.2)	(2.2)
Model 2: NO_2	66.5***	115.1***	181.4***	282.2***	41.2*	34.5*	1.4
	(22.8)	(34.3)	(52.3)	(76.2)	(24.3)	(18.2)	(2.2)
Model 3: NO_2	35.3**	38.9	75.8*	131.6***	3.6	12.2	−0.8
	(14.2)	(24.6)	(41.3)	(47.8)	(16.5)	(12.0)	(1.6)
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip codes	164	164	164	164	164	164	164
Days	1095	1,095	1,095	1,095	1,095	1,095	1,095

Notes: Table regresses zip code level sickness rates (counts for primary and secondary diagnosis codes per 10 million people) on daily instrumented pollution levels (ppb) in 2005–7. Each entry is a separate regression. Pollution is instrumented on airport congestion (taxi time) that is caused by network delays (taxi time at three major airports in the Eastern US). Model 1 assumes a uniform impact of congestion on pollution levels at all zip codes surrounding an airport, while model 2 adds an interaction with the distance to the airport, and model 3 furthermore adds interactions with wind direction and speed (columns (a)-(c) in Table 1). All regressions include weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction), temporal controls (year, month, weekday, and holiday fixed effects), and zip code fixed effects. Regressions are weighted by the total population in a zip code. Errors are two-way clustered by zip code and day. Significance levels are indicated by *** 1%, ** 5%, * 10%.

rows (labeled model 1–3) use (1) taxi time, (2) taxi time interacted with distance, and (3) taxi time interacted with distance, wind speed, and wind direction, respectively. These are the specifications outlined in equations (1), (2), and (3) above.

The estimates in Table 4 are usually an order of magnitude larger than the OLS, fixed-effects estimates from Table 3. To put the magnitudes into perspective: the average asthma sickness rate for the overall population is 339 per 10 million inhabitants (Panel A1 and A2 of Supplementary Table A3). The asthma coefficient for CO (model 3) in Table 4 implies that a one standard deviation increase in CO pollution leads to an additional $0.194 \times 368 = 71$ asthma attacks per 10 million people,[41] which is 21% of the daily mean.[42] This suggests that fluctuations in air pollution are a major cause of asthma related illnesses. For heart-related problems, the relative magnitude is 18% of the daily mean. It is important to note that the estimated CO effect may not necessarily be coming from CO itself but from some other pollutant that is co-emitted in jet exhaust that we do not observe (*e.g.* a toxic VOC or particulate matter that is emitted due to incomplete combustion). In addition to measurement error or avoidance behaviour, the fact that variation in CO comes from airplanes may be a further explanation for the discrepancy between OLS and IV estimates. However, the Federal Aviation Administration (2005) suggests that aircraft engines produce the same types of emissions as automobiles, which are the largest single source of carbon monoxide emissions in the US.

Models 2 and 3 in Table 4 estimate over-identified models instrumenting pollution with both taxi time and taxi time interactions. While estimates in model 2 are similar to those from model 1, estimates from model 3 are generally smaller. The reason for the difference in magnitudes between models 2 and 3 is not entirely clear, but we believe there are two competing explanations. The first explanation stems from the inability of models 1 and 2 to capture the spatial heterogeneity in the effect of taxi time. Recall that model 3 uses distance as well as wind direction and wind speed. Marginal impacts of airport congestion vary greatly across space as shown in Figure 3, much more than in a model that only includes distance. Failing to model this heterogeneity in pollution exposure may lead to inaccurate scaling of the reduced form relationships in our IV/2SLS setting. A competing explanation as to why model 3 estimates differ from models 1 and 2 stems from measurement error in the location of exposure. While we know the exact location of each pollution monitor and hence can correctly model the pollution surface in space, we only know the zip code of a person's residence and the hospital, not the exact location where they fell ill. As a result, all models will pair sickness counts with incorrect pollution measures if they are not close to the centroid of the zip code when they fell ill, but this might be aggravated by model 3 that explicitly uses the spatial distribution of the pollution surface. Supplementary Table A10 investigates this latter hypothesis by looking at various subsets of the data. Panel A presents our baseline results, Panel B assigns pollution data based on the zip code of the residence, while Panel C assigns pollution based on the hospital zip code. A few results are noteworthy: first, the estimates using model specification 3 are very close to the estimates using specification 1 and 2 in Panel B1 where we only count sicknesses if both the zip code of the residence and hospital are within 10 km of the same airport. On the other hand, model specification 3 diverges in Panel B2 where the hospital zip code is outside the 10 km radius from airports, perhaps because we measure exposure less

the estimated dose–response for adults are roughly comparable to the baseline estimates, which is not surprising since they are the largest share of the overall population.

41. Panel A of Supplementary Table A2 shows that the standard deviation for CO is 368.

42. This back-of-the-envelope calculation increases the pollution level in each zip code by the average 'overall' standard deviation of pollution fluctuations. Moreover, the average sickness rate is not population weighted. In subsequent sections, we increase pollution in each zip code by the zip code specific standard deviation in pollution fluctuations and calculate the population-weighted average sickness count. The relative impact decreases to 17% of the daily mean under the linear probability model and 19% under a Poisson count model.

accurately (*e.g.* the person might have been at work). In addition, panel B3 shows that there are no significant results where the hospital is within 10 km of another airport, suggesting that we are not simply picking up a daily pattern that is common to all airports.[43] As a secondary bit of evidence, model 3 in Panel B of our baseline Table 4 gives comparable point estimates to model 1 and 2 for children under the age of 5 years, whom are more likely to be at home or in a close-by day care. Due to these competing explanations for the differences across models, we continue to present all model estimates whenever possible, allowing the reader to choose their preferred estimate.

There are two additional explanations for the discrepancies between models 3 and 1 and 2 which we find less salient. First, there is a well-known bias of 2SLS estimators when instruments are weak and when there are many over-identifying restrictions (Bound *et al.*, 1995). In linear models with iid errors, Stock *et al.* (2002) propose rule-of-thumb thresholds for F-statistics for the first stage. However, in both the non-iid case (*i.e.* with clustered standard errors) and in cases with multiple endogenous variables, less is known about the relationship between the F-statistic and the properties of instrumental variables estimates. Baum *et al.* (2003) suggest comparing the test statistic to the Stock *et al.* (2002) critical values for the Cragg–Donald F statistic with a single instrument. According to this metric, results from Table 1 suggest that model 3 is a strong first-stage predictor of local pollution levels with a F-statistic that is 14 for CO pollution and to a lesser extent for NO_2 pollution (F-stat of 5). The first stage in model 3 is not as strong as in models 1 and 2, and the model is highly over-identified with twelve excluded instruments. Bound *et al.* (1995) show how the bias of 2SLS increases in the number of instruments and decreases in the strength of the first stage. The bias of 2SLS in the case of weakly identified or over-identified models is towards the OLS counterpart. Since this is consistent with model 3 estimates in Table 4 being smaller than both model 1 and 2 but still above the OLS estimates, Supplementary Table A11 estimates models 2 and 3 using Limited Information Maximum Likelihood (LIML), which is median-unbiased for over-identified, constant-effects models (Davidson and MacKinnon, 1993). Results remain similar, which suggests that weak instrument attenuation is less of a concern (Angrist and Pischke, 2008). Finally, a second alternative explanation for why model 3 gives lower point estimates is that the hourly wind data represent snapshots of the wind speed and direction and include significant measurement error. However, this is somewhat at odds with the fact that we find such significant spatial patterns in the pollution regressions.

Panels B and C of Table 4 present estimates for children and senior citizens. While the dose–response relationships are larger, so are average sickness rates. In relative terms, a one standard deviation increase in CO pollution now causes a 37% increase in asthma cases for children under 5 years compared to the average daily mean. On the other hand, a one standard deviation increase in CO pollution causes a 24% increase in heart problems for people 65 years and above. The higher 'absolute' sensitivity in Panel B and C suggests that there may exist significant heterogeneity in the population response to ambient air pollution exposure. Since the population aged 65 years and older has guaranteed access to health insurance through Medicare, they may be more inclined to visit the emergency room or hospital relative to the rest of the population, leading to larger estimated effects. On the other hand, the relative magnitude compared to average sickness rates are only slightly larger than for the overall population.

Columns (3)–(5) of each panel includes results for one of three placebos: strokes, bone fractures, and appendicitis. Both strokes and appendicitis are severe enough that people should go to the hospital. None of the results are significant for the overall population in Panel A. Consistent with the reduced form evidence in Table 2, some of the coefficients in Panel C are significant

43. If we assign pollution based on the hospital zip code in panels C, results are generally not significant.

at the 10% level. In Supplementary Table A12 we replicate the analysis using only the primary diagnosis code. None of the placebo regressions remain significant. Since we are interested in the overall effect of pollution on hospitalization rates, our baseline models continue to count total sickness counts for both primary and secondary diagnoses.

Supplementary Table A13 further investigates the sensitivity of our IV estimates to different choices of instrumental variables. As a point of comparison, Panel A replicates the baseline results of Table 4 for all ages. Panel B instruments for pollution using only the taxi time between 5 am and noon at Eastern airports to rule out endogeneity through reverse causality. The results remain robust to this change. Panel C goes one step further and instruments for taxi time at California airports using only weather measures at the three major airports in the Eastern US. While the point estimates remain comparable, the standard errors generally increase.[44]

4.3.1. Jointly estimating the effect of ambient air pollutants. A common challenge in studies linking health outcomes to pollution measures is that ambient air pollutants are highly correlated. It is therefore difficult to determine empirically which pollutant is the true cause of any observed changes in health. Our research design provides one possible solution to the identification problem. Wind speed and wind direction differentially affect both CO and NO_2 dispersion patterns. Moreover, the rate of CO and NO_2 emissions depend on the thrust produced by the engine, and higher wind speeds require more engine thrust. Wind speed hence impacts both the rate at which pollutants are produced and how they disperse. Table 5 estimates the joint effect of both CO and NO_2 on health using our first stage model with wind speed and wind direction interactions (model 3).

In all specifications for which we have multiple endogenous variables, we report the Angrist and Pischke (2008) conditional F-statistics in the tables and text, although these are somewhat hard to interpret. As mentioned above, there are no rule-of-thumb thresholds for linear models with non-iid errors or for models with multiple endogenous variables. When comparing the conditional F-statistics to the Stock *et al*. (2002), the F-statistics suggest that the first stage is "weak". Perhaps more usefully, in all specifications for which we have multiple endogenous variables, we also present two tests that are robust to issues pertaining to weak instruments, the Anderson–Rubin test statistic and the closely related Stock-Wright (2000) S statistic. The null hypothesis tested in both cases is that the coefficients of the endogenous regressors in the structural equation are jointly equal to zero, and, in addition, that the overidentifying restrictions are valid. We use a cluster-robust version of both test statistics that has the correct size even under weak identification (Chernozhukov and Hansen, 2008). The tests are equivalent to estimating the reduced form of the equation (with the full set of instruments as regressors) and testing that the coefficients of the excluded instruments are jointly equal to zero. In most specifications, inference based on the Anderson–Rubin and Stock–Wright tests are consistent with inference based on the Wald test of the same null hypothesis. This suggests that we are not drawing spurious inferences based on weak instruments. We also present results using LIML because LIML is approximately median unbiased for overidentified models, and the results are similar between 2SLS and LIML. When 2SLS is subject to weak instrument bias, the 2SLS estimand will diverge significantly from the LIML estimand toward the OLS estimand. Thus, the fact that LIML and 2SLS deliver similar results assuages our concerns pertaining to significant biases associated with weak instruments.[45]

Table 5 shows that the coefficient for CO remains comparable in size to our baseline estimates from Table 4, albeit slightly larger. Conversely, the coefficients on NO_2 switch

44. We do not estimate model 3 using weather variables as it would include 3,456 instruments.
45. A similar diagnostic exercise of this nature can be found on pages p. 213–215 of Angrist and Pischke (2008).

TABLE 5
Sickness rates regressed on instrumented pollution — joint estimation

	Asthma (1a)	Acute respiratory (1b)	All respiratory (1c)	Heart problems (2)	Stroke (3)	Bone fractures (4)	Appen-dicitis (5)
Panel A: All Ages							
Model 3: CO	0.222**	0.867***	1.189**	0.105	0.054	−0.127*	−0.006
	(0.106)	(0.313)	(0.476)	(0.139)	(0.049)	(0.073)	(0.017)
Model 3: NO_2	−2.2	−40.1	−57.4	10.7	−2.9	7.6	0.7
	(8.0)	(25.6)	(38.4)	(13.4)	(3.5)	(5.6)	(1.5)
$F_{(1st\ stage)}$ - CO	4.18	4.18	4.18	4.18	4.18	4.18	4.18
$F_{(1st\ stage)}$ - NO_2	1.57	1.57	1.57	1.57	1.57	1.57	1.57
$P_{(Anderson\text{-}Rubin)}$	0.0000	0.0000	0.0000	0.0012	0.0027	0.1418	0.0484
$P_{(Stock\text{-}Wright\ S)}$	0.0413	0.0795	0.1141	0.1232	0.5319	0.4795	0.5710
Panel B: ages below 5 years							
Model 3: CO	0.901	4.685**	5.388**	0.133	−0.073	−0.385	−0.102
	(0.599)	(2.167)	(2.483)	(0.142)	(0.051)	(0.355)	(0.072)
Model 3: NO_2	−18.8	−206.8	−237.6	−4.8	8.0*	30.9	7.7
	(46.1)	(174.9)	(199.0)	(11.5)	(4.1)	(27.3)	(6.4)
$F_{(1st\ stage)}$ - CO	3.39	3.39	3.39	3.39	3.39	3.39	3.39
$F_{(1st\ stage)}$ - NO_2	1.30	1.30	1.30	1.30	1.30	1.30	1.30
$P_{(Anderson\text{-}Rubin)}$	0.0000	0.0001	0.0000	0.0185	0.0474	0.0687	0.6100
$P_{(Stock\text{-}Wright\ S)}$	0.0588	0.0886	0.2163	0.4046	0.3567	0.3185	0.5764
Panel C: age 65 years and above							
Model 3: CO	0.268	0.775*	1.726**	1.279	0.490	0.147	−0.051
	(0.323)	(0.445)	(0.772)	(0.821)	(0.399)	(0.281)	(0.046)
Model 3: NO_2	20.1	−6.4	−25.8	60.3	−25.5	3.7	2.2
	(22.1)	(36.3)	(62.6)	(67.7)	(25.9)	(21.4)	(3.1)
$F_{(1st\ stage)}$ - CO	4.96	4.96	4.96	4.96	4.96	4.96	4.96
$F_{(1st\ stage)}$ - NO_2	2.09	2.09	2.09	2.09	2.09	2.09	2.09
$P_{(Anderson\text{-}Rubin)}$	0.1146	0.0026	0.0035	0.0009	0.0164	0.1595	0.0017
$P_{(Stock\text{-}Wright\ S)}$	0.4530	0.1484	0.2756	0.1831	0.3849	0.2969	0.1264
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip codes	164	164	164	164	164	164	164
Days	1,095	1,095	1,095	1,095	1,095	1,095	1,095

Notes: Table regresses zip code level sickness rates (counts for primary and secondary diagnosis codes per 10 million people) on daily instrumented pollution levels (ppb) in 2005–7. The effect of the two pollutants is jointly estimated for the over-identified model 3 using LIML. Pollution is instrumented on airport congestion (taxi time) that is caused by network delays (taxi time at three major airports in the Eastern US). All regressions include weather controls (quadratic in minimum and maximum temperature, precipitation and wind speed as well as controls for wind direction), temporal controls (year, month, weekday, and holiday fixed effects), and zip code fixed effects. Regressions are weighted by the total population in a zip code. Errors are two-way clustered by zip code and day. Significance levels are indicated by *** 1%, ** 5%, * 10%.

sign and are mostly negative and insignificant. We have also used the methods proposed by Chernozhukov and Hansen (2008) to build non-spherical confidence regions for the multiple endogenous variables. Within the joint parameter space of CO and NO_2, the joint confidence region lies in the quadrant where CO is positive and NO_2 is weakly negative.[46]

46. The full set of results, which consist of two- dimensional plots for each hypothesis test, are available upon request.

We interpret these findings as evidence that the returns from regulating CO exceed those from regulating NO_2, at least for the population that comprises our sample. One possible explanation for our results stems from the work done by Auffhammer and Kellogg (2011). Figure 9 of Auffhammer and Kellogg (2011) shows that the Southern California coastline, the location of most of the zip codes in our study, ozone generation seems VOC limited, *i.e.*, a reduction in VOC reduces ozone. Conversely, regions further inland and in Northern California are NO_2 limited. Reducing NO_2 in areas that are VOC limited has little effect on ozone, and this may be the reason we observe small and insignificant results for NO_2. In the remainder of the article, we therefore focus on CO.

4.3.2. Threshold effects and non-linearities in the pollution-health relationship. We explore the functional form of the dose-response function in four separate ways. First, Supplementary Table A14 estimates the relationship separately for the summer (April–September) and the winter (October–March). Each panel of the table provides the point estimates for the two seasons from a joint regression where all variables and instruments are interacted with seasonal dummies as well as the *p*-values of a test whether the coefficients are the same. Especially for the case of children under the age of 5 years, the effect seem to be significantly higher during winter months when average pollution levels are higher.

Recall that CO and NO_2 pollution are higher during the winter months, so a non-linear dose–response function that has increasing marginal damages of pollution should exhibit larger coefficients for the winter months. The coefficient for the winter months is almost always larger than for the summer months for the illnesses that are related to pollution (columns (1a)–(2)). These results are consistent with increasing marginal impacts of pollution. However, there may be other important differences in health outcomes across seasons that could explain these disparities. An obvious candidate for differences between the summer and the winter would be the level of ambient ozone concentrations which tend to be much higher in the summer than the winter. In additional results (Supplementary Table A15), we control for ozone levels as a potential confound and the results are nearly identical. One possible explanation for why ozone does not impact the baseline regression results stems from measurement error in the excluded ozone regressor and/or avoidance behaviour pertaining to ozone.

We have also explored models which estimate the possible non-linear effects of pollution on health outcomes by including higher order polynomials. Models with higher-order pollution terms increase the number of endogenous variables in our regressions, and we use the overidentified model 3 to instrument for the higher order terms. Since higher-order polynomials can be difficult to interpret, Supplementary Figure A5 plots the predicted marginal effects of the pollutant on a range of health outcomes as a function of the 'level' of the pollutant on the given day. That is, we plot the dose response function, where the *y*-axis measures the health response and the *x*-axis measures the level of pollution. Since we are fitting non-linear models, the responsiveness is allowed to vary across the *x*-axis. The dashed line displays the results from our baseline, linear dose–response model (constant marginal damage). The solid represents results from a quadratic model where the 95% confidence interval is added in grey. The four columns represent the four sicknesses that are related to pollution fluctuations (asthma, acute respiratory, all respiratory, and heart problems, respectively). The predicted marginal effect is plotted over the empirical distribution of daily pollution levels, from the 5th to the 95th percentile.[47] While there is some evidence that respiratory problems (columns 1–3) exhibit increasing marginal damages as pollution levels start to increase, again especially for children under the age of 5 years, the confidence intervals reflect

47. Supplementary Figure A4 shows the observed distribution of daily CO levels in our data set.

an inability to reject the null that the damage function is constant over the observed range of CO values.

We have investigated non-linearities in two additional ways that are broadly consistent with the findings above (results available upon request). First, we estimated models whereby we interacted our daily pollution variation of interest with the mean pollution 'level' in a zip code. This allows the dose–response curve to vary (linearly) in the level of average pollution levels of a zip code. If this interaction term is zero, this would support the hypothesis that the marginal effect of a one unit increase in emissions is the same regardless of the level of ambient air pollution (*i.e.* a constant, linear dose–response). If the coefficient on the interaction was significantly positive, then this would support the hypothesis that the marginal effect of ambient air pollution on health outcomes is progressively worse in areas with higher than average pollution levels. A challenge with this particular test is that the average level of ambient air pollution in a zip code can be correlated with many observed and unobserved factors that may contribute to heterogeneity in the dose-response relationship. For example, people in more polluted areas may lack basic preventive health services and thus be 'more' responsive to marginal changes in air pollution because of their underlying health conditions rather than any non-linearity in the dose–response. Nevertheless, results suggest (available upon request) that CO exhibits an increasing dose-response function. Second, we explored the shape of the dose-response function in an OLS, fixed effects setting. While we think the regression coefficient 'magnitudes' may be attenuated by things such as measurement error and/or avoidance type behaviours, the 'shape' of the dose-response curve is likely less sensitive to these concerns (unless of course the bias varied with the level of pollution — which might happen through avoidance behaviour such as "bad air day" alerts). We use this logic to explore the shape of the dose-response function by fitting OLS, fixed-effect regression models that include polynomials in the daily mean pollution level (*i.e.* quadratic, cubic, or quartic). We then plot the predicted marginal effects of the pollutant on a range of health outcomes as a function of the 'level' of the pollutant on the given day (as in Supplementary Figure A5). We see that for both asthma and respiratory illness, the predicted marginal effect is increasing in the level of the pollutant. The patterns suggests some sort of "threshold" by which the marginal effect of CO on health outcomes "flattens out".

While the various results in this section come from different econometric models, the conclusions pertaining to the shape of the dose-response function remain similar across the specifications. The evidence suggests that the marginal effect of pollution is increasing in the level of the pollutant, but at a decreasing rate. The diminishing marginal damages of the dose-response function is also consistent with modern evidence from epidemiology (see *e.g.* Pope *et al.* (2009) and Pope III *et al.* (2011)).

4.3.3. Potential confounding sources of variation. While our estimates suggest that CO is primarily responsible for the observed health responses, there may be other sources of unobserved, concomitant variation that may lead to similar relationships. For example, while we estimate the effect of CO and NO_2 in the same model, we do not directly control for other pollutants such as ozone. It seems unlikely that ozone O_3 is causing the observed relationship. As mentioned above, Supplementary Table A14 estimates the relationship separately for the summer (April–September) and the winter (October–March). Ozone is higher during the summer, while CO and NO_2 are higher during the winter. The observed health effects are larger and more significant during the winter time when ozone is not a big problem. We have also estimated models that directly control for ozone (Supplementary Table A15), and the results remain similar and a bit more precise than our baseline estimates. The standard errors are also much larger for the summer, especially in the case of acute respiratory problems and overall respiratory problems.

This is not surprising, because other pollutants like ozone also impact health outcomes, which will be part of the error term.

One potential omitted variable that we unfortunately cannot measure well is particulate matter, a pollutant which may emerge from combustion emissions and has been shown in the past to increase infant mortality due to respiratory causes (Currie and Neidell, 2005). Particulate matter monitors in California are limited in both their spatial and temporal coverage; readings on ambient particulate monitors are conducted every few days (as opposed to hourly data from other pollutants), and there are far fewer monitors. These limitations do not square well with our research design which relies on high-frequency, daily variation across very localized areas. Nevertheless, we have directly explored the degree to which particular matter predicts adverse health outcomes for the subsample of days and locations for which we have particulate monitor data. Table A16 presents results using the full set of particulate monitors for PM2.5.[48] Supplementary Table A16 suggests that PM does not have much explanatory power in predicting health outcomes, although the standard errors preclude definitive conclusions.[49] Recall that Los Angeles Airport is not a significant point source of particulate matter. While it is the largest point source for CO emissions in the state of California, it only ranks 2,763 and 2,782 among emissions of PM_{10} and $PM_{2.5}$. Even still, we believe that some amount of caution is warranted in interpreting CO as the unique pollutant-related causal channel leading to adverse health outcomes; there may be in fact other unobserved sources of ambient air pollution that covary with CO that may also affect health.

4.3.4. Inpatient versus outpatient data. Traditionally, studies have relied on Inpatient data sets to examine health responsiveness to various external factors such as pollution. One limitation of such data is that a person only enters the Inpatient data set if they are admitted for an overnight stay in the hospital. Many ER visits result in a discharge the same day and hence never result in an overnight stay. Starting in 2005, California began collecting Outpatient (Emergency Room) data. Previous published estimates all replied on Inpatient data only. To better understand the differences between these two datasets as well as compare our results to those from the previous literature, we replicate the analysis using sickness counts from only the Inpatient data in Panels A1–C1 in Supplementary Table A17. By the same token, Panels A2–C2 only uses the Outpatient data.[50] Not surprisingly, there is a significant relationship between pollution and heart problems (column 2) in the Inpatient data for patient ages 65 years and above (as these conditions usually require an overnight stay), but no or very limited sensitivity of asthma or overall respiratory illnesses (columns 1a and 1c) to pollution. Conversely, the Outpatient (ER) data shows a much larger sensitivity of respiratory problems to changes in pollution. These results show the importance of Outpatient (ER) data when studying the morbidity effects of ambient air pollution on health outcomes.

4.3.5. Temporal displacement and dynamics. Our baseline regression models examine only the contemporaneous effect of pollution on health. Contemporaneous estimates may lead to underestimates of the total effects of air pollution on health if health effects respond sluggishly to changes in pollution. Conversely, estimates may overstate the hypothesized effect due to

48. Unfortunately, we only observe 2 PM10 pollution monitors within 15 km of an airport (or equivalently two- zip codes) which makes our research design infeasible due to the importance of distance and wind angle/speed heterogeneity.

49. All of the estimates in Supplementary Table A16 come from limited information maximum likelihood estimates as opposed to 2SLS (although results are similar).

50. Patients that enter the ER and are later admitted for an overnight stay are dropped from the ER data to avoid double counting.

temporal displacement: if spikes in daily pollution levels make already sick people go to the hospital one day earlier, contemporaneous models overestimate the true effect associated with permanently higher pollution levels. If temporal displacement is important, the contemporaneous increase in sickness rates should be followed by a decrease in sickness rates in subsequent periods.

We investigate both of these issues by estimating a distributed lag regression model, including three lags in the pollution variable of interest. Table 6 presents the distributed lag results of pollution for the overall population. We present individual coefficients as well as the combined effect (the sum of the four) in the last row of each panel. To preserve space, we only list the results for the sickness categories that are impacted by changing CO pollution levels. Since regulatory policy is concerned with the health effects of a permanent change in pollution, we focus on cumulative effects of the model over the estimated 4 day horizon. The cumulative effect is slightly larger than the comparable baseline results in Table 4. This might be because some individuals delay hospital visits, although the exact dynamics are hard to determine empirically given the lack of significance of the individual coefficients. We have also experimented with different leads/lags (available upon request). For example, in a model with three leads and six lags, the sum of the six lags and contemporaneous terms are similar in magnitude. The three leads, on the other hand, are not jointly significant.

4.3.6. Count model. Our baseline health estimates consist of linear probability models, relating the population-scaled hospital admission rates to changes in pollution. To account for the non-negative and discrete nature of the hospital admission data, Table 7 presents estimates from a quasi-maximum likelihood, conditional Poisson IV estimator given in equation (8). In contrast to the baseline linear probability health models, these models are not weighted. In addition, since we use a control function to address issues pertaining to measurement error and omitted variables, we adjust standard errors for the first stage sampling variation using a block-bootstrap sampling procedure, resampling zip codes.[51] Analogous to the linear probability model, we find that respiratory illnesses and heart problems are sensitive to pollution fluctuations, while the three placebos are not (with the usual caveat applying to sickness counts for people aged 65 years and above).

The coefficients no longer give marginal impacts and are difficult to interpret. In order to compare the marginal impacts of pollution exposure and congestion across all of our models, Table 8 presents the predicted increase in sickness counts from (1) a one standard deviation increase in taxi time, and (2) a one standard deviation increase in pollution levels in each zip code. The results are then added for all zip codes that are within 10 km of an airport. The table also summarizes population surrounding airports. Various admission categories are given in rows, while the columns show the results for each of the twelve airports. The last column gives the combined impact among all twelve airports.

Panels A, B, and C give the predicted increase in hospital admissions using estimates from the baseline linear probability model whereby pollution is instrumented using model 3 (pollution instrumented with taxi time + interactions with distance and wind direction). These results are presented for the overall population (Panel A), children below 5 years (Panel B), and senior citizens 65 years and above (Panel C). Panel D gives the results for the overall population using the count model shown in Table 7. Impacts are evaluated at the sample mean for the nonlinear Poisson model. The results from the Poisson model are similar to those from the linear probability model

51. This is equivalent to clustering by zip code instead of two-way clustering by zip code and day. An unweighted regression of the linear probability model (available upon request) that clusters by zip code gives comparable results.

TABLE 6
Sickness rates of all ages regressed on instrumented CO pollution — lagged pollution

	Asthma	Acute respiratory	All respiratory	Heart problems
Model 1				
Pollution in t-3	0.026	0.220	0.345	0.022
	(0.096)	(0.188)	(0.250)	(0.139)
Pollution in t-2	0.130	0.109	0.023	0.003
	(0.143)	(0.246)	(0.336)	(0.255)
Pollution in t-1	−0.017	−0.060	−0.001	−0.020
	(0.132)	(0.251)	(0.289)	(0.183)
Pollution in t	0.200**	0.355	0.485	0.422***
	(0.101)	(0.263)	(0.331)	(0.134)
Cum. Effect	0.339***	0.624***	0.853***	0.427***
	(0.070)	(0.163)	(0.210)	(0.151)
Model 2				
Pollution in t-3	0.040	0.229	0.353	0.022
	(0.094)	(0.188)	(0.250)	(0.138)
Pollution in t-2	0.117	0.098	0.013	−0.002
	(0.141)	(0.245)	(0.331)	(0.250)
Pollution in t-1	−0.021	−0.062	−0.004	−0.028
	(0.133)	(0.253)	(0.291)	(0.184)
Pollution in t	0.203**	0.352	0.485	0.415***
	(0.099)	(0.262)	(0.331)	(0.132)
Cum. Effect	0.338***	0.618***	0.847***	0.408***
	(0.066)	(0.163)	(0.214)	(0.143)
Model 3				
Pollution in t-3	−0.002	0.126	0.121	0.045
	(0.041)	(0.095)	(0.124)	(0.057)
Pollution in t-2	0.079	0.023	0.020	−0.014
	(0.060)	(0.116)	(0.151)	(0.087)
Pollution in t-1	−0.059	0.008	0.020	−0.004
	(0.056)	(0.154)	(0.191)	(0.111)
Pollution in t	0.177***	0.316	0.420	0.225**
	(0.067)	(0.201)	(0.263)	(0.100)
Cum. Effect	0.195***	0.473***	0.582***	0.252***
	(0.052)	(0.115)	(0.153)	(0.067)
Observations	179,088	179,088	179,088	179,088
Zip Codes	164	164	164	164
Days	1,092	1,092	1,092	1,092

Notes: Table replicates the results of CO pollution on sickness counts for all ages in Table 4 except that three lags of the instrumented pollution levels are included. Each column in each panel presents the coefficients from one regression as well as the cumulative effect (sum of all four coefficients). Significance levels are indicated by *** 1%, ** 5%, * 10%.

in Panel A. Panel E gives the average daily sickness count in 2005–7 for the overall population for comparison.

Pollution fluctuations have a large effect on the 6 million people living within 10 km of one of the twelve airports: A one standard deviation increase in a zip-codes specific pollution fluctuations increases asthma counts for the overall population by 17% under the linear probability model

TABLE 7
Sickness counts regressed on instrumented CO pollution — poisson model

	Asthma (1a)	Acute respiratory (1b)	All respiratory (1c)	Heart problems (2)	Stroke (3)	Bone fractures (4)	Appendicitis (5)
Panel A: all ages							
Model 1: CO	0.834***	0.596***	0.577***	0.488***	0.270	−0.114	0.325
	(0.171)	(0.109)	(0.112)	(0.128)	(0.184)	(0.184)	(0.454)
Model 2: CO	0.846***	0.589***	0.573***	0.482***	0.246	−0.116	0.245
	(0.172)	(0.111)	(0.116)	(0.128)	(0.185)	(0.188)	(0.466)
Model 3: CO	0.561***	0.399***	0.378***	0.292***	0.132	−0.150	0.163
	(0.132)	(0.090)	(0.087)	(0.095)	(0.195)	(0.133)	(0.325)
Panel B: ages below 5 years							
Model 1: CO	1.202***	0.237	0.303	2.061*	3.334	0.187	−0.369
	(0.387)	(0.179)	(0.208)	(1.148)	(2.876)	(0.572)	(2.923)
Model 2: CO	1.202***	0.216	0.278	1.891*	3.347	0.233	−0.691
	(0.396)	(0.179)	(0.207)	(1.105)	(2.799)	(0.567)	(2.963)
Model 3: CO	1.133***	0.261**	0.256*	1.297	4.238*	−0.064	−1.290
	(0.287)	(0.132)	(0.143)	(0.966)	(2.480)	(0.495)	(2.643)
Panel C: ages 65 years and older							
Model 1: CO	1.287***	0.757***	0.610***	0.634***	0.397*	0.626**	1.190
	(0.364)	(0.208)	(0.173)	(0.165)	(0.219)	(0.314)	(1.247)
Model 2: CO	1.264***	0.743***	0.608***	0.630***	0.388*	0.589*	1.135
	(0.341)	(0.202)	(0.174)	(0.166)	(0.224)	(0.313)	(1.291)
Model 3: CO	0.804***	0.413**	0.397**	0.369***	0.159	0.292	−0.852
	(0.275)	(0.180)	(0.154)	(0.126)	(0.223)	(0.219)	(1.185)
Observations	179,580	179,580	179,580	179,580	179,580	179,580	179,580
Zip Codes	164	164	164	164	164	164	164
Days	1,095	1,095	1,095	1,095	1,095	1,095	1,095

Notes: Table replicates the results for regression models of CO in Table 4 except that we use a Poisson count model instead of a linear probability model. Further differences are that the regressions are unweighted and standard errors are obtained from 100 clustered bootstrap draws (drawing entire zip code histories with replacement), which is comparable to clustering by zip code in the baseline regression. Significance levels are indicated by *** 1%, ** 5%, * 10%.

and 19% under the Poisson count model.[52] Overall, a one standard deviation increase in zip code specific 'daily' pollution levels results in 107 additional admissions for respiratory problems and 49 additional admissions for heart problems, which are 17% and 9% of the daily mean. For respiratory problems, infants only account for roughly one third of the overall impacts. Studies focusing only on the impact on infants therefore would miss a significant portion of the overall impacts. Not surprisingly, the elderly are responsible for the largest share of heart related impacts.

Airport congestion significantly contributes to the overall impacts: a one standard deviation increase in taxi time increases respiratory and heart admissions by roughly 1% of the daily mean. At LAX, the largest airport in California, a one standard deviation increase in taxi time is responsible for roughly one-fourth of the effect of a one-standard deviation increase in pollution.

52. Recall that these estimates are smaller than what we reported under Table 4, where we increased pollution levels in each zip code by the average 'overall' standard deviation in pollution levels and took an average baseline sickness rate that was not population weighted.

TABLE 8
Impact of CO pollution on health (Model 3)

	LAX	SFO	SAN	OAK	SJC	SMF	SNA	ONT	BUR	SBA	LGB	PSP	Total
Panel A: linear probability model — all ages													
Population	812	182	540	448	910	41	822	454	794	59	875	93	6028
	One Standard Deviation increase in taxi time												
Asthma	1.20	0.16	0.29	0.15	0.21	0.01	0.26	0.09	0.12	0.00	0.10	0.01	2.60
Acute respiratory	2.44	0.34	0.59	0.30	0.44	0.02	0.53	0.19	0.25	0.01	0.20	0.02	5.31
All respiratory	3.18	0.44	0.77	0.39	0.57	0.03	0.68	0.25	0.32	0.01	0.25	0.03	6.91
Heart disease	1.40	0.19	0.34	0.17	0.25	0.01	0.30	0.11	0.14	0.00	0.11	0.01	3.04
	One Standard Deviation increase in pollution												
Asthma	4.80	0.52	4.00	1.37	5.96	0.18	5.13	1.98	5.92	0.18	6.45	0.15	36.63
Acute respiratory	9.82	1.06	8.17	2.80	12.18	0.36	10.49	4.04	12.10	0.36	13.19	0.30	74.87
All respiratory	12.78	1.38	10.63	3.64	15.85	0.47	13.65	5.26	15.75	0.47	17.17	0.40	97.44
Heart disease	5.61	0.61	4.67	1.60	6.96	0.21	5.99	2.31	6.92	0.21	7.54	0.17	42.79
Panel B: linear probability model — ages 5 years and below													
Population	54	11	33	32	68	4	58	35	55	3	65	6	424
	One Standard Deviation increase in taxi time												
Asthma	0.27	0.03	0.06	0.03	0.05	0.00	0.06	0.02	0.03	0.00	0.02	0.00	0.59
Acute fespiratory	0.87	0.11	0.19	0.11	0.17	0.01	0.20	0.08	0.09	0.00	0.08	0.01	1.91
All respiratory	1.00	0.13	0.22	0.13	0.20	0.01	0.23	0.09	0.10	0.00	0.09	0.01	2.20
Heart disease	0.03	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.07
	One Standard Deviation increase in pollution												
Asthma	1.14	0.11	0.84	0.33	1.54	0.06	1.24	0.53	1.42	0.03	1.68	0.03	8.96
Acute respiratory	3.69	0.37	2.73	1.08	4.98	0.19	4.03	1.73	4.59	0.09	5.43	0.10	28.99
All respiratory	4.25	0.42	3.14	1.24	5.74	0.22	4.63	1.99	5.28	0.11	6.25	0.11	33.37
Heart disease	0.13	0.01	0.09	0.04	0.17	0.01	0.14	0.06	0.16	0.00	0.19	0.00	1.00
Panel C: linear probability model — ages 65 years and above													
Population	82	26	54	51	88	3	79	34	79	12	89	18	615
	One Standard Deviation increase in taxi time												
Asthma	0.30	0.06	0.07	0.04	0.05	0.00	0.06	0.02	0.03	0.00	0.02	0.01	0.67
Acute respiratory	0.43	0.08	0.10	0.06	0.07	0.00	0.09	0.03	0.04	0.00	0.03	0.01	0.94
All respiratory	0.87	0.17	0.21	0.12	0.15	0.00	0.18	0.05	0.09	0.01	0.07	0.02	1.93
Heart disease	1.19	0.23	0.29	0.16	0.20	0.01	0.24	0.07	0.12	0.01	0.10	0.02	2.63
	One Standard Deviation increase in pollution												
Asthma	1.23	0.19	1.02	0.39	1.46	0.03	1.28	0.38	1.51	0.09	1.69	0.08	9.33
Acute respiratory	1.73	0.26	1.44	0.55	2.06	0.04	1.80	0.54	2.13	0.13	2.38	0.11	13.15
All respiratory	3.54	0.54	2.94	1.14	4.21	0.08	3.68	1.10	4.36	0.26	4.87	0.22	26.93
Heart disease	4.82	0.74	4.00	1.54	5.73	0.11	5.01	1.49	5.93	0.35	6.62	0.29	36.63

(*continued*)

On the other hand, smaller airports (*e.g.* Santa Barbara or Long Beach) are responsible for a much lower share of the overall pollution impacts.

4.3.7. Economic Cost. In order to monetize the health impacts associated with both pollution exposure, we use the diagnosis-specific reimbursement rates offered to hospitals through

TABLE 8
Continued

Panel D: Poisson model — all ages													
	One Standard Deviation increase in taxi time												
Asthma	1.41	0.21	0.35	0.24	0.16	0.01	0.16	0.09	0.11	0.00	0.11	0.01	2.87
Acute respiratory	2.62	0.40	0.61	0.42	0.34	0.02	0.43	0.20	0.26	0.00	0.23	0.03	5.55
All respiratory	3.44	0.53	0.81	0.54	0.45	0.02	0.55	0.26	0.34	0.01	0.31	0.04	7.30
Heart disease	1.59	0.28	0.40	0.23	0.22	0.01	0.26	0.11	0.16	0.00	0.14	0.02	3.42
	One Standard Deviation increase in pollution												
Asthma	6.35	0.67	5.29	2.29	5.02	0.22	3.60	2.00	6.22	0.09	8.54	0.14	40.41
Acute respiratory	11.47	1.35	8.93	3.98	10.22	0.33	9.09	4.48	13.99	0.19	16.81	0.39	81.24
All respiratory	15.00	1.76	12.01	5.07	13.26	0.41	11.79	5.91	18.16	0.28	22.70	0.57	106.93
Heart disease	6.72	0.87	5.85	2.13	6.30	0.16	5.62	2.48	8.62	0.23	10.03	0.30	49.31
Panel E: baseline average — all ages													
Asthma	33.1	7.9	22.3	25.4	24.2	1.6	18.1	14.9	26.0	0.9	36.0	3.0	213.6
Acute respiratory	87.4	21.7	55.2	63.2	71.8	3.6	66.9	48.0	85.8	3.1	104.3	11.8	623.0
All respiratory	121.3	30.3	78.2	85.2	98.7	4.7	91.6	67.2	117.9	4.6	149.0	18.0	866.8
Heart disease	72.8	20.3	50.0	46.4	61.7	2.4	56.9	36.9	73.4	5.0	86.1	12.3	524.2

Notes: Table gives population as well as daily hospital admissions for all zip codes that are within 10 km (6.2 miles) of one of the twelve major California airports. Panels A–D give predicted *changes* in sickness counts, while Panel E gives baseline averages. Panels A–C use the linear probability model 1 for CO from Table 4, while panel D uses the Poisson model 1 for CO from Table 7. Panel E gives average daily sickness counts in the data. The first twelve columns give impacts by airport, while the last column gives the total for all twelve airports. Population is in thousand. Predicted changes in hospitalization are for both inpatient as well as outpatient admissions.

Medicare.[53] We view this measure as a lower bound on the total health costs for several reasons: first, our methodology measures limited impacts on both a temporal and spatial scale. By focusing on day-to-day fluctuations, we do not address the long run, cumulative effect of pollution on health. If these are sizable relative to the contemporaneous effects, the overall cost estimate will be higher. Similarly, our focus has been on individuals living within 10 km of an airport. Some of our estimates suggest the marginal impact of taxi time extends beyond the 10 km radius, in which case we would be understating the overall effect. Second, we only count people that are sick enough to go to the hospital — anybody who sees their primary care physician or stays home feeling sick will not be counted. Recent work by Hanna and Oliva (2015) finds that pollution decreases labour supply in Mexico City, imposing real economic costs on society not measured in our analysis. Similarly, Deschênes *et al*. (2012) find that increased levels of ambient NO_2 lead to increased levels of spending on respiratory related prescription medicines, an outcome not measured in our analysis. Third, and most importantly, the marginal willingness to pay to avoid treatment is likely higher than the cost of treatment. For example, severe heart-related problems that are not treated within a narrow time frame will likely result in death. The statistical value of life that EPA uses for its benefit–cost analyses is around 6 million dollars, which is 1,000 times as larger as our medical reimbursement cost for heart-related problems. Individuals might be willing to pay significantly more than medical reimbursement rates to avoid illnesses

53. This information comes from a translation between our hospital diagnosis codes (ICD-9) and Diagnosis Related Group (DRG) codes. We used the crosswalk from the AMA Code Manager Online Elite. Using the set of DRG codes, we calculate the Medicare reimbursement rates using the DRG Payment calculator provided by TRICARE (http://www.tricare.mil/drgrates/). In accordance with Medicare reimbursement policy, we adjust the DRG payments using the average wage index in our sample. The average cost for respiratory problems and heart related admissions are US$ 2702 and 6501, respectively.

that, if not adequately treated, have dire consequences. Using the predicted increase in hospital visits under the linear probability model given in Table 8, a one standard deviation increase in pollution levels amounts to about a $540,000 increase in hospitalization payments related to respiratory and heart related hospital admissions under model 3.[54] Since a one-standard deviation change in pollution is an extrapolation from the fluctuations caused by airport congestion, we also analysis counterfactual where peak exposure levels are capped using the non-linear models of Supplementary Figure A5 but find comparable results that are available upon request.[55]

5. CONCLUSIONS

This study has shown how daily variation in ground level airport congestion due to network delays significantly affects both local pollution levels as well as local measures of health. In doing so, we develop a framework through which to credibly estimate the effects of exogenous shocks to local air pollution on contemporaneous measures of health. Daily local pollution shocks are caused by events that occur several thousand miles away and are arguably exogenous to the local area. We address several longstanding issues pertaining to non-random selection and behavioural responses to pollution. In addition, we show how newly available data on the universe of emergency room provides much cleaner insight as to the sensitivity of populations to ambient pollution levels, relative to existing Inpatient Discharge records. Our results suggest that ground operations at airports are responsible for a tremendous amount of local ambient air pollution. Specifically, a one standard deviation change in daily congestion at LAX is responsible for a 0.28 standard deviation increase in levels of CO next to the airport that fades out with distance. The average impact for zip codes within 10 km is 0.23 standard deviations.

When connecting these models to measures of health, we find that admissions for respiratory problems and heart disease are strongly related to these pollution changes. A one standard deviation increase in daily zip-code specific pollution levels increases asthma counts by 17% of the baseline average, total respiratory problems by 17%, and heart problems by 9%. Infants and the elderly show a higher sensitivity to pollution fluctuations, and marginal damages of pollution seem to be increasing in pollution for infants. At the same time, adults age 20–64 years are also impacted. For respiratory problems, the general adult population accounts for the majority of the total impacts despite the lower sensitivity to fluctuations as they are the largest share of the population. A one standard deviation increase in pollution levels is responsible for 540 thousand dollars in hospitalization costs for the 6 million people living within 10 km of one of the twelve airports of our study. This is likely a significant lower bound as the willingness to pay to avoid such illnesses will be higher than the Medicare reimbursement rates.

Examining various mechanisms for the observed pollution-health relationship, we find that CO is primarily responsible for the observed health effects as opposed to NO_2 or O_3. We find no evidence of forward displacement or delayed impacts of pollution. We also find no evidence that people in areas with larger pollution shocks are less susceptible or less responsive to pollution.

These estimates suggest that relatively small amounts of ambient air pollution can have substantial effects on the incidence of local respiratory illness, at least for the population that

54. The corresponding number under model 1 is $920 thousand. These figures are calculated by taking the estimated increase in hospital visits and multiplying it by the average Medicare reimbursement for each of the respective diagnoses.

55. Specifically, we test the sensitive of our results to assumed linear extrapolation through a counterfactual where all CO levels in 2005–7 are caped at half the observed mean, *i.e.* values that exceed half the historic mean are reduced to equal half the historic mean. The implied pollution reduction is evaluated both using our linear baseline model as well as a quadratic or cubic in pollution exposure. The predicted economic benefits are 520 thousand in the linear model and 650 thousand in the two non-linear models, suggesting that allowing for increasing marginal damages of pollution might give slightly larger damages.

comprises our study. While EPA recently decided against lowering the existing carbon monoxide standards due to lack of sufficient evidence of the harmful effects of CO at levels below current EPA mandates, we find significant impacts on morbidity. Recent research suggests that the rates of respiratory illness in the US are rising dramatically, even as ambient levels of air pollution have continued to fall (Center for Disease Control, 2011). Why asthma rates continue to rise is an open question, but the increase in asthma rates is most pronounced among African Americans who disproportionately live in densely populated, congested areas. At the same time, traffic congestion in cities has been rising dramatically. Results presented here suggest that at least part of the increased rate of asthma in urban areas can be explained by increased levels of traffic congestion. The exact mechanism remain beyond the scope of the current study, but this remains an interesting area for further research.[56]

Acknowledgments. We would like to thank Antonio Bento, Janet Currie, Ryan Kellogg, Mushfiq Mobarak, Matthew Neidell, Marit Rehavi, Jay Shimshack, and Christopher Timmins for comments on an earlier version of this article. All remaining errors are ours. Walker acknowledges support from the Robert Wood Johnson Foundation.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

ANGRIST, J. D. and PISCHKE, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*. (Princeton, NJ: Princeton University Press).
AUFFHAMMER, M. and KELLOGG, R. (2011), "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality", *American Economic Review*, **101**, 2687–2722.
BATTERMAN, S. A., ZHANG, K. and KONONOWECH, R. (2010), "Prediction and Analysis of Near-Road Concentrations Using a Reduced-Form Emission/Dispersion Model", *Environmental Health*, **9**, 1–18.
BAUM, C. F., SCHAFFER, M. E. and STILLMAN, S. (2003). "Instrumental Variables and GMM: Estimation and Testing", *Stata Journal*, **3**, 1–31.
BENSON, P. E. (1984), "CALINE4 - A Dispersion Model for Predicting Air Pollutant Concentrations near Roadways", *California Department of Transportation - FHWA/CA/TL-84/15*.
BOUND, J., JAEGER D. A. and BAKER, R. M. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak", *Journal of the American Statistical Association*, **90**, 443–450.
BRUNER, J. (2009), "America's Worst Intersections", *Forbes*.
CAMERON, A. C. and TRIVEDI, P. K. (1998), *Regression Analysis of Count Data.* (Cambridge, MA: Cambridge University Press).
CENTER FOR DISEASE CONTROL (2011), "Asthma in the US: Growing Every Year." *CDC Vital Signs*.
CHAY, K. Y. and GREENSTONE, M. (2003), "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession", *Quarterly Journal of Economics*, **118**, 1121–1167.
CHERNOZHUKOV, V. and HANSEN, C. (2008). "The Reduced Form: A Simple Approach to Inference with Weak Instruments", *Economics Letters*, **100**, 68–71.
CURRIE, J. and NEIDELL, M. (2005), "Air Pollution and Infant Health: What Can We Learn From California's Recent Experience?" *Quarterly Journal of Economics*, **120**, 1003–1030.
DAHL, G. B. and LOCHNER, L. (2012), "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit", *The American Economic Review*, **102**, 1927–1956.
DAVIDSON, R. and MACKINNON, J. G. (1993), *Estimation and Inference in Econometrics*. (Oxford: Oxofd University Press).
DESCHÊNES, O., GREENSTONE, M. and GURYAN, J. (2009), "Climate Change and Birth Weight." *American Economic Review: Papers & Proceedings*, **99**, 211–217.
DESCHÊNES, O., GREENSTONE, M. and SHAPIRO, J. (2012). "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program and Ozone Reductions" (Energy Institute at Haas Working Paper Series 233).

56. Currently, the highest rates of asthma incidence in the US are found in Bronx, New York (Garg *et al*., 2003). This area of northern New York City is bisected by five major highways, that rank among the most congested in the US (Bruner, 2009).

ENVIRONMENTAL PROTECTION AGENCY (1992), "Procedures for Emission Inventory Preparation. Volume IV: Mobile Sources (EPA420-R-92-009)".
—— (2005), "2005 National Emissions Inventory Data And Documentation".
FEDERAL AVIATION ADMINISTRATION (2005), "Aviation & Emissions: A Primer", Office of Environment and Energy.
—— (2009), *Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports.* http://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/, last accessed October 2010.
FRIEDMAN, M. S., POWELL, K. E., HUTWAGNER, L., *et al*. (2001). "Impact of Changes in Transportation and Commuting Behaviors During the 1996 Summer Olympic Games in Atlanta on Air Quality and Childhood Asthma", *Journal of the American Medical Association*, **285**, 897–905.
GAREN, J. (1984), "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica*, **52**, 1199–1218.
GARG, R., KARPATI, A., LEIGHTON, J. *et al*. (2003), *Asthma Facts.* Second ed., (New York: New York City Department of Health and Mental Hygiene).
GRILICHES, Z. and HAUSMAN, J. A. (1986), "Errors in Variables in Panel Data", *Journal of Econometrics*, **31**, 93–118.
GROSSMAN, M. (1972), "On the Concept of Health Capital and the Demand for Health", *Journal of Political Economy*, **80**, 223–255.
HANNA, R. and OLIVA, P. (2015), "The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City", *Journal of Public Economics*, **122**, 68–79.
HAUSMAN, J., HALL, B. H. and GRILICHES, Z. (1984), "Econometric Models for Count Data with an Application to the Patents-R & D Relationship", *Econometrica*, **52**, 909–938.
HECKMAN, J. and VYTLACIL, E. (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling", *Journal of Human Resources*, **33**, 974–987.
ITO, K., THURSTON, G. D. and SILVERMAN, R. A. (2007), "Characterization of $PM_{2.5}$, Gaseous Pollutants, and Meteorological Interactions in the Context of Time-series Health Effects Models", *Journal of Exposure Science and Environmental Epidemiology*, **17**, S45–S60.
KNITTEL, C. R., MILLER, D. L. and SANDERS, N. J. (2011), "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health" (NBER Working Paper 17222).
LINN, W. S., AVOL, E. L., PENG, R.-C. *et al*. (1987), "Replicated Dose-Response Study of Sulfur Dioxide Effects in Normal, Atopic, and Asthmatic Volunteers." *American Journal of Respiratory and Critical Care Medicine*, **136**, 1127–1135.
MORETTI, E. and NEIDELL, M. 2011. "Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles." *Journal of Human Resources*, **46**, 154–175.
MURPHY, K. M. and TOPEL, R. H. (2002), "Estimation and Inference in Two-Step Econometric Models", *Journal of Business & Economic Statistics*, **20**, 88–97.
NEIDELL, M. (2009), "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations", *Journal of Human Resources*, **44**, 450–478.
NEW YORK TIMES. January 27, 2012. *N.Y. Airports Account for Half of All Flight Delays*.
PEEL, J. L., TOLBERT, P. E., KLEIN, M. *et al*. (2005), "Ambient Air Pollution and Respiratory Emergency Department Visits", *Epidemiology*, **16**, 164–174.
PETERS, A., DOCKERY, D. W., MULLER, J. E. *et al*. (2001), "Increased Particulate Air Pollution and the Triggering of Myocardial Infarction", *Circulation*, **103**, 2810–2815.
POPE, C. A., BURNETT, R. T., KREWSKI, D., *et al*. (2009). "Cardiovascular Mortality and Exposure to Airborne Fine Particulate Matter and Cigarette Smoke Shape of the Exposure-Response Relationship", *Circulation*, **120**, 941–948.
POPE III, C. A., BURNETT, R. T., TURNER, M. C. *et al*. (2011), "Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships", *Environmental health perspectives*, **119**, 1616.
PYRGIOTISA, N., MALONEB, K. M. and ODONI, A. (2013), "Modelling Delay Propagation Within an Airport Network", *Transportation Research Part C: Emerging Technologies*, **27**, 60–75.
SCHILDCROUT, J. S., SHEPPARD, L., LUMLEY, T. *et al*. (2006), "Ambient Air Pollution and Asthma Exacerbations in Children: An Eight-CityAnalysis", *American Journal of Epidemiology*, **164**, 505–517.
SCHLENKER, W. and ROBERTS, M. J. (2009), "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change", *Proceedings of the National Academy of Sciences of the United States*, **106**, 15594–15598.
SCHLENKER, W. and WALKER, W. R. (2011), "Airports, Air Pollution, and Contemporaneous Health" (NBER Working Paper 17684).
SCHWARTZ, J., SPIX, C., TOULOUMI, G. *et al*. (1996), "Methodological Issues in Studies of Air Pollution and Daily Counts of Deaths or Hospital Admissions", *Journal of Epidemiology and Community Health*, **50**, S3–S11.
SEATON, A., GODDEN, D., MACNEE, W. *et al*. (1995), "Particulate Air Pollution and Acute Health Effects", *The Lancet*, **345**, 176–178.
STOCK, J. H., WRIGHT, J. H. and YOGO, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business & Economic Statistics*, **20**, 518–529.

TRANSPORTATION RESEARCH BOARD. (2008), "Aircraft and Airport-Related Hazardous Air Pollutants: Research Needs and Analysis", *National Academy of Sciences, Airport Cooperative Research Program (ACRP) Report 7*.
UNITED STATES ENVIRONMENTAL PROTECTION AGENCY. (2004), "Guidance on Airport Emissions Reduction Credits for Early Measures Through Voluntary Airport Low Emissions Programs." Air Quality Strategies and Standards Division. Office of Air Quality Planning and Standards.
WELMAN, S., WILLIAM, A. and HECHTMAN, D. (2010), "Calculating Delay Propagation Multipliers for Cost-Benefit Analysis", *Center for Advanced Aviation System Development and MITRE MP 100039*.
WILLIS, R. J. and ROSEN, S. (1979), "Education and Self-Selection", *Journal of Political Economy*, **87**, S7–S36.
WOOLDRIDGE, J. M. (1997), "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model", *Economics Letters*, **56**, 129–133.
—— (1999). "Distribution-Free Estimation of Some Nonlinear Panel Data Models", *Journal of Econometrics*, **90**, 77–97.
—— (2002), *Econometric analysis of Cross Section and Panel Data.* (MIT Press).
—— (2003), "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model", *Economics Letters*, **79**, 185–191.