Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Air quality and COVID-19 adverse outcomes: Divergent views and experimental findings

Leonardo Becchetti a,*, Gabriele Beccari a, Gianluigi Conzo a, Pierluigi Conzo b,c, Davide De Santis a, Francesco Salustri d

a University of Rome Tor Vergata, Rome, Italy
b University of Turin, Turin, Italy
c Collegio Carlo Alberto, Turin, Italy
d University College London, UK

ARTICLE INFO

Keywords:
COVID-19
Environment
Air pollution
Mortality

ABSTRACT

Background: The questioned link between air pollution and coronavirus disease 2019 (COVID-19) spreading or related mortality represents a hot topic that has immediately been regarded in the light of divergent views. A first “school of thought” advocates that what matters are only standard epidemiological variables (i.e. frequency of interactions in proportion of the viral charge). A second school of thought argues that co-factors such as quality of air play an important role too.

Methods: We analyzed available literature concerning the link between air quality, as measured by different pollutants and a number of COVID-19 outcomes, such as number of positive cases, deaths, and excess mortality rates. We reviewed several studies conducted worldwide and discussing many different methodological approaches aimed at investigating causality associations.

Results: Our paper reviewed the most recent empirical researches documenting the existence of a huge evidence produced worldwide concerning the role played by air pollution on health in general and on COVID-19 outcomes in particular. These results support both research hypotheses, i.e. long-term exposure effects and short-term consequences (including the hypothesis of particulate matter acting as viral “carrier”) according to the two schools of thought, respectively.

Conclusions: The link between air pollution and COVID-19 outcomes is strong and robust as resulting from many different research methodologies. Policy implications should be drawn from a “rational” assessment of these findings as “not taking any action” represents an action itself.

1. Introduction: pieces of the puzzle, research questions and schools of thoughts

The tragedy of the coronavirus diseases 2019 (COVID-19) pandemic has stimulated an incredibly vast number of reflections in the public opinion that ultimately turned out to become research questions for academics and medical practitioners. One of the main emerging issues explores the reason why the pandemic spread has been so uneven across different geographical areas, both across and within countries. On this point, Italy offers a classic example, with one region (Lombardy) concentrating approximately 17 percent of the population (characterized by a unique increasing trend in terms of proportion of people aged >85 on the regional population) and more than 46 percent of deaths attributed to COVID—19 from March 5th to October 15th, namely in the first epidemic wave during last winter.

The recent history of the debate and the empirical research on this question tells us that the discussion hinges around two main “schools of thought”. The first supports a purely epidemiological explanation where the only factor accounting for the observed heterogeneity is given by the frequency of physical encounters in proportion to the viral charge of each individual. According to this approach, the only variables that matter in explaining the phenomenon are those capturing the non-linear (usually “bell-shaped” curves) contagion dynamics. Variation across different geographical areas in this perspective is explained by the non-synchronous origin of the phenomenon across different regions, and crucially influenced by the presence/absence of “super-spreader”
individuals or events (e.g. the Champions League match between Atalanta and Valencia in late February, when 40,000 supporters from the neighbor province of Bergamo moved to Milan, the place of the match). A typical descriptive picture reflecting this approach is the non-synchronous overlapping of pandemic contagion curves in different countries/regions. The first day of the local epidemics at the origin of the X-axis conventionally starts from the 100th contagion case, and the abilities of different policymakers in tackling the local spread of the epidemics can be evaluated at first sight by looking at the overlapped curves.

A non-alternative but more articulated approach argues that several other factors beyond the standard expected dynamics of the contagion can help us to assess the observed heterogeneity. Air quality represents one of these factors. Over the last two months, the debate on the role of air pollution has typically opposed the two above mentioned “schools of thought” discussing the issue from two different, if not opposite, views.

The hypothesis that air quality can influence the dynamics of COVID-19 contagions and deaths finds a strong theoretical background in scientific literature about the link between air pollution and several respiratory and heart diseases. This literature suggests that quality of air can affect adverse COVID-19 outcomes in two ways: long term exposure to particulate matter (PM) may weaken health in general and lungs in particular, but at the same time it cannot be excluded that air pollutants might serve as “carriers” for the viruses.

The first hypothesis relies on the well-established link between long term exposure to PM and lungs morbidity. In particular, this exposure may have weakened lungs and alveolar reactivity to the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), thereby making severe respiratory and pulmonary consequences more likely to occur.

In the literature, the link between PM inhalation and lung diseases is hugely documented. Pope and Dockery (2006) have assessed around 200 papers focusing on adverse health effects of exposure to PM. They conclude that long term exposure produces lungs inflammation and oxidative stress, accelerating the progression and exacerbation of chronic obstructive pulmonary disease (COPD), and reducing lung function.

As it is well known, PM may have different anthropogenic (sulfates, nitrates, ammonia, carbon, lead, organics) or natural (soil, dust, seasalt, bio-aerosols) origins. PM from anthropogenic origin is made of smaller particles (usually with diameter below 2.5 μm) and it is more dangerous as it penetrates as well in small breathing passages, bronchi and air sacs, while PM particles of larger diameters (typically those from natural origins) remains in the upper respiratory views (Johnson et al., 2011). Therefore, air pollution generated by human activity is more dangerous for health than PM concentration generated by atmospheric phenomena (such as the Sahara dust carried by perturbations).

Among the researchers who support the existence of a link between PM and various morbidities before the COVID-19 pandemic, there are those identifying a link between PM2.5 and hospitalizations for pneumonia in Canada (Neupane et al., 2010), as well as between PM10 and hospital admissions for respiratory diseases in US cities (Medina-Ramon et al., 2006). Similar results have been found in China (Xu et al., 2016), the city of Boston (Zanobetti and Schwartz, 2006) and Ontario (Lognaah et al., 2005). It is noteworthy to remark that two of these studies published long time before the COVID-19 pandemic were performed in Wuhan (Zhang et al., 2015) and in Milan (Santus et al., 2012), two of the most severely hit cities by COVID-19 contagions.

A research hypothesis (i.e. carrier effect) argues that PM can carry the virus and therefore increase virus survival outside the human body. Along this line, Setti et al. (2020a) (Setti et al., 2020b) demonstrated the presence of the SARS-COV-2 viral RNA on several PM10 samples of outdoor/airborne PM10 in Bergamo, despite specific tests on vitality and infectious potential of the viral particles on PM10 were not performed due to the study design and unavailable high-security laboratory facilities. However, it can be argued that PM probably carries the coronavirus and therefore make its presence outside the human body more dangerous in terms of contagions. What is still missing in this analysis is an evaluation of the viral load and therefore of the potential contagion effects, if we consider that viruses get progressively weaker when outside the human body.

In our short survey we try to follow this discussion by illustrating the main empirical or experimental results found so far in support of both hypotheses, challenging the pure epidemiological explanation and trying to assess the relevance of air quality as a co-factor not only in terms of statistical significance but also in terms of relevance of the observed effect magnitude.

2. Materials and methods

The tools for health economists who aim at answering the proposed research question are mainly statistical and econometric methods. The inspection of a simple correlation between the two variables of interest (quality of air, on the one side, and COVID-19 contagions or deaths, on the other side) serves only as a starting point of the analysis. This is followed by the identification of all the concurring and confounding factors that may have affected the phenomenon. Thus, the analysis outlines a multivariate model that allows to test the impact (and hopefully the causality) of the main variable of interest (namely air quality) on the dependent variables, coeteris paribus.

The task is daunting since many other factors may have concurred to the observed outcome. The first candidates are time invariant structural factors related to geography, structural commuting dynamics, quality of local and regional administrations, structural characteristics of the regional health systems that in many countries are autonomous and decentralized. Together with this structural and time invariant factors (as conventionally assumed), many other time varying factors may matter. Local authorities and regulators may have had different speed and quality of reaction to the pandemic, thereby contributing to generate heterogeneity in time varying effects at regional or local level. The main candidate of time varying effects at local level is obviously represented by local mobility data that are made available by Google platforms. One of the most interesting ones is the variation of presences in “transit stations” as defined by subway, bus and train stations, sea ports, taxi stands, highway rest stops and car rental agencies. Other equally relevant mobility variables are those measuring changes of dwellers presence in urban parks or recreational premises.

Based on these considerations, the benchmark specification tested in multivariate analysis takes the standard form as follows:

\[
\text{COVID - 19 outcome}_{it} = \beta_0 + \beta_1 \text{Quality of air}_{it} + \sum_j \beta_j \text{Controls}_{it} + u_{it}
\]  

(1)

where observations are captured at time t in region m and the explanatory power of quality of air is tested after controlling for other relevant concurring factors (Controls). An important methodological issue here relates to the definition of the dependent variable. As it is well known, COVID-19 recorded cases are highly endogenous as they depend from the number of tests performed.

A more refined measure of contagion and intensity is therefore the ratio between positive cases and total number of tests. COVID-19 deaths are also measured with underlying errors and with highly heterogeneous methodologies across countries and regions for at least two reasons. The first concerns the cause of death, that is whether patients died because of the COVID-19 or with COVID-19 (patients dying for their own chronic conditions who just tested positive at the time of death) and how the two types of situations are evaluated by each local health authority, given that the distinction between the two concepts can be considered arbitrary with not so clear cut. The problem of a correct diagnosis is particularly relevant also because most of the people died with COVID-19 deaths suffered due to underlying comorbidities. As a result, some
A second problem is the lack of a proper diagnosis. When the epidemics is at peaks and intensive therapy beds in hospitals are saturated, local health officials tend to delay interventions and diagnosis for the most serious cases. As a consequence, a remarkable number of affected patients die at home without a proper diagnosis. Here again we have a problem of under-reporting of COVID-19 deaths. A solution found by many researchers to these measurement error problems is to use the “excess deaths” as dependent variable. Excess deaths can be defined as the difference between any recorded deaths in a given period (usually day or week) and the average of any deaths occurred within the same period in the previous years. The use of excess deaths has the advantage of eliminating all problems related to regional and country recording differences, because it is based on any deaths regardless the cause of the death, which are homogeneously recorded by each municipality. However, the advantage of this approach is traded off against the fact that only a measure of the “gross mortality effect” due to COVID-19 is provided. Nonetheless, this measure is interesting because it allows us to evaluate direct and indirect effects of the pandemic on morality.

There are at least three important indirect effects to be considered. First, concentration of hospital activities on the COVID-19 emergence slows down all other activities such as, for instance, follow-up visits of cancer patients. This phenomenon may eventually lead to deceases for other causes. Second, patients with other serious health emergencies (e.g. strokes, heart attacks) may delay access to hospital because they fear to be infected at hospital by COVID-19. Third, the sharp reduction of traffic during lockdown periods reduces deaths like road or workplace fatalities, especially in big cities. This third factor that reduces mortality may partially offset the first two death-increasing factors thereby making direction and sign of the difference between net and gross COVID-19 deaths ambiguous.

In order to link our benchmark model with the two competing schools of thought described in the introduction, the pure epidemiological approach estimates a reduced form where relevant controls are limited to time trends capturing the non-linear dynamics of contagion. The approach can be grossly resumed by a bell-shaped dynamics that can be captured by only three variables represented by a linear, a quadratic and a cubic time trend:

\[
\text{COVID} - 19 \text{ outcome}_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + u_{it} \tag{2}
\]

The alternative model of the school of thought advocating the role of other factors beyond non-linear epidemiological dynamics may be resumed by the assumption that these three variables do not capture all the phenomenon under investigation and the model takes instead the form of:

\[
\text{COVID} - 19 \text{ outcome}_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum \gamma_t \text{Controls}_{it} + u_{it} \tag{3}
\]

where there exists at least one \( \gamma_t \) different from zero. In the literature investigating the link between air pollution and COVID-19 cases or deaths, the main candidate is air quality.

### 2.1. Estimation approaches

The challenge between the two competing theories occurs in the domain of multivariate analysis. A typical problem in multivariate analysis is the omitted variable bias. The problem is likely to become more severe when using finer and more disaggregated administrative units (i.e. provinces or municipalities). The use of fixed effect estimates (time invariant intercepts for each administrative unit) allows to capture all unobservable time invariant local idiosyncratic factors thereby partially solving the problem. In the case of COVID-19 pandemics it may capture structural differences of local health systems (e.g. available beds in intensive care units, average distance from hospital, number of local general practitioners per person). Fixed effects cannot however capture time varying local effects such as the day-by-day reaction capacity of local authorities to the pandemics. The problem may be partially solved by using non-synchronous regional time trends starting from the first day in which contagions are more than 100 in a given region or, alternatively, by using region-week dummies. However, the most important time varying effect concerns mobility data, which have been used in the studies discussed below (see Results section). These data are crucial since they help to track the dynamic of contacts and interactions among individuals, one of the main drivers of contagion.

The significance of the \( \beta_1 \) coefficient in the multivariate analysis estimating model (1) indicates a statistically significant correlation between quality of air and the dependent variable of adverse COVID-19 outcomes. Correlation however is not causation and there are at least three different interpretations for it. First, quality of air does cause COVID-19 adverse outcomes. Second, reverse causality occurs but this cannot be applied to our case since it is hard to believe that COVID-19 mortality can cause (directly) an increase in air pollution. Third, there is an omitted driver causing both COVID-19 deaths and quality of air that produces a spurious correlation between the two variables. Candidates for this interpretation can be economic activity, population density, frequency of human interactions that cause both poor quality of air and COVID-19 adverse outcomes. A first way to control for the endogeneity problem is to include all these variables in the multivariate analysis. The pattern of these relationships can be quite complex to disentangle. In fact, it is highly plausible that interaction flows, traffic mobility and economic activity have a positive and significant effect per se on contagion, while also affecting quality of air which, in turn, negatively affects COVID-19 outcomes. A partial solution to it may be creating sample splits and testing whether the significant effect of air quality on the dependent variable persists when tested in the below median economic activity or traffic mobility sample.

Finding a solution to the endogeneity problem is not easy. A standard approach to deal with causality would rely on the first best counterfactual (i.e. a comparison between what happened with COVID-19 and what would have happened without COVID-19). Obviously, this approach is out of reach for whatever research.

The second best would be a randomized controlled trial, where the effect of pre-defined balanced treatment and control groups are tested with the difference-in-differences methodology. This approach would be impossible, too. Experimentally, we could not produce worse quality of air in some areas (treatment group) having non-significantly different ex ante characteristics compared to other areas (control group) and test whether reaction to COVID-19 epidemic is different between the two groups (it would be a quite complex experiment with a double treatment in any case and would raise ethical questions). Another usual approach in economic analysis is the instrumental variable method. An instrument is a variable that satisfies two properties: the validity, that is the variable is not directly correlated with the dependent variable, i.e. COVID-19 outcomes, and the relevance, that is the variable is significantly correlated with the instrumented driver of our interest (i.e. quality of air). Typical candidates for a valid and relevant instrument in our case are atmospheric phenomena such as wind intensity and direction, and rain precipitation that are assumed not to cause directly COVID-19 contagions or deaths while affecting significantly quality of air. Since rain precipitation may however increase indoor activities, which in turn affect virus spread, it is advisable to lag the variable and to control for time varying mobility. Lagged rain precipitation does not affect contemporary mobility while continues to affect air quality.
3. Results

3.1. Empirical findings on the nexus between quality of air and COVID-19 outcomes

Since early 2020, empirical findings rejecting the null of no incidence of quality of air on negative COVID-19 outcomes have emerged from scholars located worldwide and refer to evidences collected in different countries. Wu et al. (2020) control for a large set of observable concurring factors and find that a 1 μg/m³ is associated with an 8% increase in COVID-19 deaths in US counties. Cole et al. (2020) find a similar result for municipalities in The Netherlands, even though the quantitative effect is smaller (the change in mortality is around 3%). Carteni et al. (Carteni et al., 2020) use the number of days in 2019 with PM exceeding 50 μg/m³ as air pollution variable, and find that the impact is positive and significant. Perone (2020) finds a positive result for ozone and nitrogen dioxide together with PM. Coker et al. (2020) use municipality data and cross-sectional negative binomial models accounting for spatial autocorrelation, and find that in Northern Italy a 1 μg/m³ is associated with a 9% increase in COVID-19 deaths. Other studies finding significant effects are those of Ogen (2020), Yongijan et al. (Yongijan et al., 2020), Comunian et al. (2020), Becchetti et al. (2020a, 2020b, 2020c, 2020d, 2020e) use provincial data and test the impact of ex ante time invariant exposure to air pollution (PM2.5, PM10 and NO2) on COVID-19 cases and deaths. Their analysis shows that the impact is significant and positive when investigating the issue with different approaches. The research methodology involves first cross sectional estimates (one observation for each province) taking a static snapshot on the effect of PM concentration on cumulative cases and deaths.

Then, the methodology performs pooled and fixed effect estimates where ex ante time invariant PM exposure is interacted with epidemic time trends. Finally, the authors create an artificial experiment by predicting the dynamics of the epidemics without lockdown intervention and comparing it with what happens in the presence of the intervention. This simulated counterfactual lockdown is highly significant in reducing negative adverse outcomes and more so in provinces with poorer quality of air. Among robustness checks, the authors smooth daily into weekly averages of local rainfalls, or consider regional non-synchronous pandemic trends taking into account the strong heterogeneity of the virus spread across Italian regions. Another robustness check of their analysis consists in removing extreme rainfall events to avoid the suspicion of a direct causality between the instrument and the dependent variable. An important original contribution of this research is in the decomposition of the total effect into a time invariant and a time varying component. This decomposition aims to test simultaneously the two existing research hypotheses (i.e. long term exposure and carrier effect) on the nexus between air quality and COVID-19 adverse outcomes. To do so, the authors regress in a first stage the 11-day relevant PM moving average on the previous 2-year (time invariant) average PM concentration at municipality level. The residual of this estimate is identified as the time varying component and introduced in the benchmark estimate together with the 2-year time invariant average. Econometric findings show that both components are positive and significant therefore supporting both hypotheses. An obvious caveat of this decomposition is that the time varying component may proxy for both the carrier and the short term effect of PM variation on lung inflammation. The issue remains open to debate and to further research.

The quality of natural capital is obviously a strong antidote against air pollution. Italy represents an interesting case study because, according to data from the Ministry of the Environment and Protection of Land and Sea processed by Ancitel, in 2020 there were 2073 municipalities (around 25% of all municipalities) within protected natural areas. These are areas located within national, regional, provincial or local parks, natural reserves and sea natural areas. Around a quarter of municipalities (502) are located in national parks, while almost half share at least 45% of their surface area with parks, reserves or the so called Environmental Economic Zone (EEZ). EEZ are areas defined in 2019 by a decree-law and enjoy special economic support to preserve their natural resources. If we consider average data from the last three years until end May 2020 (thereby including the first wave of the pandemic), we find that park municipalities have on average 4 μg/m³ less of PM2.5 and PM10 and around one third of NO2. Becchetti et al. (2020a, 2020b, 2020c, 2020d, 2020e) calculate that, if we consider prudential estimates from average data from epidemiological findings, people living in “park municipalities” have around 8–10% lower mortality rate for this combined effect. Similarly, Becchetti et al. (2020a, 2020b, 2020c, 2020d, 2020e) find that better air quality reduced incidence of COVID-19 contagions and deaths in park municipalities during the first pandemic wave after controlling for all observable concurring factors.

4. Conclusions and policy implications

A number of researches in the few months after the onset of the COVID-19 pandemic have produced robust evidence on the association between air pollution and COVID-19 adverse outcomes (contagions and deaths). The set of methodologies adopted by the different contributions are extremely rich and articulated. The contemporary emergence of
significant results from different researchers located worldwide provides evidence in favor of the hypothesis of causation.

However, as explained in the methodological discussion, the first best counterfactual is not available (and can be only imperfectly simulated). Similarly, the second best of randomized experiments is out of reach when investigating a phenomenon that did not start after the organization of the experimental setting (as in any randomized control trial with treatment and control group). Therefore, we cannot confirm being one hundred percent sure the causality nexus despite the fact that the evidence presented above is quite convincing.

This does not imply however that we cannot draw policy conclusions from the existing literature, and the following analogy can be useful to understand why. Imagine you are at a dinner and you are told that, with 90 percent probability, what you are going to eat can cause you a serious illness. The instinctive, but also “rational”, reaction of each of us would be that or refusing to eat such a meal. The choice of refusing would represent our “policy decision”. The health effect of smoking is another example showing how policy interventions need to be bold and differ from academic robustness, while in dialogue with the scientific community. In fact, anti-smoking campaigns could have started before the last unempteenth evidence. In a similar manner, it is not wise not to take policy action when you know that, based on the available evidence, you are 90 percent (or almost as such) certain that quality of air has a positive effect on COVID-19 contagions and deaths.

The suggestion to reduce PM concentration stemming from this literature is not new. The World Health Organization calculates that air pollution (to whom PM concentration gives one of the main contributions) kills around 7 million people around the world. Although the sectors contributing more to pollutant emissions vary across regions, we know that overall house heating is the main responsible of PM propagation, followed by traffic mobility, energy production, industry and agriculture (Iriti et al., 2020). It is therefore urgent to replace polluting production techniques with cleaner techniques in the most pollutant sectors. This policy advice is not new but it is definitely reinforced by what found with the recent research on the determinants of COVID-19 deaths.

Declaration of competing interest

We have no interests to declare

References

Austin, Wes, et al., 2020. COVID-19 Mortality and Contemporaneous Air Pollution. No. Paper2016. International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.
Becchetti, L., Becacci, G., De Santis, D., 2020a. Lagged Particulate Matter, Contagions and Deaths: the relationship between Quality of Air and COVID-19 at European Level, vol. 159. SOAS University of London. Cefimsdp.
Becchetti, L., Becacci, G., Conzo, G., Conzo, P., De Santis, D., Salustri, F., 2020b. Particulate Matter and COVID-19 Excess Deaths: Decomposing Long-Term Exposure and Short-Term Effects. Mimeo.
Becchetti, L., Conzo, G., Conzo, P., Salustri, F., 2020c. Understanding the heterogeneity of adverse COVID-19 outcomes: the role of poor quality of air and lockdown decisions. Available at: SSRN: https://doi.org/10.2139/ssrn.3572548.