

Role of meteorological factors in the transmission of SARS-CoV-2 in the United States

Yiqun Ma ^{1,2}, Sen Pei ^{3✉}, Jeffrey Shaman ³, Robert Dubrow ^{1,2} & Kai Chen ^{1,2✉}

Improved understanding of the effects of meteorological conditions on the transmission of SARS-CoV-2, the causative agent for COVID-19 disease, is needed. Here, we estimate the relationship between air temperature, specific humidity, and ultraviolet radiation and SARS-CoV-2 transmission in 2669 U.S. counties with abundant reported cases from March 15 to December 31, 2020. Specifically, we quantify the associations of daily mean temperature, specific humidity, and ultraviolet radiation with daily estimates of the SARS-CoV-2 reproduction number (R_t) and calculate the fraction of R_t attributable to these meteorological conditions. Lower air temperature (within the 20–40 °C range), lower specific humidity, and lower ultraviolet radiation were significantly associated with increased R_t . The fraction of R_t attributable to temperature, specific humidity, and ultraviolet radiation were 3.73% (95% empirical confidence interval [eCI]: 3.66–3.76%), 9.35% (95% eCI: 9.27–9.39%), and 4.44% (95% eCI: 4.38–4.47%), respectively. In total, 17.5% of R_t was attributable to meteorological factors. The fractions attributable to meteorological factors generally were higher in northern counties than in southern counties. Our findings indicate that cold and dry weather and low levels of ultraviolet radiation are moderately associated with increased SARS-CoV-2 transmissibility, with humidity playing the largest role.

¹Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA. ²Yale Center on Climate Change and Health, Yale School of Public Health, New Haven, CT, USA. ³Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA. ✉email: sp3449@cumc.columbia.edu; kai.chen@yale.edu

Since first detected, the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has produced a major global pandemic. As of March 20, 2021, ~29.8 million COVID-19 cases and 542 thousand deaths had been reported in the U.S.¹, more than any other country. The decreased stability of SARS-CoV-2 in warmer temperatures, higher humidity, and simulated sunlight in laboratory experiments^{2–5}, and the documented seasonality of influenza^{6,7} and infections caused by other coronaviruses^{8–10}, lead to the hypothesis that lower air temperature, lower humidity, and lower ultraviolet (UV) radiation are associated with increased SARS-CoV-2 transmission. Quantifying this effect on a population level is needed to help inform public health control efforts, including transmission prevention and communication with the public¹¹.

Numerous preliminary studies have found either positive or negative associations of air temperature, humidity, and UV radiation with reported COVID-19 case numbers^{12–17}. However, given the large number of undocumented SARS-CoV-2 infections¹⁸, the variations in the lag between infection and symptom onset, and the inconsistent lag between testing and reporting, using daily new confirmed cases may not be optimal for examining meteorological effects¹⁹. As a result, a few studies have used the reproduction number to estimate SARS-CoV-2 transmissibility^{20–22}. One study reported high daily air temperature and high daily relative humidity (RH, the amount of water vapor in the air expressed as a percentage of the amount needed for saturation at a given temperature) to be associated with a reduced daily effective reproduction number (R_e , the mean number of new infections caused by a single infected person in a population in which some individuals may no longer be susceptible due to acquired immunity²³) for SARS-CoV-2 in both China and the U.S.²⁰. However, early studies focused on the first few months of the pandemic found no association between temperature, humidity, or UV radiation and the basic reproduction number (the mean number of new infections caused by a single infected person in a population in which everyone is assumed to be susceptible and no public health measures have been implemented)^{21,22}.

Early analyses, in particular, should be interpreted with caution¹¹, as the range of temperature, humidity, and UV radiation measurements during the short observation period at the beginning of the pandemic was relatively narrow in most studies^{12–15,20–22}, thus limiting the ability to detect associations between these meteorological variables and SARS-CoV-2 transmission. In addition, many previous studies (whether using COVID-19 cases or reproduction number as the outcome) controlled for no or only a few potential confounders^{12–16,21,22}, which include other environmental factors, socioeconomic factors, temporal changes in population immunity, and implementation of public health interventions.

Furthermore, although most early studies found an association between air temperature, humidity, or UV radiation and COVID-19 incidence, the fraction of cases or deaths attributable to meteorological conditions remains unclear. One modeling study predicted that as long as most of the population is susceptible to infection, any role of humidity in SARS-CoV-2 transmission would be overwhelmed by the lack of population immunity²⁴. This prediction is supported by the rapid transmission of SARS-CoV-2 regardless of climate zone, including warmer locations such as tropical Brazil, India, and southern states in the U.S. during the northern hemisphere summer¹. In addition, the relative importance of different meteorological factors needs further investigation²⁵.

Here we investigate the association between meteorological conditions, i.e., air temperature, specific humidity (SH; the mass

of water vapor in a unit mass of moist air [g kg^{-1}]), and UV radiation and SARS-CoV-2 transmission, as measured by the reproduction number R_t (the mean number of new infections caused by a single infected person, given the public health measures in place, in a population in which everyone is assumed to be susceptible). In this study, instead of R_e , we use R_t to quantify the transmission rate of SARS-CoV-2, which removes the impact of population immunity on disease transmission. We estimate R_t in the 2669 counties with at least 400 cumulative cases as of December 31, 2020 and calculate the fraction of R_t attributable to temperature, SH, or UV radiation, adjusting for a wide range of potential confounders.

Results

Distribution of meteorological factors and R_t . From March 15 to December 31, 2020, a total of 19,430,010 cases of COVID-19 were reported in the 2669 study counties (Supplementary Table 1). We estimated the county-specific R_t using a dynamic metapopulation model informed by human mobility data that represents the transmission of SARS-CoV-2 in the U.S. (see “Methods”). Mean daily R_t averaged over all counties and days during the study period was 1.49 and ranged from 0.45 to 6.62. Daily air temperature, SH, and UV radiation also ranged widely (air temperature: -22.25 – 39.98 °C; SH: 0.49 – 22.37 g kg^{-1} ; UV radiation: 1.68 – 155.61 kJ m^{-2}). Cass County, Indiana had the highest R_t averaged over the study period, and Brewster County, Texas had the lowest (Fig. 1a). Southern counties generally were hotter than northern counties (Fig. 1b). In the eastern U.S., southern counties were more humid than northern counties; western U.S. counties were drier than eastern U.S. counties, with inland western counties generally drier than coastal western counties (Fig. 1c). In the eastern U.S., UV radiation levels were higher in southern counties than in northern counties; western U.S. counties generally were exposed to higher levels of UV radiation than eastern U.S. counties (Fig. 1d).

Associations between meteorological factors and R_t . We estimated the complex non-linear and temporally delayed associations of meteorological factors with SARS-CoV-2 R_t using a generalized additive mixed model adjusting for spatiotemporal variations in R_t and potential measured confounders, described in detail in “Methods”. We then calculated the optimum values of temperature, SH, and UV radiation, which correspond to the lowest R_t , between the 1st and the 99th percentiles of the distribution of each meteorological variable. For temperature in the range of 20 – 40 °C, we found an approximately linear inverse temperature– R_t relationship (Fig. 2a), with lower air temperatures significantly associated with increased transmission of SARS-CoV-2. No significant associations were observed when temperature was below ~ 10 °C. Compared with the optimum temperature (31.23 °C), a temperature of 20 °C was associated with a 5.15% (95% CI: 2.49–7.88%) increase of R_t .

The relationship between SH and R_t was non-linear (Fig. 2b). Higher SH was significantly associated with decreased transmission, except for a stable trend from ~ 7 to 12 g kg^{-1} . Compared with the optimum value (19.21 g kg^{-1}), the 1st percentile of the distribution of SH (1.80 g kg^{-1}) was associated with a 15.20% (95% CI: 9.65–21.04%) increase of R_t . UV radiation level was unrelated to SARS-CoV-2 transmission when UV radiation was lower than ~ 100 kJ m^{-2} , but when above this level, an almost linear negative association was observed between UV radiation and R_t (Fig. 2c). A UV radiation level of 100 kJ m^{-2} was associated with a 5.18% (95% CI: 2.26–8.17%) increase of R_t over the optimal level (142.78 kJ m^{-2}).

Trends of effect estimates in the lag dimension are shown in Supplementary Fig. 1. Sensitivity analyses showed the estimated relationships between meteorological factors and R_t to be generally consistent under different modeling choices (Fig. 2a–c), except for the temperature curve when the number of degrees of freedom (df) of exposure (meteorological factors) was changed to 4, which could be a result of overfitting (Fig. 2a). The coefficient table of other covariates in the main model is shown in Supplementary Table 2. The R^2 of the main model is 0.514, and the spatial and temporal autocorrelations are insignificant (P values = 0.159 and 0.798 for spatial and temporal autocorrelations, respectively) (Supplementary Table 3).

Fractions of R_t attributable to meteorological factors. Based on the estimated associations of meteorological factors with R_t and daily county-specific R_t , we further calculated the fraction of R_t attributable to meteorological factors (i.e., the attributable fraction [AF], which can be interpreted as the fraction of R_t attributable to the deviation of temperature, SH, or UV radiation from the optimum value). Across all 2669 counties over the entire study period, the AF for temperature was 3.73% (95% empirical confidence intervals [eCI]: 3.66–3.76%), the AF for SH was 9.35% (95% eCI: 9.27–9.39%), and the AF for UV radiation was 4.44% (95% eCI: 4.38–4.47%) (Supplementary Table 4). In total, the three meteorological factors contributed to ~17.5% of R_t . Compared with the main model, models including each meteorological factor separately resulted in higher AF for temperature and SH, and lower AF for UV radiation (Supplementary Fig. 2).

The AF for temperature generally was higher in the eastern U.S. and the West Coast than in other regions (Fig. 3a). The AF for SH showed an increasing trend from south to north in the eastern U.S., whereas in the western U.S., the AF for SH was lower in counties in coastal states than in counties in interior states (Fig. 3b). The AF for UV radiation was generally higher in the eastern U.S. than in the western U.S. (Fig. 3c) and was lowest in the southwest. The total AF for all three meteorological factors combined generally was higher in northern counties than in southern counties in the eastern U.S. and was also high in most of the western U.S. (Fig. 3d). Each meteorological factor exhibited the highest AF in winter and the lowest AF in summer (Fig. 4).

Sensitivity analyses indicate that the AF for air temperature, SH, or UV radiation generally remains robust when excluding socioeconomic factors and when additionally adjusting for smoking and obesity prevalence, long-term air pollution, climate zones, or short-term air pollution (Supplementary Table 4).

Discussion

Using estimated reproduction numbers for 2669 U.S. counties and controlling for temporal and spatial trends and other potential confounders, we assessed the associations of air temperature, SH, and UV radiation with the transmission of SARS-CoV-2 and estimated the fractions of R_t attributable to meteorological factors. We found lower air temperature (within the range of 20–40 °C), lower SH, and lower UV radiation to be significantly associated with increased R_t . During the study period, meteorological factors contributed to ~17.5% of R_t : 3.73%,

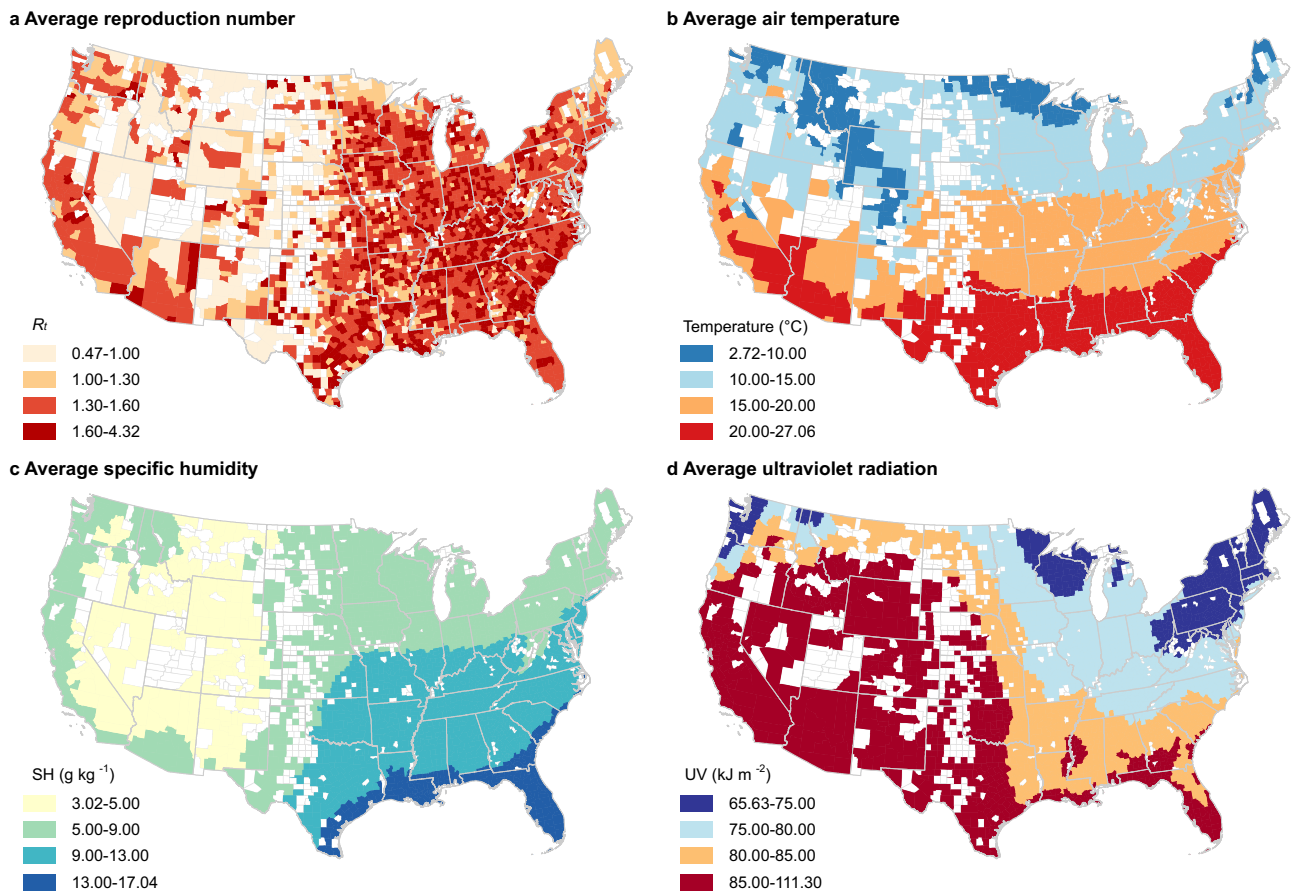


Fig. 1 Map of the distribution of reproduction number, air temperature, specific humidity, and ultraviolet radiation in study counties. This figure displays the distribution of key variables averaged over the study period in 2669 U.S. counties. **a** The distribution of the daily reproduction number (R_t). **b** The distribution of daily air temperature. **c** The distribution of daily specific humidity (SH). **d** The distribution of daily ultraviolet (UV) radiation. The shapefile in the maps was obtained from the U.S. Census Bureau.

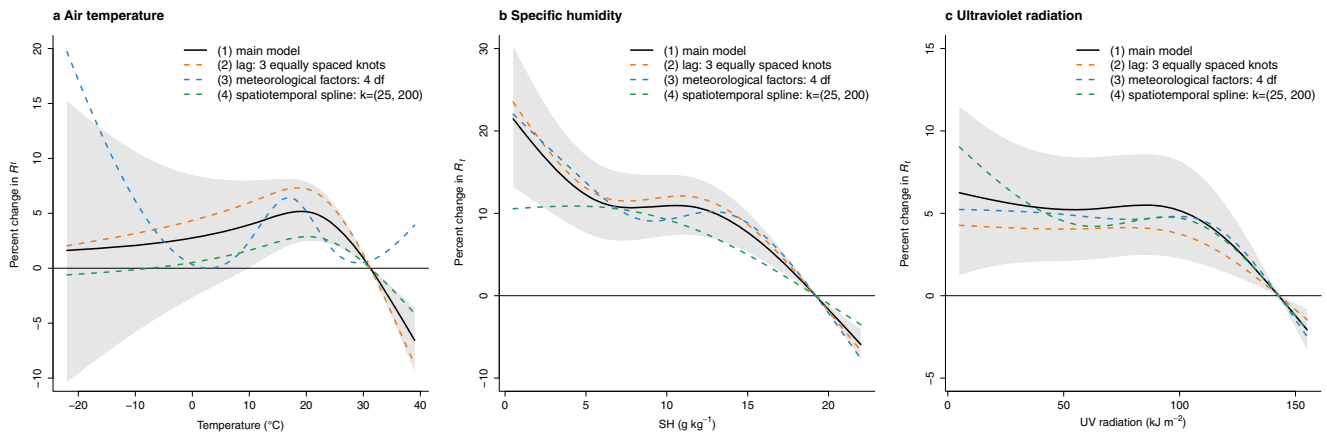


Fig. 2 The associations of air temperature, specific humidity, and ultraviolet radiation with R_t , under different choices of model. This figure shows the estimated exposure-response curves (meteorological factors vs. percent change in R_t) for the associations of air temperature (a), specific humidity (SH) (b), and ultraviolet (UV) radiation (c) with reproduction number (R_t) for SARS-CoV-2, with different modelling choices: (1) main model with 95% confidence interval (grey area): tensor product smooths to control for the temporal and spatial variations with a maximum of 30 and 200 knots (k), respectively, and cross-basis terms for meteorological factors, which are defined by natural cubic splines with 3 df for both the exposure-response and lag-response association, with a maximum lag of 13 days; (2) redefine the lag dimension using a natural cubic spline and 3 equally placed internal knots in the log scale; (3) change the df to 4 in the cross-basis terms for meteorological factors in the exposure-response function; (4) change the maximum number of knots to 25 in the flexible natural cubic spline to control time trend in the tensor product smooths.

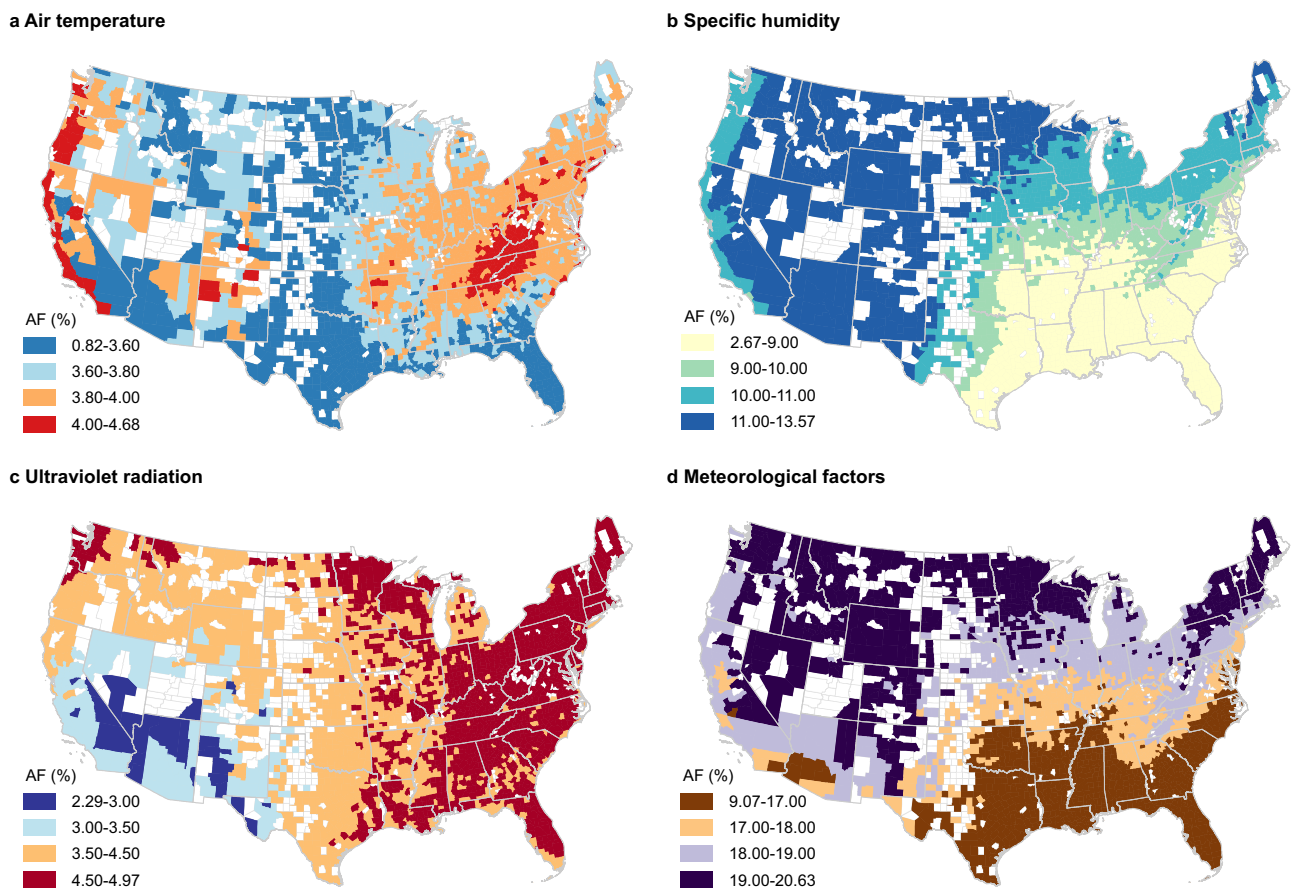


Fig. 3 Fractions of R_t attributable to meteorological factors by county. The distribution of the fraction of reproduction number (R_t) attributable to temperature (a), specific humidity (b), ultraviolet radiation (c), or the sum of the three meteorological factors (d) (i.e., attributable fraction [AF]) in each county. The shapefile in the maps was obtained from the U.S. Census Bureau.

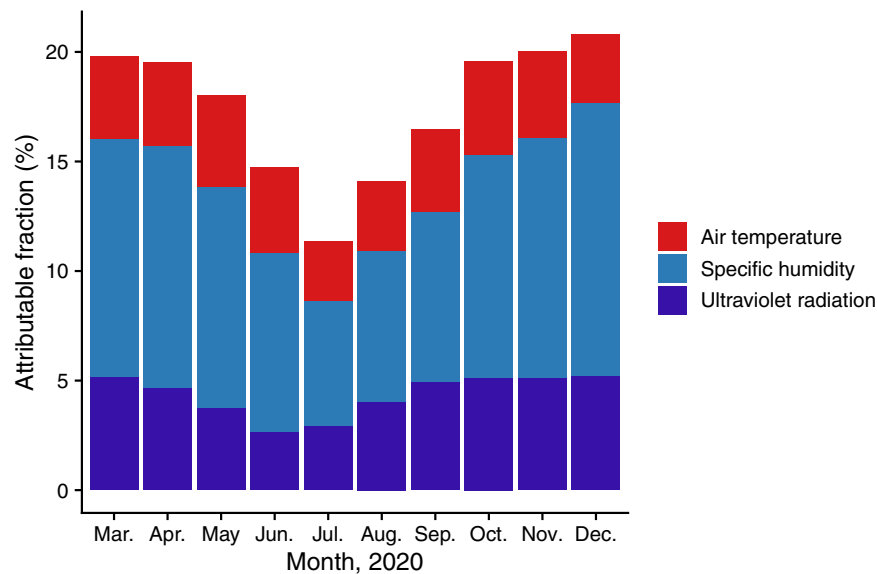


Fig. 4 Fractions of R_t attributable to meteorological factors by month. The distribution of the fraction of reproduction number (R_t) attributable to temperature, specific humidity, ultraviolet radiation, or the sum of the three meteorological factors (i.e., attributable fraction [AF]) by month in 2020.

9.35%, and 4.44% of R_t was attributable to the deviation of temperature, SH, and UV radiation from their optimum values, respectively. Meteorological factors in total contributed more to the transmission of SARS-CoV-2 in counties and months with colder and drier weather and lower levels of UV radiation than in counties and months with warmer, more humid weather and higher levels of UV radiation. In December (the month with the lowest temperature, lowest SH, and lowest UV radiation of our study period of March–December), the AF for meteorological factors was 20.8% (Fig. 4). Everything else being equal, we can anticipate the highest AF during the months with colder and drier weather and lower UV radiation in future years.

Associations of lower temperature, lower humidity, and lower amount of UV radiation with increased COVID-19 outcomes have been reported by many previous studies. Many multicity analyses in China reported such negative associations^{12,14,15}. For example, using data of daily confirmed case counts from 30 provincial capital cities of China, Liu et al. found that lower temperature and lower absolute humidity were associated with higher COVID-19 case counts¹⁴. Later, with the rapid spread of COVID-19 around the world, studies in other countries emerged^{17,26–29}, and UV radiation was considered in some studies. In the early stages of this pandemic in the U.S., a state-level study of daily COVID-19 case counts observed a declining trend of reported cases with higher UV radiation and increasing temperature up to 52 °F²⁶. Based on data from 166 countries worldwide, another study reported that a 1 °C increase in temperature and a 1% increase in RH were associated with a 3.08% and 0.85% reduction in daily new cases, respectively²⁸. Another multi-country study provided evidence for a protective role of ultraviolet-B (UVB) radiation in reducing COVID-19 deaths²⁹. However, many of these earlier studies were limited by short study periods (e.g. 1–2 months), use of daily confirmed cases or deaths across countries for which there were varying reporting biases, failure to account for the time lag between observed weather conditions and when cases or deaths were recorded, or failure to account for time delays between infection acquisition and case confirmation^{19,25}.

By representing the transmissibility of SARS-CoV-2, the estimated daily reproduction number serves as a better outcome than daily case counts. While case counts are subject to the influence of

reporting delay and underreporting, which vary across locations and are thus difficult to control, the reproduction number is a direct estimate of the transmission rate of SARS-CoV-2, quantifying the average number of infections caused by one infection in the population. A small number of studies previously analyzed the association between temperature, humidity, or UV radiation and reproduction number^{20–22,30}. Wang et al. found that a 1 °C increase in temperature was associated with a reduction in the effective reproduction number of 0.026 in China and 0.020 in the U.S., and a 1% increase in RH was associated with a reduction in the effective reproduction number of 0.0076 in China and 0.0080 in the U.S.²⁰. Adnan et al. reported a significant negative association between UV index (a standard measurement of the strength of sunburn-producing UV radiation) and basic reproduction number in major cities of Pakistan³⁰. These associations are consistent with our findings but were not supported by two studies in China that examined the basic reproduction number: the first found no association between temperature or UV radiation and SARS-CoV-2 transmission²²; the second found no association between absolute humidity and SARS-CoV-2 transmission²¹. However, these early studies were limited by short observation periods at the beginning of the pandemic, and they did not account for variations of testing capacity, reporting, human mobility, and population susceptibility in estimating SARS-CoV-2 transmissibility.

We estimated R_t using a dynamic metapopulation model informed by human mobility data. This mechanistic model accounted for unreported infections, reporting delays, and county-to-county movement. Previous estimates of R_e using reported cases did not consider underreporting of infections. Our approach mitigates this limitation by additionally modeling the transmission of unreported infections and estimating the ascertainment rate—the fraction of all infections that are confirmed cases, and the relative contagiousness of those unreported infections. Further, reported incidence is a lagged indicator of disease transmission due to the delay from infection acquisition to laboratory confirmation. We corrected for this lag using a reporting delay model informed by line-list data from the U.S. Lastly, R_e used in previous studies is determined by both the local transmission rate and population susceptibility: $R_e = R_t \times s$, where s is the fraction of the total population susceptible to infection. Analyses using R_e are complicated by the variation of

population susceptibility across U.S. counties. To address this issue, we explicitly estimated the population susceptibility in each county, and removed its influence in the calculation of R_t ³¹ (see “Methods”). The model estimating population susceptibility has been validated against independent seroprevalence study data³¹. Thus, our estimates account for spatial heterogeneity in population immunity.

Another strength of our study was adjustment for a wide range of demographic and socioeconomic factors in the main analysis, as well as for smoking and obesity, air pollution, and climate zone in sensitivity analyses. We also thoroughly controlled for spatially and temporally heterogeneous unmeasured confounders, such as implementation of and compliance with public health measures³¹, by simultaneously controlling for temporal and spatial variations (Supplementary Fig. 3) and including smooths of random effects to further account for unmeasured state- and county-level confounding (see “Methods”). This approach accounted for substantial differences in the epidemic curves among states and counties (Supplementary Fig. 4).

Our findings for air temperature and SH are supported by laboratory evidence on the stability of SARS-CoV-2 as a function of temperature and humidity. It has been reported that the virus’ half-life in human nasal mucus and sputum is shorter under conditions of higher temperature and RH than under conditions of lower temperature and RH². Similar findings were reported by other studies testing virus stability in virus transport medium³, in aerosols³², and on various surfaces³². Further, the SARS-CoV-2 half-life was found to be longer at lower temperatures, and at both 22 °C and 27 °C, the half-life decreased as RH increased from 40 to 65% but increased as RH increased from 65 to 85%³³. This result is roughly consistent with the non-linear relationship between SH and R_t observed in our study (Fig. 2b), in which there was a stable trend of R_t from 7 to 12 g kg⁻¹ of SH superimposed on the overall decreasing trend. In addition to being mediated by effects on the virus itself, the associations between temperature and humidity and SARS-CoV-2 transmissibility may be mediated by human airway antiviral defenses. Inhalation of cold and dry air can impair mucociliary clearance, a crucial mechanism for the elimination of inhaled pathogens³⁴. Further, during the colder winter months people spend more time indoors, which may facilitate virus transmission³⁵. During these months, whether indoors or outdoors, people are exposed to less UV radiation from the sun, which modulates the immune system^{36,37}.

UV radiation may affect transmission of SARS-CoV-2 through impacts on the virus and on immune function²⁵. It has been shown that higher levels of UV radiation, particularly ultraviolet-C radiation (UV light with wavelengths between 200 and 280 nm), can inactivate RNA viruses²⁵. In experimental studies, exposure to simulated sunlight resulted in rapid inactivation of infectious SARS-CoV-2 on different surfaces⁴ and in aerosols⁵. Furthermore, UV radiation can indirectly influence SARS-CoV-2 transmission through its impact on the synthesis of vitamin D and other UV-induced mediators of immune function³⁶.

We estimated that a total of ~17.5% of R_t was attributable to the three meteorological factors combined. This estimate is consistent with a previous modelling study, which found that weather (temperature, RH, and UV radiation) explained 17% of the variation in COVID-19 growth rate (i.e., the exponential increase in cases)³⁸. We found that SH contributes more to SARS-CoV-2 transmission than temperature, which is consistent with studies of influenza^{39,40}. SH is more strongly associated with the observed seasonality of influenza in temperate regions than either temperature or RH^{6,39}. In developed countries, such as the U.S., people spend ~90% of their time indoors⁴¹, especially during winter³⁵. Although indoor temperature is usually controlled, indoor humidity generally is not, and closely mirrors outdoor

levels^{42–44}, perhaps explaining why ambient outdoor SH is more strongly associated with SARS-CoV-2 transmission than ambient outdoor temperature. In addition, the large discrepancy between indoor and outdoor temperature and the high correlation between indoor and outdoor humidity explain why ambient outdoor temperature showed no association with SARS-CoV-2 transmission when lower than 20 °C in the main model (Fig. 2a), but showed a monotonically decreasing trend in the model excluding the other two meteorological factors (Supplementary Fig. 2). However, it remains unclear whether SH (versus temperature) is the causative modulator of SARS-CoV-2 transmission or is simply a useful indicator of the indoor environment and the combined effects of temperature and RH.

In the sensitivity analyses, after adjusting for long-term PM_{2.5}, the estimated AF for temperature increased by about 40% (Supplementary Table 4), indicating that long-term PM_{2.5} acted as a confounder for temperature effects. This result is consistent with a recent study that found increased COVID-19 mortality associated with increased long-term exposure to PM_{2.5}⁴⁵. In contrast, the fraction of R_t attributable to SH or UV radiation remained stable after adjusting for long-term PM_{2.5}. Although it is unclear why long-term PM_{2.5} would serve as a confounder for temperature, but not for SH or UV radiation, this result does suggest that SH and UV radiation are more robust predictors than temperature.

Several limitations of this study should be noted. First, this is an ecological rather than an individual-level study, thus making the study susceptible to the ecological fallacy. Second, due to data limitations, we were unable to explore potential heterogeneity of associations of meteorological factors with R_t for different variants of SARS-CoV-2. Future studies are needed to investigate this potential heterogeneity, as knowledge of differing meteorological impacts across variants may inform prevention strategies.

Our findings indicate that cold and dry weather and low levels of UV radiation are moderately associated with increased SARS-CoV-2 transmissibility in the U.S., with absolute humidity (i.e., SH) playing the greatest role. More extensive public health interventions are needed to mitigate the increased transmissibility of SARS-CoV-2 in winter months.

Methods

Data collection. We extracted hourly air temperature and SH from the North America Land Data Assimilation System project⁴⁶, a near real-time dataset with a 0.125° × 0.125° grid resolution. We spatially and temporally averaged these data into daily county-level records. SH is the mass of water vapor in a unit mass of moist air (g kg⁻¹). Daily downward UV radiation at the surface, with a wavelength of 0.20–0.44 μm, was extracted from the European Centre for Medium-Range Weather Forecasts ERA5 climate reanalysis⁴⁷.

Other characteristics of each county, including geographic location, population density, demographic structure of the population, socioeconomic factors, proportion of healthcare workers, intensive care unit (ICU) bed capacity, health risk factors, long-term and short-term air pollution, and climate zone were collected from multiple sources. Geographic coordinates, population density, median household income, percent of people older than 60 years, percent Black residents, percent Hispanic residents, percent owner-occupied housing, percent residents aged 25 years and over without a high school diploma, and percent healthcare practitioners or support staff were collected from the U.S. Census Bureau⁴⁸. Total ICU beds in each county were derived from Kaiser Health News⁴⁹. The prevalence of smoking and obesity among adults in each county was obtained from the Robert Wood Johnson Foundation’s 2020 County Health Rankings⁵⁰. We extracted annual PM_{2.5} concentrations in the U.S. from 2014 to 2018 from the 0.01° × 0.01° grid resolution PM_{2.5} estimation provided by the Atmospheric Composition Analysis Group⁵¹, and calculated average PM_{2.5} levels during this 5-year period for each county to represent long-term PM_{2.5} exposure (Supplementary Fig. 5). Short-term air quality data during the study period, including daily mean PM_{2.5} and daily maximum 8-h O₃, were obtained from the United States Environmental Protection Agency⁵². We categorized study counties into one of five climate zones based on the guide released by U.S. Department of Energy⁵³ (Supplementary Fig. 6).

The county-level COVID-19 case and death data were downloaded from the John Hopkins University Coronavirus Resource Center¹. The U.S. county-to-

county commuting data were available from the U.S. Census Bureau⁴⁸. Daily numbers of inter-county visitors to points of interest (POI) were provided by SafeGraph⁵⁴.

Data ethics. SafeGraph utilizes data from mobile applications of which users optionally consent to provide their anonymous location data.

Estimation of reproduction number. We estimated the daily reproduction number (R_t) in all 3142 U.S. counties using a dynamic metapopulation model informed by human mobility data^{31,55}. R_t is the mean number of new infections caused by a single infected person, given the public health measures in place, in a population in which everyone is assumed to be susceptible. In the metapopulation model, two types of movement were considered: daily work commuting and random movement. During the daytime, some commuters travel to a county other than their county of residence, where they work and mix with the populations of that county; after work, they return home and mix with individuals in their home, residential county. Apart from regular commuting, a fraction of the population in each county, assumed to be proportional to the number of inter-county commuters, travels for purposes other than work. As the population present in each county is different during daytime and night-time, we modelled the transmission dynamics of COVID-19 separately for these two time periods, each depicted by a set of ordinary differential equations (Supplementary Notes).

To account for case underreporting, we explicitly simulated reported and unreported infections, for which separate transmission rates were defined. Recent studies from several countries indicate that asymptomatic cases of COVID-19, which are typically unreported, are less contagious than symptomatic cases^{56–59}. Studies on the early transmission of SARS-CoV-2 in China⁴⁸ and the U.S.⁶⁰ also showed that undocumented infections are less transmissible than documented infections.

In order to reflect the spatiotemporal variation of disease transmission rate and reporting, we allowed transmission rates and ascertainment rates to vary across counties and to change over time. The transmission model simulated daily confirmed cases and deaths for each county. To map infections to deaths, we used an age-stratified infection fatality rate (IFR)⁶¹ and computed the weekly IFR for each county as a weighted average using state-level age structure of confirmed cases reported by the U.S. Centers for Disease Control and Prevention. We further adjusted for reporting lags using an observational delay model informed by a U.S. line-list COVID-19 data record⁶².

For the period prior to March 15, 2020, we used commuting data from the U.S. census survey to prescribe the inter-county movement in the transmission model⁴⁸. Starting March 15, the census survey data are no longer representative due to changes in mobility behavior following the implementation of non-pharmaceutical interventions. We, therefore, used estimates of the reduction of inter-county visitors to POI (e.g., restaurants, stores, etc.) from SafeGraph⁵⁴ to account for the change in inter-county movement on a county-by-county basis. Because there is no direct relationship between population-level mobility patterns and COVID-19 transmission rates⁶³, we did not model local transmission rate as a function of inter-county mobility. Instead, the SafeGraph data were only used to inform the change of population mixing across counties.

To infer key epidemiological parameters, we fitted the transmission model to county-level daily cases and deaths reported from March 15, 2020 to December 31, 2020. The estimated reproduction number was computed as follows:

$$R_t = \beta D[\alpha + (1 - \alpha)\mu], \tag{1}$$

where β is the county-specific transmission rate, μ is the relative transmissibility of unreported infections, α is the county-specific ascertainment rate, and D is the average duration of infectiousness. Note β and α were defined for each county separately and were allowed to vary over time. Unlike previous studies using effective reproduction number

$$R_e = \beta D[\alpha + (1 - \alpha)\mu]s, \tag{2}$$

where s is the estimated local population susceptibility, we used reproduction number R_t to exclude the influence of population susceptibility on disease transmission rate.

D , μ , Z (the average latency period from infection to contagiousness), and a multiplicative factor adjusting random movement (θ) were randomly drawn from the posterior distributions inferred from case data through March 13, 2020⁶⁰: $D = 3.56$ (3.21–3.83), $\mu = 0.64$ (0.56–0.70), $Z = 3.59$ (95% CI: 3.28–3.99), and $\theta = 0.15$ (0.12–0.17). Z and θ are used in ordinary differential equations used to model transmission dynamics (Supplementary Notes).

The daily transmission rate β and ascertainment rate α were estimated sequentially for each county using the ensemble adjustment Kalman filter (EAKF)⁶⁴. Specifically, parameters β_t and α_t for county i were updated each day using incidence and death data. We used the estimates on day $t - 1$ as the prior parameters on day t , and then updated the priors to posteriors using the EAKF and observations. The posteriors are the estimated parameter values on day t . To ensure a smooth parameter estimation, we imposed a $\pm 30\%$ limit on the daily change of parameters β_t and α_t . Other smoothing constraints were tested and the results were similar. To avoid possible inaccurate estimation for counties with few cases, we

inferred R_t in the 2669 U.S. counties with at least 400 cumulative confirmed cases as of December 31, 2020 (Supplementary Fig. 7).

Statistical analysis. All statistical analyses were conducted with R software (version 3.6.1) using the *mgcv* and *dlm* packages.

Association between meteorological factors and R_t . Given the potential non-linear and temporally delayed effects of meteorological factors, a distributed lag non-linear model⁶⁵ combined with generalized additive mixed models⁶⁶ was applied to estimate the associations of daily mean temperature, daily mean SH, and daily mean UV radiation with SARS-CoV-2 R_t . To quantify the total contribution, independent effects, and relative importance of meteorological factors (i.e., temperature, SH, and UV radiation), we included all three variables in the same model. To reduce collinearity, we used cross-basis terms rather than the raw variables (Supplementary Tables 5–6). The full model can be expressed as:

$$\begin{aligned} \log(E(R_{i,j,t})) = & \alpha + te(s(\text{latitude}_i, \text{longitude}_i, k = 200), s(\text{time}_t, k = 30)) + \text{cb.temperature} + \text{cb.SH} + \text{cb.UV} \\ & + \beta_1(\text{population density}_i) + \beta_2(\text{percent Black residents}_i) + \beta_3(\text{percent Hispanic residents}_i) \\ & + \beta_4(\text{percent people older than 60 years}_i) + \beta_5(\text{median household income}_i) \\ & + \beta_6(\text{percent owner - occupied housing}_i) \\ & + \beta_7(\text{percent residents older than 25 years without a high school diploma}_i) \\ & + \beta_8(\text{number of ICU beds per 10,000 people}_i) + \beta_9(\text{percent healthcare workers}_i) \\ & + \beta_{10}(\text{day when 100 cumulative cases per 100,000 people was reached}_i) + re(\text{county}_i) + re(\text{state}_j) \end{aligned} \tag{3}$$

where $E(R_{i,j,t})$ refers to the expected R_t in county i , state j , on day t , and α is the intercept. Given the distribution of R_t in our data close to a lognormal distribution (Supplementary Fig. 8), we used log-transformed R_t as the outcome variable, and the Gaussian family in the model. A thin plate spline with a maximum of 200 knots was used to control the coordinates of the centroid of each county; the time trend was controlled by a flexible natural cubic spline over the range of study dates with a maximum of 30 knots; due to the unique pattern of the non-linear time trend of R_t in each county (Supplementary Fig. 4), we constructed tensor product smooths (te) of the splines of geographical coordinates and time, to better control for the temporal and spatial variations (Supplementary Fig. 3).

Cb.temperature, cb.SH, and cb.UV are cross-basis terms for the mean air temperature, mean SH and mean UV radiation, respectively. We modeled exposure-response associations (meteorological factors vs. percent change in R_t) using a natural cubic spline with 3 degrees of freedom (df) and modeled the lag-response association using a natural cubic spline with an intercept and 3 df with a maximum lag of 13 days. We adjusted for county-level characteristics, including population density, percent Black residents, percent Hispanic residents, percent people older than 60 years, median household income, percent owner-occupied housing, percent residents older than 25 years without a high school diploma, number of ICU beds per 10,000 people, and percent healthcare workers, given their potential relationship with SARS-CoV-2 transmission^{67–70}. Day when 100 cumulative cases per 100,000 people was reached in each county was used to approximate local epidemic stage⁴⁵ (Supplementary Fig. 9). The random effects of state and county were modeled by parametric terms penalized by a ridge penalty (re), to further control for unmeasured state- and county-level confounding. Residual plots were used to diagnose the model (Supplementary Fig. 10). In additional analyses, we included air temperature, SH, and UV radiation in separate models (Supplementary Fig. 2).

Based on the estimated exposure-response curves, between the 1st and the 99th percentiles of the distribution of air temperature, SH, and UV radiation, we determined the value of exposure associated with the lowest relative risk of R_t to be the optimum temperature, the optimum SH, or the optimum UV radiation, respectively. The natural cubic spline functions of the exposure-response relationship were then re-centered with the optimum values of meteorological factors as reference values. We report the cumulative relative risk of R_t associated with daily temperature, SH, or UV radiation exposure in the previous two weeks (0–13 lag days) as the percent changes in R_t when comparing the daily exposure with the optimum reference values (i.e., the cumulative relative risk of R_t equals one and the percent change in R_t equals zero when the temperature, SH, or UV radiation exposure is at its optimum value).

Attribution of R_t to meteorological factors. We used the optimum value of temperature, SH, or UV radiation as the reference value for calculating the fraction of R_t attributable to each meteorological factor; i.e., the attributable fraction (AF). For these calculations, we assumed that the associations of meteorological factors with R_t were consistent across the counties. For each day in each county, based on the cumulative lagged effect (cumulative relative risk) corresponding to the temperature, SH, or UV radiation of that day, we calculated the attributable R_t in the current and next 13 days, using a previously established method⁷¹. Specifically, in a given county, the R_t attributable to a meteorological factor (x_t) for a given day t was defined as the attributable absolute excess of R_t ($AE_{x,t}$, the excess reproduction number on day t attributable to the deviation of temperature or SH from the optimum value) and the attributable fraction of R_t ($AF_{x,t}$, the fraction of R_t attributable to the deviation of the meteorological factor from its optimum value), each

accumulated over the current and next 13 days. The formulas can be expressed as:

$$AF_{x,t} = 1 - \exp\left(-\sum_{l=0}^{13} \beta_{x,t,l}\right) \quad (4)$$

$$AE_{x,t} = AF_{x,t} \times \sum_{l=0}^{13} \frac{n_{t+l}}{13+1}, \quad (5)$$

where n_t is the R_t on day t , and $\sum_{l=0}^{13} \beta_{x,t,l}$ is the overall cumulative log-relative risk for exposure x_t on day t obtained by the exposure-response curves re-centered on the optimum values. Then, the total absolute excess of R_t attributable to temperature, SH, or UV radiation in each county was calculated by summing the absolute excesses of all days during the study period, and the attributable fraction was calculated by dividing the total absolute excess of R_t for the county by the sum of the R_t of all days during the study period for the county. The attributable fraction for the 2669 counties combined was calculated in a similar manner at the national level. We derived the 95% eCI for the attributable absolute excess and attributable fraction by 1000 Monte Carlo simulations⁷¹. The total fraction of R_t attributable to meteorological factors was the sum of the attributable fraction for temperature, SH, and UV radiation. We also calculated the attributable fractions by month in the study period.

Sensitivity analyses. We conducted several sensitivity analyses to test the robustness of our results: (a) the lag dimension was redefined using a natural cubic spline and three equally placed internal knots in the log scale; (b) an alternative four df was used in the cross-basis term for meteorological factors in the exposure-response function; (c) the maximum number of knots was reduced to 25 in the flexible natural cubic spline to control time trend in the tensor product smooths; (d) all demographic and socioeconomic variables were excluded from the model; (e) adjustment for the prevalence of smoking and obesity among adults was included in the model; (f) adjustment for climate zone was included in the model; (g) additional adjustment was made for the average $PM_{2.5}$ concentration in each county during 2014–2018⁴⁵; (h) additional adjustment was made for daily mean $PM_{2.5}$, and daily maximum 8-h O_3 . For daily covariates with available data in only some of the counties or study period, the results of sensitivity analyses were compared to the main model re-run on the same partial dataset.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available in GitHub (<https://doi.org/10.5281/zenodo.4766014>)⁷².

Code availability

R code for this analysis is available at <https://github.com/CHENlab-Yale/COVID-Climate>.

Received: 17 November 2020; Accepted: 20 May 2021;

Published online: 14 June 2021

References

- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
- Matson, M. J. et al. Effect of environmental conditions on SARS-CoV-2 stability in human nasal mucus and sputum. *Emerg. Infect. Dis.* **26**, 2276–2278 (2020).
- Chin, A. W. H. et al. Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe* **1**, e10 (2020).
- Ratnesar-Shumate, S. et al. Simulated sunlight rapidly inactivates SARS-CoV-2 on surfaces. *J. Infect. Dis.* **222**, 214–222 (2020).
- Schuit, M. et al. Airborne SARS-CoV-2 is rapidly inactivated by simulated sunlight. *J. Infect. Dis.* **222**, 564–571 (2020).
- Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T. & Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **8**, e1000316 (2010).
- Sooryanarain, H. & Elankumaran, S. Environmental role in influenza virus outbreaks. *Annu. Rev. Anim. Biosci.* **3**, 347–373 (2015).
- Tan, J. et al. An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation. *J. Epidemiol. Community Health* **59**, 186–192 (2005).
- Abdul-Rasool, S. & Fielding, B. C. Understanding human coronavirus HCoV-NL63. *Open Virol. J.* **4**, 76–84 (2010).
- Esper, F., Weibel, C., Ferguson, D., Landry, M. L. & Kahn, J. S. Coronavirus HKU1 infection in the United States. *Emerg. Infect. Dis.* **12**, 775–779 (2006).
- Carlson, C. J., Gomez, A. C. R., Bansal, S. & Ryan, S. J. Misconceptions about weather and seasonality must not misguide COVID-19 response. *Nat. Commun.* **11**, 4312 (2020).
- Shi, P. et al. Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Sci. Total Environ.* **728**, 138890 (2020).
- Prata, D. N., Rodrigues, W. & Bermejo, P. H. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. *Sci. Total Environ.* **729**, 138862 (2020).
- Liu, J. et al. Impact of meteorological factors on the COVID-19 transmission: a multi-city study in China. *Sci. Total Environ.* **726**, 138513 (2020).
- Xie, J. & Zhu, Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Sci. Total Environ.* **724**, 138201 (2020).
- Pani, S. K., Lin, N. H. & RavindraBabu, S. Association of COVID-19 pandemic with meteorological parameters over Singapore. *Sci. Total Environ.* **740**, 140112 (2020).
- Isaia, G. et al. Does solar ultraviolet radiation play a role in COVID-19 infection and deaths? An environmental ecological study in Italy. *Sci. Total Environ.* **757**, 143757 (2021).
- Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
- Smit, A. J. et al. Winter is coming: a southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19. *Int. J. Environ. Res. Public Health* **17**, 5634 (2020).
- Wang, J. et al. Impact of temperature and relative humidity on the transmission of COVID-19: a modelling study in China and the United States. *BMJ Open* **11**, e043863 (2021).
- Poirier, C. et al. The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales. *Scientific reports* **10**, 1–11 (2020).
- Yao, Y. et al. No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *Eur. Respir. J.* **55**, 2000517 (2020).
- Gostic, K. M. et al. Practical considerations for measuring the effective reproductive number, R_t . *PLoS Comput. Biol.* **16**, e1008409 (2020).
- Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E. & Grenfell, B. T. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020).
- World Meteorological Organization (WMO). First report of the WMO COVID-19 Task Team: review on meteorological and air quality factors affecting the COVID-19 pandemic (WMO-No. 1262) (2021).
- Sehra, S. T., Salciccioli, J. D., Wiebe, D. J., Fundin, S. & Baker, J. F. Maximum daily temperature, precipitation, ultra-violet light and rates of transmission of SARS-Cov-2 in the United States. *Clin. Infect. Dis.* **71**, 2482–2487 (2020).
- Runkle, J. D. et al. Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. *Sci. Total Environ.* **740**, 140093 (2020).
- Wu, Y. et al. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci. Total Environ.* **729**, 139051 (2020).
- Moozhipurath, R. K., Kraft, L. & Skiera, B. Evidence of protective role of Ultraviolet-B (UVB) radiation in reducing COVID-19 deaths. *Sci. Rep.* **10**, 17705 (2020).
- Adnan, S. et al. Impact of heat index and ultraviolet index on COVID-19 in major cities of Pakistan. *J. Occup. Environ. Med.* **63**, 98–103 (2021).
- Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).
- van Doremalen, N. et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567 (2020).
- Morris, D. H. et al. Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses. *eLife* **10**, e65902 (2021).
- Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101 (2020).
- Shaman, J. & Galanti, M. Will SARS-CoV-2 become endemic? *Science* **370**, 527–529 (2020).
- Hart, P. H., Gorman, S. & Finlay-Jones, J. J. Modulation of the immune system by UV radiation: more than just the effects of vitamin D? *Nat. Rev. Immunol.* **11**, 584–596 (2011).
- Abhimanyu & Coussens, A. K. The role of UV radiation and vitamin D in the seasonality and outcomes of infectious disease. *Photochem. Photobiol. Sci.* **16**, 314–338 (2017).
- Merow, C. & Urban, M. C. Seasonality and uncertainty in global COVID-19 growth rates. *Proc Natl Acad Sci U S A*, 117, 27456–27464. (2020).
- Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).

40. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).
41. Klepeis, N. E. et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J. Expo. Anal. Environ. Epidemiol.* **11**, 231–252 (2001).
42. Shaman, J., Kandula, S., Yang, W. & Karspeck, A. The use of ambient humidity conditions to improve influenza forecast. *PLoS Comput Biol.* **13**, e1005844 (2017).
43. Quinn, A. & Shaman, J. Indoor temperature and humidity in New York City apartments during winter. *Sci. Total Environ.* **583**, 29–35 (2017).
44. Nguyen, J. L., Schwartz, J. & Dockery, D. W. The relationship between indoor and outdoor temperature, apparent temperature, relative humidity, and absolute humidity. *Indoor Air* **24**, 103–112 (2014).
45. Wu, X., Nethery, R. C., Sabath, M. B., Braun, D. & Dominici, F. Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. *Sci Adv* **6**, eabd4049 (2020).
46. Cosgrove, B. A. et al. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.: Atmos.* **108**, <https://doi.org/10.1029/2002JD003118> (2003).
47. Hersbach, H. et al. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (2018).
48. United States Census Bureau. Tables (2020).
49. Kaiser Health News. Millions Of Older Americans Live In Counties With No ICU Beds As Pandemic Intensifies. (2020).
50. Robert Wood Johnson Foundation. 2020 County Health Rankings (2020).
51. van Donkelaar, A., Martin, R. V., Li, C. & Burnett, R. T. 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.* **53**, 2595–2611 (2019).
52. United States Environmental Protection Agency. Outdoor Air Quality Data (2020).
53. U.S. Department of Energy. Guide to Determining Climate Regions by County. vol. 7.3 (2015).
54. SafeGraph. The Source of Truth for POI Data & Business Listings (2020).
55. Pei, S., Kandula, S., Yang, W. & Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proc. Natl Acad. Sci. USA* **115**, 2752–2757 (2018).
56. Sayamanathan, A. A. et al. Infectivity of asymptomatic versus symptomatic COVID-19. *Lancet* **397**, 93–94 (2021).
57. Bi, Q., et al. Household transmission of SARS-CoV-2: insights from a population-based serological survey. *medRxiv*, 2020.2011.2004.20225573. Preprint at <https://doi.org/10.1101/2020.11.04.20225573> (2020).
58. Gao, M. et al. A study on infectivity of asymptomatic SARS-CoV-2 carriers. *Respir. Med* **169**, 106026 (2020).
59. Byambasuren, O. et al. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *J. Assoc. Med Microbiol Infect. Dis. Can.* **5**, 223–234 (2020).
60. Pei, S. & Shaman, J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. *medRxiv*, 2020.2003.2021.20040303. Preprint at <https://doi.org/10.1101/2020.03.21.20040303> (2020).
61. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
62. Centers for Disease Control and Prevention. COVIDView: a weekly surveillance summary of U.S. COVID-19 Activity (2021).
63. Badr, H. S. & Gardner, L. M. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *Lancet Infect. Dis.* **21**, e113 (2020).
64. Anderson, J. L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903 (2001).
65. Gasparrini, A., Armstrong, B. & Kenward, M. G. Distributed lag non-linear models. *Stat. Med* **29**, 2224–2234 (2010).
66. Wood, S. N., Goude, Y. & Shaw, S. Generalized additive models for large data sets. *J. R. Stat. Soc. Ser. C. Appl Stat.* **64**, 139–155 (2015).
67. Bhadra, A., Mukherjee, A. & Sarkar, K. Impact of population density on Covid-19 infected and mortality rate in India. *Model Earth Syst. Environ.* **7**, 623–629 (2021).
68. Baena-Diez, J. M., Barroso, M., Cordeiro-Coelho, S. I., Díaz, J. L. & Grau, M. Impact of COVID-19 outbreak by income: hitting hardest the most deprived. *J. Public Health* **42**, 698–703 (2020).
69. Tanzer-Gruener, R., Li, J., Eilenberg, S. R., Robinson, A. L. & Presto, A. A. Impacts of modifiable factors on ambient air pollution: a case study of COVID-19 shutdowns. *Environ. Sci. Technol. Lett.* **7**, 554–559 (2020).
70. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med* **26**, 1205–1211 (2020).
71. Gasparrini, A. & Leone, M. Attributable risk from distributed lag models. *BMC Med. Res. Methodol.* **14**, 55 (2014).
72. CHENlab-Yale. CHENlab-Yale/COVID-Climate: First release of the public repository for the COVID-Climate project (2021).

Acknowledgements

Y.M. received funding from the China Scholarship Council (201906320022). S.P. and J.S. acknowledged funding from the National Institutes of Health (GM110748) and the National Science Foundation (DMS-2027369), as well as a gift from the Morris-Singer Foundation. R.D. received funding from the High Tide Foundation.

Author contributions

K.C. conceived of and supervised the conduct of this study and edited the manuscript. J. S. and R.D. contributed to the study design, interpretation of results, and manuscript revision. S.P. estimated the reproduction number and contributed to the writing. Y.M. conducted formal analyses and drafted the manuscript. All authors reviewed and approved the final version of this manuscript.

Competing interests

J.S. and Columbia University disclose partial ownership of SK Analytics. J.S. discloses consulting for BNI. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23866-7>.

Correspondence and requests for materials should be addressed to S.P. or K.C.

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

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

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



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

© The Author(s) 2021