

A spatial and contextual exposome-wide association study and polyexposomic score of COVID-19 hospitalization

Hui Hu ^{1*}, Francine Laden^{1,2,3}, Jaime Hart^{1,2}, Peter James^{2,4}, Jennifer Fishe⁵, William Hogan ⁶, Elizabeth Shenkman⁶, Jiang Bian ⁶

¹Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

²Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁴Department of Population Medicine, Harvard Pilgrim Healthcare, Boston, MA, USA

⁵Department of Emergency Medicine, University of Florida College of Medicine—Jacksonville, Jacksonville, FL, USA

⁶Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA

*To whom correspondence should be addressed: Email: hui.hu@channing.harvard.edu

Abstract

Environmental exposures have been linked to COVID-19 severity. Previous studies examined very few environmental factors, and often only separately without considering the totality of the environment, or the exposome. In addition, existing risk prediction models of severe COVID-19 predominantly rely on demographic and clinical factors. To address these gaps, we conducted a spatial and contextual exposome-wide association study (ExWAS) and developed polyexposomic scores (PES) of COVID-19 hospitalization leveraging rich information from individuals' spatial and contextual exposome. Individual-level electronic health records of 50 368 patients aged 18 years and older with a positive SARS-CoV-2 PCR/Antigen lab test or a COVID-19 diagnosis between March 2020 and October 2021 were obtained from the OneFlorida+ Clinical Research Network. A total of 194 spatial and contextual exposome factors from 10 data sources were spatiotemporally linked to each patient based on geocoded residential histories. We used a standard two-phase procedure in the ExWAS and developed and validated PES using gradient boosting decision trees models. Four exposome measures significantly associated with COVID-19 hospitalization were identified, including 2-chloroacetophenone, low food access, neighborhood deprivation, and reduced access to fitness centers. The initial prediction model in all patients without considering exposome factors had a testing-area under the curve (AUC) of 0.778. Incorporation of exposome data increased the testing-AUC to 0.787. Similar findings were observed in subgroup analyses focusing on populations without comorbidities and aged 18–24 years old. This spatial and contextual exposome study of COVID-19 hospitalization confirmed previously reported risk factor but also generated novel predictors that warrant more focused evaluation.

Keywords: prediction

Introduction

The 2019 novel coronavirus disease (COVID-19) is a global pandemic causing significant health, social, and economic impacts. There are over 97 million cases and 1 million deaths in the United States as of October 2022.¹ A number of risk factors of severe COVID-19 have been identified, including older age, male gender, and underlying comorbidities such as hypertension, diabetes mellitus (DM), obesity, chronic lung diseases, cardiovascular diseases (CVDs), liver and kidney diseases, cancer, clinically apparent immunodeficiencies, local immunodeficiencies, and pregnancy.² Recent studies suggested that spatial and contextual factors may be important determinants of severe COVID-19.³ Long-term exposures to a number of spatial and contextual factors (eg, air pollution and chemical exposures, climate, and the built environment) have been linked to COVID-19 severity through several major underlying mechanisms including impairing of immune system, regulating viral survival and

transport, and increasing comorbidities associated with severe COVID-19.²

However, there are several gaps in existing studies. First, despite the potential important role of spatial and contextual factors, existing risk prediction models of severe COVID-19 predominantly rely on demographic and clinical factors, and little effort has been made to consider the spatial and contextual exposome. Second, most COVID-19 studies that examined spatial and contextual factors have focused on individual exposures or exposures from a single class (eg, air pollutants),³ and very few have considered the totality of the environment or the exposome.^{4,5} Individuals are exposed to multiple spatial and contextual factors simultaneously with complex interplays among these factors, and the exposome is the ideal framework to rigorously estimate the health effects associated with multiple exposures. Third, to date, most of these COVID-19 studies on spatial

Received: December 19, 2022. **Revised:** March 22, 2023. **Accepted:** April 06, 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

and contextual factors were ecological and based on aggregated COVID-19 outcomes. While informative for hypothesis generation, ecological studies are limited by the lack of individual-level characteristics and substantial exposure misclassification.

To address these gaps, in this study, we obtained individual-level electronic health records (EHRs) data of COVID-19 patients from a large clinical research network, the OneFlorida+ network, and linked data from multiple sources to characterize patients' long-term exposures to the spatial and contextual exposome based on their geocoded residential histories. Using the agnostic and hypothesis-free exposome framework, we conducted a spatial and contextual exposome-wide association study (ExWAS) and developed polyexposomic scores (PES) of COVID-19 hospitalization.

Methods

Study population

We obtained EHR data from the OneFlorida+ Clinical Research Network, a large clinical data research network that is part of the National Patient Centered Clinical Research Network (PCORnet).^{6,7} OneFlorida+ encompasses collaborations with 14 academic institutions and health systems and comprises longitudinal and linked individual-level health care information for approximately 20 million patients. The data included in OneFlorida+ are a HIPAA-limited data set (ie, patients' geocoded residential histories are available and dates are unaltered) and follow the PCORnet Common Data Model.⁸

We obtained data from 58 482 patients with (1) a positive SARS-CoV-2 PCR/Antigen lab test or a COVID-19 diagnosis between March 2020 and October 2021, (2) at least one inpatient encounter or two outpatient encounters (at least 3 months apart) within 5 years before COVID-19 onset, and (3) geocoded residential history. We further excluded those under the age of 18 years ($n=8114$). A total of 50 368 patients were included in the analyses. This study was approved by the institutional review boards at the Mass General Brigham HealthCare System (2021P002830) and University of Florida (IRB202001831).

COVID-19 hospitalization

Consistent with other studies,⁹ we defined COVID-19 hospitalization as the first hospital admission within 7 days prior to and/or 15 days after patients' first COVID-19 positive date.

The spatial and contextual exposome

To assess patients' long-term exposures to the spatial and contextual exposome, 10 well-validated sources were used to obtain data on various measures related to the natural, built, and social environment. Patients' geocoded residential history data were obtained from the EHR, and spatiotemporally linked to exposome measures using a 250-m circular buffer around each address. Area- and time-weighted averages were generated to account for the heterogeneous spatiotemporal scale of exposome data. [Table 1](#) shows a summary of the spatial and contextual exposome data sources. A total of 268 spatial and contextual exposome factors covering 10 categories were obtained.

Natural environment

We obtained data on ambient ozone (O_3) and fine particulate matter ($PM_{2.5}$) from the US Environmental Protection Agency (USEPA)'s Fused Air Quality Surface Using downscaling (FAQSD) Files,¹⁰ which used a Bayesian space-time downscaler model to fuse 12 km gridded output from the Models-3/Community

Multiscale Air Quality model with daily O_3 and $PM_{2.5}$ stationary monitored data from the National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS).¹¹ Daily estimates were obtained at the Census tract level for 2015–2018. In addition, we obtained data on monthly $PM_{2.5}$ composition (ie, sulfate [SO_4^{2-}], ammonium [NH_4^+], nitrate [NO_3^-], organic matter [OM], black carbon [BC], mineral dust [DUST], and sea-salt [SS]) at a spatial resolution of 0.01° in longitude and latitude in 2015–2017 from the University of Washington at St. Louis Atmospheric Composition Analysis Group.¹² Extensively cross-validated geographically weighted models were used to statistically fuse satellite observations of aerosol optical depth with data from a chemical transport model (GEOS-Chem).¹² Air toxicant measures were obtained from the National Air Toxics Assessment, which was developed based on a national emissions inventory of outdoor air toxics sources.¹³ Exposure estimates of 175 air toxicants are available at the Census tract level in 2014.

Built environment

Green space was assessed using the normalized difference vegetation index from NASA's MODIS/Terra, which has been widely used in epidemiological studies.¹⁴ We obtained data in 2015–2019 with a 16-day temporal resolution and a 250-m spatial resolution. Walkability was assessed using the National Walkability Index developed by the USEPA,¹⁵ which measures walkability on a scale from 1 to 20 (ie, higher score indicates higher walkability) for each Census block group. Data on food access were obtained from the US Department of Agriculture (USDA) Food Access Research Atlas.¹⁶ We obtained Census-tract level measures in 2015 and 2019, and linear interpolation was performed to construct measures in 2016–2018. Vacant land measures at the Census-tract level in 2015–2019 were obtained from the US Department of Housing and Urban Development (USHUD) aggregated US Postal Service (USPS) administrative data.¹⁷

Social environment

Neighborhood deprivation index (NDI), a validated metric of neighborhood socioeconomic status (SES),¹⁸ was generated based on 20 Census block group-level factors (covering seven domains including poverty, occupation, housing, employment, education, racial composition, and residential stability) using data from the 2015–2019 American Community Survey. Higher NDI scores represent worse neighborhood SES. In addition, 10 contextual-level social capital measures were constructed using the 2015–2019 Census Business Pattern data at the 5-digit ZIP Code Tabulation Area (ZCTA5) level based on the North American Industry Classification System codes.¹⁹ Furthermore, we also obtained eight county-level annual crime measures from the Uniform Crime Reporting Program in 2015–2019.²⁰

Covariates

Patients' age was categorized into seven groups, with 10-year increments for those aged 25–74 years old and two additional groups for those aged between 18–24 and ≥ 75 years old. Patients' gender and race/ethnicity (ie, non-Hispanic White, non-Hispanic Black, Hispanic, or other) were also included. In addition, we obtained patients' health insurance status (ie, Medicare, Medicaid, other government programs, private insurance, no insurance, or other) and history of comorbidities (ie, atherosclerotic CVD, myocardial infarction, hypertension, peripheral vascular disease, cerebrovascular disease, DM, chronic obstructive pulmonary disease, asthma, cancer, chronic kidney disease, renal disease, and organ transplant). We also obtained time-series data

Table 1. Summary of spatial and contextual exposome measures

Category	Data source	Time period	Spatial scale	Temporal scale	Number of measures	Number of variables ^a
PM _{2.5} and O ₃	Fused Air Quality Surface Using Downscaling Files, USEPA	2015–2018	Census tract	1-day	2	2
PM _{2.5} compositions	Atmospheric Composition Analysis Group, WUSTL	2015–2017	0.01° in lon./lat.	1-month	7	7
Air toxicants	National Air Toxic Assessment, USEPA	2014	Census tract	1-year	175	103
Green space	MODIS/Terra Normalized Difference Vegetation Index, NASA	2015–2019	250 m	16-day	1	1
Walkability	Walkability Index, USEPA	2019	Census block group	Cross-sectional	1	1
Food access	Food Access Research Atlas, USDA	2015, 2019	Census tract	1-year	44	43
Vacant land	Aggregated USPS Administrative Data on Address Vacancies, USHUD	2015–2019	Census tract	3-month	19	18
Neighborhood deprivation	American Community Survey, US Census Bureau	2015–2019	Census block group	5-year	1	1
Social capital	Census Business Pattern, US Census Bureau	2015–2019	ZCTA5	1-year	10	10
Crime and safety	Uniform Crime Reporting Program, FBI	2015–2019	County	1-year	8	8

^a Number of variables after removing 57 measures with the number of unique values <0.1% of the total sample size and 17 measures with absolute correlations >0.99 with another measure.

Abbreviations: FBI, Federal Bureau of Investigation; MODIS, The Moderate Resolution Imaging Spectroradiometer; NASA, The National Aeronautics and Space Administration; USDA, United States Department of Agriculture; USEPA, United States Environmental Protection Agency; USHUD, United States Department of Housing and Urban Development; WUSTL, Washington University at St. Louis; ZCTA5, 5-digit ZIP Code Tabulation Areas.

on several county-level COVID-19-related factors and linked them to each patient based on their county of residence and COVID-19 positive date, including number of days since first COVID-19 case derived using data from Johns Hopkins University, Center for Systems Science and Engineering Coronavirus Resource Center,²¹ county-level COVID-19 vaccination rates (ie, at least one dose, fully vaccinated) obtained from the US Centers for Disease Control and Prevention,¹ and county-level hospital bed capacity.²²

Statistical analysis

Descriptive analyses by patients' sociodemographic status, comorbidities, county-level COVID-19-related factors, and COVID-19 hospitalization were performed. A total of 57 spatial and contextual exposome measures with unique values <0.1% of the sample size were removed (Table S1). All continuous exposures were standardized (ie, mean=0 and standard deviation=1). Figure 1 shows the flow chart summarizing the analysis pipeline.

The standard two-phase environment-wide association study followed by a multiple generalized linear regression (EWAS-MLR) approach was used to perform the ExWAS, with environment-wide association study followed by a multivariable regression step including the identified hits.^{23,24} Pairwise Pearson correlations were computed, and we further excluded 17 spatial and contextual measures with absolute correlations >0.99 with another measure (Table S2). We imputed missing data for all exposures and potential confounders for the ExWAS using the chained equations method with the mice package in R.²⁵ We considered variables as predictors in the imputation model if their proportion of nonmissing values among patients with missing values in the variable to be imputed was larger than 40% and they were correlated (ie, an absolute correlation value > 0.4) with the variable to be imputed or the probability of the variable being missing. A single dataset was imputed given the minimal impacts

of the imputation procedure because of the large sample size and small fractions of missing data. In Phase 1, data were randomly split into a 50% discovery set and a 50% replication set. We individually examined the associations between each of the 194 spatial and contextual exposome factors included and COVID-19 hospitalization, after accounting for multiple comparisons. Mixed-effect logistic regression models were fitted for each exposure after adjusting for all the potential confounders and with a random intercept by county. To account for the multiple testing, the Benjamin-Hochberg procedure was used to control the false discovery rate (FDR) at 5%.²⁶ A variable with FDR-adjusted P-values (or q-values) < 0.05 in both the discovery and replication sets is determined as statistically significant. A correlation heatmap was also generated to visualize pairwise Pearson correlations of the variables retained from Phase 1. In Phase 2, we used a multivariable mixed-effect logistic regression model to estimate the effect sizes, which simultaneously included all significant exposures from Phase 1 along with all the potential confounders. Variables remained significant in Phase 2 are retained. Odds ratios (ORs) and 95% confidence intervals (CIs) were reported.

To develop and validate the PES of COVID-19 hospitalization, data were randomly split into a training set (80%) and a testing set (20%). We trained prediction models using CatBoost, a high-performance open-source library for gradient boosting on decision trees. Hyperparameters were tuned using the grid search approach based on 4-fold cross-validations over the training set, and validation was performed using the testing set. Six models were trained and validated with and without spatial and contextual exposome measures in (1) all COVID-19 patients, (2) COVID-19 patients without any comorbidity, and (3) COVID-19 patients without any comorbidity and aged 18–24 years old. Feature importance measures were assessed using Shapley additive explanations (SHAP) values.²⁷ All analyses were performed using the R statistical software (version 4.1.3; R Development Core Team).

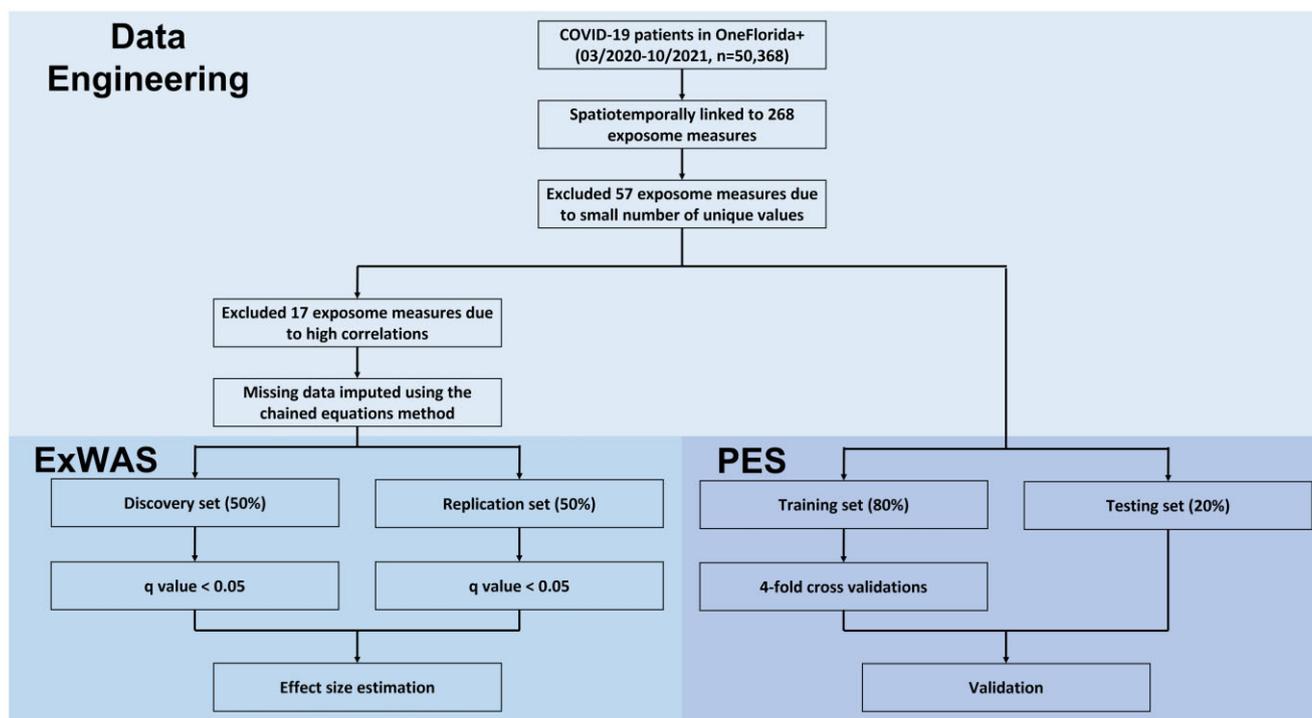


Figure 1. Flowchart of the analysis pipeline. Abbreviations: ExWAS, exposome-wide association study; PES, polyexposomic score.

Results

Of the 50 368 COVID-19 patients included in this study, a total of 12 911 (25.6%) were hospitalized. Table 2 shows the distribution of patients' characteristics by COVID-19 hospitalization. Compared with those without hospitalization, hospitalized patients were more likely to be older, male, non-Hispanic Black or Hispanic, Medicare recipient, and have comorbidities. In addition, hospitalized patients were less likely to live in counties with higher vaccination rates or hospital bed capacity. Table S3 shows the distribution of spatial and contextual exposome measures included in this study.

The volcano plot presented in Figure 2 summarizes results from Phase 1 of the ExWAS. A total of 12 and 24 exposome measures were statistically significantly associated with COVID-19 hospitalization after accounting for multiple comparisons in the discovery and replication sets, respectively. Among them, seven exposome measures were significant in both the discovery and replication sets. The effect estimates for each of the 194 exposome measures from Phase 1 were shown in Table S4.

Figure S1 shows the pairwise correlations of the seven exposome measures that were significant in both the discovery and replication sets in Phase 1. All correlation coefficients had absolute values below 0.7. In Phase 2 of the ExWAS, the seven exposome measures were simultaneously included in a multivariable mixed-effect logistic regression model after adjusting for the covariates. Table 3 shows the adjusted ORs for each standard deviation increase in these measures, along with their 95% CIs. A total of four exposome measures significantly associated with COVID-19 hospitalization were identified, including 2-chloroacetophenone (OR, 1.13, 95% CI, 1.08–1.18), percentage of low food access population with housing units without vehicle access at 1 mile (OR, 1.04, 95% CI, 1.00–1.07), NDI (OR, 1.10, 95% CI, 1.07–1.13), and density of fitness and recreational sports centers (OR, 0.95, 95% CI, 0.92–0.97).

Figure 3 shows the receiver operating characteristic (ROC) curves from the validations using the testing sets for the prediction models trained in the three target populations, with and without spatial and contextual exposome measures as predictors. In the first set of models focusing on all COVID-19 patients, we observed testing-AUCs (areas under the curve) of 0.787 (95% CI, 0.777–0.797) and 0.778 (95% CI, 0.768–0.788) with and without spatial and contextual exposome measures included as predictors, respectively. Similar patterns of AUCs were observed between models with and without exposome measures in the other two target populations, with testing-AUCs of 0.767 (95% CI, 0.751–0.785) and 0.756 (95% CI, 0.739–0.773) observed in models focusing on those without comorbidities, respectively, and testing-AUCs of 0.705 (95% CI, 0.651–0.756) and 0.669 (95% CI, 0.602–0.732) in models focusing on those without comorbidities and between 18 and 24 years old. Table S5 shows the optimal hyperparameters for each model tuned by grid searches based on cross-validations.

Figure 4 shows the beeswarm plot and mean absolute SHAP values for the top 10 predictive features from each of the models with spatial and contextual exposome measures included. In the model with all COVID-19 patients, 3 of the top 10 features were exposome measures, including dimethyl phthalate, NO_3^- , and NDI. In the model of patients without any comorbidity, 4 of the top 10 features were exposome measures, including DUST, NO_3^- , NDI, and O_3 . Similarly, 5 of the top 10 features were exposome measures in the model focusing on those without comorbidities and aged 18–24 years old, including burglary rate, density of fitness and recreational sports centers, DUST, murder rate, and hexamethylene diisocyanate.

Discussion

This is the first spatial and contextual exposome study of COVID-19 hospitalization using individual-level data. Using the ExWAS, we assessed the associations between 194 spatial and contextual

Table 2. COVID-19 patients' characteristics by COVID-19 hospitalization between March 2020 and October 2021 in OneFlorida+ [mean \pm SD/n(%)]

Characteristics	COVID-19 hospitalization		Total 50 368 (100.0)
	Yes 12 911 (25.6)	No 37 457 (74.4)	
Age (year)			
Continuous	59.5 \pm 18.2	44.0 \pm 17.6	48.0 \pm 19.0
Categorical			
18–24	446 (3.5)	6292 (16.8)	6738 (13.4)
25–34	1064 (8.2)	7639 (20.4)	8703 (17.3)
35–44	1413 (10.9)	7001 (18.7)	8414 (16.7)
45–54	2004 (15.5)	6133 (16.4)	8137 (16.2)
55–64	2782 (21.5)	5320 (14.2)	8102 (16.1)
65–74	2428 (18.8)	3040 (8.1)	5468 (10.9)
\geq 75	2774 (21.5)	2032 (5.4)	4806 (9.5)
Gender			
Male	5805 (45.0)	13 110 (35.0)	18 915 (37.6)
Female	7104 (55.0)	24 343 (65.0)	31 447 (62.4)
Missing	2 (0.0)	4 (0.0)	6 (0.0)
Race/ethnicity			
Non-Hispanic White	5079 (39.3)	15 430 (41.2)	20 509 (40.7)
Non-Hispanic Black	3547 (27.5)	9735 (26.0)	13 282 (26.4)
Hispanic	3345 (25.9)	8456 (22.6)	11 801 (23.4)
Other	729 (5.6)	2790 (7.4)	3519 (7.0)
Missing	211 (1.6)	1046 (2.8)	1257 (2.5)
Health insurance			
Medicare	2791 (21.6)	3250 (8.7)	6041 (12.0)
Medicaid	567 (4.4)	2370 (6.3)	2937 (5.8)
Other government programs	865 (6.7)	1531 (4.1)	2396 (4.8)
Private insurance	6450 (50.0)	18 995 (50.7)	25 445 (50.5)
No insurance	628 (4.9)	1893 (5.1)	2521 (5.0)
Other	1427 (11.1)	9248 (24.7)	10 675 (21.2)
Missing	183 (1.4)	170 (0.5)	353 (0.7)
Comorbidities			
Atherosclerotic CVD	2462 (19.1)	2685 (7.2)	5147 (10.2)
Myocardial infarction	533 (4.1)	548 (1.5)	1081 (2.1)
Hypertension	7507 (58.1)	12 507 (33.4)	20 014 (39.7)
Peripheral vascular disease	1453 (11.3)	1900 (5.1)	3353 (6.7)
Cerebrovascular disease	459 (3.6)	646 (1.7)	1105 (2.2)
DM	4682 (36.3)	6324 (16.9)	11 006 (21.9)
Chronic obstructive pulmonary disease	2484 (19.2)	4333 (11.6)	6817 (13.5)
Asthma	157 (1.2)	173 (0.5)	330 (0.7)
Cancer	889 (6.9)	1199 (3.2)	2088 (4.1)
Chronic kidney disease	4416 (34.2)	5483 (14.6)	9899 (19.7)
Renal disease	2407 (18.6)	2231 (6.0)	4638 (9.2)
Organ transplant	354 (2.7)	358 (1.0)	712 (1.4)
Any of the comorbidities above	8629 (66.8)	16 094 (43.0)	24 723 (49.1)
County-level COVID-19-related factors			
Number of days since first case	308.0 \pm 157.1	307.1 \pm 161.6	307.3 \pm 160.5
At least one vaccination dose (%)	21.8 \pm 25.6	22.5 \pm 25.9	22.3 \pm 25.8
Fully vaccinated (%)	17.5 \pm 21.9	18.2 \pm 22.3	18.0 \pm 22.2
Hospital bed capacity (per 100 000 population)	69.4 \pm 65.2	78.9 \pm 66.7	76.5 \pm 66.4

exposome measures and COVID-19 hospitalization in Florida, USA. After accounting for multiple testing and high correlations among the exposures, four exposome measures characterizing the natural (ie, 2-chloroacetophenone), built (ie, food access), and social environment (ie, neighborhood deprivation and density of fitness and recreational sports centers) were identified to be significantly associated with COVID-19 hospitalization. In addition, we also developed and validated PES models of COVID-19 hospitalization. While not statistically significant, prediction models with spatial and contextual measures as predictors had better performance compared with models without exposome measures.

Long-term exposures to air pollution have been widely reported to be associated with COVID-19 severity. Ecological studies conducted in the early stage of the pandemic suggested

that long-term exposures to US EPA-regulated criteria air pollutants such as PM_{2.5} and O₃ are associated with higher COVID-19 mortality.^{28–38} More recently, studies using individual-level data also found positive associations between long-term exposures to PM_{2.5} and COVID-19 hospitalization.^{9,39} In our study, neither PM_{2.5} or O₃ was significantly associated with COVID-19 hospitalization. The inconsistent findings may be partially due to different study populations (eg, Bowe et al.⁹ focused on veterans), potential exposure misclassifications (eg, Mendy et al.³⁹ estimated exposures at residential zip code level), and heterogeneous source and chemical composition of air pollution across different regions.⁴⁰ Interestingly, we found that while NO₃⁻ (ie, a major PM_{2.5} constituent) was only significantly associated with COVID-19 hospitalization in the discovery set in Phase 1 of the ExWAS (and therefore was not retained in Phase 2 of the ExWAS), it was one

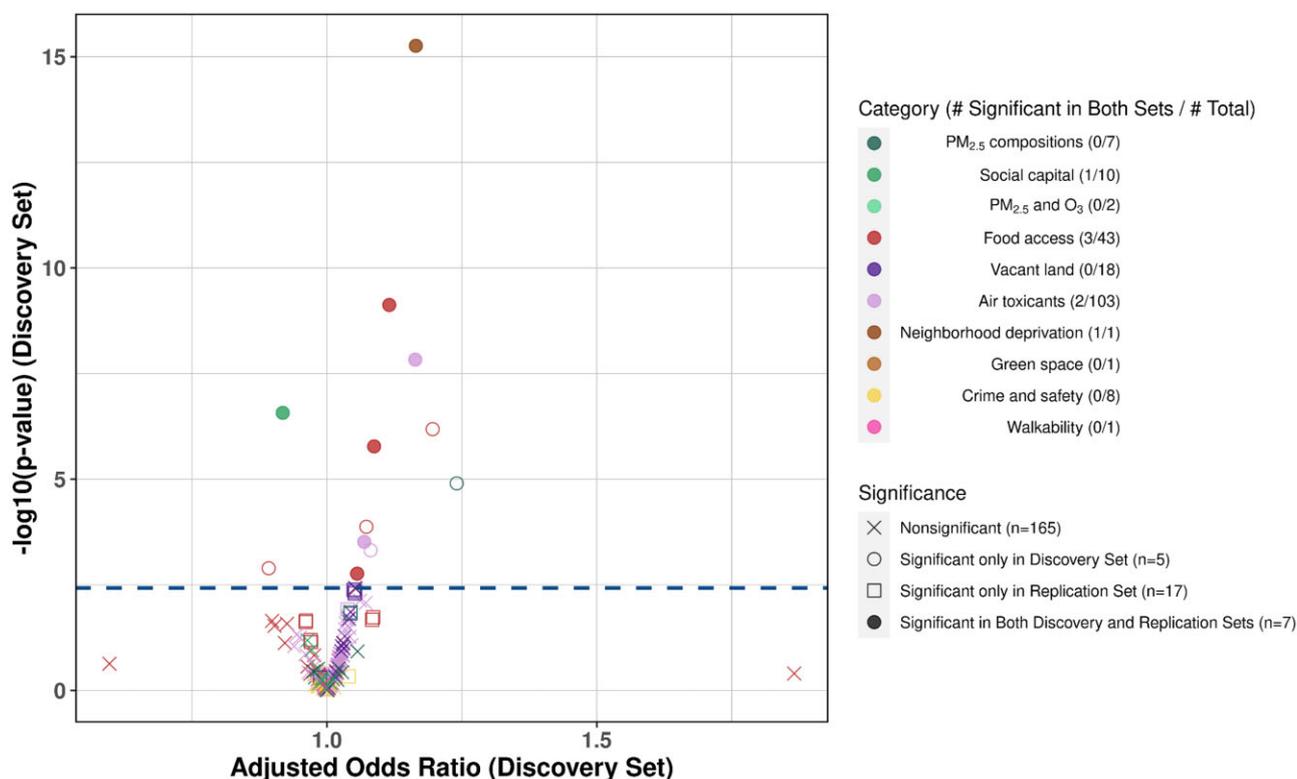


Figure 2. Volcano plot showing the results from Phase 1 of the ExWAS of COVID-19 hospitalization between March 2020 and October 2021 in OneFlorida+ ($n = 50\,368$).

of the top 10 most predictive features of COVID-19 hospitalization in the PES model focusing on all patients. Feature interaction strength measures from the CatBoost models showed that the top feature interacted with NO_3^- is age, which was also reflected in the stratified analyses focusing on those without any comorbidity and aged 18–24 years old, among whom NO_3^- was not identified as a top 10 predictive feature. These findings along with the well-known inflammatory impacts of NO_3^- suggest that long-term exposures to NO_3^- may contribute to COVID-19 severity,⁴¹ especially among older patients. On the other hand, few studies have been conducted to examine air toxicants beyond the criteria air pollutants. In this study, we found that long-term exposures to 2-chloroacetophenone were also associated with increased odds of COVID-19 hospitalization. 2-Chloroacetophenone is primarily used in tear gas and chemical mace, and previous animal studies suggested that chronic exposures to 2-chloroacetophenone have adverse respiratory effects.⁴² In addition, dimethyl phthalate (used in plastics and insect repellents and has been linked to irritation of the eyes, nose, and throat)⁴³ and hexamethylene diisocyanate (used in polyurethane paints and coatings, and has been linked to chronic lung problems)⁴⁴ were also identified among the top 10 predictive features in PES models in all patients and in patients without any comorbidity and aged 18–24 years old, respectively. Future studies are warranted to better identify and understand source and chemical composition of air pollutants that are associated with COVID-19 outcomes.

The association between neighborhood deprivation and COVID-19 hospitalization has been widely documented,^{45,46} which was also observed in this study. However, few studies have been conducted to examine the potential impacts of other built and social environmental factors on COVID-19 hospitalization independent of neighborhood deprivation. In this study, we found

that independent of neighborhood deprivation, lower food access is also associated with higher odds of COVID-19 hospitalization. This finding along with consistent results observed in previous studies suggest that long-term food insecurity may impact COVID-19 outcomes.^{5,47,48} In addition, we also found that higher long-term access to fitness and recreational sports centers is associated with lower odds of COVID-19 hospitalization. While no previous study has examined access to fitness centers and COVID-19 outcomes, the results observed in this study are consistent with the well-established protective associations between physical activity and COVID-19 outcomes.^{49,50}

In addition to the ExWAS, we also developed and validated prediction models of COVID-19 hospitalization. Leveraging the rich spatial and contextual exposome data, we developed PES models in three different target populations. While not statistically significant, we did observe better prediction performance comparing the PES with exposome factors versus prediction models without spatial and contextual exposome factors, especially in individuals without any traditional risk factors (ie, no comorbidity and aged 18–24 years old). These results suggest that spatial and contextual exposome data can provide complementary information to non-spatial factors in disease prediction. Compared with other omics data, the cost to append spatial and contextual exposome measures is relatively low. However, there are several methodological challenges to better leverage spatial and contextual exposome data for disease prediction.⁵¹ For example, the nonsignificant prediction improvements may be partially due to the use of traditional machine learning models which rely on manually spatiotemporally aggregated features. More efforts are needed to develop novel deep learning model architectures to better preserve the rich and heterogeneous spatiotemporal structures in spatial and contextual exposome data.⁵¹

Table 3. Results from the ExWAS of COVID-19 hospitalization between March 2020 and October 2021 in OneFlorida+ (n = 50 368)

Variable	Exposure Category	Standard deviation	Phase 1						Phase 2	
			Discovery set			Replication set			OR (95% CI)	P-value
			OR (95% CI)	P-value	q-value	OR (95% CI)	P-value	q-value		
2-chloroacetophenone (µg/m ³)	Air toxicant	3.42 × 10 ⁻⁶	1.16 (1.10–1.23)	1.48 × 10 ⁻⁸	9.58 × 10 ⁻⁷	1.12 (1.07–1.19)	1.48 × 10 ⁻⁵	6.63 × 10 ⁻⁴	1.13 (1.08–1.18)	8.5 × 10 ⁻⁸
Selenium compounds (µg/m ³)	Air toxicant	1.74 × 10 ⁻⁴	1.07 (1.03–1.11)	3.05 × 10 ⁻⁴	6.56 × 10 ⁻³	1.06 (1.03–1.10)	8.53 × 10 ⁻⁴	1.10 × 10 ⁻²	1.00 (0.97–1.03)	8.5 × 10 ⁻¹
% low food access population with housing units without vehicle access at 1 mile	Food access	1.37	1.06 (1.02–1.09)	1.72 × 10 ⁻³	2.78 × 10 ⁻²	1.07 (1.03–1.10)	7.32 × 10 ⁻⁵	1.78 × 10 ⁻³	1.04 (1.00–1.07)	2.95 × 10 ⁻²
% low food access population with housing units without vehicle access at 1/2 mile	Food access	2.37	1.09 (1.05–1.13)	1.67 × 10 ⁻⁶	5.40 × 10 ⁻⁵	1.09 (1.05–1.13)	1.05 × 10 ⁻⁶	1.02 × 10 ⁻⁴	0.99 (0.96–1.03)	7.46 × 10 ⁻¹
% low food access population that are low income at 1/2 mile	Food access	6.96	1.12 (1.08–1.15)	7.51 × 10 ⁻¹⁰	7.28 × 10 ⁻⁸	1.08 (1.04–1.11)	3.7 × 10 ⁻⁵	1.20 × 10 ⁻³	1.02 (0.98–1.06)	2.66 × 10 ⁻¹
Neighborhood deprivation index	Neighborhood deprivation	2.10	1.16 (1.12–1.21)	5.51 × 10 ⁻¹⁶	1.07 × 10 ⁻¹³	1.11 (1.07–1.15)	2.36 × 10 ⁻⁸	4.58 × 10 ⁻⁶	1.10 (1.07–1.13)	3.49 × 10 ⁻⁹
Number of establishments in fitness and recreational sports centers per 10 000 population	Social capital	9.92 × 10 ⁻¹	0.92 (0.89–0.95)	2.69 × 10 ⁻⁷	1.30 × 10 ⁻⁵	0.93 (0.90–0.96)	1.07 × 10 ⁻⁵	6.63 × 10 ⁻⁴	0.95 (0.92–0.97)	1.87 × 10 ⁻⁵

Odds ratio (OR) and 95% confidence interval (CI) for each standard deviation increase.

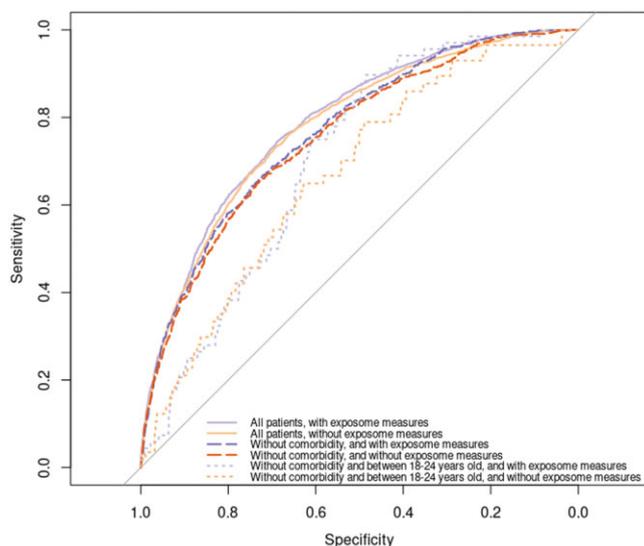


Figure 3. ROC curves of validations using the testing datasets for prediction models of COVID-19 hospitalization between March 2020 and October 2021 in OneFlorida+ ($n = 50\,368$).

Our study has several strengths. Leveraging individual-level EHR data from OneFlorida+ with detailed information on known risk factors of severe COVID-19 and potential confounders, we conducted an ExWAS to examine the associations between COVID-19 hospitalization and long-term exposures to a variety of spatial and contextual exposome factors. In addition, we also developed and validated prediction models of COVID-19 hospitalization. Furthermore, to account for residential mobility and the spatiotemporal dynamic nature of spatial and contextual exposome factors, we spatiotemporally linked exposome data based on residential histories in OneFlorida+.

Several limitations also need to be acknowledged. First, potential exposure misclassifications may exist. While residential histories were considered in exposome data linkage, we do not have information on individuals' time-activity patterns. Second, the EWAS-MLR approach used in the ExWAS did not consider potential interactions or nonlinear associations, which are methodologically challenging for exposome studies focusing on inference given the high-dimensional nature of the data.⁵¹ However, we observed consistent findings from the feature importance measures in prediction models, which was developed using gradient boosting on decision trees that can account for nonlinear associations and high-order interactions. Third, vaccination data were only available at the county-level and SHAP analysis showed that county-level vaccination rates were positively associated with higher likelihoods of COVID-19 hospitalization. Future studies with individual-level vaccination data are warranted to account for the potential ecological fallacy. Fourth, similar to other EHR- or real-world data-based studies,⁹ COVID-19 cases who were not tested or tested outside of OneFlorida+ Clinical Research Network were not included. Fifth, while many spatial and contextual factors have been included in this study, our list is not exhaustive, and future efforts are warranted in the field to continuously improve data standards and ontologies of the spatial and contextual exposome.⁵¹

Conclusions

This spatial and contextual exposome study provides new insights into the role of long-term environmental exposures in

COVID-19 hospitalization. In the ExWAS, we confirmed previously reported associations (ie, food access and neighborhood deprivation) and identified novel environmental factors (ie, 2-chloroacetophenone and access to fitness and recreational sports centers) associated with COVID-19 hospitalization. We also developed and validated prediction models of COVID-19 hospitalization and showed that the spatial and contextual exposome provides complementary predictive information to identify high-risk individuals for COVID-19 hospitalization.

Author contributions

Hui Hu (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Funding acquisition [lead], Investigation [lead], Methodology [lead], Project administration [lead], Resources [lead], Software [lead], Supervision [lead], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]), Francine Laden (Writing—review & editing [equal]), Jaime Hart (Writing—review & editing [equal]), Peter James (Writing—review & editing [equal]), Jennifer Fishe (Funding acquisition [supporting], Writing—review & editing [equal]), William Hogan (Funding acquisition [supporting], Resources [equal], Writing—review & editing [equal]), Elizabeth Shenkman (Funding acquisition [supporting], Resources [equal], Writing—review & editing [equal]), and Jiang Bian (Conceptualization [lead], Data curation [supporting], Funding acquisition [lead], Investigation [lead], Methodology [equal], Project administration [lead], Resources [lead], Software [supporting], Supervision [lead], Validation [lead], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [equal])

Supplementary material

Supplementary material is available at *Exposome* online.

Data availability

The data underlying this article were provided by the OneFlorida+ Clinical Research Network (<https://onefloridaconsortium.org/>), which are made available to researchers with an approved study protocol and data use agreement at <https://onefloridaconsortium.org/front-door/prep-to-research-data-query/>. The data will be shared on reasonable request to the corresponding author with permission of OneFlorida+ Clinical Research Network.

Funding

Research reported in this publication was supported in part by the National Institute of Environmental Health Sciences under award numbers R21ES032762 and P30ES000002; in part by OneFlorida+ Clinical Research Network, funded by the Patient-Centered Outcomes Research Institute numbers CDRN-1501-26692 and RI-CRN-2020-005; in part by the OneFlorida Cancer Control Alliance, funded by the Florida Department of Health's James and Esther King Biomedical Research Program number 4KB16; and in part by the University of Florida Clinical and Translational Science Institute and its Clinical and Translational Science Award (CTSA) hub partner, Florida State University (FSU), which are supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health grant numbers UL1TR001427, KL2TR001429, and TL1TR001428. The content is solely the responsibility of the

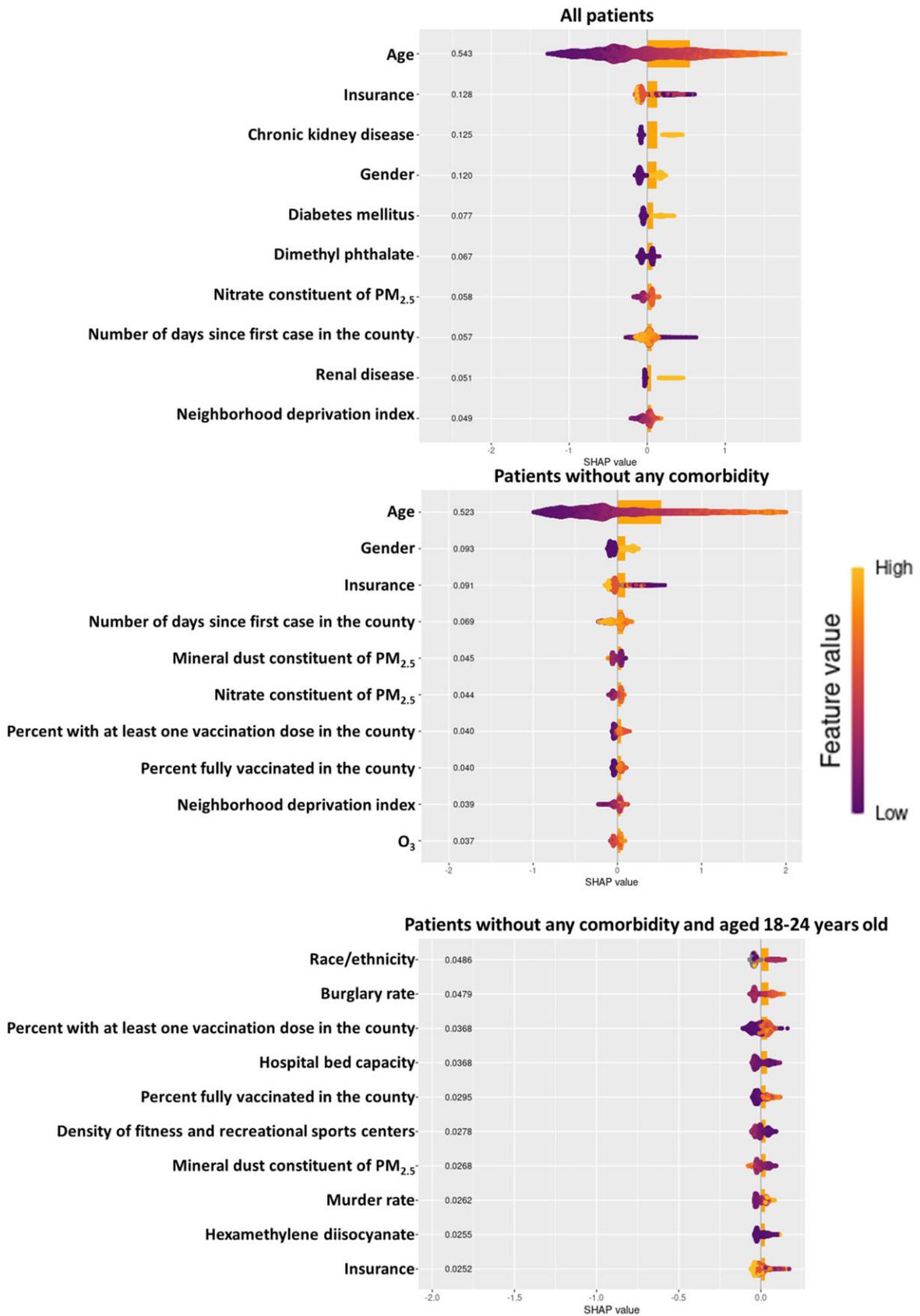


Figure 4. Beeswarm plot and mean absolute SHAP values for the top 10 predictive features from each PES models of COVID-19 hospitalization between March 2020 and October 2021 in OneFlorida+ (*n* = 50 368).

authors and does not necessarily represent the official views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology, the OneFlorida+ Clinical Research Network, the UF-FSU Clinical and Translational Science Institute, the Florida Department of Health, or the National Institutes of Health.

Conflict of interest statement

None declared.

References

1. CDC. COVID data tracker (published 2022). Accessed March 31, 2022. https://covid.cdc.gov/covid-data-tracker/#cases_casesinlast7days
2. Gao YD, Ding M, Dong X, et al. Risk factors for severe and critically ill COVID-19 patients: a review. *Allergy*. 2021;76(2):428–455.
3. Weaver AK, Head JR, Gould CF, Carlton EJ, Remais JV. Environmental factors influencing COVID-19 incidence and severity. *Annu Rev Public Health*. 2022;43(1):271–291.
4. Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*. 2005;14(8):1847–1850.
5. Hu H, Zheng Y, Wen X, et al. An external exposome-wide association study of COVID-19 mortality in the United States. *Sci Total Environ*. 2021;768:144832.
6. Shenkman E, Hurt M, Hogan W, et al. OneFlorida Clinical Research Consortium: linking a clinical and translational science institute with a community-based distributive medical education model. *Acad Med*. 2018;93(3):451–455.
7. Hogan WR, Shenkman EA, Robinson T, et al. The OneFlorida Data Trust: a centralized, translational research data infrastructure of statewide scope. *J Am Med Inform Assoc*. 2022;29(4):686–693.
8. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc*. 2014;21(4):578–582.
9. Bowe B, Xie Y, Gibson AK, et al. Ambient fine particulate matter air pollution and the risk of hospitalization among COVID-19 positive individuals: cohort study. *Environ Int*. 2021;154:106564.
10. USEPA. Technical information about fused air quality surface using downscaling tool: metadata description (published 2016). Accessed July 1, 2022. https://www.epa.gov/sites/default/files/2016-07/documents/data_fusion_meta_file_july_2016.pdf
11. USEPA. Downscaler model for predicting daily air pollution (published online July 6, 2015). Accessed June 30, 2022. https://19january2017snapshot.epa.gov/air-research/downscaler-model-predicting-daily-air-pollution_.html
12. van Donkelaar A, Martin RV, Li C, Burnett RT. 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;53(5):2595–2611.
13. Logue JM, Small MJ, Robinson AL. Evaluating the national air toxics assessment (NATA): comparison of predicted and measured air toxics concentrations, risks, and sources in Pittsburgh, Pennsylvania. *Atmos Environ* (1994). 2011;45(2):476–484.
14. Rhew IC, Vander Stoep A, Kearney A, Smith NL, Dunbar MD. Validation of the normalized difference vegetation index as a measure of neighborhood greenness. *Ann Epidemiol*. 2011;21(12):946–952.
15. Thomas J, Zeller L. National walkability index user guide and methodology (published May 17, 2021). Accessed April 18, 2022. <https://www.epa.gov/smartgrowth/national-walkability-index-user-guide-and-methodology>
16. USDA. Introduction to the food access research atlas (published 2021). Accessed June 30, 2022. <https://gisportal.ers.usda.gov/portal/apps/experiencebuilder/experience/?id=a53ebd7396cd4ac3a3ed09137676fd40>
17. Garvin E, Branas C, Keddem S, Sellman J, Cannuscio C. More than just an eyesore: local insights and solutions on vacant land and urban health. *J Urban Health*. 2013;90(3):412–426.
18. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83(6):1041–1062.
19. Rupasingha A, Goetz SJ, Freshwater D. The production of social capital in US counties. *J Socio Econ*. 2006;35(1):83–101.
20. Barnett-Ryan C. Introduction to the uniform crime reporting program. In: Lynch J., Addington L (eds). *Understanding Crime Statistics*. Cambridge, NY: Cambridge University Press 2007: 55–89.
21. JHU CSSE. Global cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) (published online 2020).
22. U.S. Department of Health, Human Services. COVID-19 reported patient impact and hospital capacity by facility (published online December 12, 2020). Accessed June 30, 2022. <https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u>
23. Patel C. Environment-wide association studies to connect multiple personal exposures to health. *Environ Health Perspect*. 2013;2013(1):5848.
24. McGinnis DP, Brownstein JS, Patel CJ. Environment-wide association study of blood pressure in the National Health and Nutrition Examination Survey (1999–2012). *Sci Rep*. 2016;6:30373.
25. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Soft*. 2011;45(3):1–67.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57(1):289–300.
27. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *arXiv [csAI]* (published online May 22, 2017). Accessed January 02, 2023, <http://arxiv.org/abs/1705.07874>
28. Coker ES, Cavalli L, Fabrizi E, et al. The effects of air pollution on COVID-19 related mortality in Northern Italy. *Environ Resour Econ (Dordr)*. 2020;76(4):611–634.
29. Konstantinoudis G, Padellini T, Bennett J, Davies B, Ezzati M, Blangiardo M. Long-term exposure to air-pollution and COVID-19 mortality in England: a hierarchical spatial analysis. *Environ Int*. 2021;146:106316.
30. Hendryx M, Luo J. COVID-19 prevalence and fatality rates in association with air pollution emission concentrations and emission sources. *Environ Pollut*. 2020;265(Pt A):115126.
31. Ogen Y. Assessing nitrogen dioxide (NO₂) levels as a contributing factor to coronavirus (COVID-19) fatality. *Sci Total Environ*. 2020;726:138605.
32. Conticini E, Frediani B, Caro D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ Pollut*. 2020;261:114465.

33. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. *Sci Total Environ*. 2020;727:138704.
34. Fattorini D, Regoli F. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. *Environ Pollut*. 2020;264:114732.
35. Cole MA, Ozgen C, Strobl E. Air pollution exposure and Covid-19 in Dutch municipalities. *Environ Resour Econ (Dordr)*. 2020;76(4):581–610.
36. Liang D, Shi L, Zhao J, et al. Urban air pollution may enhance COVID-19 case-fatality and mortality rates in the United States. *Innovation (Camb)*. 2020;1(3):100047.
37. Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. *Sci Adv*. 2020;6(45):eabd4049. <https://10.1126/sciadv.abd4049>
38. 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*. 2022;215(Pt 1):114155.
39. Mendy A, Wu X, Keller JL, et al. Air pollution and the pandemic: long-term PM_{2.5} exposure and disease severity in COVID-19 patients. *Respirology*. 2021;26(12):1181–1187.
40. Kelly FJ, Fussell JC. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos Environ (1994)*. 2012;60:504–526.
41. Liu Q, Baumgartner J, Zhang Y, Liu Y, Sun Y, Zhang M. Oxidative potential and inflammatory impacts of source apportioned ambient air pollution in Beijing. *Environ Sci Technol*. 2014;48(21):12920–12929.
42. Recer GM, Johnson TB, Gleason AK. An evaluation of the relative potential public health concern for the self-defense spray active ingredients oleoresin capsicum, o-chlorobenzylidene malononitrile, and 2-chloroacetophenone. *Regul Toxicol Pharmacol*. 2002;36(1):1–11.
43. Schettler T. Human exposure to phthalates via consumer products. *Int J Androl*. 2006;29(1):134–139; discussion 181–185.
44. Wisnewski AV, Srivastava R, Herick C, et al. Identification of human lung and skin proteins conjugated with hexamethylene diisocyanate in vitro and in vivo. *Am J Respir Crit Care Med*. 2000;162(6):2330–2336.
45. Jannot AS, Countouris H, Van Straaten A, et al.; AP-HP/Universities/INSERM COVID-19 Research Collaboration. Low-income neighbourhood was a key determinant of severe COVID-19 incidence during the first wave of the epidemic in Paris. *J Epidemiol Community Health*. 2021;75(12):1143–1146.
46. Lewis NM, Friedrichs M, Wagstaff S, et al. Disparities in COVID-19 incidence, hospitalizations, and testing, by area-level deprivation—Utah, March 3–July 9, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(38):1369–1373.
47. Ariya M, Karimi J, Abolghasemi S, et al. Food insecurity arises the likelihood of hospitalization in patients with COVID-19. *Sci Rep*. 2021;11(1):20072.
48. Huang H. Food environment inequalities and moderating effects of obesity on their relationships with COVID-19 in Chicago. *Sustain Sci Pract Policy*. 2022;14(11):6498.
49. Sallis R, Young DR, Tartof SY, et al. Physical inactivity is associated with a higher risk for severe COVID-19 outcomes: a study in 48 440 adult patients. *Br J Sports Med*. 2021;55(19):1099–1105.
50. Ezzatvar Y, Ramírez-Vélez R, Izquierdo M, Garcia-Hermoso A. Physical activity and risk of infection, severity and mortality of COVID-19: a systematic review and non-linear dose–response meta-analysis of data from 1 853 610 adults. *Br J Sports Med*. 2022;56(20):1188–1193.
51. Hu H, Liu X, Zheng Y, et al. Methodological challenges in spatial and contextual exposome-health studies. *Crit Rev Environ Sci Technol*. 2023;53(7):827–846.