

Weather, air pollution, and SARS-CoV-2 transmission: a global analysis



Ran Xu*, Hazhir Rahmandad*, Marichi Gupta, Catherine DiGennaro, Navid Ghaffarzadegan, Heresh Amini, Mohammad S Jalali

Summary

Background Understanding how environmental factors affect SARS-CoV-2 transmission could inform global containment efforts. Despite high scientific and public interest and multiple research reports, there is currently no consensus on the association of environmental factors and SARS-CoV-2 transmission. To address this research gap, we aimed to assess the relative risk of transmission associated with weather conditions and ambient air pollution.

Methods In this global analysis, we adjusted for the delay between infection and detection, estimated the daily reproduction number at 3739 global locations during the COVID-19 pandemic up until late April, 2020, and investigated its associations with daily local weather conditions (ie, temperature, humidity, precipitation, snowfall, moon illumination, sunlight hours, ultraviolet index, cloud cover, wind speed and direction, and pressure data) and ambient air pollution (ie, PM_{2.5}, nitrogen dioxide, ozone, and sulphur dioxide). To account for other confounding factors, we included both location-specific fixed effects and trends, controlling for between-location differences and heterogeneities in locations' responses over time. We built confidence in our estimations through synthetic data, robustness, and sensitivity analyses, and provided year-round global projections for weather-related risk of global SARS-CoV-2 transmission.

Findings Our dataset included data collected between Dec 12, 2019, and April 22, 2020. Several weather variables and ambient air pollution were associated with the spread of SARS-CoV-2 across 3739 global locations. We found a moderate, negative relationship between the estimated reproduction number and temperatures warmer than 25°C (a decrease of 3.7% [95% CI 1.9–5.4] per additional degree), a U-shaped relationship with outdoor ultraviolet exposure, and weaker positive associations with air pressure, wind speed, precipitation, diurnal temperature, sulphur dioxide, and ozone. Results were robust to multiple assumptions. Independent research building on our estimates provides strong support for the resulting projections across nations.

Interpretation Warmer temperature and moderate outdoor ultraviolet exposure result in a slight reduction in the transmission of SARS-CoV-2; however, changes in weather or air pollution alone are not enough to contain the spread of SARS-CoV-2 with other factors having greater effects.

Funding None.

Copyright © 2021 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

The COVID-19 pandemic has greatly challenged the global community. High-stake policy decisions depend on how environmental factors affect the transmission of the disease.¹ Given that many related viral infections such as seasonal influenza,² MERS-CoV,³ and SARS-CoV⁴ show notable seasonality, it might be expected that the transmission of SARS-CoV-2 is similarly dependent on weather. Earlier works indicate that temperature,⁵ humidity,⁶ air pressure, ultraviolet exposure, and precipitation could affect the spread of SARS-CoV-2 by changing the virus survival times on surfaces and in droplets,^{7,8} moderating the distance virions can travel through air,⁸ changing host susceptibility, and affecting individual activity patterns and immune systems.^{7,8} A few other studies suggest air pollutants could act as vectors for the virus or affect the immune system.^{9,10} Yet, there is poor agreement on the shape and magnitude of these relationships.

Although some studies have found associations between pandemic severity and variations in temperature,^{6,11–17} relative and absolute humidity,^{6,11–14,16–19} ultraviolet,¹⁶ wind speed,¹² visibility, and precipitation,¹¹ others^{20,21} indicate weaker, inconsistent, or no relationships. A 2020 review found inconclusive evidence for the role of weather in SARS-CoV-2 transmission¹ and others caution against interpreting weather as a key driver due to this uncertainty.²²

The explanation for these inconclusive results is unclear. Estimates that are based on datasets focused only on China or the USA could lack generalisability.^{11,13,14,18,19,23,24} Others have studied only a subset of meteorological measures, complicating comparisons.^{6,14,18,19,23} Most studies have not controlled for other important factors, such as varying government and public responses, population density, and cultural practices.^{6,11,12,19,24} The delay between infection and official recording of cases is a particularly understudied factor. Failure to correct for these delays,

Lancet Planet Health 2021; 5: e671–80

*Contributed equally

Department of Allied Health Sciences, University of Connecticut, Storrs, CT, USA (R Xu PhD); Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA (H Rahmandad PhD); Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA, USA (M Gupta, C DiGennaro BA, M S Jalali PhD); Department of Industrial and Systems Engineering, Virginia Tech, Falls Church, VA, USA (N Ghaffarzadegan PhD); Department of Public Health, University of Copenhagen, Copenhagen, Denmark (H Amini PhD); Department of Environmental Health, Harvard T H Chan School of Public Health, Boston, MA, USA (H Amini); MGH Institute for Technology Assessment, Harvard Medical School, Harvard University, Boston, MA, USA (M S Jalali)

Correspondence to: Mohammad S Jalali, MGH Institute for Technology Assessment, Harvard Medical School, Harvard University, Boston, MA 02114, USA msjalali@mgh.harvard.edu

Research in context**Evidence before this study**

Understanding and projecting the spread of SARS-CoV-2 requires reliable estimates of the effect of environmental factors on the transmission of the virus. Weather and ambient air pollution can affect transmission in multiple ways, including viral survival, host susceptibility, and behavioural change. Summarising evidence from the first few months of the COVID-19 pandemic, members of the National Academies of Sciences, Engineering, and Medicine's Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats highlighted inconsistent results with regard to the association between SARS-CoV-2's survival and transmissibility and weather-related factors. Several limitations of previous studies might have led to this absence of consensus: only including data from a subset of geographical locations; considering only a subset of variables (eg, temperature and humidity); not controlling for important factors, such as government and public responses, population density, and cultural practices; and not correctly accounting for delays between actual infection and official recording of cases.

Added value of this study

Addressing the current research gaps, we estimated the association of weather and SARS-CoV-2 transmission by assembling and analysing a large dataset integrating early COVID-19 infections, weather, and air pollution, and used the results to project relative transmission risk globally. Our research provides important evidence-based findings to assist global policy and decision making as states and countries balance vaccination, physical distancing, and other responses to contain the pandemic.

Implications of all the available evidence

Our findings provide evidence for relationships between several weather variables, ambient air pollution, and the spread of COVID-19. The magnitude of these associations is modest; hence, weather alone is not likely to contain SARS-CoV-2 transmission but acts as a potent force multiplier in conjunction with various policies and interventions.

estimated to be approximately 10 days,^{25,26} confounds attempts to associate daily weather conditions with recorded new cases and could partly explain the inconsistent and inconclusive findings to date. A comprehensive evaluation of factors that might affect SARS-CoV-2 transmission, stratified by location and measured over time, is essential to predict and protect against disease spread.^{27,28}

Here, we focus on the early stages of the pandemic, and assemble a dataset of the global spread of the COVID-19 pandemic up until late April, 2020, spanning 3739 locations globally. We aimed to validate and apply a statistical method to estimate the daily reproduction number in each location. Controlling for location-specific differences (eg, in population density, cultural practices, socioeconomic differences, public transportation, and age distribution) as well as time-variant responses in each location (eg, physical distancing, quarantine, lockdowns, and public space closures), we estimated the association of weather and air pollutants with the reproductive number of SARS-CoV-2 and provide year-round, global projections for May, 2020 to April, 2021.

Methods**Study design**

Our dataset includes infection data for 3739 distinct locations, spanning Dec 12, 2019, to April 22, 2020. We augment the data reported by the Johns Hopkins Center for Systems Science²⁹ with data reported by the Chinese Center for Disease Control and Prevention, Provincial Health Commissions in China, and Iran's state-level reports. We assembled disaggregate data for the spread of SARS-CoV-2 in Australia (eight states), Canada (ten

provinces), China (34 province-level administrative units and 301 individual cities), Iran (31 provinces), and the USA (3144 counties and five territories). For remaining countries or territories with reported COVID-19 cases (206 locations, see appendix p 2), country-level aggregates were used. We did not exclude any location without reported cases.

We compiled weather data from archival databases (World Weather Online and OpenWeather), and air pollution data from the European Centre for Medium-Range Weather Forecasts. Because reliable COVID-19 case data at the city level was scarce worldwide, for country-level locations that included cities with populations of 500 000 or greater, weather and pollution data were first gathered for each city and then its mean weighted by each city's population into country-level measures. The weighted variables thus better represent the weather and pollution condition of heavily populated cities in which the reported cases are predominantly concentrated. For Australian states, Canadian provinces, US counties, and any remaining countries, we used the weather and air pollution data for the coordinate of the centroid of that location (appendix p 9 shows robustness to uncertainty of case report data due to different sizes of the locations). We obtained daily data for minimum, maximum, and mean temperature, humidity, precipitation, snowfall, moon illumination, sunlight hours, ultraviolet index, cloud cover, wind speed and direction, pressure data, as well as air pollutants including PM_{2.5}, nitrogen dioxide (NO₂), ozone (O₃), and sulphur dioxide (SO₂). We used population density data from Demographia, the US Census Bureau, the Iran Statistical Centre, the UN's Projections, City Population, and official published

See Online for appendix

For **World Weather Online** see <https://www.worldweatheronline.com/developer/>

For **OpenWeather** see <https://openweathermap.org/>

For the **European Centre for Medium-Range Weather Forecasts** see <https://apps.ecmwf.int/datasets/data/cams-nrealtime/levtype=sf>

For **Chinese Center for Disease Control and Prevention** see <https://www.chinacdc.cn/en>

For **Demographia report** see <http://www.demographia.com/db-worldua.pdf>

For **US Census Bureau** see <https://www.census.gov/data/tables/time-series/dec/density-data-text.html>

For **Iran Statistical Centre** see <https://www.amar.org.ir/english>

For the **United Nation's Projections** see <https://www.un.org/development/desa/pd>

For **City Population** see <https://citypopulation.de/>

estimates for countries not covered by these sources (UN city and country density).

Estimation of the reproduction number

An essential parameter in understanding the spread of an epidemic is the effective reproduction number, R_e , the expected number of secondary cases generated by an index patient. An epidemic grows when R_e is greater than 1 and will die out once R_e stays below 1. Reproduction number can be approximated by (\hat{R}) based on the number of new infections (I_m) per currently infected individual, multiplied by the duration of illness (τ). Actual new infections on any day (I_N) are not directly observable but if an unbiased estimator, \hat{I}_N , is available, one can estimate \hat{R} using equation 1. To develop an unbiased estimator, \hat{I}_N , we note that data for measured daily infections (I_m) lag actual new infections by both the incubation period and the delay between the onset of symptoms and testing and recording of a case. We used published measures to quantify the distribution of both the incubation period (mean of between 5–6 days) and onset-to-detection delays (mean of between 4–6 days).^{25,26} Together these shape the overall detection delay. Given the variance in detection delay, a simple shift of measured infection by the mean delay (about 10 days) offers an unreliable estimate of true infections (appendix pp 5, 24). We therefore developed an algorithm to find the most likely actual daily new infections \hat{I}_N based on the observed measured infections (I_m) and the detection delay distribution (appendix p 5). We then used \hat{I}_N to estimate the reproduction number:

$$\hat{R}(t) = \frac{\hat{I}_N(t)\tau}{\sum_{s=t-\tau-1}^{s=t-1} \hat{I}_N(s)}$$

We use the daily $\hat{R}(t)$ as our dependent variable. The estimate of $\hat{R}(t)$ is robust to the existence of asymptomatic cases and under-reporting:³⁰ if actual cases are f times larger than those detected, the f factor shows up in both the numerator and denominator of equation 1 and cancels out (appendix pp 5, 26). We used a delay of $\tau=20$ days from exposure to resolution; results are robust to other durations of illness (appendix p 8). For each location, we only included days with \hat{I}_N values greater than 1. Reliability of early \hat{I}_N values for each location is affected by irregularities in early testing. Moreover, an unbiased estimate for $\hat{R}(t)$ requires τ days of previous new infection estimates. Thus, to ensure robustness we exclude the first 20 days after \hat{I}_N reaches 1 in each location (appendix p 9). This results in 19 221 observations of $\hat{R}(t)$ across the globe, which are included in our main model specification.

Controls for estimating \hat{R}

The reproduction number for SARS-CoV-2 primarily varies due to location-specific factors, from population

density, cultural practices, and public transportation use to age and comorbidity distribution, and genetic profile, among others. We controlled for these and other unobserved factors using location fixed effects (a parameter for each location estimating the effect of all those location-specific factors). Moreover, school closures, a ban on public gatherings, physical distancing, and other behavioural responses reduce R_e over time. The reproduction number might also increase if adherence to non-pharmaceutical interventions wanes (eg, when contact tracing is overwhelmed, or due to reduction of public risk perception^{31,32}). We account for such changes by estimating a location-specific time trend in R_e and assess sensitivity to non-linear trend controls in appendix p 9, and as we focus on the early history of the pandemic, more complex non-linear trajectories of time effect are not required. We also separately controlled for day of the week to account for the possible variations in human behavioural patterns in each day of the week.

Independent predictors

Previous studies^{6,13} suggest R_e might depend on various meteorological and air pollution factors through at least three pathways. First, the survival of the virus on surfaces and the spread of droplets and particles containing the virus might be affected by temperature, ultraviolet, humidity, wind, and particulate matter.^{8,10} Second, human host susceptibility might be affected because of factors modulating immune responses (eg, effect of ultraviolet exposure on serum vitamin D³⁶) and respiratory tract susceptibility to virus (eg, temperature, humidity, and air pollutants^{6,9}). Finally, the behaviour of human hosts (eg, interacting indoors and outdoors) is likely to be affected by multiple factors, ranging from temperature and precipitation to air pollutants, ultraviolet exposure, and humidity.³³ Our data do not allow us to establish these distinct pathways explicitly. Instead, we included the following potential contributing factors, focusing on their direct effects in the main specification and discussing interactions on appendix p 15: temperature (mean and diurnal temperature, difference between maximum and minimum daily temperature), relative humidity, pressure, precipitation, mean wind speed, and ultraviolet exposure (UV index). We included O₃ and SO₂, as air pollutants (PM_{2.5}, reported on appendix p 16). Beyond the main specification, we also explored a few interactions among these variables and the inclusion of other environmental variables, including absolute humidity, number of sun h received, snowfall, moon illumination, NO₂, and PM_{2.5}, and report those results on appendix p 16.

Statistical specification and validation

Given the large variations in \hat{R} estimated in this method, we used a log transformation of $\hat{R}(t)$ and linear models to predict $\ln(\hat{R})$. We designed and validated our statistical model for estimating $\ln(\hat{R})$ by testing its ability to identify true parameters in synthetic data. Specifically, we built a

For UN city and country density see <https://unstats.un.org/unsd/demographic-social/products/dyb/>

stochastic simulation model of the COVID-19 epidemic, generated synthetic infection data using historical weather inputs and presumed impact functions, and designed a statistical model that reliably found those presumed effects under an ensemble of simulated epidemics with different basic reproduction numbers, weather effects, population sizes, and test coverage, among others. We found that: (1) given actual infections (I_N), our method identifies the presumed functional form relating weather to transmission rate; (2) estimates become conservative (between null and true effects) when true infections are inferred rather than known; and (3) our method for inferring infections offers substantially better results than a simple shifting of official counts (appendix pp 5, 24).

Separately, to independently validate the resulting statistical method, NG and MG created a realistic individual-based model of disease transmission and used that to generate a separate test dataset with synthetic epidemics. Three scenarios were created using actual temperatures from a sample of 100 regions and three different functions for the temperature effect. RX, who was masked to the true functions in this synthetic dataset, successfully estimated the correct qualitative shape of those functions using our method (appendix p 28).

Building upon these findings, we predicted the logarithm of estimated reproduction number for location i on day t ($\ln[\hat{R}_{it}]$). Our main specification excluded days with \hat{I}_N less than 1 and the first 20 days after \hat{I}_N exceeds 1 for the first time. The model included location-specific fixed effects and trends and the following effects: a linear spline for the effect of mean temperature on transmission (\bar{T}), with the knot at 25°C (see appendix p 12 for

alternative knot values), diurnal temperature (ΔT), air pressure (P), relative humidity (H), linear and quadratic effects of ultraviolet index (UV), log precipitation (C), log wind speed (W), log O_3 , and log SO_2 :

$$\begin{aligned} \ln(\hat{R}_{it}) = & \alpha_i + \theta_i t + \beta_1 \min[\bar{T}_{it}, 25] + \beta_2 (\max[\bar{T}_{it}, 25] - 25) \\ & + \beta_3 \Delta T_{it} + \beta_4 (P_{it} - 1000) + \beta_5 H_{it} + \beta_6 (UV_{it} - 7 \cdot 13) \\ & + \beta_7 (UV_{it} - 7 \cdot 13)^2 + \beta_8 \ln(C_{it} + 1) + \beta_9 \ln(W_{it} + 1) \\ & + \beta_{10} \ln(O_{3it} + 1) + \beta_{11} \ln(SO_{2it} + 1) + \varepsilon_{it} \end{aligned}$$

We projected the effect of weather and air pollution on the relative risk of transmission for all locations in our sample and 1072 major urban areas (population >500 000) constituting about 30% of the world's population. A summary of results is provided in this Article and an interactive online platform offers comprehensive projections.

Role of the funding source

There was no funding source for this study.

Results

Our dataset included data collected between Dec 12, 2019, and April 22, 2020. Our main results for the association of weather on SARS-CoV-2 transmission are shown in the table. The model explains roughly three-quarters of the variance in $\ln(\hat{R})$ values (coefficient of determination $R^2=0.740$), much of which is due to fixed effects (39.2% of variance) and trends (34.6% of variance). Mean initial \hat{R} was 1.98 (IQR 0.88 to 2.49) 20 days after the first estimated case with much variation across locations (figure 1A). Initial \hat{R} was negatively correlated with epidemic start time and positively with population density. Most locations showed rapid reductions in reproduction number over time that capture the effect of policies and behavioural changes that reduce contacts. Typically, \hat{R} falls 5.8% (IQR -1.7 to 8.7) per day but with notable variability across locations (figure 1A) partly explained by locations with higher initial \hat{R} having faster subsequent reductions. For example, after excluding the first 20 days of estimated infections, New York City (NY, USA) showed an initial \hat{R} of 5.07, followed by a 7.8% daily reduction (figure 1B).

Even after controlling for these factors, mean temperature, ultraviolet index, diurnal temperature, air pressure, wind speed, precipitation, O_3 , and SO_2 were significantly associated with transmission (table). We found a robust effect of mean daily temperature, which for simplicity we characterised within two regimes, less than and more than 25°C (figure 2A). Temperatures higher than 25°C were associated with lower transmission rates (by 3.7% [95% CI 1.9–5.4] per additional degree), whereas those less than that threshold had a smaller effect (0.4% [0.14–0.66] reduction per degree).

Ultraviolet exposure had a robust U-shaped effect on the reproduction number, with a minimum ultraviolet index of around 6.3 (1 ultraviolet index equals 25 mW/m²;

For online platform see <https://projects.iq.harvard.edu/covid19>

	Mean (SD)	Coefficient (95% CI)	p value	Standardised coefficient*
Wind speed, log of km per ht	2.552 (0.444)	0.0323 (0.0079 to 0.0567)	0.010	0.0144
Precipitation, log of mm	0.785 (1.022)	0.0265 (0.0168 to 0.0361)	<0.0001	0.0271
Air pressure, millibars	1015.471 (6.077)	0.0022 (0.0005 to 0.0039)	0.013	0.0132
Humidity, %	66.946 (15.075)	-0.0006 (-0.0015 to 0.0003)	0.18	-0.0091
Mean temperature <25°C	11.351 (7.089)	-0.0040 (-0.0066 to -0.0014)	0.0025	-0.031
Mean temperature >25°C	27.761 (2.204)	-0.0377 (-0.0559 to -0.0194)	<0.0001	-0.0391
Ultraviolet index, 25 mW/m ²	7.129 (2.824)	0.0089 (-0.0083 to 0.0260)	0.31	0.0250
Square of ultraviolet index	..	0.0053 (0.0025 to 0.0081)	<0.0002	0.0702
Diurnal temperature, °C	8.757 (3.315)	0.0042 (0.0012 to 0.0072)	0.0063	0.0139
Ozone, log of ppbv	3.289 (0.667)	0.0349 (0.0161 to 0.0537)	<0.0003	0.0232
Sulphur dioxide, log of ppbv	1.139 (0.936)	0.0301 (0.0118 to 0.0485)	0.0013	0.0282

N=19 221. Fixed effects (mean 0.377 (SD 0.767)); mean of SE across locations 0.519, and trends (mean -0.060 [SD 0.157]); mean of SE across locations 0.107. R^2 (coefficient of determination)=0.740. ppbv=parts per billion by volume. *Standardised coefficients were obtained by first standardising all of the weather variables (mean 0 [SD 1]) and then re-running the analysis with our main model specification. †The outcome (reproduction number) is a log-transformed variable; hence the interpretation for the coefficient (β) of a log-transformed variable (X) can be that a 1% change in X is associated with β percentage change in estimated reproduction number.

Table: Association of weather and air pollution and SARS-CoV-2 transmission

(figure 2B). At a low or moderate ultraviolet index of 3.0, a unit higher ultraviolet index decreased \hat{R} by 3.5% (95% CI 0.4–6.4). At a high ultraviolet index of 10.0, a unit higher ultraviolet index increased \hat{R} by 4% (1.8–6.3). Although less robust across specifications (appendix p 9), we also found weak to moderate and significant positive effects of diurnal temperature, air pressure, wind speed, precipitation, O_3 , and SO_2 . A 1 SD increase in each increases \hat{R} by 1.4% (95% CI 0.4–2.4) for diurnal temperature, 1.3% (0.3–2.4) for air pressure, 1.4% (0.4–2.5) for log-transformed wind speed, 2.7% (1.7–3.8) for log-transformed precipitation, 2.4% (1.1–3.6) for log-transformed O_3 , and 2.9% (1.1–4.6) for log-transformed SO_2 . Including these covariates does not change the effect of temperature and ultraviolet (appendix p 15).

We also found that a few interactions among these predictors might be relevant in establishing transmission rates (appendix p 15). Overall, the association of various weather and air pollution variables with SARS-CoV-2 transmission is large enough to be relevant to assessing transmission risk across locations and seasons. Variations in the reproduction number associated with the combined set of predictors in our estimation dataset showed a ratio of 1.24 between the 95th and fifth percentiles despite the sample largely coming from late winter or early spring and the conservative nature of these estimates. Given that the typical reproduction number estimated for SARS-CoV-2 is in the range of 2–3,^{34,35} estimated weather effects alone might not provide a path to containing the epidemic in most locations but could notably affect the relative transmission rates.

Validation of our statistical method using synthetic data (appendix p 20) showed that: (1) our results are robust to under-reporting and changes in test coverage; (2) our method can identify the correct sign and shape for the effect of environmental variables; and (3) those estimates are potentially conservative (ie, smaller than the true effects). The conservatism is due to two factors. First, unavoidable errors in estimating daily infections from lagged official data lead to imperfect matching of independent variables and true infection rates, weakening any estimated relationship. Second, fixed effects and trends further weaken the signal used for estimation. If a region has a lower or higher baseline reproduction number, or a faster change, due to weather factors, those effects are absorbed in the fixed or trend effects; only changes not accounted for by the trend are attributed to weather effects.

We also did eight empirical tests to assess the robustness of our findings (appendix p 6). First, our results did not change with the use of different illness durations to calculate \hat{R} . Second, our main findings are robust to excluding extreme values of the dependent variables, the last few days of data, only using the US sample, and the inclusion of location-specific quadratic trend or time

fixed effects. Third, our results are largely insensitive to different exclusion criteria for initial periods of transmission per location. Fourth, when independent variables in each location are permuted and shifted in a placebo test, no effects remain, showing that results are

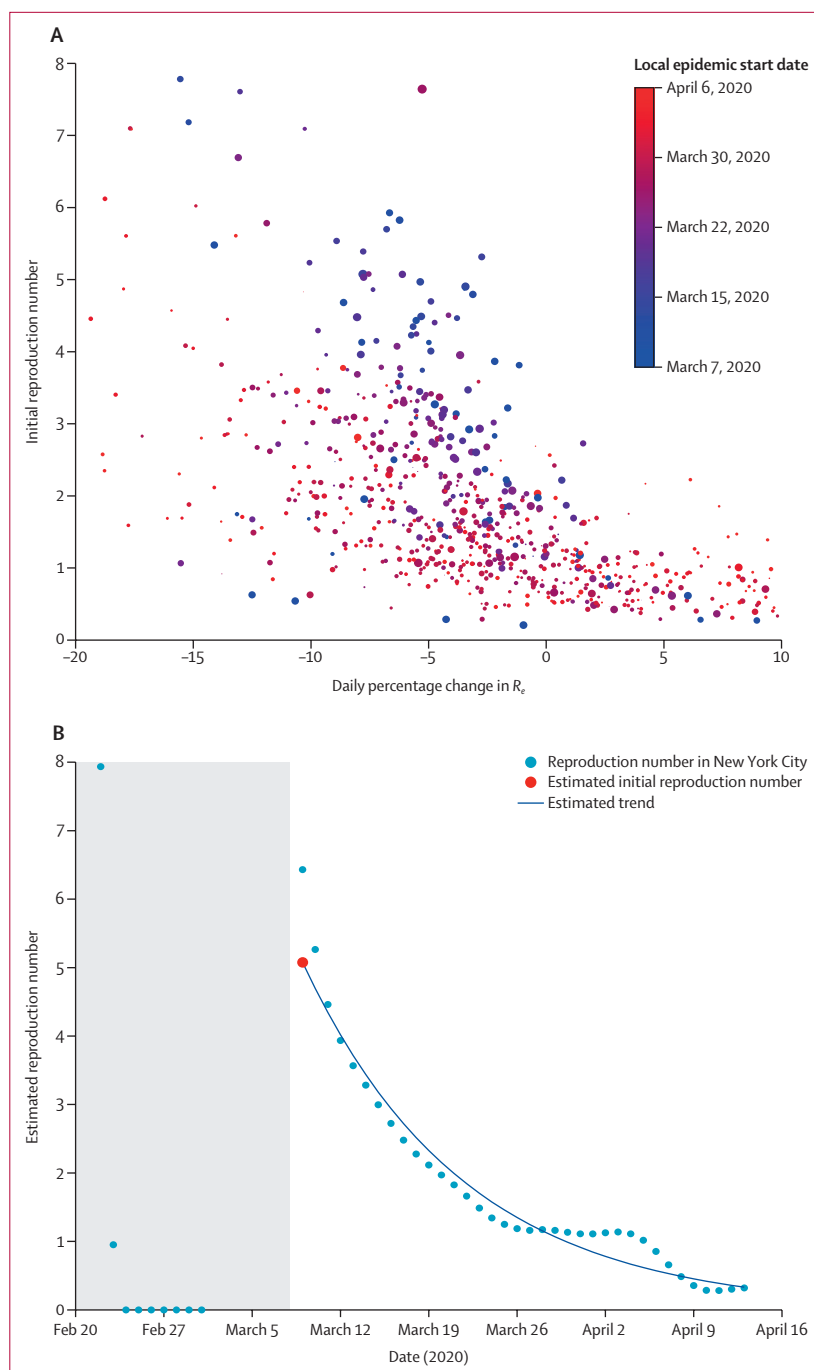


Figure 1: Estimated reproduction number over time

(A) Scatter diagram of initial reproduction number and daily percentage change across different locations, colour coded for date of local epidemic start. Size of the circle is proportional to population density in a location. (B) Estimated daily reproduction number \hat{R}_t values for New York City (20 days in grey area excluded). R_e = effective reproduction number.

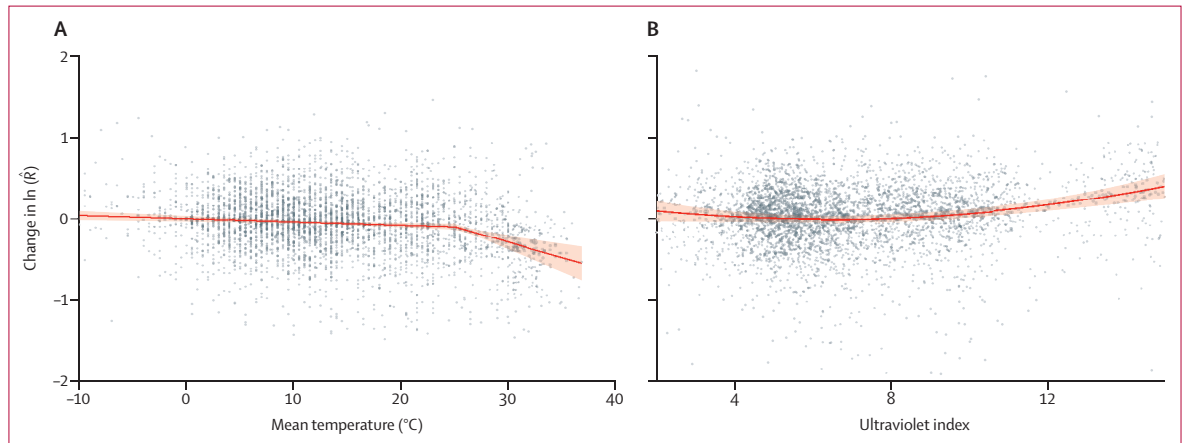


Figure 2: Association of mean temperature and ultraviolet index and SARS-CoV-2 transmission
 (A) Relationship between temperature and reproduction number (β_1 and β_2) with its uncertainty. (B) Relationship between ultraviolet index and reproduction number. A and B include (80% downsampled for temperature less than 30°C and ultraviolet index less than 14 (25 mW/m²) original data points and the relationships of interest were estimated after controlling for other factors as in equation 2. \hat{R} =estimated reproduction number.

not an artifact of the statistical method. Fifth, using different knots for the spline effect of temperature shows 25°C best separates the effect into two distinct slopes. Sixth, the estimated U-shaped effect of ultraviolet does not change when observations with high ultraviolet index are excluded. Seventh, we explored more interaction terms, and additional weather variables (eg, absolute humidity, NO₂, and PM_{2.5}), none of which are significant predictors or change the main results. Finally, overall projections of how weather and air pollution affect transmission rates using various specifications and on independent samples are consistent with our main specification.

Our results are associative and extrapolating out of sample includes unknown risks. With that caveat in mind, one can calculate the contribution of weather and air pollution to expected transmission for any vector of weather and air pollution based on results in the table. We defined relative COVID-19 risk associated with weather and air pollution (CRW) as the relative predicted risk of each weather and air pollution vector relative to the 95th percentile of predicted risk in our estimation sample (1.476). The choice of this reference point is somewhat arbitrary but makes a value of 1 a rather high-risk level. A CRW of 0.5 reflects a 50% reduction in the estimated reproduction number compared with this (high-risk) reference. Formally:

$$\begin{aligned}
 CRW = & \exp(0.0301 \ln(SO_2 + 1) + 0.0349 \ln(O_3 + 1) \\
 & + 0.0323 \ln(W + 1) + 0.0265 \ln(C + 1) \\
 & + 0.00217(P - 1000) - 0.000602H \\
 & - 0.00399 \min[\bar{T}, 25] - 0.0377 (\max[\bar{T}, 25] - 25) \\
 & + 0.00421\Delta T + 0.00886 (UV - 7.13) \\
 & + 0.00533 (UV - 7.13)^2) / 1.4757
 \end{aligned}$$

These scores do not reveal the actual value of R_t as that value is contingent on location-specific factors and

policies for which we have no data outside the estimation sample. Rather, CRW scores inform relative risks due to weather and air pollution (ie, assuming all else is equal) across locations, or within a location over time.

Figure 3 provides a visual summary of global CRW scores, with the mean calculated over the first half of July, 2020. The colour-coded scores suggest much variation in the expected risk of SARS-CoV-2 transmission across locations, with increased risks due to both low temperature (some regions in the southern hemisphere) and very high ultraviolet indexes (some locations in Central America). Additional snapshots of global CRW scores at different times of the year are shown in the appendix (p 31), and the website for weather, air pollution, and COVID-19 transmission offers week-by-week risk measures year-round.

Figure 4 shows CRW projections for five major cities in each of four regions. These projections used weather and air pollution from 2019, with means calculated over a 15-day moving window, for 2020–21 dates; as such, they do not reflect actual weather on a given date, only a historically informed projection. Live projections could be achieved using current weather data. Many large cities go through periods of higher and lower risk during the year. We cannot associate these risks with absolute reproduction numbers, and our estimates are likely to be conservative. Nevertheless, assuming typical basic reproduction rates (eg, 2–3), weather factors will not bring the reproduction number to less than 1. For example, in New York City, with estimated $\hat{R}_{New York}$ (0) of around 5, the effect of weather could lead to a 30% variation in the reproduction number (ie, the 4–6 range), requiring substantial interventions (ie, vaccination and physical distancing) to enable containment regardless of weather. The website for weather, air pollution, and COVID-19 transmission provides these projections for the 1072 largest global cities.

For the website for weather, air pollution, and COVID-19 transmission see <https://projects.iq.harvard.edu/covid19/resources>

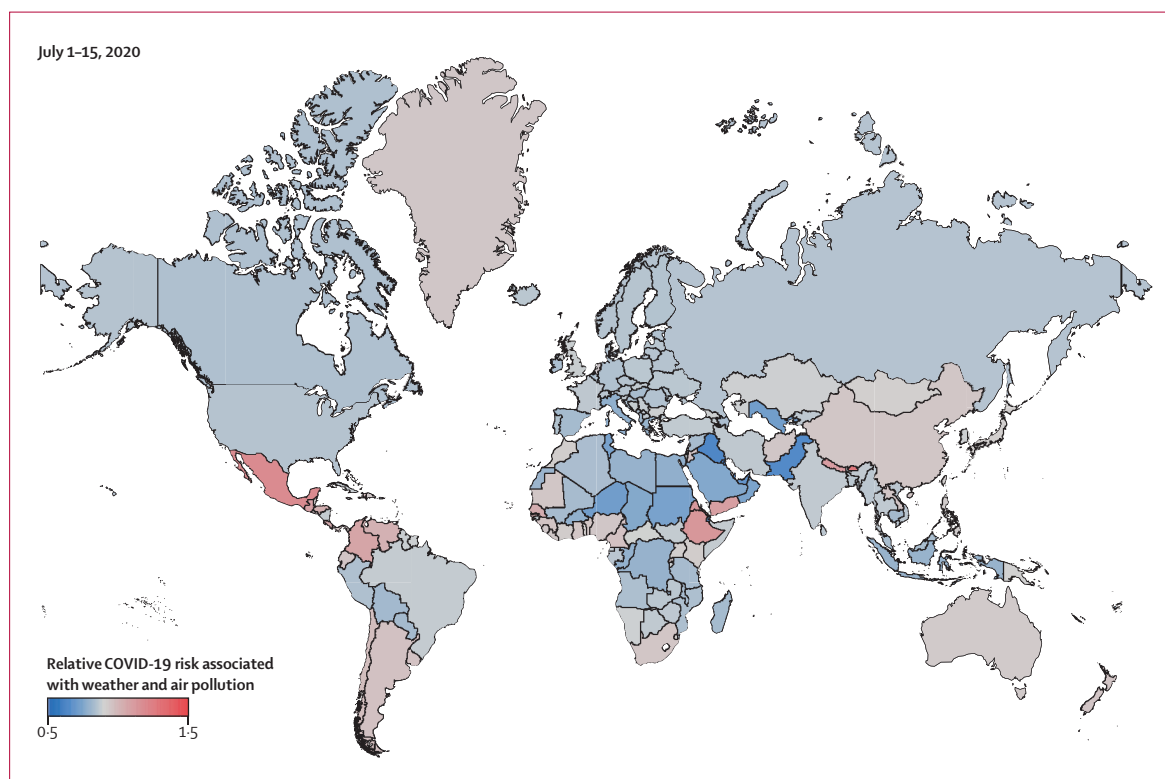


Figure 3: Relative COVID-19 risk associated with weather and air pollution globally, with means calculated over the first half of July, 2020

Discussion

This work combines an extensive dataset of early SARS-CoV-2 transmission with weather and air pollution data across the world to estimate the association of various environmental variables with the spread of SARS-CoV-2.

We found a strong association between temperatures higher than 25°C and reduced transmission rates, and a weaker effect for temperature less than 25°C. These findings suggest that many temperate zones with high population density could face larger risks in winter, whereas some warmer areas of the world might have slower transmission rates in general. The U-shaped relationship between ultraviolet index and transmission could reduce transmission in more temperate regions during summer, but lead to higher risks in equatorial regions with very high ultraviolet exposure.

Most of the associations we found are consistent with theoretical mechanisms thought to link environmental factors to transmission: the negative temperature effect on transmission, boosted at higher temperatures, is consistent with virus survival rates in experimental work;⁸ the positive effects of wind and precipitation could result from people spending more time indoors in which transmission is more probable than outdoors; and the effect of air pollutants might be related to increased susceptibility in more polluted environments.⁹

Air pollution is well known to cause a range of health outcomes, such as chronic obstructive pulmonary disease, cardiovascular diseases, and respiratory diseases, among others. All individuals with such conditions are susceptible to COVID-19 and are at-risk groups. Nevertheless, we remain mindful that our study design and data cannot establish such mechanisms empirically. For example, we hypothesised that ultraviolet exposure would reduce transmission (due to both stimulating vitamin D production and ultraviolet's disinfecting effects). Our estimates have the expected sign in the low ranges of ultraviolet, but also reveal an unexpected increase in the high ranges of ultraviolet. The unexpected increase in the high ranges of ultraviolet could be due to a shift of social interactions into higher risk, indoor, settings when ultraviolet amounts are very high; but we cannot test such explanations here.

Methodologically, we showed that accounting for the distribution of the delay between infection and detection is important. Many previous studies did not fully account for this delay, or its distribution, which might partly explain inconsistent previous results. We also showed that our methods and results are robust to significant under-counting of cases in official data, and to changes in test coverage over time, both major concerns in using official case data (appendix pp 5, 22).

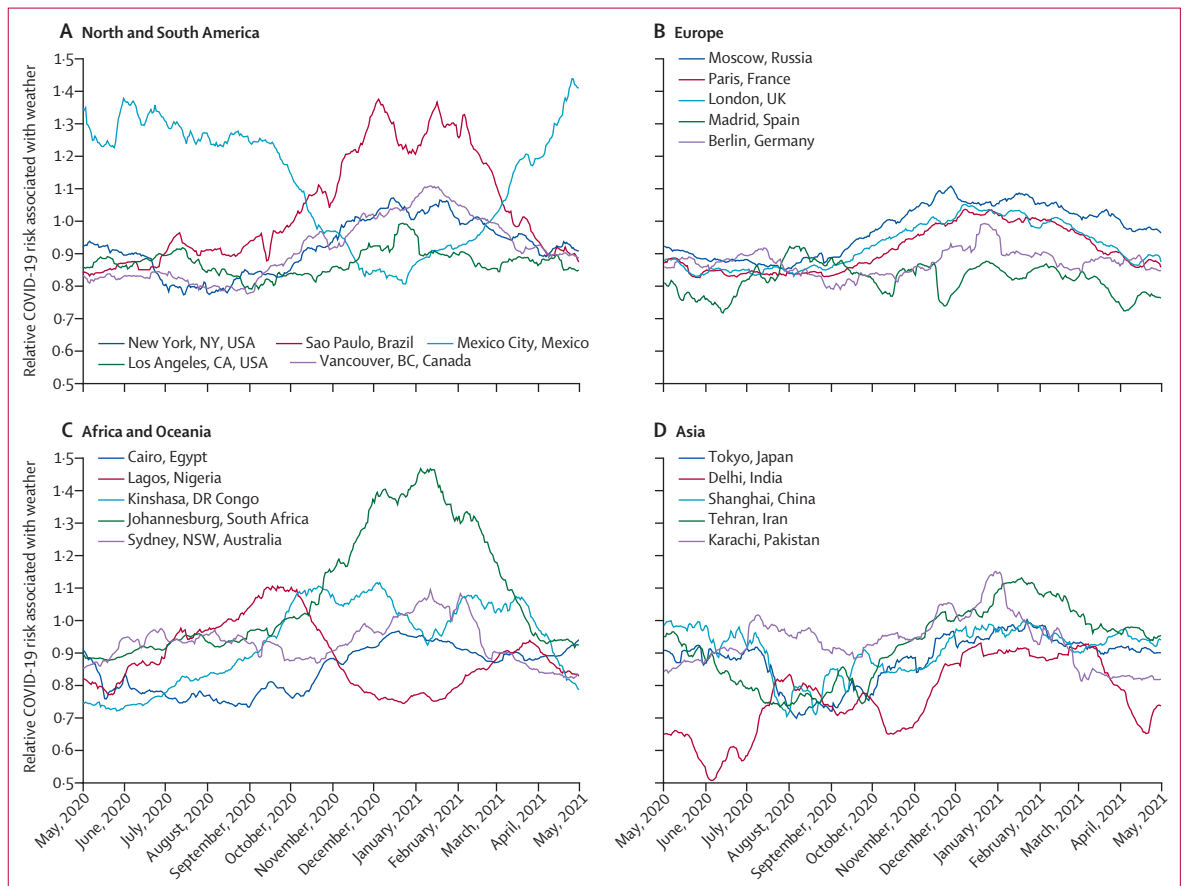


Figure 4: CRW measures over the year for major cities around the world

CRW projections for five major cities in each of the four regions of North and South America (A), Europe (B), Africa and Oceania (C), and Asia (D). CRW=relative COVID-19 risk associated with weather and air pollution.

Nevertheless, we acknowledge the inherent challenges in estimating R_t from the case report data, which can lead to conservative overall estimates as evidenced by our simulation studies. This fact should also be noted in using our projections. Other limitations include the scarcity of reliable transmission data in some regions of the world, the focus on early stages of the pandemic, the oversampling from US locations, the limited data with high temperature and ultraviolet in our estimation sample, which reduce confidence for projections when either is very high, use of 2020's weather data to project 2021's outcomes, and use of correlational evidence to inform out-of-sample projections. These items point to future avenues of research to expand this study, include more recent data, and address the study limitations.

Despite these limitations, consistent results using various conservative specifications and placebo and validation tests provide promising indications of the true effect of weather conditions on transmission. Moreover, independent validation of results based on research using our projections provides strong support for the usefulness of these projections.

The equations and CRW projections in this Article were first shared in a preprint in May, 2020,³⁶ and have since been used by several researchers. At least two of those studies provided unique opportunities to independently assess our projections. In one,³⁷ the researchers included a multiplicative term, CRW_i^* , in formulating reproduction number for country i , and estimated the γ parameter based on actual infection, death, and testing data across 92 countries over the course of 2020. If γ was estimated to be 0, CRW had no predictive value, a value of 1 means current estimates are precise, and values greater than 1 point to highly informative but conservative CRW projections. The study included endogenous changes in behaviours and policies in response to risk and accounted for under-counting, treatment effects, and the age distribution of the population. With these controls in place, the authors found an estimate of $\gamma=2.64$. These results provide strong support for the usefulness of CRW factors long after the estimates were created based only on year-long means (rather than actual values) for weather inputs. A similar study using CRW values in the context of country-level data in the USA provided the estimate of $\gamma=3.88$.³⁸ Overall, independent research provided strong

support for the CRW factors, but also points to these projections being conservative. Consistent with our theoretical arguments and synthetic data analysis reported in the appendix (p 19), our study could offer a lower bound for the true effect of weather on SARS-CoV-2 transmission.

The estimated effects suggest year-long changes in transmission rates due to weather might partly explain some waves of the pandemic across different nations and substantially moderate the risk profile over the coming year. Nevertheless, as shown in 2020, the estimated effect of weather on transmission risk is not large enough in most places to quell the epidemic alone. In fact, much of the variation in reproduction number in our sample is explained by location-specific fixed effects and trends (which include adherence to public health recommendations such as physical distancing and workplace closure), not weather. Ultimately, weather and air pollution are more likely to play a secondary role in the control of the pandemic. Policy makers and the public should remain vigilant in their responses to the pandemic, adhere to non-pharmaceutical interventions, continue to ramp up vaccination, and capitalise on the synergy between weather-induced reductions in transmission in summer, and vaccination to control the pandemic to avoid new large waves in autumn.

Contributors

HR, NG, and MSJ designed the study. RX and HR designed the methods. MG, CD, HA, and MSJ collected the data. RX analysed the empirical data. NG, MG, and HR did the simulations. CD and MSJ created online visualisations. RX, MG, and CD verified the data. All authors wrote the paper. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

All data, analysis processes, and codes are available in the appendix.

Acknowledgments

We thank Richard Larson, John Sterman, Goodarz Danaei, Neil Rubin, Mark Shumate, Pari Pandharipande, and Saeid Shahraz who provided feedback and suggestions on earlier versions of this manuscript. We thank Cynthia Dong for insights on weather data collection, Peiyi Li and Hongjin Xu who assisted us in collecting and evaluating data from China, and Maedeh Moghadam who helped us with assembling data from Iran. HA is supported by Novo Nordisk Foundation Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging (NNF17OC0027812).

Editorial note: the *Lancet* Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

References

- National Academies of Sciences Engineering and Medicine. Rapid expert consultation on SARS-CoV-2 survival in relation to temperature and humidity and potential for seasonality for the COVID-19 pandemic. 2020. <https://www.nap.edu/catalog/25771/rapid-expert-consultation-on-sars-cov-2-survival-in-relation-to-temperature-and-humidity-and-potential-for-seasonality-for-the-covid-19-pandemic-april-7-2020> (accessed Sept 17, 2021).
- Shaman J, Goldstein E, Lipsitch M. Absolute humidity and pandemic versus epidemic influenza. *Am J Epidemiol* 2011; **173**: 127–35.
- van Doremalen N, Bushmaker T, Munster VJ. Stability of Middle East respiratory syndrome coronavirus (MERS-CoV) under different environmental conditions. *Euro Surveill* 2013; **18**: 20590.
- Yuan J, Yun H, Lan W, et al. A climatologic investigation of the SARS-CoV outbreak in Beijing, China. *Am J Infect Control* 2006; **34**: 234–36.
- Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD. Effects of air temperature and relative humidity on coronavirus survival on surfaces. *Appl Environ Microbiol* 2010; **76**: 2712–17.
- Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (COVID-19). *JAMA Network Open* 2020; **3**: e2011834-e.
- Chin AWH, Chu JTS, Perera MRA, et al. Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe* 2020; **1**: e10.
- 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**: 1564–67.
- Glencross DA, Ho TR, Camiña N, Hawrylowicz CM, Pfeffer PE. Air pollution and its effects on the immune system. *Free Radic Biol Med* 2020; **151**: 56–68.
- Barakat T, Muylkens B, Su B-L. Is particulate matter of air pollution a vector of Covid-19 pandemic? *Matter* 2020; **3**: 977–80.
- Bu J, Peng D-D, Xiao H, et al. Analysis of meteorological conditions and prediction of epidemic trend of 2019-nCoV infection in 2020. *medRxiv* 2020; published online Feb 18. <https://doi.org/10.1101/2020.02.13.20022715> (preprint).
- Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; **382**: 1199–207.
- Poirier C, Luo W, Majumder MS, et al. The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales. *Sci Rep* 2020; **10**: 17002.
- Oliveiros B, Caramelo L, Ferreira NC, Caramelo F. Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *medRxiv* 2020; published online March 8. <https://doi.org/10.1101/2020.03.05.20031872> (preprint).
- Notari A. Temperature dependence of COVID-19 transmission. *Sci Total Environ* 2021; **763**: 144390.
- Merow C, Urban MC. Seasonality and uncertainty in global COVID-19 growth rates. *Proc Natl Acad Sci USA* 2020; **117**: 27456–64.
- Ficetola GF, Rubolini D. Containment measures limit environmental effects on COVID-19 early outbreak dynamics. *Sci Total Environ* 2021; **761**: 144432.
- Qi H, Xiao S, Shi R, et al. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. *Sci Total Environ* 2020; **728**: 138778.
- Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* 2020; **369**: 315–19.
- Lipsitch M. Seasonality of SARS-CoV-2: will COVID-19 go away on its own in warmer weather? Center for communicable disease dynamics. 2020 <https://ccddhsph.harvard.edu/will-covid-19-go-away-on-its-own-in-warmer-weather/> (accessed Sept 15, 2021).
- Briz-Redón Á, Serrano-Aroca Á. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. *Sci Total Environ* 2020; **728**: 138811.
- O'Reilly KM, Auzenbergs M, Jafari Y, Liu Y, Flasche S, Lowe R. Effective transmission across the globe: the role of climate in COVID-19 mitigation strategies. *Lancet Planet Health* 2020; **4**: e172.
- Shi P, Dong Y, Yan H, et al. The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak - evidence from China. *medRxiv* 2020; published online March 24. <https://doi.org/10.1101/2020.03.22.20038919> (preprint).
- Guo X-J, Zhang H, Zeng Y-P. Transmissibility of COVID-19 in 11 major cities in China and its association with temperature and humidity in Beijing, Shanghai, Guangzhou, and Chengdu. *Infect Dis Poverty* 2020; **9**: 87.
- Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020; **172**: 577–82.
- Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med* 2020; **9**: e538.

- 27 Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19—studies needed. *N Engl J Med* 2020; **382**: 1194–96.
- 28 Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate statistics on COVID-19 are essential for policy guidance and decisions. *Am J Public Health* 2020; **110**: 949–51.
- 29 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; **20**: 533–34.
- 30 Gostic KM, McGough L, Baskerville EB, et al. Practical considerations for measuring the effective reproductive number, R_t . *PLoS Comput Biol* 2020; **16**: e1008409.
- 31 Struben J. The coronavirus disease (COVID-19) pandemic: simulation-based assessment of outbreak responses and postpeak strategies. *Syst Dyn Rev* 2020; **36**: 247–93.
- 32 Ghaffarzadegan N, Rahmandad H. Simulation-based estimation of the early spread of COVID-19 in Iran: actual versus confirmed cases. *Syst Dyn Rev* 2020; **36**: 101–29.
- 33 Coelho MTP, Rodrigues JFM, Medina AM, et al. Global expansion of COVID-19 pandemic is driven by population size and airport connections. *PeerJ* 2020; **8**: e9708.
- 34 Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. *Int J Infect Dis* 2020; **93**: 201–04.
- 35 Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* 2020; eabb3221.
- 36 Xu R, Rahmandad H, Gupta M, et al. Weather conditions and COVID-19 transmission: estimates and projections. *medRxiv* 2020; published online May 11. <https://doi.org/10.1101/2020.05.05.20092627> (preprint)
- 37 Rahmandad H, Lim TY, Sterman J. Behavioral dynamics of COVID-19: estimating underreporting, multiple waves, and adherence fatigue across 92 nations. *Syst Dyn Rev* 2021; **37**: 5–31.
- 38 Rahmandad H. Behavioral responses to risk promote vaccinating high-contact individuals first. *SSRN* 2021; published online Feb 8. <https://doi.org/10.1101/2021.02.05.21251215> (preprint).