UC Davis
UC Davis Previously Published Works

Title

Associations between greenness and predicted COVID-19-like illness incidence in the United States and the United Kingdom.

Permalink

https://escholarship.org/uc/item/6xb1d1dr

Journal

Environmental Epidemiology, 7(1)

Authors

Chen, Kelly

Klompmaker, Jochem

Roscoe, Charlotte

et al.

Publication Date

2023-02-01

DOI

10.1097/EE9.0000000000000244

Peer reviewed

Associations between greenness and predicted COVID-19–like illness incidence in the United States and the United Kingdom

Kelly Chen[a], Jochem O. Klompmaker [ID][a,b,*], Charlotte J. Roscoe[a], Long H. Nguyen[c],
David A. Drew[c,d], Peter James[a,e], Francine Laden[a,b], Daniela Fecht[f], Weiyi Wang[f], John Gulliver[g], Jonathan Wolf[h],
Claire J. Steves[i], Tim D. Spector[i], Andy T. Chan[c,d,j], Jaime E. Hart[a,b]

Background: Green spaces may be protective against COVID-19 incidence. They may provide outdoor, ventilated, settings for physical and social activities and therefore decrease transmission risk. We examined the association between neighborhood greenness and COVID-19–like illness incidence using individual-level data.
Methods: The study population includes participants enrolled in the COVID Symptom Study smartphone application in the United Kingdom and the United States (March–November 2020). All participants were encouraged to report their current health condition and suspected risk factors for COVID-19. We used a validated symptom-based classifier that predicts COVID-19–like illness. We estimated the Normalized Difference Vegetation Index (NDVI), for each participant's reported neighborhood of residence for each month, using images from Landsat 8 (30 m^2). We used time-varying Cox proportional hazards models stratified by age, country, and calendar month at study entry and adjusted for the individual- and neighborhood-level risk factors.
Results: We observed 143,340 cases of predicted COVID-19–like illness among 2,794,029 participants. Neighborhood NDVI was associated with a decreased risk of predicted COVID-19–like illness incidence in the fully adjusted model *(hazard ratio = 0.965, 95% confidence interval = 0.960, 0.970, per 0.1 NDVI increase)*. Stratified analyses showed protective associations among U.K. participants but not among U.S. participants. Associations were slightly stronger for White individuals, for individuals living in rural neighborhoods, and for individuals living in high-income neighborhoods compared to individuals living in low-income neighborhoods.
Conclusions: Higher levels of greenness may reduce the risk of predicted COVID-19–like illness incidence, but these associations were not observed in all populations.

Keywords: Natural environment; Built Environment; Green space; COVID

[a]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; [b]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [c]Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [d]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [e]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts; [f]MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom; [g]Centre for Environmental Health and Sustainability, George Davies Centre, University of Leicester, Leicester, United Kingdom; [h]ZOE Global, London, United Kingdom; [i]Kings College Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; and [j]Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

Kelly Chen and Jochem O. Klompmaker contributed equally as the first author.

Editors' note: Related articles appear on pages XXX and XXX.

What this study adds

This study showed that greenness is associated with a lower risk of incident-predicted COVID-19–like illness, after adjustment for potential confounders. This study is based on individual-level data from ~2.8 million U.K. and U.S. participants from the COVID Symptom Study smartphone application study. We used a symptom-based classifier that predicts COVID-19–like illness as our outcome measure to overcome challenges regarding COVID-19 diagnosis. Stratified analyses showed protective associations among U.K. participants but not among U.S. participants. Our findings are in line with several ecological studies that examined associations between greenness and COVID-19 incidence.

Introduction

The global spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the virus responsible for COVID-19, has triggered a worldwide public health emergency.[1] As of January 2023, more than 660 million cases of COVID-19 have been documented worldwide, and more than 6.5 million deaths had been recorded.[2] Evidence from case reports and cohort studies has demonstrated that SARS-CoV-2 spreads via airborne and respiratory droplets.[3,4] Sustained use of public health measures, such as social distancing and mask use, may remain critical tools to curtail the spread of the virus even as vaccination efforts expand.[5,6]

In this context, green spaces such as public parks and gardens may alter dynamics in COVID-19 transmission.[7–10] Green spaces provide outdoor settings for physical activities and social interactions. Due to increased ventilation outdoors, the transmission risk is substantially lower. Hence, meeting individuals in green outdoor spaces instead of indoors could reduce the risk of COVID-19 infection. In addition, green spaces may affect cardiovascular and respiratory health by mitigating air pollution exposure and reducing stress,[11–15] and in turn affect COVID-19 morbidity.[16–18]

Several studies observed protective associations between greenness and risk of COVID-19 morbidity and mortality;[19–27] however, contradictory findings were reported in two studies.[28,29] A major limitation of these studies is that they relied on ecological data and can suffer from ecological fallacy. Moreover, these studies are prone to potential confounding as exposures and potential confounders may vary within geographical areas. In addition, challenges regarding COVID-19 diagnosis have led to under-reporting of COVID-19 cases.[30–32] Especially during the first months of the pandemic, health jurisdictions were overwhelmed and lacked adequate testing; hence, many COVID-19 cases have not been identified.[30–32]

In this study, we examined associations of neighborhood greenness with COVID-19–like illness incidence using individual-level data from the COVID Symptom Study (CCS) smartphone application. We included ~2.8 million participants in our cohort; all participants were encouraged to report daily on their current health condition and suspected risk factors for COVID-19. To overcome the challenges of COVID-19 diagnosis, we used a symptom-based classifier that predicts COVID-19–like illness as our outcome measure. We evaluated whether these associations with predicted COVID-19–like illness were modified by socioeconomic and lifestyle-related factors. We hypothesized that increased exposure to greenness would be associated with reduced predicted COVID-19–like illness incidence.

Methods

Study population

The study population included all users of the CSS smartphone application as of November 2020. The application is a freely available program developed by Zoe Global Ltd. in collaboration with Massachusetts General Hospital and King's College London.[33] The app launched on 24 March 2020 in the United Kingdom and on 29 March 2020 in the United States. Participant recruitment came from general and social media outreach, as well as direct invitations from investigators of long-running prospective cohorts to enrolled study participants.[34] The CSS includes all participants enrolled in the United Kingdom and the United States (n: overall = 4,633,679; United Kingdom = 4,273,668; United States = 360,011).

Participants reported demographic information and comorbidities at baseline, including their neighborhood of residence and status as frontline healthcare worker.[35] At first use and with daily reminders, participants were encouraged to report on their current health condition and a series of suspected risk factors for COVID-19 to allow for longitudinal, prospective collection of symptoms and COVID-19 testing results. These questions included items to determine if the participant felt physically normal, and if not, their symptoms, including fever, persistent cough, fatigue, loss of smell/taste, and diarrhea, as well as suspected or confirmed contact with COVID-19. Participants were asked if they had been tested for COVID-19 and, if yes, the results (negative, waiting, or positive).

We excluded individuals from the CSS cohort who did report any COVID-19 symptoms or a positive COVID-19 test within 24 hours of enrollment or had <24 hours of follow-up time (n ~ 1.5M), not reported baseline information (n ~ 0.14M), or lacked up-to-date neighborhood SES indicators (n ~ 0.36M, Northern Ireland), leaving 2,796,322 participants total.

At enrollment, participants provided informed consent to the use of aggregated information for research purposes and agreed to applicable privacy policies and terms of use. This research study was approved by the Massachusetts General Brigham Institutional Review Board and King's College London ethics committee. This protocol is registered with ClinicalTrials.gov (NCT04331509).

Outcome assessment

We assessed predicted COVID-19–like illness incidence occurring between the first response on the CCS app and 30 November 2020. We used a symptom-based classifier that predicts COVID-19 as our primary outcome measure, previously described in detail.[33] This prediction model was developed to address inadequacies in relying solely on reports of positive COVID-19 tests, which reflects both access to testing and variable delays and heterogeneities between symptoms and consequent testing.[36]

The symptom-based classifier was developed among participants who reported on symptoms, had been tested for SARS-CoV-2 by RT-PCR, and had reported tests result on the app questionnaire. The prediction model achieved a sensitivity of 0.65 (95% confidence interval [CI] = 0.62, 0.67) and a specificity of 0.78 (95% CI = 0.76, 0.80) in the United Kingdom. Additional validation in the U.S. participants indicated a comparable sensitivity of 0.66 (95% CI = 0.62, 0.69) and specificity of 0.84 (95% CI = 0.82, 0.85). The prediction model was developed before some major variants were common, therefore we ended the follow-up for this analysis in November 2020. As the prediction model was developed among participants who reported on symptoms and will likely not detect any cases that are infected with COVID-19 but do not experience any symptoms, we decided to use the term predicted COVID-19–like illness.

For comparison, we also used incident COVID-19 identified by a nasal swab PCR test. We note that relying solely on COVID-19 test results could lead to bias, as especially in the first months of the pandemic, test availability/access was limited and likely differed between participants.

Assessment of neighborhood greenness

We estimated the Normalized Difference Vegetation Index (NDVI), a satellite-derived, continuous measure of greenness, for each neighborhood of residence of the participants. NDVI is calculated as the ratio of land surface reflectance of visible and near-infrared ranges of the electromagnetic spectrum. NDVI ranges from –1 to 1, with larger values indicating higher levels of vegetative density, negative values correspond to water. In this study, we used images from Landsat 8 (Collection 1 Tier 1 of the Operation Land Imager instrument), which provides images every 16 days at the 30 m resolution.[37] We used Google Earth Engine (https://earthengine.google.com) to generate cloud-free Landsat composites for the United States and the United Kingdom after setting negative NDVI values to zero.

For each month of follow-up (March–November 2020), we calculated the spatially weighted 3-month average NDVI of each neighborhood in the United States and the United Kingdom based on satellite images from that month and the two months prior (e.g., for August, NDVI was based on images from June, July, and August) to incorporate the effect of seasonal greenness. In the United States, we calculated NDVI for Zip Code Tabulation Areas (ZCTAs) based on the U.S. Census Bureau Tiger dataset of 2010. In England, we used lower layer super output areas (LSOA) administrative boundaries, which have an average population of 1500 inhabitants and are regularly used in small-area environment and health analyses.[38] We used LSOA from 2011 Census data as the neighborhood level for participants in Wales, and Datazones based on 2011 Census data for those in Scotland. The median size of the neighborhood of residence of the participants included in this study is 0.71 km2 (25th percentile = 0.31 km^2, 75th percentile = 6.22 km^2).

Assessment of potential confounders

We obtained the following covariates of interest from the baseline CCS questionnaire for each participant: sex at birth, race/ethnicity, age, BMI (calculated from reported height and weight), current smoking status, history of comorbidities (diabetes, heart disease, lung disease, and/or kidney disease), limitations to physical activity (requiring stay-at-home, limiting activities, and/or regular use of mobility aids), and frontline healthcare worker status (defined as healthcare workers who reported direct patient contact). Missing data for categorical variables were included as a missing indicator.

We evaluated neighborhood-level socioeconomic status (SES) metrics derived from census data, based on the census unit containing the participant's reported neighborhood at baseline. For U.S. participants, we evaluated the median annual income and the proportion of individuals aged ≥ 25 years with a Bachelor's degree or higher, using U.S. Census Bureau's 2019 American Community Survey 2015–2019 5-year estimates.[39] For U.K. participants, we obtained Index of Multiple Deprivation measures for the income and education domains from the most recently available Index of Multiple Deprivation data, at the LSOA-level for England,[40] at the LSOA-level for Wales[41] and at the Datazone-level for Scotland.[42] We examined population density for all U.S. ZCTAs calculated from Census data (2019), and for all English (2019) and Welsh (2019) LSOAs and Scottish Datazones (2020). For U.S. participants, we assessed spatially weighted average ZCTA-level $PM_{2.5}$ estimates (2015) based on an ensemble model.[43] For U.K. participants, we assessed spatially weighted LSOA- and Datazone-level $PM_{2.5}$ estimates (2015) based on a chemical transport model-incorporated land use regression model.[44] Detailed information about these models can be found elsewhere.[43,44] In addition, the county of the ZCTA and the local authority of the LSOA and Datazones were linked to the participants.

Statistical analysis

We used time-varying Cox proportional hazards models to evaluate associations of NDVI with predicted COVID-19–like illness incidence. In this open cohort with left entry, follow-up time began when participants initially reported on the app. Participants were followed until they developed predicted COVID-19–like illness, or reached the end of the follow-up time (30 November 2020), whichever occurred first. We used time-varying 3-month moving average NDVI exposure data as our primary independent variable. Time-varying NDVI exposure data and corresponding follow-up time was mapped to each individual and updated each time they logged in to the app to provide updated symptom information.

In our analysis, we specified a basic and fully adjusted model and a series of basic models with separate adjustments of covariates. In the basic model, we adjusted for calendar month (categorical indicator, to control for temporal effects), and used strata for country, age (<18, 18–65, and >65 years), and calendar month at study entry, forming strata-specific baseline hazards. In the fully adjusted model, we additionally adjusted for sex, race, BMI, history of comorbidities, smoking status, frontline healthcare worker status, population density (quartiles), and neighborhood-level income and education (quartiles). We assessed potential deviations from linearity using penalized splines. Analyses were performed for the full cohort and among U.K. and U.S. participants.

We tested for differences in the relationship between greenness and predicted COVID-19–like illness incidence by conducting stratified analyses by race, residence in urban (population density > 1000 persons/km^2) or rural areas (population density ≤ 1000 persons/miles^2), frontline healthcare worker status, limitations to activity and neighborhood income. For sensitivity analyses, we specified a fully adjusted model without population density; adjusted for the season of the year instead of the month of the year; adjusted for $PM_{2.5}$; included a frailty term for county/local authority (to account for spatial clustering) and used summer (June-August) NDVI instead of a 3-month moving average NDVI. In addition, we ran a model without adjustment for calendar month. We also used incident COVID-19 identified by a PCR test. We analyzed all data using R software 4.0.2 (http://cran-r-project.org).

Results

Among the 2,794,029 eligible cohort members, we observed 143,340 cases of predicted COVID-19–like illness over 347,015,836 person-days of follow-up. In the highest NDVI quartile, we observed a higher portion of men, non-Hispanic White participants, never and former smokers, participants with heart disease, and participants that reported health problems requiring stay-at-home orders compared to the lower NDVI quartiles (Table 1). The majority of participants lived in the United Kingdom (n = 2,557,464) with a smaller number in the United States (n = 236,565, eTable 1; http://links.lww.com/EE/A218). The median (IQR) age of all participants was 46 (27) years; participants in the United States were generally older than those in the United Kingdom. In both the U.K. and U.S. participants were primarily White, women, non-frontline healthcare workers, never or former smokers, and did not report physical activity limitations. The Pearson correlation of the 3-month moving average neighborhood NDVI with population density and with summer average neighborhood NDVI was –0.51 and 0.71, respectively.

We found 22,784 incident COVID-19 cases identified by a PCR test over 357,133,583 person-days of follow-up. The median age, percent non-Hispanic White, percent never smoker, and percent frontline healthcare worker were lower among predicted COVID-19–like illness cases compared to COVID-19 cases identified by a PCR test (eTable 2; http://links.lww.com/EE/A218). The percent Mixed/Other race, percent with lung disease, percent with limitations to physical activity (regular use of mobility aid, health problems requiring stay-at-home, health problems limiting activities), and population density were higher among predicted COVID-19–like illness cases. The number of predicted COVID-19–like illness cases was highest during the start of the follow-up period (March–April), while the number of COVID-19 identified by PCR test cases was highest in October and November (eFigure 1; http://links.lww.com/EE/A218). The ~1.5 M individuals that were excluded because they had <24 hours of follow-up time were on average younger, had less greenness, higher population density, and

Table 1.

Characteristics of all CCS participants records by NDVI quartiles from 24 March 2020 to 30 November 2020 after excluding those who did not report baseline information, reported any symptoms or a positive COVID-19 test at or within 24 hours of enrollment, or had <24 hours of follow-up time (n = 2,794,029).[a,b]

	Overall Median [25, 75 pct]/%	NDVI q1 Median [25, 75 pct]/%	NDVI q2 Median [25, 75 pct]/%	NDVI q3 Median [25, 75 pct]/%	NDVI q4 Median [25, 75 pct]/%
N (person records)	14,043,710	3,510,986	3,510,886	3,510,996	3,510,842
Age	50.0 [36.0, 62.0]	45.0 [32.0, 58.0]	49.0 [35.0, 61.0]	51.0 [37.0, 63.0]	54.0 [40.0, 64.0]
Gender (%)					
Women	61.1	62	61.4	60.7	60.2
Men	38.9	38	38.5	39.3	39.8
Prefer not to say	0.0	0.1	0.0	0.0	0.0
(Missing)	0.0	0.0	0.0	0.0	0.0
Race/ethnicity (%)					
White, non-Hispanic	89.9	82.9	89.6	92.3	94.7
Hispanic/Latino	0.3	0.6	0.3	0.2	0.2
Black	0.6	0.9	0.6	0.4	0.3
Asian	2.0	3.0	2.1	1.8	1.2
Mixed/other race	2.6	3.7	2.7	2.3	1.8
Prefer not to say	0.1	0.1	0.1	0.1	0.1
(Missing)	4.5	8.7	4.6	3.0	1.7
BMI (median [IQR])	25.29 [22.5, 29.0]	25.1 [22.3, 28.8]	25.5 [22.6, 29.3]	25.4 [22.5, 29.1]	25.3 [22.5, 28.8]
Smoker status (%)					
Never smoker	43.2	35.5	42.5	45.7	49.0
Former smoker	12.9	11.1	12.8	13.2	14.3
Current smoker	3.9	4.0	4.2	3.9	3.7
(Missing)	40.0	49.4	40.5	37.2	33.0
Has diabetes (%)	3.5	3.2	3.6	3.6	3.5
Has heart disease (%)	3.2	2.7	3.2	3.4	3.8
Has lung disease (%)	10.8	10.9	11	10.8	10.6
Has kidney disease (%)	0.8	0.7	0.8	0.8	0.8
Frontline healthcare worker, yes (%)	4.2	4.2	4.3	4.1	4.2
Health problems requiring stay-at-home (%)	5.1	4.3	5.2	5.3	5.4
Regular use of mobility aid (%)	2.2	1.9	2.3	2.3	2.2
Health problems limiting activities (%)	8.6	8.1	8.9	8.7	8.6
Population density (persons/mi²)	4217.1 [750.5, 10183.5]	10630.4 [4620.2, 21554.3]	6538.7 [2136.7, 11383.5]	3591.5 [773.0, 7685.7]	655.8 [229.3, 2499.7]
Urban/rural					
Rural (≤1000 p/mi²)	24.2	8.8	14.6	24.0	49.5
Urban (>1000 p/mi²)	62.2	80.4	71.8	61.7	36.4
Missing	13.2	10.8	13.7	14.2	14.0
English IMD income	7.0 [5.0, 9.0]	6.0 [4.0, 8.0]	7.0 [5.0, 9.0]	8.0 [5.0, 9.0]	8.0 [6.0, 9.0]
English IMD Education Score	7.0 [5.0, 9.0]	7.0 [5.0, 9.0]	7.0 [5.0, 9.0]	8.0 [5.0, 9.0]	8.0 [6.0, 9.0]
U.S. Census—median household income	78650.0 [60430.0, 102242.0]	78484.0 [59702.0, 100745.0]	79873.0 [62422.0, 102650.0]	80553.0 [61983.0, 103728.0]	76569.0 [58384.0, 101518.0]
U.S. Census—percent 25+ years with bachelors	44.7 [31.0, 59.4]	45.3 [31.3, 60.8]	45.3 [32.9, 59.3]	45.8 [31.9, 60.1]	42.1 [28.3, 57.6]
PM₂.₅ (μg/m³)	9.6 [8.6, 10.2]	10.1 [9.2, 11.1]	9.8 [9.2, 10.4]	9.6 [8.8, 10.1]	8.8 [7.8, 9.4]

[a]NDVI q1: 0.000–0.358, NDVI q2: 0.358–0.455, NDVI q3: 0.455–0.547, NDVI q4: 0.547–0.827.
[b]The median (25th–75 percentile) area of the neighborhoods in urban areas is 0.45 km² (0.26–1.16) and the median area of the neighborhood in rural areas is 16.80 km² (3.95–44.12).
IMD indicates Index of Multiple Deprivation.

lower neighborhood SES than the ~2.8M participants that were included (eTable 3; http://links.lww.com/EE/A218).

We observed a linear association between neighborhood NDVI and predicted COVID-19–like illness incidence (eFigure 2; http://links.lww.com/EE/A218) and therefore present associations per 0.1 NDVI increase. In the basic model, neighborhood NDVI was associated with a decreased risk of incident-predicted COVID-19–like illness (hazard ratio = 0.928, 95% CI = 0.924, 0.932, per 0.1 NDVI increase). The effect estimate was robust to adjustment for individual-level covariates (Figure 1). After adjustment for area-level urbanization or SES, the association between neighborhood NDVI and COVID-19 slightly attenuated but remained protective. *In the fully adjusted model, we observed a hazard ratio of 0.965 (95% CI = 0.960, 0.970) per*

0.1 NDVI increase (Table 2). Associations among U.K. participants were similar to associations in the full cohort; no association was observed among U.S. participants.

In the full cohort and among U.K. participants, but not among U.S. participants, protective associations of neighborhood NDVI were slightly stronger for White individuals compared to non-White individuals, for individuals living in rural neighborhoods compared to individuals living in urban neighborhoods, and individuals living in higher income neighborhoods compared to individuals living in lower-income neighborhoods. Associations by the healthcare worker and physical activity limitation status differed between U.K. and U.S. participants.

In the full cohort and among U.K. participants, associations of neighborhood NDVI with incident COVID-19 identified by a
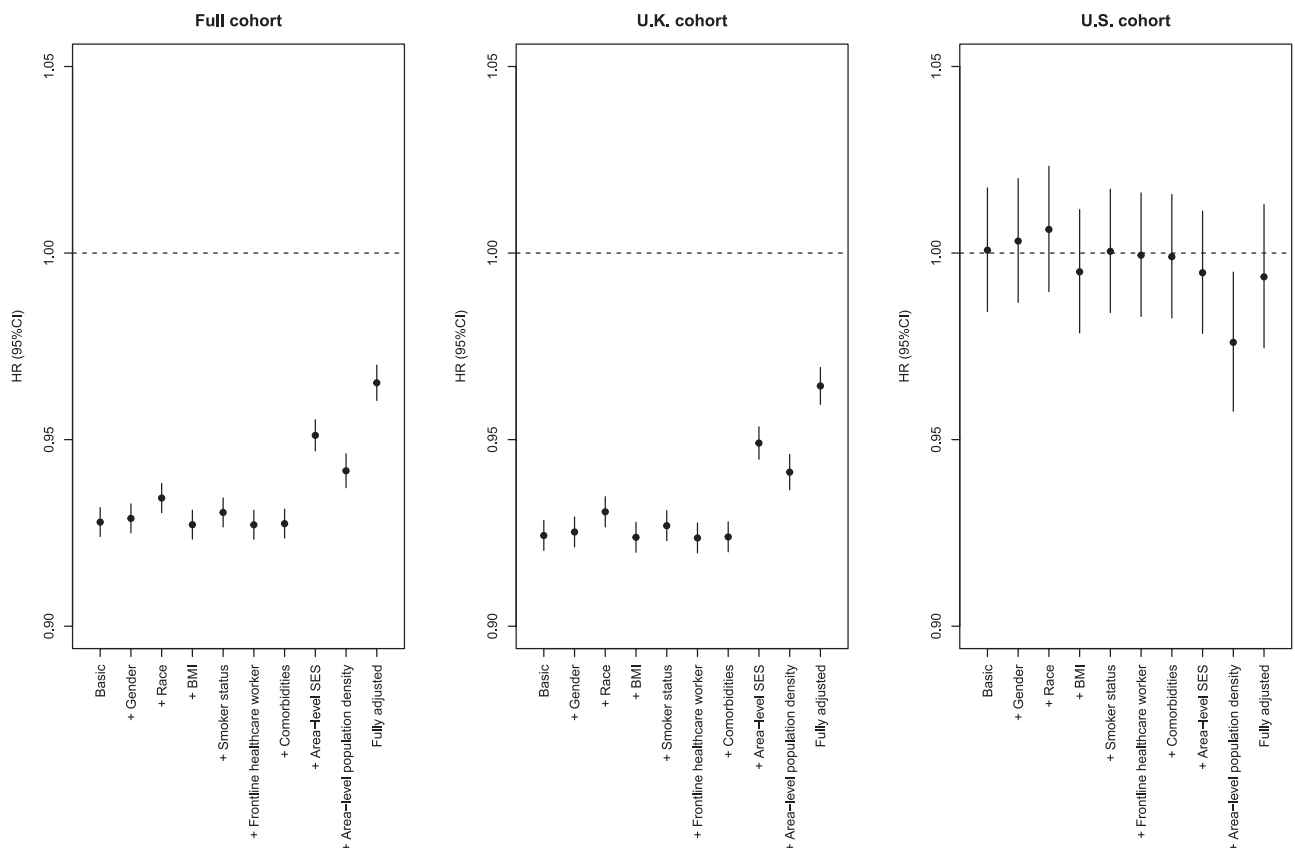
Table 2.

Hazard ratios (and 95% CIs) for predicted COVID-19–like illness incidence presented by a 0.1 increase in 3-month moving average neighborhood NDVI among participants in the CCS cohort (_n_ = 2,794,029) during follow-up from 24 March 2020 to 30 November 2020.

	Full cohort		U.K. cohort		U.S. cohort	
	Cases (person days) [a]	HR (95% CI) [b]	Cases (person days) [a]	HR (95% CI) [b]	Cases (person days) [a]	HR (95% CI) [b]
Full cohort	143,340 (347,015,836)	0.965 (0.960, 0.970)	136,682 (320,079,143)	0.964 (0.959, 0.969)	6,658 (26,936,693)	0.994 (0.975, 1.013)
Stratified by						
No limitations to physical activity	123,731 (307,698,408)	0.965 (0.960, 0.970)	118,333 (283,865,038)	0.964 (0.959, 0.969)	5,398 (23,833,370)	1.000 (0.979, 1.022)
Limitations to physical activity	19,609 (39,317,429)	0.967 (0.954, 0.980)	18,349 (36,214,105)	0.969 (0.955, 0.982)	1,260 (3,103,323)	0.961 (0.920, 1.004)
White, non-Hispanic	113,578 (320,353,202)	0.962 (0.957, 0.967)	108,594 (297,485,198)	0.962 (0.956, 0.967)	4,984 (22,868,004)	0.993 (0.971, 1.016)
Non-White	29,762 (26,662,634)	0.982 (0.971, 0.993)	28,088 (22,593,945)	0.983 (0.972, 0.994)	1,674 (4,068,689)	0.972 (0.936, 1.010)
Urban (> 1000 p/mi²)	92,106 (217,048,999)	0.970 (0.964, 0.976)	88,058 (200,588,817)	0.970 (0.964, 0.977)	4,048 (16,460,182)	0.979 (0.953, 1.006)
Rural (≤ 1000 p/mi²)	27,065 (85,637,621)	0.955 (0.944, 0.966)	24,492 (75,290,748)	0.947 (0.935, 0.960)	2,573 (10,346,873)	0.999 (0.971, 1.028)
Healthcare worker	9,733 (13,886,774)	0.981 (0.963, 1.000)	8,953 (11,761,507)	0.964 (0.945, 0.984)	780 (2,125,267)	1.040 (0.980, 1.104)
Non-healthcare worker	133,607 (333,129,063)	0.964 (0.959, 0.969)	127,729 (308,317,636)	0.964 (0.959, 0.970)	5,878 (24,811,426)	0.988 (0.968, 1.008)
Neighborhood level-income q1	31,146 (62,915,258)	0.975 (0.966, 0.985)	29,497 (58,025,771)	0.975 (0.965, 0.985)	1,649 (4,889,487)	0.989 (0.953, 1.027)
Neighborhood level-income q2	32,119 (66,271,483)	0.969 (0.959, 0.979)	30,592 (61,025,793)	0.966 (0.956, 0.976)	1,527 (5,245,690)	1.031 (0.990, 1.074)
Neighborhood level-income q3	28,330 (70,355,273)	0.966 (0.955, 0.977)	27,036 (64,919,555)	0.968 (0.957, 0.979)	1,294 (5,435,718)	0.962 (0.920, 1.005)
Neighborhood level-income q4	26,507 (72,704,329)	0.968 (0.956, 0.979)	25,289 (67,081,232)	0.969 (0.957, 0.982)	1,218 (5,623,097)	0.981 (0.936, 1.029)
Neighborhood level-income q5	25,221 (74,727,977)	0.949 (0.937, 0.961)	24,268 (69,026,793)	0.950 (0.937, 0.962)	953 (5,701,185)	0.984 (0.934, 1.037)

[a]The sum of the person-days of some strata is more than the total person-days because of rounding up follow-up time. For urban and rural and neighborhood income quintiles the cases and person days do not add up to the total person days because missing categories were used for both. The results of missing categories were not shown.
[b]Stratified by age group, country, and calendar month at study entry, and adjusted for: calendar month, gender, race, BMI, comorbidities (including diabetes, heart disease, lung disease, and kidney disease), current smoking status, frontline healthcare worker status, population density (quartiles), neighborhood level-income (quintiles), neighborhood level education (quintiles).
CI indicates confidence interval; HR, hazard ratio.

Figure 1. Hazard ratios (and 95% CIs) for predicted COVID-19–like illness incidence presented by a 0.1 increase in 3-month moving average neighborhood NDVI among participants in the full CCS cohort (_n_ = 2,794,029), among U.K. participants (_n_ = 2,557,464) and among U.S. participants (_n_ = 236,565) during follow-up from 24 March 2020 to 30 November 2020. In the basic model, we adjusted for calendar month, and stratified by country, age (<18, 18–65, and >65 years), and calendar month at study entry. In the fully adjusted model, we adjusted for all covariates.

PCR test were stronger than associations with predicted COVID-19–like illness (eTable 4; http://links.lww.com/EE/A218). Models without adjustment for population density also showed slightly stronger protective associations of neighborhood NDVI with predicted COVID-19–like illness incidence. Results were similar in models using summer average neighborhood NDVI instead of a three-month moving average and in models adjusted for county/local authority or $PM_{2.5}$. Substantially stronger protective associations were observed in models adjusted for season instead of month and in models without adjustment for month.

Among U.S. participants, we only observed protective associations in models not adjusted for month or adjusted for season instead of month.

Discussion

Exposure to neighborhood NDVI was associated with a lower risk of incident predicted COVID-19–like illness in the full cohort, among U.K. participants but not among U.S. participants. Stratified analyses showed slightly stronger protective associations for White individuals, individuals living in rural neighborhoods, and individuals living in high-income neighborhoods in the full cohort.

Our findings in the full cohort are largely consistent with a number of studies that have examined the association of greenspace and COVID-19 outcomes.[19,20,23–26] Many of these previous studies reported decreased incidence with increasing neighborhood exposures across a variety of greenspace measures; however, they relied on ecological data. In the United States, an ecological study observed a protective association of county-level greenness with COVID-19 incidence with an incidence rate ratio of 0.94 (95% CI = 0.90, 0.97) per 0.1 NDVI increase.[19] In China, an incidence rate ratio of 0.92 (95% CI = 0.90, 0.94) per 0.1 NDVI increase was observed with COVID-19 incidence.[27] In England, no associations of green space with COVID-19 incidence were found.[22] However, park use was associated with lower COVID-19 incidence.[22] In Wuhan, China, and Hong Kong, positive associations between green space metrics and COVID-19 cases were observed.[28,29] These results may reflect unique attributes of China and Hong Kong, where areas with a high density of commercial and entertainment venues tend to have higher green space and may be more frequently visited by large groups of people and used for social activities, which could increase COVID-19 transmission.[28]

We observed differences in associations among U.K. and U.S. participants. This could be due to differences in population characteristics, as the median age, the proportion of Black individuals, individuals with a BMI of 25+, individuals with diabetes and heart disease, and individuals living in rural areas were higher among U.S. participants compared to U.K. participants. In addition, the size of the neighborhoods in the United States was larger than in the United Kingdom, which may have led to differential exposure error. The null associations among U.S. participants are in contrast to several ecological studies in the United States.[19,24–26] Differences in associations might be related to study design and outcome assessment. Ecological studies used documented COVID-19 cases, while we used predicted COVID-19–like illness cases. Other factors that may play a role are differences in the study population, adjustment for potential confounders, and study period.

We observed slightly stronger protective associations in rural areas compared to urban areas among U.K. participants, but not among U.S. participants. We have no clear explanation for this difference. In urban areas, vegetation is likely composed of parks that can be accessible spaces for physical activity and may be indicative of other benefits available in urban areas. These qualities may not extend to rural areas. In the full cohort, we observed slightly stronger protective associations for White compared to non-White individuals and for individuals living in high-income neighborhoods compared to individuals in low-income neighborhoods. These differences might be related to the ability to work from home or differences in type of green space.

The underreporting of COVID-19 cases is a major limitation of most studies that evaluated associations with COVID-19 incidence. COVID-19 test availability was limited during the beginning of the COVID-19 pandemic, resulting in outcome misclassification that may differ between regions, degree of urbanization, SES classes, and race. To overcome these limitations, we primarily focused on predicted COVID-19–like illness incidence based on self-reported symptoms.[33] The prediction model addresses inadequacies in relying solely on reports of a positive COVID-19 test, which reflects both access to testing and variable delays and heterogeneities between symptoms and consequent testing. A previous study using U.K. CSS app participants found that population-level estimates of COVID-19 prevalence reported by the app reflect those obtained from studies designed to be representative of the population.[45] However, we also note that the prediction model was trained on the results of positive COVID-19 tests. Testing was more likely if an individual developed severe symptoms, had been in contact with people who have tested positive, traveled to high-risk areas, or if the individual was a healthcare worker. Access to testing may also differ between demographic groups which may have affected predictions. The sensitivity of the model was 0.65 and the specificity was 0.78 in the United Kingdom and slightly higher in the United States. The model likely missed COVID-19–infected individuals that did not experience any symptoms or experienced unusual COVID-19 symptoms; therefore, we used the term COVID-19–like illness. We note that the prediction performance of the model may have changed over time and as the exposure also changed over time, this could have affected our results.

We observed substantially more predicted COVID-19–like illness cases than incident COVID-19 cases identified by a PCR test. As expected, the percent frontline healthcare worker was higher among COVID-19 cases identified by a PCR test compared to predicted COVID-19–like illness cases, and the percent of participants with limitations to physical activity was higher among predicted COVID-19 like illness cases compared to COVID-19 cases identified by a PCR test. This is likely related to issues regarding testing availability and access. Further, we found that the protective association with predicted COVID-19–like illness was weaker than the association with incident COVID-19 identified by a PCR test.

The observed associations in this study may reflect recent findings that visits to greenspaces have generally increased during the pandemic. A study showed that stay-at-home orders and restrictions on social gatherings were associated with increased park visitations during COVID-19.[46] Parks provide places for physical activities and social gatherings outdoors while also allowing for distancing. Being outside for these activities rather than indoors might substantially reduce SARS-CoV-2 transmission risk, as increased air flow outdoors can substantially dilute levels of virus in the air.[47] *Furthermore, there is evidence that greenness is linked to increased physical activity; and inadequate physical activity has been reported as a risk factor for COVID-19 incidence in a study among 48,440 adult patients.*[48]

This study has several strengths. We included ~2.8 million individuals from two countries. We ended follow-up in November 2020, before some major variants were common and before vaccinations were available. While many previous studies were limited by ecological or cross-sectional designs, we evaluated individually reported data on COVID-19–like illness incidence. We were able to control for important individual-level covariates such as age, BMI, occupation as a frontline healthcare worker, and comorbidities. Using a 3-month moving average of NDVI, we constructed time-varying measures of exposure to greenness in the neighborhood reported by each participant that would capture potential changes in vegetation. Additionally, because participants were located in a diverse range of geographic settings across countries, we were able to test whether the relationship between greenness and COVID-19–like illness incidence was consistent in different SES neighborhoods, as well as in urban and rural locations.

This study has some key limitations in addition to those discussed above. We used neighborhood greenness and our analysis would have been strengthened if we had address-level

exposure estimates. The size of the area of neighborhoods differs between strata and may have affected our results. We also note that greenness can differ substantially within a neighborhood and participants living close to the border of their neighborhood can be exposed to natural environments in neighboring neighborhoods. In addition, we had no information on the amount of time each participant spent at home or accessing green spaces. NDVI does not measure the quality of greenness, such as whether vegetation is composed of parks, forests, agricultural land, or overgrown vacant lots. We adjusted for multiple individual-level lifestyles and health status covariates, but we did not control for individual-level income, education, and protective measures, such as social distancing and mask-wearing. Adjustment for area-level SES may have partly captured the impact of these factors. However, we note that the potential for residual confounding remains and may have resulted in an overestimation of the associations. We had no information on whether individuals lived in the same household. Hence, cases may not be independent as COVID-19–like illness cases may have clustered. We used self-reported data about potential confounders and COVID-19 symptoms which may have affected our findings. Some individuals had missing covariate data and we used a missing indicator in our analyses. Additionally, our findings may not be generalizable to the whole U.S. or U.K. population or other countries because the participants in this study who self-selected into the study may represent a limited range of age, SES, BMI, mobility, and attitudes towards COVID-19 and health compared to the wider population. Individuals excluded from our cohort because of <24 hours of follow-up time were of lower SES, had a higher population density, and were more likely to be a healthcare worker, which likely affected the generalizability of our findings as stratified analyses showed weaker associations for these groups.

Conclusions

In this CCS smartphone application cohort, we observed that those living in neighborhoods with higher greenness had a lower risk of predicted like illness COVID-19 incidence. These results underscore the potentially critical role of environments in shaping human health during the COVID-19 pandemic.

Conflict of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

References

1   Sohrabi C, Alsafi Z, O'Neill N, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg. 2020;76:71–76.
2   Johns Hopkins Coronavirus Resource Center. Available at: https://coronavirus.jhu.edu/us-map. Accessed Jan 17, 2022.
3   Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet. 2020;395:1973–1987.
4   van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. N Engl J Med. 2020;382:1564–1567.
5   VoPham T, Weaver MD, Hart JE, Ton M, White E, Newcomb PA. Effect of social distancing on COVID-19 incidence and mortality in the US. medRxiv 2020.
6   Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the united states reduced the covid-19 growth rate. Health Aff. 2020;39:1237–1246.
7   Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, Rt. PLoS Comput Biol. 2020;16:e1008409.
8   Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R0). Emerg Infect Dis. 2019;25:1–4.
9   Gao X, Wei J, Lei H, Xu P, Cowling BJ, Li Y. Building Ventilation as an Effective Disease Intervention Strategy in a Dense Indoor Contact Network in an Ideal City. PLoS One. 2016;11:e0162481.
10 Bulfone TC, Malekinejad M, Rutherford GW, Razani N. Outdoor Transmission of SARS-CoV-2 and Other Respiratory Viruses: A Systematic Review. J Infect Dis. 2021;223:550–561.
11 James P, Banay RF, Hart JE, Laden F. A Review of the Health Benefits of Greenness. Curr Epidemiol Reports. 2015;2:131–142.
12 Fong KC, Hart JE, James P. A Review of Epidemiologic Studies on Greenness and Health: Updated Literature Through 2017. Curr. Environ. Heal. 2018;5:77–87.
13 Nieuwenhuijsen MJ, Khreis H, Triguero-Mas M, Gascon M, Dadvand P. Fifty Shades of Green. Epidemiology. 2017;28:63–71.
14 Twohig-Bennett C, Jones A. The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. Environ Res. 2018;166:628–637.
15 Sarkar C, Zhang B, Ni M, et al. Environmental correlates of chronic obstructive pulmonary disease in 96 779 participants from the UK Biobank: a cross-sectional, observational study. Lancet Planet Heal. 2019;3:e478–e490.
16 Hamer M, Kivimäki M, Gale CR, Batty GD. Lifestyle risk factors for cardiovascular disease in relation to COVID-19 hospitalization: a community-based cohort study of 387,109 adults in UK. medRxiv 2020.
17 Bansal M. Cardiovascular disease and COVID-19. Diabetes Metab Syndr Clin Res Rev. 2020;14:247–250.
18 Beltramo G, Cottenet J, Mariet AS, et al. Chronic respiratory diseases are predictors of severe outcome in COVID-19 hospitalised patients: a nationwide study. Eur Respir J. 2021;58:2004474.
19 Klompmaker JO, Hart JE, Holland I, et al. County-level exposures to greenness and associations with COVID-19 incidence and mortality in the United States. Environ Res. 2021;199:111331.
20 Stieb DM, Evans GJ, To TM, Brook JR, Burnett RT. An ecological analysis of long-term exposure to PM2.5 and incidence of COVID-19 in Canadian health regions. Environ Res. 2020;191:110052.
21 Cascetta E, Henke I, Di Francesco L. The Effects of Air Pollution, Sea Exposure and Altitude on COVID-19 Hospitalization Rates in Italy. Int J Environ Res Public Health. 2021;18:452.
22 Johnson TF, Hordley LA, Greenwell MP, Evans LC. Associations between COVID-19 transmission rates, park use, and landscape structure. Sci Total Environ. 2021;789:148123.
23 Lee KS, Min HS, Jeon JH, Choi YJ, Bang JH, Sung HK. The association between greenness exposure and COVID-19 incidence in South Korea: An ecological study. Sci Total Environ. 2022;832:154981.
24 Russette H, Graham J, Holden Z, Semmens EO, Williams E, Landguth EL. Greenspace exposure and COVID-19 mortality in the United States: January–July 2020. Environ Res. 2021;198:111195.
25 Spotswood EN, Benjamin M, Stoneburner L, et al. Nature inequity and higher COVID-19 case rates in less-green neighbourhoods in the United States. Nat Sustain. 2021;4:1092–1098.
26 Lee W, Kim H, Choi HM, et al. Urban environments and COVID-19 in three Eastern states of the United States. Sci Total Environ. 2021;779:146334.
27 Peng W, Dong Y, Tian M, et al. City-level greenness exposure is associated with COVID-19 incidence in China. Environ Res. 2022;209:112871.
28 Kan Z, Kwan MP, Wong MS, Huang J, Liu D. Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong. Sci Total Environ. 2021;772:145379.
29 You H, Wu X, Guo X. Distribution of COVID-19 Morbidity Rate in Association with Social and Economic Factors in Wuhan, China: Implications for Urban Development. Int J Environ Res Public Heal. 2020;17:3417.
30 Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med. 2020;382:2081–2090.
31 Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368:489–493.
32 Villeneuve PJ, Goldberg MS. Methodological Considerations for Epidemiological Studies of Air Pollution and the SARS and COVID-19 Coronavirus Outbreaks. Environ Health Perspect. 2020;128:095001.
33 Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med. 2020;26:1037–1040.

34 Chan AT, Drew DA, Nguyen LH, et al. The coronavirus pandemic epidemiology (COPE) consortium: A call to action. Cancer Epidemiol Biomarkers Prev. 2020;29:1283–1289.

35 Drew DA, Nguyen LH, Steves CJ, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science. 2020;368:1362–1367.

36 Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 — Studies Needed. N Engl J Med. 2020;382:1194–1196.

37 Landsat 8 Data Users Handbook. Available at: https://www.usgs.gov/media/files/landsat-8-data-users-handbook. Accessed June 16, 2020.

38 Piel FB, Fecht D, Hodgson S, et al. Small-area methods for investigation of environment and health. Int J Epidemiol. 2020;49:686–699.

39 American Community Survey Data. Available at: https://www.census.gov/programs-surveys/acs/data.html. Accessed July 11, 2022.

40 English indices of deprivation 2019: mapping resources. GOV.UK. Available at: https://www.gov.uk/guidance/english-indices-of-deprivation-2019-mapping-resources. Accessed July 11, 2022.

41 Welsh Index of Multiple Deprivation (full Index update with ranks). 2019. GOV.WALES. Available at: https://gov.wales/welsh-index-multiple-deprivation-full-index-update-ranks-2019. Accessed July 11, 2022.

42 Scottish Index of Multiple Deprivation. 2020. gov.scot. Available at: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/. Accessed July 11, 2022.

43 Di Q, Amini H, Shi L, et al. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int. 2019;130:104909.

44 Wang W, Fecht D, Beevers S, Gulliver J. Predicting daily concentrations of nitrogen dioxide, particulate matter and ozone at fine spatial scale in Great Britain. Atmos Pollut Res. 2022;13:101506.

45 Varsavsky T, Graham MS, Canas LS, et al. Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application: a prospective, observational study. Lancet Public Heal. 2021;6:e21–e29.

46 Geng DC, Innes J, Wu W, Wang G. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. J For Res. 2021;32:553–567.

47 Qian H, Miao T, Liu L, Zheng X, Luo D, Li Y. Indoor transmission of SARS-Cov-2. medrxiv.org 2020;15:e0241955.

48 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;55:1099–1105.