I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15947

Air Pollution and Respiratory Infectious Diseases

Sandro Provenzano
Sefi Roth
Lutz Sager

FEBRUARY 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15947

Air Pollution and Respiratory Infectious Diseases

Sandro Provenzano
LSE

Sefi Roth
LSE and IZA

Lutz Sager
Georgetown University

FEBRUARY 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Air Pollution and Respiratory Infectious Diseases

Recent research suggests that short-term exposure to air pollution is associated with an elevated prevalence of respiratory infectious disease. We examine the relationship between the air quality index (AQI) and weekly cases of influenza-like illnesses (ILI) and COVID-19 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 ILI or COVID-19 cases.

JEL Classification: I18, Q51, Q53

Keywords: air pollution, respiratory disease, COVID-19

Corresponding author:
Sefi Roth
London School of Economics
Houghton St
London WC2A 2AE
United Kingdom

E-mail: s.j.roth@lse.ac.uk

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 & Greenstone 2003, Currie & Neidell 2005, Schlenker & Walker 2016). More recent work has linked air pollution exposure to respiratory infectious diseases such as influenza ('the flu') and COVID-19, both for long-term exposure and for short-term fluctuations (Weaver et al. 2022). From an economic viewpoint, this is important because respiratory infectious diseases generate substantial disruptions and costs. The total annual economic burden of influenza on the US economy is estimated at $87.1 billion, and COVID-19 is projected to cost the US more than $16 trillion (Molinari et al. 2007, Cutler & Summers 2020).

In this paper, we study whether fluctuations in air quality are linked to the two most common and costly respiratory infectious diseases, influenza-like illnesses (ILI) and COVID-19. 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 & 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 & 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 error. Time-varying economic activity is just one of the many possible omitted variables that is likely to affect both air pollution and the propagation of infectious diseases. In addition, air quality measurement likely suffers from measurement error due to variation within spatial units and across time. We overcome these challenges by using an Instrumental Variable (IV) approach that relies on atmospheric temperature inversions. Importantly, we document that inversions, which have previously been used as an instrumental variable for air pollution in the economics literature (e.g. Arceo et al. (2016) or Bondy et al. (2020)), are subject to seasonal patterns which raise concerns about the validity of such instrumentation in some settings. In multi-year panels, this can be accounted for by using appropriate fixed effects. Where only one year of data is available, as in our COVID-19 sample for example, using fixed effects to overcome this empirical challenge is clearly not feasible. We therefore propose an alternative specification using 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.

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 Spanish flu pandemic to overcome confounding factors. Using random variation in wind direction as an instrument, Graff Zivin et al. (forthcoming) find that elevated levels of air pollution (based on monthly AQI) significantly increase influenza hospitalizations in the U.S. Austin et al. (2020) and Isphording & Pestel (2021) apply similar IV approaches to study the impact of particulate matter (PM) concentrations on COVID-19 cases and deaths in the U.S and Germany respectively. Both studies find significant positive effects. Finally, Persico & Johnson (2021) document increased COVID-19 cases and case fatalities in the weeks following the rollback of environmental regulations in the U.S. 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 influenza and COVID-19 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 elevated levels of air pollution affect weekly influenza and COVID-19 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 on influenza-like illnesses (ILI) at the US state level. It is based on data provided by the Center for Disease Control (US CDC) listing weekly counts of ILI patients across US 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.

The second dataset is a weekly panel of COVID-19 cases and fatalities at the US county level. It is based on data collected by usafacts.org that covers 1,004 US counties representing 79.6% of the US population from January 2020 until the launch of the vaccination program in December 2020.

We complement health data with information on the Air Quality Index (AQI) from the US Environmental Protection Agency (US EPA), which we average from the daily to the weekly level. We also construct weather covariates describing surface air temperature, precipitation and relative humidity based on data from NOAA's NARR database. Finally, for our instrumental variable strategy, we construct measures of atmospheric inversion frequencies based on data from NASA's MERRA-2 database. A detailed description of sample construction is provided in Appendix A.1. Summary statistics and the distributions of the outcomes variables are presented in Table 1 and Figure A.1 respectively.

[Table 1 about here.]

3 Methodology

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

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

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, $Temp_{i,t}$ is average temperature, $RH_{i,t}$ is relative humidity, and $Rain_{i,t}$ is cumulative rainfall. We flexibly account for weather conditions in $f(Weather_{i,t})$ by including 20 temperature bins[2], relative humidity, and its' interaction with temperature, as well as rainfall and its square. We also include state fixed effects γ_i and year-week fixed effects, μ_t. We will show that these fixed effects influence the results due to the likely strong degrees of seasonality.

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 Equation 1 using the Poisson pseudo-maximum likelihood (PPML)[3] regression as proposed by Silva & 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 Equation 1 that

[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.).

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

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. 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 Equation 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 Equation 1, we employ a control function approach as proposed by Wooldridge (2015). In a first step, we estimate Equation 2 using Ordinary Least Squares (OLS) estimation. We then add the residuals from that regression, $\hat{v}_{i,t}$, to the PPML estimation of Equation 1.

4 Results

4.1 PPML Estimates

We now turn to the results, beginning with the ILI state-year sample. Results from the non-instrumented PPML regression are shown in panel (a) of Table 2. In column (1), we include weather controls but no fixed effects. As in previous studies, we find a positive association, suggesting that each 1-point increase in AQI is associated with an increase in ILI cases by 1.2%. However, once we include state and year-week fixed effects in column (2), we find a precisely estimated zero. When we add one-week lags of ILI cases and AQI to account for any potential autocorrelation across time, the coefficient remains statistically and economically insignificant.

[Table 2 about here.]

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 flu 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 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 by state and calendar week. Given the similarity to the calendar week fixed effects, it is not surprising that the coefficient is also zero. However, using long-run deviations of AQI allows us to do the same in the COVID-19 sample where we have only one year of data.

Panel (b) of Table 2 shows equivalent results for COVID-19 cases by county and week. Here, state fixed effects are replaced with county fixed effects. We limit ourselves to the time before any Covid-19 vaccines were widely available, essentially the year 2020, and are thus not able to estimate a specification with county-calendar week fixed effects. While the simple specification without fixed effects in column (1) again suggest a positive association, we detect no such relationship after accounting for time-invariant factors, common time-varying shocks, and region-specific seasonality. In fact, our results are small but the sign of the coefficient actually reverses, suggesting that higher levels of pollution reduce the number of COVID-19 cases. Given the limited testing capacity during the beginning of the pandemic and the virus' ability to spread asymptomatically, we also examine the effect of air pollution on COVID-19 fatalities, with a time lag of two 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.

4.2 Instrumental Variable Estimates

Next, we turn to the control function estimates using inversions as an instrument for air quality. The ILI results are shown in Panel (a) 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 in the bottom 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 12 and 18 points in our state-level sample. 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.

[Table 3 about here.]

Turning to the coefficient of interest, we again estimate a positive relationship between AQI and ILI cases in column (1). After including state and week fixed effects (column 2) and controlling for one-week lags for ILI cases, AQI and inversions (column 3), the coefficient of interest falls by more than half, but remains positive and statistically significant. As discussed above, we believe that it is crucial to account for region-specific seasonality. When we do so, by including state-calendar week fixed effects (column 4) or by taking deviations from the long-run mean for ILI cases, AQI and inversions (column 5), the point estimates become very small and are no longer significantly different from 0.

[Figure 1 about here.]

The control function results for the COVID-19 sample are shown in Panel (b) of Table 3. Again, inversions systematically predict AQI at the county level. And again, we fail to detect any systematic relationship between air quality and COVID-19 cases after adjusting for region-specific seasonality in column (5). Table A.2, shows similar results for COVID-19 fatalities.

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 Figure 1, we allow for a delay of up to six 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.

5 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 associated with weekly cases of ILI and Covid-19, as found in previous studies, this relationship vanishes when we use our instrumental variable approach or account for seasonality. 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. One explanation for the difference could be sampling noise, as every paper uses somewhat different data sources, time periods and geographic units. Other explanations are less sanguine, such as publication bias that may prevent null results from being circulated. Whichever the reasons, we believe that it is vital to document our null results to foster further academic investigation on this matter.

[4] We estimate equation 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 equation 1 with leads and lags, as well as residuals from all first-stage regressions.

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', *The Economic Journal* **126**(591), 257–280.

Austin, W., Carattini, S., Gomez Mahecha, J. & Pesko, M. (2020), 'Covid-19 mortality and contemporaneous air pollution', *CESifo Working Paper No. 8609* .

Bondy, M., Roth, S. & Sager, L. (2020), 'Crime is in the air: The contemporaneous relationship between air pollution and crime', *Journal of the Association of Environmental and Resource Economists* **7**(3), 555–585.

Chay, K. Y. & Greenstone, M. (2003), 'The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession', *The Quarterly Journal of Economics* **118**(3), 1121–1167.

Ciencewicki, J. & Jaspers, I. (2007), 'Air pollution and respiratory viral infection', *Inhalation Toxicology* **19**(14), 1135–1146.

Clay, K., Lewis, J. & Severnini, E. (2018), 'Pollution, infectious disease, and mortality: Evidence from the 1918 spanish influenza pandemic', *The Journal of Economic History* **78**(4), 1179–1209.

Correia, S., Guimarães, P. & Zylkin, T. (2020), 'Fast poisson estimation with high-dimensional fixed effects', *The Stata Journal* **20**(1), 95–115.

Currie, J. & Neidell, M. (2005), 'Air pollution and infant health: What can we learn from california's recent experience?', *The Quarterly Journal of Economics* **120**(3), 1003–1030.

Cutler, D. M. & Summers, L. H. (2020), 'The covid-19 pandemic and the $16 trillion virus', *Journal of the American Medical Association (JAMA)* **324**(15), 1495–1496.

Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G. & Speizer, F. E. (1993), 'An association between air pollution and mortality in six u.s. cities', *The New England Journal of Medicine* **329**(24), 1753–1759.

Graff Zivin, J., Neidell, M., Sanders, N. J. & Singer, G. (forthcoming), 'When externalities collide: Influenza and pollution', *American Economic Journal: Applied Economics* **in press**.

Isphording, I. E. & Pestel, N. (2021), 'Pandemic meets pollution: Poor air quality increases deaths by covid-19', *Journal of Environmental Economics and Management* **108**, 102448–102448.

Jans, J., Johansson, P. & Nilsson, J. P. (2018), 'Economic status, air quality, and child health: Evidence from inversion episodes', *Journal of Health Economics* **61**, 220–232.

Martelletti, L. & Martelletti, P. (2020), 'Air pollution and the novel covid-19 disease: A putative disease risk factor', *SN Comprehensive Clinical Medicine* **2**(4), 383–387.

Molinari, N.-A. M., Ortega-Sanchez, I. R., Messonnier, M. L., Thompson, W. W., Wortley, P. M., Weintraub, E. & Bridges, C. B. (2007), 'The annual impact of seasonal influenza in the us: Measuring disease burden and costs', *Vaccine* **25**(27), 5086–5096.

Persico, C. L. & Johnson, K. R. (2021), 'The effects of increased pollution on covid-19 cases and deaths', *Journal of Environmental Economics and Management* **107**, 102431.

Sager, L. (2019), 'Estimating the effect of air pollution on road safety using atmospheric temperature inversions', *Journal of Environmental Economics and Management* **98**, 102250.

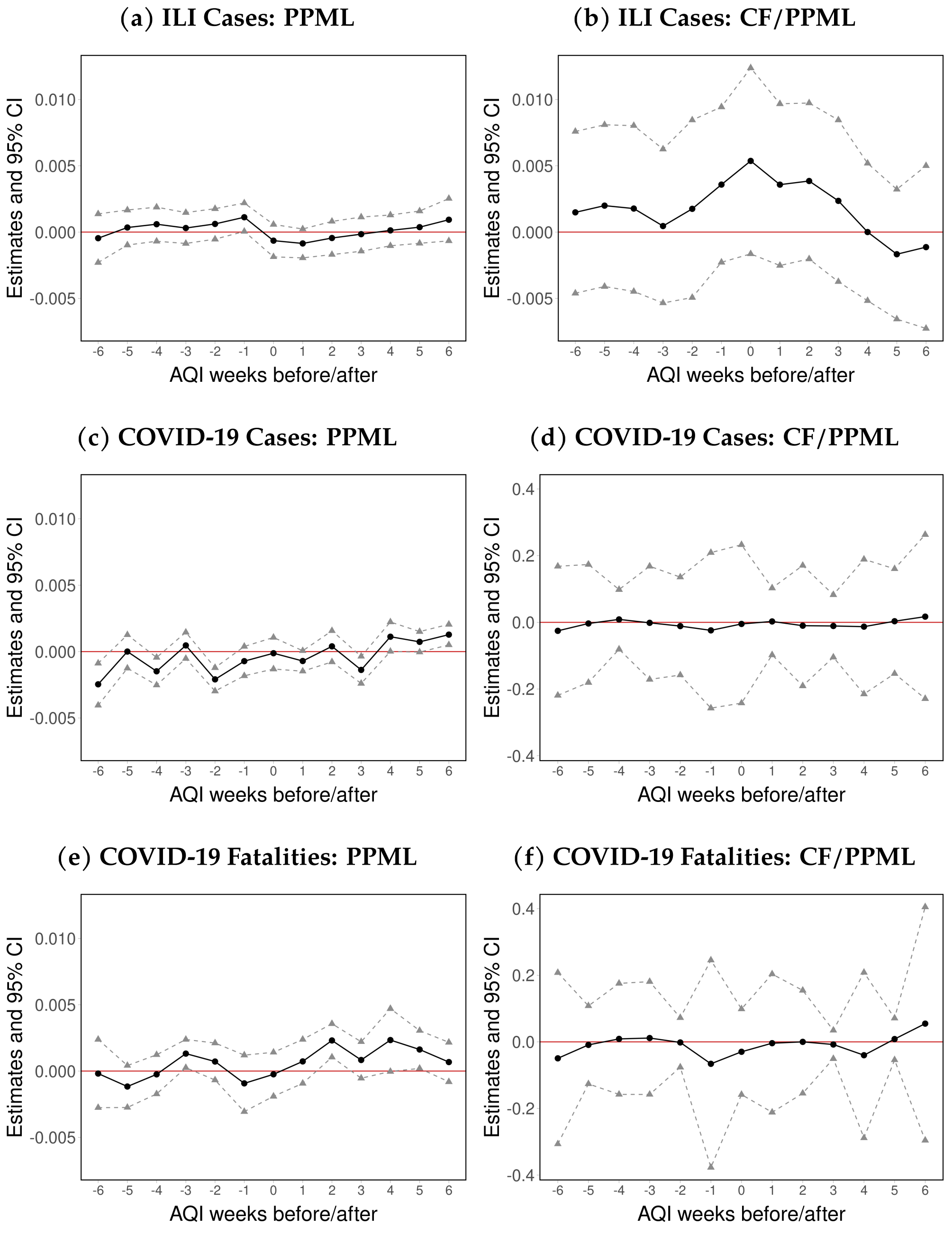
Schlenker, W. & Walker, W. R. (2016), 'Airports, air pollution, and contemporaneous health', *The Review of Economic Studies* **83**(2 (295)), 768–809.

Silva, J. M. C. S. & Tenreyro, S. (2006), 'The log of gravity', *The Review of Economics and Statistics* **88**(4), 641–658.

Weaver, A. K., Head, J. R., Gould, C. F., Carlton, E. J. & Remais, J. V. (2022), 'Environmental factors influencing covid-19 incidence and severity', *Annual Review of Public Health* **43**, 271–291.

Wooldridge, J. M. (2015), 'Control function methods in applied econometrics', *The Journal of Human Resources* **50**(2), 420–445.

Figure 1: Association between Leads/Lags of AQI Deviations and Disease



Note: The figures on the left plot the estimates based on Equation 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, 3 and A.2 respectively, but with 6 leads and lags. The 95% confidence interval is included in gray.

Table 1: Summary Statistics

	Min.	Max.	Mean	Median	SD.
A. ILI Sample (State Level)					
ILI Patient Rate	0	161.9	6.0	2.4	10.3
Air Quality Index	15.4	213.8	43.6	41.4	13.1
Inversions per Week	0	1.0	0.2	0.2	0.2
Relative Humidity	8.2	94.7	69.9	74.1	14.7
Precipitation (in mm)	0	25.5	2.6	1.8	2.7
Temperature (in °C)	-23.2	40.3	13.2	14.2	10.8
Population (in 1,000)	494	33,872	5,626	4,042	6,118
Observations:	21,519				
B. Covid Sample (County Level)					
COVID-19 Case Rate	0	5,675.4	136.4	47.2	224.4
COVID-19 Fatality Rate	0	161.5	2.1	0	4.6
Air Quality Index	0	848.6	36.9	36.4	19.6
Inversions per Week	0	1.0	0.2	0.1	0.2
Relative Humidity	9.9	96.1	69.5	74.5	16.2
Precipitation (in mm)	0	43.5	2.6	1.6	3.1
Temperature (in °C)	-17.9	40	14.6	15.2	10.2
Population (in 1,000)	0.6	9,519.3	232.4	91.8	500.6
Observations:	47,430				

Note: Summary statistics for the ILI and Covid sample respectively.

Table 2: The Association of AQI and ILI/COVID-19 Cases (PPML)

	Panel (a): ILI Cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.012^{***}	0.000	0.001	0.000	0.000
	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
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 R2	0.09	0.87	0.88	0.89	0.87

	Panel (b): Covid-19 Cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.008^{***}	-0.002^{***}	-0.001	-	-0.003^{***}
	(0.0012)	(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	47,431	47,430	46,325	-	47,430
Pseudo R2	0.11	0.90	0.90	-	0.90

Note: This table reports Poisson pseudo-maximum likelihood (PPML) estimates based on Equation 1.
Panel (a): 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.
Panel (b): 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. Standard errors in parentheses are cluster-robust to autocorrelation at the level of counties.
*p<0.1; **p<0.05; ***p<0.01

Table 3: The Association of AQI and ILI/COVID-19 Cases (CF/PPML)

	Panel (a): 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: $\hat{\rho}$	12.3^{***}	15.9^{***}	16.4^{***}	17.7^{***}	17.8^{***}
	(1.6)	(0.8)	(0.7)	(0.9)	(0.9)
IV F-stat	59.9	378.4	506.5	432.3	414.4

	Panel (b): Covid-19 Cases				
	(1)	(2)	(3)	(4)	(5)
AQI	0.126^{***}	0.019^{**}	0.023^{***}	-	-0.006
	(0.023)	(0.007)	(0.008)		(0.007)
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	47,431	47,430	46,325	-	47,430
Pseudo R2	0.12	0.90	0.90	-	0.90
First stage: $\hat{\rho}$	6.1^{***}	7.6^{***}	7.4^{***}	-	7.8^{***}
	(1.2)	(0.7)	(0.5)		(0.6)
IV F-stat	25.2	129.6	189.6	-	198.1

Note: 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. Standard errors in parentheses are bootstrapped using cluster-wise resampling at the level of flu seasons by state.
Panel (a): The dependent variable are weekly ILI cases at the US state level provided by the Center for Disease Control (US CDC).
Panel (b): The dependent variable are weekly COVID-19 cases at the US county level provided by usafacts.org.
*p<0.1; **p<0.05; ***p<0.01

APPENDIX FOR ONLINE PUBLICATION

Air Pollution and Respiratory Infectious Diseases

February 9, 2023

Sandro Provenzano, LSE
s.provenzano@lse.ac.uk

Sefi Roth, LSE and IZA
s.j.roth@lse.ac.uk

Lutz Sager, Georgetown University
lutz.sager@georgetown.edu

A.1 Detailed description of data

To study the impact of ambient air pollution on the prevalence and severity of respiratory infectious diseases, we assemble two datasets. The first dataset is a weekly panel on influenza-like illnesses (ILI) at the US state level for the study period between 2010 to 2019, covering all 50 states as well as the District of Columbia. The second dataset is a weekly panel on COVID-19 covering 1,004 US counties representing 79.6% of the US population from early until late 2020 when the vaccination program rolled out.[5] Summary statistics for our key variables in our data are presented in Appendix Table 1. In panel A, we show key statistics for our state level data on ILI and in Panel B for our county level COVID-19 sample. There are five main data inputs to create these datasets:

Influenza-like Illness Surveillance Network (ILINet):

The first data source is provided by the Center for Disease Control (US CDC) in collaboration with the state and local health departments and health care providers. From this dataset, we obtain information about weekly counts of ILI patients across US 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.[6] We exclude more recent flu seasons to avoid an overlap with the COVID-19 pandemic. The ILINet data is based on information from over 3,000 healthcare providers across the US and allows to track the weekly ILI prevalence for all states over several years. The CDC defines ILI patients as reporting symptoms of 'fever (temperature of 100°F [37.8°C] or greater) and a cough and/or a sore throat.[7]

COVID-19:

The second dataset covers weekly counts of COVID-19 cases and fatalities from usafacts.org, which is based on county-level reports by local health authorities across the US. This dataset covers the period between the start of the pandemic in the US in January 2020 until the launch of the vaccination program in December 2020. The dataset includes two measures of COVID-19 prevalence, cases and fatalities.[8] While cases might in principle be a better measure of disease prevalence, this indicator is hard to measure in practice. More specifically, the limited testing capacity, especially during the beginning of the pandemic, in conjunction with the virus' ability to spread in asymptomatic people limits the reliability of Covid cases as a main outcome of interest (Subbaraman 2020). This type of reporting error is less likely to occur when using COVID-19 fatali-

[5] The remainder of the 3,143, most of which have small populations, do not report COVID-19 counts.

[6] Following the US CDC convention of an epidemiological week, we define weeks as starting on Sundays. A flu season is defined to run for one year starting in October.

[7] More information about the ILINet data can be found at: `https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html`.

[8] More information about the COVID-19 data can be found at: `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/`.

ties instead. Yet, fatalities are also a noisy measure of disease prevalence as the time between infection and death can take several weeks and vary a lot between cases. In addition, fatalities are not only a measure of disease prevalence but also of severity. Consequently, there are advantages and disadvantages from using either cases or fatalities. For this reason, we decided to focus on cases but we will show that we obtain similar results using deaths as an alternative measure.

Air Quality Index (AQI):

We complement the health data with measures of air quality from the US Environmental Protection Agency (US EPA). The AQI is calculated on a daily basis using a variety of measures including carbon monoxide, nitrogen dioxide, ozone, sulfur dioxide as well as inhalable particulate matters (PM2.5 and PM10). For the period between 2010 and 2019, the EPA provides a daily AQI for 1,173 counties representing 83% of the US population. A higher AQI indicates a higher ambient air pollution level and ranges between 0 and 213.8 with a mean of 43.6 and standard deviation of 13.1 in our ILI sample, and between 0 and 848.6 with a mean of 36.9 and a standard deviation of 19.6 in our Covid sample (see Table 1).[9] The AQI represents as noisy measure of air pollution since the defining parameter varies across space and time, and measuring stations are capturing the air pollution at the station but not the average across the county. To combine the daily county-level AQI data with the ILI and COVID-19 data, we aggregate the AQI by week, and for ILI additionally by state weighing by county population.

Atmospheric Temperature Inversions:

To obtain quasi-random variation in local air quality, we use atmospheric temperature inversions as an instrument. These are short-term atmospheric episodes, usually occurring over a day or less, which lead to a reversal of temperature profiles that reduce atmospheric ventilation and consequently increase ground-level pollution levels. While inversion episodes tend to be associated with other atmospheric and weather conditions, they are arguably independent of human behavior on the ground which is why they have been proposed as an instrument for air pollution at the daily (Jans et al. 2018, Sager 2019) or weekly level (Arceo et al. 2016). Inversions follow cyclical patterns, which we control for. We measure the occurrence of inversions based on satellite-derived three-dimensional temperature profiles of the atmosphere, which come from the MERRA-2 reanalysis project.[10] These provide 3-hourly mean temperatures by latitude, longitude and atmospheric pressure levels[11], which correspond to altitude. We spatially match inversion grid centroids to US counties.[12] Whenever

[9] As a reference, an AQI between 0-50 is considered as 'Good'.

[10] Further information about the dataset can be found at: https://disc.gsfc.nasa.gov/datasets/M2I3NVASM_5.12.4/summary.

[11] Location by latitude and longitude is divided into grid cells of size $0.625° \times 0.5°$. Altitude is divided into 42 atmospheric pressure levels with 25hPa intervals, which corresponds to approximately 200 meters.

[12] Specifically, we assign each grid point to the county it falls into and calculating mean

the daily mean temperature at the pressure level closest to the ground is lower than the temperature at the next higher-up level (25hPa less pressure which corresponds to roughly 200m in altitude), we define an inversion in that county on that day. Our final instrument measures the share of days within each week during which such an inversion occurred. When looking at the ILI sample, we aggregate inversions to the state level using the county population share as weight.

Additional Weather Controls:

While inversions may well be independent of human behavior, they are known to covary with certain weather conditions—including precipitation patterns, frozen rain, and fog formation—that may themselves affect human behavior. Consequently, we add control variables measuring surface air temperature, precipitation as well as relative humidity, which are taken from the North American Regional Reanalysis (NARR) project by the National Oceanic and Atmospheric Administration (NOAA) and spatially matched in the same way as described above.

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', *The Economic Journal* **126**(591), 257–280.

Jans, J., Johansson, P. & Nilsson, J. P. (2018), 'Economic status, air quality, and child health: Evidence from inversion episodes', *Journal of Health Economics* **61**, 220–232.

Sager, L. (2019), 'Estimating the effect of air pollution on road safety using atmospheric temperature inversions', *Journal of Environmental Economics and Management* **98**, 102250.

Subbaraman, N. (2020), 'Why daily death tolls have become unusually important in understanding the coronavirus pandemic', *Nature* .
URL: *https://doi.org/10.1038/d41586-020-01008-1*

temperature levels in a county on a day and each pressure level.For those few counties which do not contain a grid point, we assign readings from the grid point closest to the county centroid.

A.2 Additional Tables

Table A.1: The Association of AQI and COVID-19 Fatalities (PPML)

	Dependent variable: Covid-19 Fatalities				
	(1)	(2)	(3)	(4)	(5)
AQI	0.008***	−0.001	0.000	-	−0.000
	(0.001)	(0.001)	(0.001)		(0.001)
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	44,250	44,250	42,295	-	44,250
Pseudo R2	0.10	0.75	0.76		0.75

Note: This table reports Poisson pseudo-maximum likelihood (PPML) estimates based on Equation 1 for the Covid sample. The dependent variable are weekly COVID-19 fatalities at the US county level two weeks later 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. Standard errors in parentheses are cluster-robust to autocorrelation at the level of counties. *p<0.1; **p<0.05; ***p<0.01

Table A.2: The Association of AQI and COVID-19 Fatalities (CF/PPML)

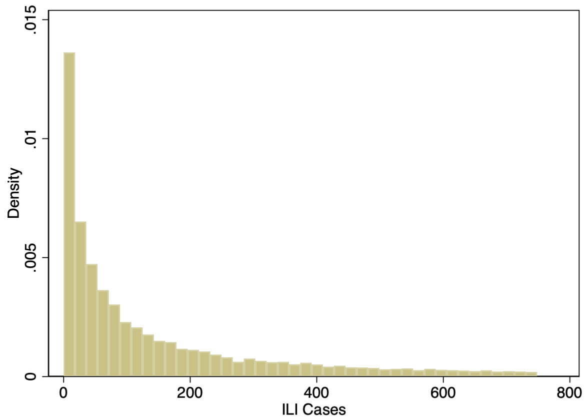
	Dependent variable: Covid-19 Fatalities				
	(1)	(2)	(3)	(4)	(5)
AQI	0.090***	0.068***	0.045***	-	−0.017
	(0.026)	(0.015)	(0.007)		(0.011)
Weather Controls	Yes	Yes	Yes	-	Yes
County FE	No	Yes	Yes	-	Yes
Week FE	No	Yes	Yes	-	Yes
Flu/AQI Lags	No	No	Yes	-	No
AQI Deviations	No	No	No	-	Yes
Observations	44,250	44,250	42,295	-	44,250
Pseudo R2	0.11	0.75	0.77		0.75
	First Stage relationship between inversions and AQI				
	(1)	(2)	(3)	(4)	(5)
Inversions	6.1***	7.6***	7.4***	-	7.8***
	(1.2)	(0.7)	(0.5)		(0.6)
IV F-stat	25.2	129.6	189.6	-	198.1
Adj. R^2	0.15	0.52	0.62	-	0.27

Note: 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)-(3)) and inversions deviations (column (5)) as instruments for air quality (columns (1)-(3)) and air quality deviations (column (5)) for the Covid sample. The dependent variable are weekly Covid fatalities at the US county level two weeks later 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. The corresponding first-stage regressions are shown in the bottom rows. Standard errors in parentheses are bootstrapped using cluster-wise resampling at the level of counties. *p<0.1; **p<0.05; ***p<0.01

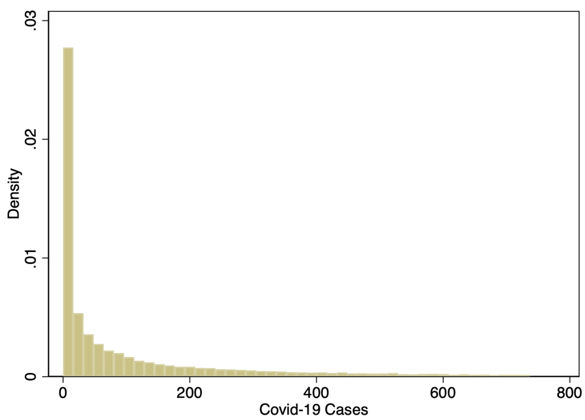
A.3 Additional Figures

Figure A.1: Distribution of outcome variables

(a) Distribution of ILI cases



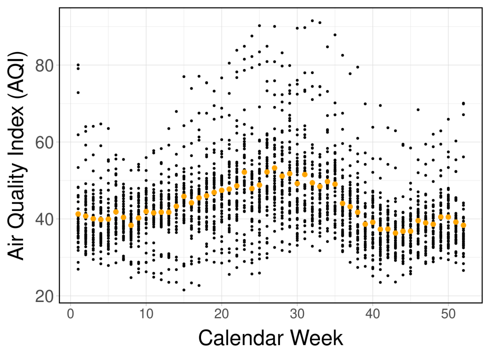
(b) Distribution of Covid-19 cases



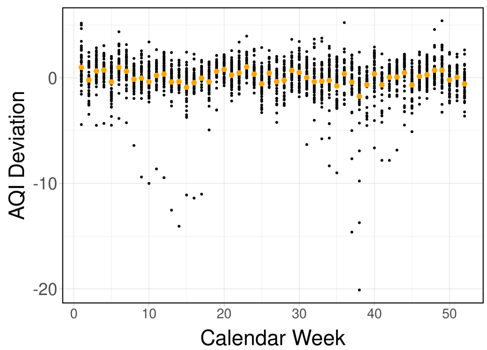
Note: Histograms showing the distribution of the two outcome variables: (a) ILI cases per week per state and (b) Covid-19 cases per week per county. Both graphs are truncated at the 90th percentile.

Figure A.2: Aggregate Seasonality Patterns

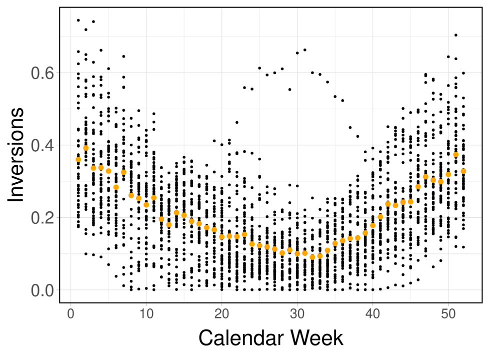
(a) AQI



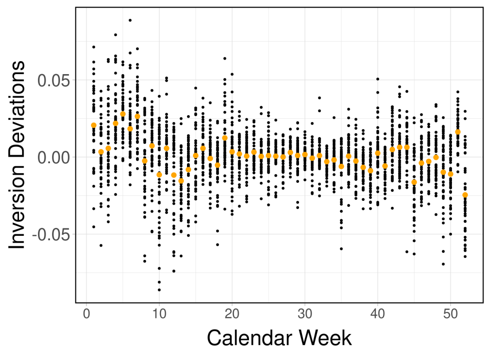
(b) AQI Deviations



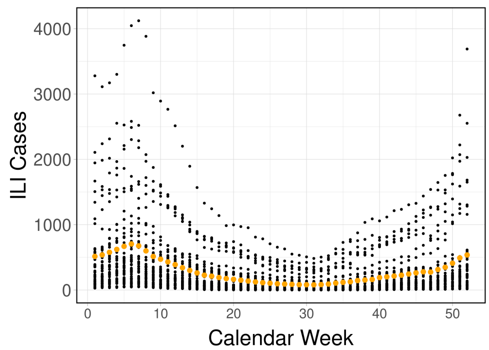
(c) Inversions



(d) Inversions Deviations



(e) ILI Cases



Note: These figures use the ILI sample to plot key variables over calendar week (the weekly bins are in orange). The AQI and inversions exhibit clear seasonal patterns, with the AQI being elevated during the summer and inversions during the winter. In contrast, taking the deviations from their respective long-term mean helps with removing the seasonality. ILI cases are displayed at the bottom and also exhibit a seasonality with elevated levels during the winter. It should be noted that these figures illustrate the aggregate seasonality across US states, and that the seasonality within states or counties is likely to be even more pronounced and vary across locations. By demeaning the AQI and inversions by state (or county) and calendar week, we remove this location specific seasonality.