# Early warning signals of emerging infectious diseases

Qinghua Zhao[1,2,3], Nuno Moniz[2], Alexis Korotasz[1], Carly Barbera[1], Jason R. Rohr[1,3*]

[1]Department of Biological Sciences, Environmental Change Initiative, Eck Institute of Global Health, University of Notre Dame, Notre Dame, IN, USA.

[2]Lucy Family Institute for Data & Society, University of Notre Dame

[3]These authors contributed equally.

[*]E-mail: jasonrohr@gmail.com.

**Abstract**

Early warning signals (EWS) provide a crucial opportunity to predict and mitigate emerging infectious disease outbreaks before they escalate[1–3]. However, their practical application remains largely theoretical, often limited to single diseases and specific locations, with little systematic evaluation of their predictive accuracy across different contexts[1]. Additionally, traditional EWS approaches fail to directly estimate outbreak timing or lead time-the period between detection of warning signals and outbreak onset-which is critical for effective intervention[1]. To address these gaps, we integrated EWS with time-to-event analysis to assess the predictive performance of 19 resilience indicators (model-free statistical properties of time series, RIs) across 184 time series encompassing 390 outbreaks of 31 infectious diseases in 134 regions. Our findings demonstrate that both traditional RI and time-to-event analyses reliably warn of upcoming outbreaks and combining the most accurate RIs significantly improves outbreak prediction, yielding an average lead time of 18-21 days-a window sufficient for implementing disease control and prevention measures. Lead time and outbreak timing were influenced by pathogen type, climate, and socio-economic factors. Temperature and precipitation exhibited unimodal effects on outbreak risk and lead time for vector-borne and viral diseases, while pathogens with longer incubation times and regions with higher human development index (an estimate of development and per capita income) experienced longer lead times and lower outbreak risk, especially for vector-borne diseases. These findings underscore the importance of integrating socio-environmental factors into outbreak prediction models and suggest that early initiation of monitoring and case reporting and an enhanced commitment to outbreak surveillance might extend lead times before an outbreak. Our study advances the field of epidemic forecasting by providing a robust framework for applying RI-based EWSs to

diverse pathogens and geographies, laying the foundation for a global disease early warning system that can guide proactive public health interventions.

**Key words**

Emerging infectious diseases, Early warning signals, Lead time, Time-to-event analysis, Socio-ecological drivers

## Introduction

Predicting infectious disease outbreaks weeks before they occur so that preventative and mitigating actions can be implemented could avert massive disease worldwide, saving millions of lives and improving quality of life globally[1]. Several approaches can theoretically be used to anticipate outbreaks[1–3], such as dynamical mathematical models, monitoring Google searches and social media posts, wastewater surveillance for pathogens, and calculating generic, model-free statistical properties of time series (e.g., lag-1 autocorrelation, variance and the first differenced variance) as early warning signals (EWSs) of complex systems approaching critical transitions or "tipping points"[4–7]. The value of using statistical properties of time series as EWSs is that they require less effort and money than parameterizing dynamical models or monitoring wastewater for pathogens, and have regularly out-performed black-box, deep-learning-based approaches and Google search and social media surveillance[1]. Additionally, across many complex systems, statistical properties of time series are well linked to critical slowing down or reduced resilience (i.e., the ability to maintain normal stabilizing dynamics when subjected to disturbances) as systems approach critical transitions[1], which is why these types of EWSs are often referred to as resilience indicators (RIs). In fact, it is well documented that there were RIs in time series of past earthquakes[8], state shifts in ocean circulation systems and climate[9], and collapses of the global finance market[10–12], societies[13–16], and ecosystems[17,18]. The application of RIs to human health has been more limited[19], but RIs have been detected for asthmatic attacks[20] and epileptic seizures[21–23].

Recently, there has been interest in applying RIs theory to anticipate infectious disease outbreaks[1,24–29]. For example, RIs were recently detected for malaria resurgence in Kenya[30] and COVID-19 waves[31]. However, the application of RIs to shifts from endemic disease-free to epidemic disease-emergence states have been predominantly theoretical and studies of real-world data have focused mostly on single diseases, regions, and RIs with inconsistent success[1,24–29]. Hence, the use of RIs to offer reliable EWSs across infectious

diseases remains equivocal[32]. Additionally, little is known about the combinations of multiple RIs that could maximize EWS performance[1,32] (Fig. 1), the performance of RIs across different types of diseases[32], or the traits of pathogens and locations that affect outbreak predictions and lead times (how early an outbreak is anticipated; Fig. 1), the latter of which is critical because it determines the amount of time communities have to prevent or mitigate an impending outbreak. Another major knowledge gap is that, unlike dynamical models or wastewater and social media monitoring, traditional RI-based approaches do not actually predict the probability or the timing of a critical transition because they never treat the timing of the actual transition (i.e., outbreak) as the dependent variable. Rather, RI-based approaches use a significant positive relationship between the RIs (dependent variable) and time before the critical transition (independent variable) as an indirect estimate of an impending state shift in the system[30,33], despite there being a well-established class of statistical analyses for studying how long it takes for a binary event to happen, which is referred to as time-to-event, failure time, or survival analyses[34,35] (Fig 1).

To fill these gaps in understanding, we analyzed a high-resolution global dataset of daily incidence of 390 outbreaks across 31 infectious diseases of humans (see Materials and Methods for a list of diseases) from 184 time series, 36 countries, and 134 regions (Fig 2). Among these 31 diseases, 64% were zoonotic, 58% were zoonotic diseases with a wildlife reservoir, 35% were vector-borne, and 54%, 35%, and 5% were caused by viruses, bacteria, and protozoa, respectively (Fig. 2). Using these 184 time series, we tested 19 RIs in total, 11 classic RIs, 6 RIs identified from machine learning and new modelling techniques, and 2 RIs specifically designed for time series influenced by seasonal forcing. To test for the significance of the RIs, we advanced a moving window forward towards the outbreak, testing for a significant positive association between the RIs and time in each window based on a Kendall $\tau$ rank correlation statistic (see Material and Methods). The performance of the RI in predicting outbreaks was evaluated based on its true positive and negative rates (i.e., using the area under the receiver operating characteristic curve (AUC)) and 5-fold cross validation. For outbreaks that could be predicted using an RI, we calculated the lead time as the days between the first instance that $\tau$ was significant and the start of the outbreak (Fig. 1). In contrast to this traditional RI approach, we also used a time-to-event, Cox proportional hazards analysis to more directly predict lead time and test predictors of the timing of outbreaks.

**Evaluating Resilience Indicators and Lead Times**

Across disease systems, we hypothesized that the traditional RI approach would provide outbreak predictions with both a high true positive and negative rate, that combinations of RIs would outperform single RIs in anticipating outbreaks, and that the best RI combinations would provide ample warning of an impending outbreak. Across all 31 diseases, the best performing single RI across moving window lengths was standard deviation (mean AUC=0.91 from cross validation, window length=13 days) (Fig 3a-d, S1). Combining standard deviation with the second highest preforming RI, the hurst exponent (mean AUC=0.88 at the same window length, Fig 3b, S1), increased the mean AUC to 0.92 (Fig 3b). However, the mean AUC from the cross validation did not increase when combining standard deviation with other top-ranking RIs across moving windows (Fig 3a), or combining the standard deviation with those RIs that were least correlated with it (Fig 3c, 3d). Hence, using the raw AUC rather than the mean AUC from cross validation can result in overfitting and thus overly optimistic estimates of model performance. For the best combination of RIs (standard deviation and hurst exponent), the mean lead time was 17.8 days (±1.3 s.e.; Fig. 4) across times series and diseases. The time-to-event analyses, which captures both the likelihood of an outbreak and its timing, also produced a dependable model, accounting for 65.8% (marginal $R^2$) of the variation in the hazard of an outbreak (Extended data Fig 1). The $\tau$ value from the best combination of RIs was negatively associated with time to an outbreak in the time-to-event analyses (Extended data Fig 1b) and thus a higher hazard of an outbreak (Extended data Fig 1a), reflecting the consistency between the traditional RI and time-to-event models. The time-to-event analysis estimated a mean lead time of 21.0 days (±1.2 s.e.; Extended data Fig 2), which is 3.2 days longer than the estimate provided by the traditional RI approach. Overall, these results suggest that both RI and time-to-event analyses reliably warn of upcoming outbreaks, combinations of RIs outperform single RIs in anticipating outbreaks, and the average warning provided by RIs of nearly 18-21 days is enough time to mitigate, and possibly even prevent, outbreaks.

Notably, compared to the original fine-scale daily data, the coarse-scale weekly data that was derived by simply averaging the daily data each week, resulted in lower AUC values (Extended Data Fig. 3) and a mean lead time of 53.4 days (Extended Data Fig. 4), which is >30 days longer than the lead time from the daily data. This indicates that temporally coarse data resolution can provide an overly optimistic estimation of lead times for outbreak mitigation and prevention. Therefore, the use of finer-scale daily data is highly recommended for more reliable EWS assessments and lead time estimates.

135

136 **How Traits of Pathogens and Outbreak Locations affect EWSs**

137 Next, using the daily data, we sought to identify the traits of pathogens and outbreak
138 locations that could explain variation in the time to outbreak occurrence and lead times.
139 We hypothesized that outbreaks caused by pathogens with shorter incubation times (the
140 interval between exposure to a pathogen and the appearance of the first symptoms) would
141 be harder to anticipate and would have shorter lead times than those with longer
142 incubation times. Temperature and precipitation were expected to exhibit hump-shaped
143 relationships with time to outbreak occurrence and lead time, as these environmental
144 factors are known for unimodal effects on transmission dynamics and pathogenesis[36,37],
145 particularly for vector-borne pathogens, all of which have ectothermic hosts that are often
146 more sensitive to climatic factors than endothermic hosts. Hence, greater deviations in
147 either direction from the optimal temperature or precipitation (anomalies) were expected to
148 decrease lead time. Additionally, given that vector-borne pathogens tend to have greater
149 frequency- than density-dependent transmission[38–40], we postulated no relationship
150 between lead time and human population density for these pathogens. Finally, given that
151 more developed areas have greater outbreak surveillance, health infrastructure, and case
152 reporting, we hypothesized that regions with a higher human development index (HDI; a
153 statistical composite of life expectancy, years of education, and per capita income)[41] would
154 have a lower hazard of outbreaks and longer lead times than regions with lower HDIs,
155 especially for vector-borne pathogens that are often effectively combatted by costly
156 insecticide applications and homes with windows or screens[42,43]. To test these *a priori*
157 hypotheses, we developed a baseline generalized linear mixed effects model for lead time
158 and a Cox proportional hazards model for time to outbreak occurrence and removed
159 nonsignificant factors through a backwards selection process to reach a final model (see
160 Materials and Methods).

161 These analyses revealed considerable variation in lead times and time to outbreak
162 occurrence across the time series and 31 diseases. For example, as predicted, pathogens
163 that spread rapidly because they have short incubation times also had short lead times,
164 with each natural log increase in days of incubation time increasing lead time by 0.47
165 ±0.10 day (mean±s.e) (Fig. 4c). Additionally, on average, vector-borne and viral pathogens
166 offered significantly shorter lead times (11.1 and 14.4 days) than non-vector-borne and
167 bacterial pathogens (23.0 and 21.4), respectively (Fig. 4). However, when temperature and
168 precipitation were near their mean values, lead times peaked for both vector-borne and

169 non-vector-borne pathogens, and vector-borne pathogens had longer lead times than non-
170 vector-borne pathogens (Fig. 4d, e). In fact, at mean temperature and precipitation, lead
171 times were 55.0 and 37.0 days, respectively, for vector-borne diseases, and approximately
172 21.0 and 18.0 days, respectively, for non-vector-borne diseases (Fig. 4d, e). Additionally,
173 vector-borne diseases occurred under more narrow temperature and precipitation
174 conditions and warmer temperatures (21.2 °C vs 15.7 °C) than non-vector-borne diseases
175 (Fig. 4d,e), consistent with the well-documented sensitivity of vector-borne pathogens to
176 environmental conditions and greater frequency of vector-borne diseases in tropical than
177 temperate locations[44].

178 Whereas lead times for both vector- and non-vector-borne diseases responded
179 unimodally to temperature, only viral pathogens were unimodally associated with
180 temperature, peaking near observed mean temperatures (16.5 °C); bacterial pathogens were
181 associated positively with temperature for both lead time and time to outbreak occurrence
182 (Fig. 4a, 5a,b). Consistent with these findings, several studies documented hump-shaped
183 relationships between arboviruses and temperature[44], whereas bacterial biological rates[45]
184 and infections in humans regularly show positive associations with environmental
185 temperature[46–49]. Precipitation exhibited a unimodal relationship with both lead time and
186 time to outbreak occurrence for vector-borne and non-vector-borne diseases (Fig. 4e, 5c)
187 and a unimodal relationship with lead time for viral pathogens (Fig. 4b), consistent with
188 previous studies documenting unimodal relationships between precipitation and these
189 pathogen types[37]. In contrast, precipitation was negatively associated with lead time for
190 bacterial pathogens (Fig. 4b), aligning with previous research documenting that higher
191 precipitation levels facilitate more rapid transmission of bacterial pathogens[49,50].

192 As predicted, lead time was associated positively with HDI, presumably because
193 wealthier locations invest more in health infrastructure and pathogen surveillance which
194 can facilitate earlier detection of human cases of disease (Fig. 4f). This positive association
195 between lead time and HDI was significant for vector-borne diseases but not for non-vector-
196 borne diseases (Fig. 4f). Similarly, the time-to-event analysis revealed an HDI-by-vector-
197 borne-disease interaction, with only vector-borne diseases showing a positive association
198 between HDI and time to outbreak occurrence (Fig. 5d). In fact, a shift from the low to high
199 HDI category more than halved the hazard of a vector-borne disease outbreak (Fig. 5d).
200 These results are consistent with vector-borne diseases being reduced by human
201 development more than non-vector-borne diseases, thus lengthening the time between
202 vector-borne disease outbreaks. For example, many vector-borne diseases can be effectively

203 managed by costly home improvements and insecticide applications[42,43]. Hence,
204 development can enable earlier detection and thus mitigation and prevention of
205 environmentally mediated diseases.

206 Time series length was associated positively with lead time (Extended Data Fig. 5a)
207 and time to outbreak occurrence (Extended Data Fig. 5b), suggesting that early initiation of
208 monitoring and case reporting[51,52], the availability of longer historical datasets[53], or an
209 enhanced commitment to outbreak surveillance[54] might enable stakeholders to extend lead
210 times before an outbreak. Additionally, the lead time for the first outbreak in a time series
211 was significantly shorter than for subsequent outbreaks (Extended Data Fig. 6), which is
212 simply an artefact of the arbitrary time of the start of the time series. Additionally, neither
213 outbreak type (first or subsequent in the time series) nor human population density had
214 significant effects on lead time or the hazard of outbreak occurrence (Table S4; Extended
215 Data Fig. S7). However, we admonish against assuming that these factors do not affect
216 these endpoints, as increased statistical power or a different subset of human diseases
217 could produce different results.

218 While our study provides critical advancements in the development and application of
219 EWSs for predicting disease outbreaks, we acknowledge potential misclassifications of case
220 data, stemming from reporting biases, diagnostic inaccuracies, or surveillance
221 inconsistencies, which may introduce uncertainties and weaken the robustness of warning
222 indicators. Importantly, these misclassifications should only amplify random error in our
223 analyses that would reduce the likelihood of detecting RIs that warn of impending
224 outbreaks. Nevertheless, we discovered EWSs despite this impediment. Another important
225 limitation of our analyses is that they are restricted to endemic states of disease
226 transmission rather than the emergence of outbreaks involving new pathogens to a region.
227 Expanding the framework to incorporate RIs for new pathogens, such as those that
228 spillover from wildlife to humans, would enhance its utility in broader public health
229 contexts.

230

231 **Significance and Conclusions**

232 Our study provides a significant advancement in the development and application of
233 EWSs for infectious disease outbreaks. By systematically evaluating multiple RIs across a
234 diverse set of diseases and geographic locations, we identified key predictive factors that
235 enhance outbreak forecasting. Shifting to time-to-event analysis from the traditional RI-

based approach improved both the accuracy of outbreak predictions and lengthened the lead time, providing a more reliable framework for early intervention.

Importantly, we demonstrate that pathogen characteristics, climate variability, and socio-economic factors significantly influence lead time and outbreak timing, highlighting the need for context-specific forecasting models. Our findings show that pathogen incubation time is associated positively with lead time. Additionally, higher HDI regions experience longer lead times and reduced outbreak risks for vector-borne diseases, suggesting that public health infrastructure and surveillance capacity play crucial roles in disease prevention. Furthermore, we emphasize the need for high-resolution temporal data, as coarse-scale weekly data may lead to overly optimistic lead time estimates, potentially compromising timely responses.

This research lays the foundation for a global disease forecasting system that integrates real-time data streams and model-free statistical techniques to enhance predictive accuracy. Future work should focus on expanding the range of predictive indicators and developing interactive platforms for real-time outbreak monitoring. By improving our ability to anticipate and mitigate outbreaks, our study contributes to strengthening public health preparedness and minimizing the societal and economic impacts of infectious diseases.

## Methods

**Disease dataset.** For our analyses, we used the subset of the Analytics for Investigation of Disease Outbreaks (AIDO) dataset (https://aido.bsvgateway.org/) with daily reported disease case data and time series ≥1 week. This resulted in 184 time series (Fig 2) across 134 regions (i.e. either country or administrative level e.g. municipality and province; Fig 2), with an arithmetic mean time series length of 117 days (±14 s.e.) (see Table S1 for the corresponding regions and time series length for each disease). This dataset includes 31 diseases: Anthrax, Brucellosis, Campylobacteriosis, Chikungunya, Cholera, Dengue, Ebola, Gastroenteritis, Lassa Fever, Leptospirosis, Malaria, Marburg, Measles, Middle East Respiratory Syndrome, Monkeypox, Nipah, Norovirus, Novel Influenza A, Pertussis, Plague, Polio, Q Fever, Rift Valley Fever, Rubella, Salmonellosis, Shiga Toxin-Producing E. Coli, Shigellosis, Tularemia, West Nile Virus, Yellow Fever, Zika.

**Socio-economic and eco-environmental factors.** We gathered covariates, all of which were matched to each time series and corresponding location in the disease dataset. Daily temperature (°C) and precipitation (mm) were retrieved from NOAA's National Centers for

Environmental Information[55]. Annual human development index (HDI) for each 5 arc-minute gridded unit (latitude and longitude) was sourced from Kummu et al[41]. and subsequently aggregated to the respective country or administrative level using arithmetic means. Yearly population density (people per km$^2$) was retrieved from the Center for International Earth Science Information Network database (2018)[56]. Across all studied locations, the yearly resolution was used for HDI and population density because this was the only temporal resolution available. For each pathogen, the incubation time (day) was retrieved from the literature. Population density and incubation time were natural log transformed to ensure normality. Precipitation data were transformed using $\ln(x + 0.1)$ to accommodate zero values, while temperature data were transformed using $\ln(x + 20)$ to adjust for temperatures below -10°C in certain regions (e.g. Nordic countries such as Finland). Last, monthly temperature anomaly was calculated as the deviation from the long-term average for the corresponding months over the past 30 years (1950–1980), i.e. the period prior to significant anthropogenic warming, from NOAA's National Centers for Environmental Information[55], with the monthly precipitation anomaly computed using the same approach. Both temperature and precipitation anomalies were derived from raw values rather than applying a natural log transformation (archived on GitHub[57]).

**Traditional RI analyses.** Using the methods of Cori et al.[58] (because it outperforms others approaches[59]; for details of these methods, see Supplementary Information 1), we first delineated the start of each outbreak in each time series as the point when $R_e$ (i.e. the average number of secondary infections caused by a single infected individual at a given time point)[58–60] was ≥1. Thus, each time series was divided into no outbreak and outbreak sections (all code archived on GitHub[57]). Of the 184 time series, 27 did not have any days where $R_e$ was ≥1, representing no outbreak time series. Of the remaining157 time series, 68 (43.3%) had a single outbreak, and 89 (56.7%) had multiple outbreaks, resulting in 390 outbreaks in total (see Fig S2 for the frequency of outbreaks). For each time series, we then utilized 11 classic RIs, 6 RIs that have never been tested for infectious diseases, and 2 RIs specifically designed for time series influenced by seasonal forcing, yielding 19 total[57] (Table S2).

Following Dakos et al.[61], we adopted a moving window approach to test for a positive association between each RI and time as it approached $R_e \geq 1$. The length of the moving window varied from 7 to 21 days (1 to 3 weeks) with increments of 1 day (Fig S1). The magnitude of the association between each RI and time in each window was measured using the nonparametric Kendall $\tau$ rank correlation statistic[61]. The significance of the $\tau$ was

determined by estimating the fraction of 100 null surrogates that performed better than the observed dataset using the method of Ebisuzaki[62]. These randomized surrogate datasets have the same power spectra and autocorrelation structure of the original data but with randomized phases to eliminate spurious results because of the correlation in time.

Next, we employed receiver operating characteristic (ROC) curves to quantify the accuracy of each RI using 100 runs of 5-fold cross validation (i.e., 80% training data and 20% testing data), performed with the caret package[63]. The cross-validation aimed to test out-of-sample model performance to determine if our models were underfit, overfit, or "well-generalized". For each of 100 runs, the area under the ROC curve (AUC) was computed to quantify the model's discrimination ability, with a larger AUC indicating higher true positive and negative rates and thus higher accuracy. The arithmetic mean AUC obtained from 100 runs of cross-validation was taken as the final performance metric to assess the accuracy of each RI. Given that the length of the moving window can impact the AUC of an RI, the optimal moving window length was chosen for each RI by selecting the window length between 7 and 21 days with the highest mean AUC in the cross validations (Fig S1).

Once the most accurate single RI was determined, we tested whether combining up to four RIs would improve model predictions based on AUC and cross validation. We implemented two important dichotomies for these analyses. For the first dichotomy, we forced all RIs to use the window length that maximized the AUC for the most accurate RI (i.e., same window length) or we used the window length that maximized the AUC for each included RI (i.e., variable window lengths). Next, we conducted Principal Component Analyses (PCA) to determine associations among the RIs for each of these two window length scenarios. For the second dichotomy, we sequentially added either the next best performing single RI or the next most orthogonal RI (i.e., least correlated based on PCA) to the best performing RI until we tested up to four total RIs in our models. Importantly, sequentially adding the next most orthogonal RI to the model ensures that the baseline AUC goes up with each RI addition, but it does not guarantee that the mean AUC from the cross validation will continually rise with each added RI (Fig 2).

Once the most accurate combination of RIs across all 31 diseases was identified, we computed and analyzed the lead time (Fig. 1). The lead time was calculated as the days between the start of an outbreak ($R_e \geq 1$) and the point that the RI or combination of RIs first became significant. Next, we employed a Generalized Linear Mixed Model (glmmTMB function in glmmTMB package[64]) with a negative binomial error distribution to test various *a prioiri* hypotheses regarding lead time highlighted in the main text. We tested the main

effects of temperature, precipitation, HDI, pathogen incubation time, human population density, time series length, two pathogen traits (vector-borne or not; virus vs. bacteria), outbreak type (first outbreak in a time series or not), and all pairwise interactions between the pathogen traits, temperature, precipitation, HDI and human population density (Table S3). Next, we applied the same model structure but replaced temperature and precipitation with their respective anomalies to assess the impact of climate anomalies on lead time (see Supplementary Information 2). The parasite taxa were limited to viruses and bacteria, because our dataset had no fungi and helminths and because the sample size for protozoa was too small to justify analyses (Table S1). We compared the effects of first versus later outbreaks within time series to account for potential differences in lead time between initial and subsequent outbreaks. See the main text for the *a prioiri* hypotheses used to justify the interactions in this model. Additionally, we fitted the temperature and precipitation as quadratic polynomial functions, because most organisms have an optimal temperature and moisture[44], and treated type of RI (e.g. Hurst exponent) and disease types nested within regions as random intercepts. Additionally, HDI was classified into low (<0.55), medium (0.55–0.699), and high (⩾0.7) categories, following the recommendation of The United Nations Development Programme[65]. Figures were generated from the statistical model using the predict function in the ggeffects package[66].

**Time-to-event analyses.** To complement the indirect approach used by traditional RI analyses, we employed a mixed effects Cox proportional hazards time-to-event analysis (coxme function in coxme package[67]) to more directly test predictions of the timing of outbreak occurrence. We used the same model structure as described for the lead time analyses (Table S3). In these analyses, the response variable was a time series of 0 values indicating non-outbreak days ($R_e < 1$) and a single 1 value indicating the outbreak point where the outbreak began ($R_e \geq 1$). Given that there is no reliable way to graph the time-to-event analyses incorporating random intercepts[68–70], we replicated the final Cox proportional hazards model in the glmmTMB package with negative binomial errors, verified that the direction of the effects and significance matched between the two models (final results see Table S3, S4), and then generated figures from the glmmTMB model[64] using the predict function in the ggeffects package[66].

To directly estimate lead time rather than relying on indirect estimation as done by traditional RIs, we employed the parametric survival regression model (*survreg* function in survival Package)[71], i.e., a time to an event model. In the model, the predictor was the tau

value ($\tau$) of the best combinations of RIs (standard deviation, window length=13 days; Hurst exponent, window length=13 days; Fig 3b). The response variable was a time series of 0 values indicating non-outbreak days ($R_e < 1$) and a single 1 value indicating the hypothetical time point at which the outbreak could be anticipated in advance. The lead time was computed as the days between the first instance that survival regression model was significant and the start of the outbreak. All analyses were conducted by R version 4.1.2[72].

**Acknowledgements**

We thank members of the Rohr Laboratory of Ecology and Public Health at University of Notre Dame for discussions. This research was supported by grants from the National Science Foundation (DEB-2017785, DEB-2109293, BCS- 2307944, ITE- 2333795), US Department of Agriculture (2021-38420-34065), the Frontier Research Foundation, and the University of Notre Dame Poverty Initiative to JRR, and a grant from the University of Notre Dame Lucy Family Institute for Data and Society to JRR and NM.

**Author Contributions**

J.R.R. and Q.H.Z conceptualized the study. J.R.R. supervised the study. N.M. developed the EWS. Q.H.Z. performed the study and analyzed the data. J.R.R. and Q.H.Z. wrote the original draft. All the authors reviewed and edited the manuscript.

**Competing interests**

The authors declare no competing interests.

**Data availability**

All daily disease data used in this study are publicly available at ADIO (https://aido.bsvgateway.org/). All produced data are available on GitHub (https://github.com/QZhao16/EWSs.across.diseases).

**Code availability**

All the code for this article has been deposited at GitHub (https://github.com/QZhao16/EWSs.across.diseases) and all codes to reproduce all figures have been attached as supplementary files.

## Figures

**Figure 1. An overview of the data and methods used in this paper**. **a.** The entire dataset of time series was divided into two groups, those with (i.e. the average number of secondary infections caused by a single infected individual at a given time point, $R_e \geq 1$; red line) and without outbreaks. **b**. Time series with outbreaks were then partitioned into sections with and without outbreaks. Next, the most accurate resilience indicator (RI) was identified by considering single or combinations of RIs (see Methods section) with accuracy determined by taking the area under the receiver operating characteristic curve of 100 runs of a 5-fold cross validation. **b.** Once the most accurate RI was established, we quantified lead time or how early an outbreak could be anticipated (i.e., days between the start of an outbreak and the point that the RI first became significant). Next, we tested whether lead time was affected by the traits of the pathogens and geographic locations of the outbreak. Last, given that traditional RI analyses only indirectly predict outbreaks by testing for correlations between RIs and time, time-to-event analyses were employed to more directly predict the timing of outbreaks.

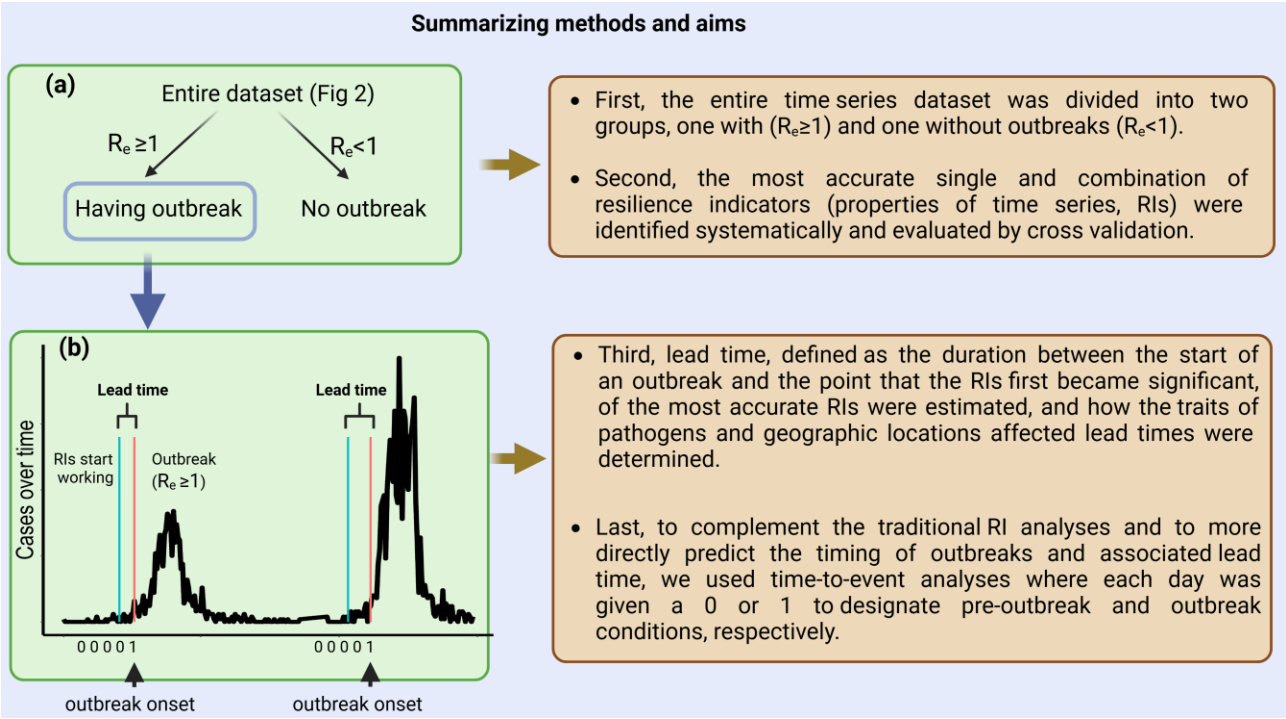


**Figure 2. The studied locations (a), total number of diseases (b), and total number of time series (c) grouped by pathogen- and disease-type categories**. See Table S1 for details of diseases, regions, and pathogen types.

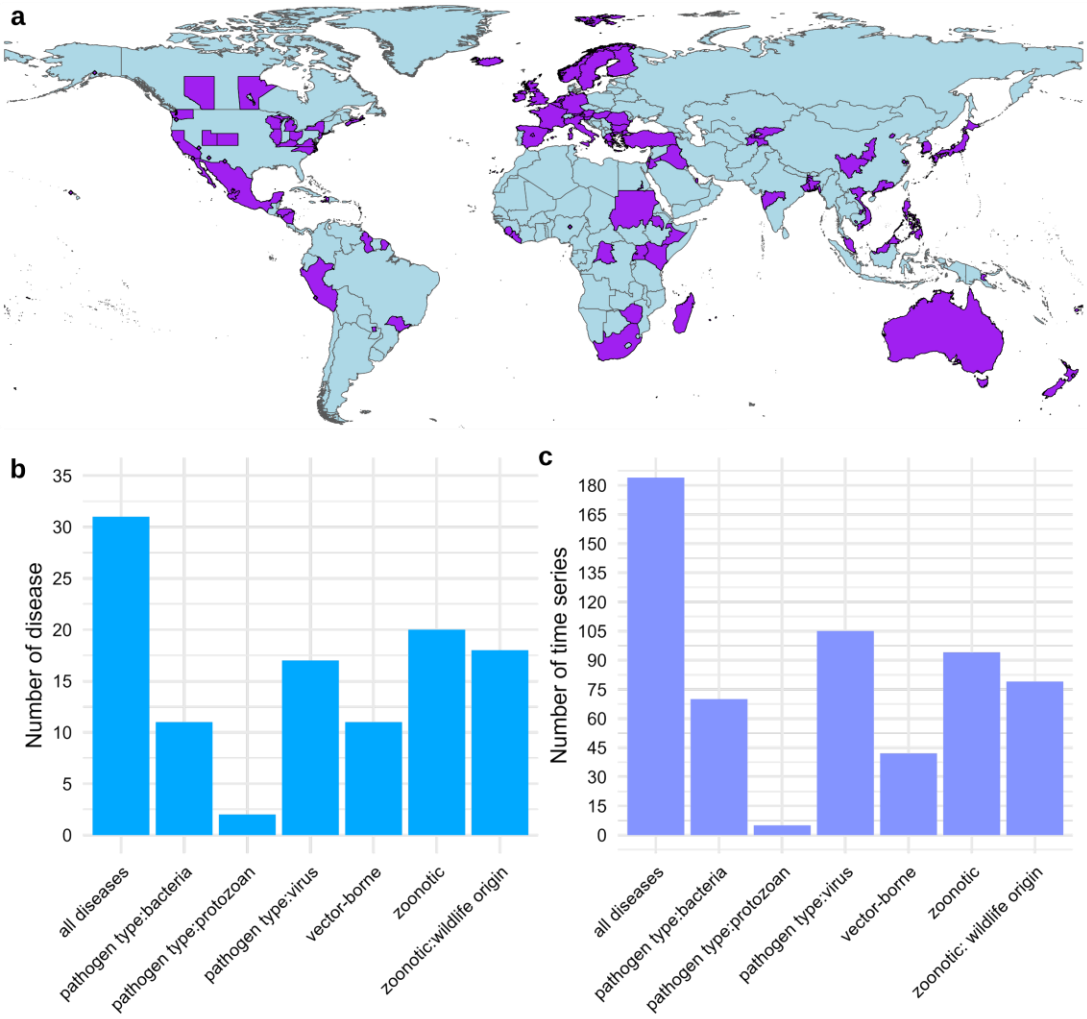


**Figure 3. The accuracy of resilience indicators (RIs) as early warning signals (EWSs) of outbreaks across 31 diseases, examined using mean area under the receiver operating characteristic curves (AUC) and cross validation**. All RIs were forced to use the window length that maximized the AUC for each RI (i.e., variable window lengths) (**a, c**) or the window length that maximized the AUC for the most accurate RI (i.e., same window length) (**b, d**). For both of these window length scenarios, we sequentially added either the next best performing single RI (**a, b**) or the next most orthogonal RI (i.e., least correlated based on Principal Component Analysis) to the best performing RI (**c, d**) until we tested up to four total RIs in our models. Importantly, sequentially adding the next most orthogonal RI to the model ensures that the baseline AUC goes up with each RI addition, but it does not guarantee that the mean AUC from the cross validation will continually rise with each added RI. Standard deviation: SD, hurst exponent: HE, relative dispersion: RD, autocovariance: AU, time series acceleration: TSA, first differenced variance: FDV, turning points: TP, kurtosis: KU and autocorrelation at lag 1: AR1.

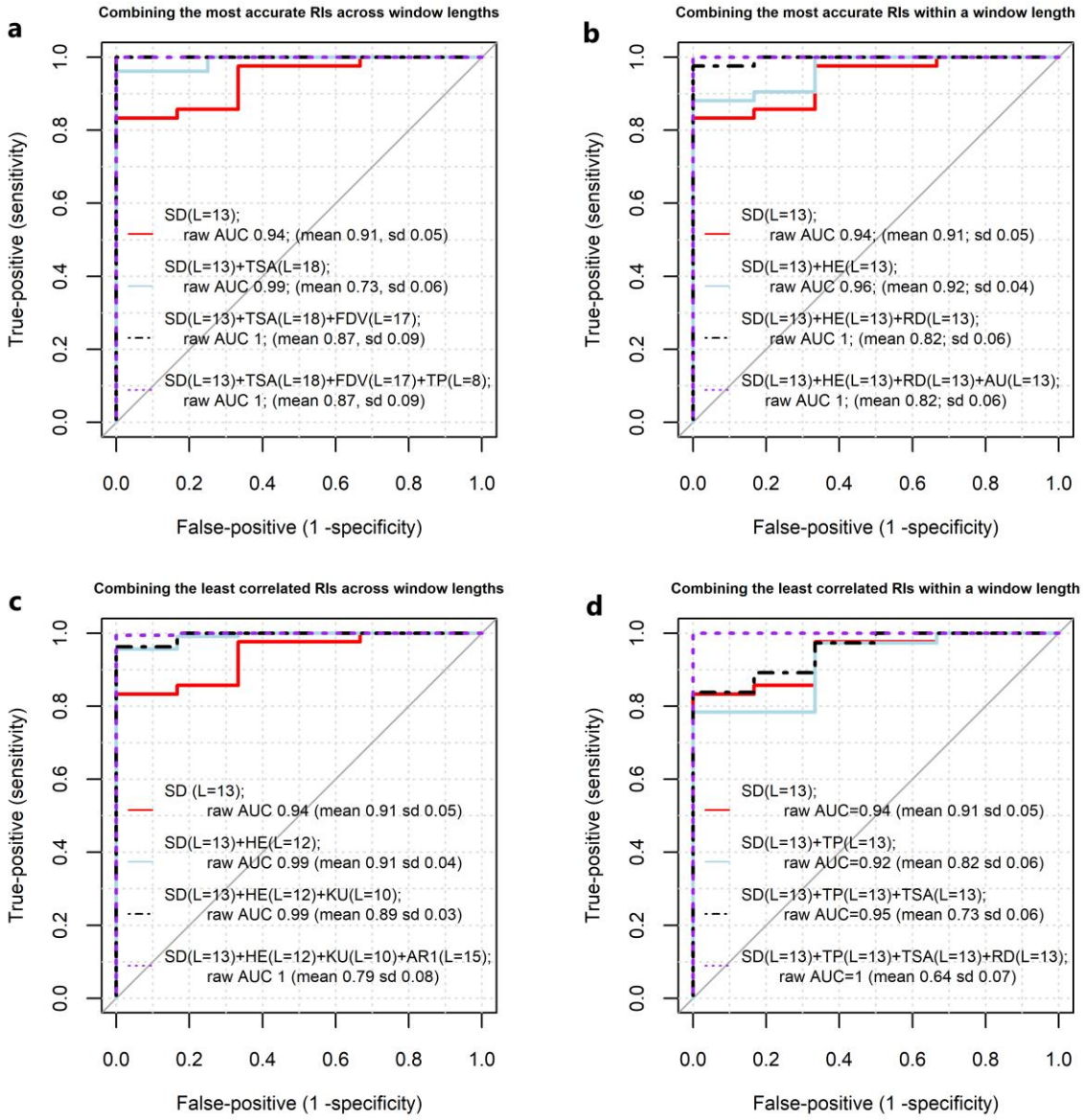


**Figure 4**. **A display of the main effects and interactions that were significantly associated with lead time.** The bold lines and gray bands depict best-fit lines and 95% confidence bands (or standard error in panel f), respectively, derived from the predicted values of the final statistical model that is available in Table S3. In panel **f**, the significance between groups was assessed using the emmeans function from the emmeans package. For the effects of population density and time series length on lead time, refer to Extended Data Fig. 1. Statistical restuls see table S3.

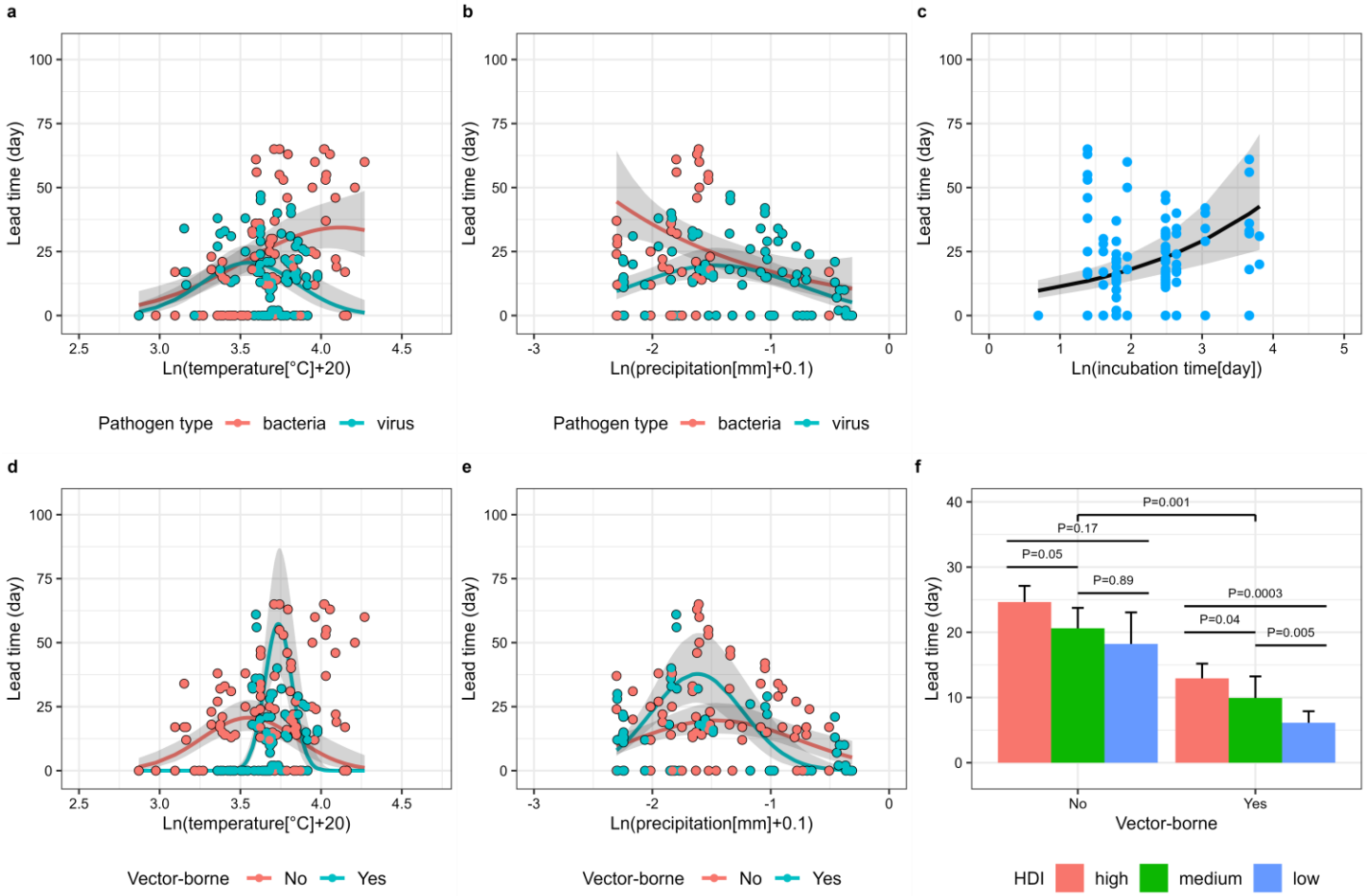
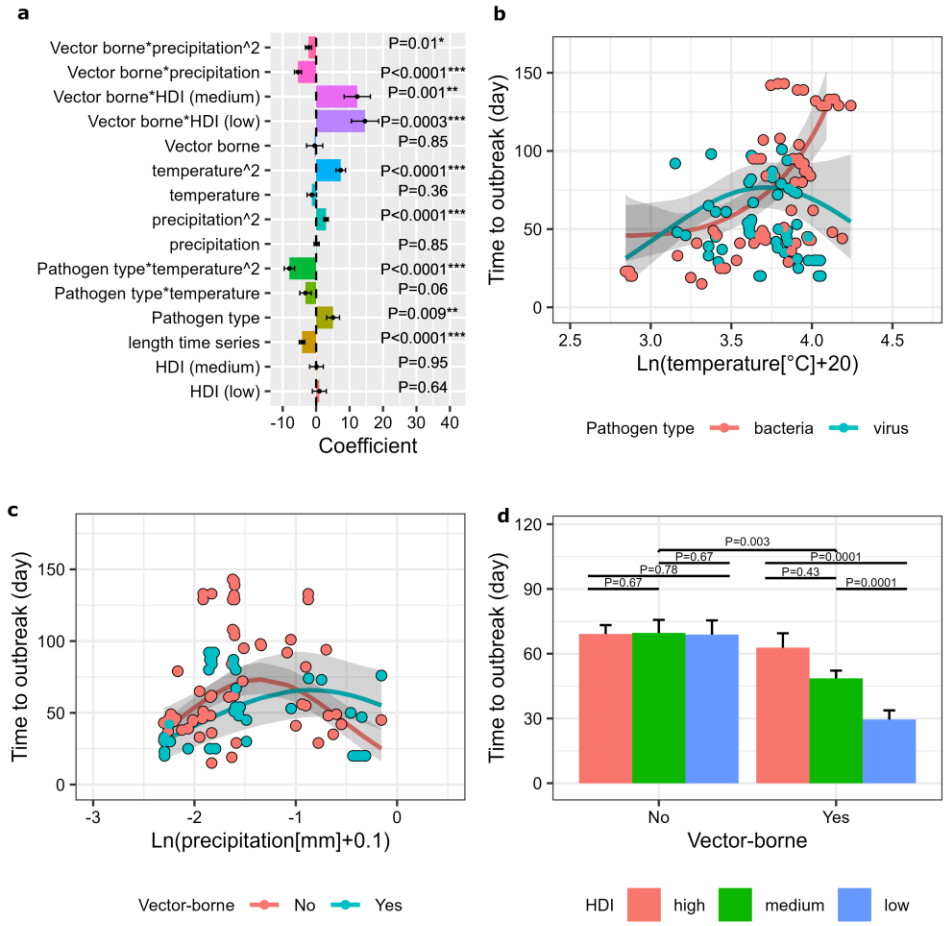
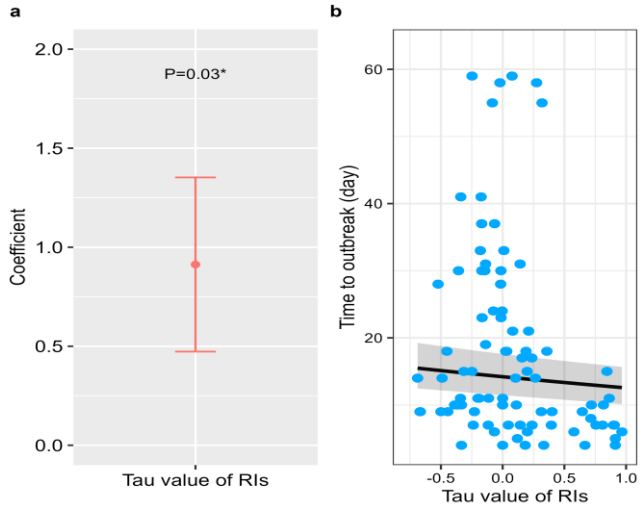


**Figure 5. Coefficients ± s.e (a) and plots of significant predictors (b-d) from a Cox proportional hazards mixed model examining the effects of various factors on time to outbreak occurrence**. **a**, A positive coefficient signifies an increased hazard of outbreak occurrence, and consequently a reduced time until an outbreak. A significant quadratic coefficient indicates a second-order polynomial function between a predictor and hazard of outbreak occurrence. **b**-**d**, the bold lines, bars, and gray bands depict best-fit lines, 95% confidence intervals, and 95% confidence bands, respectively, derived from the predicted values of the final statistical model that is available in Table S4. **d**, the significance between groups was assessed using the emmeans function from the emmeans package. Significant effects are indicated by *P <0.05, **P < 0.01, ***P < 0.001. Human development index (HDI)

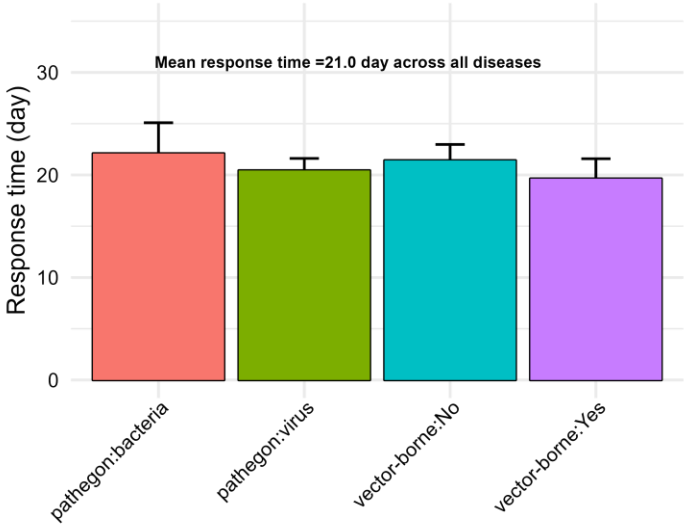


## Extended Data Figure

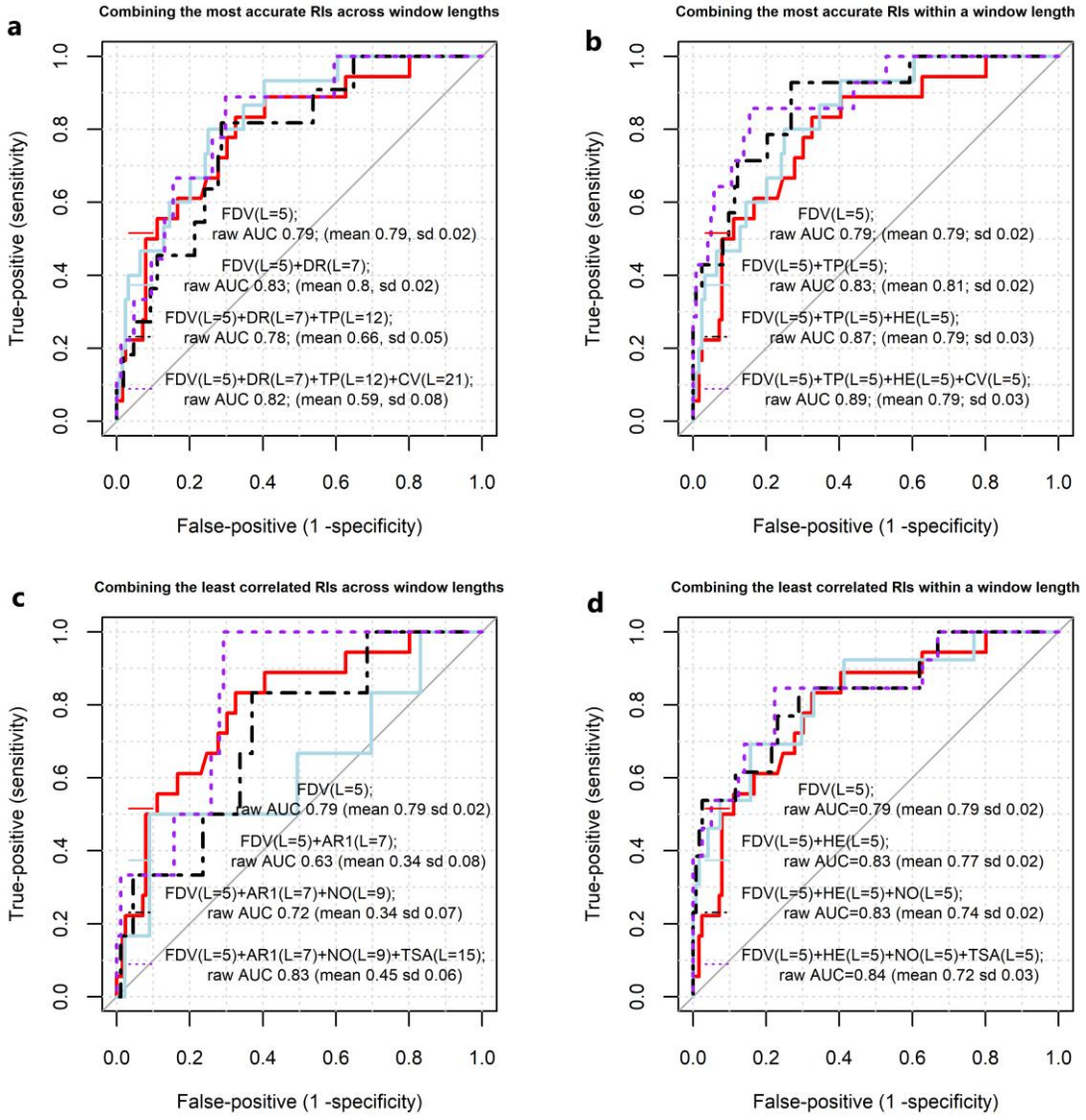
**Extended Data Fig. 1. Coefficients ± s.e (a) and plots of tau value ($\tau$) (b) from a Cox proportional hazards mixed model examining the effects of tau value ($\tau$) from the best combination of RIs (standard deviation and hurst exponent) on time to outbreak occurrence**. In the model, we treated regions nested within disease types and type of RIs (e.g. Hurst exponent) as random intercepts. The predictor was the tau value ($\tau$). The response variable was a time series of 0 values indicating non-outbreak days ($R_e < 1$) and a single 1 value indicating the outbreak point where the outbreak began ($R_e \geq 1$). **a**, A positive coefficient signifies an increased hazard of outbreak occurrence, and consequently a reduced time until an outbreak at (b). **b**, the bold lines and gray bands depict best-fit lines and 95% confidence intervals, respectively. Significant effects are indicated by *P <0.05, **P < 0.01, ***P < 0.001.



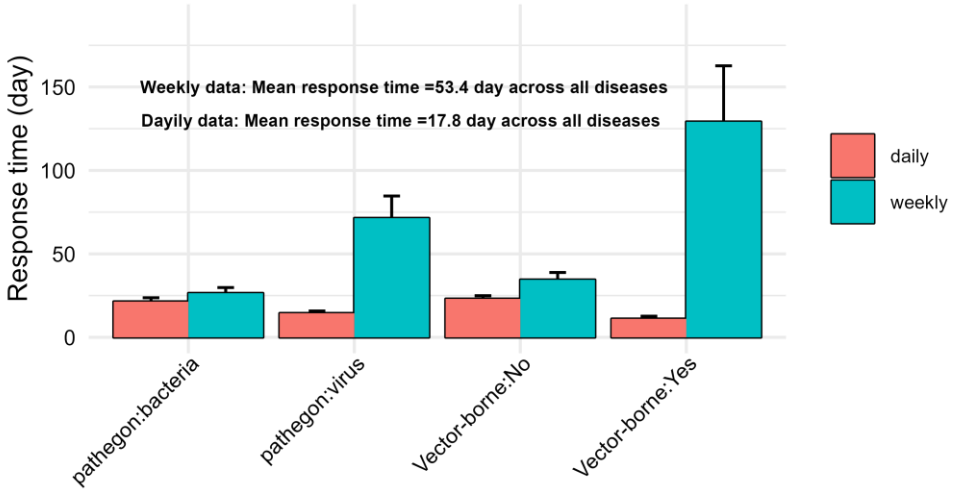
**Extended Data Fig. 2**. **The lead time estimated from the time-to-event analysis based on daily data**. The gray bar represents the mean, while the associated error bar illustrates the standard error. The lead time, i.e. days between the first instance that survival regression model was significant and the start of the outbreak, was computed by combining the best performing single RI (standard deviation, window length=13 days; Fig 3b) with second highest preforming RI (Hurst exponent, window length=13 days; Fig 3b). The predictor was the tau value ($\tau$) of the RIs. The response variable was a time series of 0 values indicating non-outbreak days ($R_e < 1$) and a single 1 value indicating the hypothetical time point at which the outbreak could be anticipated in advance.



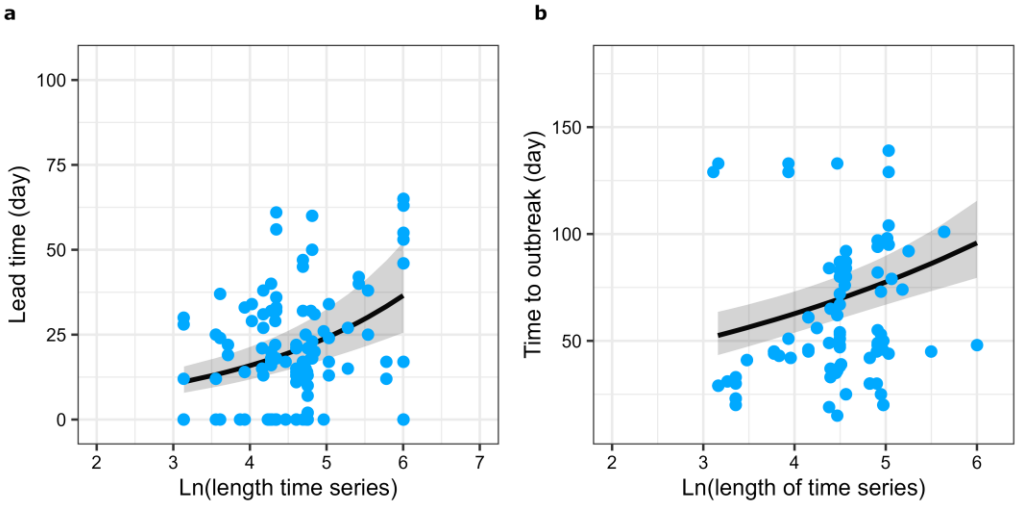
**Extended Data Fig. 3. For the aggregated weekly data, the accuracy of resilience indicators (RIs) as early warning signals (EWSs) of outbreaks across 31 diseases, examined using mean area under the receiver operating characteristic curves (AUC) and cross validation**. All RIs were forced to use the window length that maximized the AUC for each RI (i.e., variable window lengths) (**a,c**) or the window length that maximized the AUC for the most accurate RI (i.e., same window length) (**b,d**). For both of these window length scenarios, we sequentially added either the next best performing single RI (**a,b**) or the next most orthogonal RI (i.e., least correlated based on Principal Component Analysis) to the best performing RI (**c,d**) until we tested up to four total RIs in our models. Importantly, sequentially adding the next most orthogonal RI to the model ensures that the baseline AUC goes up with each RI addition, but it does not guarantee that the mean AUC from the cross validation will continually rise with each added RI. First difference variance: FDR, density ratio: DR, turning point: TP, coefficient of variation: CV, Hurst exponent: HE, autocorrelation at lag 1: AR1, number of outliers: NO, time series acceleration: TAS.



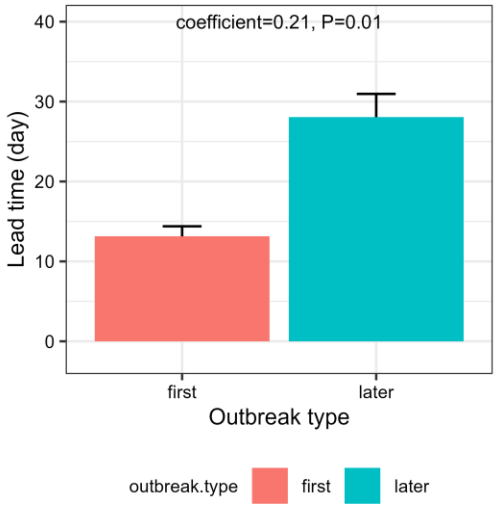
**Extended Data Fig. 4**. **The lead time estimated from RIs approach based on daily and coarse-scale weekly data**. The weekly data was derived by aggregating daily data to a weekly scale through weekly averaging. Apart from differences in data resolution, all other methods for calculating lead time for weekly data are identical to those applied to daily data using the RIs approach (see Methods). The gray bar represents the mean, while the associated error bar illustrates the standard error. The lead time for weekly data was computed by combining the best performing single RI (First differenced variance, window length=5 weeks; Extended Data Fig. 3a) with second highest preforming RI (Turning point, window length=5 weeks; Extended Data Fig. 3a).



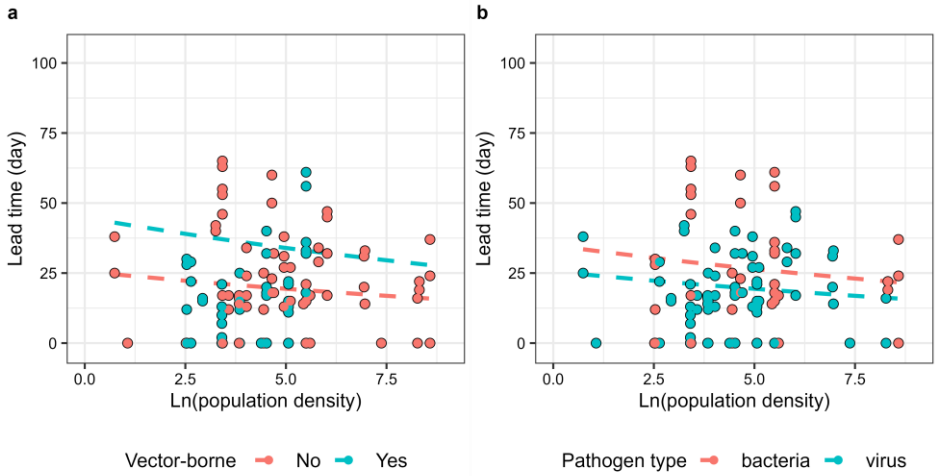
**Extended Data Fig. 5**. **Effects of time series length on (a) lead time and (b) time to outbreak occurrence**. Statistical results are presented in Table S3 and S4. The bold lines represent the significant best-fit trendlines, while the gray bands denote the corresponding 95% confidence bands. For the effects of the remaining variables on these response variables, refer to Figure 4 and 5.



**Extended Data Fig. 6. The lead time in responses to outbreak type**. First outbreaks in a time series have shorter lead times (mean ±s.e.) than subsequent outbreaks, because the arbitrary start of a time series to the first outbreak is shorter, on average, than the distance between two outbreaks in a time series. Statistical results can be found in Table S3.



537 **Extended Data Fig. 7**. **Effects of human population density (a, b) on the amount of lead time**
538 **before disease outbreaks across 31 diseases**. Dashed lines indicate non-significant trendlines.
539 Details of statistical results are available in Table S3. For the effects of the remaining variables—
540 temperature, precipitation, HDI, pathogen incubation time, two parasite-pathogen traits, and
541 outbreak type—on lead time, refer to Figure 4.




542
543
544

**Supplementary Information 1**

The method developed by Cori et al. (2013) defined the $R_e$ as the ratio of the number of new infections at time t ($I_t$) to the total infectiousness of individuals up to that point[58].

$$R_e = \frac{I_t}{\sum_{s=1}^{t} I_{t-s} w_s}$$

The total infectiousness is expressed as $\sum_{s=1}^{t} I_{t-s} w_s$, representing the sum of infection incidences from previous time steps (up to t-1), weighted by $w_s$. The $w_s$ is a probability distribution (hence summing to 1) describing the average infectiousness profile after infection, which can be directly estimated using Markov Chain Monte Carlo (dic.fit.mcmc function in coarseDataTools package)[73]. To account for the uncertainty in the distribution of $w_s$, we assume that $w_s$ follows a Gamma distribution, with its mean ($\mu$) and standard deviation ($\sigma$) varying according to truncated normal distributions[43]. Following the methodology of Cori et al. (2013)[43], we generate 1,000 pairs of mean and standard deviation, denoted as $(\sigma, \mu)^1, \ldots., (\sigma, \mu)^{1000}$, with the constraint that $\sigma < \mu$ in each pair. This constraint ensures that the Gamma probability density function of $w_s$ is null at t=0. For each pair $(\sigma, \mu)^k$, we draw 1,000 realizations of $R_e$ from its posterior distribution, conditional on the corresponding $(\sigma, \mu)^k$. This process yields a total sample size of 1,000,000 (1,000 × 1,000) from the joint posterior distribution of $R_e$. The posterior mean, standard deviation, and the 0.025, 0.05, 0.25, 0.5, 0.75, 0.95, and 0.975 quantiles of $R_e$ are eventually obtained from this sample.

Unlike methods that require fitting mechanistic dynamic models to incidence or case data for estimating the $R_e$, the approaches developed by Cori et al. (2013)[58], along with the other two approaches developed by Wallinga and Teunis (2012)[74], and Bettencourt and Ribeiro (2008)[75], are generic and commonly used model-free method that rely on case incidence data for $R_e$ estimation. However, among the three, the method by Cori et al. (2013) had higher accuracy and addressed the limitations present in the other two approaches. First, Cori et al. (2013) uses data from before time *t*, while Wallinga and Teunis (2012) require data from after time *t*[58]. This reliance makes the latter approach conceptually and methodologically less suitable for near real-time estimation, particularly when the objective is to assess the impact of interventions or other external factors on transmission dynamics at a specific point in time[58,59]. Second, the method proposed by Cori et al. (2013) can accommodate any positive, discrete generation interval distribution. In contrast, the approach by Bettencourt and Ribeiro (2008) implicitly assumes that the generation interval follows an exponential distribution-as it is derived from a Susceptible-

578 Infectious-Recovered (SIR) model, which restricts the applicability of the latter method[58,59].
579 Third, the method developed by Cori et al. (2013) demonstrates greater robustness when
580 applied to data with small aggregation time steps (e.g., daily data), enabling the generation
581 of smoother estimates[59].
582

## Supplementary figures

**Figure S1.** Across all 31 diseases, RIs accuracy (mean AUC from 100 runs of 5-fold cross validation) varies as a function of window lengths of 7 to 21 days. A higher AUC value (area under the curve) indicates a higher ability of a binary classifier to properly classify true positive and negative outcomes, thus representing higher accuracy for RIs. The red line represents the fit from a generalized additive model (gam function from mgcv package of R). The blue line represents a linear or quadratic fit from a linear regression model (lm function from the base package of R). The solid lines indicate significant associations, whereas dashed lines indicate non-significant associations.

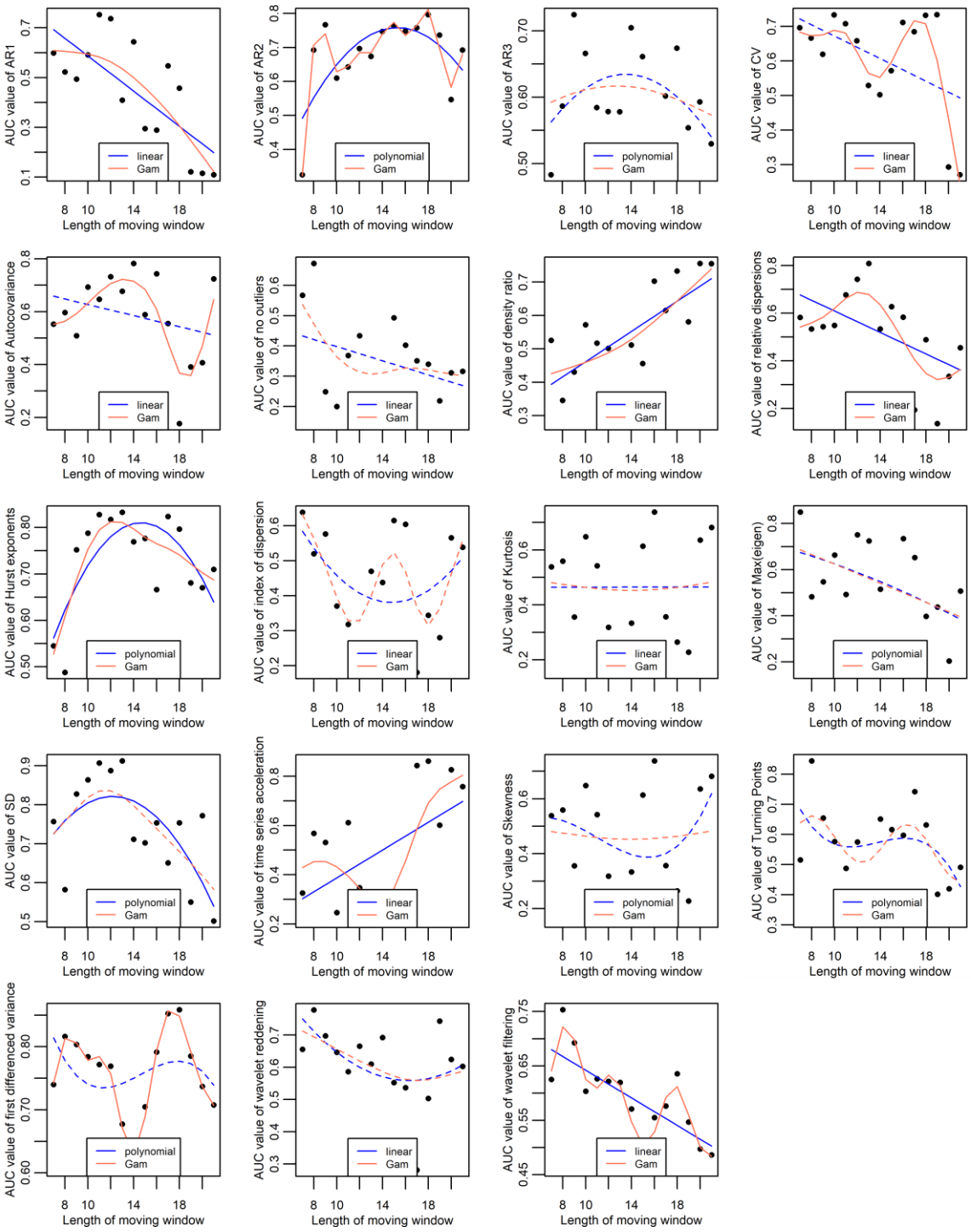
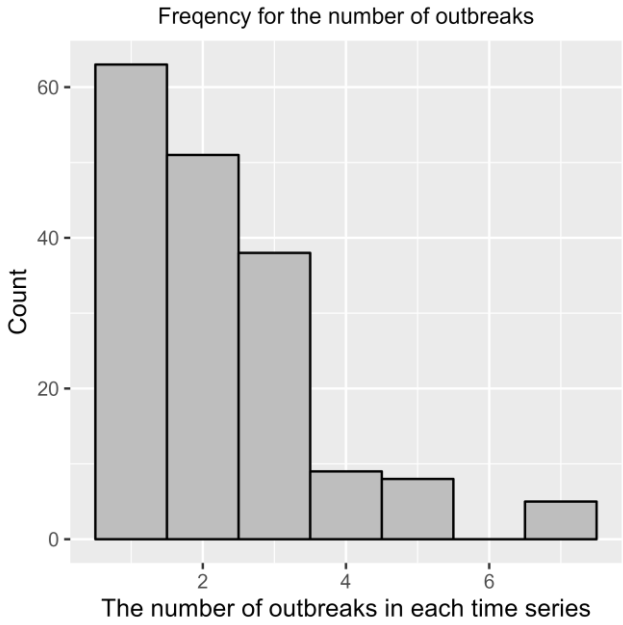


**Figure S2**. Frequency of the number of outbreaks per time series with the average number of secondary infections caused by a single infected individual at a given time point $R_e \geq 1$.



**Supplementary Information 2**

For climate anomalies, precipitation anomalies exhibited a unimodal relationship with both lead time and time to outbreak occurrence across all pathogens, peaking near the observed historical mean monthly precipitation (Figure S3b, S4b). Similarly, temperature anomalies also showed a unimodal relationship with both lead time and time to outbreak occurrence for vector-borne pathogens, with peaks occurring near the mean (Figure S3a, S4c). In contrast, for non-vector-borne pathogens, temperature anomalies were positively associated with both lead time and time to outbreak occurrence (Fig. S3a, S4c). Greater deviations from optimal climatic conditions, leading to shorter lead times and earlier outbreak occurrences, could be attributed to the well-documented unimodal effects of temperature and precipitation on the viral and vector-borne pathogen transmission dynamics[36,37]. Evidence have documented higher temperature anomaly increased the non-vector-borne pathogens, such as respiratory[76] and diarrheal diseases[50].

**Figure S3. The predicted values depicting the interaction between precipitation anomalies and pathogen type (vector-borne or not) and the interaction between temperature anomalies and pathogen type (bacteria or virus) on lead time**. The bold lines and gray bands represent best-fit lines and 95% confidence bands, respectively. Details of statistical model and results are available in Table S7.

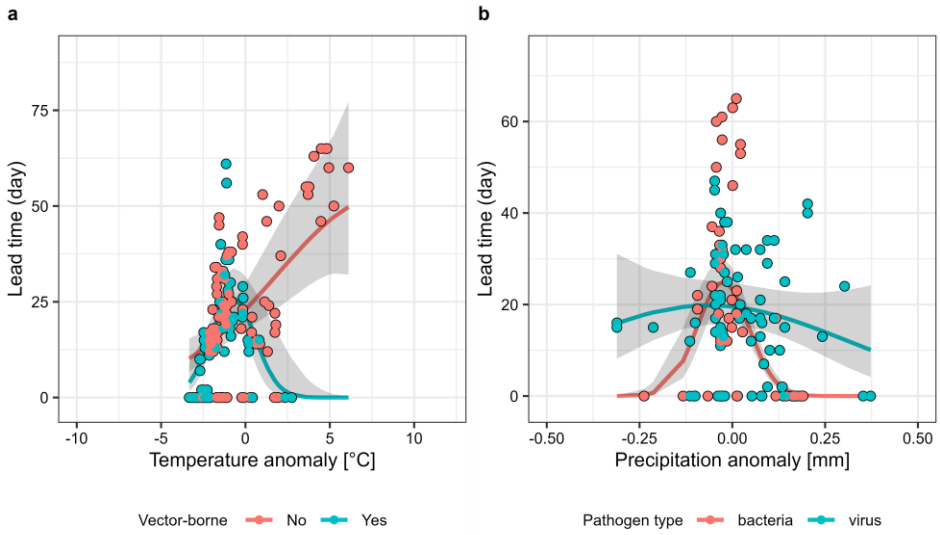
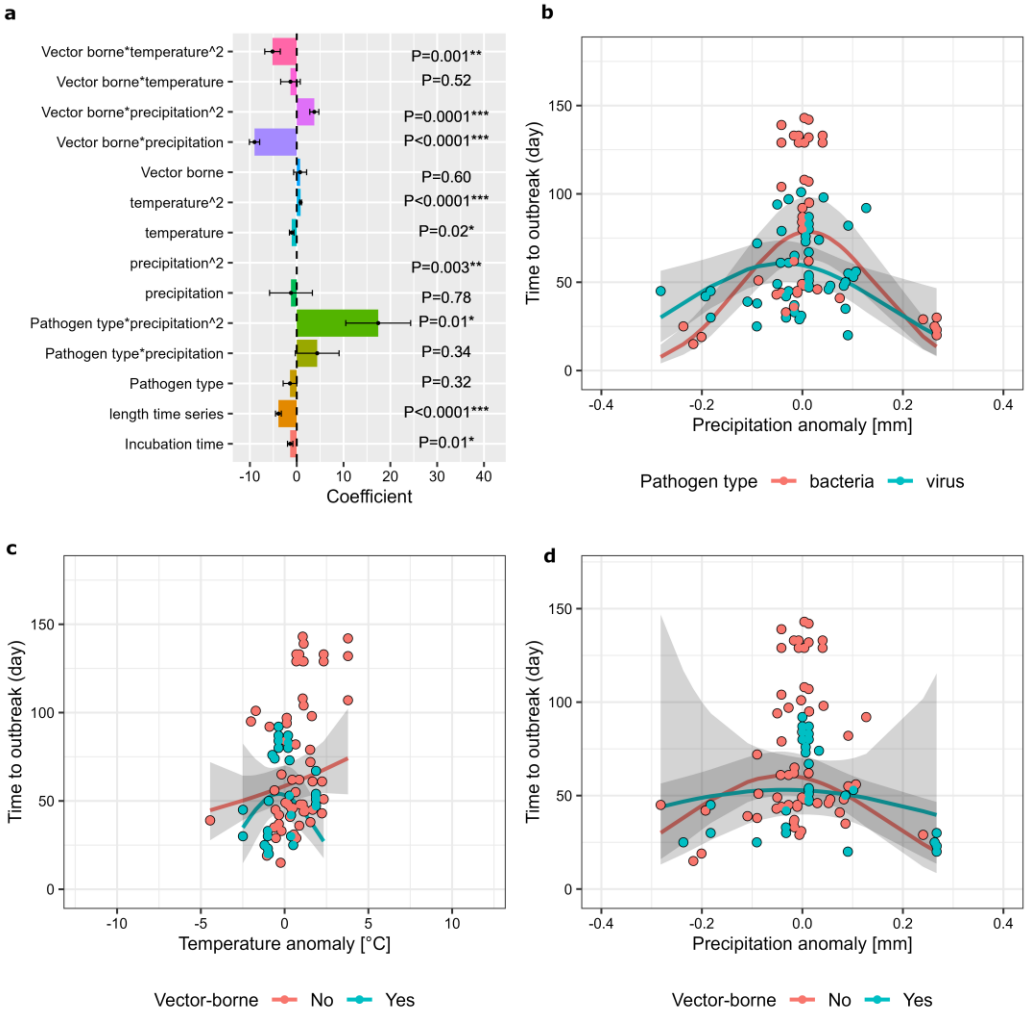


**Figure S4. Coefficients ± s.e (a) and plots of significant predictors (b-d) from a Cox proportional hazards mixed model examining the effects of various factors on time to outbreak occurrence**. **a**, A positive coefficient signifies an increased hazard of outbreak occurrence, and consequently a reduced time until an outbreak. A significant quadratic coefficient indicates a second-order polynomial function between a predictor and hazard of outbreak occurrence. **b**, the bold lines and gray bands depict best-fit lines, 95% confidence intervals, respectively, derived from the predicted values of the final statistical model that is available in Table S6. Significant effects are indicated by *P <0.05, **P < 0.01, ***P < 0.001.



## Supplementary tables

**Table S1**. All 31 diseases used in the analyses, their corresponding regions, and traits of their pathogens and time series.

| | Disease | Location | Zoonotic | Vector-borne | Pathogen type | Length of time series (mean ± s.e) |
|---|---|---|---|---|---|---|
| 1 | Anthrax | California, US<br>Muchinga, Zambia<br>Shaanxi, China<br>West Bengal, India | Yes | No | bacteria | 80±20 |
| 2 | Brucellosis | Kyustendil, Bulgaria<br>Qatar | Yes | No | bacteria | 85±15 |
| 3 | Campylobacteriosis | Australia<br>Greece<br>Hastings, New Zealand<br>Northumberland, UK<br>Kentucky, US<br>Norway<br>Barcelona, Spain<br>Stockholm, Sweden<br>Las Vegas, US<br>Adelaide, Australia<br>Anchorage, US | Yes | No | bacteria | 75±8 |
| 4 | Chikungunya | Mauritius<br>Champasak, Laos<br>Montpellier, France<br>Guangdong, China<br>Kampong Speu, Cambodia<br>Ravenna, Italy<br>West Bengal, India<br>Chennai, India | No | Yes | virus | 86±15 |
| 5 | Cholera | Vietnam<br>Sekong, Laos<br>Orissa, India<br>Haiti | No | No | virus | 64±10 |
| 6 | Dengue | Madeira, Portugal | No | Yes | virus | 76±12 |
| 7 | Ebola | Kibale, Uganda<br>Kasai-Occidental, Congo | Yes | No | virus | 112±30 |
| 8 | Gastroenteritis | Belgium<br>Pisco, Peru<br>San Vicente, El Salvador<br>Finland | Yes | No | Multiple types | 74±20 |
| 9 | Lassa Fever | Jos, Nigeria | Yes | No | virus | 82±0 |
| 10 | Leptospirosis | Illinois, US<br>Guyana<br>Fiji<br>Martinique, France<br>Peru<br>Malaysia<br>Nicaragua | Yes | Yes | protozoan | 86±16 |

**Table S1 (continued)**.

| | Disease | Location | Zoonotic | Vector-borne | Pathogen type | Length of time (mean ± s.e) |
|---|---|---|---|---|---|---|
| 11 | Malaria | Sierra Leone<br>Amhara, Ethiopia<br>Montserrado, Liberia<br>Aswan, Egypt<br>Honduras | No | Yes | protozoan | 70±6 |
| 12 | Marburg | Uganda<br>Kabale, Uganda | Yes | No | virus | 80±24 |
| 13 | Measles | California, US<br>Vancouver, US<br>Federated States of Micronesia<br>Casa Grande, US<br>Ohio, US<br>Mayuge, Uganda<br>Samoa<br>Fiji<br>Somali, Ethiopia<br>San Diego, US<br>New York, US | No | No | virus | 118±26 |
| 14 | Middle East Respiratory Syndrome (MERS-CoV) | South Korea<br>Jordan | Yes | No | virus | 76±12 |
| 15 | Monkeypox | Équateur, Congo<br>US<br>Impfondo, Congo<br>Liberia<br>Congo | Yes | No | virus | 142±28 |
| 16 | Nipah | Bangladesh<br>Philippines | Yes | No | virus | 89±40 |
| 17 | Novel Influenza A | Australia<br>Mexico<br>Kyrgyzstan<br>Jalisco, Mexico<br>Peru<br>Gelderland, Netherlands<br>Hong Kong<br>Turkey<br>Shanghai, China<br>Manitoba, Canada | Yes | No | virus | 106±30 |
| 18 | Pertussis | Spain | No | No | bacteria | 66±0 |
| 19 | Plague | Maharashtra, India<br><br>Madagascar<br>Eastern, Zambia<br>Colorado, US | Yes | Yes | bacteria | 87±5 |

**Table S1 (continued)**.

| | Disease | Location | Zoonotic | Vector-borne | Pathogen type | Length of time (mean ± s.e) |
|---|---|---|---|---|---|---|
| 20 | Norovirus | New Zealand<br>Washington, US<br>Michigan, US<br>Germany<br>Michigan, US<br>Alberta, Canada<br>Beijing, China<br>New Jersey, US<br>France<br>Malta<br>Singapore<br>Saint Thomas, United States Virgin Islands<br>UK<br>Israel<br>Wisconsin, US<br>Wuxi, China<br>California, US<br>Netherlands<br>Japan<br>Spain<br>El Paso, US<br>Colorado, US<br>Ireland | No | No | virus | 77±4 |
| 21 | Polio | Madagascar<br>Tajikistan | No | No | virus | 159±6 |
| 22 | Q Fever | Hungary<br>Nova Scotia, Canada<br>Slovenia<br>Iraq<br>UK<br>Israel<br>Bulgaria | Yes | Yes | bacteria | 64±12 |
| 23 | Rift Valley Fever | South Africa<br>Uganda<br>Madagascar<br>Kenya<br>Mayotte, France | Yes | Yes | virus | 135±57 |
| 24 | Rubella | Kenya<br>Zimbabwe | No | No | virus | 62±4 |

642 **Table S1 (continued)**.

| | Disease | Location | Zoonotic | Vector-borne | Pathogen type | Length of time (mean ± s.e) |
|---|---|---|---|---|---|---|
| 25 | Salmonellosis | Germany<br>Rhode Island, US<br>District of Columbia, US<br>Poitier, France<br>South Korea<br>UK<br>Sousse, Tunisia<br>Madrid, Spain | Yes | No | bacteria | 73±21 |
| 26 | Shiga Toxin-Producing E. Coli (STEC) | Michigan, US<br>Germany<br>UK<br>Kanazawa, Ishikawa, Japan<br>Iceland<br>Netherlands<br>Ireland | Yes | No | bacteria | 225±30 |
| 27 | Shigellosis | Kirkenes, Norway<br>Sichuan, China<br>North Carolina, US<br>Morobe, Papua New Guinea<br>Sweden<br>Virginia, US<br>Honolulu, US<br>Pemagatsel, Bhutan<br>Norway<br>Salzburg, Austria<br>Austria | No | No | bacteria | 79±16 |
| 28 | Tularemia | Germany<br>Utah, US<br>France | Yes | Yes | bacteria | 83±12 |
| 29 | West Nile Virus | Michigan, US<br>Thessaloniki, Greece<br>Romania<br>Italy | Yes | Yes | virus | 145±16 |
| 30 | Yellow Fever | Sudan<br>San Pedro, Paraguay<br>São Paulo, Brazil | Yes | Yes | virus | 92±16 |
| 31 | Zika | San Andrés, Colombia<br>Girardot, Colombia | No | Yes | virus | 120±25 |

643
644

**Table S2**. List of 19 resilience indicators (RIs) used to detect critical transitions, and their type, abbreviation, and reference.

| Indicators | Type | Abbreviation | References |
|---|---|---|---|
| 1. Autocorrelation at lag 1 | Classic indicators | AR1 | [5,6,17] |
| 2. Autocorrelation at lag 2 | | AR2 | |
| 3. Autocorrelation at lag 3 | | AR3 | |
| 4. standard deviation | | SD | |
| 5. Skewness | | SK | |
| 6. Kurtosis | | KU | |
| 7. Coefficient of variation | | CV | |
| 8. First differenced variance | | FDV | |
| 9. Autocovariance | | AU | |
| 10. Index of dispersion | | ID | |
| 11. Density ratio | | DR | |
| 12. Wavelet filtering | Wavelet spectrum analysis to handle the seasonality | WF | [77] |
| 13. Wavelet reddening | | WR | |
| 14. Relative dispersion | Automatic Feature Engineering for Forecasting | RD | [78–80] |
| 15. Hurst exponent | | HE | |
| 16. Time series acceleration | | TSA | |
| 17. Number of outliers | | NO | |
| 18. Turning Points | | TP | |
| 19.The absolute value of dominant eigenvalue (Max(eigen)) | Empirical dynamic modelling | ME | [81] |

**Table S3**. Results from a generalized linear mixed model examining the main effects of temperature, precipitation, human development index (HDI), pathogen incubation time, human population density, time series length, two pathogen traits (vector-borne or not; virus vs. bacterial pathogen), outbreak type (first or subsequent outbreak in a time series), and the pairwise interactions between the two pathogen traits and temperature, precipitation, HDI and human population density on lead time. In the model, we treated regions nested within disease types and types of RIs (e.g. Hurst exponent) as random intercepts. The parasite taxa were limited to viruses and bacteria because our dataset had no fungi or helminths (Table S1) and the sample size for protozoa was too small to justify analysis (only two parasite species). A significant quadratic coefficient indicates a second-order polynomial function between a predictor and lead time. The blue text denotes non-significant predictors that were removed from the final model through a backward sequential selection. ***$P$< 0.001, **$P$<0.01, *$P$<0.05. The model's marginal $R^2$ is 0.12, and the conditional $R^2$ is 0.15.

| **Independent variable** | **coefficient** | **s.e** | **Z** | **P value** |
|---|---|---|---|---|
| **Final model** | | | | |
| (Intercept) | 0.01 | 0.41 | 0.03 | 0.97 |
| Temperature | 4.86 | 0.84 | 5.76 | **<0.0001*** ** |
| Temperature^2 | -1.84 | 0.75 | -2.44 | **0.01*** |
| Precipitation | -5.63 | 1.35 | -4.16 | **<0.0001*** ** |
| Precipitation^2 | 0.02 | 1.13 | 0.02 | 0.98 |
| HDI (low) | 0.29 | 0.17 | 1.71 | 0.08 |
| HDI (medium) | 0.31 | 0.13 | 2.38 | **0.02*** |
| Incubation time | 0.47 | 0.10 | 4.55 | **<0.0001*** ** |
| Length time series | 0.47 | 0.07 | 7.01 | **<0.0001*** ** |
| Pathogen type | -0.90 | 0.15 | -6.16 | **<0.0001*** ** |
| Vector borne | -4.36 | 0.85 | -5.14 | **<0.0001*** ** |
| Outbreak type | 0.23 | 0.09 | 2.49 | **0.01*** |
| Temperature: Pathogen type | -6.61 | 1.69 | -3.92 | **<0.0001*** ** |
| Temperature^2: Pathogen type | -5.85 | 2.18 | -2.68 | **0.007** ** |
| Precipitation: Pathogen type | 4.31 | 1.70 | 2.53 | **0.01*** |
| Precipitation^2: Pathogen type | -4.69 | 1.88 | -2.50 | **0.01*** |
| Temperature: Vector borne | 68.40 | 8.90 | 7.68 | **<0.0001*** ** |
| Temperature^2: Vector borne | -88.04 | 15.31 | -5.75 | **<0.0001*** ** |
| HDI (low): Vector borne | 2.72 | 1.03 | 2.65 | **0.008** ** |
| HDI (medium): Vector borne | 1.12 | 0.72 | 1.56 | 0.12 |
| Precipitation: Vector borne | -11.30 | 3.93 | -2.87 | **0.004** ** |
| Precipitation^2: Vector borne | -10.61 | 2.69 | -3.94 | **<0.0001*** ** |
| **Non-significant predictors below were removed from the final model through a backward sequential selection.** | | | | |
| Population density | | | | |
| Population density: Vector borne | | | | |
| Population density: Pathogen type | | | | |

**Table S4.** Results from a mixed effect Cox proportional hazard model examining the main effects of temperature, precipitation, human development index (HDI), pathogen incubation time, human population density, time series length, two pathogen traits (vector-borne or not; virus vs. bacterial pathogen), outbreak type (first or subsequent outbreak in a time series) and the pairwise interactions between the two pathogen traits and temperature, precipitation, HDI and human population density on the timing of outbreak occurrence. In the model, we treated regions nested within disease types and type of RIs (e.g. Hurst exponent) as random intercepts. The parasite taxa were limited to viruses and bacteria because our dataset had no fungi or helminths (Table S1) and the sample size for protozoa was too small to justify analysis (only two parasite species). A positive coefficient signifies an increased hazard of outbreak occurrence, and consequently a reduced time until an outbreak. A significant quadratic coefficient indicates a second-order polynomial function between a predictor and hazard of outbreak occurrence. The blue text denotes non-significant predictors that were removed from the final model through a backward sequential selection. Model's Nagelkerke $R^2$ is 0.26. ***$P$< 0.001, **$P$<0.01, *$P$<0.05.

| Independent variable | coefficient | s.e | Z | P value |
|---|---|---|---|---|
| **Final model** | | | | |
| Temperature^2 | 7.36 | 1.44 | 5.10 | **<0.0001*** ** |
| Temperature | -1.29 | 1.44 | -0.90 | 0.36 |
| Precipitation^2 | 2.97 | 0.67 | 4.43 | **<0.0001*** ** |
| Precipitation | 0.14 | 0.76 | 0.18 | 0.85 |
| HDI (low) | 0.95 | 2.09 | 0.46 | 0.64 |
| HDI (medium) | 0.12 | 2.04 | 0.06 | 0.96 |
| Length time series | -4.21 | 0.83 | -5.05 | **<0.0001*** ** |
| Pathogen type | 5.06 | 1.94 | 2.60 | **0.009** ** |
| Vector borne | -0.45 | 2.42 | -0.19 | 0.85 |
| Temperature^2: Pathogen type | -8.01 | 1.56 | -5.15 | **<0.0001*** ** |
| Temperature: Pathogen type | -3.20 | 1.70 | -1.88 | 0.06 |
| Precipitation: Vector borne | -5.43 | 1.05 | -5.18 | **<0.0001*** ** |
| Precipitation^2: Vector borne | -2.30 | 0.93 | -2.47 | **0.01* ** |
| HDI (low): Vector borne | 14.60 | 4.09 | 3.57 | **0.0003*** ** |
| HDI (medium): Vector borne | 12.29 | 3.92 | 3.14 | **0.001** ** |
| **Non-significant predictors below were removed from the final model through a backward sequential selection.** | | | | |
| Population density | | | | |
| Population density: Pathogen type | | | | |
| Population density: Vector borne | | | | |
| Incubation time | | | | |
| Outbreak type | | | | |
| Precipitation: Pathogen type | | | | |
| Precipitation^2: Pathogen type | | | | |
| Temperature^2: Vector borne | | | | |
| Temperature: Vector borne | | | | |

**Table S5**. Given that there is currently no good way to visualize the results of a Cox proportional hazards (coxme) model, we replicated the final coxme model (Table S4) as a generalized linear mixed model (glmmTMB function in glmmTMB package) with time to an outbreak as a negative binomial dependent variable, shown below. Given that the statistical outcomes of the two models had similar coefficients and significance levels, we used the glmmTMB to generate figures using the predict function in the ggplot2 package in R.

| Independent variable | coefficient | s.e | Z | P value |
|---|---|---|---|---|
| Temperature^2 | 1.27 | 0.46 | 2.75 | **0.006**** |
| Temperature | 3.73 | 0.62 | 6.03 | **<0.0001***** |
| Precipitation^2 | -3.63 | 0.62 | -5.85 | **<0.0001***** |
| Precipitation | -0.26 | 0.63 | -0.42 | 0.67 |
| HDI (low) | -0.03 | 0.11 | -0.27 | 0.79 |
| HDI (medium) | 0.13 | 0.12 | 1.15 | 0.25 |
| Length time series | 0.21 | 0.04 | 4.77 | **<0.0001***** |
| Pathogen type | -1.10 | 0.08 | -1.14 | **0.04*** |
| Vector borne | -0.11 | 0.18 | -0.61 | 0.54 |
| Temperature^2: Pathogen type | -3.01 | 1.12 | -2.68 | **0.007**** |
| Temperature: Pathogen type | -2.32 | 1.05 | -2.12 | 0.05 |
| Precipitation: Vector borne | 3.10 | 1.26 | 2.46 | **0.01*** |
| Precipitation^2: Vector borne | 1.86 | 0.91 | 2.05 | **0.04*** |
| HDI (low): Vector borne | -0.73 | 0.22 | -3.40 | **0.0006***** |
| HDI (medium): Vector borne | 0.14 | 0.25 | 0.15 | **0.03*** |

**Table S6.** Results from a mixed effect Cox proportional hazard model examining the main effects of temperature anomaly, precipitation anomaly, human development index (HDI), pathogen incubation time, human population density, time series length, two pathogen traits (vector-borne or not; virus vs. bacterial pathogen), outbreak type (first or subsequent outbreak in a time series) and the pairwise interactions between the two pathogen traits and temperature anomaly, precipitation anomaly, HDI and human population density on the timing of outbreak occurrence. In the model, we treated regions nested within disease types and type of RIs (e.g. Hurst exponent) as random intercepts. The parasite taxa were limited to viruses and bacteria because our dataset had no fungi or helminths (Table S1) and the sample size for protozoa was too small to justify analysis (only two parasite species). A positive coefficient signifies an increased hazard of outbreak occurrence, and consequently a reduced time until an outbreak. A significant quadratic coefficient indicates a second-order polynomial function between a predictor and hazard of outbreak occurrence. The blue text denotes non-significant predictors that were removed from the final model through a backward sequential selection. Model's Nagelkerke $R^2$ is 0.25. ***$P$< 0.001, **$P$<0.01, *$P$<0.05.

| Independent variable | coefficient | s.e | Z | P value |
|---|---|---|---|---|
| **Final model** | | | | |
| Temperature^2 | 0.83 | 0.20 | 4.18 | **<0.0001*** ** |
| Temperature | -1.08 | 0.48 | -2.23 | **0.02*** |
| Precipitation^2 | -20.02 | 6.84 | -2.93 | **0.003** ** |
| Precipitation | -1.23 | 4.58 | -0.27 | 0.79 |
| Incubation time | -1.40 | 0.57 | -2.47 | **0.01*** |
| Length time series | -3.91 | 0.62 | -6.31 | **<0.0001*** ** |
| Pathogen type | -1.44 | 1.46 | -0.99 | 0.32 |
| Vector borne | 0.72 | 1.39 | 0.52 | 0.60 |
| Precipitation: Pathogen type | 4.37 | 4.67 | 0.93 | 0.35 |
| Precipitation^2: Pathogen type | 17.41 | 6.93 | 2.51 | **0.01*** |
| Temperature^2: Vector borne | -5.18 | 1.66 | -3.12 | **0.001** ** |
| Temperature: Vector borne | -1.34 | 2.09 | -0.64 | 0.52 |
| Precipitation ^2: Vector borne | 3.75 | 0.97 | 3.89 | **<0.0001*** ** |
| Precipitation: Vector borne | -9.02 | 1.10 | -8.22 | **<0.0001*** ** |
| **Non-significant predictors below were removed from the final model through a backward sequential selection.** | | | | |
| HDI (low) | | | | |
| HDI (medium) | | | | |
| Population density | | | | |
| HDI (low): Vector borne | | | | |
| HDI (medium): Vector borne | | | | |
| Precipitation^2: Pathogen type | | | | |
| Precipitation: Pathogen type | | | | |
| Population density: Pathogen type | | | | |
| Population density: Vector borne | | | | |

**Table S7**. To climate anomaly, results from a generalized linear mixed model examining the main effects of temperature anomaly, precipitation anomaly, human development index (HDI), pathogen incubation time, human population density, time series length, two pathogen traits (vector-borne or not; virus vs. bacterial pathogen), outbreak type (first or subsequent outbreak in a time series), and the pairwise interactions between the two pathogen traits and temperature anomaly, precipitation anomaly, HDI and human population density on lead time. In the model, we treated regions nested within disease types and types of RIs (e.g. Hurst exponent) as random intercepts. The parasite taxa were limited to viruses and bacteria because our dataset had no fungi or helminths (Table S1) and the sample size for protozoa was too small to justify analysis (only two parasite species). A significant quadratic coefficient indicates a second-order polynomial function between a predictor and lead time. The blue text denotes non-significant predictors that were removed from the final model through a backward sequential selection. ***$P$< 0.001, **$P$<0.01, *$P$<0.05. The model's marginal $R^2$ is 0.08, and the conditional $R^2$ is 0.12.

| Independent variable | coefficient | s.e | Z | P value |
|---|---|---|---|---|
| **Final model** | | | | |
| Intercept | 1.11 | 0.37 | 3.05 | **0.002**** |
| Temperature | 4.83 | 0.85 | 5.67 | **<0.0001***** |
| Temperature^2 | -0.77 | 0.60 | -1.28 | 0.20 |
| Precipitation | -18.28 | 2.10 | -8.69 | **<0.0001***** |
| Precipitation^2 | -26.63 | 3.73 | -7.14 | **<0.0001***** |
| HDI (low) | 0.62 | 0.19 | 3.27 | **0.001**** |
| HDI (medium) | 0.61 | 0.17 | 3.50 | **0.0004***** |
| Length time series | 0.13 | 0.08 | 1.70 | 0.08 |
| Pathogen type | 1.15 | 0.18 | 6.57 | **<0.0001***** |
| Vector borne | -0.73 | 0.33 | -2.22 | **0.03*** |
| Precipitation: Pathogen type | 17.25 | 2.15 | 8.02 | **<0.0001***** |
| Precipitation^2: Pathogen type | 25.70 | 3.76 | 6.83 | **<0.0001***** |
| Temperature: Vector borne | -26.58 | 8.53 | -3.12 | **0.001**** |
| Temperature^2: Vector borne | -16.96 | 5.47 | -3.10 | **0.001**** |
| HDI (low): Vector borne | -0.99 | 0.29 | -3.43 | **0.0005***** |
| HDI (medium): Vector borne | -0.64 | 0.27 | -2.38 | **0.01*** |
| **Non-significant predictors below were removed from the final model through a backward sequential selection.** | | | | |
| Population density | | | | |
| Incubation time | | | | |
| Temperature^2: Pathogen type | | | | |
| Temperature: Pathogen type | | | | |
| Precipitation: Vector borne | | | | |
| Precipitation^2: Vector borne | | | | |
| Population density: Pathogen type | | | | |
| Population density: Vector borne | | | | |

## References


1. Delecroix, C. *et al.* The potential of resilience indicators to anticipate infectious disease outbreaks, a systematic review and guide. *PLOS Glob Public Health* **3**, e0002253 (2023).
2. Chaves, L. F. & Pascual, M. Comparing Models for Early Warning Systems of Neglected Tropical Diseases. *PLOS Neglected Tropical Diseases* **1**, e33 (2007).
3. Grassly, N. C., Shaw, A. G. & Owusu, M. Global wastewater surveillance for pathogens with pandemic potential: opportunities and challenges. *The Lancet Microbe* **6**, (2025).
4. Scheffer, M. *Critical Transitions in Nature and Society*. (Princeton University Press, 2009). doi:10.2307/j.ctv173f1g1.
5. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
6. Scheffer, M. *et al.* Anticipating Critical Transitions. *Science* **338**, 344–348 (2012).
7. Gorban, A. N., Smirnova, E. V. & Tyukina, T. A. Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications* **389**, 3193–3217 (2010).
8. Ramos, O. Criticality in earthquakes. Good or bad for prediction? *Tectonophysics* **485**, 321–326 (2010).
9. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences* **105**, 1786–1793 (2008).
10. Kambhu, J., Krishnan, N. & Weidman, S. New directions for understanding systemic risk: a report on a conference cosponsored by the Federal Reserve Bank of New York and the National Academy of Sciences. *Economic Policy Review* **13**, (2007).
11. May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451**, 893–894 (2008).
12. Diks, C., Hommes, C. & Wang, J. Critical slowing down as an early warning signal for financial crises? *Empir Econ* **57**, 1201–1228 (2019).
13. Downey, S. S., Haas, W. R. & Shennan, S. J. European Neolithic societies showed early warning signals of population collapse. *Proceedings of the National Academy of Sciences* **113**, 9751–9756 (2016).
14. Neuman, Y., Nave, O. & Dolev, E. Buzzwords on their way to a tipping-point: A view from the blogosphere. *Complexity* **16**, 58–68 (2011).
15. Saavedra, S., Duch, J. & Uzzi, B. Tracking Traders' Understanding of the Market Using e-Communication Data. *PLOS ONE* **6**, e26705 (2011).

16. Scheffer, M. Anticipating societal collapse; Hints from the Stone Age. *Proceedings of the National Academy of Sciences* **113**, 10733–10735 (2016).

17. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).

18. Scheffer, M., Hosper, S. H., Meijer, M.-L., Moss, B. & Jeppesen, E. Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution* **8**, 275–279 (1993).

19. Trefois, C., Antony, P. M., Goncalves, J., Skupin, A. & Balling, R. Critical transitions in chronic disease: transferring concepts from ecology to systems medicine. *Current Opinion in Biotechnology* **34**, 48–55 (2015).

20. Venegas, J. G. *et al.* Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* **434**, 777–782 (2005).

21. Kramer, M. A. *et al.* Human seizures self-terminate across spatial scales via a critical transition. *Proceedings of the National Academy of Sciences* **109**, 21116–21121 (2012).

22. Litt, B. *et al.* Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* **30**, 51–64 (2001).

23. McSharry, P. E., Smith, L. A. & Tarassenko, L. Prediction of epileptic seizures: are nonlinear methods relevant? *Nat Med* **9**, 241–242 (2003).

24. Bury, T. M. *et al.* Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences* **118**, e2106140118 (2021).

25. Drake, J. M. *et al.* The statistics of epidemic transitions. *PLOS Computational Biology* **15**, e1006917 (2019).

26. Brett, T. S. *et al.* Anticipating epidemic transitions with imperfect data. *PLoS Comput Biol* **14**, e1006204 (2018).

27. Brett, T. S., Drake, J. M. & Rohani, P. Anticipating the emergence of infectious diseases. *Journal of The Royal Society Interface* **14**, 20170115 (2017).

28. O'Regan, S. M. & Drake, J. M. Theory of early warning signals of disease emergenceand leading indicators of elimination. *Theor Ecol* **6**, 333–357 (2013).

29. O'Regan, S. M., Lillie, J. W. & Drake, J. M. Leading indicators of mosquito-borne disease elimination. *Theor Ecol* **9**, 269–286 (2016).

30. Harris, M. J., Hay, S. I. & Drake, J. M. Early warning signals of malaria resurgence in Kericho, Kenya. *Biology Letters* **16**, 20190713 (2020).

31. O'Brien, D. A. & Clements, C. F. Early warning signal reliability varies with COVID-19 waves. *Biology Letters* **17**, 20210487 (2021).

32. Southall, E., Brett, T. S., Tildesley, M. J. & Dyson, L. Early warning signals of infectious disease transitions: a review. *Journal of The Royal Society Interface* **18**, 20210555 (2021).

33. Southall, E., Brett, T. S., Tildesley, M. J. & Dyson, L. Early warning signals of infectious disease transitions: a review. *Journal of The Royal Society Interface* **18**, 20210555 (2021).

34. Jr, D. W. H., Lemeshow, S. & May, S. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. (John Wiley & Sons, 2011).

35. Klein, J. P. & Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. (Springer, New York, NY, 2003). doi:10.1007/b97377.

36. Mordecai, E. A. *et al.* Thermal biology of mosquito-borne disease. *Ecology Letters* **22**, 1690–1708 (2019).

37. Caldwell, J. M. *et al.* Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents. *Nat Commun* **12**, 1233 (2021).

38. Thrall, P. H., Biere, A. & Uyenoyama, M. K. Frequency-Dependent Disease Transmission and the Dynamics of the Silene-Ustilago Host-Pathogen System. *The American Naturalist* **145**, 43–62 (1995).

39. Thrall, P. H., Antonovics, J. & Hall, D. W. Host and Pathogen Coexistence in Sexually Transmitted and Vector-Borne Diseases Characterized by Frequency-Dependent Disease Transmission. *The American Naturalist* **142**, 543–552 (1993).

40. Rudolf, V. H. W. & Antonovics, J. Species Coexistence and Pathogens with Frequency-Dependent Transmission. *The American Naturalist* **166**, 112–118 (2005).

41. Kummu, M., Taka, M. & Guillaume, J. H. A. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Sci Data* **5**, 180004 (2018).

42. Lindsay, S. W. *et al.* Changes in house design reduce exposure to malaria mosquitoes. *Tropical Medicine & International Health* **8**, 512–517 (2003).

800 43. Wilson, A. L. *et al.* The importance of vector control for the control and elimination of vector-borne diseases.
801 *PLOS Neglected Tropical Diseases* **14**, e0007831 (2020).
802 44. Mordecai, E. A. *et al.* Thermal biology of mosquito-borne disease. *Ecology Letters* **22**, 1690–1708 (2019).
803 45. Ratkowsky, D. A., Olley, J., McMeekin, T. A. & Ball, A. Relationship between temperature and growth rate of
804 bacterial cultures. *J Bacteriol* **149**, 1–5 (1982).
805 46. Guzman Herrador, B. R. *et al.* Analytical studies assessing the association between extreme precipitation or
806 temperature and drinking water-related waterborne infections: a review. *Environmental Health* **14**, 29 (2015).
807 47. Yang, K. *et al.* Global Distribution of Outbreaks of Water-Associated Infectious Diseases. *PLOS Neglected Tropical*
808 *Diseases* **6**, e1483 (2012).
809 48. Eber, M. R., Shardell, M., Schweizer, M. L., Laxminarayan, R. & Perencevich, E. N. Seasonal and Temperature-
810 Associated Increases in Gram-Negative Bacterial Bloodstream Infections among Hospitalized Patients. *PLOS ONE* **6**,
811 e25298 (2011).
812 49. Manchal, N., Young, M. K., Castellanos, M. E., Leggat, P. & Adegboye, O. A systematic review and meta-analysis
813 of ambient temperature and precipitation with infections from five food-borne bacterial pathogens. *Epidemiology &*
814 *Infection* **152**, e98 (2024).
815 50. Levy, K., Woster, A. P., Goldstein, R. S. & Carlton, E. J. Untangling the impacts of climate change on waterborne
816 diseases: A systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and
817 drought. *Environ Sci Technol* **50**, 4905–4922 (2016).
818 51. Dai, Y. & Wang, J. Identifying the outbreak signal of COVID-19 before the response of the traditional disease
819 monitoring system. *PLOS Neglected Tropical Diseases* **14**, e0008758 (2020).
820 52. Aiken, E. L. *et al.* Real-time estimation of disease activity in emerging outbreaks using internet search
821 information. *PLOS Computational Biology* **16**, e1008117 (2020).
822 53. Zheng, H., Woodall, W. H., Carlson, A. L. & DeLisle, S. Can long-term historical data from electronic medical
823 records improve surveillance for epidemics of acute respiratory infections? A systematic evaluation. *PLOS ONE* **13**,
824 e0191324 (2018).

54. Lusignan, S. de *et al.* Emergence of a Novel Coronavirus (COVID-19): Protocol for Extending Surveillance Used by the Royal College of General Practitioners Research and Surveillance Centre and Public Health England. *JMIR Public Health and Surveillance* **6**, e18606 (2020).

55. Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. NOAA National Centers for Environmental Information https://doi.org/10.7289/V5D21VHZ (2012).

56. Center for International Earth Science Information NetworkCIESIN-Columbia University. (2018) Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. (NASA Socioeconomic Data and Applications Center). https://sedac.ciesin.columbia.edu/data/collection/gpw-v4.

57. Zhao, Q. QZhao16/EWs. (2023).

58. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol* **178**, 1505–1512 (2013).

59. Gostic, K. M. *et al.* Practical considerations for measuring the effective reproductive number, Rt. *PLOS Computational Biology* **16**, e1008409 (2020).

60. Brett, T. S. & Rohani, P. Dynamical footprints enable detection of disease emergence. *PLOS Biology* **18**, e3000697 (2020).

61. Dakos, V. *et al.* Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data. *PLOS ONE* **7**, e41010 (2012).

62. Ebisuzaki, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate* **10**, 2147–2153 (1997).

63. Kuhn, M. caret: Classification and Regression Training. 6.0-94 https://doi.org/10.32614/CRAN.package.caret (2007).

64. Brooks, M. *et al.* glmmTMB: Generalized Linear Mixed Models using Template Model Builder. 1.1.9 https://doi.org/10.32614/CRAN.package.glmmTMB (2017).

65. Nations, U. *Data Reader's Guide*. *Human Development Reports* https://hdr.undp.org/reports-and-publications/2020-human-development-report/data-readers-guide.

66. Lüdecke, D., Aust, F., Crawley, S., Ben-Shachar, M. S. & Anderson, S. C. ggeffects: Create Tidy Data Frames of Marginal Effects for 'ggplot' from Model Outputs. (2025).

852 67. Therneau, T. M. coxme: Mixed Effects Cox Models. 2.2-22 https://doi.org/10.32614/CRAN.package.coxme
853 (2009).
854 68. Medina, R. G. *et al.* 2. Developmental stability, age at onset of foraging and longevity of Africanized honey bees
855 (Apis mellifera L.) under heat stress (Hymenoptera: Apidae). *Journal of Thermal Biology* **74**, 214–225 (2018).
856 69. Moore, D. F. 3. book: Multiple Survival Outcomes and Competing Risks. in *Applied Survival Analysis Using R* (ed.
857 Moore, D. F.) 113–135 (Springer International Publishing, Cham, 2016). doi:10.1007/978-3-319-31245-3_9.
858 70. McMahon, D. P. *et al.* 4.Elevated virulence of an emerging viral genotype as a driver of honeybee loss.
859 *Proceedings of the Royal Society B: Biological Sciences* **283**, 20160811 (2016).
860 71. Therneau, T. M. survival: Survival Analysis. 3.8-3 https://doi.org/10.32614/CRAN.package.survival (2001).
861 72. R Core Team. R: The R Project for Statistical Computing. http://R-project.org/. (2021).
862 73. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. EpiEstim: Estimate Time Varying Reproduction Numbers
863 from Epidemic Curves. 2.2-4 https://doi.org/10.32614/CRAN.package.EpiEstim (2022).
864 74. Wallinga, J. & Teunis, P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar
865 Impacts of Control Measures. *Am J Epidemiol* **160**, 509–516 (2004).
866 75. Bettencourt, L. M. A. & Ribeiro, R. M. Real Time Bayesian Estimation of the Epidemic Potential of Emerging
867 Infectious Diseases. *PLOS ONE* **3**, e2185 (2008).
868 76. Feng, F. *et al.* Effects of extreme temperature on respiratory diseases in Lanzhou, a temperate climate city of
869 China. *Environ Sci Pollut Res* **28**, 49278–49288 (2021).
870 77. Miller, P. B., O'Dea, E. B., Rohani, P. & Drake, J. M. Forecasting infectious disease emergence subject to seasonal
871 forcing. *Theoretical Biology and Medical Modelling* **14**, 17 (2017).
872 78. Cerqueira, V., Moniz, N. & Soares, C. VEST: automatic feature engineering for forecasting. *Mach Learn* (2021)
873 doi:10.1007/s10994-021-05959-y.
874 79. Mei, Y., He, W., Xie, X., Wan, S. & Gu, B. Increasing Long-Term Memory as an Early Warning Signal for a Critical
875 Transition. *Journal of Climate* **37**, 487–504 (2023).
876 80. Pavithran, I. & Sujith, R. I. Effect of rate of change of parameter on early warning signals for critical transitions.
877 *Chaos: An Interdisciplinary Journal of Nonlinear Science* **31**, 013116 (2021).

81. Grziwotz, F. *et al.* Anticipating the occurrence and type of critical transitions. *Science Advances* **9**, eabq4558 (2023).