Air Pollution and Respiratory Infectious Diseases

Sandro Provenzano[1] · Sefi Roth[1,2] · Lutz Sager[3]

Abstract

Recent research suggests that short-term exposure to air pollution is associated with an elevated prevalence of respiratory infectious disease. In this paper, we examine the relationship between the air quality index and weekly cases of COVID-19 and influenza-like illnesses (ILI) in the United States. We address potential bias from omitted variables and measurement error with an instrumental variable approach using atmospheric temperature inversions. Unlike other recent studies, we find no relationship between air quality and either COVID-19 or ILI cases.

Keywords Air pollution · Respiratory disease · Influenza · COVID-19

JEL Classification I18 · Q51 · Q53

1 Introduction

Air pollution exposure has been linked to a wide range of adverse health outcomes such as lower life expectancy, higher infant mortality, and more frequent emergency room visits (Dockery et al. 1993; Chay and Greenstone 2003; Currie and Neidell 2005; Schlenker and Walker 2016). More recent work has linked air pollution exposure to respiratory infectious diseases such as COVID-19 and influenza ('the flu'), both for long-term exposure and for short-term fluctuations (Weaver et al. 2022). From an economic

✉ Lutz Sager
     Lutz.Sager@georgetown.edu

     Sandro Provenzano
     s.provenzano@lse.ac.uk

     Sefi Roth
     s.j.roth@lse.ac.uk

[1] London School of Economics, Houghton St, London WC2A 2AE, UK

[2] IZA, Bonn, Germany

[3] McCourt School of Public Policy, Georgetown University, 3700 O St. NW, Washington, DC 20057, USA

viewpoint, this is important because respiratory infectious diseases generate substantial disruptions and costs. The total annual economic burden of influenza on the U.S. economy is estimated at $87.1 billion, and COVID-19 is projected to cost the U.S. more than $16 trillion (Molinari et al. 2007; Cutler and Summers 2020).

In this paper, we study whether fluctuations in air quality are linked to the two most common and costly respiratory infectious diseases, COVID-19 and influenza-like illnesses (ILI). In theory, air pollution can affect respiratory infectious diseases in three main ways: First, exposure to air pollution can affect the body directly, either by making the respiratory system more vulnerable to such diseases or by inducing inflammatory reactions which impair the immune response to new infections (Ciencewicki and Jaspers 2007). Second, the existence of pollution in the air might affect the airborne survival of respiratory viruses, allowing the virus to remain in the air for longer (Martelletti and Martelletti 2020). Third, air pollution might also lead to changes in human behavior that in turn can impact virus transmission. While the first two channels suggest that there might be a positive link between pollution and respiratory diseases, the last one is more ambiguous. We, therefore, aim to estimate the relationship between air pollution and respiratory infectious diseases empirically. Throughout, we focus on short-term links, analyzing U.S. administrative data on ambient air pollution and cases of ILI and COVID-19 at the weekly level.

Assessing the link between pollution and infectious respiratory disease is challenging due to the presence of correlated omitted variables and measurement errors. In terms of the former, the response to the COVID-19 pandemic in the U.S. involved a combination of federal, state, and local policies, regulations, and guidelines aimed at mitigating the spread of the virus. In the early stages of the pandemic, during 2020, measures such as stay-at-home orders, social distancing, contact tracing, mask mandates, and business and school closures were implemented. While the federal government, through agencies like the Centers for Disease Control and Prevention (CDC), provided recommendations, actual policies and practices varied significantly between and even within states (Hamad et al. 2022). Furthermore, testing availability, shortages of personal protective equipment (PPE), and managing the strain on healthcare systems also varied across the country. Since these measures and conditions might affect both exposure to air pollution and the propagation of infectious diseases, they are some of the many possible omitted variables that can lead to biased estimates. In addition, air quality measurement likely suffers from measurement error due to variation within spatial units and across time.

We overcome these empirical challenges by using an Instrumental Variable (IV) approach that relies on atmospheric temperature inversions that induce plausibly exogenous variation in air quality. Importantly, we document that inversions, which have previously been used as an instrument for air pollution in the economics literature (e.g., Arceo et al. 2016; Bondy et al. 2020), are subject to seasonal patterns which raise concerns about the validity of such instrumentation in some settings. In our ILI analysis, we have a multiyear panel and can account for this with appropriate unit-specific seasonality fixed effects. This does not alter the fact that our instrument is a relevant determinant of air pollution. Where only 1 year of data is available, as in our COVID-19 sample for example, using fixed effects is not feasible. We therefore propose an alternative approach that relies on deviations from long-term averages to overcome this seasonality issue. We hope that this approach may prove useful to other researchers investigating air pollution in contexts of seasonality. Overall, we believe that our preferred instrument is unlikely to be systematically correlated with omitted variables, such as local economic activity or the COVID-19 policies discussed above.

Several recent papers in the economics literature document a positive association between air pollution and respiratory infectious diseases. Clay et al. (2018) find a positive link between elevated pollution from coal plants and the number of deaths during the 1918 Spanish flu pandemic across U.S. cities, exploiting differential timing of the pandemic to overcome confounding factors. Using random variation in wind direction as an instrument, Graff Zivin et al. (2023) find that elevated levels of air pollution (monthly AQI) significantly increase influenza hospitalizations in the United States. Isphording and Pestel (2021) and Austin et al. (2023) apply similar IV approaches to study the impact of particulate matter (PM) concentrations on COVID-19 cases and deaths in Germany and the U.S. respectively. Both studies find significant positive effects. Finally, Persico and Johnson (2021) document increased COVID-19 cases and case fatalities in the weeks following the rollback of environmental regulations in some U.S. regions. These findings suggest a reinforcing relationship between these two important sources of externalities.

We contribute to this growing literature by exploiting an alternative instrument, atmospheric inversions, and by estimating similar models for both COVID-19 and influenza in the United States. Our estimates are precise, based on several time windows of exposure and are robust to different specifications. Contrary to previous studies, we find no evidence that short-term fluctuations in air pollution affect COVID-19 and influenza cases in the U.S. once we control for seasonality or instrument for pollution using temperature inversions. Considering that all other studies, without fail, find a positive association, we believe that it is vital to document our precise null results to foster further investigation on this matter.

2 Data

To study the impact of ambient air pollution on the prevalence and severity of respiratory infectious diseases, we assemble two health datasets.

The first dataset is a weekly panel of COVID-19 cases and fatalities at the U.S. county level. It is based on data collected by usafacts.org that covers 1004 U.S. counties representing 79.6% of the U.S. population from January 2020 until the launch of the vaccination program in December 2020.

The second dataset is a weekly panel on influenza-like illnesses (ILI) at the U.S. state level. It is based on data provided by the Center for Disease Control (US CDC) listing weekly counts of ILI patients across U.S. states over 9 years/full flu seasons—from the 2010/11 flu season beginning in October 2010 until the 2018/19 flu season ending in October 2019. We exclude more recent flu seasons to avoid an overlap with the COVID-19 pandemic.

We complement health data with information on the Air Quality Index (AQI) from the U.S. Environmental Protection Agency (EPA), which we average from the daily to the weekly level in most of our analysis given that the average time between COVID-19 infection and symptoms onset is about 5–6 days (Lauer et al. 2020). AQI is a summary measure of air quality taking into account multiple pollutants, but we also confirm our results using fine particulate matter ($PM_{2.5}$) concentrations only. We construct additional weather covariates describing surface air temperature, precipitation and relative humidity based on data from the NOAA's North American Regional Reanalysis (NARR) database. Finally, for our instrumental variable strategy, we construct measures of atmospheric inversion frequencies based on data from NASA's MERRA-2 database. All data used in this manuscript are

Table 1 Summary statistics

	Mean	Median	SD	Min	Max
A. ILI sample (state-week level)					
ILI cases	282.5	87	567.1	0	11,452
Air quality index (AQI)	43.55	41.35	12.99	15	165
Share of inversion days	.21	.1696	.1729	0	1
Relative humidity	69.91	74.1	14.72	8.2	95
Precipitation (in mm)	2.628	1.848	2.681	0	26
Temperature (Celsius)	13.24	14.17	10.83	− 23	40
Population (thousands)	5,626	4,042	6,118	494	33,872
Observations	21,519				
B. Covid sample (county-week level)					
Covid cases	328	39	1393	0	109,978
Covid fatalities	5.724	0	28.12	0	1608
Air quality index (AQI)	36.75	36.43	18.73	0	351
Share of inversion days	.1829	.1429	.2194	0	1
Relative humidity	69.47	74.49	16.2	9.9	96
Precipitation (in mm)	2.616	1.6	3.096	0	43
Temperature (Celsius)	14.58	15.21	10.22	− 18	40
Population (thousands)	232	91.81	500.2	.56	9519
Observations	47,513				

Summary statistics for the ILI and Covid sample respectively

publicly available. A detailed description of sample construction is provided in Appendix A1. Summary statistics and the distributions of the outcomes variables are presented in Table 1 and Figure A.1 respectively.

3 Methodology

We estimate the short-term relationship between air pollution and two measures of respiratory disease: Weekly cases of (1) COVID-19 at the county level and (2) influenza-like illness (ILI) at the state level. First, consider the expected number of ILI cases in state i during week t:

$$E(Cases_{i,t}) = exp[\beta \, AQI_{i,t} + f(Weather_{i,t}) + \mu_t + \gamma_i] \tag{1}$$

The expected number of ILI cases exponentially[1] depends on air quality, weather and additional time-invariant factors. $AQI_{i,t}$ is the average air quality index (AQI) in state i during week t. We flexibly account for weather conditions in $f(Weather_{i,t})$ by including 20 temperature bins,[2] mean relative humidity and its' interaction with temperature, as well as

---

[1] Exponential mean specifications are standard for count data with long right tails. As shown in Appendix Figures A.1, both the ILI and the COVID-19 case counts exhibit such distributions.

[2] Temperature bins are cut at: − 30, − 10, − 5, 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 45 (all in ˚C).

rainfall and its' square. We also include state fixed effects γ_i and year-week fixed effects, μ_t. We will show that the choice of fixed effects influences the results, likely due to the strong degrees of seasonality in both disease and pollution.

For our second sample, $Cases_{i,t}$ denotes the number of COVID-19 cases in county i during week t, and all other variables are also measured at the county-level. In both cases, our coefficient of interest is β, which describes the relationship between AQI and (exponential) cases of respiratory disease.

We estimate Eq. 1 using the Poisson pseudo-maximum likelihood (PPML)[3] regression as proposed by Silva and Tenreyro (2006) and implemented using the computationally efficient routine in the presence of high-dimensional fixed effects as developed by Correia et al. (2020). However, these estimates may be biased for at least for two reasons—identification and measurement. In terms of identification, estimate $\hat{\beta}$ could be biased when certain variables are omitted from Eq. 1 that affect both air quality and respiratory outcomes. The level of economic activity in a given region and during a given week is just one of the many possible candidates for such an omitted variable. The pandemic response in 2020 is an another possible source of bias in the COVID-19 analysis, as discussed above. Regarding measurement, the assignment of air quality is bound to be imprecise due to variation within spatial units and weeks, biasing estimates $\hat{\beta}$, generally towards 0.

To address these concerns, we turn to a second identification strategy that relies on atmospheric temperature inversions as an instrument to induce plausibly exogenous variation in the levels of air quality. Temperature inversions are short-term atmospheric episodes, usually occurring over a day or less, which lead to a reversal of temperature profiles that lower atmospheric ventilation and thus temporarily increase ground-level pollution levels. They are best suited as instruments for short-term fluctuations in air quality at the daily (Jans et al. 2018; Sager 2019) or weekly level (Arceo et al. 2016). Specifically, we estimate the following linear first-stage relationship:

$$AQI_{i,t} = \rho\, INV_{i,t} + \delta(Weather_{i,t}) + \eta_t + \theta_i + v_{i,t} \tag{2}$$

Air quality in a given county or state i and during week t, $AQI_{i,t}$, depends on the share of days in that week during which inversions occurred, $INV_{i,t}$, as well as the same covariates as in Eq. 1. As we will show, inversions are systematically associated with higher levels of air pollution throughout all specifications and both samples. To estimate the exponential relationship stipulated in Eq. 1, we employ a control function approach as proposed by Wooldridge (2015). In a first step, we estimate Eq. 2 using Ordinary Least Squares (OLS) estimation. We then add the residuals from that regression, $\hat{v}_{i,t}$, to the PPML estimation of Eq. 1.

4 Results

4.1 PPML Estimates

We now turn to the results, beginning with the sample of COVID-19 cases by county and week. Results from the non-instrumented PPML regression are shown in panel (a)

---

[3] PPML maintains the exponential mean structure, incorporates zero counts, and relaxes the mean-variance equivalence typical of standard Poisson models.

Table 2 The association of AQI and ILI/COVID-19 cases (PPML)

	Panel (a): Covid-19 cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.016***	− 0.002***	− 0.001	–	−0.003***
	(0.0016)	(0.0008)	(0.0008)		(0.0007)
Weather controls	Yes	Yes	Yes	–	Yes
County FE	No	Yes	Yes	–	Yes
Week FE	No	Yes	Yes	–	Yes
Flu/AQI Lags	No	No	Yes	–	No
AQI deviations	No	No	No	–	Yes
Observations	47,513	47,430	46,325	–	47,430
Pseudo R2	0.13	0.90	0.90	–	0.90
	Panel (b): ILI cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.013***	0.000	0.001	0.001	0.000
	(0.0044)	(0.0011)	(0.0008)	(0.0013)	(0.0013)
Weather controls	Yes	Yes	Yes	Yes	Yes
State FE	No	Yes	Yes	N/A	Yes
Week FE	No	Yes	Yes	Yes	Yes
Flu/AQI lgs	No	No	Yes	No	No
State-calendar week FE	No	No	No	Yes	No
AQI dviations	No	No	No	No	Yes
Observations	21,519	21,519	21,418	21,519	21,519
Pseudo R2	0.09	0.87	0.88	0.89	0.87

This table reports Poisson pseudo-maximum likelihood (PPML) estimates based on Eq. 1

Panel (a): The dependent variable are weekly COVID-19 cases at the US county level provided by usafacts. org, and the main explanatory variable is the air quality index (AQI) (in column (5) deviations) by the US Environmental Protection Agency (US EPA), with higher AQI values indicating higher air pollution. Panel (b): The dependent variable are weekly ILI cases at the US state level provided by the Center for Disease Control (US CDC), and the main explanatory variable is the air quality index (AQI) (in column (5) deviations) by the US Environmental Protection Agency (US EPA), with higher AQI values indicating higher air pollution. Standard errors in parentheses are cluster-robust to autocorrelation within each flu season by state

Standard errors in parentheses are cluster-robust at the level of (a) counties and (b) flu seasons by state respectively

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

of Table 2. In column (1), we include weather controls only. As in previous studies, we find a positive association, suggesting that each 1-point increase in AQI is associated with an increase in COVID-19 cases by 1.6%. But this association disappears when we account for time-invariant factors, common time-varying shocks, and region-specific seasonality. When we include county and week fixed effects in column (2), we find a very small coefficient. In fact, the sign of the coefficient actually reverses, suggesting that higher levels of pollution reduce the number of COVID-19 cases. When we add 1-week lags of COVID-19 cases and AQI to account for any potential autocorrelation

across time in column (3), the coefficient becomes statistically and economically insignificant.

The same holds for ILI cases by state and week, results for which are shown in panel (b) of Table 2. While the simple model in column (1) suggests a positive association (1.3% increase per extra AQI point), we estimate precise zeros when including state and year fixed effects (column 2) and 1-week lags for ILI cases and AQI (column 3).

Another concern is that pollution and disease are seasonal (as we show in Figure A.2) and that seasonality may differ across regions. For example, we may see more pollution and more respiratory disease cases during late January in states that routinely experience severe winters. This introduces a substantial risk of bias when estimating the relationship between air quality and respiratory disease without accounting for region-specific seasonality trends.

We take two approaches to region-specific seasonality. First, we include time-varying fixed effects in column (4). Because we limit our COVID-19 sample to the time before vaccines were widely available, essentially the year 2020, we cannot estimate a specification with county-calendar week fixed effects. But in the ILI sample we can include state-calendar week fixed effects in column (4) of Table 2. The coefficient of interest remains at essentially zero. Second, in column (5) we calculate AQI as the deviation from its' long-run average in each county (or state) and calendar week (e.g. mean AQI in week 2 of each year between 2010–2020). This approach using long-run deviations of AQI is feasible even for the COVID-19 sample where we only use case data for 2020. Coefficient estimates remain very close to zero.

Given the limited testing capacity during the beginning of the pandemic and the virus' ability to spread asymptomatic-ally, we also examine the effect of air pollution on COVID-19 fatalities, with a time lag of 2 weeks to allow for the delay between infection and death. The results are presented in Table A.1 and show no link between pollution and COVID-19 related mortality either.

4.2 Instrumental Variable Estimates

Next, we turn to the control function estimates using inversions as an instrument for air quality.

The COVID-19 results are shown in panel (a) and the ILI results in panel (b) of Table 3, with each column again showing equivalent specifications to those in Table 2. This approach requires that more frequent inversions are associated with higher pollution levels. Our first-stage results at the bottom of the panel show this to be the case. Increasing the share of inversion days in a week from 0 to 1 is associated with an increase in AQI of between 6 and 8 points in our county-level sample and between 12 and 18 in our state-level sample. Both times, the relationship is reasonably strong as indicated by high F-statistics. Importantly, columns (4) and (5) show that the instrument is robust to accounting for region-specific seasonality. The approach also requires that the frequency of inversions is not, after controlling for weather conditions and fixed effects, associated with any change in respiratory health other than through changes in air quality. We are not aware of any mechanism that would lead to such confounding, though we cannot be certain.

Turning to the coefficient of interest, we again estimate a positive relationship between AQI and COVID-19/ILI cases in column (1). After including county/state and week fixed effects (column 2) and controlling for 1-week lags for respiratory cases, AQI and inversions (column 3), the coefficients fall substantially, but remain positive and statistically

Table 3 The association of AQI and ILI/COVID-19 cases (CF/PPML)

	Panel (a): Covid-19 cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.114***	0.019***	0.024***	–	– 0.006
	(0.026)	(0.007)	(0.006)		(0.006)
Weather controls	Yes	Yes	Yes	–	Yes
County FE	No	Yes	Yes	–	Yes
Week FE	No	Yes	Yes	–	Yes
Flu/AQI lags	No	No	Yes	–	No
AQI deviations	No	No	No	–	Yes
Observations	47,513	47,430	46,325	–	47,430
Pseudo R2	0.14	0.90	0.90	–	0.90
First stage: $\hat{\rho}$	6.2***	7.4***	7.3***	–	7.7***
	(1.2)	(0.6)	(0.5)		(0.5)
IV F-stat	26.8	131.1	192.2	–	212.5
	Panel (b): ILI cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.063***	0.022***	0.017***	0.001	0.005
	(0.012)	(0.003)	(0.002)	(0.004)	(0.003)
Weather controls	Yes	Yes	Yes	Yes	Yes
State FE	No	Yes	Yes	N/A	Yes
Week FE	No	Yes	Yes	Yes	Yes
Flu/AQI lags	No	No	Yes	No	No
State-calendar week FE	No	No	No	Yes	No
AQI deviations	No	No	No	No	Yes
Observations	21,519	21,519	21,418	21,519	21,519
Pseudo R^2	0.09	0.87	0.89	0.89	0.87
First stage: $\hat{\rho}$	12.5***	15.8***	16.4***	17.7***	17.8***
	(1.6)	(0.8)	(0.7)	(0.8)	(0.9)
IV F-stat	61.9	382.2	512.4	442.0	422.4

This table reports Poisson pseudo-maximum likelihood (PPML) estimates based on the control function approach as proposed by Wooldridge (2015) that uses inversions (column (1)–(4)) and inversions deviations (column (5)) as instruments for air quality (columns (1)–(4)) and air quality deviations (column (5)). The corresponding first-stage regression coefficients are shown in the bottom rows

Panel (a): The dependent variable are weekly COVID-19 cases at the US county level provided by usafacts. org

Panel (b): The dependent variable are weekly ILI cases at the US state level provided by the Center for Disease Control (US CDC)

Standard errors in parentheses are bootstrapped using cluster-wise resampling at the level of (a) counties and (b) flu seasons by state respectively

$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

significant. This would seem to support the findings in the previous literature. However, as discussed above, we believe that it is crucial to account for region-specific seasonality. When we do so by taking deviations from the long-run mean (column 5) the point estimates become very small and no longer significantly different from 0 for both COVID-19 and ILI cases. The same is true when we include state-calendar week fixed effects (column 4) for the ILI sample. Table A.2, shows similar results for COVID-19 fatalities.

Taken together, the results in Table 3 suggest that the positive association between AQI and COVID-19 / ILI cases disappears when accounting for region-specific seasonality using unit-calendar-week fixed effects or equivalent deviations (column 4 and 5). Uniform week fixed effects (column 2) appear insufficient to filter out confounding seasonality. To explore how much different elements of seasonality are relevant here, we provide additional results with unit-specific seasonality effects in Appendix Table A.3. Point estimates do fall substantially upon including unit-season or unit-month fixed effects. But unit-calendar-week fixed effects seem necessary to fully capture region-specific seasonality. We believe that this is an important insight, as seasonality patterns of respiratory disease, air quality and inversions will all differ across regions of the United States.
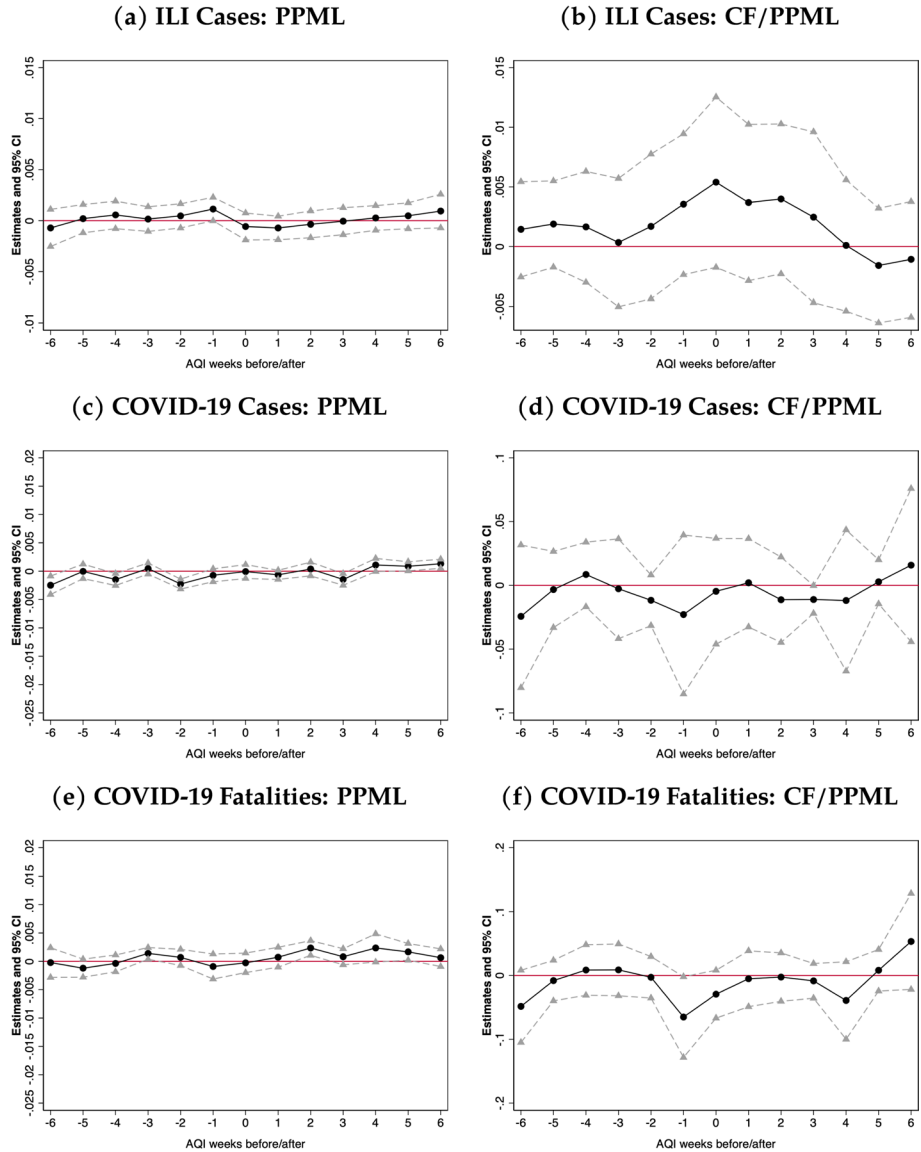
4.3 Timing and Cumulative Effects

Our results fail to support a relationship between air pollution and respiratory disease at the weekly level. However, it might be that pollution exposure takes some time to translate into higher case counts. In Fig. 1, we allow for a delay of up to 6 weeks. Panel (a) is equivalent to column (5) of Table 2 and panel (b) shows control function estimates equivalent to column (5) in Table 3, but with leads and lags.[4] In panels (c) and (d), we do the same for COVID-19 cases, and in panels (e) and (f) for COVID-19 fatalities. Throughout, we find no association to air quality in either the preceding or following weeks.

In addition to lagged effects, air pollution damages may also accumulate over time. To test for such a possibility, we adopt an approach similar to Deschenes et al. (2020), who study the relationship between obesity and cumulative $PM_{2.5}$ exposure over preceeding months. Appendix Figure A.3 shows results when respiratory outcomes in a given week are linked to the cumulative pollution exposure over 2/4/8/12 weeks (same week plus N-1 weeks prior). The results show no systematic evidence of an association over longer time horizons up to 12 weeks. While we report these results for completeness, this is not our preferred approach. Estimates become increasingly more noisy with longer time windows. As noted above, this is likely because the inversion instrument is better suited to study air quality fluctuations at the daily or weekly level rather than multi-week or longer periods.

Another possibility is that there may be an even shorter term relationship between air quality and respiratory disease than our analysis at the weekly level can uncover. To test this assumption, we use the sample of COVID-19 cases, which are available at the daily level. As shown in Appendix Table A.4, we find very similar results as in the weekly analysis, and certainly no convincing evidence of a positive effect. COVID-19 has an incubation period of multiple days, with median estimates around 5–6 days and ranges up to 14 days (Lauer et al. 2020). While contemporaneous pollution may

---

[4] We estimate Eq. 2 in each time period (same week + 6 lags + 6 leads), including each time all inversion instruments (same week + 6 lags + 6 leads). We then estimate Eq. 1 with leads and lags, as well as residuals from all first-stage regressions.

(a) ILI Cases: PPML **(b) ILI Cases: CF/PPML**

(c) COVID-19 Cases: PPML **(d) COVID-19 Cases: CF/PPML**

(e) COVID-19 Fatalities: PPML **(f) COVID-19 Fatalities: CF/PPML**

Fig. 1 Association between leads/Lags of AQI deviations and disease. *Note*: The figures on the left plot the estimates based on Eq. 1 equivalent to column (5) of Table 2, but with 6 leads and lags. The figures on the right plot the estimates based on the control function approach equivalent to column (5) of Tables 3 and A.2 respectively, but with 6 leads and lags. The 95% confidence interval is included in gray

alter behavior that changes the likelihood of detection, it is implausible that air pollution would drive actual infections recorded on the same day. To flexibly allow for longer incubation periods, we show in Appendix Figure A.4 results for specifications with lags up to 14 days. Again, we see little support of a systematic relationship between AQI and recorded COVID-19 cases.

4.4 Robustness

Besides the question of timing, we also test our control function results for alternative inversion measures used as instruments. As discussed in Sager (2019), inversions can be measured between various distances of atmospheric layers, and calculated as binary indicators or continuous temperature differences (inversion strength). In Appendix Table A.5, we test alternatives to our baseline measure, the share of 24-h average inversion periods between the bottom and the second-lowest atmospheric layer (25 hPa difference). Specifically, we show that the results remain unchanged when we use a measure of inversion strength, when we consider higher atmospheric layers (50 hPa difference) and when we look at nighttime inversions between 00:00 and 06:00 a.m. only.

Finally, we may wonder if AQI, which is a summary measure based on a combination of multiple pollutants, best captures the relevant dimensions of air quality. The most frequently analysed single pollutant is particulate matter pollution, measured as the number of particles less than 2.5 micrometers in diameter ($PM_{2.5}$). In Appendix Table A.6, we find very similar results when looking at the relationship between COVID-19 cases and $PM_{2.5}$ concentrations at the county-week level. Again, the estimates become indistinguishable from zero (or even slightly negative) after adequately controlling for location-specific seasonality. However, it is important to mention that using a single pollutant in such an IV setting might be problematic as the presence of correlated pollutants might violate the exclusion restriction, which is one reasons we prefer our main AQI estimates.

5 Discussion

As we mentioned above, to the best of our knowledge, most if not all published work on this subject finds a positive short-term correlation between air pollution and infectious disease. This raises the important question as to why our results are different. We posit three broad possible explanations: First, the difference could be due to the choice of empirical strategy. More specifically, a significant proportion of the literature only documents correlations (e.g., Bashir et al. 2020; Zhu et al. 2020), while we use quasi-experimental variation induced by the inversion instrument to establish causality. Notably we also see a positive correlation in column 1 of Table 2. However, some studies (mainly in the economics literature) also aim to establish causal links. In these cases, the difference in results might be due to the reliance on different quasi-experimental variation, i.e. instrument choice, or the chosen specification. In this context, we believe it is particularly important how one controls for seasonality, as shown in our main results.

Second, the difference could stem from each paper using different data sources with different temporal and spatial resolutions and also different periods and geographies. For example, Isphording and Pestel (2021) studies Germany rather than the U.S., and Persico and Johnson (2021) focuses on a few weeks of exposure which is a slightly different time window compared to our main results. However, as we show above, we find no evidence that different timing and cumulative effects can explain these differences within our sample. And, given that other studies also cover the same geography as our study (e.g., Austin et al. 2023), we think that these are less likely explanations.

Finally, null results may not be published due to publication bias, which is the systematic tendency of academic journals and researchers to preferentially publish studies with

statistically significant results over those with null results. This type of bias can result from various factors, including editorial and referee decisions and authors' choices in submitting or not submitting their work ("the file drawer problem"). While there is no way to test for publication bias, it is another plausible explanation as to why most published work on this subject finds a positive link.

6 Conclusion

This paper has examined the short-term relationship between air pollution and respiratory infectious diseases in the United States. While some empirical models suggest that air pollution is indeed positively associated with cases of COVID-19 and ILI, as found in previous studies, this relationship vanishes when we use our instrumental variable approach or account for the substantial degrees of seasonality present in both air quality and respiratory disease. Importantly, our null results are precise, robust to different specifications, and remain virtually the same for different time windows of exposure. We recognise that a number of contributions find a positive relationship between air pollution and infectious diseases. Indeed, we are not aware of any published work that finds an absence of such a relationship, as we do here. Our analysis suggests that one important factor is the chosen identification strategy and, in particular, how one controls for region-specific seasonality. Other explanations for this difference include the reliance of different sources of variation or the exact choice of data sources, time periods and geographic units, and less sanguine ones, such as publication bias that may prevent null results from being circulated. That is why we believe that it is vital to document our null results to foster further academic investigation on this matter.

Declarations

References

Arceo E, Hanna R, Oliva P (2016) Does the effect of pollution on infant mortality differ between developing and developed countries? Evidence from Mexico city. Econ J 126(591):257–280

Austin W, Carattini S, Gomez-Mahecha J, Pesko MF (2023) The effects of contemporaneous air pollution on COVID-19 morbidity and mortality. J Environ Econ Manag 119:102815

Bashir MF, Jiang B, Komal B, Bashir MA, Farooq TH, Iqbal N, Bashir M et al (2020) Correlation between environmental pollution indicators and COVID-19 pandemic: a brief study in Californian context. Environ Res 187:109652

Bondy M, Roth S, Sager L (2020) Crime is in the air: the contemporaneous relationship between air pollution and crime. J Assoc Environ Res Econ 7(3):555–585

Chay KY, Greenstone M (2003) The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. Q J Econ 118(3):1121–1167

Ciencewicki J, Jaspers I (2007) Air pollution and respiratory viral infection. Inhal Toxicol 19(14):1135–1146

Clay K, Lewis J, Severnini E (2018) Pollution, infectious disease, and mortality: evidence from the 1918 Spanish influenza pandemic. J Econ Hist 78(4):1179–1209

Correia S, Guimarães P, Zylkin T (2020) Fast Poisson estimation with high-dimensional fixed effects. Stata J 20(1):95–115

Currie J, Neidell M (2005) Air pollution and infant health: what can we learn from California's recent experience? Q J Econ 120(3):1003–1030

Cutler DM, Summers LH (2020) The COVID-19 pandemic and the \$16 trillion virus. J Am Med Assoc (JAMA) 324(15):1495–1496

Deschenes O, Wang H, Wang S, Zhang P (2020) The effect of air pollution on body weight and obesity: evidence from China. J Dev Econ 145:102461

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE (1993) An association between air pollution and mortality in six U.S. cities. N Engl J Med 329(24):1753–1759

Graff Zivin J, Neidell M, Sanders NJ, Singer G (2023) When externalities collide: influenza and pollution. Am Econ J Appl Econ 15(2):320–351

Hamad R, Lyman KA, Lin F, Modrow MF, Ozluk P, Azar KM, Goodin A, Isasi CR, Kitzman HE, Knight SJ et al (2022) The US COVID-19 county policy database: a novel resource to support pandemic-related research. BMC Public Health 22(1):1882

Isphording IE, Pestel N (2021) Pandemic meets pollution: poor air quality increases deaths by COVID-19. J Environ Econ Manag 108:102448–102448

Jans J, Johansson P, Nilsson JP (2018) Economic status, air quality, and child health: evidence from inversion episodes. J Health Econ 61:220–232

Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 172(9):577–582

Martelletti L, Martelletti P (2020) Air pollution and the novel COVID-19 disease: a putative disease risk factor. SN Compr Clin Med 2(4):383–387

Molinari N-AM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, Bridges CB (2007) The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 25(27):5086–5096

Persico CL, Johnson KR (2021) The effects of increased pollution on COVID-19 cases and deaths. J Environ Econ Manag 107:102431

Sager L (2019) Estimating the effect of air pollution on road safety using atmospheric temperature inversions. J Environ Econ Manag 98:102250

Schlenker W, Walker WR (2016) Airports, air pollution, and contemporaneous health. Rev Econ Stud 83(2):768–809

Silva JMCS, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658

Weaver AK, Head JR, Gould CF, Carlton EJ, Remais JV (2022) Environmental factors influencing COVID-19 incidence and severity. Annu Rev Public Health 43:271–291

Wooldridge JM (2015) Control function methods in applied econometrics. J Hum Resour 50(2):420–445

Zhu Y, Xie J, Huang F, Cao L (2020) Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. Sci Total Environ 727:138704