Drinking Water Contamination and COVID-19 Mortality in the United States

Kelly Hyde[1]

Abstract

Over 185,000 deaths have been attributed to the COVID-19 pandemic in the United States as of September 2020. There is growing evidence that the composition of these deaths reflects multiple long-standing health disparities, including environmental quality. In this paper, a county-day level panel of confirmed COVID-19 case and death counts, water quality violations, and demographic variables is constructed to estimate the association between risk of exposure to drinking water contaminants and the COVID-19 case fatality rate (CFR). Counties with more recent violations among major community water systems than average ("treated") are matched to "control" counties on key demographic and environmental variables using coarsened exact matching (CEM). Three categories of water quality violations are considered: acute health-based violations, which pose an immediate health threat to exposed individuals; health-based violations involving contaminants shown in prior literature to increase the risk of cardiovascular disease (lead, arsenic, cadmium, and copper); and all health-based violations regardless of type. The county-level COVID-19 CFR is significantly associated with acute and cardiovascular-associated health-based violations. On average, the CFR is about 18% higher (0.48 percentage points; $p < 0.01$) in counties more affected by acute violations than average and about 15% higher (0.42 percentage points; $p = 0.037$) in counties more affected by cardiovascular-associated violations. There is suggestive evidence of a linear association between the "dose" of violation exposure (the sum of the estimated percentages of the population affected by each respective violation) and the CFR.

Introduction

As of September 2020, over 185,000 deaths in the United States have been attributed to COVID-19. Recent literature demonstrates a disproportionate incidence of COVID-19 morbidity and mortality among Black, Hispanic, and Native American populations (Benitez et al., 2020, Buford and Johnson, 2020), individuals in poverty (Mollalo et al., 2020), and individuals in occupations for which social distancing is difficult (Almagro and Orane-Hutchinson, 2020, Dyal et al., 2020). While COVID-19 itself is novel, these disparities in outcomes reflect long-standing broader health disparities across subpopulations attributable to a range of structural and socioeconomic determinants (Woolf and Braveman, 2011), including environmental quality (Schaider et al., 2019, Switzer and Teodoro, 2018).

Exposure to air and water pollutants is associated with the development of both acute and chronic health conditions (Brulle and Pellow, 2006, Hunter et al., 2010, Manan et al., 2018, Troesken, 2004, Zivin and Neidell, 2013), and the presence of preexisting health conditions is strongly associated with COVID-19 mortality (Wortham et al., 2020). In terms of air pollution, recent studies have found evidence in favor of such an association (Comunian et al., 2020, Petroni et al., 2020, Wu et al., 2020). However, the possibility of a similar association between water

[1]University of Pittsburgh, email: kdh50@pitt.edu.

contamination and COVID-19 mortality has not yet been explored.[2] While an association with air pollution is perhaps more intuitive for a respiratory disease, the primary mechanism by which historical air pollution interacts with present-day COVID-19 mortality—increased prevalence of preexisting health conditions—is also plausible for water pollution.

The scope for preexisting health conditions caused by water contamination to exacerbate the mortality consequences of COVID-19 is large. Millions per year in the United States are potentially affected by health-based drinking water quality violations (Allaire et al., 2018). Depending on the contaminant, these violations can increase risk of a range of conditions affecting the digestive, cardiovascular, endocrine, nervous, and reproductive systems (US EPA, 2015). Although COVID-19 is primarily considered a respiratory disease, it has systemic effects on the cardiovascular and immune systems, and clinical outcomes are significantly worse for individuals with preexisting cardiovascular and immunosuppressive conditions (Liu Peter P. et al., 2020, Wortham et al., 2020). Thus violations of maximum contaminant levels (MCLs) or maximum residual disinfectant levels (MRDLs) for substances associated with elevated risk of cardiovascular problems are especially likely to interact with COVID-19 mortality.

Identifying subregions and subpopulations at high risk of adverse COVID-19 outcomes is essential to devise well-designed mitigation policies in the absence of a proven and widely available vaccine. Recent literature in economics has explored optimal targeted lockdown policies to achieve mitigation while minimizing the economic side effects such as income losses (Acemoglu et al., 2020, Fajgelbaum et al., 2020, Jones et al., 2020). In particular, Acemoglu et al. (2020) varies lockdown policies by age group, with the optimal strictness increasing with age as the relative risk increases. Accordingly, if the COVID-19 fatality rate is higher in areas with a history of water quality violations, the optimal mitigation policy in these areas will be stricter. Additionally, irrespective of policy, it is also likely that the economic effects of the pandemic itself on these areas will be more severe. Decreases in spending and employment have predominantly been attributed to direct effects of the pandemic rather than mitigation policies (Bartik et al., 2020, Cox et al., 2020), suggesting that the risk of negative economic consequences and their magnitude is increasing with the severity of the pandemic.

This paper aims to determine the extent to which geographic variation in the COVID-19 case fatality rate in the United States is associated with community water system health-based violations. Conditional on exposure to the virus, it is anticipated that COVID-19 mortality will be more likely in areas with more violations that are associated with increased risk of preexisting conditions that interact with COVID-19 clinical outcomes. Accordingly, two types of violations are focused on: acute health-based violations, which are most likely to produce immediate adverse health effects in exposed individuals, and cardiovascular-associated health-based violations more broadly, which may cause cardiovascular conditions that exacerbate COVID-19 morbidity (Liu Peter P. et al., 2020, Wortham et al., 2020).

Methods

Data sources

Confirmed COVID-19 case and death counts by day at the county level were obtained from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns

[2]The most closely related study is Amankwaa and Fischer (2020), which explores the correlation between poor access to water, sanitation, and hygiene services and COVID-19 mortality in sub-Saharan Africa.

Hopkins University (Dong et al., 2020). This data spans from January 22, 2020 to September 8, 2020. As of September 9, 2020, the date on which the analysis was conducted, this data included 3,068 counties, 2,483 (80.9%) of which had at least one recorded death.

Water quality violations were obtained from the Safe Drinking Water Information System (SDWIS) maintained by the U.S. EPA. This data includes both health-based (MCL, MRDL, and treatment technique) and monitoring/reporting violations from 2009 to 2019. Counties were included in the final dataset if they had at least one community water system (CWS) that serves at least 500 people (following Allaire et al. (2018), which uses this criterion because reporting requirements differ for smaller water suppliers) represented in SDWIS. This results in a sample of 2,776 counties.

Demographic variables at the county level, including sex and race/ethnicity composition, poverty rate, percentage of the population over 65 years of age, reliance on public transportation, average household size, educational attainment, prevalence of Census-recorded disabilities, and occupation shares were calculated using the 2014-2018 5-year American Community Survey (ACS). Population density was calculated using the population estimates from the ACS and geographic area from a shapefile of U.S. county boundaries (ESRI, 2012).

County-level measures of PM2.5 air pollution were obtained from Donkelaar et al. (2019) following the procedures of Wu et al. (2020). This data is used to verify that the estimated association between water quality violations and COVID-19 mortality is not simply attributable to a positive correlation between air pollution and water pollution in an area. In fact, the raw correlation coefficient between PM2.5 and the measure of water quality violation exposure used in this paper is negative (-0.06). This is largely attributable to the fact that air pollution is predominantly an urban phenomenon in the United States (Brochu et al., 2011, Sun and Zhu, 2019) while water quality violations most frequently occur in low-income rural areas (Allaire et al., 2018).

Definitions and measurements

The case fatality rate (CFR) was calculated by day for each county based on its daily cumulative count of confirmed cases and deaths. However, since the COVID-19 pandemic is ongoing and there is a natural lag between a case being confirmed and culminating in a death, studies have proposed a range of lagged measures of CFR to prevent underestimates when the growth rate of cases is high (Baud et al., 2020, Kim et al., 2020, Rosakis and Marketou, 2020). These measures are constructed by dividing the cumulative number of deaths by the cumulative number of cases N days ago, where N is the average delay between the reporting of a case and a fatality. Four different CFRs were calculated: contemporaneous, 5-day lagged, 10-day lagged, and 14-day lagged. The results reported in the paper use the 14-day lagged measure to match Baud et al. (2020), but sensitivity analyses were conducted with all four measures without significant changes in conclusions.

Water quality violation measures were constructed from the SDWIS data based on the percentage of the corresponding county's population that was affected by each violation. These measures represent the cumulative dose of water contamination events experienced by a county's population from 2009 to 2019. For example, if a CWS that serves 1,000 people in a county with 10,000 residents committed an acute health-based violation, the dose of this violation is 0.1. These doses are calculated for the following categories: all violations recorded in SDWIS, acute health-based violations, and violations regarding contaminants associated with increased risk of cardiovascular problems. For the last category, arsenic, lead, copper, and cadmium were selected based on prior literature in epidemiology (Chowdhury et al., 2018, Jomova and Valko, 2011, Lanphear et al., 2018, Yang et al., 2020).

In analyses where the water quality violation measure is dichotomized, a county is considered "treated" if its value of the associated violation measure is greater than the sample median, or greater than zero in cases where the sample median is equal to zero. This is true of cardiovascular-associated violations, which have occurred in 237 (7.7%) of the 3,068 counties in the sample, and any acute health-based violations, which have occurred in 448 (14.6%).

Data analysis

Treated counties in the sample were matched with control counties on demographics using coarsened exact matching (CEM) (Iacus et al., 2012). CEM is appealing in this context for several reasons. One, water quality violations are not randomly assigned. Using SDWIS data, Allaire et al. (2018) shows that health-based violations are most common in low-income rural areas and spatially concentrated in the Southwest and Northwest, especially in Texas, Oklahoma, and Idaho. Thus a range of confounders, including racial composition, poverty, and population density, need to be adjusted for when estimating the association between water quality violations and COVID-19 mortality. By matching observations on these confounding variables, CEM improves balance between treatment and control with limited model dependence. Two, CEM restricts the sample to observations that can be matched to at least one other observation on the selected demographics, thereby limiting the need for extrapolation (Iacus et al., 2012, Stuart, 2010). This is especially important in this context because variance in both the COVID-19 CFR and the demographic composition is higher in smaller, more rural counties, which are also among the areas most affected by water quality violations.

Standard regression estimates are presented alongside the CEM estimates. Unlike the CEM estimates, which focus on extensive-margin variation above and below a particular cutoff of water contamination exposure risk, the standard regression estimates take intensive-margin variation in the dose of exposure into account because the continuous violation measures are used. State dummies are also included in these regression models. The inclusion of state dummies addresses the potential concern of different testing or reporting practices across states at the cost of absorbing substantial geographic variation in water quality violations.

In the specifications reported in the main text, extreme outliers in the 14-day lagged CFR are trimmed by dropping the top 1% of observations (31 counties). As of September 8, 2020, this meant dropping counties with a 14-day lagged CFR of 11.7% or greater. This is done for two reasons: one, since health-based water quality violations are more common in rural areas where both absolute case and death counts are likely to be smaller, the estimates could be substantially biased upward by counties with a small number of cases and very large 14-day lagged CFR. Two, in larger counties where earlier outbreaks occurred and the estimated CFR is very high, such as Sussex County, New Jersey (14.3%) and Middlesex County, Connecticut (13.4%), extremely limited availability of COVID-19 testing in March and April likely leads to a very high degree of ascertainment bias (Golding et al., 2020).

Results

CEM estimates

Results in this subsection were obtained by applying CEM to the sample as of September 8, 2020 matching on the following county demographic variables: the estimated share of the population

Table 1: Mean Difference in COVID-19 Case Fatality Rate by Types of Health-Based Water Quality Violations as of September 8, 2020, CEM Estimates

Dependent variable: Cumulative CFR × 100, Cases Lagged by 14 Days			
	(1)	(2)	(3)
Violation type	Acute	Cardiovascular	All
Treated	0.48***	0.42**	0.01
(Violations above sample median)	(0.14)	(0.20)	(0.11)
Treated N (Match %)	268 (60%)	137 (58%)	850 (63%)
Control N	1033	788	866
Control Mean CFR	2.63	2.79	2.95

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

that is Black, Hispanic, below 200% of the federal poverty line, and over 65 years of age, respectively; the population density per square meter; and the average PM2.5 air pollution from 2000 to 2016 (Wu et al., 2020). When regression was conducted on the matched subsample, the following additional control variables were included: the estimated share of the population that is female, is Asian, is Native American, is 18 to 34 years of age, is 35 to 64 years of age, relies primarily on public transportation, is married, lives in a household with 4 or more residents, has a Bachelor's degree or higher educational attainment, has a Census-recorded disability, is employed in the manufacturing sector, is a healthcare worker, and is a food-service worker, respectively, and the number of days elapsed since the county's first confirmed COVID-19 case. The variables used for matching were also included as controls to span any remaining differences after CEM as recommended by Iacus et al. (2012).

Table 1 shows the results of this specification. There is a significant ($p < 0.05$) association between higher-than-average incidence of acute and cardiovascular-associated health-based violations, respectively, and the COVID-19 case fatality rate. The estimated difference based on acute health-based violations is about 0.48 percentage points (0.19 to 0.76, $p < 0.01$), and the estimated difference based on cardiovascular-associated health-based violations is about 0.42 percentage points (0.03 to 0.82, $p = 0.037$). For acute health-based violations, the point estimate is about 18% of the control group mean; for cardiovascular-associated health-based violations, it is about 15%. By contrast, there is no significant association between the COVID-19 CFR and the average incidence of all health-based violations. As expected, this suggests that the type of contaminant matters, and the types of contaminants significantly associated with COVID-19 mortality are those most likely to cause pre-existing conditions that exacerbate the severity of COVID-19.

Figure 1 shows how the estimated treatment effects presented in Table 1 have evolved over the course of the pandemic. The wider confidence intervals and noisier estimates in May and June reflect the fact that the spread of COVID-19 ramped up in different areas at different times, meaning that the subsample of counties for which the CFR is defined (especially with a 14-day lag on case counts) increases over time. In particular, COVID-19 arrived in rural counties systematically later (Healy et al., 2020, Thebault and Hauslohner, 2020). By July 15, the treatment effects in all three specifications converge to approximately their current values reflected in Table 1.

Figure 1: Daily Estimated Treatment Effects from May 1, 2020 to September 8, 2020

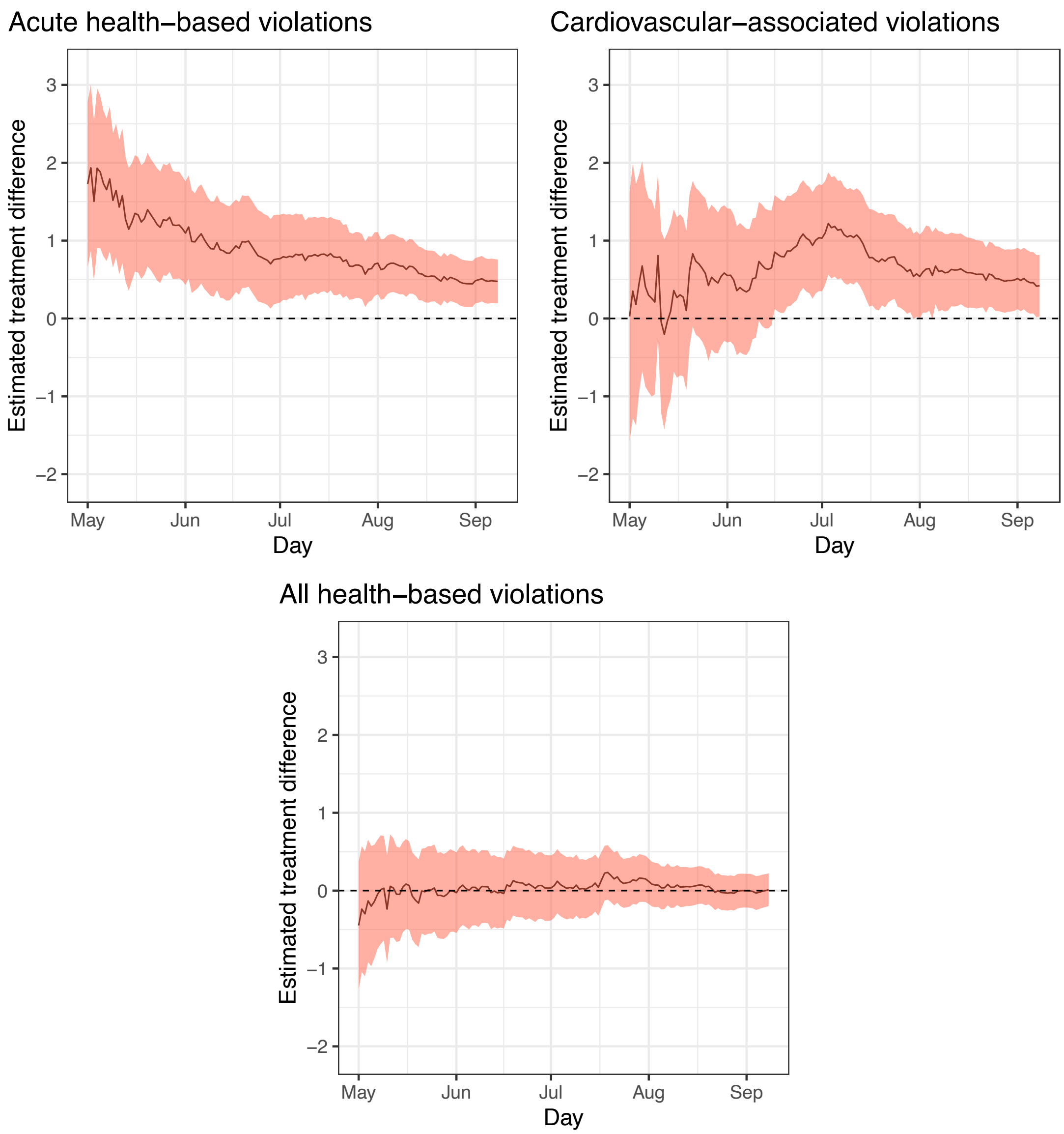

Table 2: Mean Difference in COVID-19 Case Fatality Rate by Types of Health-Based Water Quality Violations as of September 8, 2020, Regression Estimates

Dependent variable: Cumulative CFR × 100, Cases Lagged by 14 Days			
	(1)	(2)	(3)
Violation type	Acute	Cardiovascular	All
	0.752*	0.034*	0.019**
	(0.420)	(0.021)	(0.010)
State dummies included	Yes	Yes	Yes
Adjusted R^2	0.13	0.13	0.13
Total observations	2747	2747	2747
Observations with non-zero dose	383	236	1795
Sample mean CFR	2.50	2.50	2.50
Mean dose (including zeroes)	0.03	0.26	1.93

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard regression estimates

Table 2 reports the results of a standard regression without matching for the same three categories of water quality violations as Table 1. All three columns include all controls that were included in the CEM estimates (listed in the previous subsection) as well as state dummies to absorb any systematic differences across states in COVID-19 testing, reporting, and mitigation policies. Because state dummies were included, the results of Table 2 should be interpreted as conservative estimates. The state dummies may absorb some meaningful variation in water quality violations, which are highly heterogeneous by state as discussed in the data analysis subsection.

In this specification, cardiovascular-associated health-based violations and acute violations are marginally significantly associated with the CFR ($p < 0.1$) while all health-based violations, contrasting with the CEM estimates in Table 1, are significantly associated ($p < 0.05$) albeit with a smaller point estimate. The point estimates suggest that an acute health-based violation affecting the entire population of a county (adding 1 to the county dose) increases the COVID-19 CFR by 0.75 percentage points (30% of the sample mean), while any cardiovascular-associated health-based violation increases the CFR by 0.03 (1.4% of the sample mean) and any health-based violation increases the CFR by 0.02 (0.8% of the sample mean) on average. It is important to note, however, that the relative rarity of acute health-based violations means that the associated point estimate is a substantial extrapolation from the sample average and should be interpreted cautiously.

Discussion

The results in Tables 1 and 2 should be considered local average treatment effects (LATEs) applicable only to the counties that experienced acute or cardiovascular-associated health-based violations respectively. This subsample of counties is not nationally representative: in both cases, it is more White, less densely populated, less likely to be in poverty, less likely to have a Census-recorded disability, and less exposed to PM2.5 air pollution than the national average. (For a balance table, see Online Appendix.) While this may limit the generalizability of the findings, it also draws a distinction between disparities associated with historical exposure to water contamination and other documented disparities in the United States, most of which focus on urban-dwelling, non-White, or poor (and combinations thereof) individuals (e.g. Almagro and Orane-Hutchinson, 2020, Bandi et al., 2020, Benitez et al., 2020, Doumas et al., 2020, Mineo, 2020, Wu et al., 2020). This is likely in part because COVID-19 broke out in densely populated urban areas of the United States, particularly New York City, about a month earlier than it did in most rural areas (Healy et al., 2020). As the pandemic has spread over time, the burden of COVID-19 morbidity and mortality has shifted away from counties that experienced earlier outbreaks toward areas where water contamination risk is higher (Thebault and Hauslohner, 2020).

The primary limitation to the analyses presented in this paper is the unobserved heterogeneity in COVID-19 testing and reporting practices across counties. Particularly when focusing on rural areas, ascertainment bias is a concern when working with confirmed COVID-19 case and death data. In areas that do not have robust healthcare services and thus are less likely to have large-scale, conveniently accessible COVID-19 testing centers, it is possible that non-severe COVID-19 cases are significantly underreported, while cases that become severe enough to culminate in hospitalization or death are likely to be identified at a similar rate. Since the conditional CFR for severe cases only is naturally higher than the CFR for all cases, this could result in an overestimate of the CFR in rural areas where water quality violations are more likely to occur. By matching treatment and control on population density, the CEM approach used to produce Table 1 limits this concern, but the possibility that ascertainment bias is significantly positively correlated with historical water quality violations cannot be entirely ruled out since ascertainment bias is unobservable.

At the same time, it is possible that these estimates are attenuated by under-ascertainment of water quality violations. An estimated 26% of health-based water quality violations are either not reported or inaccurately reported to SDWIS (U.S. Government Accountability Office, 2011). This is especially a concern for lead contamination, since lead is most likely to enter the residential water supply through the corrosion of lead-based residential pipes and there is limited information on the location of these pipes (U.S. Government Accountability Office, 2017). States are not currently required to report the locations of lead-based pipes to SDWIS, making it impossible to ascertain whether or not there is adequate sampling of high-risk areas to facilitate enforcement of the Lead and Copper Rule. Thus there may be communities highly exposed to water contaminants associated with the risk of cardiovascular disease that are not represented in the SDWIS data.

Conclusion

Five months after the initial outbreak in New York City, the number of deaths attributed to COVID-19 in the United States continues to rise with scattered outbreaks throughout the country. Prior literature on the early dynamics of the pandemic has shown that these outbreaks pose different risks of severe morbidity and mortality to different subpopulations. Knowledge of these disparities is essential to build mitigation policies that optimally balance risk reduction with the avoidance of disruptions to economic activity and day-to-day life. All else equal, the optimal mitigation policy in an area with individuals more susceptible to severe consequences of COVID-19 is more aggressive, with tighter restrictions and more stringent criteria (e.g. the moving average positive test rate) required to ease those restrictions.

This paper demonstrates that counties with a higher likelihood of exposure to drinking water contaminants—especially those linked to the risk of cardiovascular disease—exhibit a higher case fatality rate (CFR) than otherwise comparable areas. This finding has implications for both current and long-run public health policy. In the short run, it suggests that the violation history of an area's main water suppliers should be considered in the design of optimal mitigation policies. All else equal, the consequences of relaxing restrictions on large gatherings and close-contact indoor activities will be more severe in a county with serial water quality violations which have placed its residents at a higher risk of pre-existing health conditions. In the long run, it suggests that residential water quality influences the mortality consequences of communicable disease even when that disease is not waterborne or otherwise explicitly linked to the water supply. Thus, in addition to the many direct health benefits of improving drinking water quality, investing in residential water infrastructure, monitoring, and enforcement capacity will mitigate the population health consequences of future pandemics.

References

Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. Optimal Targeted Lockdowns in a Multi-Group SIR Model. Working Paper 27102, National Bureau of Economic Research, May 2020. URL http://www.nber.org/papers/w27102. Series: Working Paper Series.

Maura Allaire, Haowei Wu, and Upmanu Lall. National trends in drinking water quality violations. Proceedings of the National Academy of Sciences, 115(9):2078–2083, February 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1719805115. URL https://www.pnas.org/content/115/9/2078. Publisher: National Academy of Sciences Section: Social Sciences.

Milena Almagro and Angelo Orane-Hutchinson. The differential impact of COVID-19 across demographic groups: Evidence from NYC. SSRN Scholarly Paper ID 3573619, Social Science Research Network, Rochester, NY, April 2020. URL https://papers.ssrn.com/abstract=3573619.

Godfred Amankwaa and Christian Fischer. Exploring the correlation between COVID-19 fatalities and poor WASH (Water, Sanitation and Hygiene) services. medRxiv, page 2020.06.08.20125864, June 2020. doi: 10.1101/2020.06.08.20125864. URL https://www.medrxiv.org/content/10.1101/2020.06.08.20125864v1. Publisher: Cold Spring Harbor Laboratory Press.

Sindhura Bandi, Michael Z. Nevid, and Mahboobeh Mahdavinia. African American children are at higher risk of COVID-19 infection. Pediatric Allergy and Immunology, n/a(n/a), 2020. ISSN 1399-3038. doi: 10.1111/pai.13298. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/pai.13298. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pai.13298.

Alexander W Bartik, Marianne Bertrand, Feng Lin, Jesse Rothstein, and Matt Unrath. Measuring the labor market at the onset of the COVID-19 crisis. Working Paper 27613, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27613. Series: Working Paper Series.

David Baud, Xiaolong Qi, Karin Nielsen-Saines, Didier Musso, Léo Pomar, and Guillaume Favre. Real estimates of mortality following COVID-19 infection. The Lancet. Infectious Diseases, 20(7):773, July 2020. ISSN 1473-3099. doi: 10.1016/S1473-3099(20)30195-X. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7118515/.

Joseph Benitez, Charles Courtemanche, and Aaron Yelowitz. Racial and Ethnic Disparities in COVID-19: Evidence from Six Large

Cities, 2020. URL https://www.iza.org/publications/dp/13521/racial-and-ethnic-disparities-in-covid-19-evidence-from-six-large-cities.

Paul J. Brochu, Jeff D. Yanosky, Christopher J. Paciorek, Joel Schwartz, Jarvis T. Chen, Robert F. Herrick, and Helen H. Suh. Particulate Air Pollution and Socioeconomic Position in Rural and Urban Areas of the Northeastern United States. American Journal of Public Health, 101(S1):S224–S230, November 2011. ISSN 0090-0036. doi: 10.2105/AJPH.2011.300232. URL https://ajph-aphapublications-org.pitt.idm.oclc.org/doi/full/10.2105/ajph.2011.300232. Publisher: American Public Health Association.

Robert J. Brulle and David N. Pellow. ENVIRONMENTAL JUSTICE: Human Health and Environmental Inequalities. Annual Review of Public Health, 27(1):103–124, March 2006. ISSN 0163-7525. doi: 10.1146/annurev.publhealth.27.021405.102124. URL https://www.annualreviews.org/doi/10.1146/annurev.publhealth.27.021405.102124. Publisher: Annual Reviews.

Talia Buford and Akilah Johnson. Early Data Shows African Americans Have Contracted and Died of Coronavirus at an Alarming Rate, April 2020. URL https://www.propublica.org/article/early-data-shows-african-americans-have-contracted-and-died-of-coronavirus-at-an-alarming- Archive Location: https://www.propublica.org/ Library Catalog: www.propublica.org Publisher: ProPublica.

Rajiv Chowdhury, Anna Ramond, Linda M. O'Keeffe, Sara Shahzad, Setor K. Kunutsor, Taulant Muka, John Gregson, Peter Willeit, Samantha Warnakula, Hassan Khan, Susmita Chowdhury, Reeta Gobin, Oscar H. Franco, and Emanuele Di Angelantonio. Environmental toxic metal contaminants and risk of cardiovascular disease: systematic review and meta-analysis. BMJ, 362, August 2018. ISSN 0959-8138, 1756-1833. doi: 10.1136/bmj.k3310. URL https://www.bmj.com/content/362/bmj.k3310. Publisher: British Medical Journal Publishing Group Section: Research.

Silvia Comunian, Dario Dongo, Chiara Milani, and Paola Palestini. Air Pollution and COVID-19: The Role of Particulate Matter in the Spread and Increase of COVID-19's Morbidity and Mortality. International Journal of Environmental Research and Public Health, 17(12), June 2020. ISSN 1661-7827. doi: 10.3390/ijerph17124487. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7345938/.

Natalie Cox, Peter Ganong, Pascal Noel, Joseph Vavra, Arlene Wong, Diana Farrell, and Fiona Greig. Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data. SSRN Scholarly Paper ID 3633008, Social Science Research Network, Rochester, NY, July 2020. URL https://papers.ssrn.com/abstract=3633008.

Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, 0(0), February 2020. ISSN 1473-3099, 1474-4457. doi: 10.1016/S1473-3099(20)30120-1. URL https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/abstract. Publisher: Elsevier.

A. Donkelaar, R.V. Martin, C. Li, and R.T. Burnett. Regional Estimates of Chemical Composition of Fine Particulate Matter using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors, Environ. Sci. Technol, 2019. doi: 10.1021/acs.est.8b06392.

Michael Doumas, Dimitrios Patoulias, Alexandra Katsimardou, Konstantinos Stavropoulos, Konstantinos Imprialos, and Asterios Karagiannis. COVID19 and increased mortality in African Americans: socioeconomic differences or does the renin angiotensin system also contribute? Journal of Human Hypertension, pages 1–4, July 2020. ISSN 1476-5527. doi: 10.1038/s41371-020-0380-y. URL https://www.nature.com/articles/s41371-020-0380-y. Publisher: Nature Publishing Group.

J.W. Dyal, M.P. Grant, and K. Broadwater. COVID-19 Among Workers in Meat and Poultry Processing Facilities — 19 States. In April 2020. MMWR Morb Mortal Wkly Rep 2020;69:557–561. 2020. doi: 10.15585/mmwr.mm6918e3. URL DOI:.

ESRI. USA Counties, 2012. URL https://www.arcgis.com/home/item.html?id=a00d6b6149b34ed3b833e10fb72ef47b.

Pablo Fajgelbaum, Amit Khandelwal, Wookun Kim, Cristiano Mantovani, and Edouard Schaal. Optimal Lockdown in a Commuting Network. Working Paper 27441, National Bureau of Economic Research, June 2020. URL http://www.nber.org/papers/w27441. Series: Working Paper Series.

Nick Golding, Timothy W. Russell, Sam Abbott, Joel Hellewell, Carl A. B. Pearson, Kevin van Zandvoort, Christopher I. Jarvis, Hamish Gibbs, Yang Liu, Rosalind M. Eggo, John W. Edmunds, and Adam J. Kucharski. Reconstructing the global dynamics of under-ascertained COVID-19 cases and infections. medRxiv, page 2020.07.07.20148460, July 2020. doi: 10.1101/2020.07.07.20148460. URL https://www.medrxiv.org/content/10.1101/2020.07.07.20148460v1. Publisher: Cold Spring Harbor Laboratory Press.

Jack Healy, Sabrina Tavernise, Robert Gebeloff, and Weiyi Cai. Coronavirus Was Slow to Spread to Rural America. Not Anymore. The New York Times, April 2020. ISSN 0362-4331. URL https://www.nytimes.com/interactive/2020/04/08/us/coronavirus-rural-america-cases.html.

Paul R. Hunter, Alan M. MacDonald, and Richard C. Carter. Water Supply and Health. PLoS Medicine, 7(11), November 2010. ISSN 1549-1277. doi: 10.1371/journal.pmed.1000361. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976720/.

Stefano M. Iacus, Gary King, and Giuseppe Porro. Causal Inference without Balance Checking: Coarsened Exact Matching. Political Analysis, 20(1):1–24, 2012. ISSN 1047-1987, 1476-4989. doi: 10.1093/pan/mpr013. URL https://www.cambridge.org/core/journals/political-analysis/article/causal-inference-without-balance-checking-coarsened-exact-matching/5ABCF5B3FC3089A87FD59CECBB3465C0. Publisher: Cambridge University Press.

Klaudia Jomova and Marian Valko. Advances in metal-induced oxidative stress and human disease. Toxicology, 283(2):65–87, May 2011. ISSN 0300-483X. doi: 10.1016/j.tox.2011.03.001. URL http://www.sciencedirect.com/science/article/pii/S0300483X11000886.

Callum J Jones, Thomas Philippon, and Venky Venkateswaran. Optimal Mitigation Policies in a Pandemic: Social Distancing and Working from Home. Working Paper 26984, National Bureau of Economic Research, April 2020. URL http://www.nber.org/papers/w26984. Series: Working Paper Series.

Dong-Hyun Kim, Young June Choe, and Jin-Young Jeong. Understanding and Interpretation of Case Fatality Rate of Coronavirus Disease 2019. Journal of Korean Medical Science, 35(12), March 2020. ISSN 1011-8934. doi: 10.3346/jkms.2020.35.e137. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7105506/.

Bruce P. Lanphear, Stephen Rauch, Peggy Auinger, Ryan W. Allen, and Richard W. Hornung. Low-level lead exposure and mortality in US adults: a population-based cohort study. The Lancet Public Health, 3(4):e177–e184, April 2018. ISSN 2468-2667. doi: 10.1016/S2468-2667(18)30025-2. URL https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(18)30025-2/abstract. Publisher: Elsevier.

Liu Peter P., Blet Alice, Smyth David, and Li Hongliang. The Science Underlying COVID-19. Circulation, 142(1):68–78, July 2020. doi: 10.1161/CIRCULATIONAHA.120.047549. URL https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.120.047549. Publisher: American Heart Association.

Norfazillah Ab Manan, Azimatun Noor Aizuddin, and Rozita Hod. Effect of Air Pollution and Hospital Admission: A Systematic Review. Annals of Global Health, 84(4):670–678, November 2018. ISSN 2214-9996. doi: 10.29024/aogh.2376. URL http://www.annalsofglobalhealth.org/articles/10.29024/aogh.2376/. Number: 4 Publisher: Ubiquity Press.

Liz Mineo. The impact of COVID-19 on Native American communities, May 2020. URL https://news.harvard.edu/gazette/story/2020/05/the-impact-of-covid-19-on-native-american-communities/. Library Catalog: news.harvard.edu Section: National & World Affairs.

Abolfazl Mollalo, Behzad Vahedi, and Kiara M. Rivera. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of The Total Environment, 728:138884, August 2020. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2020.138884. URL http://www.sciencedirect.com/science/article/pii/S0048969720324013.

Michael Petroni, Dustin Hill, Lylla Younes, Liesl Barkman, Sarah Howard, I. Brielle Howell, Jaime Mirowsky, and Mary B. Collins. Hazardous air pollutant exposure as a contributing factor to COVID-19 mortality in the United States. Environmental Research Letters, 15(9):0940a9, September 2020. ISSN 1748-9326. doi: 10.1088/1748-9326/abaf86. URL https://doi.org/10.1088%2F1748-9326%2Fabaf86. Publisher: IOP Publishing.

Phoebus Rosakis and Maria E. Marketou. Rethinking case fatality ratios for covid-19 from a data-driven viewpoint. The Journal of Infection, 81(2):e162–e164, August 2020. ISSN 0163-4453. doi: 10.1016/j.jinf.2020.06.010. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7286834/.

Laurel A. Schaider, Lucien Swetschinski, Christopher Campbell, and Ruthann A. Rudel. Environmental justice and drinking water quality: are there socioeconomic disparities in nitrate levels in U.S. drinking water? Environmental Health, 18(1):3, January 2019. ISSN 1476-069X. doi: 10.1186/s12940-018-0442-6. URL https://doi.org/10.1186/s12940-018-0442-6.

Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward. Statistical science : a review journal of the Institute of Mathematical Statistics, 25(1):1–21, February 2010. ISSN 0883-4237. doi: 10.1214/09-STS313. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/.

Zhuanlan Sun and Demi Zhu. Exposure to outdoor air pollution and its human health outcomes: A scoping review. PLOS ONE, 14(5):e0216550, May 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0216550. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216550. Publisher: Public Library of Science.

David Switzer and Manuel P. Teodoro. Class, Race, Ethnicity, and Justice in Safe Drinking Water Compliance. Social Science Quarterly, 99(2):524–535, 2018. ISSN 1540-6237. doi: 10.1111/ssqu.12397. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ssqu.12397. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ssqu.12397.

Reis Thebault and Abigail Hauslohner. A deadly 'checkerboard': Covid-19's new surge across rural America. Washington Post, May 2020. ISSN 0190-8286. URL https://www.washingtonpost.com/nation/2020/05/24/coronavirus-rural-america-outbreaks/.

Werner Troesken. Water, Race, and Disease. MIT Press, 2004. ISBN 978-0-262-20148-3. Google-Books-ID: GmgYTwfcIMsC.

OMS US EPA. Drinking Water Contaminants, August 2015. URL https://www.epa.gov/enviro/drinking-water-contaminants. Library Catalog: www.epa.gov.

U.S. Government Accountability Office. Drinking Water: Unreliable State Data Limit EPA's Ability to Target Enforcement Priorities and Communicate Water Systems' Performance. (GAO-11-381), July 2011. URL https://www.gao.gov/products/GAO-11-381.

U.S. Government Accountability Office. Drinking Water: Additional Data and Statistical Analysis May Enhance EPA's Oversight of the Lead and Copper Rule. (GAO-17-424), October 2017. URL https://www.gao.gov/products/GAO-17-424?utm_medium=email&utm_source=govdelivery.

Steven H. Woolf and Paula Braveman. Where Health Disparities Begin: The Role Of Social And Economic Determinants—And Why Current Policies May Make Matters Worse. Health Affairs, 30(10):1852–1859, October 2011. ISSN 0278-2715. doi: 10.1377/hlthaff.2011.0685. URL https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2011.0685. Publisher: Health Affairs.

J.M. Wortham, J.T. Lee, and S. Althomsons. Characteristics of Persons Who Died with COVID-19 — United States, February 12–May 18, 2020. In MMWR Morb Mortal Wkly Rep 2020;69:923-929. 2020. doi: 10.15585/mmwr.mm6928e1. URL DOI:.

Xiao Wu, Rachel C. Nethery, Benjamin M. Sabath, Danielle Braun, and Francesca Dominici. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. medRxiv, page 2020.04.05.20054502, April 2020. doi: 10.1101/2020.04.05.20054502. URL https://www.medrxiv.org/content/10.1101/2020.04.05.20054502v2. Publisher: Cold Spring Harbor Laboratory Press.

Ai-Min Yang, Kenneth Lo, Tong-Zhang Zheng, Jing-Li Yang, Ya-Na Bai, Ying-Qing Feng, Ning Cheng, and Si-Min Liu. Environmental heavy metals and cardiovascular diseases: Status and future direction. Chronic Diseases and Translational Medicine, April 2020. ISSN 2095-882X. doi: 10.1016/j.cdtm.2020.02.005. URL http://www.sciencedirect.com/science/article/pii/S2095882X20300189.

Joshua Graff Zivin and Matthew Neidell. Environment, Health, and Human Capital. Journal of Economic Literature, 51(3):689–730, 2013. ISSN 00220515. URL http://www.jstor.org.pitt.idm.oclc.org/stable/23644825.

Online Appendix

CEM Balance Tables

Below are balance tables for the CEM estimates presented in Table 1. Since each column of Table 1 involves a different treated group, a separate balance table is presented for each column. Variable names in bold are the matching variables used to create the control group with CEM. The unbolded variables were not used in matching but were included as controls when estimating the sample average treatment on the treated (SATT) in a weighted linear regression on the matched sample. The outcome variable (14-day lagged CFR), current cases per million population, and current deaths per million population are also included. In the rightmost two columns, the average of the unmatched control group is included as a reference. This serves two purposes: one, it can be used to determine whether or not CEM was successful in improving the balance of the matched subsample relative to the entire sample (for the matched variables, the matched control group average should be closer to the treated group average than the unmatched control group average), and two, it provides a benchmark to assess how the matched sample differs from the remainder of the sample.

Acute Violations Treated-Control Balance Table

	Matched control		Treated		Difference		Unmatched control	
	Mean	Std. dev.	Mean	Std. dev.	Diff.	p-value	Mean	Std. dev.
% female	51.06	1.20	51.20	1.00	0.1	0.108	51.42	1.98
Average PM2.5, 2000-2016	8.40	2.08	8.41	2.25	0.0	0.986	8.45	2.75
% Black	5.11	7.85	5.16	8.87	0.2	0.822	14.35	18.72
% Native American	0.99	1.00	1.05	1.95	0.1	0.745	1.97	6.91
% Asian	1.30	1.20	1.42	1.85	0.1	0.439	1.52	2.78
% Hispanic	6.28	5.33	6.57	8.40	0.3	0.671	14.05	18.27
% below 2x federal poverty line	8.83	2.30	8.80	2.68	-0.0	0.890	10.51	3.78
% relying on public transit	0.45	0.59	0.46	0.89	0.0	0.770	0.47	1.66
% in households with >= 4 members	7.98	1.03	8.04	1.03	0.1	0.494	7.88	1.59
% with Bachelor's degree or higher	15.64	5.03	16.39	6.42	0.7	0.163	15.27	7.03
Population density (per km^2)	51.69	58.83	73.02	132.40	21.3	0.037	114.54	860.00
% with Census-recorded disability	15.36	4.28	15.70	3.91	0.3	0.256	16.05	4.54
% education and healthcare service workers	5.95	1.69	5.88	1.65	-0.1	0.582	5.72	2.21
% accommodation, and food service workers	5.14	1.89	5.51	1.70	0.4	0.012	5.57	2.47
% agriculture and mining workers	0.57	0.68	0.50	0.83	-0.1	0.438	0.76	1.20
Days since first confirmed case (as of Sep 8, 2020)	161.38	20.98	165.67	17.63	4.3	0.007	160.83	24.49
% with high blood pressure	25.76	5.66	25.45	4.23	-0.3	0.506	26.40	5.62
% over 65 years of age	18.04	2.94	17.91	3.39	-0.1	0.637	18.13	5.36
Outcome variables								
14-day lagged CFR	2.63	2.22	3.30	2.62	0.6	0.002	2.43	2.18
Confirmed cases per million population	12358.36	8354.03	11536.19	8501.20	-822.2	0.228	19216.24	15503.23
Confirmed deaths per million population	285.48	273.43	346.83	341.70	52.5	0.032	446.00	522.88

Cardiovascular-Associated Violations Treated-Control Balance Table

	Matched control		Treated		Difference		Unmatched control	
	Mean	Std. dev.	Mean	Std. dev.	Diff.	p-value	Mean	Std. dev.
% female	50.95	1.13	51.09	1.24	0.1	0.291	51.41	1.81
Average PM2.5, 2000-2016	7.66	2.05	7.63	2.35	-0.0	0.910	8.56	2.57
% Black	4.10	5.39	4.25	6.46	0.2	0.735	12.53	17.61
% Native American	0.98	1.01	1.13	1.79	0.2	0.358	2.05	7.19
% Asian	1.37	1.33	1.58	1.88	0.3	0.246	1.38	2.52
% Hispanic	9.31	5.16	9.51	10.89	0.2	0.885	11.25	15.63
% below 2x federal poverty line	8.48	2.07	8.44	2.71	-0.0	0.899	10.36	3.52
% relying on public transit	0.48	0.55	0.54	1.19	0.1	0.463	0.44	1.50
% in households with >= 4 members	8.05	1.00	8.12	1.23	0.1	0.556	7.88	1.45
% with Bachelor's degree or higher	16.45	5.23	17.12	6.39	0.7	0.342	15.09	6.81
Population density (per km^2)	48.09	58.68	56.17	101.79	8.1	0.486	102.81	740.76
% with Census-recorded disability	15.07	3.94	14.58	3.54	-0.5	0.226	16.31	4.61
% education and healthcare service workers	5.80	1.63	5.62	1.54	-0.2	0.278	5.86	2.11
% accommodation, and food service workers	5.31	1.80	5.42	2.23	0.1	0.722	5.41	2.31
% agriculture and mining workers	0.89	0.70	0.57	1.08	-0.3	0.032	0.66	1.04
Days since first confirmed case (as of Sep 8, 2020)	161.23	18.99	164.96	23.27	3.7	0.119	160.87	23.64
% with high blood pressure	24.79	5.42	25.31	4.12	0.5	0.398	26.91	5.64
% over 65 years of age	18.24	3.05	18.02	3.99	-0.2	0.621	18.07	4.88

Outcome variables

14-day lagged CFR	2.79	2.37	3.24	2.44	0.5	0.034	2.39	2.21
Confirmed cases per million population	11361.02	7684.56	10906.03	8180.20	-455.0	0.606	17817.26	14614.24
Confirmed deaths per million population	274.52	250.31	312.56	316.32	47.0	0.150	408.80	503.81

All Health-Based Violations Treated-Control Balance Table

	Matched control		Treated		Difference		Unmatched control	
	Mean	Std. dev.	Mean	Std. dev.	Diff.	p-value	Mean	Std. dev.
% female	51.28	1.36	51.33	1.28	0.1	0.441	51.12	2.01
Average PM2.5, 2000-2016	8.81	2.28	8.81	2.21	-0.0	0.981	8.00	3.05
% Black	7.53	12.68	7.23	12.73	-0.2	0.813	11.47	16.21
% Native American	0.68	1.15	0.69	1.17	0.0	0.829	3.03	9.97
% Asian	1.05	1.44	0.97	1.50	-0.1	0.288	2.48	4.16
% Hispanic	5.06	6.19	4.99	6.64	-0.1	0.803	16.72	18.23
% below 2x federal poverty line	9.32	2.56	9.32	2.51	0.0	0.988	9.88	4.09
% relying on public transit	0.32	0.57	0.30	0.67	-0.0	0.424	0.82	2.10
% in households with >= 4 members	7.90	1.11	7.84	1.10	-0.1	0.328	7.91	1.78
% with Bachelor's degree or higher	14.39	5.38	14.46	5.45	0.1	0.800	17.79	8.44
Population density (per km^2)	43.64	85.98	42.75	84.53	-0.9	0.834	180.33	561.38
% with Census-recorded disability	15.91	3.97	16.92	4.31	1.0	<.001	14.76	4.60
% education and healthcare service workers	6.21	1.73	5.98	1.81	-0.2	0.050	5.42	2.18
% accommodation, and food service workers	5.17	1.95	5.09	2.09	-0.1	0.472	6.05	2.58
% agriculture and mining workers	0.47	0.62	0.60	0.85	0.1	0.003	0.66	1.22
Days since first confirmed case (as of Aug 18, 2020)	161.25	19.54	159.72	22.33	-1.5	0.283	163.48	26.16
% with high blood pressure	26.88	5.43	27.29	5.49	0.4	0.336	25.06	5.17
% over 65 years of age	18.45	3.52	18.43	3.39	-0.0	0.916	17.92	6.36
Outcome variables								
14-day lagged CFR	2.95	2.16	2.87	2.30	-0.1	0.508	2.32	2.19
Confirmed cases per million population	13138.91	10455.24	12895.53	10707.86	-243.4	0.693	17933.58	14330.53
Confirmed deaths per million population	331.13	339.84	345.09	384.32	7.6	0.682	403.06	506.36