The Effects of Increased Pollution on COVID-19 Cases and Deaths

Claudia L. Persico
Department of Public Administration and Policy, School of Public Affairs, American University and IZA
cpersico@american.edu

Kathryn R. Johnson
Department of Public Administration and Policy, School of Public Affairs, American University
kj5677a@student.american.edu

June 22, 2020

Abstract

The SARS-COV-2 virus, also known as the coronavirus, has spread around the world. While a growing literature suggests that exposure to pollution can cause respiratory illness and increase deaths among the elderly, little is known about whether increases in pollution could cause additional or more severe infections from COVID-19, which typically manifests as a respiratory infection. Using variation in pollution induced by a rollback of enforcement of environmental regulations by the Environmental Protection Agency (EPA) and a difference in differences design, we estimate the effects of increased pollution on county-level COVID-19 deaths and cases. We find that counties with more Toxic Release Inventory (TRI) sites saw a 13 percent increase in pollution on average following the EPA's rollback of enforcement, compared to counties with fewer TRI sites. We also find that these policy-induced increases in pollution are associated with a 38.8 percent increase in cases and a 19.1 percent increase in deaths from COVID-19.

* We are grateful to Unacast, who provided the restricted access Social Distancing Dataset. We are also grateful to Johns Hopkins University, the Environmental Protection Agency, the National Oceanic and Atmospheric Administration, and Julia Raifman for providing the publicly available data used in this analysis. We are grateful to the American University School of Public Affairs for their generous support of this research. We thank Jennifer Heissel for her helpful comments and suggestions and assistance gathering data. Any errors or conclusions are our own.

I. Introduction

The SARS-COV-2 coronavirus disease 2019 (COVID-19), which has spread throughout the United States at an alarming rate, represents a serious threat to public health and well-being. It is critically important to discover the factors that cause more cases and deaths from COVID-19, as well as why outcomes vary from place to place. COVID-19 (also knowns as the coronavirus) commonly manifests as a respiratory infection, and in severe cases, there is progressive respiratory failure leading to death (Xu et al, 2020). While a growing body of literature suggests that exposure to pollution can increase mortality and cause people to get sick with a respiratory illness (Currie et al. 2009; Deryugina et al. 2019; Jans, Johansson, and Nilsson 2014; Ransom and Pope 1992; Simeonova et al. 2019), little is known about the factors influencing how the COVID-19 virus spreads or whether pollution might be a factor in increasing the spread of the virus or deaths from COVID-19. However, recently Setti et al (2020) discovered that COVID-19's genetic material can be detected on particles of air pollution called Particulate Matter 10 (PM10). Another recent study finds that PM10 upregulates the receptor used by COVID-19 to infect host cells (Miyashita et al. 2020). This suggests that increased air pollution could increase infections and deaths from COVID-19.

On March 26, 2020, the United States Environmental Protection Agency (EPA) announced a freeze in civil enforcement of environmental regulations due to the coronavirus pandemic. In particular, the EPA stated that it does "not expect to seek penalties for violations of routine compliance monitoring, integrity testing, sampling, laboratory analysis, training, and reporting or certification obligations in situations where the EPA agrees that COVID-19 was the cause of the noncompliance" (EPA 2020).

Most of the facilities impacted by this rollback are likely to be required to report their emissions on the Toxic Release Inventory (TRI), a database maintained by the EPA on industrial or federal facilities that release toxic chemicals such as those commonly found in pollution from factories, power plants, mining, recycling and waste treatment, and other facilities. This allows us to track which US counties were most likely to see an increase in air pollution after the new policy was announced. In 2018, Toxic Release Inventory (TRI) sites alone (which represent only one type of industrial plant) released 3.85 billion pounds of (untreated) toxic chemicals in America into the air, land and water, out of 32.12 billion total pounds of toxic chemicals created in production-related wastes (EPA 2018). TRI sites are common: there are currently about 21,800 TRI sites operating across the United States and more than 221.5 million people (i.e., 2/3 of the U.S. population) had a TRI site operating in their zip code in 2016.[1]

We use EPA data on daily air quality and pollution sites to sort counties into two groups: those who are in the top third in terms of the total number of TRI sites operating in the county and those who are in the bottom two thirds. The top third of counties have 6 or more TRI sites with at least one that releases air pollution. To address concerns about the possible selection of TRI sites into counties, we limit our main sample to counties that have at least one TRI site emitting air pollution, so the comparison group includes counties with 1 to 5 TRI sits. Using the timing of these changes in short-term pollution exposure by county and data on COVID-19 cases and deaths from Johns Hopkins University, we employ a difference in differences design to estimate whether counties with more TRI sites experienced increased pollution because of the rollback of environmental regulations and see increases in COVID-19 deaths and cases, compared to counties that had fewer TRI sites. We control for social distancing, stay at home

[1] We made this calculation based on linking zip code level census counts of the population to TRI data.

orders, re-openings, days since the first COVID-19 death, weather, day of the week, cumulative confirmed cases (in the specification on deaths), and county, day of the week, and month fixed effects. We also employ an instrumental variables analysis using the policy-induced increase in pollution to adjust for the possible endogeneity of observed pollution levels and estimate the effects of predicted pollution on outcomes. To address concerns that population density or other factors might drive the findings, we also show the results with a variety of alternative comparison groups and estimate the effects of the number of TRI sites on cases and deaths nonparametrically.

This paper makes two important contributions. First, we find that that there is a large, sustained and statistically significant increase in air pollution after the rollback of environmental regulations in counties with more TRI sites. On average, counties with 6 or more TRI sites experience about 13 percent (i.e., 0.80 ug/m^3) higher Particulate Matter2.5 (PM2.5) pollution after the rollback, relative to counties in the bottom two thirds of the TRI distribution. This suggests that firms respond in the absence of regulatory incentives to increase pollution. Second, we find that increases in pollution resulting from the rollback of EPA enforcement led to large and statistically significant increases in COVID-19 cases and deaths. Counties with 6 or more TRI sites experienced a 19.1 percent increase in daily COVID-19 deaths and a 38.8 percent increase in daily confirmed COVID-19 cases after the rollback, compared to counties with 1 to 5 TRI sites. Both PM2.5 and ozone are associated with large increases in the virus's spread and deaths. In addition, we find that pollution exposure is worse for counties with a higher fraction of Black individuals, counties with higher unemployment, and counties that have a higher percentage of people in poverty. These results are robust to using weekly (rather than daily) cases and deaths, and a variety of alternative comparison groups and specifications.

This is the first paper to find that exposure to pollution worsens cases and deaths during a pandemic. This is also the first paper to document that in the response to the rollback of environmental enforcement, counties with more TRI sites saw increases in pollution, suggesting that firms respond in the absence of regulatory incentives by increasing pollution. Evidence of the extent to which air pollution affects cases and mortality from COVID-19 is important for three reasons. First, any changes in environmental policy should be informed by the costs associated with those changes, and the costs of deregulation right now could easily outweigh the benefits. Second, it informs our understanding of how pollution affects the transmission of viruses and the death toll during a pandemic, which could create opportunities for live-saving interventions. For example, air purifiers could be employed in in-patient facilities that treat COVID patients on particularly high air pollution days. Third, if reduction of pollution might assist to decrease the death rate, tangentially related interventions like stay at home orders could be modified to maximize the potential reduction in pollution.

II. Background

While some recent research suggests that pollution has decreased on average during the pandemic (Cicala et al 2020),[2] other studies find that pollution has actually increased in some areas or overall (Bekbulat et al 2020; NOAA, 2020; Schade 2020). For example, Bekbulat and colleagues (2020) find that PM2.5 concentrations are higher than expected across the United States based on long term seasonal trends. Schade (2020) finds the pollution has increased in some parts of Texas, such as Houston, which has many TRI sites. The NOAA (2020) reported that atmospheric carbon dioxide just reached the highest monthly reading ever recorded in May

[2] While Cicala and colleagues (2020) estimate that CO2 and PM2.5 emissions are projected to decline over this time period via estimates of electricity consumption and distance traveled, they also find substantial heterogeneity in their estimates and do not make use of pollution monitor data from 2020.

2020. Even though car pollution has likely decreased overall and electricity consumption may have decreased in some places (Cicala et al 2020), in counties with more TRI sites the total amount of average daily pollution as measured by EPA pollution monitors has increased. This makes sense, considering that pollution from ground transportation only makes up 20.6 percent of greenhouse gas emissions (Le Quere et al 2020) and that emissions globally are only projected to fall by 5.5 percent from 2019 levels because of COVID-19 (Evans 2020). The largest emitters of pollution come from electricity generation, industry and agriculture, jointly accounting for more than 58% of the total pollution emissions in the U.S. (EPA 2020).

Thus, while some measures of pollution may be lower in some places, this does not preclude the possibility of increases in pollution in other places. Firms may have an incentive to pollute more if reducing pollution is more costly (because, for example, monitoring pollution to be within acceptable limits requires staff hours) or the fines for exceeding standards are removed, or both. Recent survey evidence also suggests that the enforcement of environmental regulations from a traditional regulatory structure is the biggest motivator for many facilities' environmental compliance decisions (Gray and Shimshack 2011). Clay and Muller (2019) also find that pollution has increased 5.5% overall in the U.S. since 2016, which coincided with a decline in enforcement. This suggests that firms respond to regulatory incentives by changing the amount of pollution they release over time.

Little is known about the environmental factors affecting COVID-19 transmission, but there is some evidence that increased air pollution increases cases of respiratory illnesses. Jans, Johansson and Nilsson (2014) find that worsening air quality due to inversion episodes causes an increase in respiratory illnesses among children. Beatty and Shimshack (2011) find that retrofitting school buses was associated with reductions in bronchitis, asthma, and pneumonia incidence for at-risk populations. Simeonova et al. (2018) also show that the implementation of a congestion tax in Stockholm decreased the rate of acute asthma attacks among children. Research has shown that wind direction (Anderson, 2019; Deryugina et al., 2019) and airport delays (Schlenker and Walker, 2011) also can cause small changes in air pollution, which then impact mortality and hospitalizations. Overall, there is evidence that even small increases in pollution can have detrimental effects across a wide variety of health indicators and outcomes.

There is also growing evidence that days of high air pollution can cause deaths (Anderson, 2019; Schwartz Bind and Koutrakis 2017; Deryugina et al. 2019). Anderson finds that living downwind of a highway increases the mortality of persons over 75 years old. Anderson further finds that this increased mortality is from a range of different causes (Anderson, 2019). Deryugina et al (2019) find that vulnerable populations, such as the elderly or those with existing chronic conditions, are also more susceptible to the mortality increases that result from increases in pollution. Similarly, elderly people and people with existing chromic conditions have worse outcomes from COVID-19. There is also evidence that pollution can increase inflammatory cytokines, which have been implicated in deaths from COVID-19 infections due to “cytokine storms” (Tay et al 2020).

Recently Zhang et al (2020) identified airborne transmission in fine aerosols as the dominant route for the spread of COVID-19 and show that the outbreak in Wuhan, China corresponded with increased PM2.5 levels. Setti et al (2020) discovered that COVID-19’s genetic material can be detected on particles of air pollution called Particulate Matter 10. Another recent study finds that PM10 upregulates the receptor used by COVID-19 to infect host cells (Miyashita et al. 2020). Taken together, this suggests that increased air pollution could increase infections and deaths from COVID-19. However, there is currently no published causal research on this question.

III. Data

To examine how pollution affects deaths and cases from COVID-19, we use data compiled by the Johns Hopkins University Center for Systems Science and Engineering Coronavirus Resource Center (JHU) on the daily number of cases and deaths by county. While the data in its raw form is a set of cumulative totals of cases or deaths by county and day, we transform the data into the total new cases or new deaths reported each day. We match this data to daily data on air pollution by county from the EPA’s Air Quality System (AQS). We use data on the number of TRI sites by county from the EPA’s 2018 TRI Basic Data Files, which is the most recent year of TRI data. We also match these data to daily weather data from the National Oceanic and Atmospheric Administration (NOAA). The AQS has daily data on PM2.5, PM10 and ozone from January through May, 2020. We aggregate this to the county level by matching each pollution monitor to a zip code and imputing data to zip codes with missing monitors using the closest monitor and inverse distance weighting. We do not impute data from monitors more than 30 km away or for zip codes in counties without at least one monitor. Grainger and Schreiber (2019) suggest that monitors are often strategically positioned by local regulators to avoid pollution hotspots. Therefore, aggregating the pollution monitor data in this way is likely to lead to downwardly biased levels of pollution. Nevertheless, the pollution monitor data represents the best data available over a short time window.

We match additional daily data on cases and deaths in all five counties comprising New York City from the New York City Department of Health.[3] The JHU data adds county-level counts of daily COVID-19 cases and deaths starting on March 22, 2020, so the time period for the regressions on deaths and cases is limited to March 22nd through May 27th, 2020.

Because pollution emissions could be confounded by social distancing behaviors, we also use data on the degree of social distancing by county from Unacast's restricted access data, which they use to compute the Social Distancing Scoreboard. To construct this measure of social distancing, Unacast uses cell phone geolocation data on the average distance traveled from pre-COVID-19 days to estimate the percent change in total distance traveled in the four weeks before the pandemic, compared to each day during the pandemic starting on March 22, 2020 in each county. This social distancing measure is at the daily level by county across all counties in the United States from February 24, 2020 until the present.

We additionally match these data to data from the COVID-19 United States Policy Database on the exact timing of official stay at home orders and re-openings by state. We use the date businesses reopened in a state as the reopening date. We also match these data to county-level data on demographics and essential workers by industry sector from the 2018 Census and 2019 Bureau of Labor Statistics data from the Quarterly Census of Employment and Wages. Further details about the sample and weighting procedures are reported in the Data Appendix.

IV. Identification Strategies

Naïve correlations between air pollution and COVID-19 outcomes cannot be interpreted as causal because pollution is not randomly assigned. In order to disentangle the effects of

[3] Bronx, Richmond, Queens and Kings county were omitted from the John's Hopkin's data for unknown reasons, so we updated the data on all of New York City. The results are robust to the omission of this additional data.

pollution on COVID-19 outcomes from other county-specific factors that could influence COVID-related outcomes, we use two different identification strategies. First, we use a difference in differences design with county fixed effects to estimate whether being in a county with more TRI sites after the rollback led to an increase in pollution and whether increases in pollution led to increases in cases and deaths. Second, we estimate the effects of different types of pollutants on COVID-19 outcomes using an instrumental variables strategy in which we use predicted pollution increases after the rollback as an instrument for actual pollution. This allows us to avoid using potentially endogenous observational changes in pollution (for example, from measurement error) by documenting the relationship between an exogenous shock in pollution and COVID-related outcomes.

A. Comparability of Treated and Control Counties

Most of the facilities impacted by the environmental rollback are likely required to report their emissions on the Toxic Release Inventory (TRI), a database maintained by the EPA on industrial or federal facilities that release pollution over threshold amounts. The EPA's memo specifically mentions "reporting facilities", which include all TRI sites. This allows us to track which US counties were most likely to see an increase in air pollution after the new policy was announced. We use data from the 2018 TRI to sort counties into two groups: those who are in the top third in terms of the total number of TRI sites operating in the county and those who are in the bottom two thirds. In order to be considered in the top third of polluting counties, the county must have 6 or more TRI sites with at least one that releases air pollution. To address concerns about the possible selection of TRI sites into counties, we limit our main sample to counties with at least one TRI site emitting air pollution. However, the results are robust to including all counties in the United States that are represented in the Johns Hopkins data (i.e., 2777 counties), even counties with no TRI sites. The results are also robust to using other numbers of TRI sites as cutoffs and plotting the effects nonparametrically by the number of TRI sites. We discuss these specifications in section V.C. of the paper.

Table 1 presents the characteristics of all counties in the sample, as well as treatment and control counties separately. Overall Table 1 indicates that counties with 6 or more TRI sites are quite similar demographically to counties with 1 to 5 TRI sites in terms of the percentage of people who are essential workers,[4] White, Black, Hispanic, in poverty, or over 65, as well as the unemployment rate in 2018. However, counties with 1 to 5 TRI sites differ in terms of total population, the number of cases and deaths from COVID-19, and population density. Thus, we also show the characteristics of an additional control group we employ of counties with 1 to 5 TRI sites that also have population densities of more than 700 persons/mi. In addition, Table A1 shows that treated and control counties are widely distributed among nearly every state.[5]

B. Estimating Pollution Increases from the EPA's Rollback of Environmental Enforcement

Using daily data on air pollution by county from the EPA's Air Quality System (AQS) from January 1st through May 27th, 2020, we use a difference in differences design to estimate the amount pollution has increased because of the rollback. We regress the amount of pollution on an indicator for being in a treated county (with 6 or more TRI sites) after the rollback as follows:

$$(1) \qquad Pollution_{it} = \beta_1 TreatedPost_{it} + X_{it} + \sigma_i + \varphi_t + \varepsilon_{it}$$

In this equation, $Pollution_{it}$ is the daily amount of PM2.5 (or ozone) pollution in ug/m^3 (or ppm) in county *i* on day of the week *t*. $TreatedPost_{it}$ is a binary indicator for being in a

[4] See the data appendix for more on how we estimated the percentage of people by county who are likely to be essential workers.
[5] Dropping states without control counties does not change the results.

county in the top third of the distribution in terms of the number of TRI sites (with 6 or more TRI sites) after the rollback of environmental enforcement. X_{it} is a vector of daily county-level variables (i.e., whether there is a stay at home order on that day, state re-openings, and average temperature and precipitation). σ_i are state-county fixed effects, φ_t are day of the week fixed effects.[6] The amount that pollution increased within a county in the top third of the TRI distribution post-environmental regulation rollback relative to counties with fewer TRI sites is given by β_1. Standard errors are clustered at the county level. This equation is used to estimate the predicted amount of pollution in treated versus control counties that we use in our instrumental variables strategy described in the next section.

C. Estimating Cases and Deaths from the Policy-Induced Pollution Increase

While it is tempting to analyze this data cross-sectionally (for example, some studies compare counties with more versus less long-term pollution), long term pollution exposure might be associated with a variety of other characteristics of counties, such as social distancing proclivities, racial composition, employment levels, or income. Thus, there also may be selection into more polluted counties for people with worse underlying health or who practice different health behaviors related to social distancing.

Thus, our primary identification strategy is a difference in differences design in which we exploit the within-county change in pollution over time induced by the EPA's environmental rollback, controlling for county, month, and year fixed effects, as well as a variety of county-level demographic control variables and Unacast's measure of social distancing. The basic difference in differences model we will use is as follows:

[6] Note that there are no year fixed effects because the time period is constrained to only occur within the time window of the pandemic (from March 2020 onwards).

(2) $Y_{it} = \beta_1 TreatedPost_{it} + Post_t + X_{it} + \sigma_i + \varphi_t + \varepsilon_{it}$

In this equation, Y_{it} is the log of the number of daily deaths (or confirmed cases) in county *i* in time *t*. We apply the inverse hyperbolic sine (IHS) transformation to each daily count of deaths or cases to account for zeros in daily death or case values: asinh(Y_{it})= log($Y_{it} + (Y_{it}^2 + 1)^{0.5}$). The IHS transformation is approximately equal to log(2(Y_{it})), except for very small values, and can be interpreted in the same way as a logarithmic transformation (as a percent change). $TreatedPost_{it}$ is an indicator for being in a county with 6 or more TRI sites after the rollback of environmental regulation enforcement. $Post_t$ is a binary variable for being in the period after the EPA's rollback of civil enforcement, which took place on March 26th, 2020 for the entire United States. Note that a binary variable for being in the treatment group ($Treated_i$) will be omitted due to multicollinearity when using county fixed effects. X_{it} is a vector of daily county variables (i.e., daily average temperature and precipitation, whether there is a stay at home order, state re-openings, days since the first death from COVID-19, the number of confirmed cases, and daily social distancing measures). We only control for the daily number of confirmed cases by counties in regressions of the effect of environmental regulations on the log of deaths. σ_i are county fixed effects and φ_t are a vector of day of the week fixed effects and month fixed effects. Standard errors are clustered at the county level. The effect of being in a county in the top third of the TRI distribution post-environmental regulation rollback on cases or deaths is given by β_1.[7]

Counts of daily deaths from COVID-19 are likely to be less biased than counts of daily cases, since cases might be reported many days after people become sick. Thus, we think our estimates on daily cases are largely picking up COVID-19 cases that became worse because of

[7] The results are about twice as large when using county population weighting.

exposure to pollution. However, we estimate both cases and deaths using both daily and weekly measures, allowing for lags in both, later in the paper. By controlling for daily social distancing, the cumulative number of confirmed cases, and the number of days since the first death in a county, we are effectively controlling for time trends that could affect the spread and severity of cases of COVID-19. We also estimate model (2) with state fixed effects (rather than county fixed effects) and control for a variety of county level demographic variables directly (in addition to the daily controls mentioned in the previous paragraph): total population, population density, percent white, percent Black, percent Hispanic, poverty rate, the unemployment rate, median income, and percent of workers who are likely to be essential. Nevertheless, there are likely to be unobserved time invariant characteristics of counties that could leave them susceptible to worse cases of COVD-19, so our primary specification uses county fixed effects rather than state fixed effects.

Our main identifying assumption is that in the absence of the environmental rollback, outcomes in the treated counties (those with more TRI sites) would have followed a parallel trajectory to outcomes in counties in the control group. We provide evidence supporting the parallel trends assumption in several ways. First, we show that the characteristics of counties with more and fewer TRI sites are similar across a variety of different demographic characteristics in Table 1 (even if they are also different in some important ways). We also show that our main results are robust to county-specific linear time trends, as well as dropping counties with population densities of less than 700 persons/mile^2 in the control group so that the population density in control group counties are similar to those in treated counties. Third, in Section V.B. we show that these results are robust to dropping a variety of different types of counties or states that could be problematic. Fourth, we show that before the pandemic, treated and control counties show similar trends in monthly deaths overall (see Figure A2).

We also show an event study of the treatment and control counties in Figure 3 indicating that the counties are on similar trajectories in the pre-treatment period. The basic event study model we use is given by:

$$(3) \quad log\ Y_{it} = \beta_0 + \sum_{j=-2}^{6} \beta_j \mathbb{1}[\tau_{it} = j]_{it} + X_{it} + \eta_i + \theta_t + \epsilon_{it}$$

We include two days of lags and six days of leads for the treatment, where τ_{it} denotes the day relative to the rollback of the EPA's enforcement of environmental regulations. For example, a value of τ_{it} = -1 represents the deaths one day before the day the EPA released the memo saying it would not enforce environmental regulations (March 26, 2020). β is the effect of the environmental rollback on COVID-19 deaths. η_i are county fixed effects and θ_t are month and day of the week fixed effects. X_{it} is a vector of daily county variables defined above, and standard errors are again clustered at the county level.

Finally, we also estimate the effects of different types of pollution on outcomes using an instrumental variables strategy. As previously discussed, observed pollution might be endogenous due to the locations of pollution monitors and measurement error. Thus, we estimate the effects of PM2.5 and ozone using an instrumental variables strategy in which we use equation (1) to predict the policy-induced increase in pollution in treatment and control counties following the EPA's rollback of environmental enforcement. To quantify the relationship between pollution and COVID-19 deaths or cases associated with the environmental rollback we estimate systems of equations of the following form by 2SLS:

$$(1) \quad Pollution_{it} = \beta_1 TreatedPost_{it} + X_{it} + \sigma_i + \varphi_t + \varepsilon_{it}$$

$$(4) \quad Y_{it} = \beta_1 \widehat{Pollution}_{it} + \pi_{it} + \sigma_i + \varphi_t + \varepsilon_{it}$$

Our endogenous treatment variable is the amount of pollution in ug/m^3 or parts per million in county *i* at time *t* and the policy-induced predicted PM2.5 (or ozone) pollution after the rollback is the excluded instrument from the second stage. To only rely on variation within county, we use county fixed effects σ_i and day of the week fixed effects and month fixed effects φ_t. We also control for π_{it}, a vector of daily county variables (i.e., daily average temperature and precipitation, an indicator for being after the rollback, whether there is a stay at home order, state re-openings, days since the first death from COVID-19, the number of confirmed cases, and daily social distancing measures). Standard errors are clustered at the county level. We only control for the daily number of confirmed cases by counties in regressions of the effect of environmental regulations on the log of deaths. Our identifying assumption is that the only way increases in predicted pollution affects COVID-19 outcomes is through actual pollution increases from the environmental rollback. By using the exogenous policy-induced increase in pollution within a county, we are able to credibly estimate the effects of increased pollution on COVID-19 cases and deaths.

V. Results

A. Results on Pollution Increases

While some measures of pollution have been reported to be lower in some locations during the pandemic, aggregated analyses of pollution might mask important heterogeneity in pollution releases and exposure. Our analyses in Panel A of Table 2 and Figure 1 show that there is a statistically significant increase in air pollution after the rollback of enforcement of environmental regulations in counties with 6 or more TRI sites, relative to counties with 1 to 5 TRI sites. We find that counties in the top third of the TRI distribution in terms of the number of TRI sites experience about 13 percent (i.e., 0.80 ug/m^3) higher PM2.5, 5 percent higher ozone, and 14 percent higher PM10 pollution after the EPA's rollback, relative to counties in the bottom two thirds of the TRI distribution. This suggests that in the absence of regulatory incentives, firms may respond by releasing more pollution.

This increase in pollution is also unique to 2020. Panels B, C and D of Table 2 show results when running the same regressions for previous years using a March 26th cutoff in each year (and the same counties in the treated and control groups). We observe the opposite pattern of results for PM2.5, ozone, and PM10 in previous years: the coefficients in 2017, 2018 and 2019 are largely negative and not statistically significant (with one exception – ozone is positive in 2019). In addition, the event study of weekly PM2.5 in 2020 depicted in Figure 1 is a visual representation of our first stage. Figure 1 shows that pollution was lower in February and March until the week of the EPA's rollback, at which point pollution in counties with 6 or more TRI sites became higher over the next six weeks compared to pollution in control counties.[8] In contrast, the same event study in 2019 depicted in Figure A1 shows that PM2.5 pollution was largely falling over March and April, possibly due to the discontinuation of winter heating.[9]

In addition, Panel A of Figure 2 shows the association between average PM2.5 and the number of TRI sites emitting pollution by county overall in January through May of 2020. As the number of TRI sites increases, the total amount of pollution by county increases as well. There is substantial heterogeneity by county in the number of TRI sites and pollution levels, but the number of TRI sites is positively associated with total pollution by county on average. Panel B of Figure 2 plots the coefficients on the interaction between an indicated for being after the EPA's

[8] We describe our event study methodology in section IV.C. The event studies in Figures 1 and A1 differ from Figure 3 in that the outcome is weekly estimates of PM2.5 pollution, instead of daily estimates of log deaths. In addition, the event study in Figure 1 controls for only stay at home orders, state re-openings, weather and daily social distancing.
[9] Event studies in 2017 and 2018 show similar patterns of results, wherein pollution levels are largely falling over March and April in both treated and control counties.

rollback of civil enforcement ($POST_i$) with the stated bin for the number of TRI sites, controlling for stay at home orders, re-openings, temperature, precipitation and month, day of the week and county fixed effects. The omitted category is counties with 1 or 2 TRI sites. As the number of TRI sites increase in a county, so does the amount of pollution that increased after the rollback of enforcement. The overall pattern of results suggests that the EPA's rollback of environmental enforcement caused an increase in pollution in counties with more TRI sites that otherwise would not have occurred due to weather or seasonal patterns.

B. Main Results on COVID-19 Deaths and Cases

Table 3 presents the main results of our difference in differences models. Columns 1 and 5 present the results of our state fixed effects model with county-level demographic, economic, and daily controls.[10] However, even with this controls there might be other time-invariant characteristics of counties that could affect COVID-19 cases and deaths. Thus, columns 2 and 6 present the results from our preferred main specification using county fixed effects and daily controls (equation 2). Reassuringly, the results from our state fixed effects model with additional county-level controls and our county fixed effects model with daily controls are very similar – being in a treated county after the rollback led to a 19.1 percent increase in the daily death rate and a 38.8 percent increase in the daily case rate using the county fixed effects model, compared to a 16 percent increase in COVID-19 deaths and a 22.7 percent increase in cases using the state fixed effects model. While the means for daily deaths and cases are relatively low (i.e., 0.64

[10] These include controls for total population, population density, percent white, percent Black, percent Hispanic, poverty rate, the unemployment rate, median income, percent of workers who are likely to be essential, daily average temperature and precipitation, whether there is a stay at home order, state re-openings, days since the first death from COVID-19, the number of confirmed cases, daily social distancing measures at the county level, and month, state and day of the week fixed effects.

deaths/day), a 19 percent increase is still concerning. An increase of 38.8 percent in daily cases implies an additional 4.4 cases per county per day.

Columns 3 and 7 of Table 3 present results from our main specification adding a control for county-specific linear time trends to account for the fact that some places might have worse cases over time for reasons other than pollution. Because there is a concern that population density furthers the spread of the virus, columns 4 and 8 present the results of our main specification limiting the control group to more population-dense counties (with population density > 700 persons/mi^2). Note that this specification leaves less population dense counties in the treatment group.

Overall, we find substantial evidence that increases in pollution increased the conditional daily COVID-19 death rate and daily new case rate of COVID-19 – being in a county with 6 or more TRI sites after the rollback led to between a 13.6 and 19.1 percent increase in the daily death rate, compared to counties with 1 to 5 TRI sites. We also find that being in a treated county after the rollback is associated with between a 17.3 and 38.8 percent increase in daily cases. Because many people with COVID-19 do not show symptoms (Hu et al. 2020; Mizumotoet al. 2020), the rate of daily new cases likely represents more severe cases since tests are generally reserved for more severe cases (CDC 2020). Therefore, the observed increase in cases might indicate that pollution worsens existing COVID-19 cases or that pollution causes new cases. Because COVID-19 cases are likely reported with some delay and data on hospitalizations are only available at the state level, we are unfortunately unable to disambiguate fully between these two hypotheses. Nevertheless, the results are broadly similar across the different specifications. In addition, Table A2 shows that the results are very similar when we include all counties in the United States that are represented in the Johns Hopkins data (i.e., 2777 counties), even counties with no TRI sites, or estimate the results using Poisson pseudo-maximum likelihood (PPML) regressions.

Fig. 3 displays the main results of our event study of the timing of the environmental rollback on the log of COVID-19 deaths. While COVID-19 deaths were falling slightly in the treated counties in the pre-period, after the announced rollback of enforcement, deaths increased substantially in counties with more TRI sites relative to counties in the control group. Unfortunately, because the Johns Hopkins data begins on March 22 and the rollback is on March 26th, we cannot show more pre-period observations.[11] Although the data unfortunately does not exist to show a longer pre-period for COVID-19 cases and deaths for the treatment and control groups, we can use data from the CDC Wonder database on all deaths by month in 2018 to get a sense of whether deaths followed similar trajectories in the treatment and control groups.[12] The log of average deaths per month for the same treatment and control groups are shown in Figure A2. While the total number of deaths per month differs, the treatment and control groups show very similar trends over time in total deaths per month. Taken together, these figures suggest that the environmental rollback led to a large increase in the death rate above the treatment group that might not have occurred in the absence of the rollback.[13]

Lastly, we address the concern that data on daily cases and deaths could suffer from serial correlation, errors in reporting or delays before the onset of symptoms and testing by showing our main results using weekly estimates instead of daily estimates in Table 4. Table 4 depicts the results of being in a treated county after the rollback using weekly estimates of the total deaths

[11] We omit March 22nd in the event study because we have concerns that the data for the first day of the county-level data was aggregated from all previous days since the JHU data is a rolling cumulative total of deaths or cases on each day. The observations on March 22 are several times higher than on any later day in the next two weeks in each county. The results are robust to the inclusion of this date, however.
[12] 2018 is the most recent year of county-level data by month.
[13] In addition, we estimate a similar event study using the daily COVID-19 death rate per 10,000 individuals (dropping counties with less than 10,000) and find very similar results.

and cases in the same week, and deaths and cases one week later, two weeks later, and three weeks later. The estimates using weekly variation are nearly three times the size of the estimates using daily variation – we observe a 61.5 percent increase in weekly deaths and a 54 percent increase in weekly cases. The largest effects on cases and deaths occur in the same week, followed by one week later, and then two weeks later, with the smallest effects three weeks later. This is in line with a literature that suggests that it takes a few days to develop symptoms after a COVID-19 infection, and the time to taking a test might vary by the availability of the tests in a location and the severity of the infection. Some tests might occur right away, while others might occur only once someone is hospitalized after several weeks of being sick. In addition, people might become sicker with COVID-19 because of exposure to pollution but die a few weeks later after symptoms progress. The pattern of results suggests that pollution has both contemporaneous and delayed negative effects on COVID-19 cases and deaths. However, the largest results are contemporaneous, suggesting that pollution may cause existing COVID cases to worsen.

C. Effects by Pollution Type and Subgroup

Next, we turn to our analysis of COVID outcomes by pollution type. Table 5 presents the results for our instrumental variables strategy in which we estimate the effects of different types of pollutants using the predicted pollution after the rollback as our excluded instrument. Table 2 shows the first stage of these regressions. Overall, PM2.5 and ozone both have large detrimental effects on COVID-19 deaths and cases, suggesting that they might exacerbate respiratory distress and cause COVID-19 cases to worsen. A one $\mu g/m^3$ increase in PM2.5 is associated with a 68.6 percent increase in deaths. On average, counties experienced a 0.8 increase in PM2.5 after the rollback, implying a 54.9 percent increase in deaths and an 82 percent increase in cases.

Similarly, a 0.0021 ppm increase in ozone is associated with a 50 percent increase in deaths and a 73.8 percent increase in cases (coefficients are again scaled by the increases in Table 2 Panel A). We are unfortunately only able to estimate our instrumental variables strategy in counties that have pollution monitors, and there were unfortunately too few observations to reliably estimate the effects of PM10. Nevertheless, PM10 and PM2.5 are correlated in our sample (with a correlation coefficient of 0.51), so the results for PM2.5 likely reflect the effects of other types of pollution as well.

In Table 6, we also examine the results by different characteristics of counties in the 2018 American Community Survey. We find that counties with higher than median percentages of Black individuals have much worse outcomes as a result of pollution exposure after the rollback. There is some evidence that non-White individuals are overrepresented in neighborhoods around TRI sites (Currie et al. 2015), which might explain the higher death rate for counties with more Black individuals. There is also suggestive evidence that Black Americans are dying at higher rates than Whites (Garg et al. 2020), and our results suggest that pollution might be a driving factor in this.

In addition, the outcomes are worse in counties that had above-median unemployment rates according to the 2018 census. If industrial sites lower housing values, the lower income individuals might be more likely to live closer to TRI sites, resulting in a higher death rate. In addition, lower income individuals might have less access to healthcare, which could negatively affect COVID-19 outcomes. However, counties with below-median poverty rates are similarly affected to counties with higher poverty rates, so it is unclear whether income is the main driver of these findings. Surprisingly, the results are also larger for counties with a lower than median percentage of people over the age of 65, indicating that pollution might be just as harmful for younger individuals.[14]

D. Additional Robustness Checks

If our results are driven by TRI pollution, we would expect COVID-19 deaths and cases to be worse in counties with more TRI sites. Figure 4 presents results in which we estimate our main model nonparametrically by interacting our indicator for of being after the rollback ($POST_i$) with five different bins for the numbers of TRI sites counties have (3 TRI sites, 4 to 6 TRI sites, 7 to 12 TRI sites, and 13 or more TRI sites). The omitted category is counties with 1 or 2 TRI sites. We include the indicator for being after the rollback as a control, as well as our daily controls. The estimates indicate that the effect of living in a county with up to 3 TRI sites on COVID-19 outcomes is near zero. However, as the number of TRI sites increases, the effects become much larger. Overall, the main effects appear to be driven by counties with 7 or more TRI sites. Each additional TRI site a county has increases deaths from COVID-19 by 1.6%.

While our preferred model is a difference in differences design with county fixed effects, we have also estimated the effects of pollution exposure using several alternative approaches and samples. One concern is that treatment and control counties might differ in the percentages of essential workers as a fraction of the labor force, which might affect the results. Ideally, we would like to control for the percentage of essential workers by county as we do in our state fixed effects model, but unfortunately, there is no monthly within-county variation in the data on the percentages of these workers. Instead we use data from the Bureau of Labor Statistics' 2019 Census of Employment and Wages by type of industry and guidelines from the state of Massachusetts on which industries are deemed essential to limit the counties in the sample to

[14] Unfortunately data on the age of residents by county is only available from the Census for a subset of counties.

counties with a similar estimated percentage of essential workers.[15] The following industries are likely to contain significant numbers of essential workers based on guidelines released by the state of Massachusetts: agriculture, construction, manufacturing, utilities, education, health, social assistance, and public administration (local and federal government). Non-essential workers are roughly those in wholesale, retail, information, finance, professional and scientific, arts and entertainment, mining, and other services. Unfortunately, the data do not allow us to disaggregate the categories more.[16]

While the overall distribution of essential workers is roughly similar in treatment and control counties, the results in column 1 of Table 7 show the results when limiting the counties in the sample to counties with 50-70% of essential workers (as a percentage of the total workforce). These results trim the lowest 10 percent and the highest 25 percent of counties in terms of the estimated percentages of essential workers. In general, larger counties have lower percentages of essential workers, so this analysis effectively removes especially large or small counties. The results are slightly larger when limiting to counties with the same fraction of jobs that are likely to be essential by county, indicating that differences in the fraction of essential workers does not drive the main results.

Another concern is that especially population dense counties in the treatment group (with 6 or more TRI sites) might also be driving the results. So, in column 2 of Table 7 we also show that our results are robust to dropping treated counties with population density over 1400 persons/mi^2. This analysis effectively drops all treated counties in the top 20th percentile of the distribution in terms of population density, including all treated counties in New York City.

[15] We have replicated these results using 2018 Census of employment categories instead of BLS categories and the results are very similar despite some differences in the categories.
[16] Additional details about how we applied the Massachusetts guidelines on essential workers to the BLS data are available in the data appendix.

However, the control group retains counties with high population density (above 1400), including 3 counties in New York city that had large numbers of cases and deaths (i.e., Manhattan, the Bronx and Staten Island). This is a strong test of the identification strategy because retaining dense control counties and dropping dense treatment counties should bias the analysis toward finding no result. As expected, the results are smaller, yet we observe that being in a treated county after the rollback is associated with an 8.4 percent increase in daily deaths from COVID, which is statistically significant at the 5% level.

To test whether especially large or small counties in general are driving the results, we limit to counties with more than 10,000 people and less than 1.64 million people in column 4. This trims the bottom 10 percent of the control group and the top 1 percent of the treatment group so that treatment and control counties have similar distributions. The results are again similar to our main specification in column 1. Because New York City is an outlier in terms of the number of cases and deaths, Column 5 also presents the results when dropping the five counties comprising New York City. It is worth noting that New York County, Bronx County and Richmond County are in the control group in all regressions since they contain fewer than 6 TRI sites, while Queens and Kings county are in the treatment group. The results are unchanged.

We also estimate the effects on COVID-19 cases and deaths using the daily rate of deaths or cases per 10,000 individuals as the outcome instead of the log of daily deaths and cases. The results, presented in Table A3, are similar to those in our main specification. Being in a treated county after the rollback leads to a 27.9% increase in the daily death rate and an 15.6% increase in the daily case rate (per 10,000) above the mean.

It is also possible that smoke from seasonal agricultural burning in Mexico might increase the overall levels of pollution in certain states. The prevailing winds in late April and

May were potentially blowing the smoke northeast according to Figure A3 (The World Resources Institute 2020). Thus, column 6 of Table 7 presents the results when we drop all counties in Texas, Louisiana and Florida. The results are unchanged, suggesting that smoke from seasonal agricultural burning does not drive the effects of pollution on COVID-19 cases and deaths. To test whether states without control counties affect the results, column 7 presents the results when dropping states without control counties, including Massachusetts, Rhode Island, Delaware and Washington DC. The results are similar when dropping these states.

Another concern is that some TRI sites are not expected to be operating during the pandemic. Thus, column 8 presents the results when estimating only on a sample of TRI sites that have industry codes indicating that they are likely to be considered essential according to documentation from a variety of states, including Massachusetts. Such essential industries include oil, gas and coal, food and beverage, chemicals, computers, electrical equipment, hazardous waste management, machinery, wood/paper, plastics and rubber, and transportation. These constitute more than half of all TRI sites and 97 percent of the counties in the sample has at least one essential TRI site. As expected, the results when limiting to these essential TRIs are slightly larger than to the main results (but likely include some error), indicating that these TRI sites are likely to be driving the effects. The results for daily cases are similar across all specifications in Table 7 as well.

A final concern is that daily cases and deaths could suffer from serial correlation bias. Thus, Panel C of Table A2 depicts the daily results using Hausman-Taylor correlated random effects models to account for possible serial correlation. The results are all similar in magnitude to those in Table 3 and statistically significant at the $p<0.01$ level. This suggests that the results are not being driven by serial correlation bias.

VI. Discussion

Whether pollution increases the COVID-19 case rate or death rate is an extremely important question for public health, and there is a race to discover the factors that cause more deaths. This is the first paper to document that in the response to the rollback of environmental enforcement, counties with more TRI sites saw increases in pollution, suggesting that firms respond in the absence of regulatory incentives by increasing pollution. Our results show that increased pollution increases the conditional daily COVID-19 death rate by 19.1 percent and the case rate by 38.8 percent. These results are stronger for counties with higher fractions of Black individuals and unemployed individuals, suggesting that the burden of pollution exposure is unequal. Pollution might have the largest impacts on the most vulnerable members of society, causing higher death rates and more severe cases of COVID-19.

This study also suggests that deregulation efforts may come with very high costs in terms of human lives during pandemics. Using the increase in the daily death rate, our back of the envelope calculation suggests that the environmental rollback led to 10,621 additional deaths from COVID-19.[17] Our results are consistent with a broader literature that finds that pollution increases respiratory infections and mortality. However, these findings suggest that unequal pollution exposure might exacerbate preexisting inequalities in health and result in more COVID-19 deaths. This work also underscores the importance of continuing to enforce existing regulations during pandemics.

Finally, this work also suggests several opportunities for intervention. For example, Castres and colleagues (2017) find that more than half of London's National Health Service

[17] We use the increase in the daily death rate of 0.0064 deaths / 10,000 people multiplied by the 63 days in the post-rollback period to get 0.00004032. We multiply this by the number of people living in treated counties (263,420,480) to get 10,621 additional deaths.

(NHS) facilities (including hospitals and clinics) had measured indoor air pollution that exceeded legal limits. While air filters are usually employed in intensive care units and operating rooms in the United States, it is unclear the extent to which all hospitals and clinics in the United States use HEPA air filters in all areas to eliminate indoor air pollution.[18] Air purifiers could be employed in all facilities that treat COVID-19 patients and for patients at home on high air pollution days. In addition, targeted policy and regulatory efforts to reduce pollution might assist to decrease the death rate. Our pattern of results further suggest that preventative measures should be focused on vulnerable populations, who are more at risk after exposure to pollution. Further research is needed to understand the mechanisms by which reducing pollution might affect COVID-19 cases and deaths.

References

Anderson, M.L. 2019. "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality." *Journal of European Economic Association*, 0(0): 1–42.

Beatty, K. M. T., & Shimshack, J. P. 2011. "School Buses, Diesel Emissions, and Respiratory Health." *Journal of Health Economics* 30, 987–999.

Bekbulat, B., Apte, J. S., Millet, D. B., Robinson, A. L., Wells, K. C., & Marshall, J. D. 2020. "$PM_{2.5}$ and Ozone Air Pollutino Levels Have not Dropped Consistently Across the US Following Societal COVID Response." ChemRxiv, Article 3299fafedd485e00c885.

Castres, P., Dajnak, D., Lott, M., & Watts, N. 2017. "Most London Hospitals and Clinics Exceed Air Pollution Limits." *BMJ* 357:j2855.

Center for Disease Control and Prevention. 2020, June 12. "Testing for COVID-19." https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html

Cicala, S., Holland, S. P., Mansure, E. T., Muller, N. Z., & Yates, A. J. 2020. "Expected Health Effects of Reduced Air Pollution from COVID-19 Social Distancing." Working Paper 27135, National Bureau of Economic Research.

[18] We were unable to find a single study conducted on this question in the United States, though studies in the United Kingdom, China and Taiwan all find that indoor air pollution exists in hospitals and varies with the type of ventilation system used (Chien-Cheng et al 2015).

Clay, K. & Muller, N. Z. 2019. "Recent Increases in Air Pollution: Evidence and Implications for Mortality." Working Paper 26381, National Bureau of Economic Research.

Currie, J., Hanushek, E.A., Kahn, E.M., Neidell, M., & Rivkin, S.G. 2009. "Does Pollution Increase School Absences?" *Review of Econonmic Statistics*, 91(4): 682–694.

Currie, J., Davis, L., Greenstone, M., & Walker, R. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review* 105(2):678–709.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." Working Paper 22796, National Bureau of Economic Research.

EPA (Environmental Protection Agency). 2020. "COVID-19 Implications for EPA's Enforcement and Compliance Assurance Program."

EPA (Environmental Protection Agency). 2020. "Greenhouse Gas Inventory Data Explorer."

EPA (Environmental Protection Agency). 2018. "2018 TRI National Analysis."

Evans, S. 2020, April 9. "Analysis: Coronavirus Set to Cause Largest Ever Annual Fall in CO2 Emissions." *Carbon Brief*. https://www.carbonbrief.org/analysis-coronavirus-set-to-cause-largest-ever-annual-fall-in-co2-emissions

Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., …, & Fry, A. 2020. "Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019—COVID-NET, 14 States, March 1-30, 2020." *MMWR Morbity Mortality Weekly Report* 69, 458–464.

Grainger, C. & Schreiber, A. 2019. "Discriminations in Ambient Air Pollution Monitoring?" *AEA Papers and Proceedings* 109, 277–82.

Gray, W. B. & Shimshack, J. P. 2011. "The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence." *Review of Environmental Economics and Policy*, 5(1), 3–24.

Hu, Z., Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., Ma, H., Chen, W., Lin, Y., Zheng, Y., Wang, J., Hu, Z., Yi, Y., Shen, H. 2020. "Clinical Characteristics of 24 Asymptomatic Infections with COVID-19 Screened Among Close Contacts in Nanjing, China." *Science China Life Sciences* 63, 706-711.

Jans, J., Johansson, P., & Nilsson, J. P. 2014. "Economic Status, Air Quality, and Child Health: Evidence from Inversion Episodes." Discussion Paper 7929, IZA Institute of Labor Economics.

Le Quere, C., Jackson, R. B., Jones, M. W., Smith, A. J. P., Abernethy, S., …& Peters, G. P. 2020. "Temporary Reduction in Daily Global CO_2 Emissions During the COVID-19 Forced Confinement." *Nature Climate Change*.

Miyashita, L., Foley, G., Semple S., & Grigg, J. 2020. "Traffic-Derived Particulate Matter and Angiotensin-Converting Enzyme 2 Expression in Human Airway Epithelial Cells." Working Paper, bioRxiv. https://doi.org/10.1101/2020.05.15.097501

Mizumoto, K., Kagaya, K., Zarebski, A., & Chowell, G. 2020. "Estimating the Asymptomatic Proportion of Coronavirus Disease 2019 (COVID-19) Cases on Board the Diamond Princess Cruise Ship, Yokohama, Japan, 2020." *European Surveillance* 25(10), Article 2000180.

NOAA Research. June 6, 2020. "Rise of Carbon Dioxide Unabated: Seasonal Peak Reaches 417 Parts per Million at Mauna Loa Observatory." *NOAA Research News.*

Ransom, M. R., & Pope III, C. A. 1992. "Elementary School Absences and PM10 Pollution in Utah Valley." *Environmental Research* 58, 204–219.

Schade, G. W. 2020. "Houston Air Quality Assessment in Response to Coronavirus Social Distancing Measures." [Unpublished manuscript]. Department of Atmospheric Sciences, Texas A&M University.

Schlenker, W. & Walker, W. R. 2011. "Airports, Air Pollution, and Contemporaneous Health." Working Paper 17684, National Bureau of Economic Research.

Schwartz, J., Bind, M.-A., & Koutrakis, P. 2017. "Estimating Causal Effects of Local Air Pollution on Daily Deaths: Effect of Low Levels." *Environmental Health Perspectives* 125(1), 23–29.

Setti, L., Passarini, F., Gennaro, G.D., Barbieri, P., Perrone, M. G., Piazzalunga, A., Borelli, M., Palmisani, J., Di Gilio, A., Piscitelli, P, Miani, A. 2020. "The Potential Role of Particulate Matter in the Spreading of COVID-19 in Northern Italy: First Evidence-Based Research Hypotheses."medRxiv. https://doi.org/10.1101/2020.04.11.20061713.

Simeonova, E., Currie, J., Nilsson, P., & Walker, R. 2019. "Congestion Pricing, Air Pollution, and Children's Health." Working Paper 24410, National Bureau of Economic Research.

Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A., & Ng, L. F. P. 2020. "The Trinity of COVID-19: Immunity, Inflammation and Intervention." *Nature Reviews: Immunology*.

The World Resources Institute 2020, May 18. "Global Fires Watch." https://fires.globalforestwatch.org/home..

Unacast. 2020. Unacast Social Distancing Dataset. https://www.unacast.com/data-for-good. Version from 27 April 2020.

Xu, Z. Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., Tai, Y., Bai, C., Gao, T., Song, J., Xia, P., Dong, J., Zhao, J., Wang, F.-S. 2020. "Pathological Findings of COVID-19 Associated with Acute Respiratory Distress Syndrome." *Lancet Respiratory Medicine* 8: 420–422.

Zhang, R., Li, Y., Zhang, A. L., Wang, Y., & Molina, M. 2020. "Identifying Airborne Transmission as the Dominant Route for the Spread of COVID-19." *PNAS,* Article2009637117.

Table 1: Descriptive Statistics of Counties By Number of TRI sites

	(1) Characteristics of Counties in the U.S. in 2018 with 1 or More TRI sites	(2) Characteristics of Counties with 6 or More TRI sites	(3) Characteristics of Counties with 1 to 5 TRI sites	(4) Characteristics of Counties with 1 to 5 TRI sites, Limited to Population Density of >700
Total Population	159,964 [421,586]	290,830 [576,284]	39,153 [70,675]	208,598 [309,370]
Population Density	624.2 [2,315]	1,043 [2,040]	237.9 [2,481]	4,541 [13,249]
Percent Essential Workers	0.5408 [0.06173]	0.5414 [0.05946]	0.5382 [0.07288]	0.4890 [0.07426]
Percent White	0.830 [0.154]	0.805 [0.149]	0.853 [0.155]	0.799 [0.159]
Percent Black	0.0911 [0.137]	0.105 [0.130]	0.0780 [0.142]	0.108 [0.129]
Percent Hispanic	0.0963 [0.130]	0.108 [0.125]	0.0858 [0.133]	0.118 [0.172]
Percent With Less Than a High School Degree	0.203 [0.0922]	0.193 [0.0771]	0.212 [0.103]	0.165 [0.0930]
Percent Poverty	0.107 [0.0466]	0.101 [0.0406]	0.112 [0.0510]	0.0989 [0.0474]
Median Income	53,545 [13,645]	57,422 [14,470]	49,967 [11,759]	60,508 [18,162]
Unemployment Rate	0.03319 [0.01136]	0.03466 [0.00963]	0.03184 [0.01261]	0.03497 [0.01147]
Percent Over 65	0.1675 [0.04341]	0.1653 [0.04071]	0.1787 [0.05433]	0.1912 [0.07710]
Total TRI Sites	10.76 [20.55]	19.48 [27.06]	2.703 [1.335]	3.742 [1.154]
Total Confirmed Cases	808.9 [3,615]	1,541 [5,037]	133.5 [880.8]	1,496 [4,568]
Total Deaths	44.29 [225.4]	84.64 [310.2]	7.038 [77.50]	110.0 [413.2]
Number of Counties	1,777	853	924	31

Notes: This table shows the average characteristics of counties in the sample with standard deviations in brackets below each mean. Column 1 shows characteristics of all counties in the United States with at least one TRI site releasing air pollution. Column 2 shows characteristics of treated counties (with more than 6 TRI sites). Column 3 shows characteristics of control counties (with 1 to 5 TRI sites). Column 4 shows characteristics of control counties (with 1 to 5 TRI sites) limited to those with population density of more than 700 persons/mi.

Table 2: Difference in Differences Results for Being in County with 6 or more TRI sites on Pollution Levels After the EPA's Rollback of Enforcement Compared with Placebo Years

	(1) Daily Mean PM2.5 Concentration	(2) Daily Mean Ozone Concentration	(3) Daily Mean PM10 Concentration
		Panel A: 2020 (Treatment Year)	
Treated County Post Rollback (March 26, 2020)	0.7960*** (0.1619)	0.0021*** (0.0002)	2.2131*** (0.7107)
		Panel B: 2019 (Placebo Year)	
Treated County Post March 26, 2019	-0.3249* (0.1964)	0.0017*** (0.0002)	0.5188 (0.5920)
		Panel C: 2018 (Placebo Year)	
Treated County Post March 26, 2018	0.2602 (0.1654)	-0.0008*** (0.0003)	-0.8180 (0.7210)
		Panel D: 2017 (Placebo Year)	
Treated County Post March 26, 2017	-0.0185 (0.2378)	-0.0005 (0.0003)	-0.8370 (0.7523)
Mean of Dependent Variable	6.035	0.0424	14.902
Observations	60280	74931	27122

Notes: This table shows the effects of being in a county with 6 or more TRI sites after March 26^{th} in different years, compared to before March 26^{th}. Panel A presents the results of being in a county with 6 or more TRI sites after the rollback of environmental enforcement on March 26^{th}, 2020, compared to being in a county with 1 to 5 TRI sites. Panels B, C and D present the results of a series of placebo tests using other years. We use the same difference in differences specification in equation (1) and regress PM2.5 levels on an indicator for being in a county with 6 or more TRI sites after March 26^{th} through the end of May in each of the years indicated. Notably, pollution only increases on March 26^{th} in the year of the environmental rollback (2020). All models control for temperature, precipitation, month, day of the week fixed effects and county fixed effects. The models in Panel A additionally control for stay at home orders and reopenings. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 3: The Effects of Pollution on Deaths and Cases

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log COVID-19 Deaths	Log COVID-19 Deaths	Log COVID-19 Deaths	Log COVID-19 Deaths	Log Confirmed COVID-19 Cases	Log Confirmed COVID-19 Cases	Log Confirmed COVID-19 Cases	Log Confirmed COVID-19 Cases
Treated Counties in Post Period	0.1599***	0.1910***	0.1361***	0.1417**	0.2268***	0.3882***	0.2129***	0.1726
	(0.0146)	(0.0148)	(0.0137)	(0.0583)	(0.0329)	(0.0310)	(0.0298)	(0.1605)
With State Fixed Effects and controls	X				X			
With County Fixed Effects and daily controls		X				X		
With County Fixed Effects and County-Specific Linear Time Trends			X				X	
Limited to Counties with Population Density >700 in the Control Group				X				X
Mean of the Dependent Variable	0.194	0.194	0.194	0.194	1.148	1.148	1.148	1.148
County-Day Observations	105260	105618	105618	53771	105258	105618	105618	53771

Notes: Columns 1-4 present the results for different regression specifications with the log of COVID-19 deaths as the outcome. Columns 4-6 present the same specifications with the log of confirmed COVID-19 cases as the outcome. Columns 1 and 5 show the results for the whole sample of counties with 1 or more TRI sites using state fixed effects, controlling for total population, population density, percent white, percent Black, percent Hispanic, poverty rate, the unemployment rate, median income, and the percent of workers who are likely to be essential. Columns 2 and 6 show the results with county fixed effects. Columns 3 and 7 show the results with. county-specific linear time trends. Columns 4 and 8 show the results when limiting the control group to counties with a population density of more than 700 persons/mi. All models also control for an indicator for being after the EPA's rollback, social distancing, stay at home orders, re-openings, days since the first COVID death, weather, and day of the week, county and month fixed effects. Columns 1-4 additionally control for daily confirmed COVID-19 cases. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 4: Results for Weekly COVID-19 Death and Cases in the Same Week and Allowing for a Delay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log Weekly COVID-19 Deaths (same week)	Log Weekly COVID-19 Deaths 1 Week Later	Log Weekly COVID-19 Deaths 2 Weeks Later	Log Weekly COVID-19 Deaths 3 Weeks Later	Log Weekly COVID-19 Cases (same week)	Log Weekly COVID-19 Cases 1 Week Later	Log Weekly COVID-19 Cases 2 Weeks Later	Log Weekly COVID-19 Cases 3 Weeks Later
Treated Counties in Post Period	0.6151*** (0.0389)	0.5226*** (0.0386)	0.2962*** (0.0357)	0.1486*** (0.0345)	0.5407*** (0.0632)	0.2952*** (0.0647)	0.2072*** (0.0598)	0.1943*** (0.0560)
Whole Sample with 1 or More TRIs, using County and Month Fixed Effects	X	X	X	X	X	X	X	X
County-Week Observations	15638	15638	14165	12634	15638	15638	14165	12634

Notes: Columns 1-4 present the results for the effects of being in a treated county after the rollback with the log of weekly COVID-19 cases as the outcome. Columns 1 and 5 presents the effects on confirmed deaths or cases in the same week. one week later, Columns 2 and 6 present the effects one week later, Column 3 and 7 present the effects 2 weeks later, and Columns 4 and 8 present the effects 3 weeks later. All models control for an indicator for being after the EPA's rollback, social distancing, stay at home orders, re-openings, days since the first COVID death, weather, and county and month fixed effects. Columns 1-4 additionally control for daily confirmed COVID-19 cases. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 5: Instrumental Variables Results for the Effects of Different Types of Predicted Pollution on Deaths and Cases

	(1)	(2)	(3)	(4)
	Log COVID-19 Deaths	Log COVID-19 Deaths	Log Confirmed COVID-19 Cases	Log Confirmed COVID-19 Cases
Predicted PM2.5 Pollution (μg/m^3)	0.6863* (0.3641)	—	1.0224** (0.5007)	—
Predicted Ozone (ppm)	—	238.0023** (115.6981)	—	348.2819** (164.3792)
Sample of Counties with Pollution monitors using 2SLS regression	X	X	X	X
Observations	30594	36238	30594	36238

Notes: Columns 1 and 2 present the effects of different pollution types on the log of COVID-19 deaths. Columns 3 and 4 present the same specifications with the log of confirmed COVID-19 cases as the outcome. Columns 1 and 3 present results in which predicted PM2.5 pollution is the excluded instrument from the second stage. Columns 2 and 4 present the results in which predicted ozone is the excluded instrument from the second stage. All models use 2SLS regressions with predicted pollution after the rollback used as an instrument for actual pollution. All regressions control for an indicator for being after the EPA's rollback, social distancing, stay at home orders, re-openings, days since the first COVID death, weather, and day of the week, county and month fixed effects. Columns 1 and 2 additionally control for daily confirmed COVID-19 cases. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 6: Heterogeneity by County Characteristics in 2018

	Log COVID-19 Deaths							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	County is Below Median Percent Black	County is Above Median Percent Black	County is Below Median Percent Unemployed	County is Above Median Percent Unemployed	County is Below Median Percent Poverty	County is Above Median Percent Poverty	County is Below Median Percent Over 65	County is Above Median Percent Over 65
Treated Counties in Post Period	0.0441*** (0.0114)	0.2611*** (0.0229)	0.1305*** (0.0169)	0.2405*** (0.0231)	0.2000*** (0.0198)	0.1658*** (0.0204)	0.2329*** (0.0452)	0.1682*** (0.0377)
Observations	56789	48829	57270	48348	57095	48523	21958	23347
Average of dependent variable	0.10	0.10	0.033	0.033	53,959	53,959	0.167	0.167

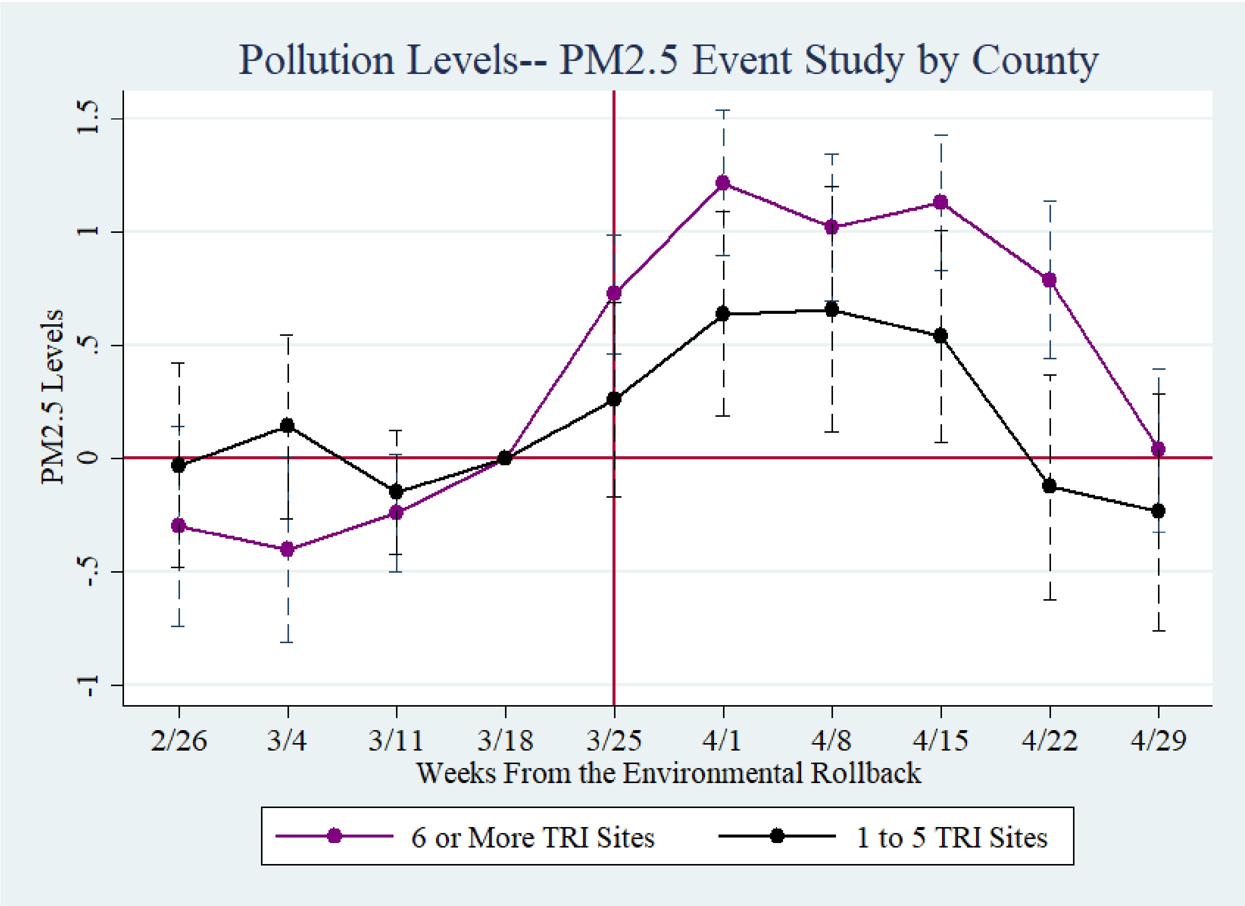
Notes: Each column presents the results for a different subgroup with the log of COVID-19 deaths as the outcome. All models control for an indicator for being after the EPA's rollback, social distancing, stay at home orders, re-openings, days since the first COVID death, weather, daily confirmed COVID-19 cases, and day of the week, county and month fixed effects. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table 7: Additional Robustness and Validity Tests for the Log of Deaths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log COVID-19 Deaths							
	Baseline Model	Limiting to Comparable Counties with 50% to 70% Essential Workers	Limited to Counties with Population Density <1400 in the Treatment Group	Limiting to Counties with Population 10,000 - 1.6 million	Dropping New York City Counties	Dropping States Near the Mexican Border With Possible Smoke Exposure	Limiting to Counties in States with Both Treated and Control Counties	Limiting to Only Essential TRIs
Treated Counties in Post Period	0.1910***	0.2636***	0.0843***	0.1695***	0.1659***	0.1746***	0.1587***	0.2131***
	(0.0148)	(0.0225)	(0.0110)	(0.0143)	(0.0140)	(0.0156)	(0.0140)	(0.0197)
With County, Month, and Day of Week Fixed Effects	X	X	X	X	X	X	X	X
Observations	105198	95621	93757	96853	105488	93858	104444	105618

Notes: Columns 1-8 present the results for being in a county with 6 or more TRI sites after the EPA's rollback with the log of COVID-19 deaths as the outcome. Column 1 replicates our results from Table 3. Column 2 presents the results when limiting to counties with similar percentages of essential workers. Column 3 presents estimates in which we drop treated counties with population densities of more than 1400. Column 4 presents results when limiting to counties with population between 10,000 and 1,600,000. The results in Column 5 drop all five counties in New York City, and column 6 drops states near the Mexican border. Column 7 limits to counties in states with both treated and control counties. Column 8 estimates the results using only essential TRI sites. All models control for being after the rollback, social distancing, stay at home orders, re-openings, days since the first COVID death, daily number of confirmed COVID-19 cases, weather, and day of the week, county and month fixed effects. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

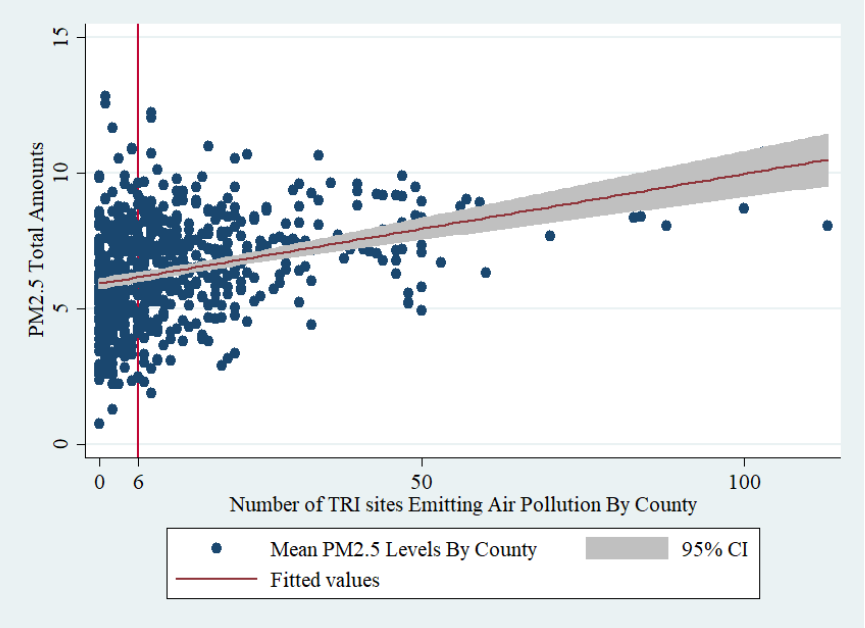
Figure 1: Event Study of Weekly Particulate Matter2.5 by County



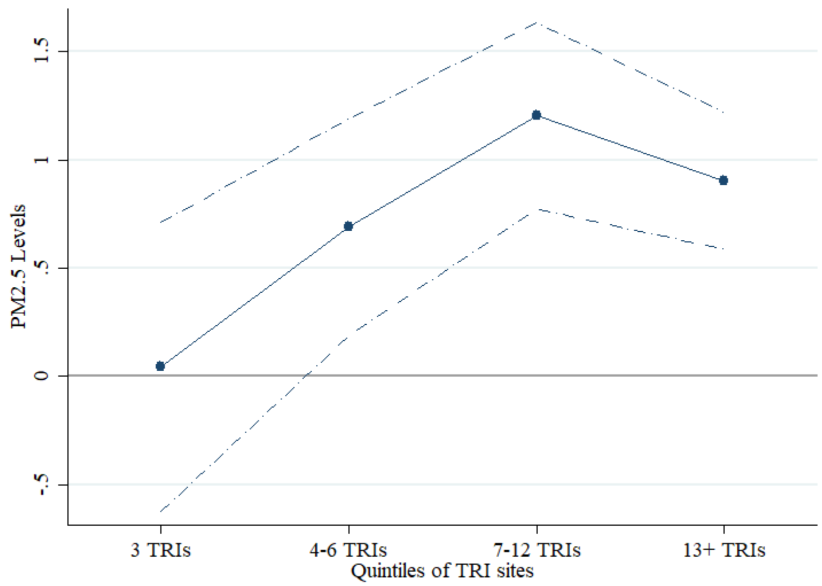
Notes: Figure 1 plots the coefficients from an OLS effects regression of weekly mean level of PM2.5 on leads and lags of time from the rollback of environmental laws on March 26, 2020 using pollution data from February through early May 2020. The red line marks the week of March 26, 2020 and all coefficients are normalized such that the coefficient in the week prior to the rollback (3/18) is zero. Dotted lines represent 0.95 confidence intervals for the coefficients. The regression controls for stay at home orders, re-openings, social distancing measures, and county fixed effects. Standard errors are clustered at the county level.

Figure 2: Association Between PM2.5 and the Number of TRI Sites Emitting Air Pollution by County

Panel A

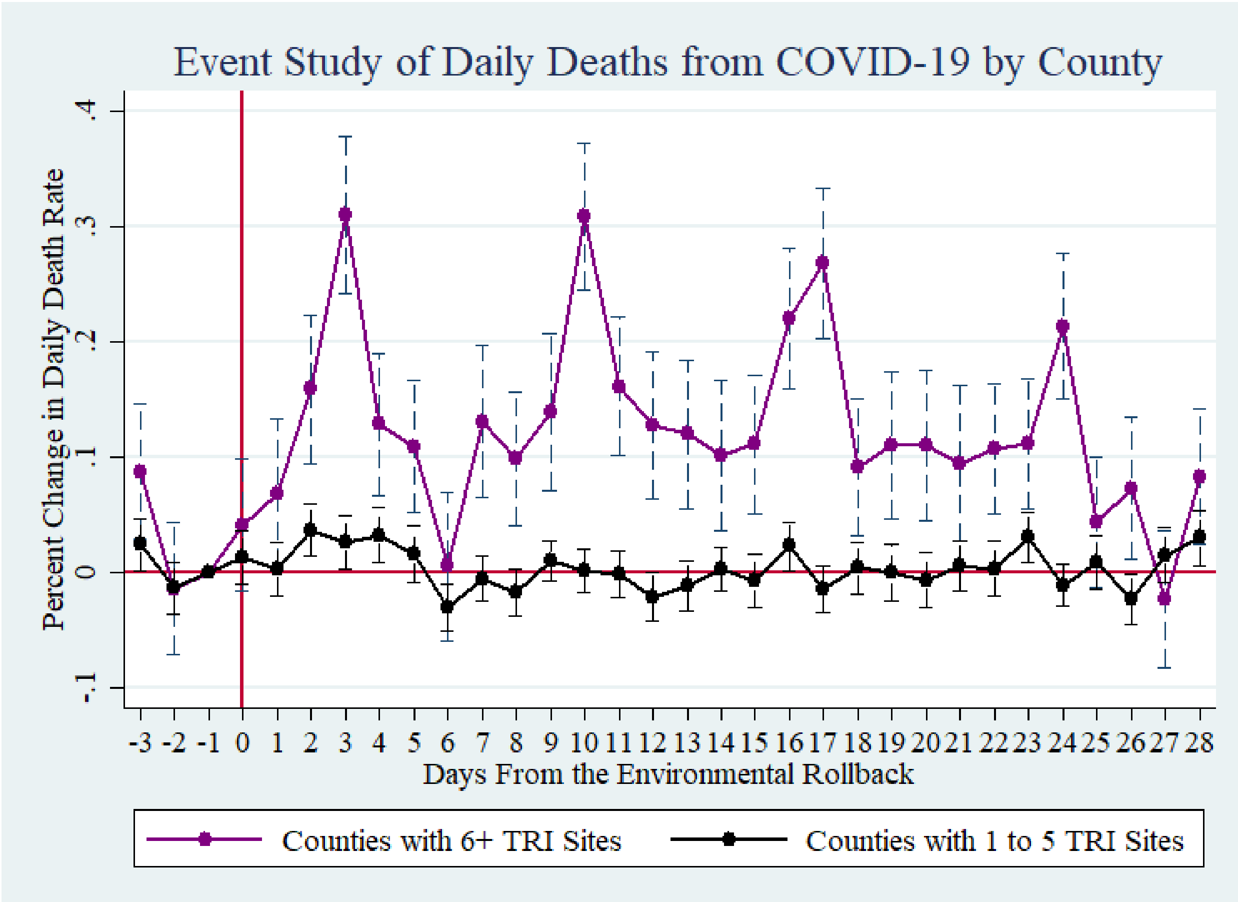


Panel B



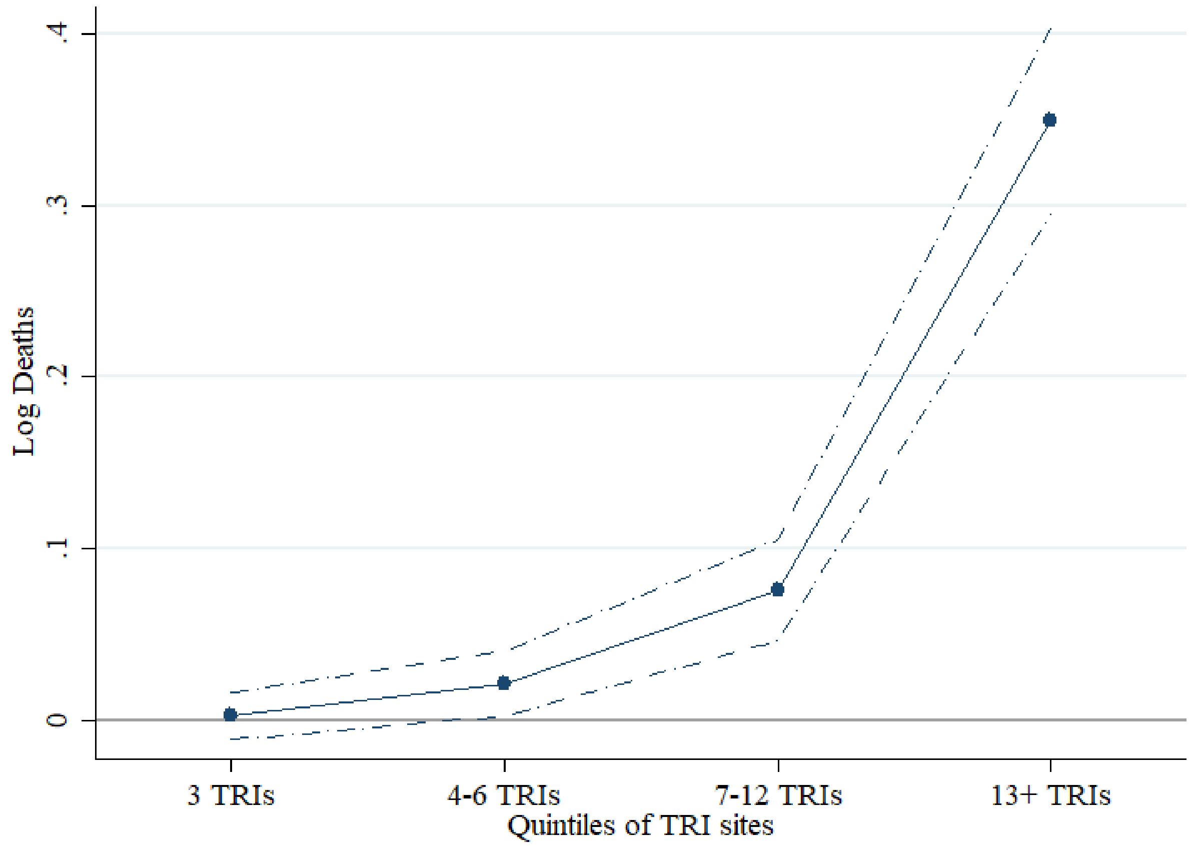
Notes: Panel A plots the relationship between the number of TRI sites in a county and the observed levels of PM2.5 pollution. As there are more TRI sites, counties experience higher pollution on average. Counties with more than 6 TRI sites comprise the treatment group in our analysis. Panel B shows how PM2.5 pollution after the rollback of environmental enforcement varied by counties based on the number of TRI sites in that county. The Y-axis plots the coefficients on the interaction between being after the EPA's rollback of civil enforcement with the stated bin for the number of TRI sites. This model includes stay at home orders, re-openings, temperature, precipitation, county fixed effects, month fixed effects and day of the week fixed effects. Includes 95% confidence intervals based on standard errors clustered on county.

Figure 3: Event Study of Daily Deaths from COVID-19 by County



Notes: Figure 3 plots the coefficients from an OLS effects regression of the log of daily COVID-19 deaths on leads and lags of time from the rollback of environmental laws on March 26, 2020 using pollution data from March and April 2020. Time 0 is March 26, 2020 and all coefficients are normalized such that the coefficient in the day prior to the rollback (-1) is zero. Dotted lines represent 0.95 confidence intervals for the coefficients. The regression controls for stay at home orders, re-openings, social distancing measures, days since the first death by county, number of confirmed cases by county, temperature, precipitation, county fixed effects, month fixed effects and day of the week fixed effects. Standard errors are clustered at the county level.

Figure 4: Effect of Pollution on Log Deaths from COVID-19 by Number of TRI sites



Notes: This figure shows how our estimates vary based on the effect of having more TRI sites in a county. The Y-axis plots the coefficients on the interaction between being after the EPA's rollback of civil enforcement with the stated bin for the number of TRI sites. This model includes stay at home orders, re-openings, social distancing measures, days since the first death by county, number of confirmed cases by county, temperature, precipitation, county fixed effects, month fixed effects and day of the week fixed effects. Includes 95% confidence intervals based on standard errors clustered on county.

Appendix Tables and Figures

Table A1: Number of Treated and Control Counties by State

(1) State	(2) Number of Control Counties	(3) Number of Treated Counties	(4) Total
Alabama	28	31	59
Alaska	4	2	6
Arizona	8	7	15
Arkansas	30	24	54
California	18	26	44
Colorado	18	11	29
Connecticut	0	8	8
Delaware	0	3	3
District of Columbia	0	1	1
Florida	25	28	53
Georgia	68	38	106
Hawaii	2	2	4
Idaho	23	4	27
Illinois	49	28	77
Indiana	38	45	83
Iowa	57	26	83
Kansas	32	16	48
Kentucky	49	21	70
Louisiana	31	21	52
Maine	9	5	14
Maryland	8	11	19
Massachusetts	0	12	12
Michigan	30	32	62
Minnesota	47	21	68
Mississippi	48	18	66
Missouri	47	22	69
Montana	14	1	15
Nebraska	35	10	45
Nevada	6	7	13
New Hampshire	3	5	8
New Jersey	4	17	21
New Mexico	9	4	13
New York	26	33	59
North Carolina	34	47	81
North Dakota	18	4	22
Ohio	28	55	83
Oklahoma	32	16	48

Oregon	14	13	27
Pennsylvania	17	45	62
Rhode Island	2	3	5
South Carolina	18	27	45
South Dakota	16	6	22
Tennessee	43	33	76
Texas	84	63	147
Utah	8	8	16
Vermont	7	3	10
Virginia	52	32	84
Washington	15	15	30
West Virginia	30	9	39
Wisconsin	25	35	60
Wyoming	12	4	16
Total	1,221	958	2,179

Notes: This table shows the number of treated counties (with 6 or more TRI sites) in Column 2, control counties (with 1 to 5 TRI sites) in Column 3, and total counties in Column 4 in the sample by state (Column 1). Note that all counties with 0 TRI sites are dropped in the main analysis.

Table A2: Results using All Counties, Including Counties with No TRI Sites and Estimates using PPML and Hausman-Taylor Correlated Random Effects

	(1)	(2)
	Log COVID-19 Deaths	Log Confirmed COVID-19 Cases
	Panel A: Results using All Counties, Including Counties with No TRI Sites	
Treated Counties in Post Period	0.2019***	0.4383***
	(0.0147)	(0.0280)
Observations	136757	136757
Predicted PM2.5 Pollution 2SLS	0.7451*	1.2023**
	(0.3884)	(0.5724)
Observations	136757	136757
	Panel B: PPML Estimates For Counties with 1 or More TRI Sites	
Treated Counties in Post Period	0.6394***	0.1399
	(0.1920)	(0.0955)
	Panel C: Hausman-Taylor Correlated Random Effects Estimates For Counties with 1 or More TRI Sites	
Treated Counties in Post Period	0.1914***	0.3915***
	(0.0148)	(0.0312)
County Fixed Effects Regression	X	X
Observations	105620	105620

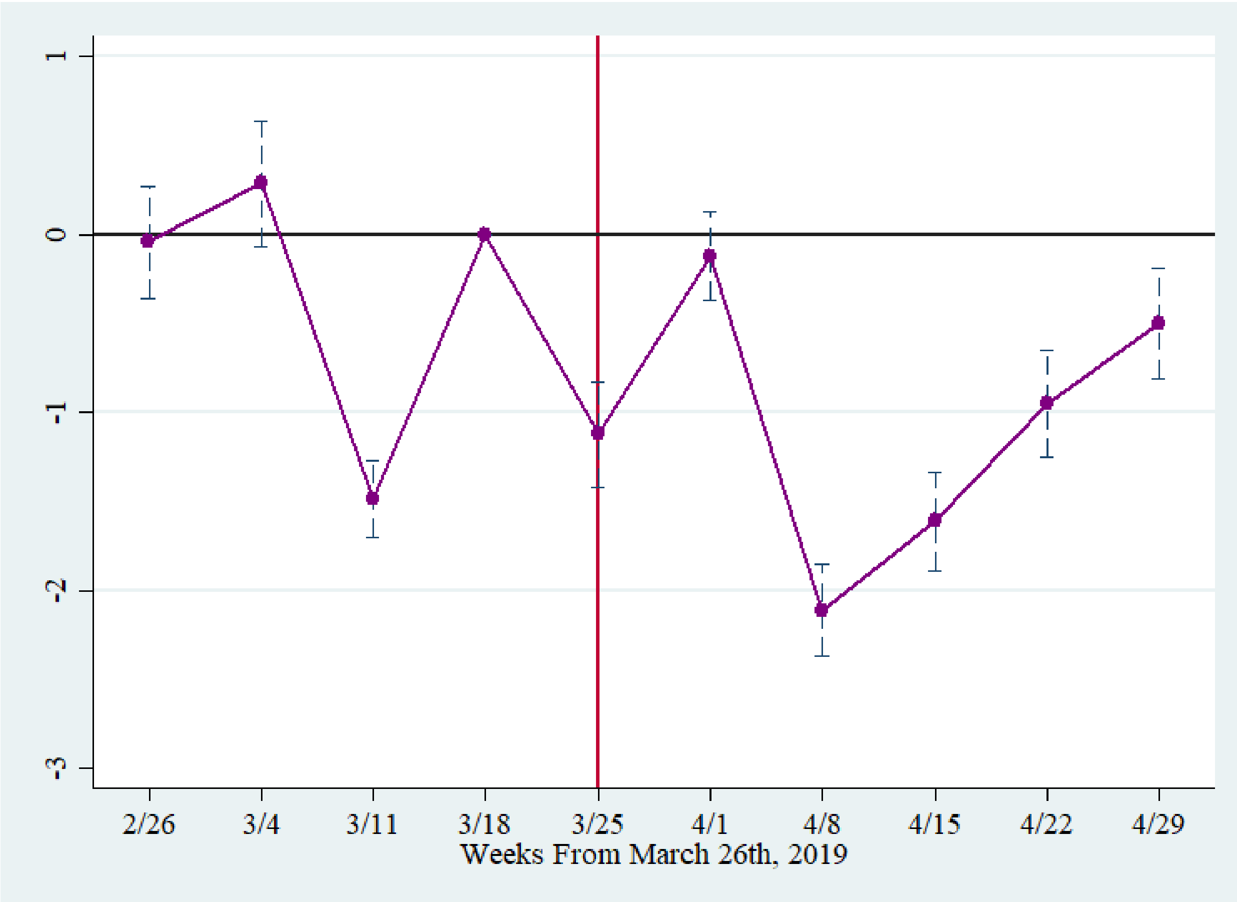
Notes: Panel A presents the results for the effects of predicted pollution or being in a treated county after the rollback with the log of COVID-19 deaths or cases as the outcome using all counties in the United States, including those with no TRI sites. Panel B presents the results of county-population-weighted Poisson pseudo-maximum likelihood regressions (PPML) with multi-way fixed effects using our primary sample of counties with one or more TRI sites. Panel C presents results when using the Hausman-Taylor random effects panel data model accounting for possible serial correlation. All models use county fixed effects and control for social distancing, stay at home orders, days since the first COVID death, weather, and day of the week and month fixed effects. Column 1 additionally controls for daily confirmed COVID-19 cases. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Table A3: The Effects of Pollution on Deaths and Cases Using Daily Rates of COVID-19 Deaths and Cases per 10,000 People (per county)

	(1) COVID-19 Daily Death Rate Per 10,000	(2) COVID-19 Daily Case Rate Per 10,000
Treated Counties in Post Period	0.0064***	0.0817*
	(0.0019)	(0.0439)
Whole Sample of Counties with TRIs using County fixed effects	X	X
Average of the dependent variable	0.0229	0.5238
Percent increase above the mean	27.9%	15.6%
Observations	98162	98162

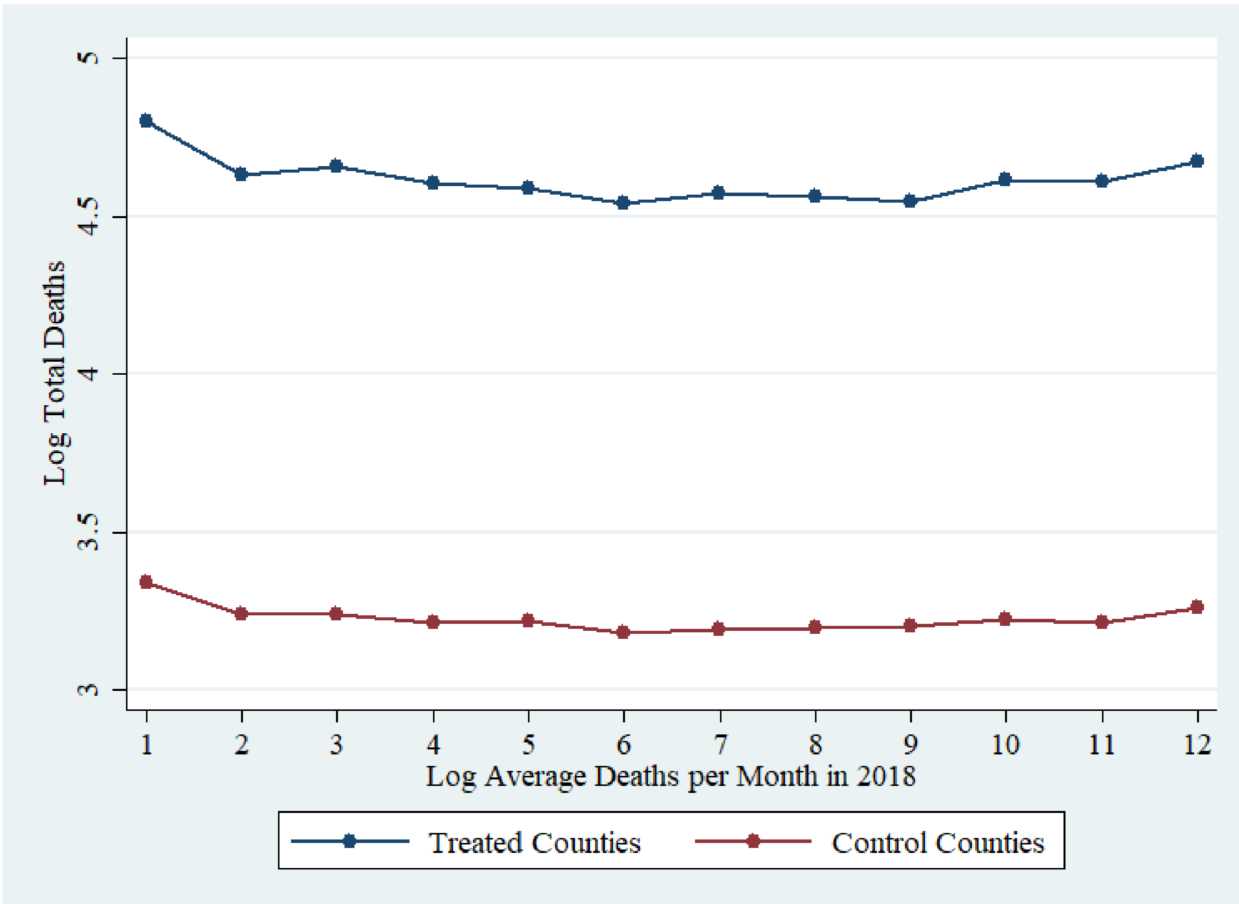
Notes: Column 1 shows the results of being in a treated county after the rollback on the daily COVID-19 death rate per 10,000 people. Column 2 shows the results of being in a treated county after the rollback on the daily COVID-19 case rate per 10,000 people. All models control for social distancing, stay at home orders, days since the first COVID death, weather, and day of the week and month fixed effects. Column 1 additionally controls for daily confirmed COVID-19 cases. These regressions only include counties with 10,000 or more individuals. Standard errors are clustered at the county level and are in parenthesis. Coefficients labeled as ***, **, and * are statistically significant at the 1, 5, and 10 percent levels, respectively.

Figure A1: Event Study for PM2.5 in Counties in 2019



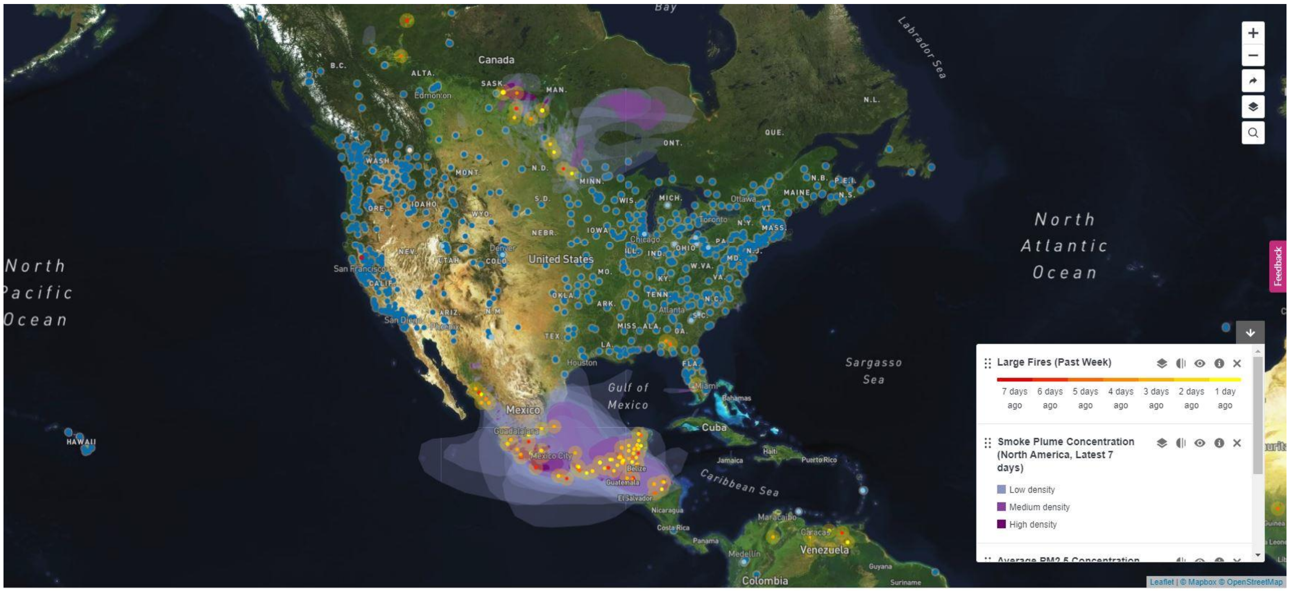
Notes: Figure A1 plots the coefficients from an OLS effects regression of weekly mean level of PM2.5 on leads and lags of time from March 26, 2019 using pollution data from January through May 2019 in all counties. The red line marks the week of March 26, 2019 and all coefficients are normalized such that the coefficient in the week prior to the rollback (3/18) is zero. Dotted lines represent 0.95 confidence intervals for the coefficients. The regression controls for weather and county fixed effects. Standard errors are clustered at the county level.

Figure A2: Log of Average Deaths Per Month in 2018 in Treated and Control Counties



Notes: Figure A2 shows the logarithm of monthly deaths in treated (with 6 or more TRI sites) and control (with 1 to 5 TRI sites) counties in 2018. Treated and control counties show similar patterns of deaths in the time before the rollback of environmental regulations and COVID-19.

Figure A3: Fires in Mexico in as of May 18, 2020



Notes: Figure A3 shows the locations of fires in Mexico on May 18, 2020 and in the previous week according to The World Resources Institute (2020). Plumes of smoke are modeled according to wind direction.

Online Data Appendix

The Unacast data on social distancing is calculated as follows. The base line comparison is against the 4 weeks prior to 03-08-2020. Unacast took the average of each weekday and each county for those 4 weeks. They assign a person to a county based on the total duration of the identifier in the county on that day. The county in which we see the device the most that day is the county they assign it to. There are 15-17 million identifiers per day in the entire dataset. The social distancing measure we use is calculated as: (average post-period distance traveled - average pre-period distance traveled) / average pre-period distance traveled.

Additional data on stay at home orders by state was gathered by the COVID-19 State Policy database, sponsored by Julia Raifman at the Boston University School of Public Health. We use pollution monitor data from the Environmental Protection Agency's Air Quality System and AirNow (when the AQS data is unavailable). The raw data is available by day, location, and pollutant. For PM2.5, the EPA has 1,069 monitors. For PM10, the EPA has 267 monitors. For ozone, the EPA has 1,195 monitors.

We assigned each monitor to the zip code in which it is located. Next, we identified zip codes that did not contain a pollutant monitor. We then assigned the nearest monitor to those zip codes, noting the distance between the monitor and the zip code center. To ensure that we were not imputing pollution measurements from monitors that were very far away from the zip code to which they were assigned, we dropped zip codes in counties without any monitors and we dropped zip codes that were more than 30 kilometers from the assigned monitor. Once we had narrowed the data to only zip codes located in counties with at least one monitor, we imputed a pollution measurement for each zip code based on the pollution measured at the nearest monitor. We created a weight based on the inverse distance between the zip code and the nearest monitor. If the zip code contained a monitor, those measurements were assigned a value of one. If the zip code had a pollution measure assigned to it, that level was multiplied by the inverse of the distance between the zip code and its assigned monitor. The data were then collapsed by county to generate a mean pollution level for each county each day.

The percentages of essential workers were derived from publicly available data from the Bureau of Labor Statistics (BLS). We used the annual average data by county from the Quarterly Census of Employment and Wages (QCEW). The data provides a county level breakdown of the annual average number of workers in different industry categories. The information is collected from establishments reporting on filled jobs. We also consulted Census data on employment, however, those data only covered a small subset of the counties we examine. Where the two data sets overlapped, our analysis resulted in similar results.

We arrived at the percentages of essential workers per county by dividing the industry designations between essential and non-essential, totaling the workers in each industry category, and dividing the sum by the total number of workers in the county. We then dropped counties that were dropped from our main analysis due to never having reported cases of COVID-19. The resulting data provides the percentage of essential workers in all counties used in our analysis. The results do contain some outliers with either very large or very small percentage of essential workers. This appears to be the result of missing data as well as certain counties were one industry appears to dominate the economy.

Employment is measured by employer type (federal government, state government, local government, or private employer). Private employment is then broken up into 12 categories. The overall industry categories measured by the BLS are “Natural Resources and Mining”, “Construction”, “Manufacturing”, “Trade, Transportation, and Utilities”, “Information”, “Financial Activities”, “ Professional and Business Service”, “Education and Health Services”, "Leisure and Hospitality", "Other Services", "Public Administration", and "Unclassified." Guided by the Massachusetts standards, we designate as "essential" these categories: Local Government, Federal Government, "Natural Resources and Mining", "Construction", "Manufacturing", "Trade, Transportation, and Utilities", and "Education and Health Services."

Every industry category likely contains some jobs that would be considered essential under the Massachusetts guidelines. We made our designations by attempting to account for the proportion of "essential" jobs in each category. For example, some "financial activities" jobs are considered essential—specifically those required to maintain financial transactions infrastructure. In contrast, "manufacturing" is considered essential if it is required for any of the other essential activities. Therefore, it is likely that a higher proportion of manufacturing jobs would be essential, as compared to "financial activities." This analysis, while imperfect, gives a sense of the types of industries that might continue to require workers to report to work during the pandemic.