Anticipating disease emergence and elimination: a test of early warning signals using empirically based models

Andrew T. Tredennick[1,2,*], Eamon B. O'Dea[1,2], Andrew Park[1,2], TBD[3], Pejman Rohani[1,2,4], John M. Drake[1,2]

[1]*Odum School of Ecology, University of Georgia, Athens, GA 30602, USA*
[2]*Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA 30602, USA*
[3]*A University of Somewhere*
[4]*Department of Infectious Diseases, University of Georgia, Athens, GA 30602, USA*

Abstract

Forecasts of the emergence, re-emergence, and elimination of human infectious diseases would allow for proactive, rather than reactive, decisions that could save lives. Recent theory suggests that a generic feature of dynamical systems approaching a tipping point – critical slowing down – can anticipate disease emergence and elimination. Empirical demonstrations of critical slowing down in real disease dynamics are scarce, but are essential before we can implement model-independent outbreak detection systems. Here, we use empirically-based, mechanistic models of measles transmission in four Nigerien cities to detect critical slowing down through statistical early warning signals. We find that several early warning signals accurately anticipate measles re-emergence and elimination, suggesting that critical slowing down can be detected before tipping points in real disease dynamics. Broadly, our findings suggest that early warning signals, coupled with decision-support algorithms and expert judgment, could provide the basis for outbreak early detection systems.

Keywords: critical slowing down, early warning signals, epidemiology, measles, infectious disease

Introduction

Forecasts of the emergence and re-emergence of infectious diseases have the potential to save lives, money, and human productivity by allowing for proactive, rather than reactive, preparedness measures [1]. Similarly, indicators of the elimination of infectious diseases can signal the effectiveness of "end game" strategies aimed at disease eradication [2]. Predicting (re)emergence and elimination is possible with complex mathematical models of disease transmission, but their success relies on detailed understanding of the underlying transmission dynamics and adequate data [3]. We often do not have enough information (or time) to parameterize such models. An alternative approach is to use model-independent statistical signals that portend infectious disease (re)emergence and elimination by detecting critical slowing down as the system approaches a critical transition [4].

Emergence and elimination of an infectious disease both involve a critical transition (technically, a *transcritical bifurcation*). The transition typically occurs at the critical point where the basic reproduction number (R_0, the number of secondary cases that arise from a single infected case in a fully susceptible population) is equal to one [5]. Thus, subcritical ($R_0 < 1$) and supercritical ($R_0 > 1$) systems represent alternative modes of fluctuation [4,6,7].

*Current address: Western EcoSystems Technology, Inc., 200 S. Second Street, Laramie, WY 82070, USA
Email addresses: atredenn@gmail.com (Andrew T. Tredennick), jdrake@uga.edu (John M. Drake)

Critical transitions in stochastic systems, such as systems of disease transmission, are often associated with critical slowing down, a reduction in the resilience of a system to perturbations [8,9]. Critical slowing down (CSD), in turn, is associated with changes in the dynamical features of the system: early warning signals (EWS) such as an increase in the variance and autocorrelation [6,10]. Recent theoretical work suggests that CSD occurs as disease dynamics approach $R_0 = 1$ from below (emergence) [4,11] and from above (elimination) [2,4,12], and that several EWS anticipate the critical transition [13–15]. These findings suggest that early warning signals could be operationalized to develop early warning systems of disease emergence and elimination, or outbreaks of endemic diseases.

However, operationalizing EWS, and deploying early warning systems based on them, faces many challenges [1,16]. For example, using EWS in an "online mode" requires choosing temporal windows over which EWS are calculated. These moving windows should be long enough to provide reliable statistics, but short enough to forget the past so as to not overwhelm information contained in new observations. This will be especially important for diseases that fluctuate seasonally, where EWS might always increase and then decrease over the course of the year, requiring the end-user to reset computations each season. Another challenge is defining thresholds for detection of an upcoming tipping point. Detection thresholds can be based on the absolute value of an EWS (e.g., warning if variance exceeds some value), the trend in an EWS over time (e.g., warning if the correlation of variance with time exceeds some value), or an algorithmic combination of many factors (e.g., variance and autocorrelation increases above some value three observation periods in a row). Developing (semi)objective algorithms for combining EWS, their values, and their trends to best detect disease emergence and elimination is a surmountable task, but one that will take considerable research effort.

Given the amount of research still required to operationalize EWS, it is imperative that we stress test the hypothesis that EWS anticipate critical transitions before we allocate time and resources to the task of deploying early warning systems. One way to stress test EWS is through empirical case studies: Can EWS anticipate emergence and elimination in real disease time series? However, uncritical application of EWS to observed data could lead to researchers getting the right answer for the wrong reasons. EWS might perform well for a given time series, but for reasons having nothing to do with critical slowing down. Likewise, critical transitions may occur in the absence of early warning signals [17]. Indeed, without known the critical point (i.e., when $R_0(t) = 1$) it is impossible to know if EWS are in fact sending us the signal we think they are.

Another option is to use empirically based models of disease transmission to test EWS. This offers several advantages. First, using a model to simulate time series of cases means we also have access to a time series of $R_0(t)$, which allows us to know precisely when the critical transition occurs. This means we know whether we are getting the right answer for the right reason. Second, we can simulate replicate time series to account for the inherent stochasticity of disease transmission. This means our conclusions are not based on one-off events that bias conclusions [18]. Third, we can specifically simulate emergence and elimination events. This means we can separate the stress testing of EWS from the research necessary to operationalize EWS. Thus, we have all the flexibility of a theoretical model, but we remain tethered to reality because the model parameters are fitted to real data.

Here, we use empirically based model simulations of measles dynamics to test whether CSD anticipates critical transitions in real disease dynamics. We focus on two scenarios: the re-emergence of measles following a large outbreak, a situation typical of measles dynamics in sub-Saharan Africa [19], and the elimination of measles by a vaccination campaign. We seek to answer two related questions. First, can CSD distinguish between time series of disease incidence when the underlying dynamics are far from and near to a critical transition? If so, then CSD can anticipate disease re-emergence and elimination. Second, how does the distance to and the rate of approaching the threshold impact the anticipatory skill of CSD?

To answer these questions, we fit mechanistic models of disease transmission to time series of measles incidence in four Nigerien cities [19,20]. We then use the fitted models to perform model experiments designed to test the performance of several EWS, which quantify CSD, at anticipating re-emergence and elimination. Our results confirm theoretical expectations about several EWS and associated CSD. In particular, we show that CSD before a critical transition is detectable by several EWS in realistic scenarios, and they do so using much shorter time series than used in theoretical studies. However, our study highlights the limitations of EWS in situations where disease re-emergence and elimination occurs rapidly. Moreoever, and contrary to theoretical expectations [4], we find that EWS perform better at detecting CSD before re-emergence than before elimination.

Materials and methods

Data

We used weekly measles case report data from four Nigerien cities: Agadez, Maradi, Niamey, and Zinder (figure 1*a*). The data were collected over an 11 year period from 1995-2005 (figure 1*b*). These data are ideal for testing theory on CSD in disease dynamics because each city has different population sizes, has different dynamics in terms of outbreak sizes and length of inter-epidemic periods, and each time series has different amounts of demographic stochasticity due to differences in population size. Such differences provide a natural gradient of "noise" that may influence CSD [21–24]. The data are provided by the Niger Ministry of Health [25].

Stochastic SEIR *model*

The model is a discrete-time approximation of a continuous-time SEIR model with limited demography, specified as a set of difference equations,

$$S_{t+dt} = n_{S,t} - n_{E,t} \tag{1}$$

$$E_{t+dt} = n_{E,t} - n_{I,t} \tag{2}$$

$$I_{t+dt} = n_{I,t} + n_{O,t} - n_{R,t}, \tag{3}$$

where $\mathbf{n}_t$ are random variables representing the number of individuals transitioning into or out of each class at each timestep $t \to t+dt$. n_S is the number of births, n_E is the number of newly infected individuals that have the disease but are not infectious, n_I is the number of newly infectious indiviuals, n_O is the number of imported infections, and n_R is the number of newly recovered individuals who are no longer infectious and have life-long immunity. The stochastic random variables are specified as follows:

$$n_{S,t} \sim \text{Poisson}(\mu_t N_t \times dt) \tag{4}$$

$$n_{E,t} \sim \text{Binomial}(\lambda_{E,t}, S_t) \tag{5}$$

$$n_{I,t} \sim \text{Binomial}(\lambda_{I,t}, E_t) \tag{6}$$

$$n_{O,t} \sim \text{Poisson}(\psi \times dt) \tag{7}$$

$$n_{R,t} \sim \text{Binomial}(\lambda_{R,t}, I_t), \tag{8}$$

where μ_t is the birth rate at time t, ψ is the rate of imported infections, and λ_E, λ_I, and λ_R are the probabilities of exposure, becoming infectious, and recovery, respectively. These probabilities reflect the processes of tranmission, transition from the latent period to the infectious period, and recovery, which we model as:

$$\lambda_{E,t} = 1 - e^{-\frac{\beta_t I_t dt}{N_t}} \tag{9}$$

$$\lambda_{I,t} = 1 - e^{-\eta E_t dt} \tag{10}$$

$$\lambda_{R,t} = 1 - e^{-\gamma I_t dt}, \tag{11}$$

where β_t is time-varying rate of transmission, η is time-invariant rate from the exposed class to the infectious class, and γ is time-invariant recovery rate. We model rate of transmission as,

$$\beta_t = \beta \left(1 + \sum_{i=1}^{6} q_i \xi_{i_t}\right) \Gamma_t, \tag{12}$$

where β is the minimum transmission rate over the season, ψ accounts for measles infections from external sources that are not part of the local dynamics, and the term $\sum_{i=1}^{6} q_i \xi_{i_t}$ is a B-spline to model seasonality in transmission. The B-spline bases (ξ_{i_t}) are periodic with a 1 year period. The transmission rate (β_t) is also subject to stochastic process noise at each time step, Γ_t, which we model as gamma-distributed white (temporally uncorrelated) noise with mean 1 and variance σ^2 [26].

We do not include a death process in the model because the rate of infection is much faster than the rate of death. Excluding deaths means we can avoid making further assumptions about demographic rates – we are already making assumptions about birth rates (e.g., the rate is the same across cities, but with city-specific population size). We model demographic stochasticity in births and imported infections by drawing time-specific values from Poisson distributions. In this model, the effective reproduction number at time t is: $R_E(t) = \frac{\beta_t}{\gamma} \frac{S_t}{N_t}$.

We assume observed case reports ($\mathbf{y}$) are drawn from a Negative Binomial distribution subject to a constant reporting fraction (ρ) and dispersion parameter τ,

$$y_t \sim \text{Negative Binomial}\left(\rho x_t, \tau\right), \tag{13}$$

where x_t are the accumulated cases that transition from the infected class to the recovered class in a one week period.

Model fitting and inference

We fit the SEIR model to time series of case reports from each of our focal cities using Maximization by Iterated particle Filtering (MIF). We estimated 14 parameters for each city: six seasonal transmission parameters (q_i), mean transmission rate (β), three initial conditions $\left(S_{(t=0)}, E_{(t=0)}, I_{(t=0)}\right)$, the number of imported infections (ψ), reporting fraction (ρ), one parameter accounting for process noise (σ), and one parameter accounting for measurement noise (τ). To ensure identifiability, and to make the model easier to fit, we assumed the infectious period was fixed at $1/\eta = 8$ days and the recovery period was fixed at $1/\gamma = 5$ days. The birth rate (μ_t) was multiplied by 0.3 to account for the reported 70% vaccination coverage [19].

MIF relies on particle filtering, which estimates the likelihood of fixed parameters by integrating state variables of a stochastic system. To narrow in on the maximum likelihood estimates, MIF lets parameters take a random walk during

the filtering process and selectively propagates forward parameter sets (i.e., particles) with the highest likelihood. The variance of the random walk decreases at each iteration of MIF, where a MIF iteration means one filtering pass through the time series. This procedure converges toward the maximimum likelihood estimates (MLEs), in theory.

We used the IF2 algorithm [27] implemented in the R [28] package `pomp` version 1.18 [29,30] to conduct the MIF procedure. To initialize MIF, we generated 5000 parameter sets using Latin Hypercube Sampling over large ranges of the parameter values. We then performed two rounds of MIF, each for 100 iterations, with 10000 particles, and geometric cooling. For the first round of MIF we set `cooling.facter = 1`. For the second round, which was initialized using the collection of parameter sets from the end of the first round, we set `cooling.factor = 0.9`. We computed the log likelihood of 5000 final MIF parameter sets (i.e., parameter sets collected after 200 MIF iterations) as the log of the mean likelihoods of 50 replicate particle filters with 10000 particles each. At this stage, we assume the parameter set with highest log likelihood is the MLE.

We used a bootstrapping approach to estimate approximate 95% confidence intervals for all parameters. The procedure, which was conducted for each city independently, is as follows. First, we simulated 100 realizations from the fitted model using the MLE parameters. Second, we fitted the SEIR model to each of the 100 bootstrap simulations using the same MIF procedure described above, except we initiated the parameter search from 50 parameter sets rather then 5000. We reduced the number of parameter sets due to the computational constraints of fitting 100 simulated data sets for each of the four cities. Third, we identified the MLE parameter set for each of the 100 bootstrap simulations from among the 50 MIF paramete sets. Last, we calculated summary statistics (mean, median, quantiles) from the distribution of 100 MLE parameters (SI text).

Model assessment

We used the MLE parameter sets to make one-week-ahead predictions and compared observed and expected case counts. To make one-week-ahead predictions, we used particle filtering with 50000 particles and retained the mean and standard deviation of all latent states across all particles before they were filtered at each time step. We used the mean predictions ($\mathbb{E}(\text{cases}_t)$) to assess model fit using a generalized coefficient of determination, calculated as: $R^2 = 1 - \frac{\sum_t [\mathbb{E}(\text{cases}_t) - \text{cases}_t]^2}{\sum_t [\text{mean(cases)} - \text{cases}_t]^2}$ [31]. In addition to comparing model expectations to in-sample observed data, we also compared our fitted SEIR models to two benchmarking models: a negative binomial sampling model that assumes independent and identically distributed observations and a seasonal moving average model (SI text). We fit the benchmarking models to the observed data and then compared models using Akaike's Information Criterion (AIC) to account for differences in the number of model parameters [32].

Simulating re-emergence

To simulate re-emergence of measles, we manipulated the initial size of the susceptible pool to simulate an increase from low $R_E(t)$ to high $R_E(t)$. Doing so allows us to test whether EWS can distinguish between windows of time when $R_E(t)$ is far from a critical transition and when $R_E(t)$ is near a critical transition. We reduced the initial fraction of susceptible individuals by multplying the MLE for $S_{(t=0)}$ by six discounting factors: 1e-4, 0.1, 0.2, 0.3, 0.4, and 0.5. These discounting factors represent situations of susceptible depletion after outbreaks of various size. After defining $S_{(t=0)}$ based on the discounting factor, we then set the initial number of recovered individuals to $R_{(t=0)} = N - S_{(t=0)}$ and set the initial number of exposed and infected individuals to zero. Population size, N, was set to the mean population size for each city over the 1995-2005 time period. We then simulated the model forward for forty years using mean birth rate for the entire country ($\mu = 0.05$) and setting the death rate equal to the birth rate ($\mu = \nu = 0.05$) to achieve a constant eqiulibrium total population size over the course of the simulation (total population size does vary, though, because of

stochasticity in the model). Forty years was long enough for $R_E(t)$ to reach or exceed 1 for each city. Because the model is stochastic, we repeated these simulations 500 times for each city-susceptible discount combination.

Next, we split each simulated time series into null and test intervals. First, averaging across all simulations for a city-susceptible discount combination, we found the simulation year in which $R_E(t)$ reaches or exceeds 1 and excluded years past that year (SI text). We split the remaining time series into two windows of equal length (figure 3*a*). The null interval is the first window, where $R_E(t)$ is increasing but far from 1. The test interval is the second window, where $R_E(t)$ is increasing and approaching 1. We did this for each city and for each level of susceptible depletion. We calculated EWS over null and test intervals separately.

Simulating elimination

To simulate elimination, we simulated a vaccine campaign in which vaccination coverage linearly increased over time to eventually reach 100%, i.e. eradication (figure S3). We ran simulations for 100 years, starting with 50 years of dynamics at the baseline vaccine coverage reported for Niger of 70%, $p = 0.7$ [19]. Note that vaccination coverage is included in our model by discounting the birth rate of susceptibles by $1 - p$. At year 50, we initiated the vaccination campaign and let the model run for another 50 years. We ran simulations across six vaccination "speeds" (the rate at which $p \to 1$; SI text), simulating situations of slow and fast approaches to elimination. As in the re-emergence simulations, we set the birth rate equal to the death rate to achieve a constant equilibrium population size.

We then split each time series into null and test intervals for calculating EWS. We define the test interval as the window of time between the start of the vaccination campaign (year 50) and the time at which vaccination coverage reached the vaccination threshold of $1 - 1/R_0$ (figure 4*a*). R_0 was calculated for each city using the MLE parameters as: $R_0 = [\eta\beta] / [(\eta + \nu)(\gamma + \nu)]$, where β is the maximum seasonal transmission rate. We define the null interval as the window of time that ends at the start of the vaccination campaign (year 49) and starts at a time that results in an interval equal in length the test interval (figure 4*a*). EWS were then calculated for each interval.

Calculating early warning signals

We considered nine candidate early warning signals (Table S1). We used the `spaero::get_stats()` function [33] in R [28] to calculate EWS according to the formulas in Table S1. All EWS except the coefficient of variation are expected to increase as $R_E(t)$ approaches 1 from below [4,12,14]. Less is known about the behavior of EWS as $R_E(t)$ approaches 1 from above. But, theory does tell us that, for SIR models, the mean should decrease, autocorrelation should increase, and the variance should decrease [4].

For each simulation of re-emergence and elimination, we calculated EWS for the time series of expected cases in the null and test intervals. This yielded a distribution of EWS over the 500 null and test intervals. We assessed the performance of each EWS using the Area Under the Curve (AUC) statistic. Specifically, we use AUC to calculate the amount of overlap between the distributions of each EWS from the null and test intervals. Values of AUC far from 0.5 (i.e., close to 0 or 1) indicate a greater degree of separation and thus better performance of a particular EWS in terms of classifying whether $R_E(t)$ is close to a critical transition. We calculated AUC as: $\text{AUC} = [r_{\text{test}} - (n_{\text{test}} + 1)/2] / (n_{\text{test}} n_{\text{null}})$ where r_{test} is the sum of the ranks of test set EWS statistics in a combined set of null and test statistics (lower numbers have lower ranks), n_{test} is the number of test of statistics and n_{null} is the number of null statistics. The AUC of an EWS is the probability that a randomly chosen EWS value from the test set is higher than an EWS value randomly chosen from the null set [34]. Therefore, AUC should be high (closer to 1) when an EWS is expected to increase as a critical transition is approached, whereas AUC should be low (closer to 0) when an EWS is expected to decrease.

Results

The fitted models adequately reproduce observed dynamics (figure 1*b*), with in-sample R^2s from one-week-ahead predictions ranging from 0.54 for Agadez to 0.89 for Maradi (figure 2*a*). The fitted models also had lower AIC values than two benchmarking models (SI table Sx). Stochastic simulations of the models displayed dynamics typical of each city (figure S1), including the decline in seasonality amplitude as population size decreases (figure 2*b*) [19]. Our model for Agadez performs poorly relative to the other cities, but still better than non-mechanistic models (SI table Sx). Maximum likelihood estimates and bootstrapped 95% confidence intervals for all parameters are in the SI text (tables Sx-Sz).

The EWS generally perform as expected by theory on the approach to re-emergence. Most EWS increased as the critical transition is approached, resulting in positive AUC values near 1 (figure 3). Skewness, kurtosis, and coefficient of variation performed poorly across all levels of susceptible depletion in all cities. Thus, these metrics are unreliable.

Variance, mean, index of dispersion, autocovariance, and autocorrelation all perform equally well at predicting re-emergence (figure 3*c*). Their performance declines as the amount of susceptible depletion decreases. This is expected because more rapid returns to $R_E(t) = 1$ result in shorter null and test intervals, making estimates of EWS less precise [13]. Moreover, as the time to reach $R_E(t) = 1$ decreases, the chance of a bifurcation delay increases because of the changing equilibrium of R_0 combined with demographic stochasticity [11,13]. Thus, re-emergence may prove difficult to anticipate in "fast" transmission systems, as demonstrated theoretically by [13] and seen here when susceptible depletion was relatively small (figure 3*c*).

The EWS did not perform as well when anticipating elimination, relative to emergence (figure 4). Only three metrics are reliable: mean, autocovariance, and variance. All three metrics decreased as $R_E(t)$ approached the critical transition (figure S5). As in the case of anticipating elimination, AUC values decreased as the speed of the vaccine campaign increased (figure 4*c*). Again, this is because of shorter null and test intervals.

In all, the suite of EWS suggest that critical slowing down does occur in measles dynamics as a critical transition is approached. We found similar results for the approach to elimination when calculating EWS over a moving window of 35 weeks in the null and test intervals (SI text, figure S7). But, all EWS performed worse when predicting the approach to emergence over the moving window (figure S7).

Discussion

Using empirically-based disease transmission models, we found evidence of critical slowing down before critical transitions to re-emergence and elimination of measles. This evidence comes from the fact that several EWS accurately anticipate the critical transition.

Evidence of CSD is strongest for scenarios of re-emergence. This is because the autocorrelation and other EWS increased on the approach to $R_E = 1$ (fig. 3*c*). Increases in EWS besides the autocorrelation, such as the variance, are necessary but insufficient evidence of CSD because such EWS can also increase when a critical transition does not exist (**CITATION?**). However, an increase in the autocorrelation, coupled with coincident increases in other EWS, provides stronger evidence of CSD (**CITATION?**). JOHN: I remember you making the points in the previous two sentences in a meeting once. Can you provide citations?

The autocorrelation was not a strong indicator of the transition to elimination (fig. 4*c*), suggesting weaker evidence for CSD. Other EWS (variane, autocovariance, and mean) did show altered behavior as the transition to elimination was approached, but these EWS were less sensitive under elimination scenarios compared to re-emergence scenarios. These

results suggest that CSD might not preempt disease elimination or that CSD is harder to detect when a disease is fading out rather than emerging. We have reasons to expect CSD at the elimination threshold based on theoretical modeling studies [2,4,12], but most theory has been developed for SIR models rather than the SEIR model we employed here. Still, we think it is more likely that CSD is simply harder to detect in noisy systems approaching elimination. EAMON: Can you add text on why this might be the case?

A potential limitation of our findings is that the levels of susceptible depletion we modeled (figure 3*c*) might be lower than the levels that occur in reality. To check the relevance of this limitation, we calculated the level of susceptible depletion after outbreaks (defined as years where the total number of cases reached 80% of the maximum observed) across one hundred replicate simulations (SI text). We found that susceptible depletion was less than 0.5, the smallest susceptible depletion level we tested, for 0.9% of outbreaks in Agadez, 21% of outbreaks in Maradi, 100% of outbreaks in Niamey, and 26% of outbreaks in Zinder. These statistics do not detract from our main findings of CSD in measles dynamics, but they do suggest that EWS might be less useful in some cases than in others. For example, AUC values for emergence at the 0.5 level of susceptible depletion are already low for most cities (figure 3*c*). Thus, EWS are not practical for cities that rarely experience levels of susceptible depletion below 0.5 (e.g., Agadez).

Our results should encourage efforts to develop model-independent early warning systems for infectious diseases [1]. We have shown that critical slowing down precedes tipping points in real disease dynamics, but how to operationalize the phenomenon of critical slowing down remains an open research area [35]. Emerging technologies like artifical intelligence might offer new ways to find optimal detection thresholds for early warning signals. But there will always be a role for expert judgement. Early warning signals, though powerful and now accompanied with empirical support, will likely be just one part of a decision-support toolkit.

Acknowledgements

We thank Tobias Brett comments on the modeling approach and Paige Miller and Éric Marty for helpful comments on early versions of this work.

Funding

This research was funded by the National Institute of General Medical Sciences of the National Institutes of Health (Award Number U01GM110744). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work was done on the Olympus High Performance Compute Cluster located at the Pittsburgh Supercomputing Center at Carnegie Mellon University, which is supported by National Institute of General Medical Sciences Modeling Infectious Disease Agent Study (MIDAS) Informatics Services Group grant 1U24GM110707.

Authors' contributions

ATT concieved of the study, designed the study, carried out the statistical and modeling analysis, and drafted the manuscript; EBO participated in the design of the study, participated in statistical and modeling analysis, and critically revised the manuscript; AP participated in the design of the study, guided the model analysis, and critically revised the manuscript; PR participated in the conception and design of the study, guided the statistical and modeling analysis, and helped draft and revise the manuscript; JMD coordinated the study, participated in the conception and design of the

study, guided the statistical and modeling analysis, and helped draft and revise the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

References

1. Han BA, Drake JM. 2016 Future directions in analytics for infectious disease intelligence. *EMBO reports*, e201642534. (doi:10.15252/embr.201642534)

2. Drake JM, Hay SI. 2017 Monitoring the path to the elimination of infectious diseaes. *Tropical Medicine and Infectious Disease* **2**, 20.

3. Metcalf CJE, Lessler J. 2017 Opportunities and challenges in modeling emerging infectious diseases. *Science* **357**, 149–152. (doi:10.1126/science.aam8335)

4. O'Regan SM, Drake JM. 2013 Theory of early warning signals of disease emergenceand leading indicators of elimination. *Theoretical Ecology* **6**, 333–357. (doi:10.1007/s12080-013-0185-5)

5. Heffernan JM, Smith RJ, Wahl LM. 2005 Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface* **2**, 281–293. (doi:10.1098/rsif.2005.0042)

6. Scheffer M *et al.* 2009 Early-warning signals for critical transitions. *Nature* **461**, 53–59. (doi:10.1038/nature08227)

7. Scheffer M *et al.* 2012 Anticipating critical transitions. *Science* **338**, 344–348. (doi:10.1126/science.1225244)

8. Wissel C. 1984 A universal law of the characteristic return time near thresholds. *Oecologia* **65**, 101–107. (doi:10.1007/BF00384470)

9. Nes EH van, Scheffer M. 2007 Slow Recovery from Perturbations as a Generic Indicator of a Nearby Catastrophic Shift. *The American Naturalist* **169**, 738–747. (doi:10.1086/516845)

10. Carpenter SR, Brock WA. 2006 Rising variance: A leading indicator of ecological transition. *Ecology Letters* **9**, 311–318. (doi:10.1111/j.1461-0248.2005.00877.x)

11. Dibble CJ, O'Dea EB, Park AW, Drake JM. 2016 Waiting time to infectious disease emergence. *Journal of the Royal Society Interface* **13**, 20160540. (doi:10.1098/rsif.2016.0540)

12. O'Regan SM, Lillie JW, Drake JM. 2016 Leading indicators of mosquito-borne disease elimination. *Theoretical Ecology* **9**, 269–286. (doi:10.1007/s12080-015-0285-5)

13. Brett TS, Drake JM, Rohani P. 2017 Anticipating the emergence of infectious diseases. *Journal of the Royal Society Interface* **14**, 20170115. (doi:10.1098/rsif.2017.0115)

14. Brett TS, O'Dea EB, Marty É, Miller PB, Park AW, Drake JM, Rohani P. 2018 Anticipating epidemic transitions with imperfect data. *PLoS Computational Biology* **14**, e1006204. (doi:10.1371/journal.pcbi.1006204)

15. Miller PB, O'Dea EB, Rohani P, Drake JM. 2017 Forecasting infectious disease emergence subject to seasonal forcing. *Theoretical Biology and Medical Modelling* **14**, 17. (doi:10.1186/s12976-017-0063-8)

16. Boettiger C, Hastings A. 2013 Tipping points: From patterns to predictions. *Nature* **493**, 157–158.

17. Boettiger C, Hastings A. 2012 Early warning signals and the prosecutor's fallacy. *Proceedings of the Royal Society B: Biological Sciences* **279**, 4734–4739. (doi:10.1098/rspb.2012.2085)

18. Boettiger C, Hastings A. 2012 Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface* **9**, 2527–2539. (doi:10.1098/rsif.2012.0125)

19. Ferrari MJ, Grais RF, Bharti N, Conlan AJ, Bjørnstad ON, Wolfson LJ, Guerin PJ, Djibo A, Grenfell BT. 2008 The dynamics of measles in sub-Saharan Africa. *Nature* **451**, 679–684. (doi:10.1038/nature06509)

20. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. 2011 Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* (doi:10.1126/science.1210554)

21. Hastings A, Wysham DB. 2010 Regime shifts in ecological systems can occur with no warning. *Ecology Letters* **13**, 464–472. (doi:10.1111/j.1461-0248.2010.01439.x)

22. Dakos V, Van Nes EH, D'Odorico P, Scheffer M. 2012 Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology* **93**, 264–271. (doi:10.1890/11-0889.1)

23. O'Dea EB, Park AW, Drake JM. 2018 Estimating the distance to an epidemic threshold. *Journal of the Royal Society Interface* **15**, 20180034. (doi:10.1098/rsif.2018.0034)

24. O'Regan SM, Burton DL. 2018 How Stochasticity Influences Leading Indicators of Critical Transitions. *Bulletin of Mathematical Biology* **80**, 1630–1654. (doi:10.1007/s11538-018-0429-z)

25. Niger Ministry of Health. 2008 Weekly measles case reports, 19 95–2004.

26. Bretó C, Ionides EL. 2011 Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications* **121**, 2571–2591. (doi:10.1016/j.spa.2011.07.005)

27. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. 2015 Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences* **112**, 719–724. (doi:10.1073/pnas.1410597112)

28. R Core Team. 2017 R: A language and environment for statistical computing.

29. King AA, Nguyen D, Ionides EL. 2016 Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal Of Statistical Software* **69**, 1–43. (doi:10.18637/jss.v069.i12)

30. King AA *et al.* 2018 pomp: Statistical Inference for Partially Observed Markov Processes (R package, version 1.18).

31. Martinez-Bakker M, King AA, Rohani P. 2015 Unraveling the transmission ecology of polio. *PLoS Biology* (doi:10.1371/journal.pbio.1002172)

32. Akaike H. 1973 Information theory and an extension of the maximum likelihood princilple. In *Proceedings of the 2nd international symposium on information theory*, pp. 267–281. (doi:10.1007/978-0-387-98135-2)

33. O'Dea EB. 2018 spaero: Software for Project AERO (R package version 0.3.0).

34. Fawcett T. 2006 An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861–874. (doi:https://doi.org/10.1016/j.patrec.2

35. Shmueli G, Burkom H. 2010 Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics* **52**, 39–51. (doi:10.1198/TECH.2010.06134)

List of Figures

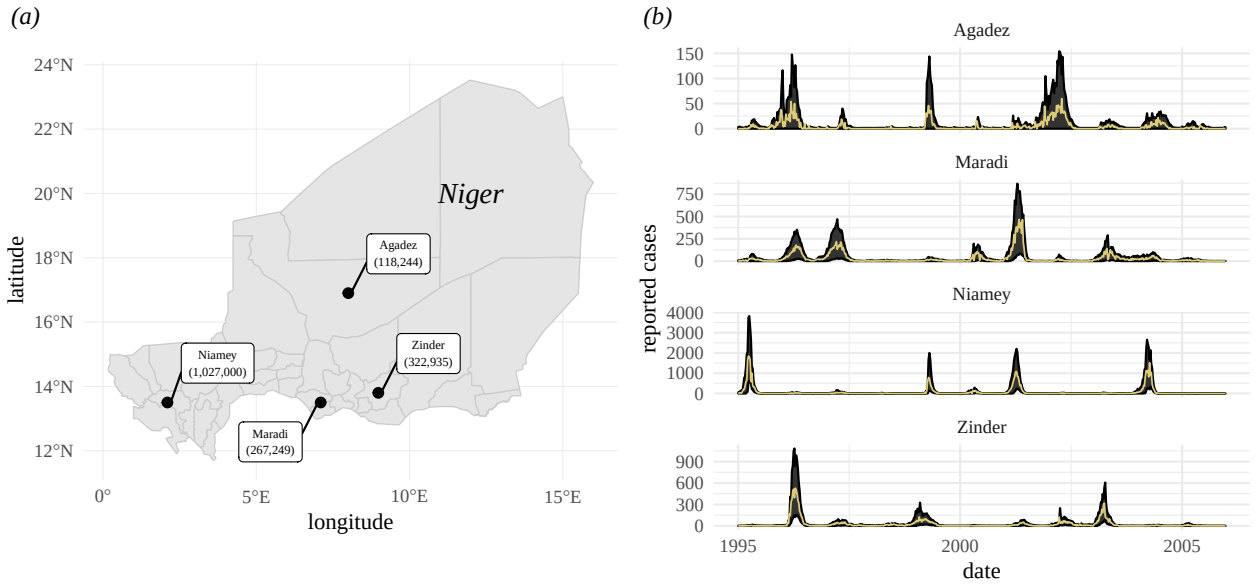

Figure 1: Locations of data sources and observed and predicted measles dynamics. (*a*) Locations and population sizes (in parantheses) of our four focal cities in Niger. (*b*) Time series of weekly reported cases (yellow solid lines) and the 95% prediction intervals (black ribbons) for one-week-ahead predictions from our fitted SEIR models for each city.

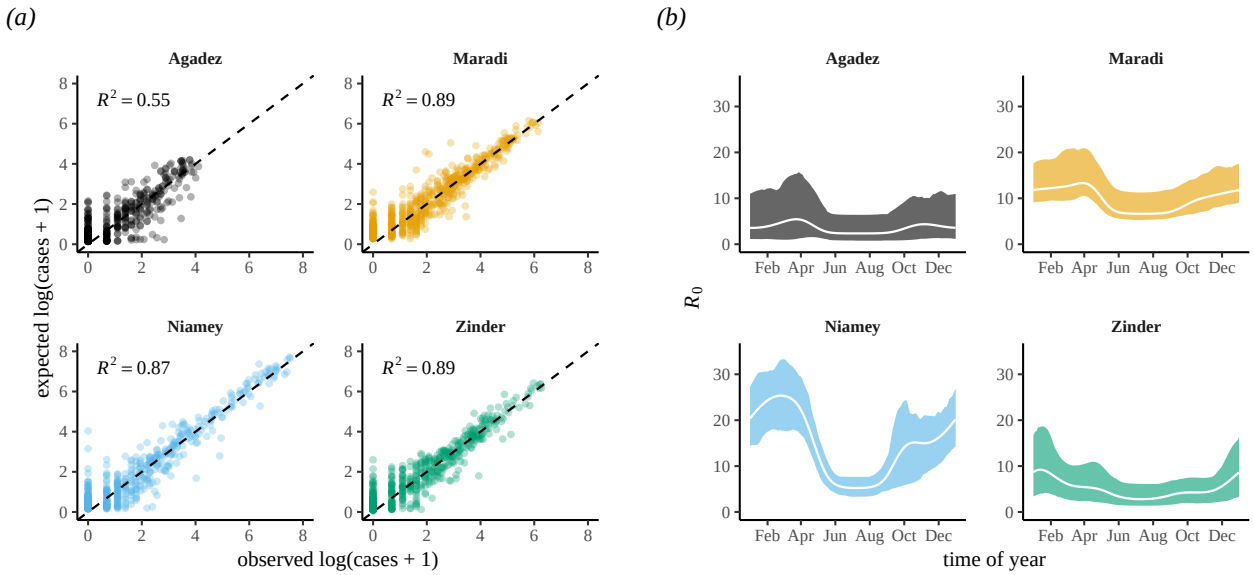

Figure 2: Accuracy of the fitted *SEIR* models and estimated seasonality. (*a*) Comparison of in-sample model predictions and observations for each city. Expected cases are one-week-ahead predictions from the fitted models. The dashed line shows 1:1. Coefficients of determination (R^2) were calculated as the reduction in the sum-of-squared errors from model predictions relative to a null model of the mean number of cases (SI text). (*b*) The estimated seasonality of the basic reproductive ratio (R_0) for each city. R_0 was calculated as: $\frac{\eta\beta_t}{(\eta+\nu)(\gamma+\nu)}$, where $1/\eta$ is the infectious period, $1/\gamma$ is the recovery period, β_t is the time-specific rate of transmission, and ν is the death rate. Only β_t is estimated by our model. We set $1/\eta = 8$ days, $1/\gamma = 5$ days, and $\mu = \nu = 0.05$ for calculating R_0 as shown in this figure. The white line is R_0 calculated using the MLE parameters; shaded regions are the bootstrapped 95% confidence intervals.

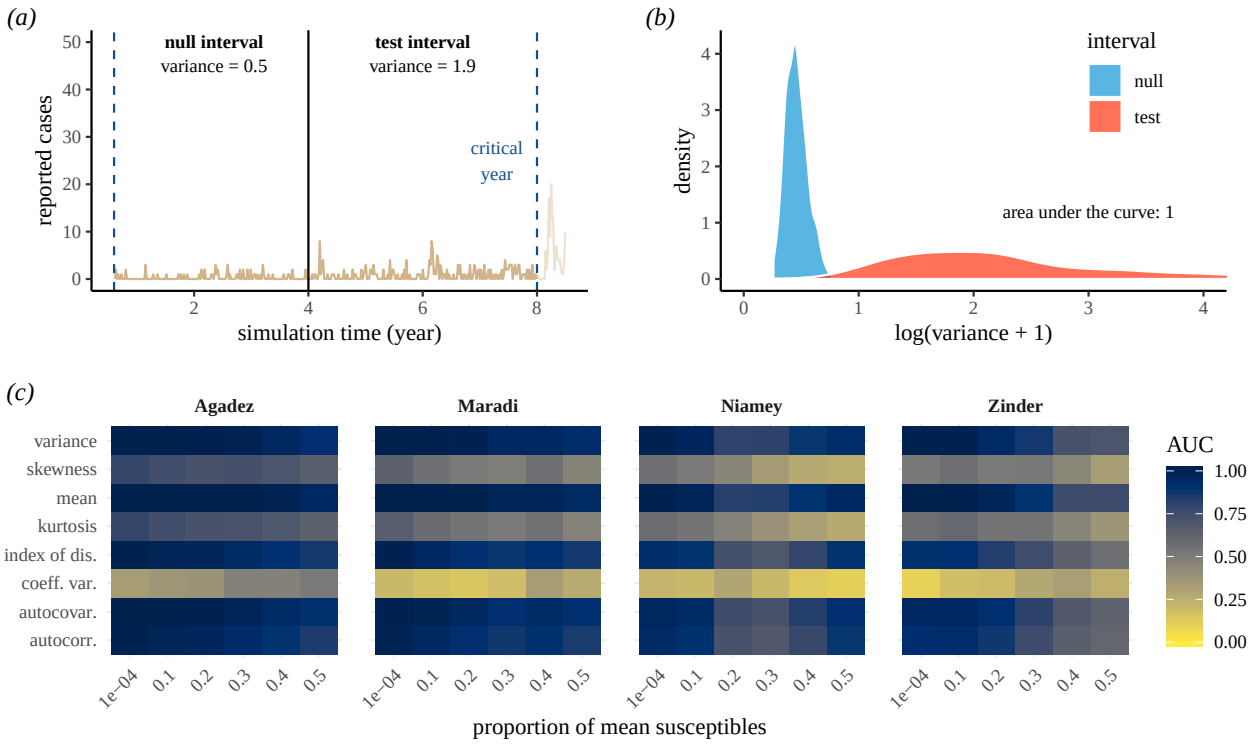

Figure 3: Performance of early warning signals (EWS) over fixed windows on the approach to emergence. (*a*) A typical example of an emergence simulation for Maradi. The two vertical blue lines indicate the start (left-most line) and end (line for critical year) of the full window. The black line demarcates the division between the equal-length null and test intervals, in which we show the calculated variance. (*b*) Empirical densities of variance in the null and test intervals across 500 simulations and the associated area under the curve (AUC) statistic. (*c*) Heatmap of AUC statistics for each EWS at each level of susceptible discount factor. AUC values closer to 0 or 1 indicate higher ability to distinguish among time series near and far from a critical transition. See Fig. Sx for visualization of how susceptible discounting factor maps to number of weeks in the null and test intervals.

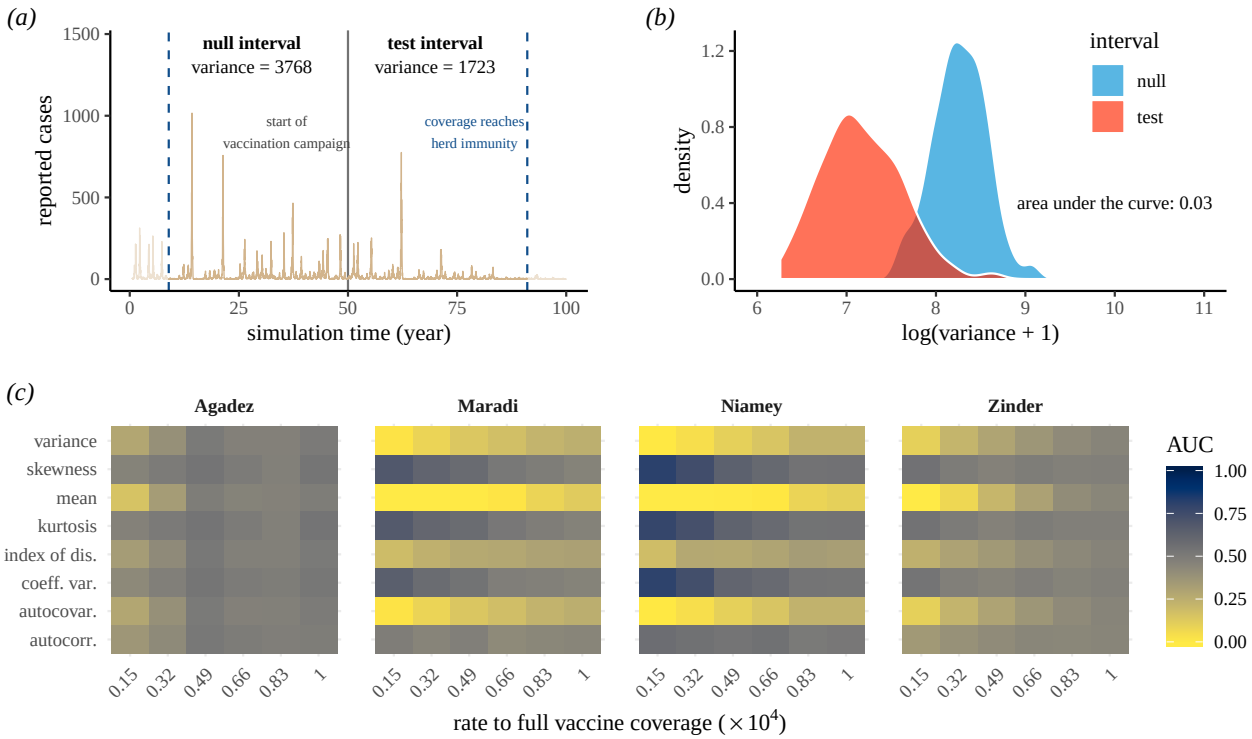

Figure 4: Performance of early warning signals (EWS) over fixed windows on the approach to elimination. (*a*) A typical example of an elimination simulation for Maradi. The two vertical blue lines indicate the start (left-most line) and end (line for critical year) of the full window. The black line demarcates the division between the equal-length null and test intervals, in which we show the calculated variance. (*b*) Empirical densities of variance in the null and test intervals across 500 simulations and the associated area under the curve (AUC) statistic. (*c*) Heatmap of AUC statistics for each EWS at each speed of approach to herd immunity. AUC values closer to 0 or 1 indicate higher ability to distinguish among time series near and far from a critical transition. See Fig. Sx for visualization of how vaccination speed maps to number of weeks in the null and test intervals.