XINN, Volume *1*

Supplemental Information

Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States

Donghai Liang, Liuhua Shi, Jingxuan Zhao, Pengfei Liu, Jeremy A. Sarnat, Song Gao, Joel Schwartz, Yang Liu, Stefanie T. Ebelt, Noah Scovronick, and Howard H. Chang

Supplementary Appendix for

Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States

Donghai Liang[1]*, PhD, Liuhua Shi[1]*, ScD, Jingxuan Zhao[2], MPH, Pengfei Liu[3], PhD, Jeremy A Sarnat[1], ScD, Song Gao[4], PhD, Joel Schwartz[5], PhD, Yang Liu[1], PhD, Stefanie T Ebelt[1], ScD, Noah Scovronick[1], PhD, Howard H Chang[6], PhD

1 Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

2 Surveillance and Health Services Research Program, American Cancer Society, Atlanta, Georgia, USA

3 School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

4 GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

5 Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

6 Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Correspondence to
Donghai Liang, PhD
Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University
1518 Clifton Rd NE, Atlanta, GA 30322
Telephone: (404)712-9583
E-mail: donghai.liang@emory.edu

* These authors contributed equally to this work

Supplementary Table

Table S1 Moran's I test for spatial autocorrelation in residuals from tri-pollutant models for COVID-19 Case-fatality Rate and Mortality Rate for each US state

	Case-fatality Rate		Mortality Rate	
State	**Moran's I**	***p*-value**	**Moran's I**	***p*-value**
Alabama	0.21	0.01	0.18	0.01
Alaska	-0.02	0.63	-0.02	0.63
Arizona	0.07	0.35	0.10	0.27
Arkansas	0.11	0.07	0.08	0.19
California	0.01	0.77	0.01	0.72
Colorado	0.06	0.33	-0.05	0.62
Connecticut	0.17	0.09	0.24	0.07
Delaware	-	-	-	-
District of Columbia	-	-	-	-
Florida	-0.06	0.57	-0.03	0.89
Georgia	0.17	0.01	0.10	0.02
Hawaii	-	-	-	-
Idaho	0.02	0.21	0.02	0.28
Illinois	0.03	0.45	0.04	0.37
Indiana	0.01	0.69	-0.03	0.73
Iowa	0.11	0.03	0.07	0.15
Kansas	0.04	0.37	0.13	0.02
Kentucky	0.09	0.08	0.08	0.10
Louisiana	0.21	0.01	0.17	0.02
Maine	0.02	0.23	-0.09	0.70
Maryland	0.07	0.51	0.19	0.15
Massachusetts	0.09	0.41	-0.10	0.91
Michigan	0.08	0.20	0.08	0.16
Minnesota	0.02	0.65	0.04	0.44
Mississippi	0.06	0.28	0.05	0.35
Missouri	0.04	0.42	0.02	0.68
Montana	0.11	0.12	0.10	0.02
Nebraska	-0.02	0.84	-0.02	0.80
Nevada	-0.08	0.82	-0.06	0.92
New Hampshire	0.41	0.01	0.13	0.06
New Jersey	0.54	0.01	0.42	0.01
New Mexico	-0.17	0.16	-0.06	0.78
New York	-0.02	0.95	0.07	0.28
North Carolina	0.15	0.01	0.12	0.04
North Dakota	-0.13	0.17	-0.15	0.09
Ohio	0.09	0.12	0.06	0.18
Oklahoma	0.11	0.10	0.05	0.34
Oregon	0.27	0.01	0.28	0.01
Pennsylvania	0.14	0.03	0.41	0.01
Rhode Island	0.22	0.24	0.29	0.15
South Carolina	-0.07	0.63	-0.10	0.42
South Dakota	-0.06	0.36	-0.10	0.22
Tennessee	-0.05	0.51	-0.01	0.99
Texas	0.10	0.01	0.10	0.01
Utah	0.10	0.23	0.12	0.16
Vermont	0.07	0.38	0.09	0.35
Virginia	0.07	0.20	0.07	0.18
Washington	0.00	0.76	0.20	0.04
West Virginia	0.08	0.26	-0.01	0.90
Wisconsin	-0.06	0.45	-0.05	0.58
Wyoming	-0.08	0.74	-0.07	0.81

Appendix-Technical Appendix

COVID-19 case-fatality rate: We obtained the number of daily county-level COVID-19 confirmed cases and deaths that occurred from January 22, 2020, the day of first confirmed case in the US, through July 17, 2020 in the US from three databases: the New York Times[1], the USAFACTS[2], and 1Point3Acres.com[3]. Each of these databases provide real-time data by retrieving information on official reports from state and local health agencies. After data acquisition from these sources, we compared the number of confirmed COVID-19 cases and deaths in each US county (identified by the Federal Information Processing Standards, FIPS code) across all databases for accuracy and consistency. In case of discrepancy, county-level case and death number were corrected by manually checking the data reported from the corresponding state and local health department websites. We calculated county-level COVID-19 case-fatality rate by dividing the number of deaths over the number of people diagnosed with COVID-19 for each US county with at least 1 or more confirmed case, as reported by July 17, 2020. Of all the data reported as of July 17, 2020, confirmed cases and deaths with unassigned counties were excluded in the analysis.

Air pollution: Three major criteria ambient air pollutants were included in the analysis, including NO_2, a traffic-related air pollutant and a major component of urban smog, $PM_{2.5}$, and O_3. We recently estimated daily ambient $PM_{2.5}$, NO_2, and O_3 levels at 1 km^2 spatial resolution across the contiguous US an ensemble machine learning model with ground measurements, satellite-data products, chemical transport model output, meteorological and land-use information as predictors[4,5]. We calculated the daily average for each county based on all covered 1 km^2 grid cells, and then further calculated the annual mean (2010-2016) for $PM_{2.5}$ and NO_2 and the warm-season mean (2010-2016) for O_3, defined as May 1 to October 31, as surrogates for long-term $PM_{2.5}$, NO_2, and O_3 exposures, respectively. More recent exposure data were not available at the time of this analysis. However, county-specific mean values of an air pollutant among different years are highly correlated.

Covariates and Data sources on covariates: We adjusted for three county-level healthcare capacity covariates, including the number of intensive care unit (ICU) beds, hospital bed, and active medical doctor per 1000 people. Number of ICU beds were based on Kaiser Health News analysis of 2018 and 2019 hospital cost reports filed to the Centers for Medicare & Medicaid Services. Numbers of active medical doctors and hospital beds of 2017 were obtained from the Area Health Resources Files[6]. State-level number of COVID-19 tests performed up to July 17, 2020 was derived from the Covid Tracking Project, based on which we calculated the positive rate in each state, i.e. the percentage of tests performed that are positive for COVID-19. The travel distance mobility data were released from the Descartes Labs and mapped by the GeoDS Lab using anonymized location data from smartphones[7,8]. The travel mobility index was a measure to compare the daily individual-level travel distance pattern to that in February. To enhance privacy, individual data are de-identified and aggregated to the county level. We calculated the county-level mean mobility index from March 1, 2020 to July 17, 2020 to represent the dramatic mean human mobility changes in reaction to the COVID-19. County-level socioeconomic status (SES) in 2015 was measured by social deprivation index (SDI)[9], which is a composite measure of area-level deprivation based on seven characteristics, including income, education, employment, housing, household characteristics, transportation, and demographics. SDI has commonly served as an area-level composite measure of SES in other studies of health and health outcomes. County-level sociodemographic covariates in 2017 such as percentage of elderly (age≥60), percentage of male, percentage of Black, and percentage of Hispanic were derived from Area Health Resource Files, and population density was derived from the 2018 US Census. County-level behavioral risk factors, including population mean BMI (an indicator of obesity) and percentage of ever smokers, were derived from the 2011 US CDC Behavioral Risk Factor Surveillance System (BRFSS). From Phase 2 of the North American Land Data Assimilation System (NLDAS-2), we acquired hourly 1/8th degree gridded near-surface air temperature and specific humidity data from January 22, 2020 through July 17, 2020[10], based on which we calculated the mean temperature and relative humidity for each 1/8th degree grid. We linked each county's centroid to the nearest 1/8th degree grid and assigned the mean temperature and relative humidity.

Statistical methods: We fit zero-inflated negative binomial mixed models (ZINB) to examine the associations between long-term exposure to $PM_{2.5}$, NO_2, and O_3 and COVID-19 case-fatality rate or mortality. The ZINB model comprises a negative binomial log-linear count model and a logit model for predicting excess zeros. The former was used to describe the associations between air pollutants and COVID-19 case-fatality rate among counties with at least one reported COVID-19 case. The latter can account for excess zeros in counties that have not observed a COVID-19 death as of July 17, 2020. We fit single-pollutant, bi-pollutant, and tri-pollutant models, with all analyses conducted at the county level. For the negative binomial count component, results are presented as percent change in case-fatality rate or mortality rate per interquartile range (IQR) increase in each air pollutant concentration. IQR was calculated on national levels. Similar results are presented as odds ratio for the excess zero component. We included a random intercept for each state because observations within the same state tend to be correlated due to similar COVID-19 responses, quarantine and testing policies, healthcare capacity, sociodemographic, and meteorological conditions.

As different testing practices may bias outcome ascertainment, we adjusted for state-level COVID-19 test positive rate (i.e. high positive rate might imply that the confirmed case numbers were limited by the ability of testing, and the case-fatality can be biased high). To model how different counties may be at different time points of the epidemic curve (i.e., phase-of-epidemic), we adjusted for days both since the first case and since the 100^{th} case (i.e., case counts reaching 100) within a county through July 17 as a measure of epidemic timing. In addition, we considered potential confounding by county-level healthcare capacity, population travel mobility index, sociodemographic, SES, behavior risk factors, and meteorological factors. Because county-specific population densities span 5 orders of magnitude, we adjusted for density using a logarithmic transformation. To control for potential residual spatial trends and confounding, we included spatial smoothers within the model using natural cubic splines with 5 degrees freedom for both county centroid latitude and longitude. We further calculated Moran's I of the standardized residuals of tri-pollutant main models for each state, to examine the presence of spatial autocorrelation in the residuals.

Sensitivity analyses: We also conducted a series of sensitivity analyses to test the robustness of our results to outliers, confounding adjustment, and epidemic timing (Figures 4 and 5). Given that New York city has far higher COVID-19 cases and deaths than any other regions in the US, which can be a very influential observation, we excluded all five counties within New York city and repeated the analysis. In another set of sensitivity analyses, we restricted the study only to the most recent 4 weeks (June 20 to July 17), when the case count and death count may be more reliable and accurate than earlier periods and when COVID-19 tests were more available. We also conducted sensitivity analysis by using air pollution data averaged between 2000 to 2016. To assess the impact of potential bias of individual covariates, we fit models by omitting a different set of covariates for each model iteration while comparing effect estimates. Statistical tests were 2-sidedand statistical significance was determined with an alpha of 0.05. All statistical analyses were conducted used R version 3.4.

Technical Appendix Reference

1. The New York Times; 2020. https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html. Accessed 04/19/2020.
2. USAFACTS; 2020. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/. Accessed 04/19/2020.
3. 1Point3Acres.com; 2020. https://coronavirus.1point3acres.com/. Accessed 04/19/2020.
4. Di Q, Amini H, Shi L, et al. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment international.* 2019a;130:104909.
5. Di Q, Amini H, Shi L, et al. Assessing NO2 Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. *Environmental science & technology.* 2019b;54(3):1372-1384.
6. Area Health Resources Files; 2019. https://data.hrsa.gov/topics/health-workforce/ahrf. Accessed 04/19/2020.
7. Warren, M.S. and Skillman, S.W., 2020. Mobility Changes in Response to COVID-19. *arXiv preprint arXiv:2003.14228*.
8. Gao, S., Rao, J., Kang, Y., Liang, Y. and Kruse, J., 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *Available at SSRN 3570145*.
9. Social Deprivation Index. Robert Graham Center; 2019. https://www.graham-center.org/rgc/maps-data-tools/sdi/social-deprivation-index.html. Accessed 04/25/2019.
10. Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, J. Geophys. Res., 117, D03109, doi:10.1029/2011JD016048.
11. Fang R. *Zero-Inflated Negative Binomial (Zinb) Regression Model for over-dispersed count data with excess zeros and repeated measures, an application to human microbiota sequence data*, University of Colorado Anschutz Medical Campus. Strauss Health Sciences Library; 2013.