



A case-crossover analysis of short-term PM_{2.5} exposure and COVID19-related hospitalizations in the National Clinical Cohort Collaborative



Cavin K. Ward-Caviness¹, Alyssa Platt^{2,3}, Elliott Fisher², Anne M. Weaver¹, Michelle L. Bell^{1,4} & Mercedes A. Bravo¹✉

Abstract

Background Studies show associations between air pollution exposure and coronavirus 2019 (COVID19) hospitalizations, but have not substantially explored regional differences. In this study, we estimate associations between shorter-term exposure to fine particulate matter (PM_{2.5}) and hospitalization among individuals with SARS-CoV-2 infection.

Methods This study utilized data from 72,385 patients (78,504 hospitalizations) with a hospital-confirmed SARS-CoV-2 infection between January 1, 2020 and December 31, 2020. Daily PM_{2.5} concentrations from ground-based monitors were averaged to generate 2, 5, and 21-day average exposures prior to hospitalization. We used a time-stratified case-crossover approach to estimate associations between PM_{2.5} and COVID19-related hospitalizations in 57 Core Based Statistical Areas (CBSAs) across the United States (US). We subsequently conducted nationwide and region-specific random effects meta-analysis.

Results In the random effects meta-analysis, a 1 $\mu\text{g}/\text{m}^3$ increase in 2, 5, and 21-day average PM_{2.5} are associated with a 0.61% (95% Confidence Interval [CI] = 0.12, 1.11); 0.91% (CI = 0.15, 1.67); and 0.04% (CI = -2.70, 2.85) increase in COVID19-related hospitalization risk, respectively. We observe substantial heterogeneity in the associations by region with the largest adverse associations in the South.

Conclusions Higher concentrations of PM_{2.5} are associated with higher risk of COVID19-related hospitalizations. Given the geographic heterogeneity observed, studies exploring factors, such as PM_{2.5} exposure, that could explain differences in COVID19 risks may help to understand the COVID19 pandemic and aid in preparing for future ones.

Plain Language Summary

Exposure to air pollution is shown to have a negative impact on human health. For example, air pollution increases the risk of cardiovascular- and respiratory-related hospitalizations and deaths. Little is known about how air pollution might affect the health of individuals with COVID19. Using hospitalization data from 57 cities across the United States (US), we assess whether individuals who tested positive for COVID19 were more likely to be hospitalized after multiple days of higher air pollution exposures. We find that higher levels of fine particulate matter, a common air pollutant, are associated with increased likelihood of COVID19-related hospitalization, and that the relationship between particulate matter and COVID19 differs across the US regions and cities we examine. The short- and long-term health impacts of air pollution exposure in individuals with COVID19 merits further research and should be considered in public health interventions and planning health care capacity.

The public health impact of the SARS-CoV-2 global pandemic has been tremendous, and specific populations have been disproportionately impacted¹. In the United States (US) alone, there have been 96 million confirmed COVID19 cases and over 1 million COVID19-related deaths². At different phases of the pandemic, populations including, elderly individuals³ and rural communities⁴ in the US have borne a

disproportionate burden of the health consequences of COVID19 in terms of viral infections, hospitalizations, and mortality^{5,6}. Many of these same populations may also be exposed to environmental hazards such as air pollution⁷, vulnerable to adverse health effects of environmental exposures^{8,9}, or both¹⁰. As a ubiquitous and potentially modifiable exposure^{11,12}, understanding the role of air pollution with respect to

¹Center for Public Health and Environmental Assessment, US Environmental Protection Agency, Chapel Hill, NC, USA. ²Global Health Institute, School of Medicine, Duke University, Durham, NC, USA. ³Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA. ⁴School of the Environment, Yale University, New Haven, CT, USA. ✉e-mail: mercedes.bravo@duke.edu

COVID19 may help address the public health burden of the COVID19 pandemic^{11,12}.

Exposure to ambient air pollution adversely affects respiratory, cardiovascular, and immune health^{13–16} and may exacerbate severity of SARS-CoV-2 infection, worsen prognosis, and/or increase risk of long-term health impacts¹⁷. A limited number of US-based studies have examined associations between air pollution and health among individuals with COVID19 (COVID19 positive/COVID19+); generally, findings suggest air pollution is related to health outcomes (e.g., hospitalizations, mortality) in those with COVID19^{18–22}. For example, a study in Southern California found that near-roadway air pollution exposures in the year prior to COVID19 diagnosis were associated with hospitalization and mortality among COVID19+ individuals²⁰. Similarly, PM_{2.5} exposure in the year prior to diagnosis was associated with increased risk of hospital admission and mortality among COVID+ individuals in New York City²¹. There have been a few studies of COVID19 and shorter-term air pollution exposure. Kim et al (2022) observed positive associations between 21-day average PM_{2.5} and COVID19 mortality in Chicago¹⁹. A similar study done in Queens, New York but only examining the time period of March 1 – April 20, 2020 also chose to examine 21-day average PM_{2.5} and found a negative association between PM_{2.5} and COVID19 cases and mortality. This study also examined ozone and found a negative association with mortality but a positive association with COVID19 case counts in Queens County²³. This study did have the limitation of only having county-level information. In a nationwide study of county-level data, positive associations were seen between PM_{2.5} (up to 14-day lag) and SARS-CoV-2 infection²⁴. Interestingly, associations were stronger after excluding counties in New York state, which they excluded due to the severity of the pandemic and the fact that including severely affected areas in previous, similar analyses had been shown to strongly alter observed associations.

Most studies of COVID19 and air pollution (including those cited above) have not used individual-level health data, or have been conducted in geographically constrained, predominantly urban samples that do not explore regional differences in associations, which may be driven by demographics, air pollution exposure distribution or composition, and other factors^{21,22,24,25}. To date, nationwide investigations of short-term ambient air pollution exposure and COVID19-related health outcomes in the US are limited to ecological studies of particulate matter $\leq 2.5 \mu\text{m}$ (PM_{2.5}) and COVID19-related case counts or mortality^{24,26,27}. However, ecological studies cannot adjust for individual-level risk factors and are subject to the ecological fallacy²⁷. Relationships between air pollution and health in COVID19+ individuals have not yet been studied with individual-level data at a national scale in the US, and this represents a critical gap in research.

The National Clinical Cohort Collaborative (N3C) (previously the National COVID19 Cohort Collaborative) was created to accelerate understanding of COVID19 by developing a centralized, harmonized, secure, national-in-scope electronic health record (EHR) data resource from health systems across the US²⁸. The N3C includes patient-level information on COVID19 test results, medications, co-morbidities, vital signs, laboratory values, demographics, and 3 or 5-digit ZIP code of residence on >6 million people in the US who have tested positive for COVID19 since March 2020. While N3C has gained traction as a key data resource, evidenced by a growing list of publications^{28–38}, it has not yet been utilized to examine health effects of environmental exposures such as air pollution.

We hypothesize that short-term PM_{2.5} exposure is associated with increased risks of hospitalization among individuals with a recent, confirmed SARS-CoV-2 infection. Given the likelihood of geographic heterogeneity in such associations we also explored region specific associations. Geographic heterogeneity in associations could arise due to differing composition of PM_{2.5}, differences in the demographic makeup of populations or even differences in behaviors that could act to alter air pollution-related health risks. To conduct these analyses, we linked individual-level data from N3C participants with air pollution exposure data and examined associations between air pollution and COVID19 by Core Based Statistical

Area (CBSA, an indicator for metropolitan and surrounding areas). Using a time stratified case-crossover design, we evaluate whether individual-level characteristics (e.g., age, pre-existing co-morbidities) modify observed associations, a question that has not been thoroughly explored in previous studies that lack individual-level data. Results from each CBSA are meta-analyzed to produce nationwide and region-specific associations. Overall, we find evidence that PM_{2.5} exposure is associated with increased COVID19-related hospitalization risk and observe substantial heterogeneity in PM_{2.5}-COVID19 associations by region, with the largest associations in the South.

Methods

Health data

National Clinical Cohort Collaborative (N3C). The National Clinical Cohort Collaborative (N3C) or “N3C Enclave” is a centralized, harmonized, secure, national-in-scope clinical EHR data resource with embedded analytical capabilities. It houses EHR data from 78 distinct health systems (e.g., hospitals, hospital-affiliated primary care providers, etc.) across the US²⁸. All health systems included in N3C are located in the contiguous 48 states. However, due to the rapidly unfolding COVID19 pandemic there was not a systematic selection of hospital systems based on a priori selection criteria as might be ideally done. Instead, the N3C represents a nationwide surveillance of COVID19 based on those hospitals that agreed to participate. Thus, the N3C has broad geographical representation across the US but should not be considered to be a perfectly representative sample of the US. For participating institutions, N3C includes COVID19+ individuals (i.e., individuals who tested positive for SARS-CoV-2 in a hospital setting or had symptoms consistent with COVID19 from January 1, 2020 onward) and matched controls (i.e., individuals that were not COVID19+ but matched on race, sex, and hospital system to someone with COVID19 included in N3C). N3C includes records of inpatient and outpatient visits, and variables such as COVID19 test results and dates, vaccination status, symptoms, medications, co-morbidities, sex, age, race, ethnicity, and 5-digit ZIP code (for last known residence), among others²⁸. The N3C also has “lookback” data, i.e., encounters in the same source health system beginning on or after January 1, 2018. N3C uses centrally maintained shared logic sets for common diagnostic and phenotype definitions. EHR data in the N3C Enclave are harmonized in the Observational Medical Outcomes Partnership (OMOP) common data model, v5.3.1^{28,39}. Additional information on the design and infrastructure of the N3C is available elsewhere³⁸.

Identifying COVID19-positive individuals. The N3C definition of COVID19+ is publicly available at <https://national-clinical-cohort-collaborative.github.io/guide-to-n3c-v1/> and includes patients in the health systems with any encounter after January 1, 2020, with (i) one of a set of a priori-defined SARS-CoV-2 laboratory tests; (ii) a strong positive diagnostic code; or (iii) two weak positive diagnostic codes during the same encounter or on the same date before May 1, 2020 (i.e., between January 1, 2020 and May 1, 2020)⁴⁰. In the construction of our COVID19+ patient hospitalization dataset, individuals age ≥ 18 years with a positive nucleic acid amplification test (e.g., PCR) or positive antigen test were identified as having a COVID infection. Thus, our definition of “COVID19+” is more stringent than that required for inclusion in the N3C Enclave as COVID19+²⁸. Specifically, we do not include individuals solely based on positive diagnostic codes or positive antibody test, which may detect remote infection rather than active COVID19 infection.

Identifying COVID19-related hospitalizations. Following the work of others⁴¹, we define a COVID19-related hospitalization as an inpatient hospitalization with an admission date that is 16 days or less following a positive PCR or antigen test, or an inpatient admission date that is one day before a positive test.

Air pollution exposure

Air pollution data consist of daily average concentrations of particulate matter (PM_{2.5}) recorded at ambient air quality monitors at National, State and Local Air Monitoring Stations. We generated daily estimates of air pollutant concentrations at 5-digit ZIP Code Tabulation Areas (ZCTA) using daily monitoring data obtained from the US Environmental Protection Agency (EPA). A pollution monitor was assigned to a ZCTA if: (i) the monitor was located within the ZCTA; or (ii) the monitor was within 20 km of the ZCTA centroid. The 20 km distance, and general methodology, is based on previous studies, and is designed to maximize the number of individuals included in the study for whom we can reasonably estimate pollution exposure⁴². Daily ZCTA-level PM_{2.5} concentrations were obtained from all monitors which provided 24 h PM_{2.5} measurements. For hourly monitors we required that a monitor have recorded PM_{2.5} for each of 24 hours in a given day for data to be considered valid for that day to offset the risk of using PM_{2.5} values from a monitor that was malfunctioning and to avoid bias towards portions of the day. For a given ZCTA, we averaged PM_{2.5} values across all monitors and assigned the PM_{2.5} value for that day as the average across all monitors.

Individuals residing in a ZCTA with PM_{2.5} monitoring data (i.e., at least one monitor located within the ZCTA or within 20 km of the ZCTA centroid) were assigned PM_{2.5} exposures based on their ZIP code of residence using an existing ZIP-to-ZCTA crosswalk (<https://udsmapper.org/zip-code-to-zcta-crosswalk/>). For each inpatient encounter, daily ZCTA-level PM_{2.5} concentrations were averaged to create 21-day lagged exposure estimates for the 21 days preceding the individual's hospitalization. Additionally, 2-day and 5-day lagged exposure estimates were also generated. For 2- and 5-day average exposures we required the relevant monitor(s) to have measurements for all days in the averaging period; for 21-day averages, a maximum of 3 missing days were allowed.

Covariate Data

Meteorological data. Meteorological conditions such as temperature and humidity are potential confounders of the association between short-term air pollution exposure and hospitalizations. Daily estimates of temperature and humidity were obtained from the Parameter-elevation Relationships on Independent Slopes Model (PRISM). PRISM uses variables such as measured temperature and humidity at weather stations, location, elevation, coastal proximity, and topographic features to estimate temperature, humidity, and precipitation across the conterminous US⁴³⁻⁴⁵ and has been used in previous epidemiological studies^{46,47}. Daily estimates of temperature, precipitation, and humidity reported at the 5-digit ZIP code level were joined with individuals based on their ZIP code of residence. Meteorological covariates were chosen to adjust for potential confounding by weather patterns which may covary with both PM_{2.5} and COVID19-related hospitalizations.

US Census data. The Census classifies geographic areas into Core Based Statistical Areas (CBSAs) as well as Regions and Divisions. CBSAs consist of 560 micropolitan and 362 metropolitan statistical areas based on location within county or county equivalents in the US. Metropolitan statistical areas have at least one urbanized area with a population of 50,000 or more with adjacent territory measured by commuting ties. Micropolitan statistical areas have at least one urban cluster with a population between 10,000 and 50,000 with socially and economically integrated adjacent territory as measured by commuting ties⁴⁸. For CBSAs in the analysis, we obtained the percentage of individuals with incomes below the federal poverty line and percentage of individuals with a high school or more education from the 2020 Census. Census regions consisted of North, South, Midwest, and West as defined by the US Census (https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf).

Urbanicity. Rural-Urban Commuting Area (RUCA) codes were used to classify ZIP codes as urban, suburban, or rural using 10 primary and 21 secondary designations⁴⁹.

Counts of COVID19 cases. National daily county-level COVID19 community transmission data, including counts of new COVID19 cases and running totals from January 22, 2020 forward were obtained from the USA Facts database^{50,51}.

Statistical analysis

Differences between inpatient hospitalizations included in the final analytic cohort and those that were excluded were characterized using stratified descriptive statistics tables. We report means and standard deviations for continuous variables and counts and percentages for categorical variables. Reported p-values are based on t-tests (for continuous variables) and Chi-square tests (for categorical variables). In large datasets such as those in N3C, p-values may be deceptively small. Thus, standardized mean differences (SMD) were also calculated to quantify the magnitude of differences between strata^{52,53}. The level of observation for the case-crossover analysis was the patient encounter, meaning individual patients could enter the analysis at multiple timepoints. We restricted our analyses to 2020 as this was the timeframe of available air pollution data in N3C at the time of data analysis and as this was the timeframe prior to widespread vaccinations. As individual-level vaccination information is only available for a small subset of N3C we considered this restriction necessary.

Exposures considered were 2, 5, and 21-day average PM_{2.5} concentrations. Each exposure was associated with COVID19-related hospitalization using conditional logistic regression models (*survival* package) implemented in R software version 3.5.1⁵⁴⁻⁵⁶, within a time-stratified case-crossover design. For each patient, the case day was defined as the date of admission for each COVID19-related inpatient hospitalization. Repeat admissions determined to be part of the overall same hospitalization (termed macrovisit) are grouped into a single hospitalization within N3C, preventing repeat analyses of the same overall hospitalization. Details on the creation of the macrovisit concept in N3C have been published elsewhere⁵⁷. Control days were defined as the same day of the week and calendar month of the case day (index). Case-control groups were accounted for as strata in the regressions while repeated measures for multiple hospitalizations for the same patient were accounted for with cluster robust standard errors, with the cluster defined as the individual patient. While time-invariant variables do not confound in case-crossover studies by design, exposures that are time-varying (e.g., meteorology) can still confound associations. Therefore, we adjusted for mean temperature, dewpoint, and precipitation corresponding to the case and control dates within the models.

To account for potential confounding by local COVID19 dynamics, we included a one-week lagged average count of new infections in the county of residence starting from the date 16 days prior to the index case (or control date), with the 16-day lookback meant to account for the period of time in which the individual may have been infected prior to hospitalization. As the COVID19 pandemic did not begin everywhere at the same time we also adjusted for a CBSA-specific measure of days from the first case reported in the patient's CBSA.

Initial analyses indicated significant heterogeneity in associations across the nation and potentially region-specific associations which have not been previously explored. To address this, we performed CBSA-specific analyses followed by random effects meta-analysis (as implemented in the *metafor* R package)⁵⁸ to account for heterogeneity in the associations across CBSAs. For data privacy and regression stability, we required that CBSAs have at least 50 index cases to be included in the CBSA-specific analysis. This follows similar approaches used in the literature to understand spatial heterogeneity in case-crossover analyses^{59,60}. Point estimates and 95% confidence intervals from the random effects meta-analysis were compared with those from the individual-level data analysis (which did not account for heterogeneity between CBSAs) to understand how heterogeneity impacts observed associations. We also performed Census region (Northeast, Midwest, South, and West) specific random effects meta-analyses to address the unexplored research gap of region-specific associations between PM_{2.5} and COVID19-related hospitalizations. For Census region meta-analyses

those CBSAs that spanned multiple census regions were placed within the census region containing the metropolitan core.

In addition to region specific analyses we also performed subgroup analyses to ascertain whether relationships between short-term $PM_{2.5}$ exposure differed for individuals with risk factors for severe COVID19 infection, including: race (Black, White, Asian or Pacific Islander [API], Alaskan Native or American Indian [ANAI], and Other or Multiple Races), age (<65 years, ≥ 65 years); and pre-infection histories of medical conditions, namely, cardiovascular diseases (CVD), respiratory disease (RD), or metabolic disease (MD). For CVD, we included a broad spectrum of cardiac and vascular diseases to broadly capture vascular dysfunction which may exacerbate COVID19-related health risks. CVD was defined as the presence of any of the following chronic cardiac or vascular diseases: cardiomyopathy, cerebrovascular disease, congestive heart failure, coronary artery disease, hypertension, previous myocardial infarction, or peripheral vascular disease. RD was defined as the presence of emphysema or chronic bronchitis. MD was defined as the presence of Type 1 or 2 diabetes or obesity. CBSA-specific subgroup analyses were meta-analyzed to provide overall and Census region-specific results requiring 50 index cases per CBSA as before. To formally test for a difference in the regression coefficients between the subgroups we used a two-sample Z-test as done in similar analyses of COVID19 health effects and $PM_{2.5}$ exposure¹⁹. Initial analyses included continuous confounders (temperature, dewpoint, precipitation, county of new COVID19 infections, and days since first reported COVID19 case in CBSA) as linear terms. However, it is possible these confounders have non-linear relationships with the outcome, thus as a sensitivity analysis we re-ran the models using a natural cubic spline with three degrees of freedom for all continuous confounders except precipitation. We attempted to evaluate models with a natural cubic spline for precipitation, but such models did not converge and thus precipitation was left as a linear term even in the sensitivity analyses.

We used a two tailed $\alpha < 0.05$ as evidence of statistical significance for subgroup differences. All results are reported as the percent increase in COVID19-related hospitalizations per $1 \mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ along with the associated 95% confidence interval (95% CI). Code files for cohort construction and linking to environmental data are archived in Zenodo, an open research repository that facilitates preservation and dissemination of research output (<https://doi.org/10.5281/zenodo.15990139>)⁶¹.

The N3C data transfer to the National Center for Advancing Translational Sciences (NCATS) is performed under a Johns Hopkins University Reliance Protocol # IRB00249128 or individual site agreements with US National Institutes of Health (NIH). The N3C Data Enclave is managed under the authority of the NIH; information can be found at <https://ncats.nih.gov/n3c/resources>. The Duke University Health System Institutional Review Board completed an administrative review and found the consent form for the study to be consistent with DUHS IRB policies. As the research posed no more than minimal risk and informed consent of individuals was waived. Additionally, we obtained an approved data use request (DUR) with N3C (DUR B5-D391).

Results

Study cohort description

In the N3C, there were 1,563,240 observed COVID-19 cases in 2020. Of these, 174,383 individuals were hospitalized (11.2%). Of these 174,383 hospitalized individuals, 19,747 were excluded for missing age or invalid age data, age <18 years, or hospitalization occurring in 2021 or later (and thus outside of the air quality data available for our study). A further 9864 were excluded due to the hospital system performing date shifting to obscure exact dates which were required to link hospitalizations to air quality data. As air quality data was linked by residential ZIP code, 66,081 individuals were excluded for lack of available ZIP code data or residing in a ZIP code >20 km from an air quality monitor. Finally, 1885 individuals were excluded due to missing meteorology confounders. No other confounders had missing data. This left us with a final sample of 78,504 COVID-19 cases from 48 participating hospital systems with all 4 Census regions and 57 CBSAs

represented. A STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) diagram of patient exclusions is given in Fig. 1 and a map of the COVID-19 cases is given in Supplemental Fig. 1. A density plot of the timing of COVID19-related hospitalizations is given in Supplemental Fig. 2 with a further breakdown by CBSA given in Supplemental Fig. 3.

The average age at the time of COVID19 positive test was 62 years ($SD = 18$ years). Encounters were approximately equal across the sexes (47.9% female, 52.1% male). Most encounters involved patients that were either White (50.9%) or Black (23.6%), and 20.1% of encounters were for patients identified as Hispanic or Latino in their health record. Most encounters were for patients residing in the Midwest (47.6%) and Northeast (33.9%) Census regions and the vast majority (99.0%) lived in ZIP code areas that were classified as urban based on RUCA codes. A majority of encounters (55.1%) involved patients that had evidence of pre-existing CVD, with 41.1% having pre-existing MD (Table 1). The median 2-day, 5-day, and 21-day average $PM_{2.5}$ among study participants was 7.0, 7.1, and $7.6 \mu\text{g}/\text{m}^3$ respectively.

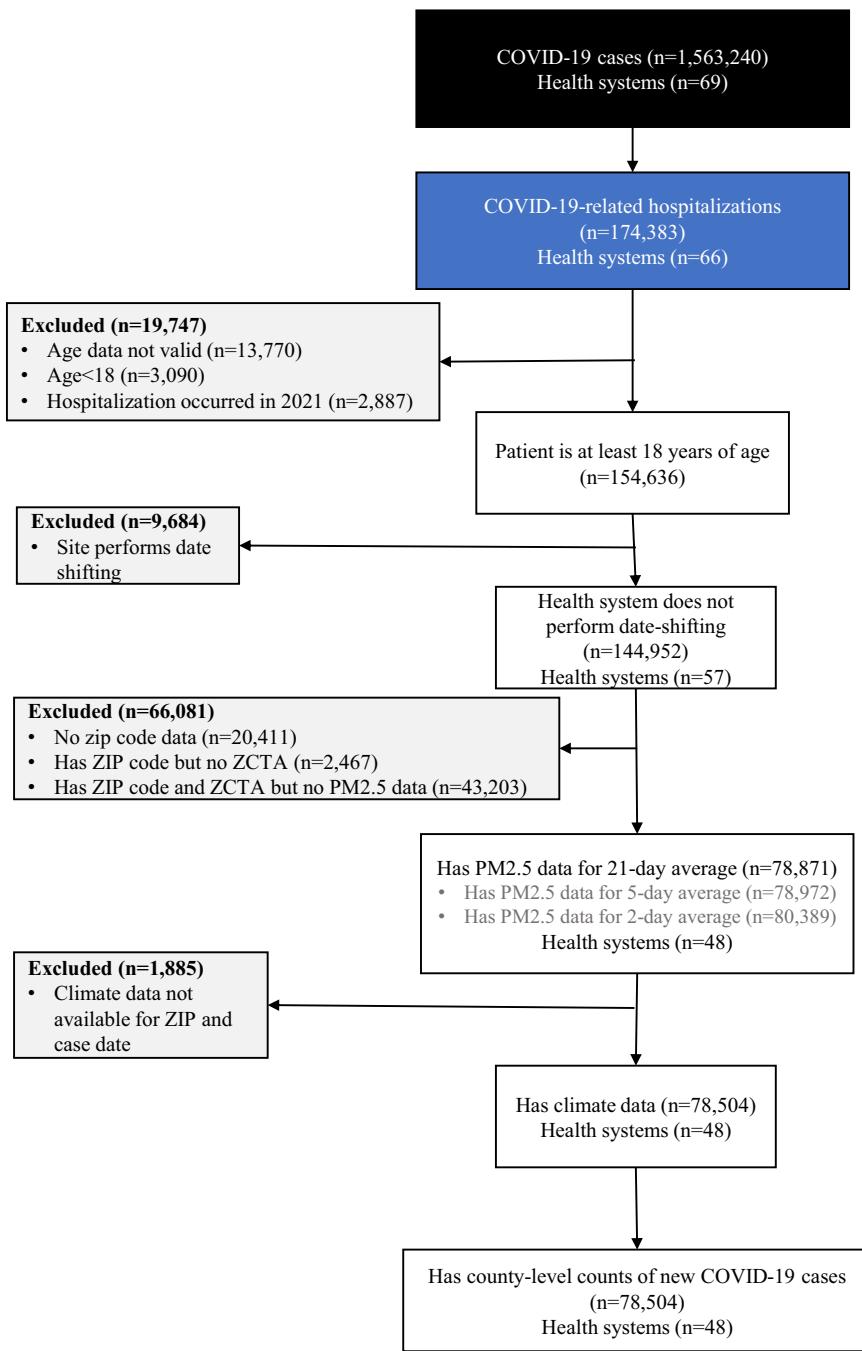
We compared characteristics between adult COVID19-related inpatient encounters that were included in our final analytic sample to those that were excluded (e.g., due to date shifting, lack of location data, or lack of $PM_{2.5}$ data) to examine characteristics associated with the probability of selecting into our final sample. There were some differences in racial (SMD = 0.442) and ethnic (SMD = 0.160) composition between the initial sample and final analytic sample (Supplemental Table 1). Further, patients with valid $PM_{2.5}$ data were more likely to reside in areas classified as urban (SMD = 1.016), due to the existing monitoring network, which has more coverage for urban areas. Patients with valid $PM_{2.5}$ data were also more likely to be Asian/Pacific Islander, Black, Other/Multiple Races, and Hispanic/Latino (SMD = 0.548 for race and SMD = 0.320 for ethnicity). (Supplemental Data 1).

Case-crossover analysis & meta-analysis

Without accounting for heterogeneity between CBSAs, in the full analytic patient population, a $1 \mu\text{g}/\text{m}^3$ unit increase in 2-day average $PM_{2.5}$ was associated with a 0.2% increase in COVID19 hospitalization risk (95% CI = 0.0, 0.4). For 5-day average $PM_{2.5}$ associations were 0.0% ($-0.3, 0.3$), and 21-day average $PM_{2.5}$ associations were -3.5% (95% CI: $-4.2, -2.8$) per $1 \mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ for each averaging period. When we stratified these results by urbanicity and observed positive associations between $PM_{2.5}$ and COVID19-related hospitalizations in urban and suburban areas for 2-day average $PM_{2.5}$ but a negative association for 21-day average $PM_{2.5}$ in urban areas (Supplemental Table 2). However, these overall estimates mask substantial variability between CBSAs in associations between air pollution and COVID19 related hospitalizations for all exposures time periods (Figs. 2–4). This variability is highlighted by heterogeneity seen through statistics such as Cochran's Q and I^2 which was observed for nearly all exposures time periods and regions examined (Supplemental Table 3). Since this study used a case-crossover design and all observations were hospital confirmed cases of SARS-CoV-2 infection & COVID19 this heterogeneity cannot be explained by at home testing differences or self-reporting of COVID19.

As detailed in the Methods, we performed a random effects meta-analysis of all CBSAs to better understand the relationship between $PM_{2.5}$ and COVID19 hospitalizations at national and regional scales while accounting for heterogeneity across CBSA-specific estimates. We required a minimum of 50 study participants in a CBSA to analyze its specific effects. There were 57 CBSAs meeting this requirement which represented 97.8% of study participants and 97.5% of hospitalizations. When all CBSAs were meta-analyzed Cochran's Q was significant ($p < 0.0001$) for all three air pollution averaging periods (2-, 5-, and 21-day). The I^2 statistic (representing the percentage of total variability attributable to intra-CBSA heterogeneity) was 68.0% for 2-day average $PM_{2.5}$, 72.6% for 5-day average $PM_{2.5}$, and 90.8% for 21-day average $PM_{2.5}$, further motivating the random effects meta-analysis and highlighting the need to account for heterogeneity across CBSAs.

Fig. 1 | STROBE diagram of cohort selection. The STROBE (Strengthening the Reporting of Observational studies in Epidemiology) diagram illustrates the sample size of the initial cohort and sample size of the final analytical sample after implementing multiple exclusion criteria.



In the meta-analyses heterogeneity was observed for both averaging period and region. In the meta-analysis of all 57 CBSAs, we observed associations for both 2-day (0.61%; 95% CI = 0.12, 1.11, $P = 0.02$; Fig. 2) and 5-day average PM_{2.5} (0.91%; 95% CI = 0.15, 1.67; $P = 0.02$; Fig. 3), but did not observe an association for 21-day average PM_{2.5} (0.04%; 95% CI = -2.70, 2.85; $P = 0.98$; Fig. 4).

Regional and Subgroup Analyses

The strongest positive associations between PM_{2.5} and COVID19-related hospitalizations were observed in the South region, where the association's magnitude increased with increasing averaging time (Table 2) and positive associations were observed for all averaging periods. Other regions showed greater variability in associations, with some averaging periods having null or positive associations and a few having negative associations, particularly for the Northeast (5-day and 21-day average PM_{2.5}) and West regions (21-day average PM_{2.5}).

We observed several subgroup specific associations that were suggestive of a difference in associations based on race, pre-existing conditions (CVD, MD, or RD), and age at COVID19 diagnosis. In particular, individuals with CVD showed strong associations overall as well as in the South region whereas those without CVD had attenuated or null associations (Supplemental Data 2). When formally testing for differences in the regression coefficients, however, there was limited evidence for statistically significant differences between coefficients. The only statistically significant differences by clinical or demographic factors were seen for CVD, Black race, and Other race (representing study participants who did not self-report as Black, White, Asian or Pacific Islander, or American Indian or Alaskan Native; Supplemental Data 3). Of these factors, only Black race was seen to modify associations when considering all CBSAs while other factors were only seen to modify associations in specific regions (Supplemental Data 3).

We performed a sensitivity analysis evaluating if relaxing the assumption of linearity for all continuous confounders (with the exception

Table 1 | Sociodemographic and clinical characteristics of patients included in analytic cohort

Variable	Summary
Patients	78,504
Sociodemographic characteristics	
Age (years) at positive test date, mean (SD)	62.3 (18.5)
Gender, n (%)	
Female	37,575 (47.9)
Male	40,920 (52.1)
Race, n (%)	
American Indian Or Alaska Native, n (%)	314 (0.4)
Asian Or Pacific Islander	3647 (4.6)
Black	18,540 (23.6)
White	39,993 (50.9)
Other or Multiple Races	11,586 (14.8)
Missing/Unknown	4424 (5.6)
Ethnicity, n (%)	
Hispanic or Latino	15,756 (20.1)
Not Hispanic or Latino	58,712 (74.8)
No matching concept, Other, or Unknown	4036 (5.1)
Co-morbidities at time of positive COVID test	
Cardiovascular diseases, n (%)	43,244 (55.1)
Respiratory disease, n (%)	17,122 (21.8)
Metabolic disease, n (%)	32,245 (41.1)
Area level characteristics	
PM2.5 on day of hospitalization - Mean (SD)	8.0 (5.1)
21-day PM2.5, Mean (SD)	7.9 (2.7)
5-day PM2.5, Mean (SD)	8.0 (3.8)
2-day PM2.5, Mean (SD)	8.0 (3.8)
Precipitation, mean (SD)	3.2 (7.9)
Mean temperature (Celsius), mean (SD)	10.7 (8.6)
Dewpoint, mean (SD)	4.4 (8.8)
7-day average of new COVID-19 cases from 16 days prior to hospitalization date, mean (SD)	489.5 (684.6)
Percent of population in ZCTA with poverty status, mean (SD)	16.3 (9.6)
Census Region, n (%)	
Midwest	37,336 (47.6)
Northeast	26,641 (33.9)
South	8811 (11.2)
West	5716 (7.3)
Rural-Urban Commuting Area classification, n (%)	
Rural	214 (0.3)
Suburban	579 (0.7)
Urban	77,704 (99.0)

SD standard deviation.

of precipitation due to convergence issues) altered associations. Results for these sensitivity analyses were slightly attenuated from the primary analysis but still showed positive associations between COVID19-related hospitalizations particularly for 2-day average PM_{2.5} (0.50%; 95% CI = -0.01, 1.02) and 5-day average PM_{2.5} (0.76%; 95% CI = -0.07, 1.59) though these associations were not statistically significant. Associations remained positive and statistically significant in the South region for all lag periods even after allowing for non-linear relationships for meteorological confounders (Supplemental Table 4).

To better visualize trends in the distribution of confounders across regions which may drive regional differences we computed the percentile ranking for all 57 CBSAs for each of the confounders and calculated the mean percentile ranking for each confounder for each Census region (Supplemental Table 5). While the West region had CBSAs that ranked the highest in PM_{2.5} concentration for all averaging periods the percentile rankings were all near 0.50 suggesting relatively little difference across the regions in PM_{2.5} exposure. In contrast the South region had CBSAs with the highest proportion of Black study participants, and the lowest proportion of White study participants and there was substantial difference in the mean percentile rankings. The South region also had the highest average CBSA percentile rankings for all three chronic disease categories examined (CVD, RD, and MD), and it contained the CBSAs with the second lowest average age at COVID19 diagnosis.

Discussion

Here, we explored associations between short-term PM_{2.5} exposure and COVID19-related hospitalizations in a demographically and geographically diverse sample with individual-level data. This study encompasses 78,504 unique individual patients representing 48 health systems and 57 CBSAs spread across all four US Census regions. To the best of our knowledge, this is the first multi-city study of PM_{2.5} and COVID19-related hospitalizations with national-in-scope individual level data in the US. Across all observations and without accounting for geographic heterogeneity as done in the random effects meta-analysis we observed a positive association between 2-day average PM_{2.5} and COVID19-related hospitalizations, no association with 5-day average PM_{2.5}, and an inverse association with 21-day average PM_{2.5}.

However, associations were much more consistent after accounting for heterogeneity across the CBSAs via a random effects meta-analysis (Table 2; Figs. 2–4), an approach that has proven useful in previous case-crossover studies^{59,60}. Factors which may have contributed to regional and CBSA heterogeneity in associations include differences in air pollution composition, temporal trends in COVID19-related hospitalizations, demographic differences (including differences in the health systems demographics), differences in underlying chronic conditions among the population, or differences in public health policy implementation/adherence which has been shown to be a factor in regional heterogeneity⁶². Previous studies of air pollution-related health effects have also observed differences between geographic regions driven by some of the same aforementioned factors (e.g., population demographics and chemical composition of air pollution)^{63–65}. Our results further highlight that failing to account for regional heterogeneity could cause misleading inference of the association between air pollution and health effects – particularly in the context of COVID19-related hospitalizations.

In the US, only a handful of studies have examined short-term air pollution exposure and COVID19-related hospitalizations. In a study of Cook County Illinois, positive associations were seen between short-term PM_{2.5} exposure and COVID19 mortality, with the strongest associations seen for cumulative PM_{2.5} exposure over 21 days¹⁹. In a study of county-level COVID19 case counts and short-term PM_{2.5} exposure researchers were able to combine COVID19 case count and PM_{2.5} data for 554/3143 (17.6%) of US counties²⁴. Though limited in spatial resolution and lacking individual-level data this study showed positive associations between PM_{2.5} exposure and county-level COVID19 cases, with positive associations seen for cumulative exposures ranging from 2 to 14-day average PM_{2.5}. The authors observed that associations were strengthened with the removal of New York state which they justified citing the severity of the COVID19 pandemic in New York state during their study time period. A 2020 study of COVID19 cases and deaths in Queens county, NY examined associations between 21-day average PM_{2.5} and both COVID19 case counts and death counts in the county and also observed negative associations between PM_{2.5} and both outcomes. In our study we also saw an unexpected negative association in New York which when taken in context of this prior study suggests that

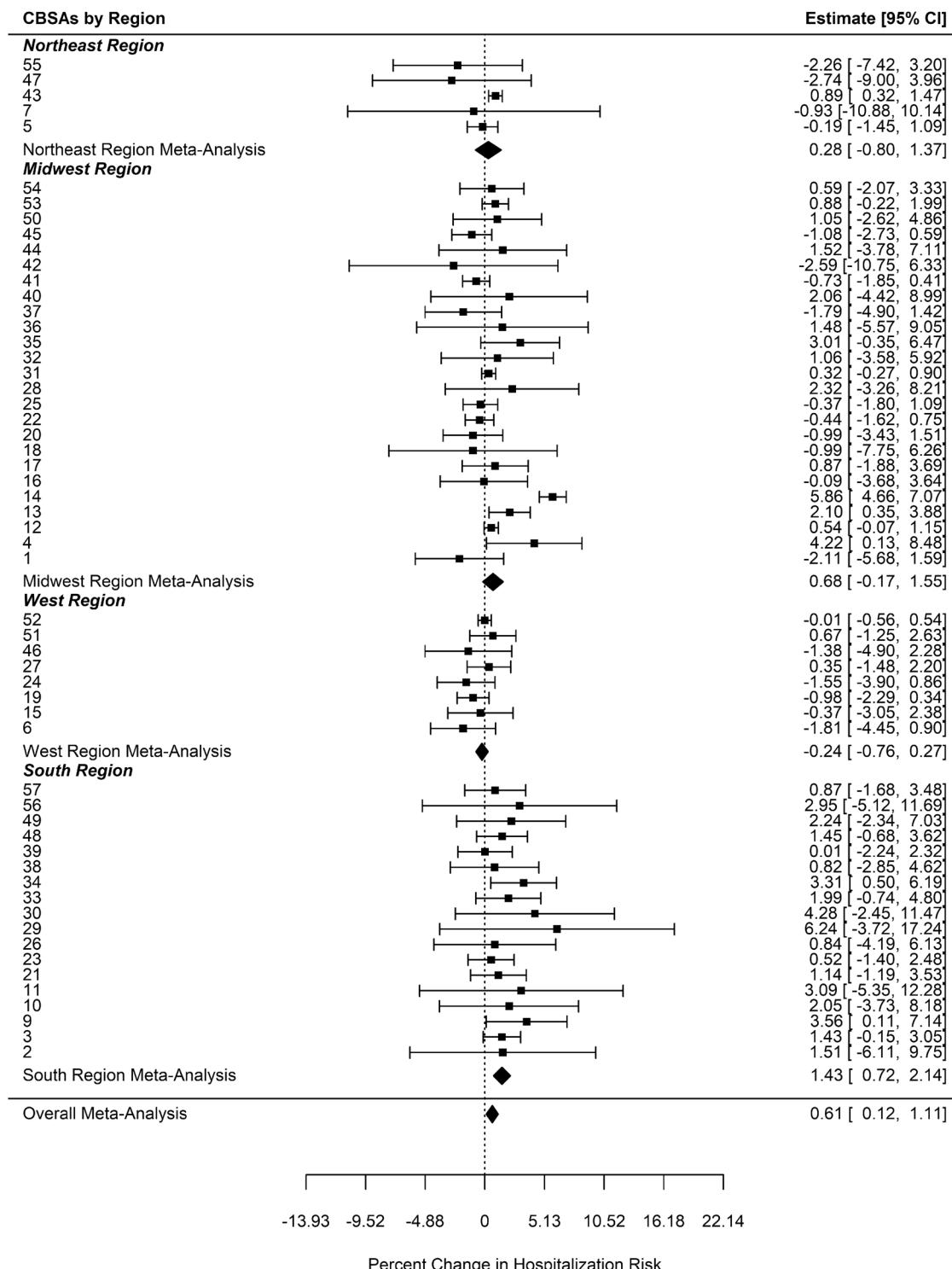


Fig. 2 | Regression estimates for the association between 2-day mean PM_{2.5} and COVID-19-related hospitalization, by CBSA and Census Region. Estimates are presented as percent increase in COVID19-related hospitalizations per 1 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5}. CBSA names are obscured to comply with privacy requirements.

evaluations of COVID19 and environmental health risks in New York may be subject to confounding or other factors not seen across the US and thus extra consideration may be warranted when comparing environmental health risks in New York to areas with differing COVID19 severity and temporal trends.

Long-term exposure studies of air pollution and COVID19 in the US have also faced some of the limitations of short-term exposures studies including limited geographical coverage or ecological analyses based on county-level counts of COVID19 cases or deaths. For example, a study in

Southern California cohort found that 1-year PM_{2.5} exposure was associated with increased odds of hospitalization, as well as other measures of infection severity among COVID19+ patients (e.g., intensive respiratory support, intensive care unit admission)²². A related study, also in Southern California, observed that 1-month and 1-year PM_{2.5} exposures were associated with increased risk of hospitalization, and this association persisted even in fully vaccinated patients⁶⁶. A study in Cincinnati, Ohio, observed associations between long-term (10-year) PM_{2.5} exposures and elevated odds of hospitalization among individuals who tested positive between March and

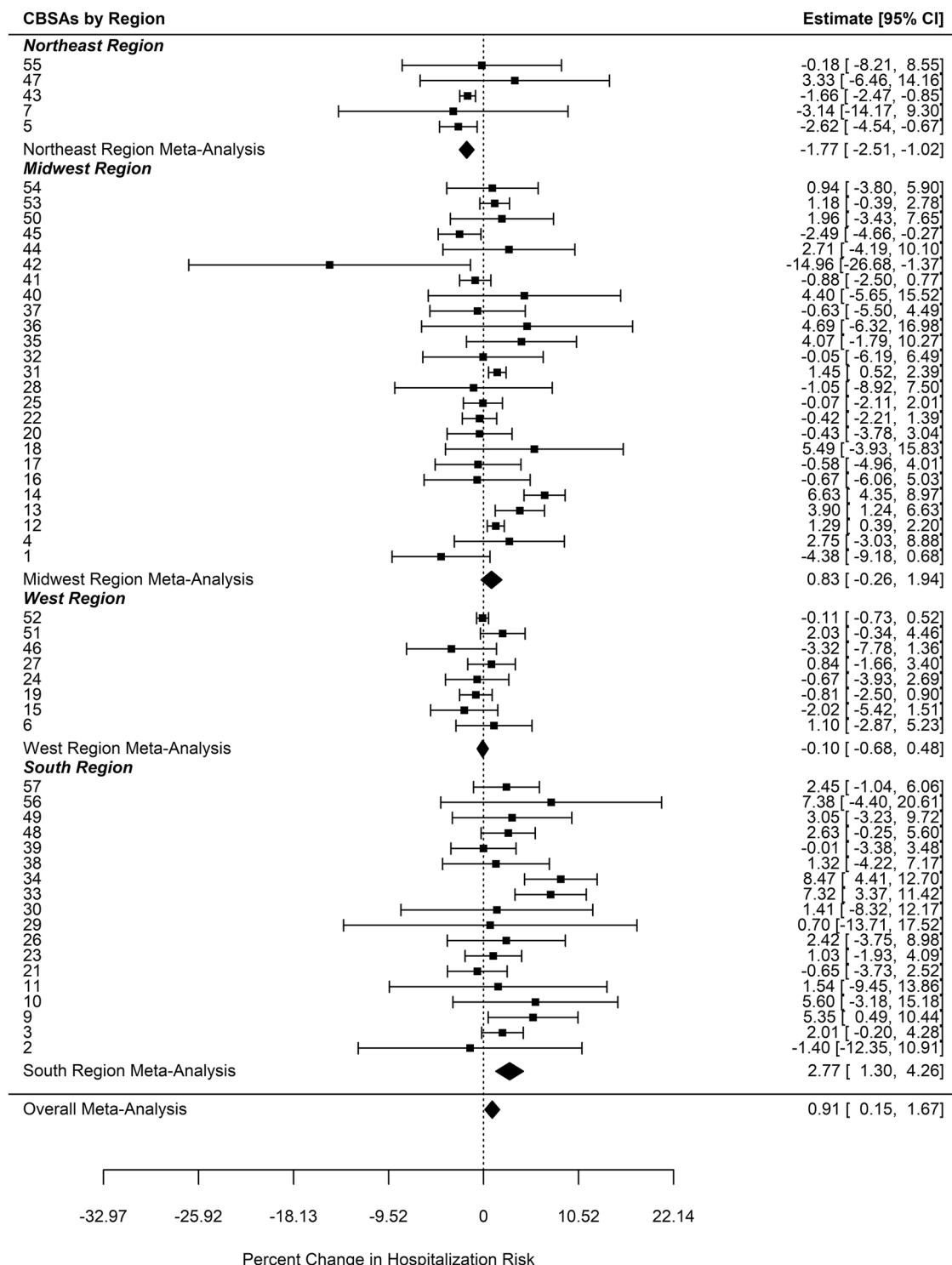


Fig. 3 | Regression estimates for the association between 5-day mean PM_{2.5} and COVID-19-related hospitalization, by CBSA and Census Region. Estimates are presented as percent increase in COVID19-related hospitalizations per 1 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5}. CBSA names are obscured to comply with privacy requirements.

September 2020⁶⁷. Air pollution may impact COVID19 in multiple ways including its incidence (risk of developing COVID) or severity. A recent review noted that of 116 articles on COVID19 and air pollution only 11 examined non-fatal severity (as we have here) and of those none were nationwide studies⁶⁸. Importantly there was significant heterogeneity among these studies with some only observing associations in individuals with chronic disease²⁵ and others noting significant regional heterogeneity as we also observed⁶⁹.

A biological explanation for associations seen among individuals with COVID19 is increased biological frailty due to environmental exposures, e.g. a weakened cardiovascular system making COVID19 cardiovascular effects more severe. This has also been called the double hit hypothesis and put forward in a previous study of associations between air pollution and mortality among COVID19 patients⁷⁰. This previous study of long-term air pollution exposure specifically hypothesized that PM_{2.5} exposure induced overexpression of the ACE-2 receptor. A study of short-term PM_{2.5}

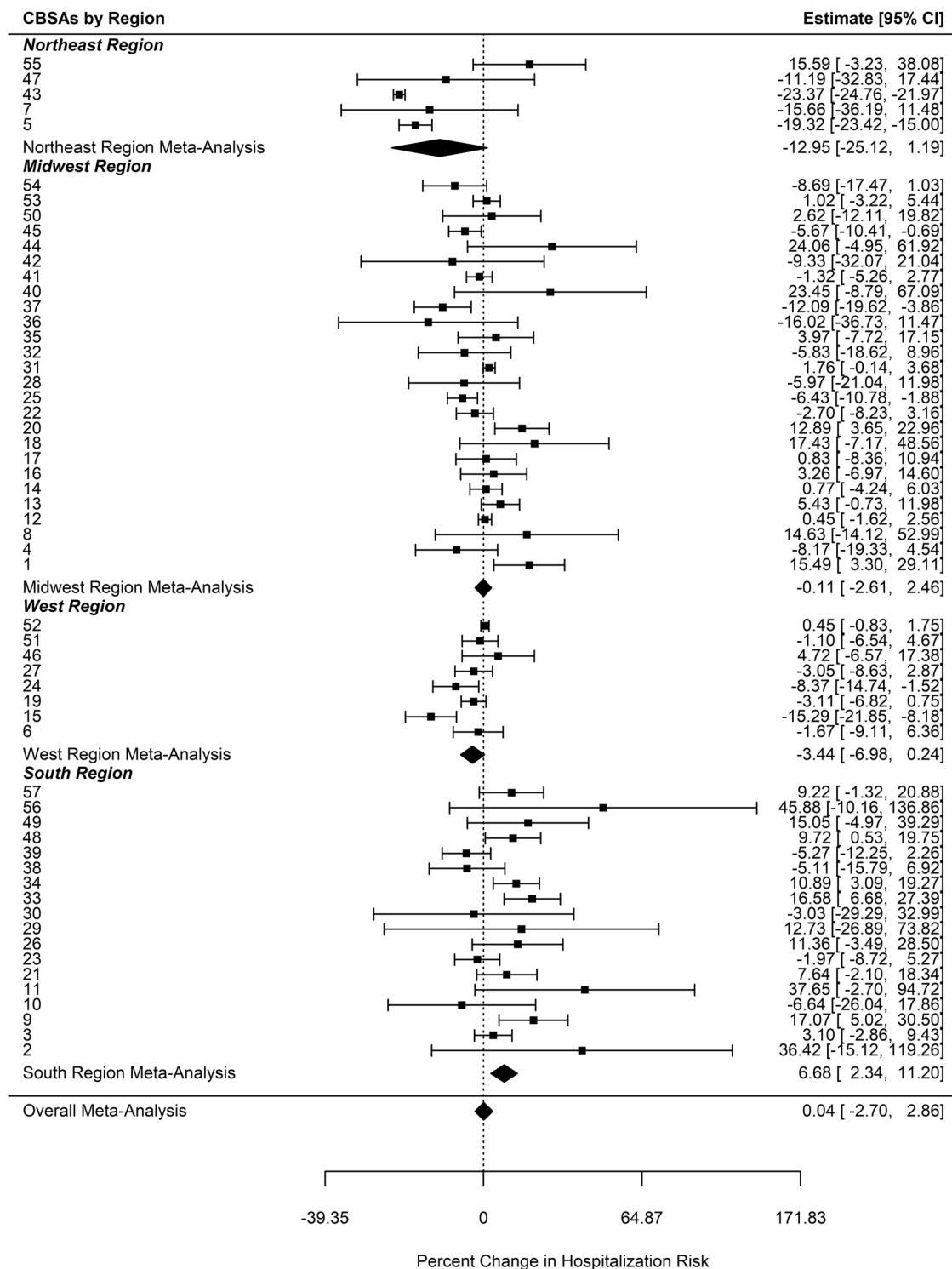


Fig. 4 | Regression estimates for the association between 21-day mean PM_{2.5} and COVID-19-related hospitalization, by CBSA and Census Region. Estimates are presented as percent increase in COVID19-related hospitalizations per 1 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5}. CBSA names are obscured to comply with privacy requirements.

exposure in mice also showed increased expression of ACE-2 receptor in the lung⁷¹. Increased ACE-2 expression was also observed when cells tank from the lungs of individuals with idiopathic pulmonary fibrosis (a progressive lung disease) were exposed to PM_{2.5}, further suggesting that short-term PM_{2.5} exposure may induce ACE-2 receptor expression in individuals with a compromised pulmonary system⁷². In addition to upregulating ACE-2 expression in the lung, short-term PM_{2.5} exposure is known to increase inflammation as well as trigger acute respiratory and cardiovascular responses which may also contribute to the associations observed here⁷³⁻⁷⁵.

Given the often limited sample size and ecological study design for existing studies of short-term PM_{2.5} and COVID19, few have examined how associations differ by region or geographic group. In a 21-day average PM_{2.5} study of COVID19 mortality in Cook County, Illinois authors observed that associations differed by race and ethnicity with higher estimated effects for Black individuals (particularly Black males)¹⁹. We also observed larger associations with increasing length of lag period in some census regions (South, Midwest), and found some evidence for differences in associations by race and pre-existing conditions. For example, consistent with the

Table 2 | Random Effects Meta-analysis of all CBSAs in Study Cohort

Region	2-day PM _{2.5}	5-day PM _{2.5}	21-day PM _{2.5}
Overall	0.61 (95% CI = 0.12, 1.11; $P = 0.02$)	0.91 (95% CI = 0.15, 1.67, $P = 0.02$)	0.04 (95% CI = -2.70, 2.85; $P = 0.98$)
South	1.43 (95% CI = 0.72, 2.14, $P = 7.1 \times 10^{-5}$)	2.77 (95% CI = 1.30, 4.26; $P = 2.0 \times 10^{-4}$)	6.67 (95% CI = 2.34, 11.2; $P = 0.002$)
Midwest	0.68 (95% CI = -0.17, 1.55, $P = 0.12$)	0.83 (95% CI = -0.26, 1.94; $P = 0.14$)	-0.11 (95% CI = -2.61, 2.46; $P = 0.93$)
West	-0.24 (95% CI = -0.76, 0.27, $P = 0.35$)	-0.10 (95% CI = -0.68, 0.48, $P = 0.73$)	-3.44 (95% CI = -6.98, 0.24; $P = 0.07$)
Northeast	0.28 (95% CI = -0.80, 1.37, $P = 0.61$)	-1.77 (95% CI = -2.51, -1.02; $P = 3.7 \times 10^{-6}$)	-13.0 (95% CI = -25.1, 1.19; $P = 0.07$)

Results are given as a percentage change as described in the Methods. CI confidence interval

Chicago-based study, we found larger effect sizes for PM_{2.5} and COVID19-related hospitalizations among Black study participants compared to White individuals (Supplemental Data 2; Supplemental Data 3). Additionally, associations between PM_{2.5} and COVID19-related hospitalizations among individuals with CVD were larger in magnitude than associations among other subgroups examined, particularly for 5-day average PM_{2.5} exposure in the South. We did not see differences by pre-existing respiratory or metabolic disease or age at COVID19 infection, highlighting the need for deeper investigations into the factors that may modify associations between PM_{2.5} and COVID19. To better understand whether there were differences in the confounders which may explain some of the regional differences seen in this study we examined the mean percentile ranking for each confounder at the CBSA level. The South region was enriched for CBSAs that had the highest proportion of Black study participants as well as the highest rates of chronic disease (Supplemental Table 5). The West region had the highest PM_{2.5} concentrations and the Northeast region the lowest. Chronic disease is known to increase sensitivity to air pollution exposure⁷⁶, and minority populations are often impacted by multiple social stressors which may also act to increase sensitivity to poor air quality^{77,78}. From a public health standpoint, understanding how poor air quality can contribute to worsened health outcomes during the COVID19 pandemic can have important implications for responding to and managing the next pandemic. Understanding how clinical and demographic factors, may cause environmental risks to be worsened in some areas of the country as compared to others also has important implications for predicting how the health effects from pandemics manifest across the country and the best means to allocate limited resources in order to minimize adverse health effects. While specific policy recommendations are outside of the scope of this manuscript, we believe the data presented here support the idea that environmental risks contributed to health effects during the COVID19 pandemic and should be considered in responses to future pandemics.

This study is, to the best of our knowledge, the first study to utilize EHR from the N3C to study environmental health risks. The N3C has become a widely utilized resource to study health outcomes related to the COVID19 pandemic, however this study establishes that its utility also extends into the environmental domain, giving environmental epidemiologists a new tool to understand how the environment shaped the pandemic. This study is not without limitations. Demographically, the N3C contains data from across the US but is not a perfect representation of the US population. Previous N3C studies have shown that there is underrepresentation of the Western US, Southwestern US, and rural areas in general³⁸. Another limitation is that although there were individuals residing in non-urban areas in the N3C EHR, the limited availability of monitoring data in less urban areas precluded our ability to perform informative analyses focused on rural areas. Additionally, EHR data is generally skewed towards patients who more frequently use health care systems and is further skewed towards individuals with more severe symptoms or disease as they are more likely to use the hospital system. Uninsured patients, those with limited access to care, or those with historical distrust of hospital systems may be underrepresented in any EHR cohort. Despite these limitations, studies using EHRs have been shown to be consistent with findings from data resources with regional or national representativeness⁷⁹, making them a valuable tool for environmental health studies. Although this study population was not drawn from a

nationally representative population, we replicated results seen in previous studies including positive associations nationally and unexpected negative associations in NY which has also been seen in studies just focusing on NY²³. Within our study, we evaluated differences between individuals included vs excluded from the final sample and found the two to be generally concordant (Supplemental Table 1 and Supplemental Data 1).

The analytic approach (time stratified case-crossover) has both strengths and limitations. A strength of this approach is it effectively controls for time-invariant and long-term time trends by design. A limitation of this approach is that without continuous serum sampling (unavailable for virtually all studies) we cannot verify that people were continuously infected with COVID19 for all referent time periods, leading to the potential sampling of non-at risk time periods. This is a known bias within such designs (and common in case-crossover studies of mortality where the post-death time period may be sampled), and has been shown to be typically small and less than the bias introduced by using unbalanced, i.e. unidirectional, designs^{80,81}. Additionally, given the dynamics of COVID19 infections it is likely that our referent selection covers the at-risk period for many participants⁸². However, this remains a limitation that could be addressed using alternative study designs with different structures in future analyses. Also, as stated before our study observed similar associations as seen in studies which did not use a time-stratified case-crossover design suggesting these associations may be robust to the analytic model chosen.

A limitation of this study is that we utilized ambient monitoring data, a choice necessitated by the lack of high-resolution (e.g., 1 km²) daily ambient air pollution exposure models covering the same time period as our data. However, at least one previous study found that associations between air pollution and COVID19 mortality were similar when pollution was estimated at ZIP code vs. residential latitude/longitude¹⁹. Finally, we specifically focused on hospitalizations in 2020, which pre-dated mass vaccinations within the US population; vaccination information and COVID19 strain are available for only a subset of the N3C population which means including these variables would have limited our sample size. However, a recent study in Southern California found that vaccination did not impact associations between PM_{2.5} and COVID19-related hospitalizations, suggesting that poor air quality may impact COVID19 health risks via a pathway independent of those acted on by vaccines⁶⁶. There are a number of avenues for future work, including the potential impact of vaccination and COVID19 strain on COVID19-related environmental health risks. Future work should consider other health endpoints beyond hospitalization and mortality, as well as exposure to: air pollutants beyond PM_{2.5}; air pollutant mixtures; and interactions between air pollutants and social stressors, especially at the individual level.

Air pollution – and PM_{2.5} specifically – is a ubiquitous and potentially modifiable exposure^{11,12} that is harmful to human health; understanding its role with respect to COVID19-related health outcomes may help address the public health burden of COVID19, and prepare for future pandemics. There continues to be a need for robust, large-scale studies of air pollution and COVID19 with demographically and geographically diverse individuals which can reveal the linkages between environmental exposures and exacerbation of COVID19 as well as other potential factors that also contributed to the severity of the COVID19 pandemic.

Data availability

The health data for this study is stored in the secure environment for the National Clinical Cohort Collaborative (N3C) and cannot be exported per data access agreements. All data from N3C used in this study is available through the N3C Enclave to approved users. See <https://covid.cd2h.org/for-researchers> for instructions on how to access the data. We used N3C data from version 152 (2023-12-07). Environmental data, including air pollution and temperature data; Census data; urbanicity data; and information on COVID19 cases are publicly available and obtained from sources as described in the manuscript.

Code availability

The code to construct the cohort and perform the analyses is available at: https://github.com/acp18/1_Base_Covid_Positive; https://github.com/acp18/2_Covid_Positive_Person_Fact; https://github.com/acp18/4_Visits; https://github.com/acp18/5_Local_Infections_and_Policies; https://github.com/acp18/6_Pollutants; https://github.com/acp18/7_Cohort_Building; https://github.com/acp18/8_Cohort_Narrowing_and_Analysis_v2. Code files used to generate the cohort and attach air pollution estimates are available in Zenodo: <https://doi.org/10.5281/zenodo.15990139>.

Received: 19 April 2024; Accepted: 22 July 2025;

Published online: 07 November 2025

References

- Wang, L., Wang, Y., Ye, D. & Liu, Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int. J. Antimicrobial Agents* **55**, 105948 (2020).
- The Lancet Public Health. COVID-19 pandemic: What's next for public health. *Lancet Public Health* **7**, e391 (2022).
- Mueller, A. L., McNamara, M. S. & Sinclair, D. A. Why does COVID-19 disproportionately affect older people? *Aging* **12**, 9959–9981 (2020).
- Anzalone, A. J. et al. Higher hospitalization and mortality rates among SARS-CoV-2-infected persons in rural America. *J. Rural Health* **27**, 39–54 (2022).
- Magesh, S. et al. Disparities in COVID-19 Outcomes by Race, Ethnicity, and Socioeconomic Status. *JAMA Netw. Open* **4**, e2134147 (2021).
- Tai, D. B. G., Sha, A., Doubeni, C. A., Sia, I. G. & Wieland, M. L. The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States. *Clin. Infect. Dis.* **72**, 703–706 (2021).
- Tessum, C. W. et al. PM_{2.5} polluters disproportionately and systemically affect people of color in the United States. *Sci. Adv.* **7**, eabf4491 (2021).
- Shumake, K. L., Sacks, J. D., Lee, J. S. & Johns, D. O. Susceptibility of older adults to health effects induced by ambient air pollutants regulated by the European Union and the United States. *Aging Clin. Exp. Res* **25**, 3–8 (2013).
- Wu, X., Braun, D., Schwartz, J., Kioumourtzoglou, M. A. & Dominici, F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci. Adv.* **6**, eaba5692 (2020).
- Di, Q. et al. Air Pollution and Mortality in the Medicare Population. *N. Engl. J. Med* **376**, 2513–2522 (2017).
- Bell, M. L. & Ebisu, K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environ. Health Perspect.* **120**, 1699–1704 (2012).
- Miranda, M. L., Edwards, S. E., Keating, M. H. & Paul, C. J. Making the environmental justice grade: the relative burden of air pollution exposure in the United States. *Int J. Environ. Res Public Health* **8**, 1755–1771 (2011).
- Day, D. B. et al. Association of Ozone Exposure With Cardiorespiratory Pathophysiologic Mechanisms in Healthy Adults. *JAMA Intern. Med.* **177**, 1344–1353 (2017).
- U.S. Environmental Protection Agency (EPA). Integrated Science Assessment for Particulate Matter (Final Report). (Washington DC, 2019).
- Pope, C. A. III, Coleman, N., Pond, Z. A. & Burnett, R. T. Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environ. Res.* **183**, 108924 (2020).
- Ural, B. B. et al. Inhaled particulate accumulation with age impairs immune function and architecture in human lung lymph nodes. *Nat. Med.* <https://doi.org/10.1038/s41591-022-02073-x> (2022).
- Benmarhnia, T. Linkages between air pollution and the health burden from COVID-19: Methodological challenges and opportunities. *Am. J. Epidemiol.* **189**, 1238–1243 (2020).
- Mendy, A. et al. Air pollution and the pandemic: Long-term PM_{2.5} exposure and disease severity in COVID-19 patients. *Respirology* **26**, 1181–1187 (2021).
- Kim, H., Samet, J. M. & Bell, M. L. Association between Short-Term Exposure to Air Pollution and COVID-19 Mortality: A Population-Based Case-Crossover Study Using Individual-Level Mortality Registry Confirmed by Medical Examiners. *Environ. Health Perspect.* **130**, 117006 (2022).
- Bozack, A. et al. Long-Term Air Pollution Exposure and COVID-19 Mortality: A Patient-Level Analysis from New York City. *Am. J. Crit. Care Med.* **205**, 651–662 (2022).
- Chen, Z. et al. Near-roadway air pollution associated with COVID-19 severity and mortality – multiethnic cohort study in Southern California. *Environ. Int.* **157**, 106862 (2021).
- Chen, Z. et al. Ambient air pollutant exposures and COVID-19 severity and mortality in a cohort of patients with COVID-19 in Southern California. *Am. J. Respiratory Crit. Care Med.* **206**, 440–448 (2022).
- Adhikari, A. & Yin, J. Short-Term Effects of Ambient Ozone, PM_{2.5}, and Meteorological Factors on COVID-19 Confirmed Cases and Deaths in Queens, New York. *Int. J. Environ. Res. Public Health* **17**, 4047 (2020).
- Xu, L., Taylor, J. E. & Kaiser, J. Short-term air pollution exposure and COVID-19 infection in the United States. *Environ. Pollut.* **292**, 118369 (2022).
- Mendy, A. et al. Long-term exposure to fine particulate matter and hospitalization in COVID-19 patients. *Respiratory Med.* **178**, 106313 (2021).
- Liang, D. et al. Urban Air Pollution May Enhance COVID-19 Case-Fatality and Mortality Rates in the United States. *Innovation* **1**, 100047 (2020).
- Wu, X., Nethery, R. C., Sabath, M. B., Braun, D. & Dominici, F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Sci. Adv.* **6**, eabd4049 (2020).
- Haendel, M. A. et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *J. Am. Med. Inform. Assoc.* **28**, 427–443 (2021).
- Pfaff, E. R. et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digital Health* **4**, E532–E541 (2022).
- Klann, J. G. et al. Distinguishing Admissions Specifically for COVID-19 From Incidental SARS-CoV-2 Admissions: National Retrospective Electronic Health Record Study. *J. Med Internet Res* **24**, e37931 (2022).
- Hill, E. et al. Risk Factors Associated with Post-Acute Sequelae of SARS-CoV-2 in an EHR Cohort: A National COVID Cohort Collaborative (N3C) Analysis as part of the NIH RECOVER program. *medRxiv*, <https://doi.org/10.1101/2022.08.15.22278603> (2022).
- Casiraghi, E. et al. A Methodological Framework for the Comparative Evaluation of Multiple Imputation Methods: Multiple Imputation of Race, Ethnicity and Body Mass Index in the U.S. National COVID Cohort Collaborative. *arXiv* **2206.06444** (2022).
- Bradwell, K. R. et al. Harmonizing units and values of quantitative data elements in a very large nationally pooled electronic health record (EHR) dataset. *J. Am. Med. Inf. Assoc.* **29**, 1172–1182 (2022).
- Deer, R. R. et al. Characterizing Long COVID: Deep Phenotype of a Complex Condition. *EBioMedicine* **74**, 103722 (2021).

35. Borland, D. et al. Enabling Longitudinal Exploratory Analysis of Clinical COVID Data. *ArXiv* (2021).

36. Rando, H. M. et al. Challenges in defining Long COVID: Striking differences across literature, Electronic Health Records, and patient-reported information. *medRxiv*, <https://doi.org/10.1101/2021.03.20.21253896> (2021).

37. Wong, R. et al. Effect of SARS-CoV-2 Infection and Infection Severity on Longer-Term Glycemic Control and Weight in People With Type 2 Diabetes. *Diab. Care* **45**, 2709–2717 (2022).

38. Bennett, T. D. et al. Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative. *Jama Netw. Open* **4**, e2116901 (2021).

39. OMOP. Vol. Version 5.3.1 (2021).

40. Phenotype_Data_Acquisition. *Github*, https://github.com/National-COVID-Cohort-Collaborative/Phenotype_Data_Acquisition (2020).

41. Ge, J., Pletcher, M. J. & Cai, J. C. Outcomes of SARS-CoV-2 Infection in Patients With Chronic Liver Disease and Cirrhosis: A National COVID Cohort Collaborative Study. *Gastroenterology* **161**, 1487–1501 (2021).

42. Bell, M. L. et al. Seasonal and regional short-term effects of fine particles on hospital admissions in 202 U.S. counties. *Am. J. Epidemiol.* **168**, 1301–1310 (2008).

43. Daly, C. et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**, 2031–2064 (2008).

44. Stoklosa, J., Daly, C., Foster, S. D., Ashcroft, M. B. & Warton, D. I. A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods Ecol. Evolution* **6**, 412–423 (2014).

45. PRISM Climate Group. PRISM Climate Data.

46. Son, J.-Y., Choi, H. M., Miranda, M. L. & Bell, M. L. Exposure to heat during pregnancy and preterm birth in North Carolina: Main effect and disparities by residential greenness, urbanicity, and socioeconomic status. *Environ. Res.* **204**, 112315 (2022).

47. Lim, C. C. et al. Long-Term Exposure to Ozone and Cause-Specific Mortality Risk in the United States. *Am. J. Resp. Crit. Care Med.* **200**, 1022–1031 (2018).

48. US Census Bureau. *Housing Patterns and Core-Based Statistical Areas*, <https://www.census.gov/topics/housing/housing-patterns/about/core-based-statistical-areas.html> (2021).

49. Hall, S. A., Kaufman, J. S. & Ricketts, T. C. Defining Urban and Rural Areas in U.S. Epidemiologic Studies. *J. Urban Health* **83**, 162–175 (2006).

50. USA Facts. *US COVID-19 cases and deaths by state* <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/> (2020).

51. Center for Disease Control and Prevention. *United States COVID-19 County Level of Community Transmission*, <https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-County-Level-of-Community-T/nra9-vzzn> and https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series (2022).

52. Flury, B. K. & Riedwyl, H. Standard Distance in Univariate and Multivariate Analysis. *Am. Statistician* **40**, 249–251 (1986).

53. Austin, P. C. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Commun. Stat. - Simul. Comput.* **38**, 1228–1234 (2009).

54. R Core Team. (ed R Foundation for Statistical Computing) (Vienna, Austria, 2019).

55. A Package for Survival Analysis in R (2023).

56. Therneau, T. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model.* (Springer, 2000).

57. Leese, P. et al. Clinical encounter heterogeneity and methods for resolving in networked EHR data: a study from N3C and RECOVER programs. *J. Am. Med. Inform. Assoc.* **30**, 1125–1136 (2023).

58. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

59. Yang, Y. et al. Spatial distribution and driving factors of the associations between temperature and influenza-like illness in the United States: a time-stratified case-crossover study. *BMC Public Health* **23**, 1403 (2023).

60. Eric, L. et al. Short-term exposure to ambient air pollution and individual emergency department visits for COVID-19: a case-crossover study in Canada. *Thorax* **78**, 459 (2023).

61. Zenodo. <https://doi.org/10.5281/zenodo.15990139> (Github, 2023).

62. Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L. & Samet, J. M. Spatial and Temporal Variation in PM_{2.5} Chemical Composition in the United States for Health Effects Studies. *Environ. Health Persp.* **115**, 989–995 (2007).

63. Pope, C. A. et al. Mortality Risk and Fine Particulate Air Pollution in a Large, Representative Cohort of U.S. Adults. *Environ. Health Perspect.* **127**, 077007 (2019).

64. Tian, Y. et al. Ambient air pollution and daily hospital admissions: A nationwide study in 218 Chinese cities. *Environ. Pollut.* **242**, 1042–1049 (2018).

65. Jin, T. et al. Associations between long-term air pollution exposure and the incidence of cardiovascular diseases among American older adults. *Environ. Int.* **170**, 107594 (2022).

66. Chen, Z. et al. The Independent Effect of COVID-19 Vaccinations and Air Pollution Exposure on Risk of COVID-19 Hospitalizations in Southern California. *Am. J. Respir. Crit. Care Med.* **207**, 218–221 (2023).

67. Mendy, A. et al. Air pollution and the pandemic: Long-term PM(2.5) exposure and disease severity in COVID-19 patients. *Respirology* **26**, 1181–1187 (2021).

68. Hernandez Carballo, I., Bakola, M. & Stuckler, D. The impact of air pollution on COVID-19 incidence, severity, and mortality: A systematic review of studies in Europe and North America. *Environ. Res.* **215**, 114155 (2022).

69. Linares, C. et al. Impact of environmental factors and Sahara dust intrusions on incidence and severity of COVID-19 disease in Spain. Effect in the first and second pandemic waves. *Environ. Sci. Pollut. Res. Int.* **28**, 51948–51960 (2021).

70. Frontera, A., Cianfanelli, L., Vlachos, K., Landoni, G. & Cremona, G. Severe air pollution links to higher mortality in COVID-19 patients: The “double-hit” hypothesis. *J. Infect.* **81**, 255–259 (2020).

71. Sagawa, T. et al. Exposure to particulate matter upregulates ACE2 and TMPRSS2 expression in the murine lung. *Environ. Res.* **195**, 110722 (2021).

72. Li, H.-H. et al. Upregulation of ACE2 and TMPRSS2 by particulate matter and idiopathic pulmonary fibrosis: a potential role in severe COVID-19. *Part. Fibre Toxicol.* **18**, 11 (2021).

73. Tang, H. et al. The short- and long-term associations of particulate matter with inflammation and blood coagulation markers: A meta-analysis. *Environ. Pollut.* **267**, 115630 (2020).

74. de Bont, J. et al. Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. *J. Intern. Med.* **291**, 779–800 (2022).

75. Xing, Y.-F., Xu, Y.-H., Shi, M.-H. & Lian, Y.-X. The impact of PM2.5 on the human respiratory system. *J. Thorac. Dis.* **8**, E69–E74 (2016).

76. Sacks Jason, D. et al. Particulate Matter–Induced Health Effects: Who Is Susceptible? *Environ. Health Perspect.* **119**, 446–454 (2011).

77. Fuller, C. H., Feeser, K. R., Sarnat, J. A. & O'Neill, M. S. Air pollution, cardiovascular endpoints and susceptibility by stress and material resources: a systematic review of the evidence. *Environ. Health* **16**, 58 (2017).

78. Weaver, A. M. et al. Neighborhood Sociodemographic Effects on the Associations Between Long-term PM2.5 Exposure and Cardiovascular Outcomes and Diabetes Mellitus. *Environ. Epidemiol.* **3**, e038 (2019).

79. Ward-Caviness, C. K. et al. Long-Term Exposure to Particulate Air Pollution Is Associated With 30-Day Readmissions and Hospital Visits

Among Patients With Heart Failure. *J. Am. Heart Assoc.* **10**, e019430 (2021).

80. Lumley, T. & Levy, D. Bias in the case–crossover design: implications for studies of air pollution. *Environmetrics* **11**, 689–704 (2000).

81. Janes, H., Sheppard, L. & Lumley, T. Overlap bias in the case–crossover design, with application to air pollution exposures. *Stat. Med.* **24**, 285–300 (2005).

82. Farina, N. et al. COVID-19: Pharmacology and kinetics of viral clearance. *Pharmacol. Res.* **161**, 105114 (2020).

Acknowledgements

This work does not necessarily represent the views or policies of the US Environmental Protection Agency. Funding for this work was provided by the US Environmental Protection Agency. We wish to acknowledge support from the Biostatistics, Epidemiology and Research Design (BERD) Methods Core funded through Grant Award Number UL1TR002553 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH). The N3C Publication committee confirmed that this manuscript (MSID:1684.601) is in accordance with N3C data use and attribution policies; however, this content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the N3C program. The analyses described in this publication were conducted with data or tools accessed through the NCATS N3C Data Enclave <https://covid.cd2h.org> and N3C Attribution & Publication Policy v 1.2-2020-08-25b supported by NCATS U24 TR002306, Axle Informatics Subcontract: NCATS-P00438-B, and EPA Contract 68HE0B22P0097. This research was possible because of the patients whose information is included within the data and the organizations (<https://ncats.nih.gov/n3c/resources/data-contribution/data-transfer-agreement-signatories>) and scientists who have contributed to the on-going development of this community resource [<https://doi.org/10.1093/jamia/ocaa196>]. Individual acknowledgements for core contributors and data partners with released data are provided in the acknowledgement section of the Supplementary Material.

Author contributions

Dr. Ward-Caviness and Dr. Bravo led hypothesis generation and defined the research question, with input from Dr. Weaver. Dr. Bravo oversaw the development of the analytical plan, led by Ms. Platt, and with input from Dr. Ward-Caviness and Dr. Weaver. Mr. Fisher was primarily responsible for data management and processing to create analysis-ready datasets, with oversight from Dr. Bravo and Dr. Ward-Caviness, and input from Dr. Weaver and Ms. Platt. Dr. Ward-Caviness and Dr. Bravo drafted the initial manuscript, with support from Ms. Platt in specific sections; Dr. Bell and Dr. Weaver

reviewed and edited multiple drafts of the manuscript, in addition to answering specific questions posed by Dr. Ward-Caviness and Dr. Bravo throughout the research process. All authors approved of the final submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-01063-7>.

Correspondence and requests for materials should be addressed to Mercedes A. Bravo.

Peer review information *Communications Medicine* thanks Jiangshao Gu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025