



Causal inference concepts can guide research into the effects of climate on infectious diseases

Received: 12 February 2024

Accepted: 31 October 2024

Published online: 25 November 2024

Check for updates

Laura Andrea Barrero Guevara^{1,2}, Sarah C. Kramer¹, Tobias Kurth¹ & Matthieu Domenech de Cellès¹

A pressing question resulting from global warming is how climate change will affect infectious diseases. Answering this question requires research into the effects of weather on the population dynamics of transmission and infection; elucidating these effects, however, has proved difficult due to the challenges of assessing causality from the predominantly observational data available in epidemiological research. Here we show how concepts from causal inference—the sub-field of statistics aiming at inferring causality from data—can guide that research. Through a series of case studies, we illustrate how such concepts can help assess study design and strategically choose a study's location, evaluate and reduce the risk of bias, and interpret the multifaceted effects of meteorological variables on transmission. More broadly, we argue that interdisciplinary approaches based on explicit causal frameworks are crucial for reliably estimating the effect of weather and accurately predicting the consequences of climate change.

A key question ensuing from global warming is how climate change may impact the population dynamics of infectious diseases^{1–3}. Indeed, observations of large climatic variability in the distribution and seasonality of multiple infectious diseases worldwide—including major causes of death like malaria⁴, cholera⁵ and influenza⁶—suggest that many pathogens are sensitive to environmental conditions such that climate change could modify their ecology and epidemiology. Accordingly, predictive studies, based on numerical simulations combining models of global climate and infectious diseases under different scenarios of greenhouse gas emissions, suggest that climate change will affect many infections. These include infections with indirect transmission through intermediate, climate-sensitive stages involving a vector, such as mosquito-borne diseases like malaria and dengue, or the environment, including water-borne diseases like cholera. All of these infections are predicted to shift their geographical range under continued global warming^{7–9}. Fewer studies have focused on directly transmitted pathogens, but it has been suggested that climate change could also alter the transmission dynamics of respiratory syncytial

viruses in the United States and Mexico¹⁰ and varicella zoster viruses in Mexico¹¹. Although such predictions cannot yet be evaluated, earlier research has already documented the impact of past climate warming³, for example, the increased altitudinal range of malaria in the highlands of Ethiopia and Colombia¹² and the increased risk of *Vibrio* disease in Northern Europe, coinciding with the warming of the Baltic Sea's surface¹³.

A prerequisite to predicting the long-term consequences of climate change is to elucidate the effect of weather on infectious diseases. Even though effects of weather on infection dynamics (and the resulting ‘calendar of epidemics’¹⁴) have long been observed, there are persisting uncertainties about the direct causes and mechanisms for even well-researched pathogens like influenza viruses^{15,16}. Perhaps the most robust evidence for this effect is afforded by experimental studies, which demonstrate that environmental variables like temperature and humidity tightly modulate transmission parameters (such as pathogen survival time or infectivity) of viral^{17,18}, bacterial^{19,20} and parasitic^{21,22} infections. Although such evidence is useful for population-based

¹Max Planck Institute for Infection Biology, Infectious Disease Epidemiology Group, Campus Charité Mitte, Berlin, Germany. ²Institute of Public Health, Charité–Universitätsmedizin Berlin, Berlin, Germany. e-mail: domenech@mpiib-berlin.mpg.de

research (in particular for postulating causal environmental factors), it remains too limited to estimate the causal impact of weather at the scale of human populations for at least three reasons. First, the endpoints measured in experimental studies—such as pathogen survival time—can be challenging to translate into meaningful epidemiological quantities, such as transmissibility. Second, because of differences in infection biology between species, the results from animal studies may not generalize to humans²³. Third, experimental studies cannot recapitulate all the mechanisms whereby weather affects infection, especially those operating at the population level—for example, weather causes behavioural changes in people, resulting in seasonal changes in social contacts²⁴. Hence, observational studies remain necessary to estimate the multifaceted effects of weather on human infectious diseases.

However, a well-known shortcoming of observational studies is their tendency to misidentify causes because found associations do not always imply causation for observational data. This problem is also expected when inferring the effect of weather, which is characterized by meteorological variables generally highly correlated with one another and potentially many other seasonal causes of infectious diseases. Here we discuss and demonstrate how causal inference—a methodological framework aiming at inferring causes from observed data—offers a principled approach to tackle these issues and strengthen evidence in observational research^{25,26}.

Causal inference in climate–infectious disease research

Causal inference frameworks and their tools are increasingly used to analyse data and guide study design in epidemiology²⁷ and beyond^{28,29}, and may also be useful for experimental research²⁹. The impact of such tools is illustrated by the fact that causal frameworks like target emulation trials may provide evidence as robust as that from randomized trials³⁰, thereby expanding the scope of observational research for answering causal questions.

Despite these advances, the use of causal methods—in the form of mechanistic models or statistical models based on causal reasoning—remains limited in the field of weather–or climate–infectious disease research³¹. To assess how regularly causal methods are used in these fields, we re-analysed 33 studies previously assessed in a review^{32–64}. All of these studies used time series regression models to evaluate the association between weather and cases of dengue, influenza, cholera or malaria cases (Supplementary Table 1). Although more causally principled methods for time series analysis exist^{65,66}, the standard time series design remains widely used, as evidenced by numerous recent applications for SARS-CoV-2⁶⁷. Four of the 33 studies had an explicitly predictive objective and did not address causality^{38,41,43,58}. Of the remaining 29 studies that addressed causality, only one derived the statistical model from explicit causal assumptions⁴⁰. By contrast, the other 28 studies neither explicitly mentioned causal reasoning to formulate their research question nor used causal graphs for study design or statistical analysis. Based on this assessment, we conclude that applying causal inference methods may help strengthen the evidence in this field.

Here we aim to demonstrate how using a causal inference framework can improve research on the effects of weather and climate on infectious diseases. Our focus is on the conceptual aspects, while we recommend other studies for methodological details^{65,68,69}. Throughout our analysis, we illustrate the different causal inference concepts through a series of short case studies ('vignettes'), all based on a causal inference framework described below.

Results

A causal inference framework for a model of infectious disease transmission

A causal inference framework can be broadly defined as a systematic approach to identifying the causal effects of an input variable (for

example, a weather variable) on an outcome variable of interest (for example, infectious disease risk)^{70,71}. In this framework, a central first step is to explicitly represent cause-and-effect relationships using a causal diagram called a directed acyclic graph (DAG; see glossary Fig. 1)⁷². The benefits of this representation are double: conceptual, as they help investigators clarify their hypotheses and assumptions and highlight potential sources of bias; and methodological, as given a DAG, causal inference theory prescribes a set of rules to determine the form of a model (if any) that can answer the causal question of interest. In a DAG, one should also carefully consider the measurement process to distinguish between observed and unobserved variables. This distinction is central for infectious diseases, as variables that are typically unobserved—such as population immunity, often portrayed as the 'dark matter' of epidemics⁷³—sensitively control transmission dynamics. We illustrate the application of a causal inference framework with a simple mathematical model representing the population-level dynamics of an acute infection spread by direct contact between susceptible and infected hosts. Although we do not focus on a specific pathogen in the rest of this analysis, this model—known as Susceptible–Infected–Recovered–Susceptible (SIRS) in the field of infectious disease modelling⁷⁴—may be considered realistic for respiratory viruses with short generation times, like influenza viruses⁷⁵, respiratory syncytial viruses⁷⁶ and SARS-CoV-2⁷⁶.

To describe the effect of weather on infection dynamics, we incorporated an environmental model representing the joint causal effect—dictated by physical laws^{77,78}—of ambient air temperature (T_e) and dew point temperature (a measure of absolute humidity) on relative humidity (RH). We then assumed a direct, negative effect of temperature and relative humidity on transmission (β)—that is, transmission decreased as either climatic variable increased. Finally, we incorporated an observation model representing the causal link between the true and observed incidence rates, assuming a surveillance system with perfectly specific but incompletely sensitive case detection, resulting in case under-reporting. Of note, we assumed this observation model was completely random, and we thus did not consider the potential bias resulting from systematic differences between the true and the observed incidence rates^{79–81}. Strategic simplifying assumptions (in particular, of a discrete-time model with fixed generation time and time step of 1 week) then allowed us to represent the full model with a simple DAG (Fig. 2). We focus on this representation instead of the conventional compartmental model diagram used in infectious disease modelling because DAGs more concisely convey causal concepts. Still, we note that both representations are causal diagrams (see ref. 82 for a more extensive discussion of the correspondence between both diagrams). Full model details—including equations, numerical implementation and further discussion of our assumptions about the effects of weather—can be found in Methods.

In the rest of this analysis, we use this DAG and its underlying model to illustrate four causal inference concepts: descendants and measurement bias (vignette 1); natural experiments (vignette 2); confounders and confounding bias (vignette 3); and mediators (vignette 4). In so doing, we emphasize two key points: first, causal inference frameworks are useful—and, indeed, required—to assess the effects of weather on infectious diseases; second, transmission models are valuable to encapsulate this framework, as they specify explicit causal mechanisms for the observed and unobserved variables that underlie infectious disease data.

Causal inference concepts—illustrations with four vignettes

Vignette 1 on descendants, measurement bias and the intricate association between environmental variables and incidence rate. Time series regression analysis of observed incidence rates is a frequent study design in environmental epidemiology^{68,83}. The implicit assumption of such studies is that statistical quantities derived from regression models—typically, regression coefficients—will accurately

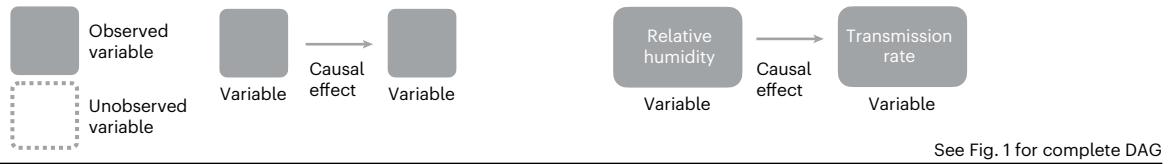
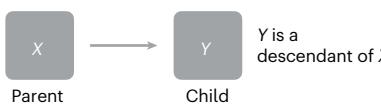
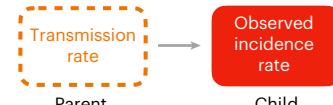
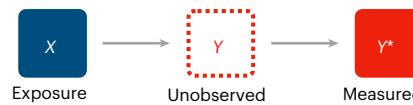
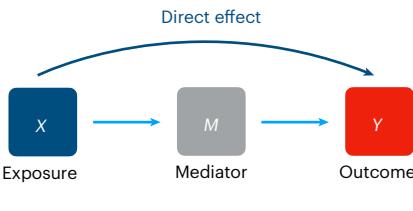
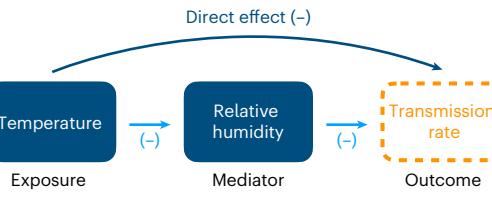
Concept	Definition	Section
DAG	A causal diagram where the nodes are variables and the edges are direct causal effects between variables. Unobserved variables are represented by dashed lines.	Vignette 1–4. Weather (here, relative humidity) can affect the transmission rate. 
Descendant	A variable Y that is directly or indirectly caused by another variable ($X \rightarrow Y$).	Vignette 1. The observed incidence rate is a descendant of the transmission rate.  
Measurement bias	A bias on the estimated effect of exposure X on outcome Y resulting from the process by which the variables are measured (X^* or Y^*).	Vignette 1. The estimated effect of weather (here, relative humidity) on the unobserved transmission rate can be biased as it is measured as the observed incidence rate.  
Confounding bias	A bias on the estimated effect of an exposure X on outcome Y when the exposure and the outcome share a common cause C ($X \leftarrow C \rightarrow Y$). The common cause is often referred to as confounder.	Vignette 2. The estimated effect of spatial spread between two locations can be biased when the incidence rate in both locations share a common cause (here, climate).  
Quasi-experiments	Leveraging randomness that occurs naturally in observed data to evaluate causal effects.	Vignette 3. Naturally occurring randomness in weather in tropical regions allows evaluating causal effects.  
Mediator	A middle variable M through which an exposure X affects indirectly an outcome Y ($X \rightarrow M \rightarrow Y$). The total effect of X on Y comprises the direct and indirect effects.	Vignette 4. Relative humidity is a mediator through which temperature affects the transmission rate. The total effect of temperature on the transmission rate comprises its direct and indirect effects.  

Fig. 1 | Glossary of causal inference concepts. Definitions of causal inference concepts with causal graphs illustrating the topic discussed in each vignette.

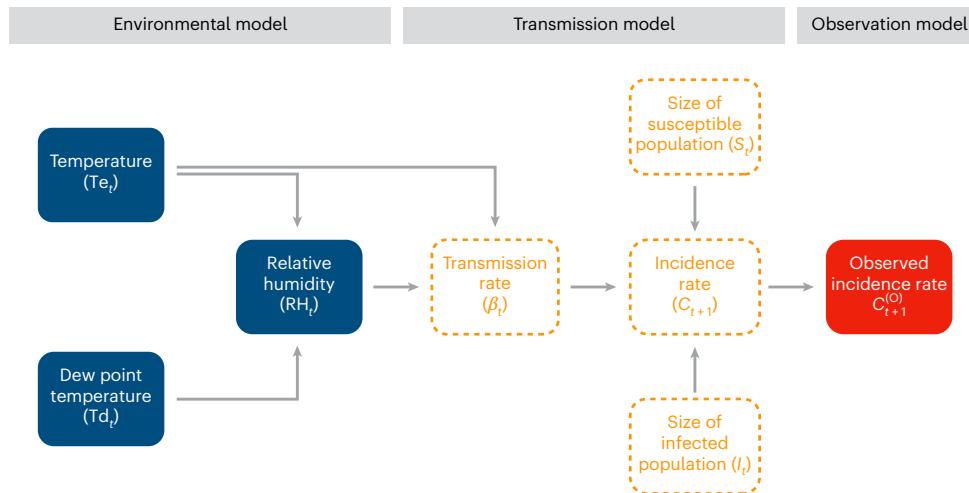


Fig. 2 | Causal graph for the illustrative transmission model. At a given time t , physical laws^{77,78} dictate that the ambient temperature and the dew point temperature jointly determine the relative humidity of the air. We assume that both temperature and relative humidity directly affect the transmission rate of the pathogen: $\log \beta_t = \log \beta + \delta_{Te}(Te'_t - 1) + \delta_{RH}(RH'_t - 1)$, where β represents the average transmission rate, Te'_t and RH'_t the rescaled environmental variables (with unit mean), and $\delta_{Te} = \delta_{RH} = -0.2$ their causal effects on transmission. The causal link between the transmission rate and the incidence rate—defined as the number of new cases per unit of time—is governed by a deterministic nonlinear model (here assumed to be a simple SIRS model⁷⁴) representing the pathogen’s

epidemiological dynamics, in which new infections arise from contact between susceptible and infected individuals. Because of observation error (for example, imperfect test sensitivity resulting in case under-reporting), the observed incidence rate differs from the actual incidence rate; a stochastic observation model (here assumed to be a negative binomial model) represents the causal link between the two rates. Mathematically, this causal diagram translates into a discrete-time model that iterates the epidemic dynamic from one generation of infection to the next (see Methods for full information about the model’s formulation and implementation). Variables surrounded by dashed lines are assumed to be unobserved.

capture the causal effect of meteorological variables. However, as shown in our causal graph (Fig. 2) and as expected in practice, the causal chain relating weather to observed incidence rates is indirect and complex. A key issue arises from the fact that, although weather directly affects the transmission rate, the observed incidence rate is located two causal links downstream from the transmission rate; in causal inference language, the latter rate is described as a descendant (that is, a consequence; Fig. 1) of the former. Of these two causal links, the one relating the transmission and incidence rates may be challenging to recapitulate with regression models because unobserved variables like the size of the population susceptible to infection (inversely related to the population, or herd, immunity, which controls epidemic thresholds) induce nonlinearities that may result in marked dissimilarities between these two rates^{69,74}. Hence, measurement bias—that is, the bias arising from non-random differences between the targeted (here, transmission) and the observed endpoints (here, observed incidence; Fig. 1)⁸⁴—may distort causal inference from time series regression models. This potential bias has been recognized in environmental epidemiology, as reflected in recommendations to include additional covariates for capturing temporal variations in population immunity or other long-term trends⁶⁸. However, because of the above complexities, such additions, depending on the underlying causal structure and available information, are not guaranteed to reduce measurement bias.

To illustrate, we generated model simulations for a pathogen with low, medium or high transmissibility (basic reproduction number of 1.25, 2.5 or 5, respectively), with meteorological data from a temperate climate (Lübeck, Germany; Supplementary Table 2) resulting in a seasonally forced transmission rate with a single peak every winter (Fig. 3a). Under the medium-transmissibility scenario (Fig. 3b, middle panel), the epidemiological dynamic displayed annual periodicity, with winter seasonality in the incidence rate that broadly matched that of the transmission rate (Spearman’s correlation coefficient: $r_s = 0.63$). In marked contrast, lower transmissibility resulted in biennial epidemics showing little correlation with the seasonal transmission rate ($r_s = 0.07$; Fig. 3b, top panel). This phenomenon—called sub-harmonic resonance⁷⁴—resulted from the higher susceptibility threshold needed

to trigger epidemics and the longer time required to replenish the pool of susceptible individuals (via births and waning immunity) to exceed that threshold. Finally, the opposite phenomenon of super-harmonic resonance was observed in the high-transmissibility scenario, which resulted in biannual epidemics (Fig. 3b, bottom panel). These simple numerical experiments illustrate the complex dynamic of infectious diseases and the potent but sometimes counter-intuitive footprint that weather—or, for that matter, any other source of seasonal forcing—can have on this dynamic⁶⁹.

Next, we generated 100 replicate time series of observed incidence rates for each scenario to assess the reliability of time series regression models. For every replicate, as a control, we first fitted a negative binomial generalized additive model (GAM) with 1-week-lagged weather variables and the susceptible and infected population sizes as covariates and the observed incidence rate as endpoint—the true candidate model for our application, as we show in the Methods. As expected, the causal effects of temperature and relative humidity on the transmission rate were estimated, on average, without bias for this model (Supplementary Fig. 1). In practical applications, however, the susceptible and infected population sizes would be unobserved. Therefore, we next fitted a comparable model with a flexible smooth of time to try to capture variations in these unobserved variables. As shown in Fig. 3c, because of measurement bias, estimation performance was, overall, poor. For low transmissibility, the causal effect of temperature on the transmission rate was estimated with a substantial bias (mean absolute bias (AB): 0.08, 40% relative error in comparison to the actual value of -0.2) and imprecision (mean standard error of estimates across simulations (SE): 0.07). The bias was even more substantial in the medium- (mean AB: 0.16) and high-transmissibility scenarios (mean AB: 0.22). Because relative humidity had a high correlation with temperature but lower variability, its estimated effect was marred with large uncertainty, which exceeded, on average, the absolute effect size in every scenario (mean SE: 0.21–0.35, to be compared with the true effect size of -0.2; Supplementary Fig. 1).

In an additional analysis, we tested time series regression models of the effective reproduction number, another outcome that can be

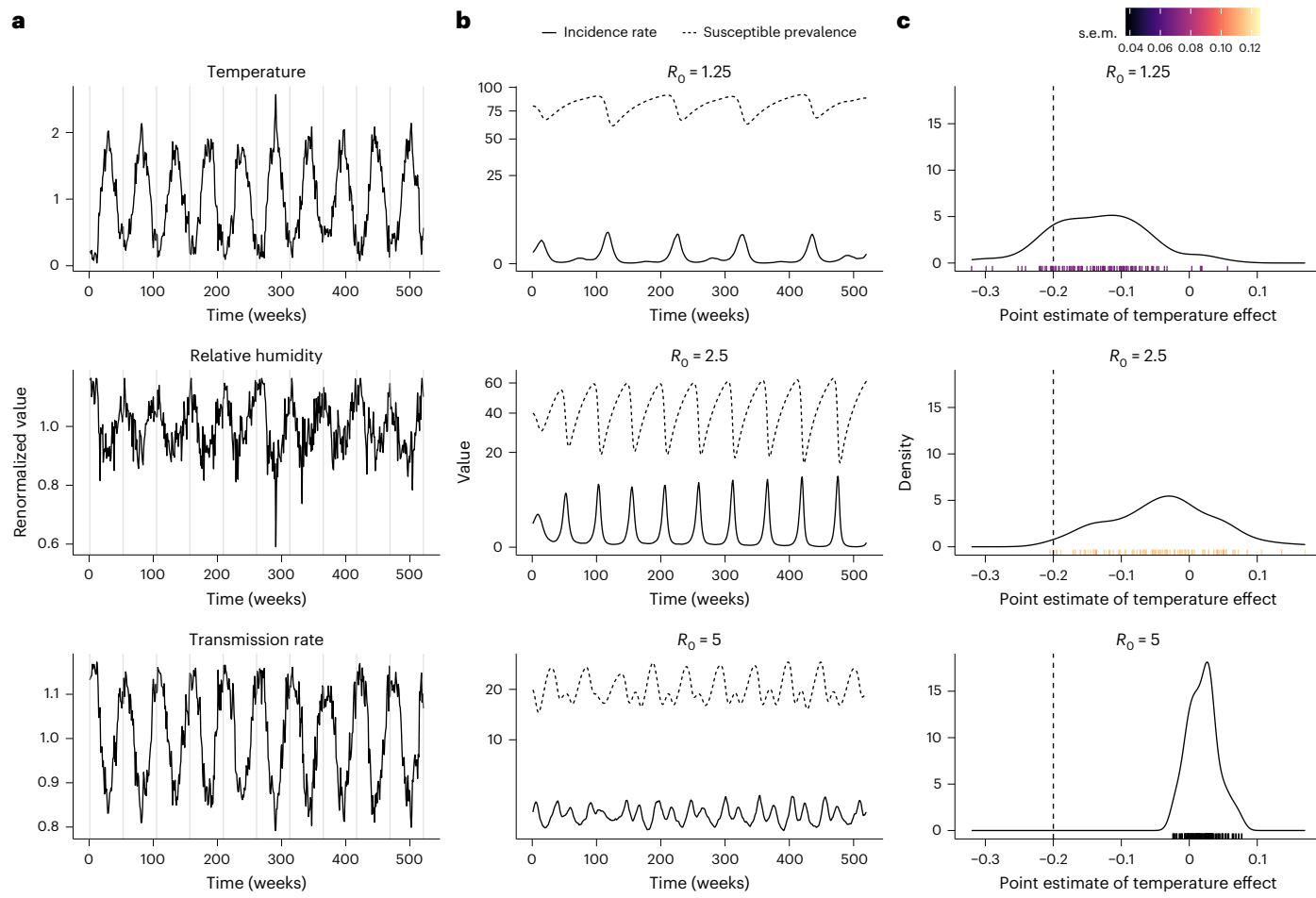


Fig. 3 | Measurement bias and the intricate association between environmental variables and incidence rate (vignette 1). **a**, Time series of temperature (top panel) and relative humidity (middle panel, both climatic variables renormalized to have unit mean) during 2013–2022 in Lübeck, Germany. The bottom panel displays the resulting seasonal component of the transmission rate, assuming a negative effect of both climatic variables on transmission. In the three panels, the vertical grey lines mark the beginning of every calendar year (week no. 1). **b**, Dynamics of susceptible prevalence ($100 \times S/N$; dashed lines)

and incidence rate ($100 \times C/N$; solid lines) for three different values of the basic reproduction number (R_0) and an average duration of immunity of 1 year (other model parameters fixed to the values in Supplementary Table 3). **c**, Distribution of point estimates for the effect of temperature estimated from a negative binomial GAM regression model fitted to each of 100 replicate time series of the observed incidence rate. The marks on the x axis indicate the point estimates, with the colours representing their standard errors. The vertical dashed line indicates the true effect fixed in the transmission model for generating the observations.

considered to assess the effects of weather. The corresponding DAG (Supplementary Fig. 2) shows that this outcome depends on only one unobserved variable, while the incidence rate depends on two (Fig. 2). As a result, we found that time series regression generally performed better for this outcome, even though the bias remained very large in all scenarios (Supplementary Fig. 1). We note that this better performance may not generalize to all settings. In practice, the effective reproduction number is not directly observable and must first be reconstructed from incidence time series⁸⁵. In our deliberately simple application, we assumed a small amount of noise but no systematic bias in this reconstruction, an assumption that may be too optimistic for more realistic models.

Although not intended to be exhaustive, this simple simulation study echoes earlier discussions of measurement bias as a major concern for study designs based on time series regression, particularly when population immunity varies over fast time scales^{68,83}. Relating to the central thread of this analysis, we note that causal reasoning—particularly the causal diagram representing the effect of weather—allowed us to identify, *a priori*, the relevant theoretical issues and propose simple numerical experiments to assess their practical relevance. This vignette thus illustrates the value of causal reasoning not only as a methodological tool but also as a theoretical tool to guide study design and analysis.

Vignette 2 on climate variability as natural experiments to estimate the individual effect of meteorological variables. Owing predominantly to latitudinal gradients in solar radiation and other factors like altitude and proximity to the sea, the Earth displays a large variability of climates⁸⁶. This variability is reflected in the Köppen–Geiger system, which classifies worldwide climates into 5 main types and 30 sub-types based on seasonal averages of precipitation and temperature⁸⁷. Because these different climates exhibit diverse seasonal patterns of variation in weather variables and correlations between them, they may be regarded as a range of ‘natural experiments,’ conceptually equivalent to manipulating specific weather variables to identify their causal effects. More broadly, the strategy of leveraging randomness that occurs naturally in observed data (that is, quasi-experiments; Fig. 1) is increasingly advocated for when inferring causality in predominantly observational research fields like economics²⁸, ecology⁸⁸ and epidemiology⁸⁹. As an example of such quasi-experiments, previous studies analysed large-scale, irregular oceanic phenomena such as the El Niño–Southern Oscillation to evaluate the effects of ‘climate change-like shocks’^{90,91}.

Of particular interest for environmental epidemiological research is the contrast between tropical climates (where temperature generally varies little and thus only slightly affects relative humidity) and temperate climates (where the opposite is typically observed), which may be

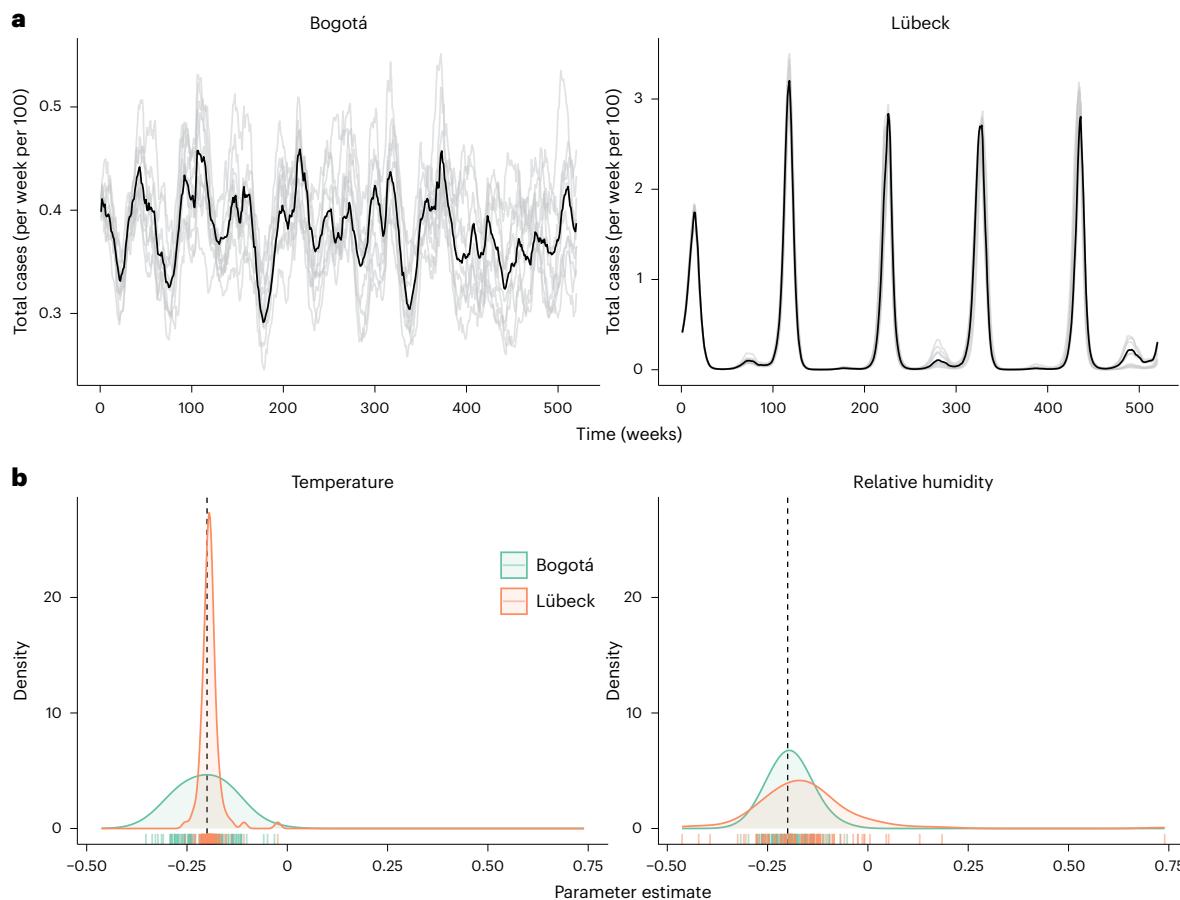


Fig. 4 | Climate variability as natural experiments to estimate the individual effect of meteorological variables (vignette 2). **a**, Simulated incidence rate (C_t) in Bogotá, Colombia (left panel) and Lübeck, Germany (right panel). Model parameters: basic reproduction number of 1.25, average duration of immunity of 1 year and other parameters as in Supplementary Table 3. In each panel, the black line shows the dynamic from the deterministic transmission model. The grey lines represent simulations (10 displayed out of 100 overall) from a stochastic transmission model, in which noise was added to the transmission rate at every time point (see Methods for full details). We implemented this stochastic variant only for this vignette to add realism in the form of model misspecification during

estimation. **b**, Distribution of the 100 maximum likelihood estimates of the effect of temperature (left panel) and relative humidity in Bogotá and Lübeck. These estimates were obtained by direct maximization of the log-likelihood (that is, trajectory matching) by fitting the misspecified model—in which the transmission model was deterministic and the observation model stochastic—to data generated from the fully stochastic model—in which both the transmission and the observation models were stochastic. The dashed vertical lines indicate the true parameter value (-0.2) fixed in all model simulations. See Methods for further details about the estimation procedure.

leveraged to isolate the effects of temperature and relative humidity. To test this hypothesis, we used our illustrative model to run a range of numerical experiments in a pair of locations where this contrast was marked: Lübeck, Germany (53.9°N latitude, coefficient of variation (CV; standard deviation/mean) of temperature $\text{CV}(\text{Te}) = 0.59$, $\text{CV}(\text{RH}) = 0.07$, $r_s(\text{Te}, \text{RH}) = -0.48$) and Bogotá, Colombia (4.7°N latitude, $\text{CV}(\text{Te}) = 0.04$, $\text{CV}(\text{RH}) = 0.07$, $r_s(\text{Te}, \text{RH}) = -0.10$). By back-fitting our transmission model to 100 replicate time series of observed incidence rates it generated, we gauged how well we could estimate the effects of temperature and relative humidity (as well as other model parameters that would be unknown in real-world applications; Methods) in the two climates. In Lübeck, as expected for a climate characterized by low RH variability and large RH–Te correlation, the effect of temperature was estimated with more accuracy than that of relative humidity (mean AB of 0.02 and 0.09, respectively; Fig. 4). Of note, these results are reminiscent of those of vignette 1, except that the parameters estimated in this vignette originated from the true causal transmission model (Fig. 2) and, therefore, did not suffer from measurement bias. In Bogotá and its climate with low Te variability and almost null RH–Te correlation, the opposite result held, with higher accuracy for relative humidity than for temperature (mean AB of 0.04 and 0.05, respectively).

This simulation study thus suggests the scope for strategic choices regarding a study's location, where the local climate's properties can help estimate the effect of the weather variable of causal interest. Whenever more data are available, an alternative strategy is to use multilevel models, which provide a principled way to pool information while modelling variation across multiple locations⁸⁴. Multilevel extensions are now routine for standard regression models but more challenging for the complex—typically nonlinear, stochastic and partially observed⁹²—models needed to capture infectious disease dynamics. Nevertheless, recent statistical advances permit the estimation of such multilevel models^{93,94}, opening an avenue for large-scale dynamical modelling studies that harness information from multiple natural experiments in different climates. More broadly, this vignette underscores the critical importance of considering the causal mechanisms underlying weather dynamics.

Vignette 3 on confounding bias and how climate variability can masquerade as spatial spread. Spatial heterogeneities are commonly observed for infectious diseases^{95–99}. Such heterogeneities can result from two broad classes of mechanisms, depending on whether they involve spatial interactions through the movement of individuals

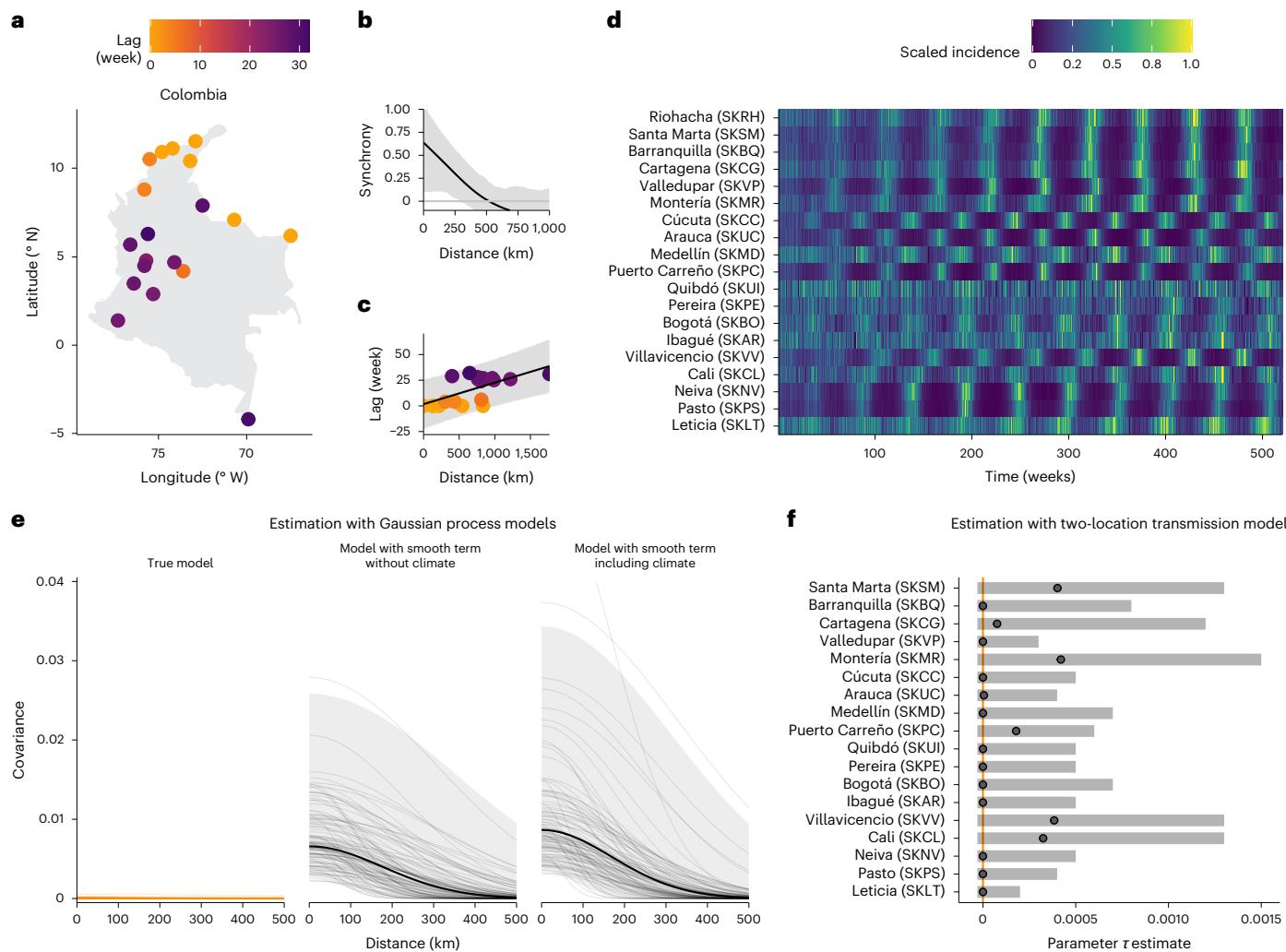


Fig. 5 | Confounding bias and how climate variability can masquerade as spatial diffusion in Colombia (vignette 3). **a,d**, Considering 19 locations across Colombia (**a**), we simulated incidence with no spatial diffusion but a common effect of climate on transmission (**d**; see Methods for full details). **b**, Pairwise epidemic synchrony between locations, with the fitted covariance function shown as a line and the 95% confidence interval shaded in grey. **c**, Relative timing of the simulated epidemic peaks across Colombia, referenced to the northernmost location (Riohacha). The colour indicates the time difference between the epidemic peaks of each site and those of the reference site. The line represents the line posterior mean from a Bayesian linear model, from which we estimated the travelling wave, and the grey envelope represents the 95% credible

interval. **e**, Estimated spatial spread from Gaussian process models, with the dark lines representing the estimated mean covariance function, light lines representing 100 draws from the posterior distribution and shaded envelopes (orange and grey) indicating the 95% credible interval. **f**, Estimated spatial spread from two-location transmission models assuming no effect of climate, where the orange line denotes the true value of spatial spread, $\tau = 0$. The points represent the maximum likelihood estimate of the spatial spread, and the grey intervals represent the 95% confidence interval. Model parameters: basic reproduction number of 2.5, average duration of immunity of 2 years and other parameters as in Supplementary Table 3.

(that is, spatial spread) or spatial variation in some other variable (for example, climate)¹⁰⁰. For example, spatial variations in climate across multiple locations may result in spatial covariance between these locations, even in the absence of spatial spread. This common cause of spatial variability may thus result in spurious associations that confound the estimated effect of spatial spread on observed incidence—that is, confounding bias (Fig. 1; see also Supplementary Fig. 3 for a DAG illustrating the problem in two locations).

To illustrate, we considered a scenario with seasonal transmission forced by weather but no spatial spread in two distinct countries (Supplementary Table 2): one with a definite latitudinal gradient in climate (Colombia¹⁰¹) and another with little spatial variability in climate (Spain). We simulated the resulting dynamics of our transmission model and assessed epidemic synchrony¹⁰² across various locations in these two countries. In Colombia, the simulated incidence displayed diverse seasonal patterns that followed a latitudinal gradient broadly

matching that of the climate (Fig. 5b). In contrast, the low climatic variability in Spain resulted in tightly synchronous epidemics across the locations (Supplementary Fig. 4b). Hence, despite the absence of mechanisms causing spatial spread in our model, the shared effect of climate between locations resulted in marked spatial correlation in observed incidence, up to ~250 km in Colombia and more extensively throughout Spain (Fig. 5d and Supplementary Fig. 4d).

To further characterize this spatial correlation, we estimated the speed of the—spurious—travelling wave under the incorrect assumption of spatial spread being the sole cause of spatial heterogeneity⁹⁶. Speed estimates were near infinite in Spain (Supplementary Fig. 4a,c), suggesting either confounding by climate or extremely strong coupling between the locations⁷⁴. In Colombia, however, the speed was estimated at 218 km per month (95% credible interval (CI): 121–444 km per month; Fig. 5a,c), a value consistent with that documented for real travelling waves—for example, 110–320 km per month for pertussis

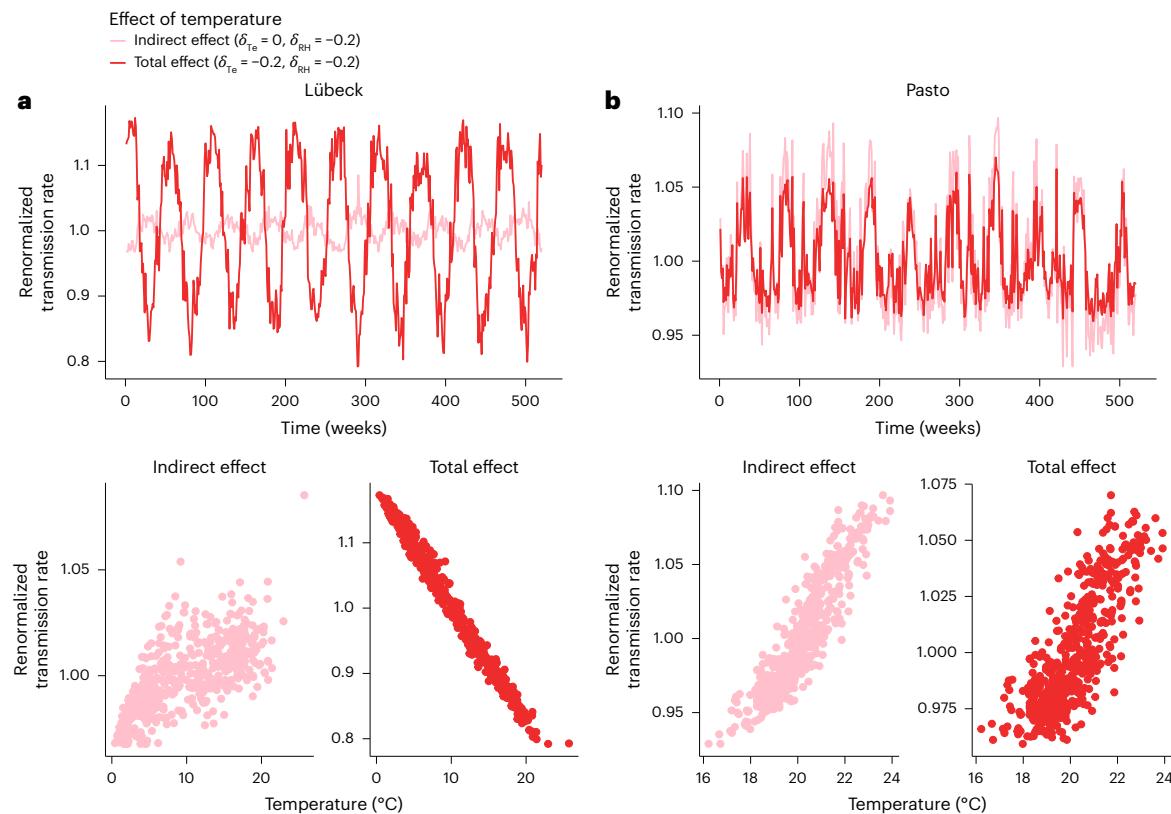


Fig. 6 | Mediation and the direct and indirect causal effects of temperature on transmission (vignette 4). **a,b,** Simulated transmission rate (top panels) in Lübeck, Germany (**a**) and Pasto, Colombia (**b**), from models including the total effect of temperature (dark lines, where the direct effect of temperature, δ_{Te} , and the direct effect of RH, δ_{RH} , were -0.2) or only the indirect effect of temperature mediated through humidity (light lines, where $\delta_{Te} = 0$ and $\delta_{RH} = -0.2$; see

Methods for full details). The bottom panels show the indirect effect of temperature through relative humidity on transmission (left panels) and the total effect of temperature (right panels). Model parameters: basic reproduction number of 1.25, average duration of immunity of 1 year and other parameters as in Supplementary Table 3.

during 1951–2010 in the United States⁹⁸ and 150 km per month for dengue during 1983–1997 in Thailand¹⁰³.

To assess whether a statistical approach can address this confounding bias, we fitted a series of Gaussian process models to estimate the covariance in observed incidence rates between all locations in each country (Methods). As a control, we first tested the true model implied by the DAG, which included the sizes of the susceptible and infected populations, in addition to 1-week-lagged weather variables as covariates. As expected, this model correctly identified the absence of spatial spread, with the maximum covariance between locations being negligible (η (95% CI) = 0.01 (0–0.02) in Colombia and 0 (0–0.02) in Spain; Fig. 5e and Supplementary Fig. 4e left panel). Next, we fitted models with smooths of time to try to capture the variations in the susceptible and infected population sizes, which would be unobserved in practice. We found that a model omitting weather variables (that is, the confounded model) estimated a spurious spatial covariance between the locations (η (95% CI) = 0.09 (0.05–0.16) in Colombia and 0.38 (0.22–0.70) in Spain; Fig. 5e and Supplementary Fig. 4e middle panel). Furthermore, a model including the weather variables also led to biased estimates of spatial covariance (η (95% CI) = 0.11 (0.06–0.18) in Colombia and 0.39 (0.23–0.74) in Spain; Fig. 5e and Supplementary Fig. 4e right panel). This finding thus re-emphasizes the difficulty of analysing incidence data because of measurement bias (vignette 1) and suggests the need for explicit models for capturing the unobserved variables.

Next, we thus designed a simple transmission model with climate forcing and spatial spread (described by a single coupling parameter τ) between two locations. Using the data generated from the model with no spatial spread (Fig. 5d), we estimated the value of the coupling

parameter (Methods). Here, the model including the weather variables—and thus controlling for the shared weather between the two locations—correctly revealed the absence of spatial spread (Fig. 5f and Supplementary Fig. 4f, with all confidence intervals of the maximum likelihood estimate for τ including 0). Although used for illustration here, this deliberately simple model could be extended to include more realistic features of spatial spread, such as multiple locations with different population sizes, potentially resulting in source–sink dynamics between urban and rural areas⁹⁵. More generally, this vignette highlights the importance of integrating explicit causal models with transmission models to disentangle the mechanisms underlying spatial heterogeneities.

Vignette 4 on mediation and the direct and indirect causal effects of temperature on transmission. Weather is the consequence of a complex web of causally related environmental variables that can affect pathogen transmission through multiple pathways⁶⁵. For instance, as illustrated in our simple causal graph and supported by experimental evidence¹⁸, both temperature and relative humidity may directly affect transmission. However, because temperature also impacts relative humidity, its total causal effect may comprise a direct effect ($Te \rightarrow \beta$) and an indirect effect mediated by relative humidity ($Te \rightarrow RH \rightarrow \beta$). Importantly, but counter-intuitively, the direct and indirect effects may act in opposite directions depending on the causal relationship between the parent variable (for example, temperature) and its mediator (for example, relative humidity; Fig. 1) in the environmental model.

To illustrate how these two effects play out in different settings, we simulated our model in Lübeck, Germany, and Pasto, another

Colombian city with relative humidity variability higher than in Bogotá. Owing to climatic differences causing temperature to be more variable than relative humidity in Lübeck ($CV(Te) = 0.59$, $CV(RH) = 0.07$) but less variable in Pasto ($1.4^\circ N$ latitude, $CV(Te) = 0.07$, $CV(RH) = 0.19$), we hypothesized the total impact of temperature would differ between the two cities. In each city, we considered two scenarios. In the first scenario, we set the two climatic parameters (see legend of Fig. 2) to their baseline values to capture the total effect of temperature—that is, the direct effect and the indirect effect mediated by relative humidity. In the second scenario, we set the climatic parameter of temperature to 0 to capture only its indirect effect, mediated by relative humidity.

In both cities, because of a combination of two negative effects (of temperature on relative humidity and relative humidity on transmission), higher temperature increased transmission through the indirect pathway ($r_s(Te, \beta) = 0.91$ in Pasto and 0.72 in Lübeck; Fig. 6a,b). In Lübeck, however, the higher temperature variability caused the direct effect to outweigh this indirect effect so that, overall, transmission decreased as temperature increased ($r_s(Te, \beta) = -0.99$; Fig. 6a). By contrast, the higher variability in relative humidity reversed the total effect in Pasto, where transmission increased with temperature ($r_s(Te, \beta) = 0.80$; Fig. 6b). Despite this overall positive effect, adjusting for relative humidity revealed the negative direct effect of temperature in Pasto (partial Spearman's rank correlation coefficient: $r_s(Te, \beta | RH) = -0.56$; Fig. 6b). Of note, echoing the results of vignette 1, the effects of temperature on the observed incidence rates were less definite because of measurement bias (Supplementary Fig. 5). Hence, despite identical causal mechanisms, climatic differences resulted in divergent effects of temperature in the two cities.

These conceptual insights have practical implications for interpreting the association between environment and transmission rates. Specifically, when evaluating the effect of temperature in a DAG similar to that in Fig. 2, models adjusting for relative humidity would identify the direct effect of both variables. In contrast, models without adjustment would only identify the total effect of temperature. The lack of clear causal frameworks may thus lead to misinterpreting model outputs, a risk described by earlier research as the ‘table 2 fallacy’¹⁰⁴. Hence, this vignette re-emphasizes the critical importance of causal reasoning and careful interpretation when probing the effect of climate on infection dynamics.

Discussion

Here we aimed to show how causal inference concepts—such as descendants and mediators, confounding and measurement biases, and quasi-experiments—can guide research into the effects of climate on infectious diseases. Through a series of case studies, we illustrated how such concepts could help assess study design (vignette 1); strategically choose a study’s location to achieve the set-up of a natural experiment (vignette 2); evaluate the risk of confounding bias (vignette 3); and interpret the direct and—sometimes paradoxical—indirect effects of meteorological variables on transmission (vignette 4). In addition, we showed that transmission models offer a principled and parsimonious tool to capture infectious disease dynamics and encapsulate causal frameworks. More broadly, seconding earlier calls in the epidemiological field²⁷, we argue that such frameworks are necessary for inferring the effect of weather and subsequently predicting the consequences of climate change on infectious diseases.

Because of this study’s conceptual focus, we sidestepped the many methodological technicalities that inevitably arise in practice. In addition to mere data problems (for example, a mismatch between the time scales of observed data and infection dynamics), the effect of weather on infectious diseases can be more intricate than our simple model suggests. First, variables other than temperature and relative humidity may directly affect transmission, such that disentangling their direct and indirect effects may be more challenging than in vignette 4. For example, if one assumes a direct effect of dew point temperature on

transmission in the DAG represented in Fig. 2, then a causal mediation analysis would be required to estimate these different effects for all the environmental variables. Second, because of interindividual variability in the period separating exposure from infectiousness (that is, the latent period), the effect of weather on transmission is expected to be lag-distributed, resulting in a more complex causal diagram than Fig. 2. In this case, a standard compartmental model diagram may be more suited to depict the causal processes, with the added benefit that such diagrams can explicitly represent interactions and interference⁸². Third, this effect may be non-continuous and non-monotonic, as illustrated by recent experimental evidence showing a V-shape threshold association between relative humidity and survival time of coronaviruses^{18,105}. Fourth, although the causal graph of Fig. 2 is realistic for pathogens causing acute, directly transmitted infections (such as respiratory viruses), the weather may have multiple effects on pathogens with more complex relationships with their human hosts. Such multiple effects are, for example, expected for invasive bacteria with prolonged carrier states (like the pneumococcus¹⁰⁶), with weather affecting not only transmission of carriage but also progression from carriage to invasive disease^{107,108}. Finally, the poor correlation between indoor and outdoor meteorological variables (especially observed for temperature and relative humidity¹⁰⁹) has led to discussions about which measure of weather is more appropriate for causal inference, with indoor data argued to represent the bulk of weather exposures^{16,110}. This problem may be viewed as another form of measurement bias and treated in a causal framework by modelling the causal link between indoor variables and their outdoor counterparts, with some recent research in this direction¹¹⁰. More generally, our simple causal framework could be similarly extended to tackle the other complexities listed above, as well as the potential biases illustrated in the vignettes simultaneously.

In conclusion, the expanding field of causal inference offers opportunities to strengthen evidence derived from observed data. This study thus presents an early effort to integrate this field with infectious disease epidemiology and climatology, with the ultimate research aims of elucidating how climate affects pathogens and predicting the consequences of climate change. Given that climate is an ever-present component of the environment, this research will also advance our understanding of the ecology of infectious diseases.

Methods

Model formulation

Meteorological model. The daily average records of dew point temperature (Td , expressed in $^\circ C$) and ambient temperature (Te , expressed in $^\circ C$) were extracted using the WeatherData function in Mathematica¹¹¹. The data covered the period 2013–2022 in multiple weather stations located near the major cities of Colombia, Spain and Germany (Supplementary Table 1). Because the transmission model had a time step of 1 week, we calculated weekly averages of Te and Td for inclusion as covariates in the model. In case of missing daily records for a given week, we calculated the weekly average based on the records observed within that week in other years.

The relative humidity (RH, defined as the actual amount of water moisture in the air compared with the total amount that the air can hold at a given temperature) was then calculated using the formula^{77,78}:

$$\log RH = \frac{\alpha Td}{\lambda + Td} - \frac{\alpha Te}{\lambda + Te}$$

where $\alpha = 17.625$ and $\lambda = 243.04$ $^\circ C$ are the revised Magnus coefficients¹¹². This association can be understood intuitively as follows: relative humidity increases as the absolute moisture in the air (quantified by Td) increases, while it decreases as ambient temperature increases (because of the physical property that the maximum moisture air can hold increases exponentially with temperature). To verify the adequacy

of this formula, we also extracted actual RH records from the weather stations: the agreement between predictions and measurements was excellent (>90% correlation in all the locations considered), even though the predicted RH was overestimated at low ambient temperatures in temperate climates, such as in Germany.

In the following, we denote by Te_t and RH_t the weekly time series of ambient temperature and relative humidity, $\bar{\text{Te}}$ and $\bar{\text{RH}}$ their temporal averages (over the entire time series), and $\text{Te}'_t = \frac{\text{Te}_t}{\bar{\text{Te}}}$ and $\text{RH}'_t = \frac{\text{RH}_t}{\bar{\text{RH}}}$ their renormalized values.

Transmission model. To illustrate the different causal inference concepts, we formulated a discrete-time SIRS⁷⁴ model with the transmission rate β_t forced by the two climatic variables:

$$\log \beta_t = \log \beta + \delta_{\text{Te}}(\text{Te}'_t - 1) + \delta_{\text{RH}}(\text{RH}'_t - 1)$$

where β is the average transmission rate, δ_{Te} the effect of ambient temperature and δ_{RH} the effect of relative humidity. Based on experimental evidence on respiratory viruses^{18,113}, we assumed a small negative effect of both climatic variables: $\delta_{\text{Te}} = \delta_{\text{RH}} = -0.2$. In other words, we assumed that transmission decreased as either climatic variable increased.

To derive the equations of the discrete-time model, we first write the system of ordinary differential equations for the continuous-time model:

$$\frac{dS}{dt} = \mu N + \alpha R - (\lambda(t) + \mu)S$$

$$\frac{dI}{dt} = \lambda(t)S - (\gamma + \mu)I$$

$$\frac{dR}{dt} = \gamma I - (\alpha + \mu)R$$

where $\lambda(t) = \beta(t)I(t)/N$ is the force of infection, N is the population size, μ is the birth/death rate, γ^{-1} is the generation time, and α^{-1} is the average duration of protection. All parameters are listed in Supplementary Table 3. Assuming a fixed time step $\Delta t = 1$ week and a fixed generation time equal to this time step ($\gamma^{-1} = \Delta t = 1$ week), we discretized the system of ordinary differential equations (using the approximation $\frac{dx}{dt} \approx \frac{\tilde{x}_{t+\Delta t} - x_t}{\Delta t}$) to get the equations of the discrete-time model:

$$S_{t+1} = S_t + \mu N + \alpha R_t - (\lambda_t + \mu)S_t$$

$$I_{t+1} = \lambda_t S_t - \mu I_t$$

$$R_{t+1} = R_t + I_t - (\alpha + \mu)R_t$$

For all simulations, we initialized the state variables to their equilibrium values for the model with no seasonal forcing ($\delta_{\text{Te}} = \delta_{\text{RH}} = 0$). For the discrete-time model, these equilibria are given by $S^* = \frac{N}{R_0}$, $I^* = \frac{\alpha+\mu}{\alpha+\mu+1}(N - S^*)$ and $R^* = N - S^* - I^*$, where $R_0 = \frac{\beta}{\mu+1}$ is the basic reproduction number.

Observation model. To complete the model formulation, we specified a stochastic observation model to generate observed data from the transmission model's outputs. Let

$$C_t = \lambda_{t-1} S_{t-1} = \beta_{t-1} \frac{I_{t-1}}{N} S_{t-1}$$

represent the (true) incidence rate (Fig. 2), defined as the weekly number of new cases. We then used a negative binomial (NB) model to sample the observed incidence rate:

$$C_t^{(0)} \sim \text{NB}(\mu_t = \bar{\rho} C_t, \rho_k)$$

where $\bar{\rho}$ is the mean reporting probability and ρ_k the reporting over-dispersion, representing extra variability in the mean reporting probability.

Complete model and causal graph. The complete model thus consisted of the discrete-time transmission model and the observation model described above. Because of our simplifying assumptions (in particular, a fixed generation time and an immediate effect of climate on transmission), this model was exactly represented by the causal graph displayed in the main text (Fig. 2).

Numerical implementation. The model was implemented in the R package pomp¹¹⁴, operating in R version 4.4.1¹¹⁵. All figures were created with the R package ggplot¹¹⁶, and the data for the maps was obtained from Natural Earth (<https://www.naturalearthdata.com>). Other packages used for specific vignettes are cited below.

Vignette 1 on descendants and measurement bias

Simulation details. The simulations for this vignette were based on climatic data in Lübeck, Germany, a location with a temperate oceanic climate (Köppen–Geiger classification: Cfb) characterized by large seasonal variability in temperature (CV: 0.59), little variability in relative humidity (CV: 0.07) and marked correlation between the two ($r_s = -0.48$). The model was simulated for three different values of the basic reproduction number (1.25, 2.5 and 5) and an average duration of immunity of 1 year; the other parameters were fixed to the values indicated in Supplementary Table 3.

Regression model for time series of observed cases. To identify a candidate regression model for the variable $C_t^{(0)}$ (observed incidence rate), we first log-transformed the variable C_t (true incidence rate):

$$\log C_t = \log \left(\beta_{t-1} \frac{I_{t-1}}{N} S_{t-1} \right) = C + \delta_{\text{Te}} \text{Te}'_{t-1} + \delta_{\text{RH}} \text{RH}'_{t-1} + \log(I_{t-1} S_{t-1})$$

where C is a constant. This equation, alongside the negative binomial observation model connecting $C_t^{(0)}$ and C_t , shows that a natural candidate model for $C_t^{(0)}$ is a negative binomial regression model with log-link, and Te'_{t-1} and RH'_{t-1} as covariates. In practice, the variable $I_{t-1} S_{t-1}$ is unobserved but may be captured by including a function of time as a covariate⁶⁸. Here, we fitted a negative binomial GAM¹¹⁷ that included the climatic covariates and a smooth of time to capture temporal variations in this variable. To give the regression model enough flexibility to capture these variations (which may occur over fast time scales; see Fig. 3), we set the basis dimension of the smooth to 50, or approximately 1 degree of freedom per 10 weeks of data. As a control, we also verified that the exact model with $\log(S_{t-1} I_{t-1})$ as a covariate yielded, on average, unbiased estimates of δ_{Te} and δ_{RH} (Supplementary Fig. 1). All the regression models were fitted using the mgcv package (version 1.8-42) in R¹¹⁷.

Regression models for time series of effective reproduction numbers. By definition, the time-varying effective reproduction number $R_{e,t}$ equals:

$$R_{e,t} = \frac{R_{0,t} S_t}{N} \approx \frac{\beta_t S_t}{N}$$

where $R_{0,t} = \frac{\beta_t}{(\mu+1)}$ is the time-varying basic reproduction number and $\mu \ll 1$ (per week) is the death rate. Taking logs, we find that:

$$\log R_{e,t} = C + \delta_{\text{Te}} \text{Te}'_t + \delta_{\text{RH}} \text{RH}'_t + \log(S_t)$$

where C is a constant. The corresponding DAG, shown in Supplementary Fig. 2, differs from the DAG for the incidence rate (Fig. 2) in that

the effective reproduction number depends on just one unobserved variable (S_t), while the incidence rate depends on two (S_t and I_t).

To derive an estimator for $R_{e,t}$, we first express it as a function of the incidence rates C_t . Specifically (see ‘Model formulation’), $C_{t+1} = \lambda_t S_t = \frac{\beta_t I_t S_t}{N} \approx R_{e,t} I_t$ and $I_t = C_t - \mu I_{t-1} \approx C_t$ (neglecting mortality). Hence, we find the following equation for $R_{e,t}$:

$$R_{e,t} \approx \frac{C_{t+1}}{C_t}$$

In other words, the effective reproduction number is approximately equal to the epidemic growth rate, as expected intuitively because of our assumption of a 1-week fixed generation time.

In practice, because of under-reporting, the true values of $R_{e,t}$ are unobserved. However, estimation is possible via renewal models that back-calculate $R_{e,t}$ values from the generation time distribution and the observed incidence rates. Denoting by $R_{e,t}^{(0)}$ this estimator, we generated it as follows:

$$\log R_{e,t}^{(0)} \sim N(\mu = \log R_{e,t}, \sigma = 0.1)$$

In other words, we assumed a small dose of noise (approximately 10% around the true value) but no systematic bias in the estimation $R_{e,t}$.

We then fitted normal (N) regression models of $\log R_{e,t}^{(0)}$ (outcomes) with $T_{e,t}'$, RH_t' and a smooth of time as covariates (Supplementary Fig. 1). As a control, we also verified that the true regression model with $T_{e,t}'$, RH_t' and $\log S_t$ as covariates yielded, on average, unbiased estimates (Supplementary Fig. 1).

Vignette 2 on climate variability as natural experiments

Simulation details. The simulations for this vignette were based on climatic data in Bogotá, Colombia, and Lübeck, Germany. In marked contrast to Lübeck’s climate (see above), Bogotá’s climate is classified as warm and temperate (Köppen classification: Csb), with little seasonality in temperature because of proximity to the Equator (CV: 0.04) but larger variability in relative humidity (CV: 0.07) and decoupling between the two climatic variables ($r_s = -0.1$).

To introduce model misspecification during the estimation of model parameters, we implemented a stochastic transmission model where the deterministic transmission rate β_t was multiplied at every time step by gamma white noise with mean 1 and standard deviation 0.02 (ref. 118). The complete model was, therefore, fully stochastic, with noise affecting both the transmission and the observation models. The model parameters were set as follows: basic reproduction number of 1.25, average duration of immunity of 1 year and other parameters fixed to the values indicated in Supplementary Table 3.

Parameter estimation protocol. The following six model parameters were assumed unknown and estimated from the data: basic reproduction number (R_0), waning immunity rate (α), mean reporting probability (ρ), reporting over-dispersion (ρ_k) and climatic parameters (δ_{Te} and δ_{RH}). To generate the synthetic data for estimation, we first generated 100 replicate time series of observed weekly cases ($C_t^{(0)}$) from the fully stochastic model. For every replicate time series, we then fitted the misspecified model—with a deterministic transmission model and stochastic observation model—using trajectory matching⁹². Specifically, we used the Nelder–Mead algorithm¹¹⁹ (initialized at the true parameter values) to maximize the log-likelihood and identify the maximum likelihood parameter estimates. All the parameters were estimated on an unconstrained scale using log (parameters R_0 , α and ρ_k) or logit (parameter ρ) transformations.

Vignette 3 on confounding bias

Simulation details. The simulations for this vignette were based on climatic data from 15 weather stations in Spain and 19 in Colombia,

located near the major cities of both countries. Continental Spain exhibits relatively uniform temperate climates with consistent seasonal variations of temperature and humidity seasonality across the country. In contrast, Colombia displays a range of tropical climates, with diverse seasonal patterns of precipitation along a latitudinal gradient¹⁰¹. The models were simulated without incorporating spatial spread between the locations and with a basic reproduction number of 2.5 and an average duration of immunity of 2 years to achieve yearly epidemics in all the locations. The other parameters were fixed to the values indicated in Supplementary Table 3.

Epidemic synchrony and assessment of spatial spread. To assess spatial synchrony between locations, we estimated the non-parametric (cross-) correlation function (NCF) of the simulated time series using the NCF package¹⁰⁰. We estimated the 95% confidence intervals using 500 bootstraps.

To diagnose the risk of confounding bias caused by climate, we estimated the speed of the potential travelling wave that would have resulted from spatial diffusion. To do so, we first estimated the difference in epidemic peak timing as the lag (in weeks), maximizing the cross-correlation function between every location and a reference location (Riohacha for Colombia and Gijón for Spain). We then regressed this difference against the geographical distance between locations and estimated the speed wave as the inverse of the regression coefficient.

To further evaluate how climate can confound the estimate of spatial spread between locations, we fitted negative binomial Gaussian process models, that assumes a joint multivariate normal (MVN) distribution, to the time series of observed incidence from all locations in Colombia:

$$\begin{pmatrix} \log C_t^{(0)(1)} \\ \log C_t^{(0)(2)} \\ \vdots \\ \log C_t^{(0)(19)} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \log \mu_t^{(1)} \\ \log \mu_t^{(2)} \\ \vdots \\ \log \mu_t^{(19)} \end{pmatrix}, K \right)$$

The Gaussian process captures the covariance $K^{(ij)}$ between the incidence of any pair of locations i and j as a function of the geographic distance between locations, defined by the exponential quadratic kernel:

$$K^{(ij)} = \eta^2 \exp \left(-\frac{\|x^{(i)} - x^{(j)}\|^2}{2\varphi^2} \right)$$

where $\|x^{(i)} - x^{(j)}\|$ is the Euclidean distance between two locations, η is the maximum covariance between locations and φ is a characteristic distance that controls the spatial scale over which the covariance varies. Based on the proposed ‘true’ and ‘smooth’ regression models described in the methods for vignette 1, we fitted models including the size of the infected and susceptible population or a smooth of time with or without the environmental variables. For example, for the true model, we wrote $\log \mu_t^{(1)} \approx \log(S_{t-1}^{(1)} I_{t-1}^{(1)}) + Te_{t-1}^{(1)} + RH_{t-1}^{(1)}$ in location 1. Finally, we calculated the correlation matrix between the incidence of all locations from the estimated covariance matrix K . All models were fitted using the brms package¹²⁰ with four Markov chains, each run for 10,000 iterations, assuming uninformative priors.

Two-location transmission model with spatial diffusion. As purely statistical approaches resulted in confounded estimations of spatial spread, we moved on to estimate spatial spread with transmission models. We extended our climate-forced SIRS model to include spatial diffusion by dividing the population into two coupled locations, for $i = 1, 2$ (refs. 121, 122):

$$S_{t+1}^{(i)} = S_t^{(i)} + \mu N^{(i)} + \alpha R_t^{(i)} - (\lambda_t^{(i)} + \mu) S_t^{(i)}$$

$$I_{t+1}^{(i)} = \lambda_t^{(i)} S_t^{(i)} - \mu_t^{(i)}$$

$$R_{t+1}^{(i)} = R_t^{(i)} + I_t^{(i)} - (\alpha + \mu) R_t^{(i)}$$

with the force of infection in each location i given by:

$$\lambda^{(i)}(t) = \frac{\beta^{(ii)} I^{(1)}(t)}{N^{(1)}} + \frac{\beta^{(i2)} I^{(2)}(t)}{N^{(2)}}$$

We assumed a symmetric scenario where the transmission rate is the same within patches ($\beta^{(ij)} = \beta$ if $i=j$) but differs between patches ($\beta^{(ij)} = \tau\beta$ if $i \neq j$, where $0 \leq \tau \leq 1$ is the coupling strength). In this case, each location had the same population ($N^{(1)} = N^{(2)} = N$), and the endemic equilibria were independent of the location, given by $S^* = \frac{N}{R_0}$, $I^* = \frac{\alpha+\mu}{\alpha+\mu+1} (N - S^*)$, and $R^* = N - S^* - I^*$, where $R_0 = \frac{\beta(1+\tau)}{\mu+1}$.

Parameter estimation protocol. As in vignette 2, we assumed the same six parameters unknown and estimated from the simulated data without spatial spread. We fitted the two-location transmission model with spatial diffusion for every location and a reference location (Riohacha for Colombia and Gijón for Spain) using trajectory matching⁹². Again, we used the Nelder–Mead algorithm¹¹⁹ (initialized at the true parameter values) to maximize the log-likelihood and identify the maximum likelihood parameter estimates. Then, we generated likelihood profiles to estimate the spatial spread parameter τ by varying the parameter while maximizing the likelihood over the remaining parameters to obtain likelihood-ratio-test-based confidence intervals. All the parameters were estimated on an unconstrained scale using log (parameters R_0 , α and ρ_k) or logit (parameter $\bar{\rho}$ and τ) transformations.

Vignette 4 on mediation, direct and indirect causal effects

Simulation details. The simulations in this vignette were based on climatic data from Pasto, Colombia, and Lübeck, Germany. As described before, in Lübeck, temperature displays larger seasonal variability than relative humidity. In contrast, in Pasto, seasonal variability in relative humidity is larger (CV: 0.19) than in temperature (CV: 0.07).

As temperature affects humidity, the total causal effect of temperature comprises a direct effect and an indirect effect mediated by relative humidity. Thus, we simulated a model representing the total effect of temperature (with climatic parameters to $\delta_{Te} = -0.2$ and $\delta_{RH} = -0.2$) and another model representing only the indirect effect of temperature (with climatic parameters fixed to $\delta_{Te} = 0$ and $\delta_{RH} = -0.2$). For these simulations, we fixed the other parameters as follows: basic reproduction number of 1.25, average duration of immunity of 1 year and other parameters as indicated in Supplementary Table 3.

Reporting summary

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

Data availability

All data are available at https://github.com/DomenechLab/Causality_Seasonality and stored in Edmond, the open research data repository of the Max Planck Society, at <https://doi.org/10.17617/3.9CWN7W>. All weather data were extracted using the WeatherData function in Mathematica (<https://reference.wolfram.com/language/ref/WeatherData.html>), with the weather station names indicated in Supplementary Table 2.

Code availability

All R programming codes are available at https://github.com/DomenechLab/Causality_Seasonality and stored in Edmond, the open research data repository of the Max Planck Society, at <https://doi.org/10.17617/3.9CWN7W>.

References

- Our Risk for Infectious Diseases is Increasing Because of Climate Change (National Center for Emerging and Zoonotic Infectious Diseases, 2021).
- Impact of Climate Change on Infectious Diseases and Antimicrobial Resistance – Part 1 of the German Status Report on Climate Change and Health 2023 (Robert Koch Institute and Statistisches Bundesamt, 2023); https://www.rki.de/EN/Content/Health_Monitoring/Health_Report/GBEDownloads/JHealthMonit_2023_S3_Status_report_climate_change_health_part1.html
- Cissé, G. et al. in *Climate Change 2022: Impacts, Adaptation and Vulnerability* (eds Pörtner, H. O. et al.) 1041–1170 (Cambridge Univ. Press, 2022).
- The Global Health Observatory Estimated Number of Malaria Deaths (World Health Organization, 2024); <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimated-number-of-malaria-deaths>
- Cholera Worldwide Overview: Geographical Distribution of Cholera Cases Reported Worldwide (European Centre for Disease Prevention and Control, 2024); <https://www.ecdc.europa.eu/en/all-topics-z/cholera/surveillance-and-disease-data/cholera-monthly>
- Global Influenza Programme: Burden of Disease (World Health Organization, 2024); <https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring/burden-of-disease>
- Colón-González, F. J. et al. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. *Lancet Planet. Health* **5**, e404–e414 (2021).
- Escobar, L. E. et al. A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* **149**, 202–211 (2015).
- Kruger, S. E., Lorah, P. A. & Okamoto, K. W. Mapping climate change's impact on cholera infection risk in Bangladesh. *PLoS Glob. Public Health* **2**, e0000711 (2022).
- Baker, R. E. et al. Epidemic dynamics of respiratory syncytial virus in current and future climates. *Nat. Commun.* **10**, 5512 (2019).
- Baker, R. E., Mahmud, A. S. & Metcalf, C. J. E. Dynamic response of airborne infections to climate change: predictions for varicella. *Clim. Change* **148**, 547–560 (2018).
- Siraj, A. S. et al. Altitudinal changes in malaria incidence in highlands of Ethiopia and Colombia. *Science* **343**, 1154–1158 (2014).
- Baker-Austin, C. et al. Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nat. Clim. Change* **3**, 73–77 (2012).
- Martinez, M. E. The calendar of epidemics: seasonal cycles of infectious diseases. *PLoS Pathog.* **14**, e1007327 (2018).
- Shaman, J. & Kohn, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc. Natl Acad. Sci. USA* **106**, 3243–3248 (2009).
- Marr, L. C., Tang, J. W., Van Mullekom, J. & Lakdawala, S. S. Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence. *J. R. Soc. Interface* **16**, 20180298 (2019).
- 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).
- 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).
- Huq, A., West, P. A., Small, E. B., Huq, M. I. & Colwell, R. R. Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl. Environ. Microbiol.* **48**, 420–424 (1984).

20. Jusot, J.-F. et al. Airborne dust and high temperatures are risk factors for invasive bacterial disease. *J. Allergy Clin. Immunol.* **139**, 977–986.e2 (2017).
21. Paaijmans, K. P. et al. Influence of climate on malaria transmission depends on daily temperature variation. *Proc. Natl Acad. Sci. USA* **107**, 15135–15139 (2010).
22. Bayoh, M. N. & Lindsay, S. W. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). *Bull. Entomol. Res.* **93**, 375–381 (2003).
23. van der Worp, H. B. et al. Can animal models of disease reliably inform human studies? *PLoS Med.* **7**, e1000245 (2010).
24. Willem, L., Van Kerckhove, K., Chao, D. L., Hens, N. & Beutels, P. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS ONE* **7**, e48695 (2012).
25. Hernán, M. A. & Robins, J. M. *Causal Inference: What If* (Chapman & Hall/CRC, 2020).
26. Neal, B. *Introduction to Causal Inference from a Machine Learning Perspective* (2020); https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf
27. Kurth, T. Continuing to advance epidemiology. *Front. Epidemiol.* **1**, 782374 (2021).
28. Liu, T., Ungar, L. & Kording, K. Quantifying causality in data science with quasi-experiments. *Nat. Comput. Sci.* **1**, 24–32 (2021).
29. Collazo, A., Kuhn, H.-G., Kurth, T., Piccininni, M. & Rohmann, J. L. Rethinking animal attrition in preclinical research: expressing causal mechanisms of selection bias using directed acyclic graphs. *J. Cereb. Blood Flow Metab.* <https://doi.org/10.1177/027168X241275760> (2024).
30. Hernán, M. A., Wang, W. & Leaf, D. E. Target trial emulation: a framework for causal inference from observational data. *JAMA* **328**, 2446–2447 (2022).
31. Liang, L. & Gong, P. Climate change and human infectious diseases: a synthesis of research findings from global and spatio-temporal perspectives. *Environ. Int.* **103**, 99–108 (2017).
32. Kim, Y.-M., Park, J.-W. & Cheong, H.-K. Estimated effect of climatic variables on the transmission of *Plasmodium vivax* malaria in the Republic of Korea. *Environ. Health Perspect.* **120**, 1314–1319 (2012).
33. Jusot, J.-F. & Alto, O. Short term effect of rainfall on suspected malaria episodes at Magaria, Niger: a time series study. *Trans. R. Soc. Trop. Med. Hyg.* **105**, 637–643 (2011).
34. Haque, U. et al. The role of climate variability in the spread of malaria in Bangladeshi highlands. *PLoS ONE* **5**, e14341 (2010).
35. Xiao, D. et al. Spatiotemporal distribution of malaria and the association between its epidemic and climate factors in Hainan, China. *Malar. J.* **9**, 185 (2010).
36. Olson, S. H. et al. Links between climate, malaria, and wetlands in the Amazon Basin. *Emerg. Infect. Dis.* **15**, 659–662 (2009).
37. Hashizume, M., Terao, T. & Minakawa, N. The Indian Ocean Dipole and malaria risk in the highlands of western Kenya. *Proc. Natl Acad. Sci. USA* **106**, 1857–1862 (2009).
38. Teklehaymanot, H. D., Schwartz, J., Teklehaymanot, A. & Lipsitch, M. Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia II. Weather-based prediction systems perform comparably to early detection systems in identifying times for interventions. *Malar. J.* **3**, 44 (2004).
39. Teklehaymanot, H. D., Lipsitch, M., Teklehaymanot, A. & Schwartz, J. Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia I. Patterns of lagged weather effects reflect biological mechanisms. *Malar. J.* **3**, 41 (2004).
40. Abeku, T. A. et al. Effects of meteorological factors on epidemic malaria in Ethiopia: a statistical modelling approach based on theoretical reasoning. *Parasitology* **128**, 585–593 (2004).
41. Hii, Y. L., Zhu, H., Ng, N., Ng, L. C. & Rocklöv, J. Forecast of dengue incidence using temperature and rainfall. *PLoS Negl. Trop. Dis.* **6**, e1908 (2012).
42. Gomes, A. F., Nobre, A. A. & Cruz, O. G. Temporal analysis of the relationship between dengue and meteorological variables in the city of Rio de Janeiro, Brazil, 2001–2009. *Cad. Saude Publica* **28**, 2189–2197 (2012).
43. Lowe, R. et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in southeast Brazil. *Stat. Med.* **32**, 864–883 (2013).
44. Hashizume, M., Dewan, A. M., Sunahara, T., Rahman, M. Z. & Yamamoto, T. Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: a time-series study. *BMC Infect. Dis.* **12**, 98 (2012).
45. Ernest, A., Tan, S. B. & Wilder-Smith, A. Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiol. Infect.* **140**, 1244–1251 (2012).
46. Pham, H. V., Doan, H. T. M., Phan, T. T. T. & Minh, N. N. T. Ecological factors associated with dengue fever in a Central Highlands province, Vietnam. *BMC Infect. Dis.* **11**, 172 (2011).
47. Pinto, E., Coelho, M., Oliver, L. & Massad, E. The influence of climate variables on dengue in Singapore. *Int. J. Environ. Health Res.* **21**, 415–426 (2011).
48. Shang, C.-S. et al. The role of imported cases and favorable meteorological conditions in the onset of dengue epidemics. *PLoS Negl. Trop. Dis.* **4**, e775 (2010).
49. Chen, S.-C. et al. Lagged temperature effect with mosquito transmission potential explains dengue variability in southern Taiwan: insights from a statistical analysis. *Sci. Total Environ.* **408**, 4069–4075 (2010).
50. Tipayamongkhogkul, M., Fang, C.-T., Klinchan, S., Liu, C.-M. & King, C.-C. Effects of the El Niño–Southern Oscillation on dengue epidemics in Thailand, 1996–2005. *BMC Public Health* **9**, 422 (2009).
51. Lu, L. et al. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health* **9**, 395 (2009).
52. Johansson, M. A., Dominici, F. & Glass, G. E. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl. Trop. Dis.* **3**, e382 (2009).
53. Thammapalo, S., Chongsuwiwatwong, V., McNeil, D. & Geater, A. The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. *Southeast Asian J. Trop. Med. Public Health* **36**, 191–196 (2005).
54. Hashizume, M. et al. The Indian Ocean Dipole and cholera incidence in Bangladesh: a time-series analysis. *Environ. Health Perspect.* **119**, 239–244 (2011).
55. Rajendran, K. et al. Influence of relative humidity in *Vibrio cholerae* infection: a time series model. *Indian J. Med. Res.* **133**, 138–145 (2011).
56. Hashizume, M., Faruque, A. S. G., Wagatsuma, Y., Hayashi, T. & Armstrong, B. Cholera in Bangladesh: climatic components of seasonal variation. *Epidemiology* **21**, 706–710 (2010).
57. Paz, S. Impact of temperature variability on cholera incidence in southeastern Africa, 1971–2006. *Ecohealth* **6**, 340–345 (2009).
58. Constantin de Magny, G. et al. Environmental signatures associated with cholera epidemics. *Proc. Natl Acad. Sci. USA* **105**, 17676–17681 (2008).
59. Martinez-Urtaza, J. et al. Emergence of Asiatic *Vibrio* diseases in South America in phase with El Niño. *Epidemiology* **19**, 829–837 (2008).
60. Luque Fernández, M. A. et al. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Trans. R. Soc. Trop. Med. Hyg.* **103**, 137–143 (2009).

61. Hashizume, M. et al. The effect of rainfall on the incidence of cholera in Bangladesh. *Epidemiology* **19**, 103–110 (2008).
62. Huq, A. et al. Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* **71**, 4645–4654 (2005).
63. Hu, W. et al. Did socio-ecological factors drive the spatiotemporal patterns of pandemic influenza A (H1N1)? *Environ. Int.* **45**, 39–43 (2012).
64. Jusot, J.-F., Adamou, L. & Collard, J.-M. Influenza transmission during a one-year period (2009–2010) in a Sahelian city: low temperature plays a major role. *Influenza Other Respir. Viruses* **6**, 87–89 (2012).
65. Runge, J. et al. Inferring causation from time series in Earth system sciences. *Nat. Commun.* **10**, 2553 (2019).
66. Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S. & Sugihara, G. Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* **113**, 13081–13086 (2016).
67. Moazeni, M., Rahimi, M. & Ebrahimi, A. What are the effects of climate variables on COVID-19 pandemic? A systematic review and current update. *Adv. Biomed. Res.* **12**, 33 (2023).
68. Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P. & Hashizume, M. Time series regression model for infectious disease and weather. *Environ. Res.* **142**, 319–327 (2015).
69. Metcalf, C. J. E. et al. Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead. *Proc. R. Soc. B* **284**, 20170901 (2017).
70. Pearl, J. *Causality: Models, Reasoning, and Inference* (Cambridge Univ. Press, 2009).
71. Pearl, J. 3. The foundations of causal inference. *Sociol. Methodol.* **40**, 75–149 (2010).
72. Digitale, J. C., Martin, J. N. & Glymour, M. M. Tutorial on directed acyclic graphs. *J. Clin. Epidemiol.* **142**, 264–267 (2022).
73. Mina, M. J. et al. A Global Immunological Observatory to meet a time of pandemics. *Elife* **9**, e58989 (2020).
74. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton Univ. Press, 2008).
75. Kramer, S. C. & Shaman, J. Development and validation of influenza forecasting for 64 temperate and tropical countries. *PLoS Comput. Biol.* **15**, e1006742 (2019).
76. Weber, A., Weber, M. & Milligan, P. Modeling epidemics caused by respiratory syncytial virus (RSV). *Math. Biosci.* **172**, 95–113 (2001).
77. Singh, P. *Relative Humidity Calculator* (Omni Calculator, 2022); <https://www.omnicalculator.com/physics/relative-humidity>
78. Lawrence, M. G. The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications. *Bull. Am. Meteorol. Soc.* **86**, 225–234 (2005).
79. Johndrow, J., Ball, P., Gargiulo, M. & Lum, K. Estimating the number of SARS-CoV-2 infections and the impact of mitigation policies in the United States. *Harv. Data Sci. Rev.* **11**, 202–224 (2017).
80. Osthus, D., Hickmann, K. S., Caragea, P. C., Higdon, D. & Del Valle, S. Y. Forecasting seasonal influenza with a state-space SIR model. *Ann. Appl. Stat.* **11**, 202–224 (2017).
81. van Smeden, M., Lash, T. L. & Groenwold, R. H. H. Reflection on modern methods: five myths about measurement error in epidemiological research. *Int. J. Epidemiol.* **49**, 338–347 (2020).
82. Ackley, S. F. et al. Compartmental model diagrams as causal representations in relation to DAGs. *Epidemiol. Methods* **6**, 20060007 (2017).
83. Imai, C. & Hashizume, M. A systematic review of methodology: time series regression analysis for environmental factors and infectious diseases. *Trop. Med. Health* **43**, 1–9 (2015).
84. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* 2nd edn (Chapman and Hall/CRC, 2020).
85. Nash, R. K., Nouvellet, P. & Cori, A. Real-time estimation of the epidemic reproduction number: scoping review of the applications and challenges. *PLOS Digit. Health* **1**, e0000052 (2022).
86. Madeleine, C. T. & Simon, J. M. *Climate Information For Public Health Action* (Routledge, 2018).
87. Beck, H. E. et al. Present and future Köppen–Geiger climate classification maps at 1-km resolution. *Sci Data* **5**, 180214 (2018).
88. Butsic, V., Lewis, D. J., Radloff, V. C., Baumann, M. & Kuemmerle, T. Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic Appl. Ecol.* **19**, 1–10 (2017).
89. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Primers* **2**, 6 (2022).
90. Hsiang, S. M., Meng, K. C. & Cane, M. A. Civil conflicts are associated with the global climate. *Nature* **476**, 438–441 (2011).
91. Pascual, M., Rodó, X., Ellner, S. P., Colwell, R. & Bouma, M. J. Cholera dynamics and El Niño–Southern Oscillation. *Science* **289**, 1766–1769 (2000).
92. King, A. A., Nguyen, D. & Ionides, E. L. Statistical inference for partially observed Markov processes via the R package pomp. *J. Stat. Softw.* **69**, 1–43 (2016).
93. Lavielle, M. *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools* (CRC Press, 2014).
94. Bretó, C., Ionides, E. L. & King, A. A. Panel data analysis via mechanistic models. *J. Am. Stat. Assoc.* **115**, 1178–1188 (2019).
95. Grenfell, B. T., Bjørnstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
96. Martinez-Bakker, M., King, A. A. & Rohani, P. Unraveling the transmission ecology of polio. *PLoS Biol.* **13**, e1002172 (2015).
97. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
98. Choisy, M. & Rohani, P. Changing spatial epidemiology of pertussis in continental USA. *Proc. R. Soc. B* **279**, 4574–4581 (2012).
99. Barrero Guevara, L. A. et al. Delineating the seasonality of varicella and its association with climate in the tropical country of Colombia. *J. Infect. Dis.* **228**, 674–683 (2023).
100. Bjørnstad, O. N., Ims, R. A. & Lambin, X. Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends Ecol. Evol.* **14**, 427–432 (1999).
101. Urrea, V., Ochoa, A. & Mesa, O. Seasonality of rainfall in Colombia. *Water Resour. Res.* **55**, 4149–4162 (2019).
102. Bjørnstad, O. N. *Epidemics: Models and Data Using R* (Springer, 2018).
103. Cummings, D. A. T. et al. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* **427**, 344–347 (2004).
104. Westreich, D. & Greenland, S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am. J. Epidemiol.* **177**, 292–298 (2013).
105. Yang, W., Elankumaran, S. & Marr, L. C. Relationship between humidity and influenza A viability in droplets and implications for influenza's seasonality. *PLoS ONE* **7**, e46789 (2012).
106. Weiser, J. N., Ferreira, D. M. & Paton, J. C. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat. Rev. Microbiol.* **16**, 355–367 (2018).
107. Opatowski, L. et al. Assessing pneumococcal meningitis association with viral respiratory infections and antibiotics: insights from statistical and mathematical models. *Proc. R. Soc. B* **280**, 20130519 (2013).
108. Domenech de Cellès, M. et al. Unraveling the seasonal epidemiology of pneumococcus. *Proc. Natl Acad. Sci. USA* **116**, 1802–1807 (2019).

109. Nguyen, J. L. & Dockery, D. W. Daily indoor-to-outdoor temperature and humidity relationships: a sample across seasons and diverse climatic regions. *Int. J. Biometeorol.* **60**, 221–229 (2016).
110. Verheyen, C. A. & Bourouiba, L. Associations between indoor relative humidity and global COVID-19 outcomes. *J. R. Soc. Interface* **19**, 20210865 (2022).
111. WeatherData: Wolfram Language Function (Wolfram Research, 2014); <https://reference.wolfram.com/language/ref/WeatherData.html>
112. Alduchov, O. A. & Eskridge, R. E. Improved magnus form approximation of saturation vapor pressure. *J. Appl. Meteorol. Climatol.* **35**, 601–609 (1996).
113. Prussin, A. J. II et al. Survival of the enveloped virus Phi6 in droplets as a function of relative humidity, absolute humidity, and temperature. *Appl. Environ. Microbiol.* **84**, e00551-18 (2018).
114. King, A. A. et al. pomp: statistical inference for partially-observed Markov processes. *Github* <https://kingaa.github.io/pomp/> (2023).
115. R Core Team. R: a language and environment for statistical computing. *The R Foundation* <https://www.R-project.org/> (2023).
116. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
117. Wood, S. N. *Generalized Additive Models: An Introduction with R* (CRC Press/Taylor & Francis Group, 2017).
118. He, D., Ionides, E. L. & King, A. A. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J. R. Soc. Interface* **7**, 271–283 (2010).
119. Nelder, J. A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
120. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
121. Lloyd, A. L. & May, R. M. Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**, 1–11 (1996).
122. Brauer, F., Castillo-Chavez, C. & Feng, Z. Spatial structure in disease transmission models. *Math. Models Epidemiol.* **69**, 457 (2019).

Acknowledgements

This work was funded by the Max Planck Society through the core funding of M.D.d.C.'s Max Planck Research Group at the Max Planck Institute for Infection Biology.

Author contributions

L.A.B.G. and M.D.d.C. conceptualized the project and designed the methods. Model implementation and analyses were carried out by L.A.B.G. and M.D.d.C. L.A.B.G. and M.D.d.C. wrote the paper, with

support from S.C.K. and T.K. M.D.d.C. supervised the project. All authors approved the final version of the paper.

Funding

Open access funding provided by Max Planck Society.

Competing interests

T.K. reports outside the submitted work, having received research grants from the Gemeinsamer Bundesausschuss (G-BA—Federal Joint Committee, Germany). He has also received personal compensation from Eli Lilly and Company, Novartis, the BMJ and Frontiers. M.D.d.C. received postdoctoral funding (2017–2019) from Pfizer and consulting fees from GSK. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02594-3>.

Correspondence and requests for materials should be addressed to Matthieu Domenech de Cellès.

Peer review information *Nature Ecology & Evolution* thanks Rachel Baker, Paul Ferraro and David Fisman for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The simulated data was generated from a mathematical model of transmission using the package pomp (version 5.11) in R (version 4.4.1). The daily average records of dew point temperature (T_d , expressed in °C) and ambient temperature (T_e , expressed in °C) were extracted using the WeatherData function in Mathematica (version 12.0).

Data analysis

All analyses were conducted in R (version 4.4.1). For reproducibility, the renv package (version 1.0.7) was used to keep track of all the packages' version. The model was coded and run using the package pomp (version 5.11). Further analyses were performed using the packages mgcv (version 1.9-1), brms (version 2.21.0), and ncf (version 1.3-2). All code can be found at: https://github.com/DomenechLab/Causality_Seasonality and are stored in Edmond, the open research data repository of the Max Planck Society, <https://doi.org/10.17617/3.9CWN7W>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data and R programming codes are available at https://github.com/DomenechLab/Causality_Seasonality and stored in Edmond, the open research data repository of the Max Planck Society, <https://doi.org/10.17617/3.9CWN7W>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study explores how concepts from causal inference can be applied to understand the effects of weather on the transmission and infection dynamics of infectious diseases in the context of climate change. We use mathematical models of transmission to generate simulations and demonstrate how these concepts improve study design, reduce bias, and enhance the interpretation of the impacts of climatic variables on disease transmission.
Research sample	We generated simulations from a mathematical model of transmission forced by weather (temperature and relative humidity). We used weather data from major cities of Colombia, Spain, and Germany. We chose these locations according to the characteristic of the climate.
Sampling strategy	This is a simulation study that used previously-collected climatic data, we did not conduct any sampling.
Data collection	The climatic data was collected from multiple weather stations located near the major cities of Colombia, Spain, and Germany.
Timing and spatial scale	The simulations were generated using weekly climatic data covering the period 2013–2022 in multiple weather stations located near the major cities of Colombia, Spain, and Germany.
Data exclusions	We did not exclude data for this study.
Reproducibility	All code to simulate the data and repeat the analyses has been made publicly available and the <code>renv</code> package (version 1.0.7) was used to keep track of all the packages' version.
Randomization	We simulated data from a mathematical model of transmission; no randomization was conducted.
Blinding	This is a modeling study, in which we used a transmission model to simulate data; blinding is not relevant to this study design.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|-------------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | Antibodies |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | Animals and other organisms |
| <input checked="" type="checkbox"/> | Clinical data |
| <input checked="" type="checkbox"/> | Dual use research of concern |
| <input checked="" type="checkbox"/> | Plants |

Methods

- | | |
|-------------------------------------|------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | ChIP-seq |
| <input checked="" type="checkbox"/> | Flow cytometry |
| <input checked="" type="checkbox"/> | MRI-based neuroimaging |

Plants

Seed stocks

NA

Novel plant genotypes

NA

Authentication

NA