Airlines, Pollution, and Fertility*

Xinming Du†

Charles A Taylor‡

July 2024

Abstract

This paper demonstrates a large but little-known negative externality of the aviation industry. Using a new instrument for air pollution from aircraft cruising, we show that pollution is higher beneath overhead flight routes in ways uncorrelated with local pollution. We combine this cross-sectional variation with the launch of new flight routes to establish several findings. First, aircraft cruising persistently elevates local PM2.5 by 1-3 μg/m^3. Second, PM2.5 has adverse impacts on infant health via lower birth weights, including in 44 developing countries where data are scarce. Third, we leverage the fact that propeller planes still use leaded fuel to show that 1 ng/m^3 ambient lead reduces fertility rates by 0.19%. Fourth, we generalize this in relation to the historical phase-out of leaded fuel in vehicles, which our analysis suggests added over 2 million people per year to the global population—making it among the most material public health interventions. We provide this global gridded airline data product for use in future research.

*We thank Doug Almond, Samuel Bazzi, Susanna Berkouwer, Roger Fouquet, Meredith Fowlie, Ludovica Gazze, Shanjun Li, Jessica Reyes, Sefi Roth, Alberto Salvo, Edson Severnini, Wolfram Schlenker, and participants at the AERE, LSE/Imperial Workshop on Environmental Economics, Econometric Society North American meeting, and AEW for helpful suggestions. Raphael Perot and Jiaheng Zhao provided excellent research assistance.

†National University of Singapore, Department of Economics, xdu@nus.edu.sg

‡Harvard University, Harvard Kennedy School, ctaylor@hks.harvard.edu

1 Introduction

As a major user of fossil fuels, aviation's contribution to global warming is well understood. However, much less is known about its local pollution effects, particularly in areas beyond airports. This paper focuses on the pollution externality of cruising, the overhead flying of aircraft between locations. This setting is ideal for studying pollution's impact across a range of outcomes (e.g., infant health, fertility) and sub-populations, including a host of developing countries—providing useful insights into how pollution impacts differ by pollutant type, baseline pollution levels, demographics, and other characteristics.

Air pollution is highly studied. Its large and negative effects on human health and productivity are well documented. But a challenge shared by all studies of the causal impact of pollution is the non-randomness of exposure. Pollution intensity is correlated with socioeconomic factors that drive health outcomes; for instance, at-risk populations are more likely to live and work in highly polluted areas. The empirical literature has addressed these selection concerns through the use of plausibly random variation exposure due to wind (Deryugina et al., 2019), traffic patterns (Currie and Walker, 2011; Schlenker and Walker, 2015), as well as variation across time and space in regulatory policy (Greenstone and Hanna, 2014) and industrial entry and exit (Currie et al., 2015).

These innovative research designs, however, align to specific time periods, geographies, and policy contexts, and thus may be limited in their external validity—particularly in the developing world. And in some settings, concerns remain about spatial sorting and endogenous drivers of regulation, as well as endogenous responses by individuals to regulatory change.

To this end, we introduce a new instrument for air pollution from aircraft cruising derived from the global airline network, which can be visualized in Figure 1. Aviation accounts for over 2% of global annual CO_2 emissions,[1] and the fuel combustion that emits CO_2 also produces air pollution. Much of this pollution is emitted during the cruising phase, which accounts for over two-thirds of fuel consumption, as opposed to the takeoff and landing phases (OAG, 2022).[2] But unlike takeoff and landing, which occur near population centers, cruising emissions are largely unregulated and unmonitored. We show that air pollution is persistently elevated among populations beneath overhead flight routes across the

[1] Aviation has contributed to 4% of global warming—a higher proportion due to non-CO_2 emissions of nitrogen oxides and water vapor (Klöwer et al., 2021).

[2] While some particulate matter from cruising aircraft will be disbursed broadly via atmospheric currents and weather conditions, on average, much will land in the vicinity directly beneath the flight path, especially due to wet deposition (i.e., binding to rain and falling to the ground).

world. Given the well-established adverse impacts of air pollution, this represents a large but little-known negative externality of the aviation industry.

A key insight for identification is that flight paths are generally determined by the shortest geodesic distances (i.e., "great circle route") between two airports in such a way that a given location's proximity to a flight path is random and thus exogenous to factors correlated with pollution exposure. For intuition, Figure A1 shows the flight route between Los Angeles and New York, which has dozens of direct flights per day. The shortest route happens to transect a certain part of Kansas. The idea is that the areas just north and just south of this line should not be systematically different, which we verify later.

One concern is that areas near airports where flights are concentrated and pollution may be higher for non-aviation reasons. Most empirical studies on aviation-based pollution focus on areas near airports, and thus air pollution from aircraft takeoff and landing—rather than from cruising (Schlenker and Walker, 2015; Zahran et al., 2017). This poses spatial sorting concerns related to the composition of neighborhoods near airports. Areas surrounding airports tend to be less urban with lower income and education levels compared to other parts of the same city, a fact partly attributed to noise and air pollution disamenities (Nelson, 1979; Sobotta et al., 2007; Tonne et al., 2018). For across-city comparisons, airports tend to locate in more developed cities that benefit from connectivity and have generally higher pollution (Cidell, 2015).

To address such selection concerns, we drop locations within 100km of airports in all our analyses, though we show our results are robust to their inclusion.[3] To empirically test the randomness of overhead airline assignment, we compare pixels with high and low airline intensity by income, education, and other variables. Tables A18 (global) and A19 (US) confirm that there is no systematic correlation with observables that may otherwise relate to pollution exposure in areas away from airports—providing evidence that a location's exposure to overhead airline flights is effectively random.

To establish the validity of our instrument, we demonstrate the strong first-stage relationship between overhead flight route intensity and ground-level measures of air pollution (PM2.5). Figure 2 visualizes the variation in the raw data, whereby PM2.5 is persistently 1-3 μg/m^3 higher, on average, in high versus low flight intensity locations. This represents a material increase over average PM2.5 levels in our sample areas of 10 μg/m^3 (US) and 19 μg/m^3 (global). Combining cross-sectional variation in airline intensity across space with the launch of flight routes over time, we confirm the first stage holds using several

[3] Figure A3 shows 100km buffers near airports that are dropped from our sample in the robustness check.

empirical designs: i) a panel model approach with the universe of all airline routes, ii) an event study design involving the launch of new flight routes, and iii) a natural experiment involving the grounding of flights after 9/11.

We confirm that our first stage results hold with the following robustness tests: using different distance cutoffs ranging from 0km (i.e., no observations dropped) to 1,000km from airports; controlling for aircraft altitudes that may affect surface air pollution, as well as wind and turbulence conditions that may affect pollution dispersion; using alternative air pollution proxies such as aerosol levels as an outcome variable. And finally, to address concerns that our simulated shortest-distance flight paths between origin and destination may not precisely reflect actual airline routes, we replicate results using overhead aircraft counts obtained from OpenSky sensor data.

Using our instrument, we estimate the impacts of air pollution on health across the world, including in 44 developing countries where data are scarce. We first perform an IV regression to estimate the impact of in-utero pollution exposure on infant health outcomes using USAID Demographic and Health Surveys (DHS). We find that an increase in PM2.5 by $1\mu g/m^3$ (2.3% over the DHS sample mean) increases the incidence of low birth weight by one percentage point, which is 40% above the baseline incidence of 2.3%. We confirm that this effect is not driven by potential confounders such as noise pollution from overhead cruising aircraft.

In the second part of the paper, we turn from PM2.5 to a specific air pollutant, lead. We use an identification strategy that leverages both our instrument and the type of aircraft engine. Approximately 40% of today's aircraft have piston-driven engines (i.e., propeller planes) that use leaded fuel, while 60% have jet engines that use lead-free kerosene fuels. We thus differentiate between propeller plane flight routes (i.e., lead emitters) from all others, which can be visualized in the map at the bottom of Figure 1. Using the US EPA lead monitors, we find a strong first stage where propeller routes are associated with higher ambient lead levels.

Focusing on fertility, an outcome linked to lead exposure in the literature, we find that a 1 ng/m^3 ambient lead reduces fertility rates by 0.19%. This approach allows us to isolate the effect of lead exposure on fertility after controlling for pollution and other unobservables that may be associated with aircraft fuel combustion more generally (e.g., PM2.5, carbon dioxide, nitrogen oxides, and carbon monoxide).

We conduct several robustness checks and extensions that confirm our core results. These include: varying the model specification, weighting approach, and demographic controls

and time trends employed; dropping DHS clusters and air lead monitors close to airports in much the same spirit as the first stage; and accounting for road intensity as an additional control to account for vehicular emissions. We then use blood lead data from to provide evidence of a direct lead exposure channel.

To show our results are not an artifact of DHS survey data, we analyze the fertility response in the US using county-level birth rates derived from vital statistics. We find similar patterns whereby proximity to leaded airline routes leads to a decline in birth rates in the US. Finally, we investigate how the link between lead exposure and fertility varies across countries. Running our model separately for each country, we find that results hold in all but three countries, minimizing concerns about external validity.

In a third part of the paper, we generalize our finding of lead's impact on fertility beyond aviation in the context of a historical policy. We leverage the timing of lead gasoline bans for automobiles in countries around the world, starting with Japan in 1986 and ending with Algeria in 2021. As with our analysis of PM2.5's effect on infant health, we use DHS data across 44 countries representing a population of 3.37 billion[4]. We calculate a measure of fertility at the individual-year level. For identification, we use an approach similar in spirit to Clay et al. (2021), computing for each DHS cluster the number of roads across a 0.1-degree grid. We find the fertility impact of the lead ban to be a function of road intensity, confirming the intuition that lead exposure (and the benefit of a ban) would be concentrated among those in closer proximity to roads where vehicles combusted leaded fuel. A back-of-the-envelope calculation implies that this policy increased global population by 2.2 million people per year—making it among the most material public health interventions in history.

Our paper makes several contributions to the literature. At a high level, we uncover a large but little-known negative externality of the aviation industry: air pollution from aircraft cruising. We document aviation's effect on infant health across the developing world via elevated PM2.5, as well as reduced fertility via airborne lead exposure from propeller planes. Outside of the immediate airport vicinity, these sources of pollution are largely unregulated.

This paper also contributes to understanding the unprecedented decline in fertility across the world. This trend extends to developing countries, where in recent decades fertility rates have converged with those of rich countries. While the secular decline is driven primarily by economic and social changes, in many places there has emerged a gap between

[4] Total based on the 2022 population of the 44 countries in the DHS sample.

desired and attained fertility (Beaujouan and Berghammer, 2019). This paper sheds light on the potential environmental factors that contribute to this "fertility gap" by impairing fecundity (i.e., reduced ability to conceive due to physical issues)—an area of research with relatively little quasi-experimental evidence.

We are the first to estimate the fertility impact of the ongoing use of leaded fuel in the aviation industry. Others have identified the link between exposure to propeller planes and blood lead levels (Zahran et al., 2017, 2023) in populations in close proximity to specific airports—but no one has established the link between overhead propeller flights on lead exposure in the general population, and then further linked this to fertility outcomes. The large fertility effect we find is relevant for the policy discussion around the phase-out of leaded fuel in aviation.[5] Additionally, extending on Clay et al. (2021), our paper is the first to estimate the impact of vehicle lead bans globally, with a particular focus on developing countries.

We also plan to release our airline flight route instrument as a gridded global data product. As discussed above, a key challenge for empirical health researchers seeking to establish causal effects of pollution is the non-randomness of exposure. We hope this exogenous source of variation in pollution exposure——encompassing both common air pollutants like PM2.5 as well as lead——can be used in other research contexts, particularly where data are scarce. Given the minimal research on the subject of air pollution from aircraft cruising, and the corresponding lack of awareness of its potential risk among the general population, we argue that residential sorting along overhead airline routes is not driving this relationship.

Our paper proceeds as follows: Section 2 provides background on global fertility trends and their potential connection to pollution and other environmental factors. We also discuss the literature on lead, specifically, and its impact on fertility and health outcomes. Section 3 describes how we construct our measures of fertility, pollution, and the airline route instrument, as well as other datasets used in this analysis. Section 4 establishes the strength and validity of our airline instrument and describes our identification strategies. Section 5 presents our regression results along with robustness tests, and Section 6 expands our fertility findings in relation to the global phase-out of leaded vehicle fuel. Results are discussed in Section 7 by exploring the scientific and policy implications of our study—as well as further avenues of research using the data products we developed.

[5] See recent EPA on leaded fuel in aviation: https://www.epa.gov/newsreleases/epa-determines-lead-emissions-aircraft-engines-cause-or-contribute-air-pollution.

2 Background

2.1 Global fertility trends

The total fertility rate (TFR) is defined as the average number of children that a woman would bear throughout her lifetime, assuming she experienced current age-specific fertility rates (i.e., the number of live births per woman across age cohorts spanning 15-49). Fertility rates influence both the size and composition of a population. Over the past 250 years since the Industrial Revolution, fertility, together with economic outputs, has increased rapidly worldwide. However, in the past 50 years, the global fertility rate has halved. In 2000, the world's fertility rate was 2.7 births per woman, comfortably above the "replacement rate" of 2.1, at which a population remains stable. Today it has decreased to 2.3 and continues to decline, with over half of the global population residing in regions with fertility rates below replacement level.[6]

Fertility rates are shaped by a complex set of social, economic, cultural, and environmental factors. Economic development generally spurs a decline in fertility (Hafner and Mayer-Foulkes, 2013), as observed in more developed economies where urbanization, women's education, and changing lifestyles contribute to more family planning (Upadhyay et al., 2014; Atake and Gnakou Ali, 2019). Cultural and religious beliefs, social norms, government policies, and female empowerment all play important roles in shaping reproductive choices. One discernible trend is the shift in fertility schedules towards increasingly higher maternal ages, with a considerable proportion of women having their first kid at age 30 and older.[7] Infant and child mortality rates further influence population dynamics, and improved access to reproductive and maternal healthcare increases child survival rates and can change perceptions of the ideal family size (Ackerson and Zielinski, 2017).

Fertility change has far-reaching socioeconomic impacts. In terms of demographic shifts, a decline in fertility leads to an aging population, presenting challenges such as increased de-

[6] Fertility patterns vary across countries and over time. In 2021, Niger has the highest fertility rate at 6.9, indicating that the average woman in Niger will bear seven children in her lifetime. Most countries with the highest fertility rates are located in Africa. In contrast, South Korea has the lowest fertility rate at 0.84. Several of the world's most populous nations, including China, India, and the US, have below-replacement levels of fertility, a trend observed in parts of Europe and North America since the 1970s.

[7] Among US mothers in the 1930s, 2.5% had a first birth after 35. The proportion increases to 9-10% for cohorts born between 1956 and 1960 (Mathews and Hamilton, 2014). By 2016, the fertility rate for the 30–34 age group was higher than that of 25–29 for the first time. 30% of all first births in the US occurred among women over 30 (Martin et al., 2019). Similar increases in maternal ages are documented in the EU (Eurostat, 2021), UK (Goisis, 2023), Asia (Kato et al., 2017; OECD, 2019), Africa (Negash and Asmamaw, 2022), and other countries covered by the DHS (Bongaarts and Blanc, 2015).

pendency ratios and support systems for the elderly. In contrast, a sudden surge in fertility can result in a young population, affecting education systems and necessitating investments in infrastructure and job creation. Fertility change also plays a role in shaping economic dynamics: lower fertility rates usually lead to higher levels of female participation in the workforce (Ahn and Mira, 2002; Adsera, 2004), contributing to economic growth. Additionally, family size impacts household expenditure patterns, influencing demand for goods and services. On a broader scale, fertility change affects national and global population sizes, affecting innovation and investment, resource utilization, environmental sustainability, migration flows, and social welfare programs.

2.2 Pollution and fertility

Environmental factors such as pollution are increasingly recognized as potential contributors to fertility patterns (Clay et al., 2021; Gao et al., 2022). Many studies have explored the physiological and epidemiological links between pollution and fertility, which vary by sex. For females, Canipari et al. (2020) classify environmental factors affecting fertility into heavy metals, air pollutants, and endocrine disruptors.[8] Heavy metals adversely affect steroidogenic function, leading to fetal abnormalities and embryotoxicity (Aquino et al., 2012), and oxidative stress induced by heavy metals alters hormone function and embryo quality, contributing to female infertility (Rzymski et al., 2015).[9] Air pollutants from non-metal sources can lead to abnormal gametogenesis and diminished reproductive performance (Carré et al., 2017; Conforti et al., 2018). For PM2.5, a $10\mu g/m^3$ increase is associated with a 2% increase in female infertility (Xue and Zhang, 2018).[10]

[8] Endocrine disruptors, chemicals that interfere with endocrine system function, can mimic or block the action of endogenous hormones. Such disruptions increase the risks of cancer, birth defects, and developmental disorders. Chlorinated hydrocarbons, a type of endocrine disruptor, can disturb follicular steroidogenesis by modifying hormonal properties, synthesis, and function. These compounds are found in various human reproductive components, and usually lead to decreases in estradiol secretion by antral follicles (Gregoraszczuk et al., 2013). Exposure to polychlorinated biphenyls, a specific class of chlorinated hydrocarbons, has been linked to decreased concentrations of anti-Mullerian hormone and interference with oocyte quality, fertilization, implantation, and embryo quality (Karwacka et al., 2019).

[9] Exposure to heavy metals during pregnancy results in placental oxidative stress, which is linked to preterm birth (Singh et al., 2020). Metals can cross the placental barrier and pose risks to the fetus. Studies show an association between heavy metal concentration in the placenta and abnormal fetal growth and development, as well as fetal damage (Falcon et al., 2003; Llanos and Ronco, 2009).

[10] For air pollution, adverse effects include abnormal endocrine function, oxidative stress, and inflammation, which primarily affect ovarian function (De Coster et al., 2012). Studies reveal that women residing in highly-industrialized areas experience a notable reduction in antral follicle count, a decrease in the number of fertilizable oocytes, and a higher incidence of implantation failure compared to those in less polluted areas (Conforti et al., 2018). Additionally, PM2.5 has been identified as compromising oocyte quality and endocrine distributors.

Environmental pollution also affects male fertility. Kumar and Singh (2022) reviews the literature, categorizing pollution into hazardous chemicals, air pollution, working environment, and radiation.[11] Air pollution is negatively associated with semen volume, sperm concentration, motility, and normal morphology, leading to elevated sperm DNA fragmentation and decreased fertility (Jurewicz et al., 2018; Zhang et al., 2020). Gaseous pollutants like SO_2 and NO_2 negatively affect sperm quality. Ozone exposure decreases the proportion of sperm with normal morphology, contributing to the increased incidence of abnormal sperm morphology in males seeking infertility treatment (Wdowiak et al., 2019). PM2.5 has been negatively linked to sperm motility, concentration, total count, head morphology, and overall semen quality (Hansen et al., 2010).

2.3 Related literature on lead impacts

Lead exposure affects health, birth outcomes, and fertility. Historical records in the US from the late 19^{th} to early 20^{th} century show a 39% higher infant mortality rate in areas with leaded water pipes compared with those without (Troesken, 2008). Analyzing data from 172 cities between 1900 and 1920, Clay et al. (2014) find that the average water lead content in regions with leaded pipes increased infant mortality by 19% relative to the mean mortality rate. Recent years have seen a notable decrease in fugitive lead emissions from 1988 to 2018 due to policies like the toxic release inventories, which has led to a reduction in infant deaths (Clay et al., 2022).

In terms of fertility, the Flint water crisis, which was characterized by lead exposure, reduced the number of births by 7.5 per 1,000 women, constituting 12% of the average fertility rate (Grossman and Slusky, 2019). The elimination of lead from gasoline has been linked to increased fertility in the US: the observed reduction in airborne lead corresponded with four additional births per 1,000 women, accounting for 6% of the mean fertility rate (Clay et al., 2021). Furthermore, lead exposure affects birth outcomes. The Flint water crisis is associated with a reduction in birth weights by 32 to 49 grams (Abouk and Adams, 2018; Wang et al., 2021). Additionally, a shift in US policy stringency regarding ambient

[11] For males, environmental pollution compromises reproduction processes including spermatogenesis, steroidogenesis, sertoli cell, and sperm functions (Selvaraju et al., 2021). Male reproductive organs are particularly vulnerable to contaminants from environmental chemicals, leading to male infertility (Dissanayake et al., 2019). A recent cross-sectional study investigated maternal occupational exposure to potential endocrine-disrupting chemicals during pregnancy, notably pesticides, phthalates, and heavy metals, and their impact on the semen quality of their adult sons. The study revealed a significant correlation between maternal occupational exposure and low semen volume, as well as a reduced total sperm count in their sons. A notable association was found between maternal heavy metal exposure and low sperm concentration (Istvan et al., 2021).

airborne lead pollution, leading to the relocation of lead battery recycling from the US to Mexico, resulted in a 24-gram decrease in birth weight for infants born within two miles of Mexican recycling plants (Tanaka et al., 2022).

Apart from health, lead exposure also affects education and learning performance. Empirical evidence has shown the detrimental effects of lead on aptitude test scores, educational attainment, and labor outcomes (Ferrie et al., 2012; Reyes, 2015; Aizer and Currie, 2019; Grönqvist et al., 2020). Early-life exposure to lead has also been investigated as a potential factor contributing to shifts in violent crime trends and antisocial behaviors (Reyes, 2007, 2015; Aizer and Currie, 2019).

Variation in anthropogenic sources of lead, the primary contributor to human exposure, has been exploited for natural experiments. In the context of water, studies have exploited the presence of lead pipes (Ferrie et al., 2012; Clay et al., 2014), examined lead-related drinking water crises in places like Flint and Newark (Grossman and Slusky, 2019; Dave and Yang, 2022), and explored the effects of lead pipeline replacement (Marcus, 2023). Others have focused on lead paint in a school context: studies have identified lead impacts using classroom-level lead variation (Sauve-Syed, 2023) and comparing within-siblings performance (Gazze et al., 2021). Focusing on industrial sources, studies have examined locations in proximity to plants with toxic release inventories (Clay et al., 2022) or near battery recycling sites (Tanaka et al., 2022) to explore lead impacts.

Closer to our paper, the existing literature has examined traffic-related lead emissions. Some studies have taken advantage of regulatory changes in NASCAR gasoline content (Hollingsworth and Rudik, 2021; Hollingsworth et al., 2020) or proximity to airports (Zahran et al., 2017) as sources of variation. Other papers concentrate on the phase-out of the ban on leaded gasoline in developed nations such as the US (Aizer and Currie, 2019; Clay et al., 2019, 2021) and Sweden (Grönqvist et al., 2020). In contrast, our paper is the first to explore the global phase-out of leaded gasoline, with a particular emphasis on its effects in developing countries.

3 Data

We combine several datasets from satellite, ground-based monitors, radar trackers, socioeconomic surveys, and vital statistics. This enables us to construct measures and test robustness for our airline intensity treatment, first stage pollution effects, and downstream health outcomes. Table 1 reports the time range, space coverage, and frequency of the dif-

ferent datasets.

Fertility and infant health: The Demographic and Health Survey (DHS) includes birth records from 44 countries obtained through 69 surveys, with a total number of 1,528,451 female respondents. Each interviewee's record includes her month of birth, the month of birth of her children, health conditions, and household characteristics. Our fertility outcome of interest is constructed using birth month to create a birth dummy, and our main infant health outcome is birth weight.

Household locations are recorded as coordinates, representing centroids of 10km clusters. Households within a 10km radius of each centroid share identical coordinates. Our analysis on the impact of leaded gasoline bans specifically focuses on births occurring within 3 years preceding the ban to 5 years following the ban, a total of 9 years. Figure A4 visualizes the locations of DHS clusters in our sample. The clusters are mainly located in Southeast Asia, Africa, and Latin America, and there are notable variations in coverage within each country.

In addition, we incorporate fertility rates in the US from the National Vital Statistics System (NVSS). We aggregate individual-level birth records at the county-year level for the extensive period from 1968 to 2021. Our primary outcomes of interest include the birth rate, defined as the total number of births divided by the female population aged 15-49. Furthermore, we use the infant mortality rate to investigate the impacts of air pollution.

Airline route intensity: Historical airline route information was purchased from OAG Aviation, an aggregator of digital flight information. We obtain annual lists of all airlines, encompassing both passenger and freight carriers, 1997-2023. Each record includes the airline number, operating company, aircraft type, origin airport, destination, and frequency (the annual flight count). Airline routes are simulated based on the shortest trajectory between origin and destination airports. There are over 65,000 distinct geographic flight routes in a given year.

We observe aircraft engine models and categorize engines into propeller and jet planes. Because propeller planes still use leaded fuel and jet engines use unleaded fuel, we differentiate between these two types of routes and classify them as leaded and unleaded routes. In Figure 1, we plot the locations of 58,613 unleaded and 7,899 leaded routes. Most unleaded routes are domestic and have short distances.

Figure A5 illustrates the changes in airlines over time. In the blue lines, we drop duplicated airlines and focus on extensive margins, i.e., the presence of an airline operating between origin-destination airports. There is a consistent growth in airport connectivity,

particularly between 2004 and 2019. Connectivity decreased during COVID but recovered in 2022. Red lines illustrate the intensive margin by summing the frequency of all airlines. The increase in frequency corresponds closely with the airline count but experiences a more pronounced COVID-related drop in 2020. This implies that despite airlines continuing operations between two airports, there was a notable reduction in the overall number of flights during the COVID period.[12] The bottom panel focuses on new airlines operating in a given year that were not operational in the previous year. The variation in the number of new airlines exhibits a consistent pattern over time, with a slightly higher number between 2003 and 2008. This pattern implies a steady expansion of connectivity, consistent with the slope in the left figure.

We complement our simulated routes with data from OpenSky, which monitors the real-time locations of operating aircraft. The dataset encompasses over 30 trillion ADS-B, Mode S, TCAS, and FLARM messages collected from a network of more than 6,000 sensors worldwide. Historical data are available from 2016, and the time frequency is at the millisecond level. In each record, we observe each aircraft's coordinates, altitude, speed, heading direction, engine type, airline number, and operating company. We aggregate the microdata as the count of overhead aircraft at the grid-week level for each 0.1-degree pixel ($\sim$10km^2) to match the resolution of the global PM2.5 dataset described next.

Air pollution: We obtain PM2.5 data from the US EPA. The data are available at the monitor-hour level 1998-2023. There are 2,344 monitors distributed across 1,020 counties, covering all states. Locations of the EPA monitors are presented in Figure A2.

We also use global PM2.5 data from van Donkelaar et al. (2021). The authors employed satellite-based aerosol optical depth (AOD) as inputs. To extrapolate near-surface PM2.5 concentrations, a chemical transport model was used to simulate the relationship between AOD and ground-level PM2.5. The final dataset was produced by statistically fusing this modeled data with ground-based monitor data. The PM2.5 data are presented at the pixel-month level 1998-2021, with a spatial resolution of 0.1 degree. For robustness we also use NASA's MERRA-2 AOD product as an alternative measure of global air pollution.

Environmental lead: Ambient air lead data come from the US EPA monitor. There are 2,467 air quality monitors measuring suspended lead particles. Locations of the EPA air lead monitors are shown in Figure A2.

[12] Some routes are classified as having the same origin and destination airport, which may reflect training flights, private charter flights, or site-seeing flights, for which there is no obvious way to designate a flight path. However, such flights only account for 48,000 of the over 825 million individual flights logged in the OAG dataset, or less than 0.006%.

We also use soil lead data sourced from the USGS Background Soil Lead Survey. This cross-sectional survey spans 4,857 labs or sampling sites. Each dataset entry includes sample collection time, collection site coordinates, layer designation (Top5, Ahorizon, Chorizon[13]), lead concentration, and concentrations of other metals.

Blood lead: Blood lead information is obtained from the Department of Public Health of Massachusetts, where the state mandates reporting of blood lead testing results (105 CMR §460.070) under its Childhood Lead Poisoning Prevention Program. The department conducts annual testing on a representative sample of children aged 9-12 months to 47 months. Specimens are analyzed at laboratories operating under Clinical Laboratories Improvement Act standards for blood lead analyses. In cases where children either test above a specified threshold or reside in high-risk communities, an additional follow-up test is conducted at the age of four.

Our analysis focuses on the detect rate of high blood lead, defined as the number or proportion of cases with confirmed blood lead levels exceeding 10μg/dL. This threshold is also used by the department to identify high-risk sub-counties. Geolocations are recorded as census tracts.

Road intensity: Road intensity data comes from the Global Roads Open Access Data Set (gROADSv1), created by the Center for International Earth Science Information Network (CIESIN), Columbia University. The dataset gathers the most reliable road data from various countries into a comprehensive global road coverage. The road networks of all countries have been topologically linked at their borders, and internal topology has been adjusted for many countries. The representation of road networks spans from the 1980s to 2010, contingent upon the country (with most countries lacking a confirmed date), and spatial accuracy levels may vary.

Figure A4 presents a visualization of the road data. Each road segment is represented as spatial lines. There is high road intensity in developed countries and regions, while lower intensity is observed in forest or desert areas. Despite detailed locations, road characteristics such as age, highway, width, or pavement type are not observed in the dataset.

[13] Top5 indicates the soil sample is collected from a depth of 0 to 5 centimeters; Ahorizon represents the topsoil where primarily minerals from the parent material are mixed with organic matter; Chorizon refers to the deposit at Earth's surface from which the soil developed.

4 First stage: Airline routes as pollution instrument

4.1 Validity of instrument

The paper proposes a new instrument to address concerns about the non-randomness of pollution exposure. Pollution intensity is correlated with socioeconomic factors that drive health outcomes, such that at-risk populations are more likely to live in highly polluted areas and work in occupations with high pollution exposure (Shapiro, 2022; Currie et al., 2023). The empirical literature has addressed these selection concerns through the use of plausibly-random variation exposure due to wind (Schlenker and Walker, 2015; Deryugina et al., 2019), traffic patterns (Currie and Walker, 2011), as well as variation across time and space in regulatory policy (Greenstone and Hanna, 2014) and industrial entry and exit (Currie et al., 2015).

These innovative research designs, however, encompass specific time periods and geographies, and thus may be limited in their external validity—particularly in the developing world. And in some settings, concerns may remain about spatial sorting (Heblich et al., 2021; Chen et al., 2022) and endogenous drivers of regulation, as well as endogenous responses by individuals to regulatory change (e.g. Zou, 2021; Axbard and Deng, 2024).

To this end, we propose a new instrument, overhead airline intensity, as a source of exogenous pollution exposure. Airline routes should be associated with high air pollution due to fuel combustion. Flight cruising accounts for the majority of aircraft fuel use (62-92% of total), as opposed to the takeoff and landing phases (OAG, 2022). Engineering papers have sought to estimate aviation air pollution and its impacts. Barrett et al. (2010) show that cruising above 35,000 feetaccounts for 80% of aviation's air pollution impact and 1% of air quality-related premature mortality from all sources. An increase in airborne ultrafine particles has been documented along airline routes in areas far from airports and in both the upwind and downwind directions (Austin et al., 2021), suggesting that cruising emissions influences pollution levels along airline trajectories.

Since the chemical composition and size of PM influences its toxicity, the external validity of an aviation instrument for air pollution depends on the similarity of aircraft-generated PM relative to other commonly-studied sources (e.g., vehicle exhaust, industrial emissions, biomass burning). While we do not explicitly quantity these differences, we note that the fossil fuels used in aircraft engines (formulations of kerosene for jets and gasoline for propeller planes) produce PM precursor pollutants similar to other combustion sources. Thus the health impacts of the secondary aerosols (H_2SO_4, HNO_3, NH_3) are likely comparable

assuming no differential toxicity (Barrett et al., 2010).

Most empirical studies on air pollution from aviation, on the other hand, focus on areas near airports, and thus pollution from aircraft takeoff and landing–rather than from cruising (Schlenker and Walker, 2015; Zahran et al., 2017). This poses concerns about spatial sorting and the composition of neighborhoods near airports, which tend, on average, to be less urban with lower income and education compared with other neighborhoods in the same city. To address such selection concerns, we drop locations within 100km of airports in all our analyses, though we show that results are robust to their inclusion, as well as more distant cut-off thresholds.

We first look to the raw data to assess the strength of relationship between overheard flight intensity and pollution. Figure 2 is a binscatter plot for the world (left) and the US (right) using PM2.5 estimates from the satellite-derived gridded product (van Donkelaar et al., 2021) and EPA's on-the-ground monitors, respectively. In both cases the data show a clear positive relationship, whereby high intensity grid cells have PM2.5 levels 1-3 μg/m^3 higher than low intensity grid cells. The relationship holds when including all grid cells in the sample, or when including only those located over 100km from airports.

We argue that once away from airports, a location's exposure to overhead airline flights is effectively random. This is true because air pollution along cruising routes does not factor into the decision-making processes of either airlines or air traffic control entities. When designing routes, connectivity between two cities is their primary objective. After the origins and destinations are set, routes are chosen based on fuel consumption where the shortest distance is generally optimal,[14] as well as horizontal and vertical separation from other flight routes. The intuition behind our research design is illustrated in Figure A1. While our primary analysis computes overhead flight intensity using the geodesic distance between each origin and destination (i.e., the shortest path), we replicate our analyses with an alternative dataset compiled from actual airline routes derived from transponders, finding similar results.

To empirically test the randomness of overhead airline assignment, we compare pixels with high and low airline intensity. To do so, we classify DHS clusters into two groups with airline intensity above or below the median and compare income, education, and nutrition-related variables. In Table A18, Panel A shows that respondents living in high airline intensity clusters have a slightly higher wealth index by 10%. For other characteristics,

[14] Airline companies fly the shortest flight route possible. These routes are usually the fastest and have lower fuel consumption because the plane is flying less distance. Source: https://www.oag.com/blog/great-circle-routes-flight-paths

high-intensity clusters have slightly lower education and nutrition levels but better access to healthcare compared with those living in low airline intensity clusters; however, none of these three differences are statistically significant. In Panel B, when dropping clusters close to airports, there is no significant difference in any variables. Thus, we conclude that overhead airline intensity is not systematically correlated with observable characteristics that may be otherwise correlated with pollution exposure.

In a similar vein, we examine the similarity of covariates in US census tracts with high and low airline intensities, using data from the American Community Survey. T-test results are presented in Table A19. When compared to census tracts with low overhead airline routes, those with high intensity show no differences in working hours, educational attainments, demographic composition, and health status, but have higher income levels. In Panel B, after excluding census tracts with airports or those within 100km of any airports, sample sizes decrease due to the substantial number of airports in the US. We still observe large p-values for all covariates, suggesting that the difference between high and low intensity census tracts is not statistically significant. These balance tests support our hypothesis that overhead airline intensity is as good as randomly assigned in the US and DHS countries.

More generally, given the minimal research on the subject of air pollution from aircraft cruising, and the corresponding lack of awareness of potential exposure risks among the general population, we argue that there is unlikely to be any sorting among residents based on overhead airline route intensity.

We next provide evidence for the strength of our instrument by empirically testing the first stage relationship between overhead airline route intensity and local air pollution using i) a panel model approach involving all airline routes, and ii) an event study design involving the launch of new flight routes. We close this section 4 by applying our instrument in an IV regression to estimate the impact of in-utero pollution exposure on infant health outcomes globally.

4.2 Panel model approach

We first use a panel model to assess the impact of airline routes on air pollution with the following specification:

$$Y_{ijt} = \beta AllRoutes_{it} + \gamma_j + \eta_t + g(Trend_{jt}) + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} represents the concentration of $PM_{2.5}$ reported from US EPA monitor i located in state j on day t. We employ a 0.1-degree grid ($\sim$10km^2) around monitor i to compute the count of airline routes. $AllRoutes_{it}$ is the airline intensity in grid cell i in year t. Moreover, we add state, year, month, and day-of-week fixed effects and quadratic time trends to account for state-specific time-invariant unobservables that may affect pollution, as well as temporal patterns and trends.

The coefficient β captures the impact of airline intensities on the surrounding PM2.5 levels. We use both the full sample and a subset of grids excluding those near airports. The identifying assumption is that overhead airline intensity is as good as randomly assigned, as described in the section above and confirmed in Table A18.

AllRoutes is calculated using two steps. First, we simulate airline routes between origin and destination airports using OAG data based on the shortest distance. Then we map these routes on a 0.1-degree map, calculate the number of routes in each pixel, and rescale the value between 0 and 1. We plot the distribution of *AllRoutes* for all 6.5 million 0.1-degree grid cells worldwide in Figure A7. In the left panel, the histogram exhibits two prominent spikes at 0 and 1. 30% of grid cells have 0 values, indicating areas with no overhead airline routes. The other large fraction of grids have 1 values, signifying grids with high airline intensity and suggesting a clustering of busy airline routes in specific regions. The right panel illustrates the distribution of airline intensity after removing these two spikes. Within the range of 0.1 to 0.9, a uniform distribution prevails, with a slight peak around 0.3.

Table 2 shows the results of regressing air pollution on overhead flight intensity using Equation 1. The first two columns focus on the US using *in-situ* PM2.5 data from EPA monitors. Column (1) includes year, month, day-of-week, and state fixed effects, and quadratic time trends, while Column (2) replaces state fixed effects with county fixed effects. Coefficients on *AllRoutes* are positive and significant. Using the more spatially-granular specification in Column (2), as *AllRoutes* increases by 1 unit (moving from the lowest to highest possible intensity), the surrounding PM2.5 along airline trajectories increases by $1.35\mu g/m^3$. This increase in PM2.5 is equivalent to 13.5% of the mean and 18.4% of the standard deviation.

We supplement US pollution monitor data with a global satellite-derived PM2.5 product. We add country-specific quadratic trends to account for diverse patterns of PM2.5 over time in each country, and include country fixed effects to address characteristics specific to each country that remain constant over time. In Table 2 Column (3), we find a posi-

tive and significant coefficient on *AllRoutes*. We observe a higher Y-mean, indicating that average PM2.5 levels globally are nearly twice as large as those in the US. Specifically, as *AllRoutes* increases by 1 unit, surrounding PM2.5 increases by $3.9\mu g/m^3$, constituting a 20% rise relative to the average PM2.5 levels.[15] In Column (4), coefficient on *AllRoutes* remains robust when Adm1 (the first administrative division within countries) fixed effects are added. The global results are larger in both absolute and relative terms than the US results, but of quantitatively similar magnitudes. Together, these results confirm our hypothesis that overhead airline intensity significantly raises surrounding air pollution levels.

As discussed in Section 4.1, locations near airports may be systematically different than other areas in terms of both population characteristics and pollution exposure. For the latter, airline intensity in pixels near airports is mechanically higher due to airplane clustering, and air pollution is also higher due to emissions from airplane taxiing and road vehicles. To address this potential concern that our observed effects may be driven by airports, we drop all pixels that are within 100km of an airport, as visualized in Figure A3.[16] We replicate our analysis using this sub-sample in Table A1. We find positive and significant estimates on *AllRoutes* with magnitudes that are modestly larger than those in Table 2.

In Figure A8, we vary the distance cutoffs from 0 to 1000km. Results remain strongly robust regardless of the distance choice. Effect magnitudes and precisions decrease gradually due to smaller sample sizes, as more pixels are dropped when using a larger cutoff to filter remote pixels. This suggests a substantial impact of aircraft emissions from cruising on PM2.5 levels along trajectories beyond the vicinity of airports.

Robustness and heterogeneity

We conduct several robustness checks. First, we use a discrete value to code *AllRoutes* by replacing all positive values with 1. Here we focus on the extensive margin. In other words, as long as the grid is covered by at least one airline, we ignore the number of flights. Results are presented in Table A2 Panel A. Coefficients on *AllRoutes* remain positive and significant in all columns, affirming the robustness of our findings. The magnitudes are consistently smaller compared to those in Table 2. This suggests that variation within the 0 and 1 range contributes to changes in PM2.5 to some extent. Neglecting these differ-

[15] The R^2 in Column (3) is 0.713, much larger than those in Columns (1) and (2). Within R^2 also increases from 0.0023 in Column (2) to 0.0143 in Column (3). This suggests that the predictive power of airline intensity and country-specific controls is more substantial in the global analysis than in the US analysis.

[16] This drops 38.4% of grid cells for the US analysis, and 36.8% for the global analysis.

ences and assigning a value of 1 to all positive instances leads to the underestimation of *AllRoutes*.

In a second robustness check, we exclude smaller values of *AllRoutes*. We categorize all positive values into four quartiles and replace the bottom quartile with zeros. This practice focuses on grids with more substantial airline coverage. As the formation of PM2.5 typically requires both time and emission accumulation, larger emissions are more likely to exert a pronounced impact on ambient PM2.5 levels. Results in Table A2 Panel B still demonstrate positive, significant, and big point estimates on *AllRoutes*. Magnitudes are also stable with this adjustment, which implies that the main results are predominantly driven by grids with higher airline intensities.

For the third robustness check, we take into account airline frequency. We duplicate airline lists based on their frequency in each year and then re-simulate trajectories. This adjustment considers instances where the same origin-destination pair undergoes multiple treatments and involves a greater number of flights in each year. The new airline intensity remains scaled between 0 and 1. Grids featuring the same airline but with multiple frequencies are naturally assigned higher values than those in the main result. Estimates in Table A2 Panel C show that in comparison to grids with zero airlines, those with higher frequencies experience higher PM2.5 levels by $0.82\mu g/m^3$ in the US and $3.6\mu g/m^3$ globally, equivalent to 8.2% and 18.8% of the average PM2.5 levels respectively. Similar magnitudes suggest that high frequency considering duplicated airlines is correlated with high intensity considering the airline list, i.e., one airline per route.

Fourth, we show in Table A3 that our results are robust to varying our treatment of standard errors. For results using the US EPA data, we cluster standard errors at the county level, year level, and state and year level in Panels A, B, and C, respectively. Results stay strongly robust compared with those in Table 2. For global results, we use grid, year, and country and year as clustering units in Column (3). Standard errors have minor changes when clustering at alternative geographic units and get larger when clustering at the year level. This suggests serial correlations in pollution across space within each year.

Satellite-derived measures of air pollution can introduce prediction error when studying the impact of air pollution (Fowlie et al., 2019). For robustness, we use NASA's MERRA-2 AOD reanalysis product as an alternative measure of air pollution to complement our primary gridded PM2.5 product from van Donkelaar et al. (2021), which uses satellite signals and global ground-based air quality monitors for calibration of results. MERRA-2 focuses on aerosol optical depth and uses a different set of satellite measures, on-the-

ground measurements, and modelling approach. In Table A8, panel regression results using MERRA-2 data are similar to those using the van Donkelaar et al. (2021) product, indicating the robustness of our results to using an alternative pollution data source.

We study the heterogeneity of the airline intensity-PM2.5 relationship across baseline PM2.5 and GDP levels. The intuition is that low-polluting areas with fewer other emission sources may provide a clean first stage relationship. In an extreme case, in isolated and undeveloped areas like the polar regions, the only pollution source is likely the airline trajectory across them. In contrast, we are more likely to encounter confounders if we run the analysis in high-polluting areas. In Table A9 Panel B, we separately estimate the first stage relationship using grids in four quartiles based on baseline PM2.5. Estimates on *AllRoutes* indicate that airline intensity moving from 0 to 1 increases surrounding PM2.5 by 16.3%, 13.7%, 2.1%, and 1.3% if we focus on pixels in the lowest to the highest quartile. The precision of estimates also gets worse as baseline PM2.5 levels increase. In a similar idea, Table A9 Panel B displays heterogeneity across GDP. High magnitudes and precision are found in both the high and low GDP groups. The nonmonotonic pattern may be due to the fact that poorer areas have less development and lower pollution levels, while richer areas tend to have cleaner air quality. These patterns suggest that our proposed instrument performs better in areas with fewer other sources of air pollution.

One other potential concern is the diversion of airline routes due to unfavorable weather, which can divert flight from the shortest distance path. We directly address this concern in the next subsection using actual aircraft locations from OpenSky rather than the simulated routes from OAG and find similar patterns between airline intensity and PM2.5. But in addition, here we drop pixel-months with high turbulence or control for turbulence for robustness checks. We use turbulent surface stress data from the ERA5 dataset. In Table A10 Column (1), we drop the top 10% pixel-months with the highest turbulence levels in the sample. Using the remaining sample, estimate on *AllRoutes* remains positive and significant, confirming the robustness of our results. The magnitude of the effect increases, likely because high turbulence often occurs simultaneously with high winds, leading to stronger pollution dispersion and lower ambient PM2.5 levels with the same level of aircraft emissions. In Column (2), we use the full sample and add turbulence as an additional control variable. We still find that airline intensity significantly increases surrounding PM2.5 concentrations along the trajectory.

We next test if altitude plays a role in the airline intensity-air pollution relationship. First, we test the heterogeneity across aircraft altitudes, which is also observable from the OpenSky data. In Table A15, we separately examine the relationship between PM2.5 and low-

and high-altitude intensity using 30,000 feet as the cutoff. Panel A uses van Donkelaar et al. (2021) data as outcomes and displays similar contributions of airline intensity to surrounding PM2.5 whether aircraft fly at high or low altitude. In Panel B, using EPA PM2.5 monitor reports, we find low-altitude intensity has larger effects on surrounding pollution than high-altitude intensity. This is consistent with the intuition that low-altitude emitters may affect surface pollution more, while van Donkelaar et al. (2021) data uses the amount of aerosol in the vertical column as inputs and is less sensitive to emitter altitude.

We also add altitude as an additional control and check how the first stage relationship changes. In Table A16, we use each pixel-week's average altitude across all covering aircraft as an additional control. We find negative and significant estimates on *Altitude*, suggesting low aircraft contribute more to surrounding air pollution. Estimates on $All\ aircraft$ remain positive and significant, and magnitudes are similar to those in Table A15 Column (1), confirming our results are robust with altitude controlled.

In addition, we control for road intensity due to the large contribution of on-road vehicles to air pollution. We add each pixel's road segment length to the right-hand side. Results in Table A17 show a positive *Road* coefficient, as expected, suggesting that vehicles increase ambient PM2.5. But our main relationship remains robust whereby airline intensity increases surrounding PM2.5 levels.

Alternate airline dataset

We test the validity of our instrument by replicating our analysis using a different airline dataset and airline intensity aggregation approach. As discussed in Section 4.1, there is a potential concern that our simulated routes, derived solely from origin and destination information from OAG, may not precisely reflect actual airline routes. To this end, we use overhead aircraft counts obtained from OpenSky sensor data. We aggregate the data at the pixel-week level for each 0.1-degree pixel ($\sim$10km^2), which is the same resolution as our main analysis using OAG data. The treatment variable is the total aircraft count. The econometric specification is detailed below:

$$Y_{it} = \beta_1 AllCount_{it} + \tau_t + \gamma_i + \varepsilon_{it} \tag{2}$$

where Y_{it} is the US EPA PM2.5 in monitor i on day t. $AllCount_{it}$ is the aircraft count in the 0.1-degree grid where monitor i is located in the week of day t. For the global analysis, Y_{it} is the van Donkelaar et al. (2021) PM2.5 in grid i in year t. $AllCount_{it}$ denotes the

airplane count in grid i, averaged over all weeks in year t.

In Table A4, we find positive and significant estimates on *AllCount*. Specifically, as the overhead aircraft count increases by 1,000,000 during the specified week, ambient PM2.5 exhibits a rise of $0.44\mu g/m^3$ within the US and $2.55\mu g/m^3$ on a global scale. These results affirm that our simulated routes effectively capture both the actual aircraft locations and their impacts on surrounding PM2.5 levels.

4.3 Event study approach

One may still be concerned that our approach captures an unobservable characteristic that both drives pollution exposure and happens to be correlated with a location's overhead flight intensity. To this end, we adopt an event study model to study the impact of introducing new airline routes globally. We define treated grid cells as those that are covered by at least one *new* airline route in a given year using OAG data. We then generate annual airline intensity maps at the grid-year level, which we denote as event time zero. Treated grid cells are assigned an intensity value of zero in years -3 to -1, and a positive value in years 0 to 5.

Figure A6 Panel A shows treated grids in the year 2000.[17] There are discontinuous grids where old and new airlines intersect. In other words, if two adjacent grids are both covered by a new airline and one was previously covered by an old airline, it is not categorized as a treated grid due to a non-zero value in the pre-period. There are few treated grids over most US land areas. This is likely due to pre-existing airlines that commenced operations prior to 2000. Panel B depicts treated grids in 2014. We see an increased number compared with 2000, with the majority situated in oceanic regions and spanning across continents.

After generating event grid-years, we construct a balanced sample at the grid-year level from three years before to five years after the airline opening, nine years in total. We use alternative event windows as robustness checks. The econometric specification is shown below:

$$Y_{it} = \sum_{k=-3}^{5} \beta^k D_{it}^k + \tau_t + \gamma_i + \varepsilon_{it} \tag{3}$$

[17] Treated grids are not present in the years 1997-1999, as these periods serve as baselines with zero airline intensity. During these three years, we don't know whether airline intensity is zero or unobserved in the years -3 to -1. Similarly, treated grids are not available in 2015-2019, as the operational status of airlines in the entire post-period after 2020 is unknown.

where Y_{it} is airline intensity in grid i in year t in the first stage. We use global PM2.5 from van Donkelaar et al. (2021) as the outcome variable in the reduced form analysis. D_{it}^{k} is a dummy that equals one if year t is k relative to the airline start year and zero otherwise. Year -1 serves as baseline period and D_{it}^{-1} is dropped. We add grid and year fixed effects to account for grid-specific time-invariant characteristics and global time differences.

Coefficients β^0 to β^5 capture the impact of airline opening on surrounding airline intensity and PM2.5 along the route. They are hypothesized to be positive. The identifying assumption is that the opening of an airline route is not correlated with other unobserved events along the route trajectory that might also influence air pollution. We test this hypothesis by estimating β^{-3} and β^{-2}, the impact in the pre-period. They are expected to be not statistically different from zero.

We conduct a first stage to assess the impact of new airline openings on overhead airline intensity. Figure 3 Panel A shows zero airline intensity in years -3 and -2, followed by a sharp increase in airline intensity in the post-period—a result expected by construction. The intensity value rises from 0 to 0.2 immediately after the opening of a new airline. The estimates are precise with small standard errors. The event study figure serves as a validation of our definition of new airlines.

For our main analysis, we create a similar event study figure using PM2.5 as the dependent variable. We add grid and year fixed effects, and plot residuals in Figure 3 Panel B. In the pre-period, the difference in PM2.5 relative to year -1 is not statistically different from zero, confirming the absence of a pre-trend. This suggests that the opening of an airline route is an exogenous event not correlated with other factors related to pollution in grid cells along the airline. In the post period, PM2.5 increases by $0.1\mu g/m^3$ in year 0, with a further increase to 0.3-0.4$\mu g/m^3$ in years 1 to 5. This implies that the launch of new airline routes leads to an increase in pollution along the flight path.

Tabular results are reported in Table 4 Panel A. In Column (1), a new airline opening mechanically leads to an increase in airline intensity by 0.19, 1.23 times the mean and 1.06 times the standard deviation. As van Donkelaar et al. (2021) PM2.5 is not available in all grids and has more missing data over the ocean, we rerun the first stage in Column (2) using pixels with non-missing PM2.5 data. The estimate on *Post* is almost the same as that using the full sample. In Column (3), the reduced form analysis shows a PM2.5 increase of $0.38\mu g/m^3$ following the new airline opening. The observed effect represents a grid-year level average in years 0 to 5, taking all days together. Since pollution emissions primarily take place during airline operations, the daily effects are more striking on airline operating

days or during busy air travel seasons.

As in Table A1 using the panel approach, we drop pixels within 100km to airports and rerun the estimation. Figure A3 shows locations of 100km buffers surrounding airports. Pixels within these red dots are dropped in this robustness check. Results in A6 Panel A remain strongly stable. The newly treated grids with new airlines observe an aviation intensity increase of 0.21 and a PM2.5 increase of $0.28\mu g/m^3$. In Figure A10, we present the event study figure by excluding grids within 100 km of airports. Patterns are similar to those in Figure 3. We see little evidence of pre-trends before the airline opening and a PM2.5 increase of $0.3\mu g/m^3$ in the post-period, although this gets noisy in year 3. Overall, this suggests that the observed pollution increase extends along the entire flight path *vis-a-vis* aircraft cruising rather than being driven by areas surrounding airports.

Alternate approach to defining treatment

As a robustness check, we use an alternative method to define treated grids based on the raw OAG data, annual airline information. We use lists of actual origin-destination-year and consider treated origin-destination pairs with zero airlines in years -3 to -1 and at least one airline in years 0 to 5. After treated origin-destination pairs are filtered, we simulate actual airlines for the treated group during the period spanning years 0 to 5, with covered grids being identified as treated grids. Figure A6 Panel C and D display treated grids based on the alternative definition. Because grids are chosen based on airline lists, we have continuous routes within the treated group. Similar to the first definition, we see substantial differences in treated grids in 2000 compared with those in 2014, reflecting a large number of new airlines opening in 2014.

Figure A9 replicates the event study Figure 3 using this alternative definition. Panel A focuses on response in terms of airline intensity. In the pre-period, residuals of airline intensity in years -3 and -2 reveal an absence of any discernible pre-trend, reaffirming the exogeneity of airline route openings. As expected, there is a distinct and substantial increase in intensity from year 0 to 5. Panel B shows the air pollution response. PM2.5 exhibits a significant increase of 0.5 in the post-period compared with the baseline year -1. Though magnitudes are similar, the standard errors are larger than those in Figure 3 Panel B. This imprecision may be attributed to the impact of pre-existing operating airlines within the grid that started their operations in the preceding years. In summary, both definitions of treated grids yield similar patterns in which there is a significant rise in air pollution fol-

lowing the introduction of new airlines.[18]

Another robustness test is to employ an alternative event window. Specifically, we consider the period from 4 years before to 6 years after the initiation of a new airline and re-choose treated grids based on simulated airline intensity. The alternative event window results in a smaller sample size, as additional years are required to ensure the absence of flight operations beforehand. In Figure A11, both panels depict similar event study figures with no discernible pre-trends. There is a substantial increase in airline intensity by 0.2 and PM2.5 by $0.3\mu g/m^3$. In Table A7, estimates on *Post* are positive and precise. The magnitudes in aviation intensity and PM2.5 increase exhibit a magnitude similar to that observed in the main specification.

Natural experiment

We complement our event study of new flight routes with a natural experiment of aircraft intensity reduction due to the flight ban in September 2001. Following the terrorist attacks on September 11, the Federal Aviation Administration (FAA) issued a nationwide ground stop, which led to the suspension of all commercial flights from September 11-13, 2001. We investigate the impact of this flight ban on airlines' surrounding PM2.5 levels in the US. The hypothesis is that the flight ban leads to a decrease in PM2.5, particularly in regions characterized by high overhead airline intensity. To test this hypothesis, we use the OAG data and a difference-in-difference design built on Equation 1. We add *Event* on the right-hand side, a dummy that equals one on these three days and zero otherwise. We also add *AllRoutes* and its interaction with *Event* as treatment variables.

In Table A5, we find similar estimates on *AllRoutes* with and without event variables, affirming the robustness of our identified relationship between airline routes and air pollution. Estimate on *Event* is negative and significant, which suggests a national decrease in PM2.5 on these three days regardless of locations within the US. Estimate on the interaction term $Event \times AllRoutes$ is negative, significant, and of substantial magnitude, indicating a more pronounced decrease in areas characterized by high overhead airline intensity. This flight ban natural experiment confirms our finding that airline intensity contributes to an increase in surrounding PM2.5 levels.

[18] For completeness, we also re-estimate Equation 3 using the alternative definition of treated grids and report results in Table 4 Panel B. We find positive and significant estimates on *Post* in all three columns. Table A6 Panel B confirms that when dropping grid cells within 100km from an airport, our results are unchanged when using this alternative treatment definition, suggesting that the observed pollution response is driven by aircraft cruising rather than airports.

Estimating the overall magnitude

How big is the air pollution impact from the launch of a new airline route? Over each year between 2000 and 2014, a total of 5,497 new airline routes (origin-destination pairs) were established, as shown in Figure A5.[19] On average, each airline covers 66,526 grids along the trajectory, with each 0.1-degree grid encompassing about 121km^2. Using event study results in Table 4, we observe an increase in PM2.5 levels by 0.38μg/m^3. The increase due to each new airline is equivalent to 0.9% of the average PM2.5 levels in areas located beneath the flight path. Our findings underscore the substantial externality associated with airline operations.

As a sanity check, we conduct a back-of-the-envelope calculation based on total fuel consumption, airline route length, and the emission intensity of fuel combustion to validate the increase in pollution within airline surrounding grids. Figure A12 presents the fuel burn data for ten sample airlines departing from London Heathrow Airport (LHR). In the left figure, it is evident that cruising accounts for 62-92% of the fuel consumption, surpassing the combined fuel usage during taxiing, claiming, and approaching phases. This underscores the significance of potential air pollution along airline trajectories. The right figure illustrates that a typical aircraft burns 2.7-13.9 liters of fuel per kilometer, depending on the aircraft model and seating capacity. That said, covering one 0.1-degree grid results in fuel consumption ranging from 30 to 153 liters. According to engineering estimates, air pollutant emissions from aviation fuel combustion amount to 1.2g/kg-fuel (Owen et al., 2022). Consequently, we anticipate a PM2.5 increase of 0.36-1.83μg/m^3. This range aligns with our point estimate detailed in Table 4.

4.4 IV: Health impacts of PM2.5

Our results demonstrate a robust link between overhead airline routes and local air pollution that extends beyond areas near airports. Our event studies show air pollution along trajectories is significantly affected by new the opening of new airline routes and the flight ban in September 2011. The pollution externality from airline cruising is meaningfully large.

In this section, we examine the health impacts of air pollution resulting from overhead airline activity. To do so, we use geocoded birth outcome data from the Demographic and Health Survey (DHS) of developing countries, as detailed in Section 3. Each birth outcome

[19] Figure A13 shows a similar time series for propeller-based airlines. There is a slight decreasing trend in airlines operating with leaded fuel, though it is still a substantial number.

is coded with airline intensity and PM2.5 levels in the DHS cluster locations. For treatment timing, we focus on pollution exposure in the in-utero period, average PM2.5 in the 9 months leading up to the birth.

In Table 3 Panel A Column (1), the reduced form analysis shows airline increases the occurrence of low birth weight. Specifically, as *AllRoutes* increases from 0 to 1, the probability of having a low birth weight increases by 0.018, relative to a mean of 0.022 (which means that 2.2% of sample births are categorized as having a low birth weight). This is consistent with existing findings that PM2.5 increases low birth weight occurrence (Gehrsitz, 2017; Jones, 2020; Alexander and Schwandt, 2022) in the context of low emission zones, dust storms, and diesel car emission scandals. In Column (2), the incidence of *very* low birth weight increases by 0.005 as *AllRoutes* increases from 0 to 1.

We also run an IV regression in Table 3 Panel B. PM2.5 instrumented by airline intensity causes adverse birth outcomes. F-statistics are over 10, suggesting airline routes are suitable instruments for PM2.5. As PM2.5 increases by $1\mu g/m^3$, or 2.3% over the sample mean, the incidence of low and very low birth weight increases by 0.01 and 0.003.

As a robustness check in light of the borderline F-statistic above, we use the MERRA-2 AOD product during the in-utero period as the endogenous regressor to replace van Donkelaar et al. (2021) PM2.5 product. Table A11 displays similar IV regression results using in-utero airline intensity as an instrument and low birth weight as outcomes. We find that first stage strengths improve and that the F-statistic increases from 11 to 17. We also observe positive and significant estimates on $MERRA-2\ AOD$, suggesting that in-utero particle pollution worsens birth outcomes. These findings indicate that PM2.5 emitted by airlines has a substantial externality on human health along the route.

Noise pollution

One confounding factor is noise pollution from aircraft. High airline intensity is correlated with high noise exposure. Residents along the flight path may be negatively affected by noise, which can also lead to worse birth outcomes independent of the air pollution channel. To address this concern, we collect aircraft model-specific noise information from the European Union Aviation Safety Agency (EASA) noise database. We classify aircraft models into noisy and non-noisy categories using 90dB (the 3rd quartile) as the cutoff.[20] We then separately simulate more noisy and less noisy airline intensity.

[20] We use 90dB to classify aircraft noise because noise regulations in construction, entertainment, and boating use a 90dB threshold.

We estimate the first-stage relationship between airline intensity and PM2.5 for more and less noisy routes. In Table A12 Panel A, effects on EPA monitor-reported PM2.5 are similar for more and less noisy routes. Point estimates are very similar when separately estimating these two or running a horse-race. In Panel B, results using data from van Donkelaar et al. (2021) show PM2.5 effects are slightly higher from less noisy aircraft, though the two estimated coefficients are not statistically different from each other.

In Table A13, we estimate reduced form effects of airline intensity on birth outcomes. Both noisy and less noisy routes increase the incidence of low birth weight. Effect sizes are larger for less noisy routes, suggesting that adverse health outcomes are unlikely to be driven solely by high-noise aircraft.

5 Lead and fertility

5.1 Model

In the next section of the paper, we investigate whether our airline instrument also works for airborne lead. Fuel combustion emits pollutants, and the effect size and chemical components of air pollution should differ across aircraft due to variations in fuel quality and combustion efficiency. More importantly, propeller planes still use leaded fuel due to engine design and engineering benefits[21] whereas jet planes use unleaded fuel derived from kerosene. Therefore, we apply our airline instrument separately to propeller-based routes and other routes as an instrument for airborne lead, and then study the relationship between lead exposure and fertility outcomes.

We first test whether overhead leaded airline intensity is as good as randomly assigned. Similar to Table A18, we check socioeconomic variables in DHS clusters but now focus on clusters covered by many leaded routes that are operated by propeller planes and clusters covered by many unleaded routes. In Table A34 Panel A, clusters with high leaded airline intensity operated by propeller planes witness lower wealth index, education length, and vaccination rates. In Panel B, after excluding clusters within 100km of airports, as we do in our earlier analysis, statistical differences between the two groups are not significant, suggesting a balanced distribution of covariates. This pattern supports the assumption that overhead leaded airline intensity difference is not correlated with different socioeco-

[21] Regarding engineering benefits, lead was added to fuel to improve octane rating. Lead also helps protect exhaust valves and their valve seats.

nomic variables after accounting for airport effects.

To investigate the impact of propeller airline routes on ambient air lead level, we conduct estimate the following:

$$Y_{ijt} = \alpha_1 LeadedRoutes_{it} + \alpha_2 AllRoutes_{it} + State_j + Time_t + g(Trend_{jt}) + \varepsilon_{ijt} \quad (4)$$

where Y_{ijt} is the air lead measurement obtained from US EPA monitor i located in state j on day t. $LeadedRoutes_{it}$ represents the count of propeller airline routes around monitor i during year t. We adopt a 0.1-degree buffer around the coordinates of monitor i to calculate the density of airline routes. For other controls, we add state fixed effects, capturing state-specific time-invariant variations; year, month, and day-of-week fixed effects, accommodating national temporal distinctions; and quadratic time terms, which assimilate temporal trends inherent in air lead levels.

The coefficient of interest, α_1, represents the effect of propeller airlines on the ambient air lead levels. α_2 is characterized as the "horse-race" term that increases with the intensities of both propeller and jet airlines. Since only propeller routes contribute to lead emissions, we hypothesize that α_1 is positive and α_2 approximates zero.

We employ the same empirical approach to examine the impact of propeller airline intensity on soil lead levels, children's blood lead levels, and fertility rates. For soil lead, coefficient α_1 encapsulates the first-stage impact of propeller routes on soil lead concentrations. For children's blood lead and female respondents' fertility rates, coefficient α_1 shows the reduced-form estimate on downstream outcomes.

5.2 Results

Estimation results of Equation 4 using the US EPA data are presented in Table 5. In Column (1), coefficient on *LeadedRoutes* is positive and statistically significant, indicating that higher propeller plane intensity contributes to higher ambient air lead. Specifically, as the propeller intensity increases from 0 to 1, the air lead concentration rises by $0.018 \mu g/m^3$, representing 13.2% of the mean and 2.7% of the standard deviation. In Column (2), we add both *LeadedRoutes* and *AllRoutes* as right-hand variables. Point estimate on *LeadedRoutes* remains stable and statistically significant, while the estimate on *AllRoutes* is insignificant.

In Column (3), the estimate on *UnleadedRoutes* is not statistically distinguishable from

zero. This finding aligns with the use of current deleaded fuel in jet engines and serves as a placebo test, indicating that unleaded routes have no discernible effect on surrounding lead levels.[22]

Given that ambient lead eventually falls to the ground through deposition, we assess whether propeller planes also contribute to increased soil lead levels. Table A20 presents our findings. Large estimate on *LeadedRoutes* suggests the number of leaded routes operated by propeller planes results in high soil lead levels. Specifically, for every increase of 1,000 in leaded route count, the soil lead collected by the surrounding laboratory rises by 473.9 wt%.[23] The consistency between air and soil results concludes that emissions from propeller planes have a substantial and statistically significant impact on environmental lead levels.

We next analyze health outcomes that are plausibly affected by environmental lead exposure. We conduct a reduced form analysis using DHS fertility data in 44 developing countries as outcome variables. Results are reported in Table 6. After controlling for country-specific quadratic time trends, we find female respondents are less likely to give birth in areas with a high intensity of propeller planes. In Column (1), an increase in leaded route intensity from 0 to 1 causes a 0.07 decrease in the probability of giving birth. In Column (2) and (3), no discernible effect on fertility is found from unleaded routes operated by jet planes. This finding is consistent with medical evidence and papers on lead exposure and fertility decline discussed in Section 2.

We also study the fertility response in the US using county-level birth rates from the NVSS as the outcome variable. Results presented in Table A21 exhibit similar patterns. Leaded airline routes are associated with decreased birth rates, while unleaded routes show no effect. Specifically, a 1-unit increase in propeller airline intensity leads to a decline in the birth rate by 2.2 percentage points, representing a 3.2% decrease relative to the mean birth rate.

We next run an IV regression using airline intensity, ambient lead levels, and birth rates in the US.[24] In Table A23 Column (1), the OLS regression shows a negative and imprecise es-

[22] Two connected airports are likely to operate planes of any type, so the intensity of leaded and unleaded airline routes is highly correlated. This may lead to positive point estimates for *UnleadedRoutes*. We verify this in Table A22. Panel A tests the correlation between unleaded airline route intensity and leaded route intensity using all pixels. In Panel B, we only use pixels from the sample in Table 5 with EPA lead monitors. Both correlations are positive and significant.

[23] The unit of soil lead pollution is wt%, weighted percent.

[24] We end up with a small sample size since not all counties have lead monitors or complete readings over time. We conduct monitor-day level analysis in our first stage and county-year level analysis in the reduced form, both having large sample sizes. Now our IV regression is at the county-year level and only

timate on *Lead*. In Column (2), after using propeller intensity in the in-utero period as an instrument, the point estimate remains negative, and significance level improves. As ambient lead increases by 1 ng/m^3, birth rate decreases by 0.167, equivalent to 0.19% relative to the average birth rate. Additionally, F-statistics are moderately large, confirming propeller intensity is a suitable instrument for ambient lead level. In Column (3), we use both propeller and all airline intensity as IVs, and point estimate on *Lead* is similar to that in Column (2). The F-statistic becomes smaller, which suggests all routes perform worse in predicting surrounding lead than propeller intensity alone.

While we cannot run a similar IV using DHS data outside the US due to the lack of lead monitors, we can compare our reduced form estimates of the effect of leaded airline intensity on fertility (Table 6 (global DHS) vs. Table A21 (US)). Although the fertility measures are different, the effect size relative to the mean is smaller in the US than in the DHS countries. This may be due to two reasons. First, higher nutrition levels contribute to the mitigation of lead poisoning effects on human health. In the presence of the same level of overhead airline activity of propeller planes, the enhanced nutritional status in the US may play a role in alleviating fertility decline. Second, residents in developed countries may have access to and implement better defensive measures, thereby shielding themselves more effectively from lead exposure.

We test whether leaded and unleaded routes contribute differently to PM2.5 levels along trajectories. The intuition is that propeller engines using different fuels may have different emission factors compared to jet engines. Leaded routes may not only have higher lead contamination but also higher general PM2.5 levels. In Table A14, we find that both leaded and unleaded routes contribute to significant increases in surrounding PM2.5 levels. The two estimates are not statistically different, suggesting that the two types of aircraft have similar emission intensity for general particle pollution. This gives us higher confidence that the reduced form estimates and effects on fertility reduction are driven by lead pollution, rather than higher PM2.5 pollution.

Heterogeneity

We explore the heterogeneity of fertility responses across different age groups. Results are shown in Table A24. Similar point estimates on *LeadedRoutes* suggest that the lead effect remains consistent regardless of the respondents' ages. For teenage pregnancies below 19 years old, an increase in propeller plane intensity causes a reduction in the likelihood of

includes those county-years with lead readings.

birth by 0.06. For older mothers above 36, the effect size is comparable, with a decrease in the likelihood of birth by 0.058. This suggests that the impact of lead exposure on fertility is widespread and does not exhibit significant variation across different age cohorts.

We also examine heterogeneity across nutrition levels, motivated by distinct responses observed in DHS countries and the US. To measure nutrition levels, we use female respondents' hemoglobin concentration and categorize them into four quartiles. The subgroup estimation results presented in Table A25 reveal that the lead effect remains consistent across these nutrition quartiles. This could be attributed to the similarity in nutrition and socioeconomic conditions among the respondents in the DHS data.

Robustness

We conduct several robustness checks. First, we employ unweighted regression in the DHS analysis, treating each female respondent with equal weight. The results are presented in Table A26, revealing patterns very similar to those in the main specification. Second, we conduct a regression without age controls, excluding both linear and quadratic terms from our model. The results are reported in Table A27, affirming the robustness of our findings across alternative age specifications.

Next, we drop DHS clusters and air lead monitors close to airports that host at least one propeller plane in our sample period. This speaks to the potential concern that our observed effects are solely driven by areas in close proximity to airports. We use 100km as a cutoff distance and only keep remote areas. Results in Table A28 and A29 show that both air lead and fertility effects remain stable with the subsample analysis. We use an alternative cutoff, 200km, to define remote areas. In Table A30 and A31, estimates on *LeadedRoutes* are quite similar to those in the main analysis.

Furthermore, we add road intensity as an additional control given the lead emission from leaded fuels on roads. Given the late phase-out of leaded gasoline in developing countries, on-road lead emissions could affect our analysis. In Table A32, air lead results using the US EPA data show the inclusion of road intensity variable does not affect the air lead findings. Estimates on *Road* are small and imprecise. This is consistent with the earlier removal of leaded gasoline in the US around the 1980s, and in our study period road emissions no longer include lead and would not affect ambient air lead. In Table A33, focusing on the DHS countries, road intensity leads to a decrease in fertility, captured by negative estimates on *Road*. Estimates on *LeadedRoutes* are similar with and without road intensity controlled, confirming our findings are not affected by on-road lead emissions.

Discussion on mechanism and magnitude

The mechanism of our findings is rooted in direct lead exposure. Human exposure to lead has known adverse effects on reproductive functions, leading to a natural decrease in the probability of giving birth. We use blood lead data to verify the exposure channel using detect rate data from Massachusetts. In Table A35, high numbers of overhead propeller planes lead to a higher detect rate of blood lead. As leaded routes move from 0 to 1, the detect rate rises by 2.0‰, equivalent to 33% of the average and 9.8% of the standard deviation. In contrast, unleaded routes operated by jet engines do not have significant effects on blood lead levels.

How does the link between lead exposure and fertility vary across countries? To answer this question, we estimate α_1 separately for each country in the DHS dataset and the US. Results are visualized in Figure 4. Negative estimates indicate that all but three countries experience persistent declines in fertility due to high propeller plane intensity. Our estimated link between propeller plane emissions, lead pollution, and fertility reduction is consistent with the medical literature, and is similar across the world. Given the robust patterns and extensive country coverage, our findings gave little concern about external validity.

In terms of the magnitude difference, the fertility decline due to a 1-unit increase in leaded airline emissions is more pronounced in East and Northwest Africa compared to the US and South Asia. Apart from the nutrition channel discussed in Table A25, another possible explanation is the nonlinear impact of lead exposure on fertility. East and Northwest Africa typically have higher background levels of lead from traffic and food sources (Harper et al., 2003; Ford and Stein, 2016). If lead exposure has a convex toxicity towards reproductive functions, the marginal increase in lead emission has a more severe effect on fertility reduction, captured by bigger magnitudes in African countries. One major source of lead emissions is the leaded gasoline in on-road vehicles. While the US implemented leaded gasoline bans in 1990, these bans were gradually phased out in Africa until 2021, leading to generally higher background lead levels in these developing countries. We further evaluate the impact of the leaded gasoline ban in Section 6.

6 Policy extensions: Leaded fuel ban

6.1 Model

In this section, we generalize our findings in relation to the global phase-out of leaded gasoline in road vehicles over the course of 35 years, which started with Japan in 1986 and ended with Algeria in 2021. To do so, we obtain the timing of leaded fuel bans at the country level and employ an event study design to examine the effects of fuel bans on fertility rates. The econometric specification is as follows:

$$Y_{ijt} = \beta Post_{jt} + f(Age_{it}) + Country_j + Time_t + g(Trend_{jt}) + \varepsilon_{ijt} \tag{5}$$

where Y_{ijt} represents the birth indicator for individual i in country j during year t. Our analysis focuses on female respondents of reproductive ages, 15 to 49, surveyed from three years before to three years after the enactment of the fuel ban. We construct a panel spanning seven years surrounding the year of the ban and code Y_{ijt} based on female respondents' decisions of giving birth in year t. On the right-hand side, variable $Post_{jt}$ is an indicator that equals one if the leaded fuel ban has been implemented in country j in year t and zero otherwise. Variable $f(Age_{it})$ incorporates quadratic terms of the age of female i in year t. To account for country-specific, time-invariant factors and global temporal disparities, we introduce country fixed effects and year fixed effects. Furthermore, we introduce country-specific quadratic trends in $g(Trend_{jt})$ that capture the distinctive fertility patterns of each country.

The coefficient of interest is β, which captures the effect of the leaded fuel ban on the fertility rate at the country-year level. β is the average treatment effect taking all 44 countries across the three-year post-ban periods together.

Given the potential different impacts of the ban on fertility rates across countries, we extend our investigation by assessing within-country variations. We compare DHS clusters across different levels of road intensity and employ a difference-in-difference framework, as outlined below:

$$\begin{aligned} Y_{ijkt} = \beta_1 Post_{jt} + \beta_2 Road_k + \beta_3 Post_{jt} \times Road_k + Country_j + Time_t \\ + f(Age_{it}) + g(Trend_{jt}) + \varepsilon_{ijt} \end{aligned} \tag{6}$$

where Y_{ijkt} is the same group of female respondents in DHS cluster k. $Road_k$ is the count

of road segments near cluster k. We draw a buffer of 0.1 degrees around the geographical coordinates of cluster k to calculate road count. Other controls are the same as those in Equation 5. β_3 captures the impact of the leaded gasoline ban between clusters with a higher density of road segments compared to those with a lower density.

6.2 Results

Results of estimating Equation 5 are reported in Table 7 Panel A. In Column (1), our sample includes both geocoded and non-geocoded respondents. The estimate on *Post* is positive and statistically significant. Consistent with intuition, there is a substantial increase in the likelihood of giving birth after the leaded fuel ban by 0.072 units, equivalent to a 24% increase relative to the average birth probability. In Column (2), we only use respondents with geocodes. The estimate on *Post* remains positive and precise, and the magnitude becomes smaller. There is a 0.045 unit increase in birth likelihood after the fuel ban, 16.4% relative to the mean. In Column (3) and (4), we add administrative 1 area fixed effects and implement an alternative clustering method. Both practices yield robust results of fertility increases after the ban.

We explore the within-country variations by estimating Equation 6. Results shown in Table 7 Panel B reveal positive and statistically significant estimations for *Post*. The effect magnitudes are very similar to those in Panel A, underscoring the robustness of our results at the country-year level after the leaded fuel ban. Focusing on the interaction term, estimates on $Post \times Road$ are positive and precise. This suggests that DHS clusters with high road intensities derive more substantial benefits from the leaded gasoline ban in comparison to those with lower road intensities. For each unit increase in road count, the likelihood of giving birth among female respondents rises by 0.023 units. The finding is consistent with the intuition that areas with higher road intensity experience greater exposure to leaded gasoline pollution before the ban's implementation. As a result, female respondents within these regions are more significantly affected by pollution and, consequently, stand to gain more from the removal of leaded gasoline.

We investigate the heterogeneity across different age groups of female respondents. In Table A36, the most striking effects are observed in the younger age group, while the impact of the leaded fuel ban on the older group exhibits a flipped sign. This indicates that the phase-out of leaded gasoline has a bigger impact on teenage pregnancies. The probability of teenage childbirth has risen significantly by 0.226 units, accounting for 88.3% of the average. This effect size is smaller for individuals aged 20-35, where the increase amounts to

0.032 units, equivalent to 11.4% of the average value.

Furthermore, we explore the heterogeneity across different nutrition levels. This practice is inspired by medical literature suggesting that lead exposure might have a lesser impact on individuals with better nutritional status. Consequently, the prohibition of leaded gasoline should potentially yield more pronounced benefits for those with lower nutrition levels. To gauge nutritional status, we employ a proxy indicator based on the DHS question regarding blood protein levels. The findings presented in Table A37 indicate that female respondents in the highest quartile of blood protein levels experience the least pronounced increase in the likelihood of childbirth. Specifically, this effect has a magnitude of 0.03 units, equivalent to 11.4% of the mean value. Conversely, the remaining three quartiles exhibit a comparable rise in the probability of childbirth, amounting to 0.05 units or 18.7% relative to the average likelihood.

Our main specification employs weighted regression using DHS sample weights. To assess the robustness of our findings, we use unweighted regression and find similar estimates in Table A38. Furthermore, another robustness check is to exclude age controls. Table A39 reveals consistent and stable results, suggesting leaded gasoline bans significantly increase fertility rates especially in areas with high road intensities.

Apart from the DHS, we also use fertility data from the US National Vital Statistics System (NVSS). The birth rate at the county-year level covers 1980 to 2021.[25] Results of estimating the single difference model are reported in Table A40 Panel A. Similar to the pattern using the DHS data, positive and significant estimates on *Post* suggest that there is a substantial increase in birth probabilities at the country-year level subsequent to the removal of lead content from vehicle gasoline. Coefficients remain robust with state or county fixed effects, national or state-specific time trends, and alternative clustering methods of standard errors.

The measured effect size displays an increase in birth probability by approximately 0.028 units or 4.2% compared to the average birth rate. This difference in magnitude may be attributed to two factors. First, the heterogeneity across nutrition levels in Table A37 suggests that female groups with better nutritional status exhibit smaller benefits from the prohibition of leaded gasoline. In contrast to the developing nations in South Asia and Africa that constitute our DHS sample, residents of the US have relatively better nutritional levels, consequently experiencing reduced adverse impacts from lead exposure. As a result, there is a relatively smaller impact of lead content removal on US births. Second,

[25] We conduct a robustness check using all NVSS data between 1968 to 2021. Table A41 shows strongly robust results as those using the shorter sample.

vehicle gasoline emissions might constitute a significant proportion of lead sources within developing countries. In contrast, the US witnesses other notable sources including industrial emissions, aviation, lead-based paints, and water pipelines. Therefore, the restriction on fuel content assumes a relatively more pronounced role in shaping lead exposure and its repercussions within developing countries.

In Table A40 in Panel B, we explore within-country variation by computing road intensities at the county level and adding the interaction term. Results displayed in Table 7 Panel B support the observed pattern that high road intensity leads to a more substantial increase in birth rates after lead removals. For each 0.1 unit rise in road count, equivalent to 1.09 times the average count and 0.854 times the standard deviation, the corresponding birth dummy experiences a statistically significant increase of 0.0036 units or 0.539% relative to the average. Our findings show that both economically advanced and developing countries reap positive outcomes from the eradication of leaded gasoline, leading to observable enhancements in fertility rates. The global phase-out of leaded fuel emerges as a beneficial intervention, effectively countering the prevailing trend of declining fertility rates.

6.3 Effects on population

What is the magnitude of the effect of the global phase-out of leaded gasoline? We obtain population data at the country-year level from the World Bank, World Development Indicators. We conduct a similar event study using Equation 5. Estimation results are presented in Table A42. The population increase following leaded gasoline bans averages 4.5% over a post period of 15 years. Based on this point estimate of percentage change and the actual population figures, the gasoline ban is estimated to result in an additional 2.2 million people per year globally from the year after the ban to 15 years later. This increase could be attributed to a surge in new births, as discussed in the fertility effect in Section 5, and a reduction in deaths since lead also affects adult health. The substantial magnitude of the population increase underscores the long-term impact of the gasoline ban on macro socioeconomic outcomes for future research.

We also compare the magnitude of fertility impacts resulting from the ban on leaded gasoline with those from the intensity of propeller airlines. The prohibition of leaded gasoline causes an increase in the likelihood of giving birth by 0.044, equivalent to 16.1% of the mean (0.274). In contrast, a 1-unit increase in propeller plane intensity leads to a change in the fertility dummy of 0.071 or 25.9% of the average (0.274). The corresponding in-

crease in ambient air lead due to propeller plane intensity is 0.018 out of 0.152μg/m^3, 11.8% relative to the mean. Assuming a linear relationship between lead exposure and fertility, the global phase-out of leaded gasoline is equivalent to an average reduction in air lead by 0.011μg/m^3, or 7.4% of current air lead levels. Furthermore, the fertility impact of the leaded gasoline ban is equivalent to the removal of half of the global propeller plane airlines. These findings affirm the effectiveness of the leaded gasoline ban in reducing environmental lead levels. We also underscore the importance of future policies to address lead emissions from the aviation sector.

6.4 Effects on teen pregnancy

We further assert that the lead ban can explain the increase in teen pregnancy. Teen pregnancy is important given its public health implications. Adolescents are often not fully physically or emotionally developed, so teen pregnancy could lead to increased maternal and infant health risks. Also, teen pregnancy affects young parents' educational attainment, employment, financial stability, and long-term economic struggles. From a societal perspective, teen pregnancy rates affect long-term demographic trends in a population. In the US, the recent temporal trends in teen pregnancy have been a public health puzzle: the temporary reversal of the half-century secular decline in US teen pregnancy rates, when rates increased by 20% in the late 1980s and early 1990s–the exact period in which the EPA effectively reduced the amount of lead allowed in gasoline to zero.

To examine the impact of leaded gasoline bans on teen pregnancy, we obtain data from the United Nations Population Division, World Population Prospects. It measures teen pregnancy rates for mothers 15-19 at the country-year level for 237 countries or areas, 1960-2021. We use these rates as the dependent variable and re-estimate the single difference model in Equation 5. Table A43 displays the estimation results. Following the leaded gasoline ban, teen pregnancy shows an increase of 2.6 births per 1,000 young women, which is 3.5% relative to the average rate. Compared with the findings in Table A42 and 7, the magnitude is slightly smaller, indicating that the high-age group is more susceptible to lead exposure and experiences greater benefits from on-road lead removal.

7 Discussion

This paper addresses a critical challenge in the study of the causal effects of air pollution by introducing a novel instrument derived from the global airline network. Our analysis

demonstrates that air pollution is persistently elevated beneath overhead flight routes, offering a unique source of exogenous variation in pollution exposure. By leveraging this instrument, we make several important contributions to the economics literature.

We uncover a previously overlooked negative externality of the aviation industry, air pollution from aircraft cruising. We examine two pollution constituents: particulate matter and lead. For the former, we provide robust evidence of the adverse impacts of PM2.5 on infant health, including in 44 developing countries where data are limited. This finding aligns with previous research on the harmful effects of air pollution on birth outcomes. For lead pollution, we leverage differential pollution exposure patterns based on aircraft engine type. We demonstrate a large fertility impact among the general population in areas beneath propeller plane routes via increased exposure to airborne lead.

We then generalize our findings regarding lead's impact on fertility to the context of historical lead gasoline bans in countries worldwide. Our analysis suggests that these bans, which aimed to reduce lead exposure from vehicles, had a substantial positive impact on fertility rates and, as a result, global population growth—making them among the most material public health interventions in recent memory.

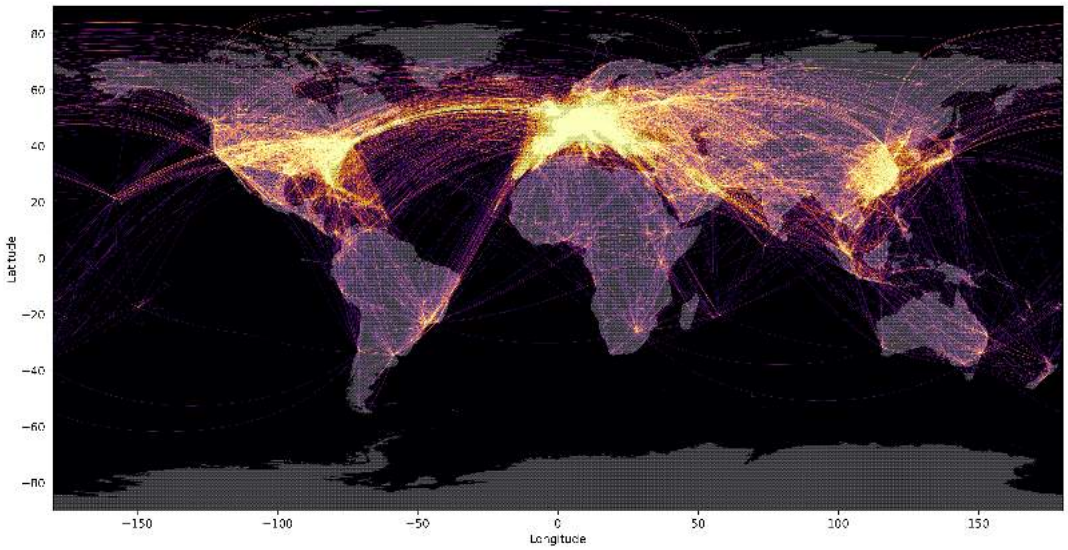
In addition to these specific contributions, we provide a global gridded airline data product that can be used in future research to address the non-randomness of pollution exposure. This data product offers researchers a tool to investigate various aspects of air pollution's impact on health and environmental outcomes, particularly in parts of the world where data are lacking and quasi-experimental settings are difficult to find.

Our findings suggest that even minimal exposure to PM2.5 and lead from airplane emissions can have significant health impacts, evidenced by lower birth weights and reduced fertility rates. Combined with the growing body of evidence on the high cost of air pollution, this paper underscores the need for policies to mitigate aviation-related pollution. This is particularly germane to the EPA's recent effort to address leaded emissions in aviation (link). Moreover, our findings contribute to the broader discussion on global fertility trends and the role of environmental factors in shaping population dynamics. Utilizing our airline instrument, we hope that further research will provide additional insights into the complex interplay between pollution, health, and demographics.

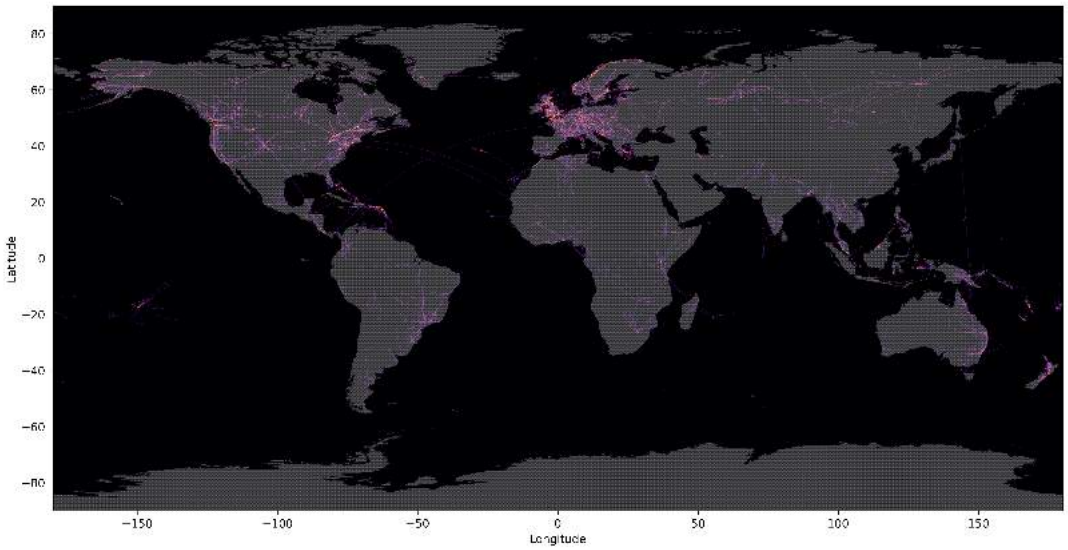
Figures and Tables

Figure 1: Visualization of airline routes

A: All flight routes

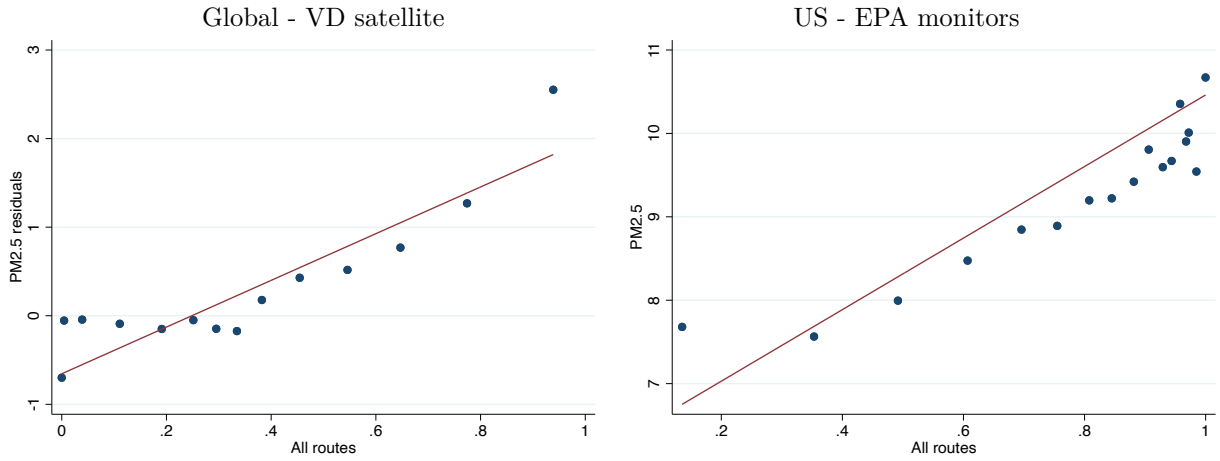


B: Propeller plane routes (leaded fuel)



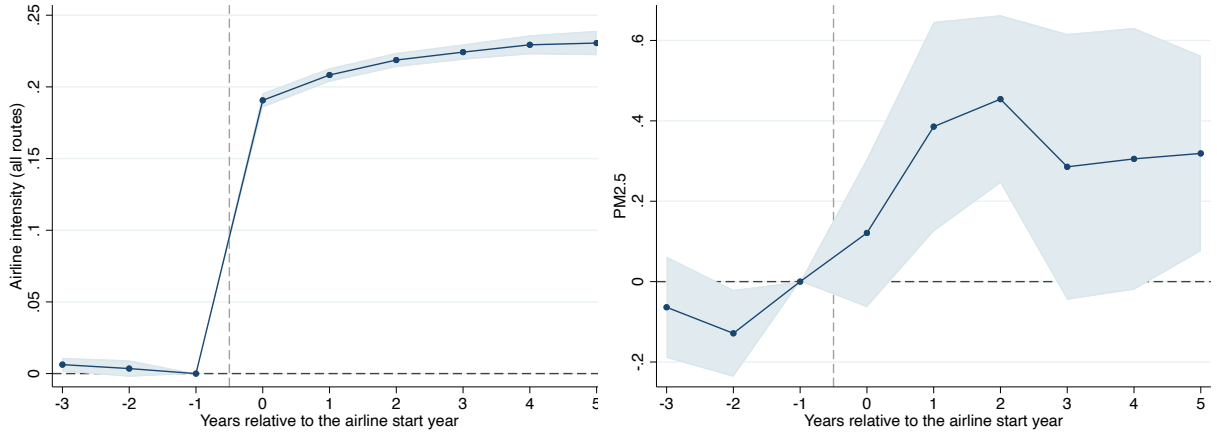
Notes: The top panel shows trajectories of 66,512 unique airline routes. The bottom panel shows trajectories of 7,899 leaded routes, mainly domestic airlines with shorter distances.

Figure 2: Binscatter plot of overhead flight intensity and PM2.5 levels



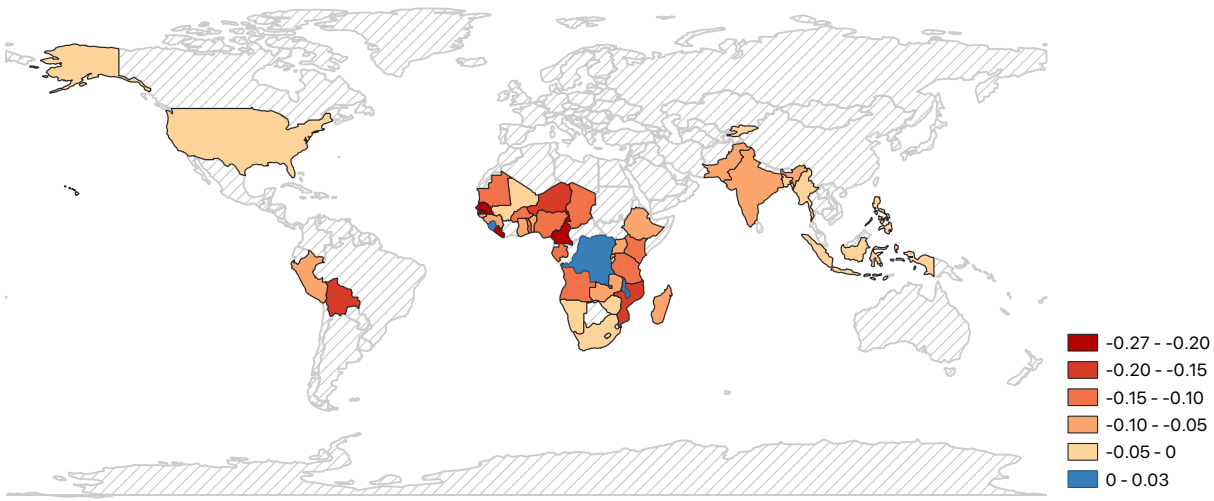
Notes: This figure shows average PM2.5 in each airline intensity bin. Global analysis (left) includes country FEs. Grid cells located with 100km of airports are dropped from analysis.

Figure 3: Event study, defined by simulated airline intensity



Notes: Panel A plots residuals of airline intensity. Panel B plots residuals of PM2.5 from van Donkelaar et al. (2021). Both panels include year and grid fixed effects.

Figure 4: Impacts of leaded airline intensity on fertility



Notes: This figure plots point estimates on α_1 in Equation 4 in the US and each individual country in the DHS sample. Gray areas are not covered by the DHS.

Table 1: Data summary

	Area	Time	Frequency
Treatment			
OAG airline route list (flight count per year)	Global	1997-2023	Yearly
OpenSky aircraft location (count per 0.1° grid)	Global	2016-2019	Weekly
Air pollution			
EPA monitor PM2.5 (μg/m^3)	US	1998-2023	Daily
van Donkelaar et al. (2021) PM2.5 (μg/m^3)	Global	1998-2021	Monthly
MERRA-2 aerosol optical depth (AOD)	Global	1997-2023	Monthly
Lead pollution			
EPA monitor lead (μg/m^3)	US	1997-2023	Daily
Soil lead (mg/kg)	US	2007-2010	Cross-sectional
Downstream outcomes			
Fertility in DHS countries (birth dummy)	44 countries	1992-2021	Microdata
Birth weight in DHS countries (grams)	44 countries	1992-2021	Microdata
Fertility in the US (birth rate per county)	US	1968-2023	Microdata
Blood lead (detect rate ‰)	Massachusetts	2012-2019	Yearly

Table 2: Airline intensity and air pollution

	PM2.5 (μg/m^3)			
	US		Global	
	(1)	(2)	(3)	(4)
All routes	1.94**	1.35***	3.90**	2.27***
	(0.83)	(0.52)	(1.56)	(0.78)
Observations	3,541,822	3,541,822	28,411,131	28,411,131
R-square	0.151	0.187	0.713	0.834
Y-mean	10.04	10.04	19.38	19.38
Y-sd	7.38	7.38	17.07	17.07
Year, Month	Y	Y	Y	Y
DOW FEs	Y	Y		
State FEs	Y			
County FEs		Y		
Country FEs			Y	
Adm1 FEs				Y
Trend	Quadratic	Quadratic	Country-specific Quadratic	Country-specific Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-day level, 1998-2019; Sample in Column (3) and (4) is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the state level in Column (1) and (2) and at the country level in Column (3) and (4). Significance: * 0.10, ** 0.05, *** 0.01.

Table 3: Airline intensity, air pollution, and birth outcomes

	Panel A: Reduced form	
	Low birth weight	Very low birth weight
	(1)	(2)
All routes	0.018**	0.005*
	(0.007)	(0.003)
Observations	194,712	194,712
R-square	0.065	0.019
Y-mean	0.022	0.004
Y-sd	0.147	0.059
	Panel B: IV regression	
PM2.5	0.010**	0.004**
	(0.004)	(0.002)
Observations	188,178	188,178
R-square	-0.255	-0.220
F-stat	11.81	11.81
Y-mean	0.023	0.004
Y-sd	0.149	0.060
X-mean	44.19	44.19
X-sd	23.89	23.89
Adm1 FEs	Y	Y
Country-specific trend	Quadratic	Quadratic

Notes: Each observation is a birth record in the DHS data. Low birth weight is a dummy that equals one if the raw birth weight is below 2500 grams and zero otherwise. Very low birth weight equals one if the birth weight is below 1500 grams. Sample covers year 1998-2019. PM2.5 levels are average pollution during in-utero periods, using data from the gridded van Donkelaar et al. (2021) product. Regression is weighted by survey sample weight, i.e., $v005$. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table 4: Event study, new airlines and air pollution

	Panel A: Defined by simulated airline intensity		
	Aviation intensity		PM2.5
	(1)	(2)	(3)
Post	0.191***	0.208***	0.380***
	(0.000)	(0.000)	(0.006)
Observations	4,437,387	1,816,233	1,816,233
R-square	0.598	0.587	0.973
Y-mean	0.155	0.166	19.232
Y-sd	0.181	0.193	17.501
	Panel B: Defined by airline list		
Post	0.019**	0.015**	0.391**
	(0.007)	(0.007)	(0.161)
Observations	5,733,918	3,378,727	3,306,453
R-square	0.888	0.876	0.679
Y-mean	0.547	0.615	20.860
Y-sd	0.356	0.342	17.113
Year FEs	Y	Y	Y
Grid FEs	Y	Y	Y

Notes: The analysis is at the grid-year level from 3 years before to 5 years after each grid's event year when new airline routes opened. The sample covers year 1998-2019. The outcome variable in Columns (1) and (2) is airline intensity, ranging between 0 and 1. The outcome variable in Column (3) is PM2.5 level from the gridded van Donkelaar et al. (2021) product. Column (1) includes all grid-years. Column (2) includes grid-years with non-missing PM2.5 data in the land area and no ocean area, the same sample as Column (3). Standard errors are clustered at the grid level. Significance: * 0.10, ** 0.05, *** 0.01.

Table 5: Aviation and air lead

	Lead (μg/m^3)		
	(1)	(2)	(3)
Leaded routes	0.018***	0.015***	
	(0.007)	(0.005)	
All routes		0.036	
		(0.038)	
Unleaded routes			0.032
			(0.031)
Observations	318,856	318,856	318,856
R-square	0.093	0.093	0.093
Y-mean	0.152	0.152	0.152
Y-sd	0.731	0.731	0.731
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
State FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the monitor-day level. Outcome variables are air lead levels from the EPA monitors, unit μg/m^3. The sample covers the years 1998-2019. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the state-year level. Significance: * 0.10, ** 0.05, *** 0.01.

Table 6: Aviation lead and fertility in DHS countries

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.071***	-0.068***	
	(0.009)	(0.008)	
All routes		-0.003	
		(0.009)	
Unleaded routes			-0.013
			(0.009)
Observations	7,096,998	7,096,998	7,096,998
R-square	0.013	0.013	0.012
Y-mean	0.274	0.274	0.274
Y-sd	0.446	0.446	0.446
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table 7: Fertility impacts of leaded gasoline bans

	Birth dummy			
	Panel A: Single difference			
	(1)	(2)	(3)	(4)
Post	0.072***	0.045***	0.044***	0.044***
	(0.006)	(0.009)	(0.009)	(0.006)
Observations	7,859,861	7,248,099	7,248,099	7,248,099
R-square	0.069	0.015	0.021	0.021
Y-mean	0.296	0.274	0.274	0.274
Y-sd	0.457	0.446	0.446	0.446
	Panel B: Double difference			
Post		0.044***	0.044***	0.044***
		(0.009)	(0.005)	(0.006)
Road		-0.021**	-0.018***	-0.018***
		(0.009)	(0.006)	(0.006)
Post × Road		0.023	0.023**	0.023*
		(0.015)	(0.011)	(0.012)
Observations		7,248,099	7,248,099	7,248,099
R-square		0.015	0.021	0.021
Y-mean		0.274	0.274	0.274
Y-sd		0.446	0.446	0.446
Country FEs	Y	Y		
Adm1 FEs			Y	Y
Age, Age^2	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from three years before to five years after the leaded fuel ban in each country. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1992-2021. Regression is weighted by survey sample weight, i.e., $v005$. Column (1) includes both geocoded and non-geocoded respondents, and Column (2)-(4) only includes geocoded respondents. Standard errors are clustered at the country level in Column (1)-(3) and at the adm1 level in Column (4). Significance: * 0.10, ** 0.05, *** 0.01.

References

Abouk, R. and S. Adams (2018). Birth outcomes in flint in the early stages of the water crisis. *Journal of Public Health Policy 39*, 68–85.

Ackerson, K. and R. Zielinski (2017). Factors influencing use of family planning in women living in crisis affected areas of sub-saharan africa: A review of the literature. *Midwifery 54*, 35–60.

Adsera, A. (2004). Changing fertility rates in developed countries. the impact of labor market institutions. *Journal of population economics 17*, 17–43.

Ahn, N. and P. Mira (2002). A note on the changing relationship between fertility and female employment rates in developed countries. *Journal of population Economics 15*(4), 667–682.

Aizer, A. and J. Currie (2019, 10). Lead and Juvenile Delinquency: New Evidence from Linked Birth, School, and Juvenile Detention Records. *The Review of Economics and Statistics 101*(4), 575–587.

Alexander, D. and H. Schwandt (2022). The impact of car pollution on infant and child health: Evidence from emissions cheating. *The Review of Economic Studies 89*(6), 2872–2910.

Aquino, N. B., M. B. Sevigny, J. Sabangan, and M. C. Louie (2012). The role of cadmium and nickel in estrogen receptor signaling and breast cancer: metalloestrogens or not? *Journal of Environmental Science and Health, Part C 30*(3), 189–224.

Atake, E.-H. and P. Gnakou Ali (2019). Women's empowerment and fertility preferences in high fertility countries in sub-saharan africa. *BMC women's health 19*(1), 1–14.

Austin, E., J. Xiang, T. R. Gould, J. H. Shirai, S. Yun, M. G. Yost, T. V. Larson, and E. Seto (2021). Distinct ultrafine particle profiles associated with aircraft and roadway traffic. *Environmental Science & Technology 55*(5), 2847–2858.

Axbard, S. and Z. Deng (2024). Informed enforcement: Lessons from pollution monitoring in china. *American Economic Journal: Applied Economics 16*(1), 213–252.

Barrett, S. R., R. E. Britter, and I. A. Waitz (2010). Global mortality attributable to aircraft cruise emissions. *Environmental science & technology 44*(19), 7736–7742.

Beaujouan, E. and C. Berghammer (2019). The gap between lifetime fertility intentions

and completed fertility in europe and the united states: A cohort approach. *Population Research and Policy Review 38*, 507–535.

Bongaarts, J. and A. K. Blanc (2015). Estimating the current mean age of mothers at the birth of their first child from household surveys. *Population health metrics 13*, 1–6.

Canipari, R., L. De Santis, and S. Cecconi (2020). Female fertility and environmental pollution. *International journal of environmental research and public health 17*(23), 8802.

Carré, J., N. Gatimel, J. Moreau, J. Parinaud, and R. Léandri (2017). Does air pollution play a role in infertility?: a systematic review. *Environmental Health 16*, 1–16.

Chen, S., P. Oliva, and P. Zhang (2022). The effect of air pollution on migration: Evidence from china. *Journal of Development Economics 156*, 102833.

Cidell, J. (2015). The role of major infrastructure in subregional economic development: an empirical study of airports and cities. *Journal of Economic geography 15*(6), 1125–1144.

Clay, K., M. Portnykh, and E. Severnini (2019). The legacy lead deposition in soils and its impact on cognitive function in preschool-aged children in the united states. *Economics & Human Biology 33*, 181–192.

Clay, K., M. Portnykh, and E. Severnini (2021). Toxic truth: Lead and fertility. *Journal of the Association of Environmental and Resource Economists 8*(5), 975–1012.

Clay, K., E. Severnini, and X. Wang (2022). Airborne lead pollution and infant mortality.

Clay, K., W. Troesken, and M. Haines (2014). Lead and mortality. *Review of Economics and Statistics 96*(3), 458–470.

Conforti, A., M. Mascia, G. Cioffi, C. De Angelis, G. Coppola, P. De Rosa, R. Pivonello, C. Alviggi, and G. De Placido (2018). Air pollution and female fertility: a systematic review of literature. *Reproductive Biology and Endocrinology 16*(1), 1–9.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Currie, J., J. Voorheis, and R. Walker (2023). What caused racial disparities in particulate exposure to fall? new evidence from the clean air act and satellite-based measures of air quality. *American Economic Review 113*(1), 71–97.

Currie, J. and R. Walker (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics 3*(1), 65–90.

Dave, D. M. and M. Yang (2022). Lead in drinking water and birth outcomes: A tale of two water treatment plants. *Journal of Health Economics 84*, 102644.

De Coster, S., N. Van Larebeke, et al. (2012). Endocrine-disrupting chemicals: associated disorders and mechanisms of action. *Journal of environmental and public health 2012*.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019, December). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Dissanayake, D., W. Keerthirathna, and L. D. C. Peiris (2019). Male infertility problem a contemporary review on present status and future perspective.

Eurostat (2021). Women in the eu are having their first child later.

Falcon, M., P. Vinas, and A. Luna (2003). Placental lead and outcome of pregnancy. *Toxicology 185*(1-2), 59–66.

Ferrie, J. P., K. Rolf, and W. Troesken (2012). Cognitive disparities, lead plumbing, and water chemistry: Prior exposure to water-borne lead and intelligence test scores among world war two u.s. army enlistees. *Economics Human Biology 10*(1), 98–111.

Ford, N. D. and A. D. Stein (2016). Risk factors affecting child cognitive development: a summary of nutrition, environment, and maternal–child interaction indicators for sub-saharan africa. *Journal of developmental origins of health and disease 7*(2), 197–217.

Fowlie, M., E. Rubin, and R. Walker (2019). Bringing satellite-based air quality estimates down to earth. In *AEA Papers and Proceedings*, Volume 109, pp. 283–288. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Gao, X., R. Song, and C. Timmins (2022). The fertility consequences of air pollution in china. Technical report, National Bureau of Economic Research.

Gazze, L., C. L. Persico, and S. Spirovska (2021). The long-run spillover effects of pollution: How exposure to lead affects everyone in the classroom.

Gehrsitz, M. (2017). The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management 83*, 121–144.

Goisis, A. (2023). Maternal age at first birth and parental support: Evidence from the uk millennium cohort study. *Population Research and Policy Review 42*(5), 75.

Greenstone, M. and R. Hanna (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review 104*(10), 3038–3072.

Gregoraszczuk, E. L., A. Ptak, et al. (2013). Endocrine-disrupting chemicals: some actions of pops on female reproduction. *International journal of endocrinology 2013*.

Grönqvist, H., J. P. Nilsson, and P.-O. Robling (2020). Understanding how low levels of early lead exposure affect children's life trajectories. *Journal of Political Economy 128*(9), 3376–3433.

Grossman, D. S. and D. J. Slusky (2019). The impact of the flint water crisis on fertility. *Demography 56*(6), 2005–2031.

Hafner, K. A. and D. Mayer-Foulkes (2013). Fertility, economic growth, and human development causal determinants of the developed lifestyle. *Journal of Macroeconomics 38*, 107–120.

Hansen, C., T. J. Luben, J. D. Sacks, A. Olshan, S. Jeffay, L. Strader, and S. D. Perreault (2010). The effect of ambient air pollution on sperm quality. *Environmental health perspectives 118*(2), 203–209.

Harper, C. C., A. Mathee, Y. von Schirnding, C. T. De Rosa, and H. Falk (2003). The health impact of environmental pollutants: a special focus on lead exposure in south africa. *International journal of hygiene and environmental health 206*(4-5), 315–322.

Heblich, S., A. Trew, and Y. Zylberberg (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy 129*(5), 1508–1552.

Hollingsworth, A., M. Huang, I. J. Rudik, and N. J. Sanders (2020). A thousand cuts: Cumulative lead exposure reduces academic achievement. Technical report, National Bureau of Economic Research.

Hollingsworth, A. and I. Rudik (2021). The effect of leaded gasoline on elderly mortality: Evidence from regulatory exemptions. *American Economic Journal: Economic Policy 13*(3), 345–373.

Istvan, M., R. Rahban, B. Dananche, A. Senn, E. Stettler, L. Multigner, S. Nef, and R. Garlantézec (2021). Maternal occupational exposure to endocrine-disrupting chemicals during pregnancy and semen parameters in adulthood: results of a nationwide cross-sectional study among swiss conscripts. *Human Reproduction 36*(7), 1948–1958.

Jones, B. A. (2020). After the dust settles: The infant health impacts of dust storms. *Journal of the Association of Environmental and Resource Economists 7*(6), 1005–1032.

Jurewicz, J., E. Dziewirska, M. Radwan, and W. Hanke (2018). Air pollution from natural and anthropic sources and male fertility. *Reproductive Biology and Endocrinology 16*, 1–18.

Karwacka, A., D. Zamkowska, M. Radwan, and J. Jurewicz (2019). Exposure to modern, widespread environmental endocrine disrupting chemicals and their effect on the reproductive potential of women: an overview of current epidemiological evidence. *Human Fertility 22*(1), 2–25.

Kato, T., T. Yorifuji, M. Yamakawa, S. Inoue, H. Doi, A. Eboshida, and I. Kawachi (2017). Association of maternal age with child health: A japanese longitudinal study. *PLoS One 12*(2), e0172544.

Klöwer, M., M. Allen, D. Lee, S. Proud, L. Gallagher, and A. Skowron (2021). Quantifying aviation's contribution to global warming. *Environmental Research Letters 16*(10), 104027.

Kumar, N. and A. K. Singh (2022). Impact of environmental factors on human semen quality and male fertility: a narrative review. *Environmental Sciences Europe 34*, 1–13.

Llanos, M. N. and A. M. Ronco (2009). Fetal growth restriction is related to placental levels of cadmium, lead and arsenic but not with antioxidant activities. *Reproductive toxicology 27*(1), 88–92.

Marcus, M. (2023). Burying the lead: Effects of public lead service line replacements on blood lead levels and property values.

Martin, J. A., B. E. Hamilton, M. J. Osterman, and A. K. Driscoll (2019). Births: final data for 2018.

Mathews, T. J. and B. E. Hamilton (2014). *First births to older women continue to rise*. Number 2014. US Department of Health and Human Services, Centers for Disease Control and ....

Negash, W. D. and D. B. Asmamaw (2022). Time to first birth and its predictors among reproductive age women in high fertility countries in sub-saharan africa: Inverse weibull gamma shared frailty model. *BMC Pregnancy and Childbirth 22*(1), 1–10.

Nelson, J. P. (1979). Airport noise, location rent, and the market for residential amenities. *Journal of environmental economics and management 6*(4), 320–331.

OAG (2022). Which part of a flight uses the most fuel?

OECD (2019). *Rejuvenating Korea: Policies for a Changing Society*. Organisation for Economic Co-operation and Development.

Owen, B., J. G. Anet, N. Bertier, S. Christie, M. Cremaschi, S. Dellaert, J. Edebeli, U. Janicke, J. Kuenen, L. Lim, et al. (2022). Particulate matter emissions from aircraft. *Atmosphere 13*(8), 1230.

Reyes, J. W. (2007). Environmental policy as social policy? the impact of childhood lead exposure on crime. *The B.E. Journal of Economic Analysis Policy 7*(1).

Reyes, J. W. (2015, 03). Lead Policy and Academic Performance: Insights from Massachusetts. *Harvard Educational Review 85*(1), 75–107.

Rzymski, P., K. Tomczyk, P. Rzymski, B. Poniedzialek, T. Opala, and M. Wilczak (2015). Impact of heavy metals on the female reproductive system. *Annals of agricultural and environmental medicine 22*(2).

Sauve-Syed, J. (2023). Lead exposure and student outcomes: A study of flint schools. *Health Economics*.

Schlenker, W. and W. R. Walker (2015, 10). Airports, Air Pollution, and Contemporaneous Health. *The Review of Economic Studies 83*(2), 768–809.

Selvaraju, V., S. Baskaran, A. Agarwal, and R. Henkel (2021). Environmental contaminants and male infertility: Effects and mechanisms. *Andrologia 53*(1), e13646.

Shapiro, J. S. (2022). Pollution trends and us environmental policy: Lessons from the past half century. *Review of Environmental Economics and Policy 16*(1), 42–61.

Singh, L., M. Anand, S. Singh, and A. Taneja (2020). Environmental toxic metals in placenta and their effects on preterm delivery-current opinion. *Drug and Chemical Toxicology 43*(5), 531–538.

Sobotta, R. R., H. E. Campbell, and B. J. Owens (2007). Aviation noise and environmental justice: The barrio barrier. *Journal of Regional Science 47*(1), 125–154.

Tanaka, S., K. Teshima, and E. Verhoogen (2022). North-south displacement effects of environmental regulation: The case of battery recycling. *American Economic Review: Insights 4*(3), 271–288.

Tonne, C., C. Milà, D. Fecht, M. Alvarez, J. Gulliver, J. Smith, S. Beevers, H. R. Anderson, and F. Kelly (2018). Socioeconomic and ethnic inequalities in exposure to air and noise pollution in london. *Environment international 115*, 170–179.

Troesken, W. (2008). Lead water pipes and infant mortality at the turn of the twentieth century. *Journal of Human Resources 43*(3), 553–575.

Upadhyay, U. D., J. D. Gipson, M. Withers, S. Lewis, E. J. Ciaraldi, A. Fraser, M. J. Huchko, and N. Prata (2014). Women's empowerment and fertility: a review of the literature. *Social science & medicine 115*, 111–120.

van Donkelaar, A., M. S. Hammer, L. Bindle, M. Brauer, J. R. Brook, M. J. Garay, N. C. Hsu, O. V. Kalashnikova, R. A. Kahn, C. Lee, R. C. Levy, A. Lyapustin, A. M. Sayer, and R. V. Martin (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology 55*(22), 15287–15300. PMID: 34724610.

Wang, R., X. Chen, and X. Li (2021). Something in the pipe: the flint water crisis and health at birth. *Journal of Population Economics*, 1–27.

Wdowiak, A., E. Wdowiak, A. Bień, I. Bojar, G. Iwanowicz-Palus, and D. Raczkiewicz (2019). Air pollution and semen parameters in men seeking fertility treatment for the first time. *International journal of occupational medicine and environmental health 32*(3).

Xue, T. and Q. Zhang (2018). Associating ambient exposure to fine particles and human fertility rates in china. *Environmental Pollution 235*, 497–504.

Zahran, S., T. Iverson, S. P. McElmurry, and S. Weiler (2017). The effect of leaded aviation gasoline on blood lead in children. *Journal of the Association of Environmental and Resource Economists 4*(2), 575–610.

Zahran, S., C. Keyes, and B. Lanphear (2023). Leaded aviation gasoline exposure risk and child blood lead levels. *PNAS nexus 2*(1), pgac285.

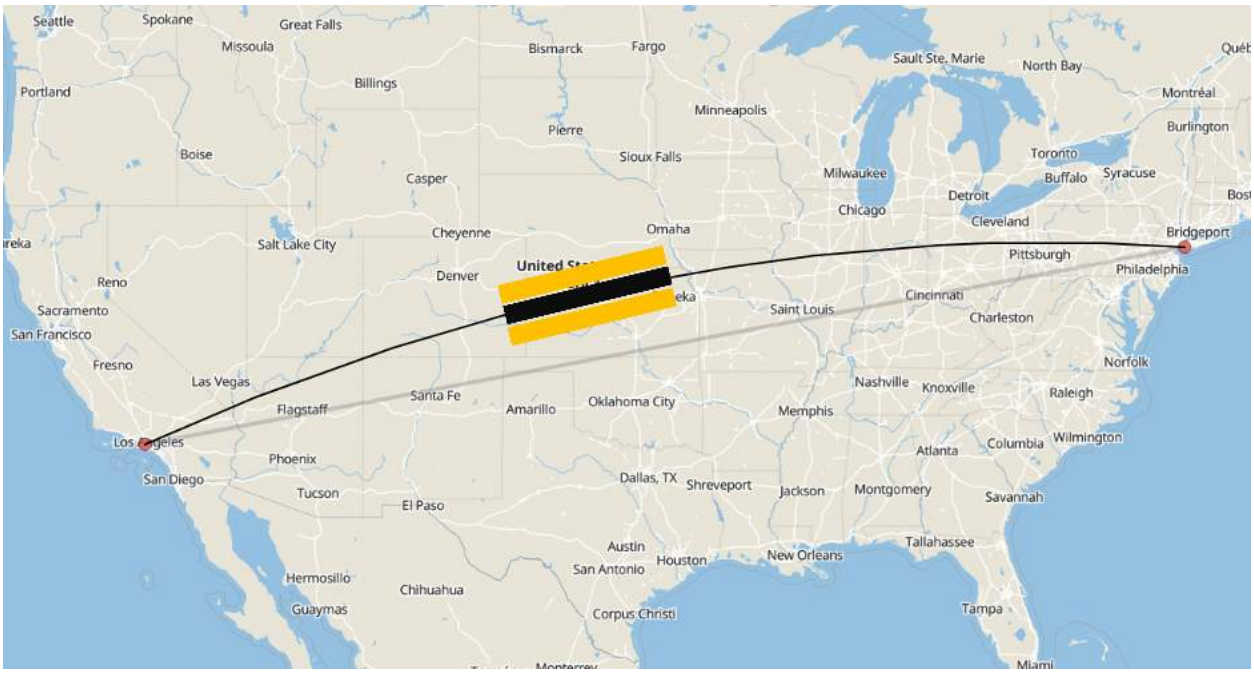
Zhang, J., Z. Cai, C. Ma, J. Xiong, H. Li, et al. (2020). Impacts of outdoor air pollution on human semen quality: a meta-analysis and systematic review. *BioMed research international 2020*.

Zou, E. Y. (2021). Unwatched pollution: The effect of intermittent monitoring on air quality. *American Economic Review 111*(7), 2101–2126.

Online Appendix

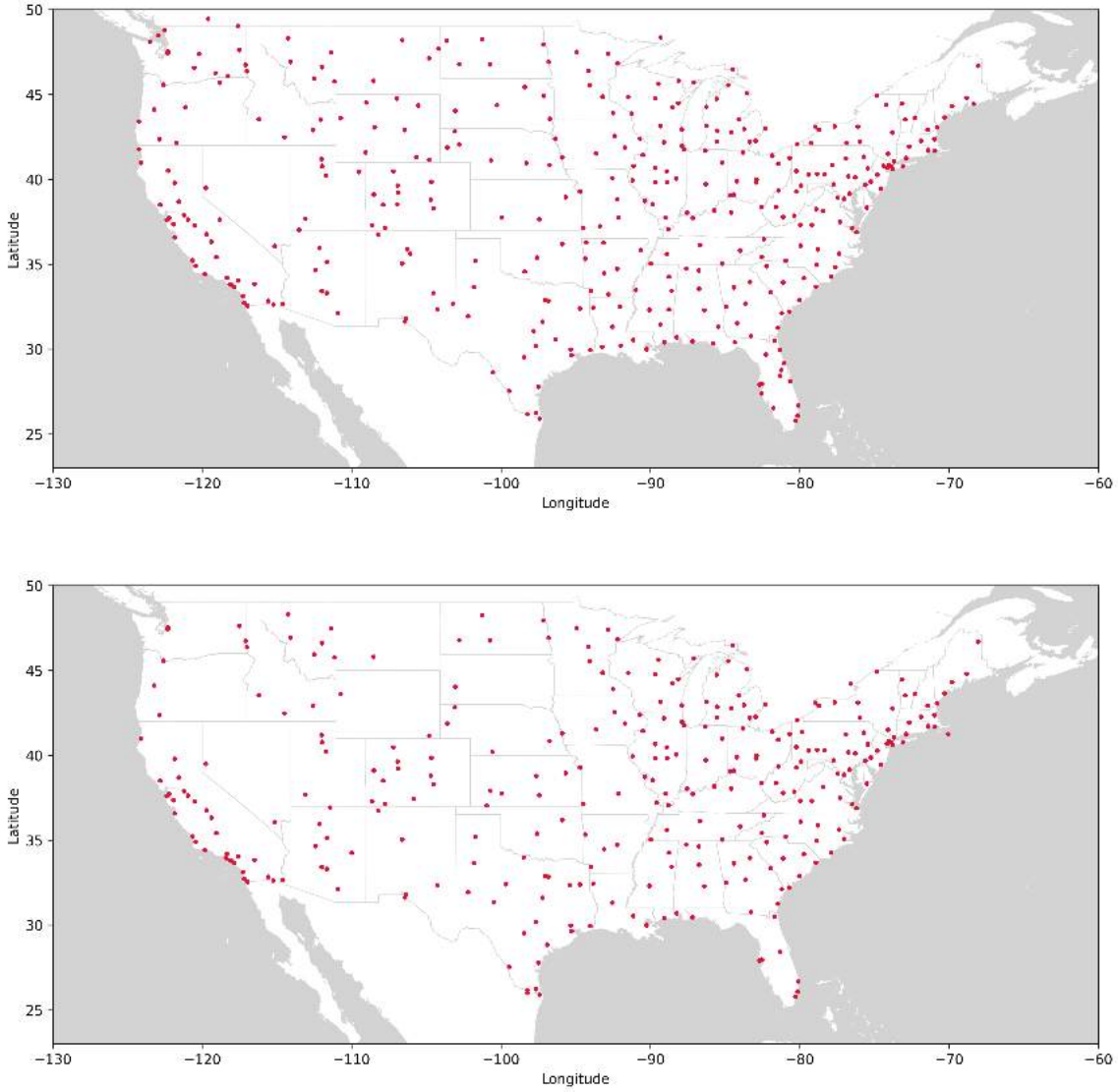
A1 Additional Figures

Figure A1: Illustration of research design using JFK to LAX flight path over Kansas



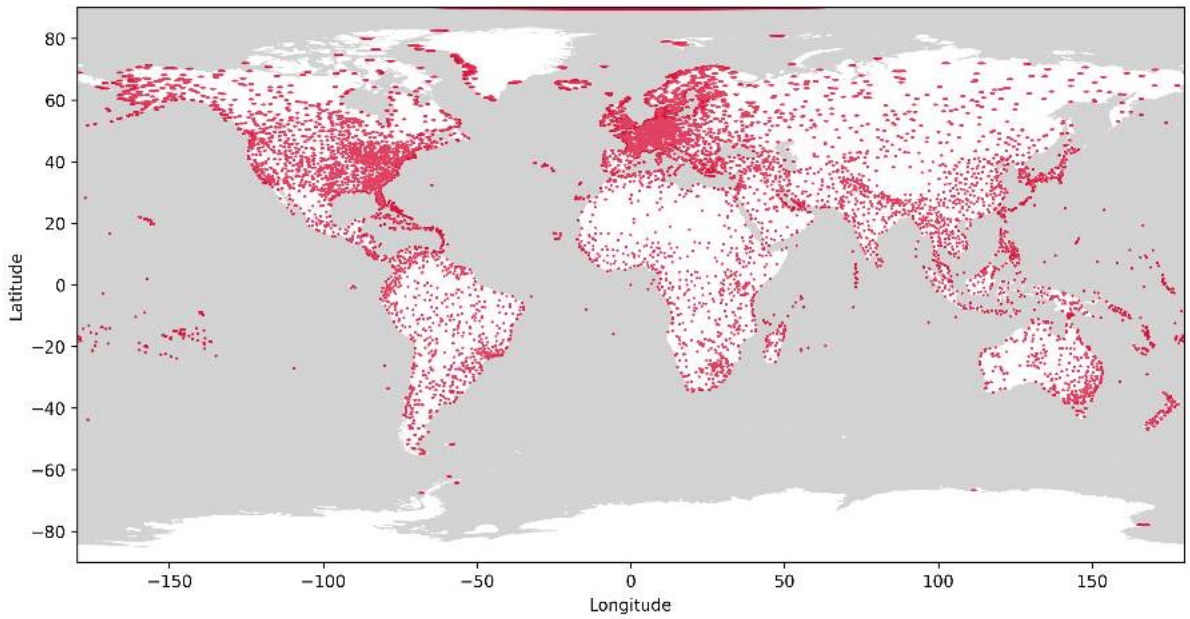
Notes: Empirical approach assumes that characteristics do not systematically differ between higher and lower flight intensity areas, as illustrated by the black and orange bands, respectively. Map produced using https://www.distance.to/LAX/JFK.

Figure A2: Locations of EPA PM2.5 (top) and air lead monitors (bottom)



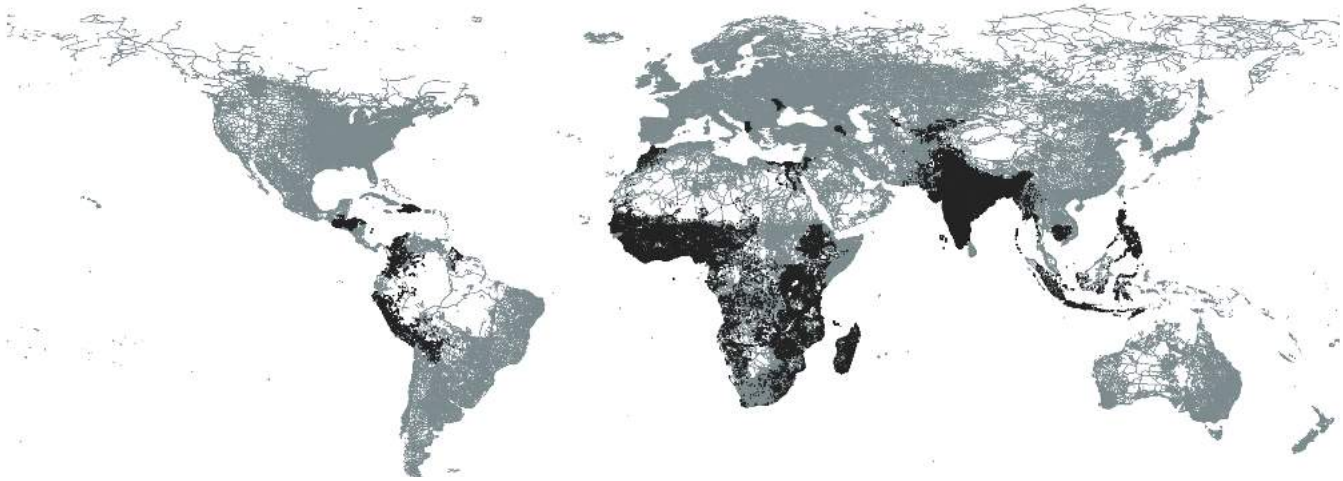
Notes: Locations of 2,344 PM2.5 monitors and 2,467 air lead monitors are shown in red dots. Each monitor has at least one report 1997-2023.

Figure A3: Buffers around airports



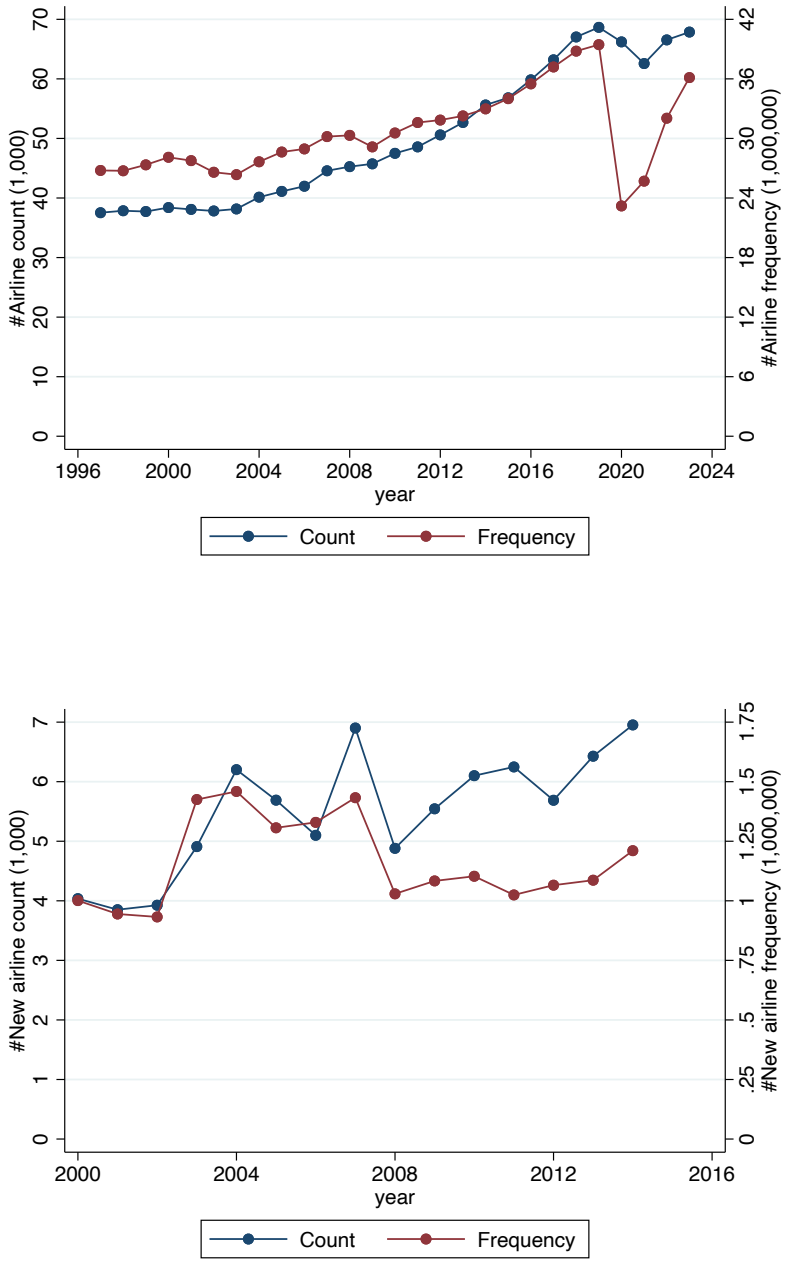
Notes: Locations near 7,698 airports within 100km are shown in red dots. These airports hosted at least one airline in 1997-2023.

Figure A4: Locations of DHS clusters and road segments



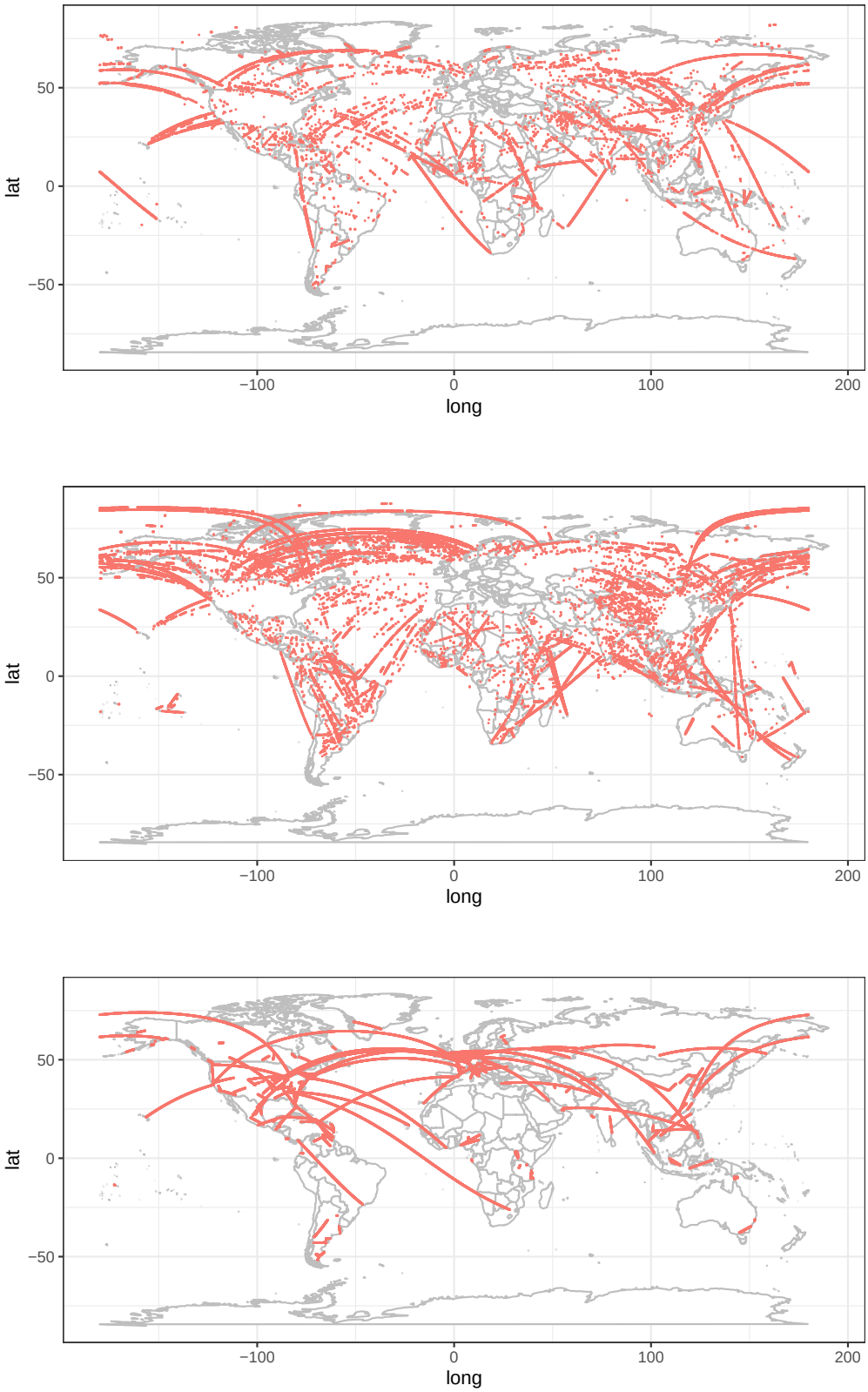
Notes: Locations of DHS clusters are shown in block dots. There are 142,842 clusters in 44 countries. Locations of road segments are from the Global Roads Open Access Dataset (gRoads) from the Socioeconomic Data and Applications Center, Columbia University, and plotted in blue lines.

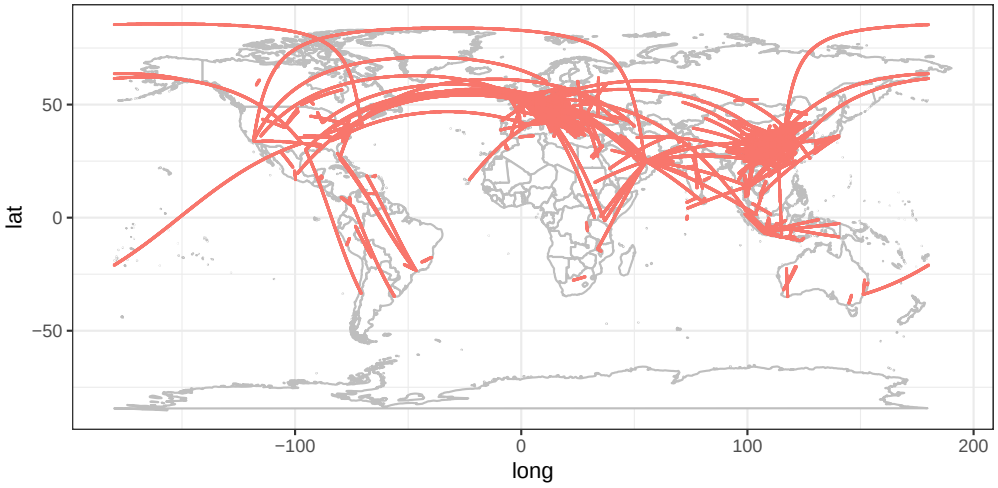
Figure A5: Airline count and frequency over time



Notes: We count non-duplicated airlines in blue and aggregate airline frequency in red. The top figure displays level changes over the period 1997-2023. The bottom figure specifically enumerates new airlines that were not in operation the previous year.

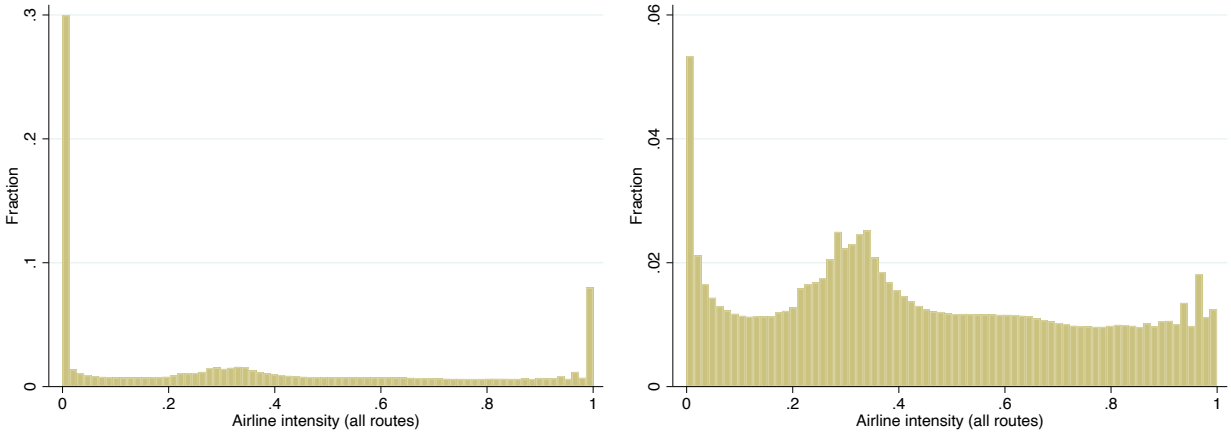
Figure A6: Treated grids with new airlines





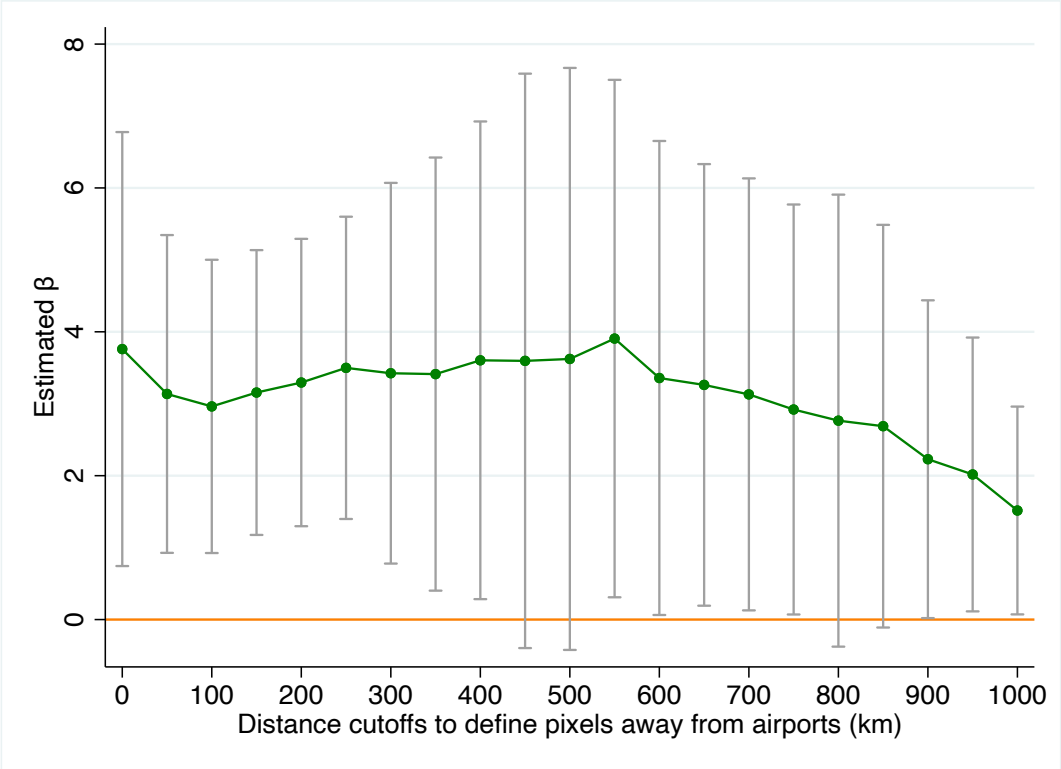
Notes: Panel A: Treated grids in 2000, defined by simulated airline intensity; Panel B: Treated grids in 2014, defined by simulated airline intensity; Panel C: Treated grids in 2000, defined by airline list; Panel D: Treated grids in 2014, defined by airline list.

Figure A7: Distribution of airline intensity



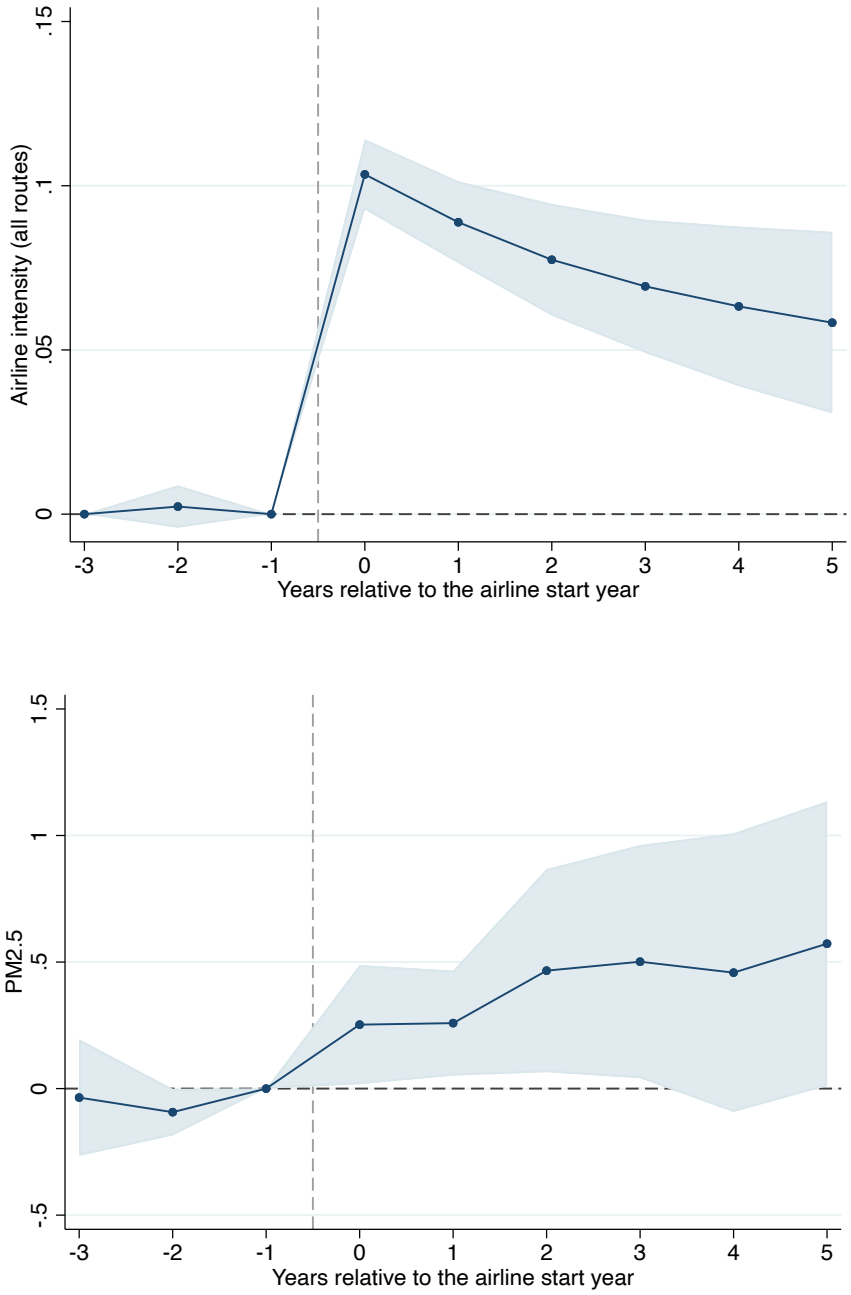
Notes: Airline intensity is a value from 0 to 1. The left figure is histogram of airline intensity using all grids. The right figure drops grids with 0 or 1 airline intensity.

Figure A8: Effects by distance cutoffs



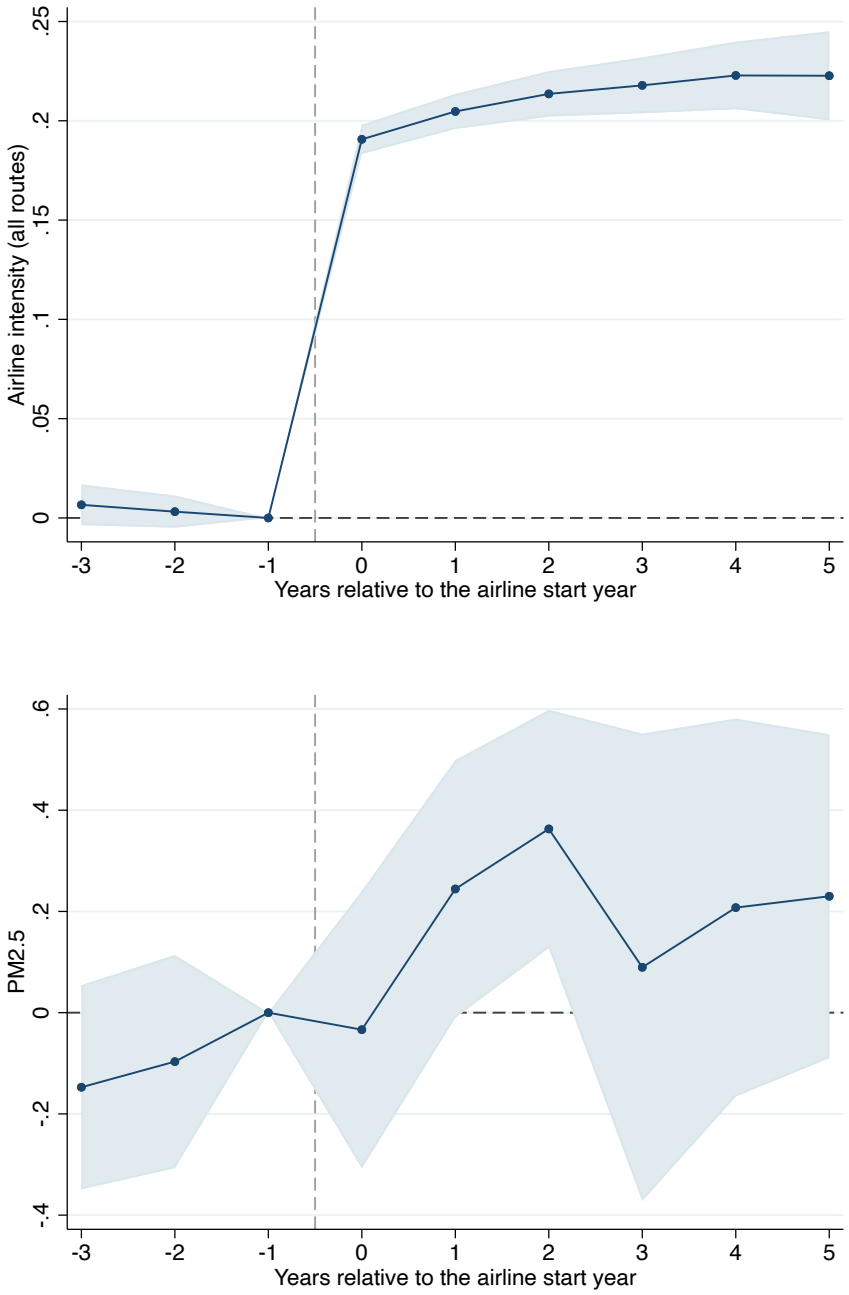
Notes: This figure shows estimated effect size of $\hat{\beta}$ in Equation 1 using different samples. We use distance cutoff to define remote areas far from airports: 0km means all pixels are kept, same as Table 2; 100km means only pixels far from airports by 100km or farther are kept, same as Table A1. Each green dot is one regression point estimate, and each gray bar is one confidence interval.

Figure A9: Event study, defined by airline list



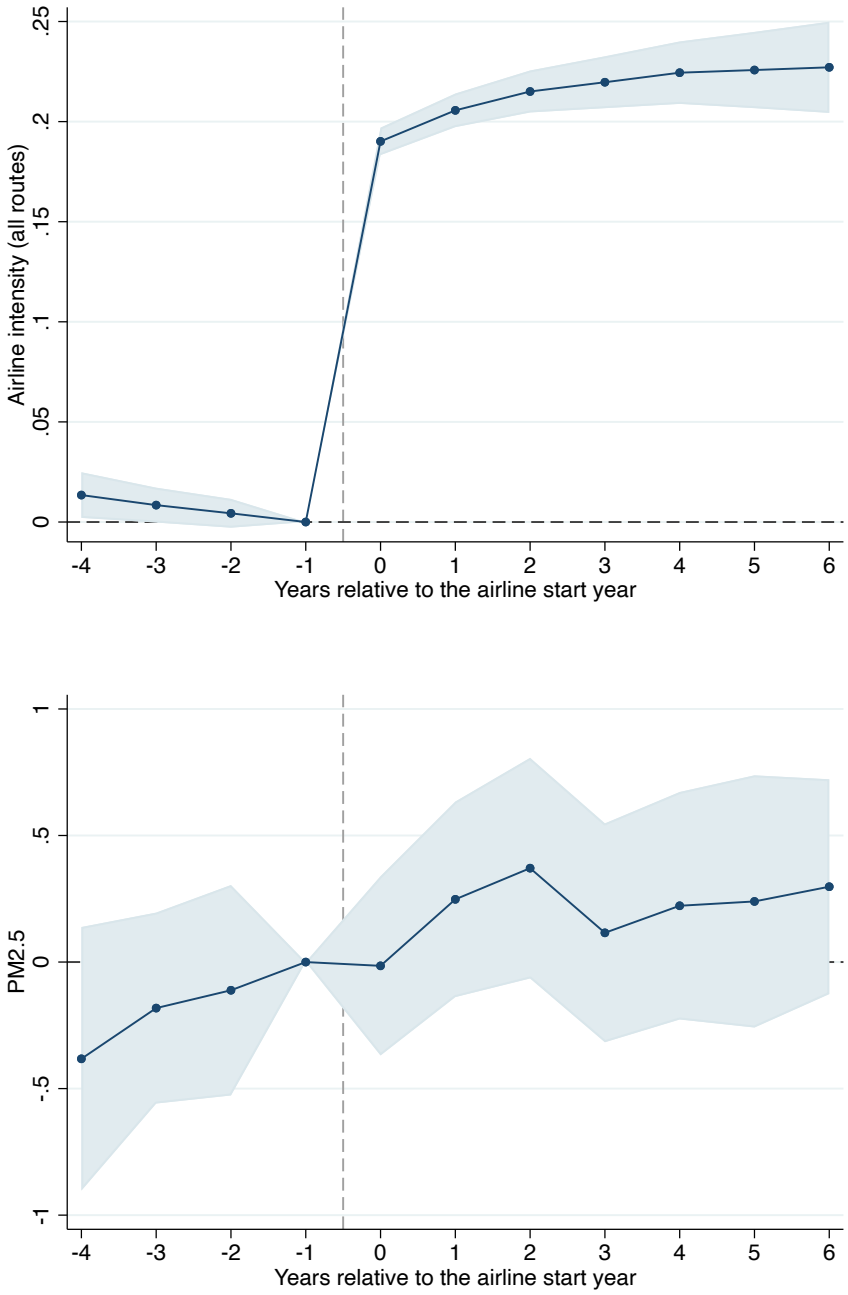
Notes: Panel A plots residuals of airline intensity. Panel B plots residuals of PM2.5 from van Donkelaar et al. (2021). Both panels include year and grid fixed effects.

Figure A10: Event study, dropping grids near airports



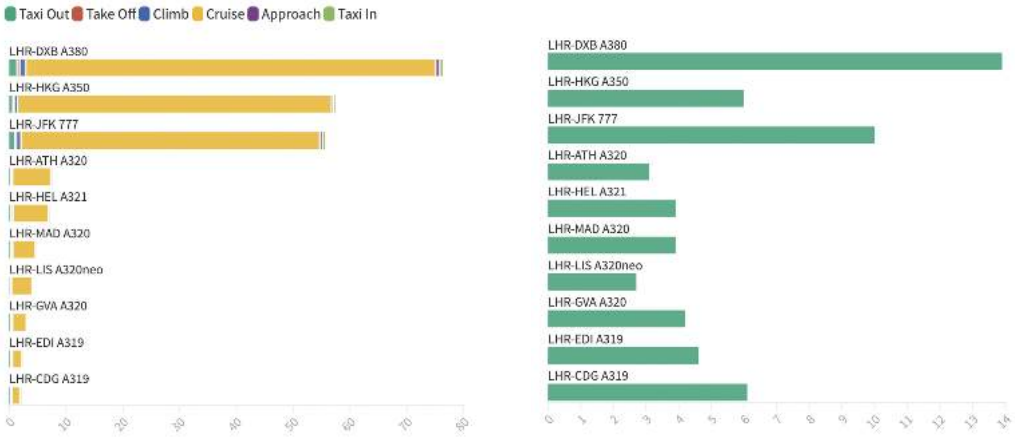
Notes: We drop pixels within 100km of any airports before estimating the residuals. Panel A plots residuals of airline intensity. Panel B plots residuals of PM2.5 from van Donkelaar et al. (2021). Both panels include year and grid fixed effects.

Figure A11: Event study, using alternative event window



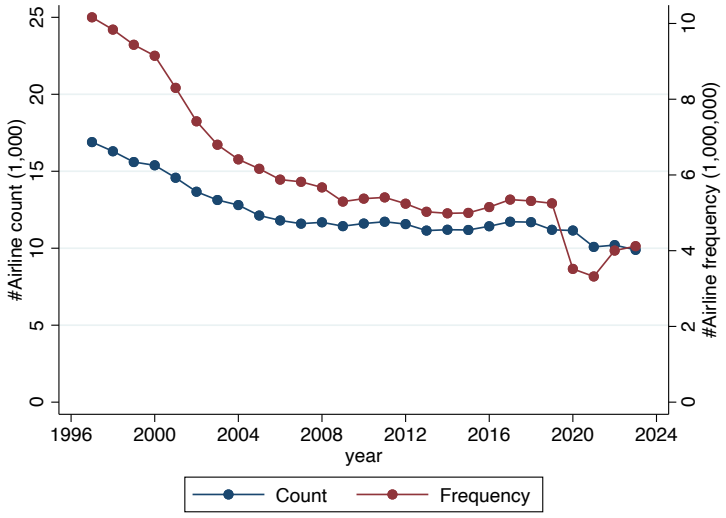
Notes: Panel A plots residuals of airline intensity. Panel B plots residuals of PM2.5 from van Donkelaar et al. (2021). Both panels include year and grid fixed effects.

Figure A12: Fuel use for ten sample airlines



Notes: The left figure shows estimated fuel burn for 10 flights departing from London Heathrow Airport (LHR) to Paris Charles de Gaulle Airport (CDG), Hong Kong (HKG), Athens (ATH), Dubai (DXB), Edinburgh (EDI), Geneva (GVA), Helsinki (HEL), Lisbon (LIS), Madrid (MAD) and New York (JFK). Unit: ton; The right figure displays fuel burn per kilometer. Unit: ton/km. Source: OAG.

Figure A13: Propeller airline count and frequency over time



Notes: We count non-duplicated airlines that are operated by propeller planes in blue and aggregate airline frequency in red. The figure displays level changes over the period 1997-2023.

A2 Additional Tables

A2.1 Airline routes as pollution instrument

Table A1: Airline intensity and air pollution, dropping pixels close to airports, 100km as cutoff distance

	PM2.5 (μg/m^3)		
	US		Global
	(1)	(2)	(3)
All routes	2.035***	1.460**	3.094***
	(0.460)	(0.705)	(1.159)
Observations	2,179,304	2,179,304	10,459,281
R-square	0.152	0.179	0.719
Y-mean	10.058	10.058	18.688
Y-sd	7.096	7.096	18.114
Year, Month	Y	Y	Y
DOW FEs	Y	Y	
State FEs	Y		
County FEs		Y	
Country FEs			Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-day level, 1998-2019; Sample in Column (3) is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the state level in Column (1) and (2) and at the country level in Column (3). Significance: * 0.10, ** 0.05, *** 0.01.

Table A2: Airline intensity and air pollution, alternative specifications

	Panel A: Using discrete intensity		
	PM2.5 (μg/m^3)		
	US		Global
All routes	0.595***	0.287***	1.199*
	(0.124)	(0.104)	(0.664)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.150	0.187	0.710
	Panel B: Dropping small intensity		
All routes	1.012***	0.446***	3.659**
	(0.278)	(0.149)	(1.479)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.152	0.187	0.713
	Panel C: Duplicated airlines based on frequency		
All routes	1.553**	0.824***	3.649**
	(0.637)	(0.292)	(1.471)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.151	0.187	0.713
Y-mean	10.036	10.036	19.381
Y-sd	7.375	7.375	17.068
Year, Month	Y	Y	Y
DOW FEs	Y	Y	
State FEs	Y		
County FEs		Y	
Country FEs			Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-day level, 1998-2019; Sample in Column (3) is at the grid-month level, 1998-2019. In Panel A, we use a dummy to code *AllRoutes* by replacing all positive values with 1. In Panel B and C, X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the state level in Column (1) and (2) and at the country level in Column (3). Significance: * 0.10, ** 0.05, *** 0.01.

Table A3: Airline intensity and air pollution, alternative clustering

	Panel A: Clustering at smaller geographic units		
	PM2.5 (μg/m^3)		
	US		Global
All routes	1.935***	1.349**	3.901***
	(0.564)	(0.558)	(0.024)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.151	0.187	0.713
Y-mean	10.036	10.036	19.381
Y-sd	7.375	7.375	17.068
	Panel B: Clustering at the year level		
All routes	1.935***	1.349***	3.901***
	(0.198)	(0.235)	(0.261)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.151	0.187	0.713
	Panel C: Two way clustering		
All routes	1.935**	1.349**	3.901**
	(0.835)	(0.570)	(1.573)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.151	0.187	0.713
Year, Month	Y	Y	Y
DOW FEs	Y	Y	
State FEs	Y		
County FEs		Y	
Country FEs			Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-day level, 1998-2019; Sample in Column (3) is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. In Column (1) and (2), standard errors are clustered at the county level in Panel A, year level in Panel B, and state and year level in Panel C. In Column (3), standard errors are clustered at the grid level in Panel A, year level in Panel B, and grid and year level in Panel C. Significance: * 0.10, ** 0.05, *** 0.01.

Table A4: Airline intensity and air pollution, using OpenSky data

	PM2.5 (μg/m^3)		
	US		Global
	(1)	(2)	(3)
All aircrafts	0.531***	0.443**	2.554***
	(0.186)	(0.183)	(0.068)
Observations	673,436	673,436	352,873
R-square	0.086	0.125	0.166
Y-mean	8.102	8.102	5.991
Y-sd	6.275	6.275	1.713
Year, Month	Y	Y	Y
DOW FEs	Y	Y	
State FEs	Y		
County FEs		Y	
Country FEs			Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-week level, 2016-2019; Sample in Column (3) is at the grid-week level, 2016-2019. X-unit is the number of overhead aircraft (count over each grid-week) divided by 1,000,000. Standard errors are clustered at the state level in Column (1) and (2) and at the country level in Column (3). Significance: * 0.10, ** 0.05, *** 0.01.

Table A5: Flight ban during September 2001

	PM2.5 (μg/m^3)	
	(1)	(2)
All routes	4.067***	4.068***
	(0.644)	(0.644)
Event		-0.850**
		(0.330)
Event × All routes		-5.508***
		(1.749)
Observations	1,276,884	1,276,884
R-square	0.119	0.119
Y-mean	12.244	12.244
Y-sd	8.213	8.213
DOW FEs	Y	Y
Month FEs	Y	Y

Notes: PM2.5 levels are from EPA monitors. Sample is at the grid-day level in 2001. *AllRoutes* is airline intensity ranging between 0 and 1. *Event* is a time dummy that equals one if this day is between September 11 to 13, 2001, and zero otherwise, and is the same for all grids. Standard errors are clustered at the grid level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A6: Event study, new airlines and air pollution, dropping areas close to airports

	Panel A: Defined by simulated airline intensity		
	Aviation intensity		PM2.5
Post	0.185***	0.205***	0.280***
	(0.000)	(0.000)	(0.008)
Observations	3,168,711	869,043	869,043
R-square	0.605	0.593	0.976
Y-mean	0.151	0.165	18.544
Y-sd	0.176	0.190	18.832
	Panel B: Defined by airline list		
Post	0.025***	0.022**	0.376**
	(0.009)	(0.009)	(0.156)
Observations	2,519,253	809,487	801,658
R-square	0.852	0.815	0.656
Y-mean	0.393	0.436	18.767
Y-sd	0.321	0.310	16.654
Year FEs	Y	Y	Y
Grid FEs	Y	Y	Y

Notes: The analysis is at the grid-year level from 3 years before to 5 years after each grid's event year when new airline routes opened. The sample covers year 1998-2019. The outcome variable in Columns (1) and (2) is airline intensity, ranging between 0 and 1. The outcome variable in Column (3) is PM2.5 level from the gridded van Donkelaar et al. (2021) product. Column (1) includes all grid-years not close to airports. Column (2) includes grid-years with non-missing PM2.5 data in the land area and no ocean area and not close to airports, the same sample as Column (3). Standard errors are clustered at the grid level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A7: Event study, using alternative event window

	Aviation intensity		PM2.5
Post	0.202***	0.222***	0.382***
	(0.000)	(0.000)	(0.007)
Observations	3,925,075	1,556,686	1,556,686
R-square	0.607	0.598	0.972
Y-mean	0.149	0.163	19.363
Y-sd	0.181	0.196	17.828
Year FEs	Y	Y	Y
Grid FEs	Y	Y	Y

Notes: The analysis is at the grid-year level from 4 years before to 6 years after each grid's event year when new airline routes opened. The sample covers year 1998-2019. The outcome variable in Columns (1) and (2) is airline intensity, ranging between 0 and 1. The outcome variable in Column (3) is PM2.5 level from the gridded van Donkelaar et al. (2021) product. Column (1) includes all grid-years. Column (2) includes grid-years with non-missing PM2.5 data in the land area and no ocean area, the same sample as Column (3). Standard errors are clustered at the grid level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A8: Results using MERRA-2 data

	MERRA-2 AOD
	Panel A: Panel model
All routes	0.046*
	(0.025)
Observations	28,411,131
R-square	0.668
Y-mean	0.185
Y-sd	0.112
Year, Month	Y
Country FEs	Y
Trend	Country-specific quadratic
	Panel B: Event study
Post	0.002***
	(0.000)
Observations	1,282,536
R-square	0.940
Y-mean	0.184
Y-sd	0.111
Year FEs	Y
Grid FEs	Y

Notes: The analysis is at the grid-year level. In Panel A, the sample covers 1997-2019. In Panel B, the sample covers 3 years before to 5 years after each grid's event year when new airline routes opened. The outcome variable is aerosol optical depth level from the MERRA-2 product. In Panel A, X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the country level. In Panel B, standard errors are clustered at the grid level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A9: Heterogeneity across baseline pollution and GDP

	PM2.5 (μg/m^3)			
	Panel A: Baseline PM2.5			
	Lowest quartile	2nd	3rd	Highest quartile
All routes	0.717***	1.393***	0.421	0.573
	(0.136)	(0.462)	(0.386)	(1.671)
Observations	7,103,061	7,103,649	7,102,410	7,102,011
R-square	0.257	0.227	0.294	0.317
Y-mean	4.406	10.134	19.666	43.321
Y-sd	1.913	3.140	4.941	15.637
	Panel B: GDP			
	Lowest quartile	2nd	3rd	Highest quartile
All routes	6.591***	2.224	3.484***	2.107***
	(2.287)	(3.601)	(0.969)	(0.649)
Observations	7,200,081	7,162,890	7,094,430	6,920,235
R-square	0.446	0.522	0.752	0.878
Y-mean	32.799	23.628	11.098	9.573
Y-sd	18.905	12.371	9.328	14.247
Year, Month	Y	Y	Y	Y
Country FEs	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are the gridded van Donkelaar et al. (2021) product. Sample is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A10: Dropping pixels with high turbulence or adding turbulence as control

	PM2.5 (μg/m^3)	
All routes	3.641**	3.753**
	(1.507)	(1.463)
Turbulence		-3.520
		(2.171)
Observations	23,774,441	28,411,131
R-square	0.727	0.716
Y-mean	20.850	19.381
Y-sd	17.803	17.068
Year, Month	Y	Y
Country FEs	Y	Y
Country-specific trend	Quadratic	Quadratic

Notes: PM2.5 levels are the gridded van Donkelaar et al. (2021) product. Sample is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. Column (2) include all grid-months. Column (1) drop the top 10% pixel-months with the highest turbulence levels in the sample. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A11: IV regression using MERRA AOD

	Low birth weight	Very low birth weight
MERRA-2 AOD	1.328167***	0.630982***
	(0.511187)	(0.243989)
Observations	182,414	182,414
R-square	-0.057	-0.098
F-stat	17.440	17.440
X-mean	0.185	0.185
X-sd	0.112	0.112
Y-mean	0.023	0.004
Y-sd	0.149	0.061
Adm1 FEs	Y	Y
Country-specific trend	Quadratic	Quadratic

Notes: Each observation is a birth record in the DHS data. Low birth weight is a dummy that equals one if the raw birth weight is below 2500 grams and zero otherwise. Very low birth weight equals one if the birth weight is below 1500 grams. Sample covers year 1998-2019. Air pollution levels are average pollution during in-utero periods, using data from the MERRA-2 aerosol optical depth product. Regression is weighted by survey sample weight, i.e., $v005$. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A12: PM2.5 from more and less noisy airline routes

	Panel A: EPA PM2.5		
More noisy routes	1.342**		0.813*
	(0.513)		(0.430)
Less noisy routes		1.544***	0.978***
		(0.518)	(0.357)
Observations	3,541,822	3,541,822	3,541,822
R-square	0.151	0.151	0.152
	Panel B: van Donkelaar et al. (2021)		
More noisy routes	3.442**		1.216**
	(1.347)		(0.601)
Less noisy routes		4.503**	3.827**
		(1.936)	(1.875)
Observations	28,411,131	28,411,131	28,411,131
R-square	0.711	0.713	0.713

Notes: PM2.5 levels are from EPA monitors in Panel A and the gridded van Donkelaar et al. (2021) product in Panel B. Sample in Panel A is at the monitor-day level, 1998-2019; Sample in Panel B is at the grid-month level, 1998-2019. In Panel A, controls include year, month, day-of-week, and state fixed effects. Standard errors are clustered at the state level. In Panel B, controls include year, month, and country fixed effects. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A13: Low birth weight from more and less noisy airline routes

	Low birth weight		
More noisy routes	0.0203**		0.0080
	(0.0077)		(0.0059)
Less noisy routes		0.0247***	0.0194***
		(0.0074)	(0.0059)
Observations	186,861	186,861	186,861
R-square 0.065	0.065	0.065	0.065
Y-mean	0.023	0.023	0.023
Y-sd	0.150	0.150	0.150

Notes: Each observation is a birth record in the DHS data. Low birth weight is a dummy that equals one if the raw birth weight is below 2500 grams and zero otherwise. Sample covers year 1998-2019. Airline intensity is average intensity during in-utero periods. Regression is weighted by survey sample weight, i.e., $v005$. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A14: Effects of any, leaded, and unleaded airline intensity on PM2.5

	PM2.5 (μg/m^3)			
All routes	3.901**			
	(1.562)			
Leaded routes		2.990**		1.356**
		(1.169)		(0.648)
Unleaded routes			3.959**	3.851**
			(1.552)	(1.532)
Observations	28,411,131	28,411,131	28,411,131	28,411,131
R-square	0.713	0.710	0.713	0.714
Y-mean	19.381	19.381	19.381	19.381
Y-sd	17.068	17.068	17.068	17.068
Year, Month	Y	Y	Y	Y
Country FEs	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from the gridded van Donkelaar et al. (2021) product. Sample is at the grid-month level, 1998-2019. X-unit is aviation intensity or leaded and unleaded intensity ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A15: Heterogeneity across aircraft altitudes

	Panel A: van Donkelaar et al. (2021) data PM2.5 (μg/m^3)		
All aircrafts	10.223***		
	(0.678)		
Low aircrafts		9.342***	
		(0.514)	
High aircrafts			15.087***
			(1.824)
Observations	357,036	357,036	357,036
R-square	0.188	0.182	0.179
Y-mean	5.986	5.986	5.986
Y-sd	1.715	1.715	1.715
Year, Month	Y	Y	Y
Country FEs	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic
	Panel B: EPA monitor data PM2.5 (μg/m^3)		
All aircrafts	0.539***		
	(0.181)		
Low aircrafts		2.465***	
		(0.363)	
High aircrafts			0.272**
			(0.127)
Observations	876,607	876,607	876,607
R-square	0.082	0.084	0.082
Y-mean	7.965	7.965	7.965
Y-sd	6.053	6.053	6.053
Year, Month	Y	Y	Y
State FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors in Panel B and the gridded van Donkelaar et al. (2021) product in Panel A. Sample in Panel B is at the monitor-day level, 2016-2019; Sample in Panel A is at the grid-month level, 2016-2019. X-unit is the number of overhead aircraft (count over each grid-week) divided by 1,000,000. Standard errors are clustered on the country level in Panel A and at the state level in Panel B. Significance: * 0.10, ** 0.05, *** 0.01.

Table A16: Add average altitude as control

	PM2.5 (μg/m^3)	
	van Donkelaar et al. (2021)	EPA monitor
All aircrafts	10.8401***	0.5530***
	(0.3028)	(0.1287)
Altitude	-0.0002**	-0.0004***
	(0.0001)	(0.0001)
Observations	357,574	876,607
R-square	0.179	0.083
Y-mean	5.985	7.965
Y-sd	1.715	6.053
Year, Month	Y	Y
Country FEs	Y	
State FEs		Y
Country-specific trend	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors in Column (2) and the gridded van Donkelaar et al. (2021) product in Column (1). Sample in Column (2) is at the monitor-day level, 2016-2019; Sample in Column (1) is at the grid-month level, 2016-2019. X-unit is the number of overhead aircraft (count over each grid-week) divided by 1,000,000. Standard errors are clustered at the country level in Column (1) and at the state level in Column (2). Significance: * 0.10, ** 0.05, *** 0.01.

Table A17: Control for road intensity

	PM2.5 (μg/m^3)		
	US		Global
All routes	1.371*	1.137**	3.837**
	(0.802)	(0.502)	(1.498)
Road	5.996***	8.070***	0.529
	(0.894)	(1.324)	(1.188)
Observations	3,541,822	3,541,822	28,411,131
R-square	0.156	0.188	0.713
Y-mean	10.036	10.036	19.381
Y-sd	7.375	7.375	17.068
Year, Month	Y	Y	Y
DOW FEs	Y	Y	
State FEs	Y		
County FEs		Y	
Country FEs			Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: PM2.5 levels are from EPA monitors (US) and the gridded van Donkelaar et al. (2021) product (Global). Sample in Column (1) and (2) is at the monitor-day level, 1998-2019; Sample in Column (3) is at the grid-month level, 1998-2019. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the state level in Column (1) and (2) and at the country level in Column (3). Significance: * 0.10, ** 0.05, *** 0.01.

Table A18: Balance table, high vs. low airline intensity

	Panel A: All clusters		
	Geo-clusters with airline routes		Difference
	All routes below median	All routes above median	
	(1)	(2)	(3)
Wealth index	2.724	2.996	0.272**
	[0.946]	[1.032]	(0.085)
Education in years	1.673	1.892	-0.218
	[2.093]	[2.878]	(0.193)
BMI	22.28	22.35	-0.059
	[2.375]	[2.169]	(0.199)
Had vaccination	0.736	0.707	0.029
	[0.272]	[0.321]	(0.016)
#Respondents	35,715	34,698	

	Panel B: Remote clusters 100km away from airports		
	Geo-clusters with airline routes		Difference
	All routes below median	All routes above median	
Wealth index	2.857	3.017	0.159
	[0.969]	[1.023]	(0.105)
Education in years	1.591	1.251	0.341
	[1.974]	[2.423]	(0.213)
BMI	22.31	22.41	-0.099
	[2.372]	[2.204]	(0.246)
Had vaccination	0.745	0.727	0.018
	[0.271]	[0.317]	(0.014)
#Respondents	19,776	14,330	

Notes: The median for *AllRoutes*, *LeadedRoutes*, and *UnleadedRoutes* are 0.364, 0.0468, and 0.350, respectively. Standard deviations are reported in brackets. Standard errors of the differences are reported in parentheses.

Table A19: Balance table for US census tracts, high vs. low airline intensity

	Panel A: All census tracts		
	All routes below median	All routes above median	Difference
	(1)	(2)	(3)
Income	84087.5	71365.6	12721.9***
	[41222.6]	[32990.1]	(304.5)
Working hours per week	38.60	38.24	0.355
	[20.53]	[20.88]	(0.169)
High school degree proportion	56.87	57.57	-0.700
	[54.71]	[54.63]	(0.445)
Female proportion	50.38	50.59	-0.211
	[17.61]	[18.64]	(0.147)
Disability proportion	20.76	20.68	0.080
	[9.801]	[9.103]	(0.077)
#Census tracts	55,862	55,862	

	Panel B: Remote census tracts 100km away from airports		
	All routes below median	All routes above median	Difference
Income	77930.1	77874.6	55.48
	[38236.9]	[37894.6]	(312.7)
Working hours per week	38.59	38.57	-0.023
	[2.515]	[2.508]	(0.020)
High school degree proportion	58.61	58.51	0.099
	[160.6]	[165.7]	(1.339)
Female proportion	50.93	50.94	-0.010
	[3.768]	[3.638]	(0.031)
Disability proportion	20.08	20.13	-0.041
	[7.061]	[7.150]	(0.062)
#Census tracts	36,368	36,367	

Notes: Standard deviations are reported in brackets. Standard errors of the differences are reported in parentheses.

A2.2 Lead and fertility

Table A20: Airline impacts on soil lead

	Soil lead (wt%)	
	(1)	(2)
Leaded routes	473.867*	206.999
	(279.933)	(226.429)
All routes		42.536***
		(13.105)
Observations	14,290	14,290
R-square	0.040	0.066
Y-mean	18.441	18.441
Y-sd	10.107	10.107
Leaded routes-mean	0.0007	0.0007
Leaded routes-sd	0.0018	0.0018
All routes-mean	0.039	0.039
All routes-sd	0.041	0.041
State FEs	Y	Y
Depth FEs	Y	Y

Notes: X-unit is the number of airline route times 1000. The unit of soil lead is weighted percent. Depth FEs indicate three soil groups. Standard errors are clustered at the state level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A21: Aviation lead and fertility in the US

	Birth rate		
	(1)	(2)	(3)
Leaded routes	-2.160***	-2.204***	
	(0.798)	(0.991)	
All routes		-0.149	
		(2.420)	
Unleaded routes			0.329
			(1.299)
Observations	226,440	226,440	226,440
R-square	0.326	0.326	0.325
Y-mean	69.156	69.156	69.156
Y-sd	15.891	15.891	15.891
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: Outcome variable birth rate is at the county-year level, calculated as the number of newborns divided by the female population aged 15-49. The sample covers 1998-2019. Standard errors are clustered at the state level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A22: Correlation between leaded and unleaded airline intensity

	Panel A: All pixels		
	Leaded routes		
Unleaded routes	0.107***	0.112***	0.021***
	(0.000)	(0.000)	(0.000)
Observations	28,411,131	28,411,131	28,411,131
R-square	0.079	0.084	0.639
Y-mean	0.047	0.047	0.047
Y-sd	0.134	0.134	0.134
Year, Month		Y	Y
Grid FEs			Y

	Panel B: EPA air lead monitors' surrounding pixels		
Unleaded routes	0.425***	0.450***	0.294***
	(0.004)	(0.004)	(0.004)
Observations	462,283	462,283	462,283
R-square	0.029	0.100	0.248
Y-mean	0.323	0.323	0.323
Y-sd	0.374	0.374	0.374
Year, Month		Y	Y
State FEs			Y

Table A23: IV regression, in-utero lead and fertility in the US

	Birth rate		
	OLS	IV	IV
	(1)	(2)	(3)
Air lead	-0.002	-0.167**	-0.130**
	(0.001)	(0.073)	(0.052)
Observations	9,473	9,473	9,473
R-square	0.98	-1.61	-0.97
F-stat		23.04	12.92
Y-mean	83.95	83.95	83.95
Y-sd	159.89	159.89	159.89
County FEs	Y	Y	Y
Year FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: Outcome variable birth rate is at the county-year level, calculated as the number of newborns divided by the female population aged 15-49. The sample covers 1998-2019. Endogenous regressor, air lead, is from the EPA monitor, and unit is mug/m^3. Column (2) uses propeller airline intensity as IV, and Column (3) uses two instruments, propeller airline intensity and all airline intensity as IV. Standard errors are clustered at the state level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A24: Airline impacts on fertility, heterogeneity across ages

	Birth dummy		
	15-19	20-35	36-49
	12.4% of the sample	79.5%	8.1%
	(1)	(2)	(3)
Leaded routes	-0.061***	-0.074***	-0.058***
	(0.012)	(0.010)	(0.008)
Observations	873,934	5,639,178	583,886
R-square	0.008	0.014	0.014
Y-mean	0.257	0.281	0.231
Y-sd	0.437	0.450	0.422
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A25: Airline impacts on fertility, heterogeneity across nutrition levels

	Birth dummy			
	1st quartile	2nd	3rd	4th
	(1)	(2)	(3)	(4)
Leaded routes	-0.066***	-0.068***	-0.069***	-0.071***
	(0.007)	(0.004)	(0.007)	(0.009)
Observations	959,510	889,936	982,244	954,576
R-square	0.008	0.008	0.009	0.013
Y-mean	0.274	0.274	0.268	0.263
Y-sd	0.446	0.446	0.443	0.440
Country FEs	Y	Y	Y	Y
Age, Age^2	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is not weighted. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A26: Airline impacts on fertility, unweighted results

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.065***	-0.069***	
	(0.012)	(0.008)	
All routes		0.004	
		(0.010)	
Unleaded routes			-0.003
			(0.010)
Observations	7,104,615	7,104,615	7,104,615
R-square	0.012	0.012	0.012
Y-mean	0.277	0.277	0.277
Y-sd	0.448	0.448	0.448
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A27: Airline impacts on fertility, drop age controls

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.070***	-0.067***	
	(0.009)	(0.008)	
All routes		-0.002	
		(0.008)	
Unleaded routes			-0.012
			(0.009)
Observations	7,096,998	7,096,998	7,096,998
R-square	0.012	0.012	0.012
Y-mean	0.274	0.274	0.274
Y-sd	0.446	0.446	0.446
Country FEs	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A28: Airline and air lead, dropping areas close to airports

	Lead (μg/m^3)		
	(1)	(2)	(3)
Leaded routes	0.014*	0.014**	
	(0.007)	(0.006)	
All routes		0.003	
		(0.040)	
Unleaded routes			0.016
			(0.033)
Observations	218,893	218,893	218,893
R-square	0.086	0.086	0.086
Y-mean	0.189	0.189	0.189
Y-sd	0.856	0.856	0.856
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
State FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the monitor-day level. Outcome variables are air lead levels from the EPA monitors, unit μg/m^3. The sample covers the years 1998-2019. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the state-year level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A29: Airline lead and fertility in DHS countries, dropping areas close to airports

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.071***	-0.066***	
	(0.009)	(0.007)	
All routes		-0.005	
		(0.008)	
Unleaded routes			-0.015*
			(0.008)
Observations	6,081,236	6,081,236	6,081,236
R-square	0.014	0.014	0.013
Y-mean	0.276	0.276	0.276
Y-sd	0.447	0.447	0.447
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A30: Airline and air lead, dropping areas close to airports within 200km

	Lead (μg/m^3)		
	(1)	(2)	(3)
Leaded routes	0.027***	0.026***	
	(0.007)	(0.006)	
All routes		0.012	
		(0.042)	
Unleaded routes			0.034
			(0.033)
Observations	100,133	100,133	100,133
R-square	0.080	0.080	0.080
Y-mean	0.108	0.108	0.108
Y-sd	0.454	0.454	0.454
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
State FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the monitor-day level. Outcome variables are air lead levels from the EPA monitors, unit μg/m^3. The sample covers the years 1998-2019. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the state-year level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A31: Airline lead and fertility in DHS countries, dropping areas close to airports within 200km

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.059***	-0.048***	
	(0.013)	(0.008)	
All routes		-0.011	
		(0.009)	
Unleaded routes			-0.017*
			(0.009)
Observations	4,386,553	4,386,553	4,386,553
R-square	0.017	0.017	0.016
Y-mean	0.276	0.276	0.276
Y-sd	0.447	0.447	0.447
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A32: Airline and air lead, adding road intensity

	Lead (μg/m^3)		
	(1)	(2)	(3)
Leaded routes	0.021***	0.018***	
	(0.007)	(0.006)	
All routes		0.044	
		(0.038)	
Unleaded routes			0.037
			(0.031)
Road	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Observations	318,856	318,856	318,856
R-square	0.093	0.093	0.093
Y-mean	0.152	0.152	0.152
Y-sd	0.731	0.731	0.731
Year FEs	Y	Y	Y
Month FEs	Y	Y	Y
DOW FEs	Y	Y	Y
State FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the monitor-day level. Outcome variables are air lead levels from the EPA monitors, unit μg/m^3. The sample covers the years 1998-2019. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the state-year level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A33: Airline lead and fertility in DHS countries, adding road intensity

	Birth dummy		
	(1)	(2)	(3)
Leaded routes	-0.066***	-0.065***	
	(0.009)	(0.008)	
All routes		-0.001	
		(0.009)	
Unleaded routes			-0.011
			(0.009)
Road	-0.001*	-0.001*	-0.001**
	(0.000)	(0.000)	(0.000)
Observations	7,096,998	7,096,998	7,096,998
R-square	0.013	0.013	0.013
Y-mean	0.274	0.274	0.274
Y-sd	0.446	0.446	0.446
Country FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from the survey year to five years before. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1998-2019. The regression is weighted by survey sample weight, i.e., $v005$. Treatment variables are aviation intensity, ranging between 0 and 1. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A34: Balance table, high leaded vs. high unleaded airline intensity

	Panel A: All clusters		
	Geo-clusters with more routes		Difference
	Unleaded routes above median	Leaded routes above median	
	(1)	(2)	(3)
Wealth index	2.983	3.294	-0.312***
	[1.192]	[1.243]	(0.074)
Education in years	5.632	6.097	-0.465**
	[3.718]	[4.033]	(0.235)
BMI	22.84	22.60	0.240
	[2.448]	[2.044]	(0.176)
Had vaccination	0.800	0.833	-0.033*
	[0.286]	[0.280]	(0.019)
#Respondents	444,740	64,421	
	Panel B: Remote clusters 100km away from airports		
	Geo-clusters with more routes		Difference
	Unleaded routes above median	Leaded routes above median	
Wealth index	2.712	2.555	0.157
	[1.022]	[1.076]	(0.132)
Education in years	3.891	3.141	0.750
	[3.692]	[3.087]	(0.465)
BMI	22.21	21.78	0.435
	[2.081]	[1.502]	(0.265)
Had vaccination	0.708	0.703	0.005
	[0.322]	[0.325]	(0.042)
#Respondents	321,469	7,288	

Notes: Standard deviations are reported in brackets. Standard errors of the differences are reported in parentheses.

Table A35: Airline impacts on blood lead

	Detect rate above 10μg/dL (‰)		
Leaded routes	2.043***	2.001***	
	(0.641)	(0.635)	
All routes		1.638	
		(3.098)	
Unleaded routes			-0.857
			(1.741)
Observations	26,259	26,259	26,259
R-square	0.194	0.194	0.194
Y-mean	6.187	6.187	6.187
Y-sd	20.763	20.763	20.763
Year FEs	Y	Y	Y
Census tract FEs	Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the census tract-year level. The outcome variable is high blood lead detect rate, calculated as the number of tested kids with blood lead above 10μg/dL divided by the total number of tested kids. The sample covers the years 2012-2019. X-unit is aviation intensity ranging between 0 and 1. Standard errors are clustered at the tract-year level. Significance: * 0.10, ** 0.05, *** 0.01.

A2.3 Leaded fuel ban

Table A36: Fertility impacts of leaded gasoline bans, heterogeneity across female ages

	Birth dummy		
	Panel A: Single difference		
	15-19	20-35	36-49
	12.4% of the sample	79.5%	8.1%
Post	0.226***	0.032***	-0.113***
	(0.006)	(0.005)	(0.005)
Observations	895,695	5,765,796	586,608
R-square	0.079	0.022	0.036
Y-mean	0.256	0.281	0.231
Y-sd	0.437	0.449	0.422
	Panel B: Double difference		
Post	0.225***	0.031***	-0.113***
	(0.006)	(0.005)	(0.005)
Road	-0.016***	-0.019***	-0.011***
	(0.006)	(0.007)	(0.003)
Post × Road	0.027*	0.026**	0.004
	(0.014)	(0.013)	(0.005)
Observations	895,695	5,765,796	586,608
R-square	0.079	0.022	0.036
Y-mean	0.256	0.281	0.231
Y-sd	0.437	0.449	0.422
Adm1 FEs	Y	Y	Y
Age, Age^2	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from three years before to five years after the leaded fuel ban in each country. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1992-2021. Regression is weighted by survey sample weight, i.e., $v005$. Column (1) includes both geocoded and non-geocoded respondents, and Column (2)-(4) only includes geocoded respondents. Standard errors are clustered at the country level in Column (1)-(3) and at the adm1 level in Column (4). Significance: * 0.10, ** 0.05, *** 0.01.

Table A37: Fertility impacts of leaded gasoline bans, heterogeneity across nutrition levels

	Birth dummy			
	Panel A: Single difference			
	1st quartile	2nd	3rd	4th
	(1)	(2)	(3)	(4)
Post	0.051***	0.052***	0.050***	0.030***
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	960,743	891,683	984,932	961,217
R-square	0.016	0.017	0.017	0.018
Y-mean	0.274	0.274	0.268	0.263
Y-sd	0.446	0.446	0.443	0.440
	Panel B: Double difference			
Post	0.051***	0.052***	0.050***	0.029***
	(0.008)	(0.008)	(0.008)	(0.008)
Road	-0.010***	-0.013***	-0.009**	-0.015***
	(0.003)	(0.004)	(0.004)	(0.004)
Post × Road	0.008**	0.009	0.011*	0.017**
	(0.004)	(0.005)	(0.006)	(0.008)
Observations	960,743	891,683	984,932	961,217
R-square	0.016	0.017	0.017	0.018
Y-mean	0.274	0.274	0.268	0.263
Y-sd	0.446	0.446	0.443	0.440
Adm1 FEs	Y	Y	Y	Y
Age, Age^2	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from three years before to five years after the leaded fuel ban in each country. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1992-2021. We use DHS blood protein data to classify respondents into four quartiles. Those without protein measurements are dropped from this table. Regression is weighted by survey sample weight, i.e., $v005$. Column (1) includes both geocoded and non-geocoded respondents, and Column (2)-(4) only includes geocoded respondents. Standard errors are clustered at the country level in Column (1)-(3) and at the adm1 level in Column (4). Significance: * 0.10, ** 0.05, *** 0.01.

Table A38: Fertility impacts of leaded gasoline bans, unweighted results

	Birth dummy			
	Panel A: Single difference			
	(1)	(2)	(3)	(4)
Post	0.075***	0.048***	0.048***	0.048***
	(0.004)	(0.011)	(0.011)	(0.005)
Observations	7,867,478	7,255,716	7,255,716	7,255,716
R-square	0.068	0.014	0.021	0.021
Y-mean	0.300	0.277	0.277	0.277
Y-sd	0.458	0.447	0.447	0.447
	Panel B: Double difference			
Post		0.048***	0.047***	0.047***
		(0.011)	(0.005)	(0.005)
Road		-0.016***	-0.014***	-0.014***
		(0.005)	(0.004)	(0.004)
Post × Road		0.016*	0.016**	0.016*
		(0.009)	(0.007)	(0.008)
Observations		7,255,716	7,255,716	7,255,716
R-square		0.015	0.021	0.021
Y-mean		0.277	0.277	0.277
Y-sd		0.447	0.447	0.447
Country FEs		Y		
Adm1 FEs			Y	Y
Age, Age^2	Y	Y	Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from three years before to five years after the leaded fuel ban in each country. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1992-2021. Regression is not weighted. Column (1) includes both geocoded and non-geocoded respondents, and Column (2)-(4) only includes geocoded respondents. Standard errors are clustered at the country level in Column (1)-(3) and at the adm1 level in Column (4). Significance: * 0.10, ** 0.05, *** 0.01.

Table A39: Fertility impacts of leaded gasoline bans, drop age controls

	Birth dummy			
	Panel A: Single difference			
	(1)	(2)	(3)	(4)
Post	0.072***	0.045***	0.044***	0.044***
	(0.006)	(0.009)	(0.009)	(0.006)
Observations	7,859,861	7,248,099	7,248,099	7,248,099
R-square	0.069	0.014	0.021	0.021
Y-mean	0.296	0.274	0.274	0.274
Y-sd	0.457	0.446	0.446	0.446
	Panel B: Double difference			
Post		0.044***	0.044***	0.044***
		(0.009)	(0.005)	(0.006)
Road		-0.021**	-0.017***	-0.017***
		(0.009)	(0.006)	(0.006)
Post × Road		0.023	0.023**	0.023*
		(0.015)	(0.011)	(0.012)
Observations		7,248,099	7,248,099	7,248,099
R-square		0.014	0.021	0.021
Y-mean		0.274	0.274	0.274
Y-sd		0.446	0.446	0.446
Country FEs		Y		
Adm1 FEs			Y	Y
Country-specific trend	Quadratic	Quadratic	Quadratic	Quadratic

Notes: The analysis is at the individual-year level. For each female respondent aged 15-29, we code her birth decision from three years before to five years after the leaded fuel ban in each country. The outcome variable is a dummy that equals one if this female respondent gives birth in a given year and zero otherwise. The sample covers the years 1992-2021. Regression is weighted by survey sample weight, i.e., $v005$. Column (1) includes both geocoded and non-geocoded respondents, and Column (2)-(4) only includes geocoded respondents. Standard errors are clustered at the country level in Column (1)-(3) and at the adm1 level in Column (4). Significance: * 0.10, ** 0.05, *** 0.01.

Table A40: Fertility impacts of leaded gasoline bans in the US

	Birth rate			
	Panel A: Single difference			
	(1)	(2)	(3)	(4)
Post	0.028***	0.028***	0.028***	0.028***
	(0.005)	(0.005)	(0.002)	(0.002)
Observations	209,775	209,775	209,775	209,775
R-square	0.154	0.511	0.511	0.541
Y-mean	0.668	0.668	0.668	0.668
Y-sd	0.156	0.156	0.156	0.156
	Panel B: Double difference			
Post	0.021***	0.021***	0.021***	0.022***
	(0.006)	(0.006)	(0.003)	(0.003)
Road	-0.037	0.000	0.000	0.000
	(0.024)	(.)	(.)	(.)
Post × Road	0.044*	0.045**	0.045***	0.036***
	(0.022)	(0.022)	(0.011)	(0.011)
Observations	209,775	209,775	209,775	209,775
R-square	0.154	0.511	0.511	0.541
Y-mean	0.668	0.668	0.668	0.668
Y-sd	0.156	0.156	0.156	0.156
State FEs	Y			
County FEs		Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic	State-specific Quadratic

Notes: The analysis is at the county-year level. Outcome variable, birth rate is at the county-year level, calculated as the number of newborns divided by the female population aged 15-49. The sample covers 1980-2021. Road is the number of road segments at the county level divided by 1000. Standard errors are clustered at the state level in Column (1) and (2) and at the county level in Column (3) and (4). Standard errors are clustered at the state level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A41: Fertility impacts of leaded gasoline bans in the US, using data 1968-2021

	Birth rate			
	Panel A: Single difference			
	(1)	(2)	(3)	(4)
Post	0.033***	0.033***	0.033***	0.033***
	(0.007)	(0.007)	(0.002)	(0.002)
Observations	246,721	246,721	246,721	246,721
R-square	0.229	0.526	0.526	0.548
Y-mean	0.689	0.689	0.689	0.689
Y-sd	0.171	0.171	0.171	0.171
	Panel B: Double difference			
Post	0.029***	0.029***	0.029***	0.026***
	(0.007)	(0.007)	(0.002)	(0.002)
Road	-0.057**	0.000	0.000	0.000
	(0.023)	(.)	(.)	(.)
Post × Road	0.036	0.038*	0.038***	0.078***
	(0.022)	(0.023)	(0.010)	(0.014)
Observations	246,721	246,721	246,721	246,721
R-square	0.230	0.526	0.526	0.548
Y-mean	0.689	0.689	0.689	0.689
Y-sd	0.171	0.171	0.171	0.171
State FEs	Y			
County FEs		Y	Y	Y
Trend	Quadratic	Quadratic	Quadratic	State-specific Quadratic

Notes: The analysis is at the county-year level. Outcome variable, birth rate is at the county-year level, calculated as the number of newborns divided by the female population aged 15-49. The sample covers 1968-2021. Road is the number of road segments at the county level divided by 1000. Standard errors are clustered at the state level in Column (1) and (2) and at the county level in Column (3) and (4). Standard errors are clustered at the state level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A42: Leaded gasoline bans and population

	ln(Population)	
	-8 to 16	-5 to 10
	(1)	(2)
Post	0.045***	0.016**
	(0.012)	(0.007)
Observations	2,975	1,936
R-square	0.998	0.999
Y-mean	16.160	16.127
Y-sd	1.847	1.832
Year FEs	Y	Y
Country FEs	Y	Y
Country-specific trend	Quadratic	Quadratic

Notes: We require a balanced sample at the country-year level from 8 years before to 16 years after the ban in Column (1) and from 5 years before to 10 years after in Column (2). The number of countries is 119 in Column (1) and 121 in Column (2). Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.

Table A43: Leaded gasoline bans and teen pregnancy

	Adolescent fertility rate	
	(1)	(2)
Post	5.698**	2.600***
	(2.624)	(0.989)
Observations	16,430	16,430
R-square	0.891	0.982
Y-mean	74.366	74.366
Y-sd	51.280	51.280
Year FEs	Y	Y
Country FEs	Y	Y
Trend		Country-specific Quadratic

Notes: Outcome is the number of births per 1,000 women ages 15-19 at the country-year level. Standard errors are clustered at the country level. Significance: * 0.10, ** 0.05, *** 0.01.