



# The probability of epidemic burnout in the stochastic SIR model with vital dynamics

Todd L. Parsons<sup>a</sup> , Benjamin M. Bolker<sup>b,c</sup>, Jonathan Dushoff<sup>b,d</sup> , and David J. D. Earn<sup>c,d,1</sup>

Edited by Alan Hastings, University of California, Davis, CA; received August 9, 2023; accepted November 17, 2023

We present an approach to computing the probability of epidemic “burnout,” i.e., the probability that a newly emergent pathogen will go extinct after a major epidemic. Our analysis is based on the standard stochastic formulation of the Susceptible-Infectious-Removed (SIR) epidemic model including host demography (births and deaths) and corresponds to the standard SIR ordinary differential equations (ODEs) in the infinite population limit. Exploiting a boundary layer approximation to the ODEs and a birth-death process approximation to the stochastic dynamics within the boundary layer, we derive convenient, fully analytical approximations for the burnout probability. We demonstrate—by comparing with computationally demanding individual-based stochastic simulations and with semi-analytical approximations derived previously—that our fully analytical approximations are highly accurate for biologically plausible parameters. We show that the probability of burnout always decreases with increased mean infectious period. However, for typical biological parameters, there is a relevant local minimum in the probability of persistence as a function of the basic reproduction number  $\mathcal{R}_0$ . For the shortest infectious periods, persistence is least likely if  $\mathcal{R}_0 \approx 2.57$ ; for longer infectious periods, the minimum point decreases to  $\mathcal{R}_0 \approx 2$ . For typical acute immunizing infections in human populations of realistic size, our analysis of the SIR model shows that burnout is almost certain in a well-mixed population, implying that susceptible recruitment through births is insufficient on its own to explain disease persistence.

epidemics | stochastic processes | SIR model | extinction

It is well known that solutions of the standard ordinary differential equations (ODEs) describing a Susceptible-Infectious-Removed (SIR) epidemic with host births and deaths (aka “vital dynamics” or “demography”) eventually converge on a globally asymptotically stable equilibrium (1). Approach to the endemic equilibrium (EE) typically occurs via damped oscillations, motivating the use of the SIR model with demography as a basis for models of observed recurrent epidemics of childhood infections such as measles (2–6). For many biologically reasonable parameter values and population sizes, however, the troughs of these oscillations pass through infectious-host densities corresponding to a small fraction of an individual—the so-called “attoxox problem” (7)—calling into question the appropriateness of the deterministic SIR model.

Here, we estimate the probability that a pathogen disappears at the end of a major epidemic in a stochastic individual-based SIR model, in a population of finite size. In the large population limit, the densities of each type ( $S$ ,  $I$ ,  $R$ ) are asymptotically deterministic and governed by the standard SIR ODEs (8). We will refer to pathogen extinction soon after introduction as *fizzle*, whereas if the pathogen escapes fizz, we will refer to extinction at the end of a major epidemic as *epidemic burnout*,\* following the terminology of Dushoff (9). We will say that the pathogen *persists* if it has a subsequent epidemic wave, although it is worth mentioning that we always expect eventual extinction in a stochastic model with a finite population (10); the time to extinction of a pathogen that has survived to a state near the endemic equilibrium is considered in, e.g., refs. 11–13. Fig. 1 shows sample paths of the proportion of infectious individuals for the stochastic SIR model (together with the trajectory obtained from the ODE), illustrating fizz, burnout, and persistence.

## Significance

If a new pathogen causes a large epidemic, then it might “burn out” before causing a second epidemic. The burnout probability can be estimated from large numbers of computationally intensive simulations, but an easily computable formula for the burnout probability has never been found. Using a conceptually simple approach, we derive such a formula for the standard SIR epidemic model with vital dynamics (host births and deaths). With this formula, we show that the burnout probability is always smaller for diseases with longer infectious periods, but is bimodal with respect to transmissibility (the basic reproduction number). Our analysis shows that the persistence of typical human infectious diseases cannot be explained by births of new susceptibles, clarifying an important epidemiological puzzle.

**Author contributions:** All authors contributed to the conception of the research; T.L.P. and D.J.D.E. carried out the analysis, simulations, and figure creation, with input from B.M.B. and J.D.; T.L.P. and D.J.D.E. wrote the manuscript; and all authors revised the manuscript and approved the final version.

The authors declare no competing interest.

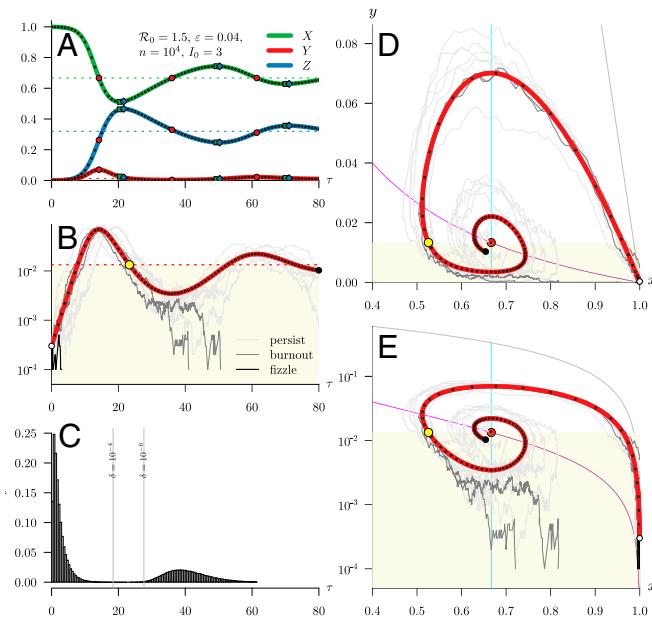
This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

\*To whom correspondence may be addressed. Email: earn@math.mcmaster.ca.

Published January 22, 2024.

<sup>a</sup>While “fade-out” (or “fadeout”) is commonly used to describe this extinction, e.g., ref. 4, §2.3, we find it conceptually useful to follow ref. 9 in distinguishing between extinction after a first major epidemic versus that occurring after multiple epidemics, and reserve the term fadeout for the latter.



**Fig. 1.** Sample paths of the stochastic SIR model (Fig. 2) and the ODE (Eq. 5) showing fizzle, burnout, and persistence. (A) The frequencies of susceptible, infectious, and removed individuals in the ODE (symbols indicate the critical points of the curve of the corresponding colour). Dashed lines indicate the endemic equilibrium (Eq. 8) of the deterministic model (Eq. 5). (B) The proportion of infectious individuals as a function of time. The boundary layer—inside which we approximate the stochastic dynamics with a birth-death process—is shaded in yellow, and the first point at which the deterministic trajectory enters the boundary layer is indicated with a heavy yellow dot. (C) Probability density of the time to extinction, estimated from  $10^6$  realizations of the stochastic process. The vertical lines show the time  $\tau_\delta$  (Eq. 76) for which the probability of fizzle after  $\tau_\delta$  is less than  $\delta$  (the lines correspond to  $\delta = 10^{-4}$  and  $10^{-6}$ ). (D and E) Trajectories in the susceptible-infectious phase plane with the nullclines; the vertical scale is linear in (D) and logarithmic in (E). The thin black curve is the boundary of the deterministically accessible region, defined by  $S + I = 1$ . Small yellow dots along trajectories are spaced by one time unit (the mean infectious period).

The problem of epidemic burnout has been of ongoing interest (4, 14, 15), e.g.,

“The question ‘will the agent go extinct after the first outbreak?’ cannot be answered within the context of a deterministic description. So we would like to be able to switch back to a stochastic description at the end of the epidemic outbreak. While it is well known how to calculate the probability of extinction from a branching process in a constant environment..., it seems difficult to do so when environmental quality (from the point of view of the agent, i.e., the presence of susceptibles!) is improving linearly at a certain rate.” (14, p. 42)

and has been previously approached via perturbation methods (16, 17) and by hybrid analytical-numerical approaches (18):

1. van Herwaarden (16) (henceforth vanH) starts from a large population diffusion approximation to the Markov chain formulation of the SIR model (*Model*). Under the assumption that the individual mortality rate is low, a highly accurate approximation to the solution of the infinite-population limit SIR ODEs is obtained, which is in turn used to estimate the point of entry to a boundary layer where the number of infectious individuals is very small. In the boundary layer, the backward equation for the diffusion approximation<sup>†</sup> is

tractable and is used to obtain an analytical approximation to the burnout probability [vanH, Eq. (5.13)], which requires the numerical evaluation of an integral. It is, to quote Diekmann and Heesterbeek (14, p. 42), “an ingenious piece of work,” although it is challenging to interpret for non-experts.

2. By contrast, Meerson and Sasorov (17) (henceforth MS) retain the discrete population model. They estimate the probability of extinction as the probability of reaching the state with only one infective individual (weighted by the expected number of returns to this state<sup>‡</sup>) times the probability that a single infective recovers before transmitting to any other individuals. They approximate this probability by the product of the expected total time (summed over multiple returns) in the state with a single infectious individual and the disease recovery rate (which is the rate of going extinct given that there is only one infectious individual). The time in the single-infective state is characterized by linear equations obtained by integrating the forward equations (see, e.g., ref. 19, §14.2) for all transient states over all time, for which an approximate solution is found via a WKB ansatz (see, e.g., ref. 22, Chapter 10) in the large population limit (i.e., a diffusion approximation is introduced implicitly). Under these assumptions, the burnout probability is shown to decay exponentially in the population size, with a constant of proportionality that is approximated analytically in the parameter regime where the initial exponential growth rate of infectious individuals greatly exceeds the per capita turnover rate (equivalent to  $\beta - \gamma \gg \mu$  in our formulation below). While providing coarser estimates than vanH, this approach yields a deterministic approximation to the most probable trajectory to pathogen extinction via a Hamiltonian formalism (see, e.g., refs. 23 and 24, Exercise 5.7.36). Like vanH, the approximation of MS involves an integral that cannot be evaluated analytically and presents a non-trivial numerical problem due to singularities in the integrand.
3. More recently, after identifying discrepancies between the analytical results of vanH and MS and the results of simulations, especially at smaller values of the expected population size  $n$ , (18) introduced a computational approach that scales as  $\mathcal{O}(n^2)$ . As in the previous approaches, Ballard et al. (18) use the solution of the SIR ODEs—now evaluated numerically and summed with a higher-order Gaussian correction (8, Theorem 11.2.3)—to identify the point of entry into a boundary layer, where a simplified form of the Markov chain is then simulated to estimate the probability of burnout.

The approximations of vanH and MS are summarized in §2.3 of ref. 18. We compare the performance of these approximations with that of an analytical approximation that we have derived in the spirit of the quote from ref. 14 above. Like vanH and ref. 18, we use the SIR ODEs to approximate the stochastic SIR trajectories outside a boundary layer. Then, inside the boundary layer, we use a time-inhomogeneous birth-and-death process that approximates the true stochastic dynamics more accurately than the diffusion approximation of vanH (in *Boundary Layer Independent Estimates*, we obtain the expression from vanH as an approximation to ours). Our approach is simpler and more intuitive than the diffusion approximation, and—in contrast to all previous work—we obtain fully analytical expressions that are numerically stable and can be computed without recourse to numerical evaluation of integrals. Our approach yields expressions for the probability of persistence after any

<sup>†</sup>See, e.g., refs. 19 or 20 for a discussion of the forward and backward diffusion equations; ref. 21 is an excellent introduction to boundary-layer methods for Markov chains.

<sup>‡</sup>In practice, there is negligible probability of returning to the state with one infective after an excursion to a state with many infectives.

number of epidemic waves and is also more amenable to generalizations than singular perturbation analysis of diffusion approximations; indeed, while we do not discuss the matter in detail here, the boundary-layer diffusions of vanH correspond to large population approximations for the branching processes we consider here (similar to limits in refs. 25 and 26).

## Approach and Analysis

**Model.** We consider the spread of an infectious disease in a discrete population in which births balance deaths on average, so there is a well-defined **expected population size**  $n$ . We consider a sequence of models indexed by  $n$ , and for the  $n$ th model denote by  $S_n(t)$ ,  $I_n(t)$  and  $R_n(t)$  the numbers of individuals at time  $t$  who are susceptible, infectious, and removed, respectively. The total population size is

$$N_n(t) = S_n(t) + I_n(t) + R_n(t). \quad [1]$$

Births and immigration of new susceptible individuals occur at constant rate  $\mu n$ , while deaths occur at per capita rate  $\mu$ , independent of disease status. Thus, at every time  $t$ , we have

$$\mathbb{E}[N_n(t)] = n, \quad [2]$$

where the expectation is taken over realizations of the stochastic process. Infectious individuals recover at rate  $\gamma$ , and new infections occur according to the law of mass action in a well-mixed population, i.e., at rate

$$\frac{\beta S_n(t) I_n(t)}{n}. \quad [3]$$

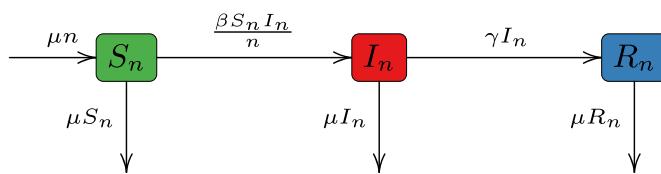
Since the demographic and epidemiological rates depend only on the state of the system at the current time, our sequence is an ensemble of Markov chain models (indexed by the expected total population size  $n$ ).

Following a common convention in probability theory, we use upper case for functions and lower case for indices and the values of functions at a given time. We index the functions by expected population size because we need to consider the limit of the sequence of functions as  $n \rightarrow \infty$ , whereas we use subscripts on function values to specify time, e.g.,  $s_0 = S_n(0)$ .

The model structure is indicated in a compartmental transfer diagram in Fig. 2, and the nature and rates of each type of event are summarized in Table 1.

**Deterministic Approximation.** In the limit of large population size, the stochastic SIR model (Fig. 2 and Table 1) is well-approximated by deterministic ODEs. More precisely, writing

$$X_n = \frac{S_n}{n}, \quad Y_n = \frac{I_n}{n}, \quad Z_n = \frac{R_n}{n}, \quad [4]$$



**Fig. 2.** Compartmental model for an SIR epidemic with vital dynamics. Labels on the arrows correspond to individual jump rates between states. For simplicity, the model is defined so that births/immigrations on average balance deaths, so that the expected total population size ( $\mathbb{E}[N_n(t)] = n$ ) is fixed.

**Table 1. Event types in the stochastic SIR model**

Event type	Rate	Transitions
Birth/immigration	$\mu N_n$	$S_n \rightarrow S_n + 1$
Transmission	$\beta S_n I_n / N_n$	$S_n \rightarrow S_n - 1, I_n \rightarrow I_n + 1$
Recovery	$\gamma I_n$	$I_n \rightarrow I_n - 1, R_n \rightarrow R_n + 1$
Susceptible death	$\mu S_n$	$S_n \rightarrow S_n - 1$
Infectious death	$\mu I_n$	$I_n \rightarrow I_n - 1$
Removed death	$\mu R_n$	$R_n \rightarrow R_n - 1$

in the limit  $n \rightarrow \infty$ , the frequencies  $(X_n(t), Y_n(t), Z_n(t))$  converge [almost surely on finite time intervals (8)] to the solution  $(X(t), Y(t), Z(t))$  of the ODEs,

$$\frac{dX}{dt} = \mu(1 - X) - \beta XY, \quad [5a]$$

$$\frac{dY}{dt} = (\beta X - \gamma - \mu)Y, \quad [5b]$$

$$\frac{dZ}{dt} = \gamma Y - \mu Z. \quad [5c]$$

Formally, to make this connection, one must be careful to have a sensible relationship between the initial conditions for the stochastic processes and the initial conditions for the ODEs. For example, given an initial state  $(X(0), Y(0), Z(0))$  for the ODEs, if one takes

$$(X_n(0), Y_n(0), Z_n(0)) = \frac{1}{n}(\lfloor nX(0) \rfloor, \lfloor nY(0) \rfloor, \lfloor nZ(0) \rfloor), \quad [6]$$

then the theorem applies. More generally, one must choose initial conditions  $(X_n(0), Y_n(0), Z_n(0))$  for the stochastic processes such that the limits  $\lim_{n \rightarrow \infty} X_n(0)$ , etc. exist, and one must take these limits as initial conditions for the ODEs (see Theorem 11.2.1 in ref. 8, p. 456); Example B on p. 453 of Ethier & Kurtz (8) illustrates how the SIR model without demography relates to the hypotheses of the theorem, and Chapter 5 in ref. 27 provides a pedagogical introduction to Kurtz's results in the context of epidemic models.

The trajectories of the deterministic SIR model (Eq. 5) always converge to a globally asymptotically stable (GAS) equilibrium point, which can be shown via a combination of the Poincaré–Bendixson Theorem and Dulac's criterion (28) or via a Lyapunov function (29). The nature of the asymptotic state is determined by the **basic reproduction number** (the expected total number of new infections caused by a single infective individual introduced into a naive population),

$$\mathcal{R}_0 = \frac{\beta}{\gamma + \mu}. \quad [7]$$

If  $\mathcal{R}_0 \leq 1$ , then the GAS fixed point is the **disease free equilibrium**,  $(x, y) = (1, 0)$ , whereas if  $\mathcal{R}_0 > 1$ , then all solutions converge—either via damped oscillations or monotonically—to an **endemic equilibrium**,

$$(x_\star, y_\star) = \left( \frac{1}{\mathcal{R}_0}, \epsilon \left( 1 - \frac{1}{\mathcal{R}_0} \right) \right), \quad [8]$$

where

$$\epsilon = \frac{\mu}{\gamma + \mu} \quad [9]$$

gives the mean infectious period as a fraction of the mean host lifetime. Our analysis requires that  $\epsilon$  is small but not too small

$(\frac{1}{n} \ll \epsilon \ll 1)$ , which is true for a wide variety of common acute immunizing infections (Table 2). The upper bound ( $\epsilon \ll 1$ ) is essential so we can justify perturbation expansions in  $\epsilon$ . The lower bound ( $\frac{1}{n} \ll \epsilon$ ) is equivalent to  $ne \gg 1$ , which ensures that the number of infectives at equilibrium ( $ny_*$ ) is substantially greater than 1 (from Eq. 8,  $ny_* \sim n\epsilon$ ). The ODEs continue to provide a good approximation to the epidemic dynamics until the **prevalence**  $y$  (the proportion of hosts that are infectious) becomes small; we take “small” to mean that  $y$  is less than the equilibrium prevalence  $y_*$  (Eq. 8). Thus, we take the **boundary layer**—within which the dynamics must be treated stochastically—to be the region of the phase plane where  $y < y_*$  (in *Boundary Layer Independent Estimates*, we also give approximations independent of the specific choice of boundary layer).

The need to analyze the dynamics differently within the boundary layer is especially clear if we consider the introduction of a single infectious individual into a fully susceptible population. If  $\mathcal{R}_0 > 1$ , then in the ODE system (Eq. 5)  $Y(t)$  will deterministically increase, whereas in the stochastic model,  $Y_n(t)$  will fizz with probability  $1/\mathcal{R}_0$  (30); i.e., the ODE (Eq. 5) fails to capture the dynamics of the stochastic model (Fig. 2) when there are few infectives. We therefore use a birth-and-death process to approximate the dynamics of the number of infectious hosts when that number is small (in contrast, susceptibles can be assumed to remain sufficiently abundant that we can always use the deterministic approximation  $X(t)$ ).

**Birth-and-Death Process Heuristic.** New infections occur at rate

$$\frac{\beta S_n(t)}{n} I_n(t) = \beta X_n(t) I_n(t) \approx \beta X(t) I_n(t), \quad [10]$$

while the number of infectious hosts decreases by one due to recovery or death at rate

$$(\gamma + \mu) I_n(t). \quad [11]$$

When there are few infectious hosts ( $I_n(t) < ny_*$ ), we approximate  $I_n(t)$  by a linear birth and death process with time-inhomogeneous per capita rates  $b(t)$  and  $d(t)$ , where

$$b(t) = \beta X(t), \quad [12]$$

$$d(t) = \gamma + \mu. \quad [13]$$

Note that when  $X(t)$  equals  $x_*$  (the classical herd immunity threshold),  $b(t) = d(t)$ , and the birth and death process transitions from subcritical to supercritical. Unlike in models without demography, the birth of new susceptible individuals ensures that a population will eventually cross the herd immunity threshold. Therefore, even if the number of infectious hosts initially declines it can eventually grow exponentially, if the infection survives until  $X(t) > x_*$ .

We can estimate the survival probability for this branching process, and thus, the persistence probability, using the following result.

**Theorem 1 [Kendall (31)].** Let  $\mathcal{K}(t)$  be a birth and death process with time-inhomogeneous per-capita birth rate  $b(t)$  and death rate  $d(t)$ . The probability of eventual extinction starting from one individual at time 0 is

$$q = \left( 1 + \frac{1}{\int_0^\infty e^{-\int_0^t [b(s)-d(s)] ds} d(t) dt} \right)^{-1}. \quad [14]$$

The extinction probability starting from  $k$  individuals is  $q^k$ .

Consequently, the probability of indefinite persistence (a branching process will either go extinct or grow indefinitely), starting from  $k$  individuals at time 0, is

$$\mathbb{P}\{\mathcal{K}(\infty) > 0\} = 1 - q^k. \quad [15]$$

To complete our persistence probability estimate, we need an expression for the proportion susceptible at time  $t$  ( $X(t)$ ) in Eq. 12). As suggested visually by the example shown in Fig. 1, inside the boundary layer ( $y < y_*$ ), both the deterministic and the stochastic trajectories spend most of their time at prevalences much lower than  $y_*$  [note the log scale in the subfigures (B) and (E)]. Consequently, we can approximate  $X(t)$  by solving (Eq. 5) with  $Y(0) = 0$ . Thus, we set

$$\frac{dX}{dt} \approx \mu(1 - X), \quad [16]$$

and solve this approximate equation as if it were exact to obtain

$$X(x_0, t) \approx 1 - (1 - x_0)e^{-\mu t}. \quad [17]$$

Here,  $x_0$  is the fraction susceptible at the initial time  $t = 0$ , and we write  $X(x_0, t)$  to emphasize the dependence on the initial state. We also write  $q(x_0)$  for the value of  $q$  in Eq. 14 obtained by taking  $b(t) = \beta X(x_0, t)$ .

We first apply this branching process approximation to a population at the disease-free equilibrium (DFE). Thus, we set  $x_0 = 1$  in (Eq. 17), which yields  $X(1, t) \equiv 1$ ; hence, we have a time-homogeneous branching process in this case, and the integral in (Eq. 14) is easily evaluated and yields  $q(1) = \frac{1}{\mathcal{R}_0} = x_*$ . Considering a small number of initially infective individuals,  $I_n(0) = k$ , we recover the classical expression for the establishment probability (30), that is, the probability that the pathogen does not fizz:

$$p_k = 1 - x_*^k. \quad [18]$$

We now use Kendall's  $q$  (Eq. 14) to compute the burnout probability. Assuming that the pathogen does not fizz, the number of infectious hosts will rapidly exceed  $ny_*$  individuals,<sup>§</sup> at which point the densities of both susceptible and infective hosts are well approximated by the ODEs (Eq. 5). To proceed, we need a formula for the fraction of hosts that are susceptible when the trajectory enters the boundary layer at the end of an epidemic; we denote this fraction  $x_{in}$  to emphasize that it refers to the susceptible proportion upon entry *into* the boundary layer (the point  $(x_{in}, y_*)$  is indicated by a heavy yellow dot in Fig. 1). In ref. 34, assuming  $\epsilon$  is small,<sup>¶</sup> we derive an approximate expression for the fraction susceptible,  $X(y, x_i)$ , as a function of the fraction infectious ( $y$ ) and the initial fraction susceptible ( $x_i$ ). Using that approximation, we have

$$x_{in} = X(y_*, x_i) \approx -x_* W_0 \left( -\mathcal{R}_0 x_i e^{-\mathcal{R}_0 (x_i - y_*)} \right) + \epsilon e^{\mathcal{R}_0 y_*} (E_1(\mathcal{R}_0 y_*) - E_1(\mathcal{R}_0 \bar{y}_0)). \quad [19]$$

<sup>§</sup>More precisely, for any  $y < \bar{y}_0$  (Eq. 21), conditional on not fizzing, the probability that  $I_n(t)$  hits 0 before hitting  $yn$  is exponentially small in  $n$  with exponential rate depending on  $y$  [for a rigorous demonstration see ref. 32 (Supplementary Information §8.2); ref. 33 gives explicit higher-order terms for the SIS model].

<sup>¶</sup>In ref. 34, we use  $\epsilon = \epsilon/\mathcal{R}_0$  rather than  $\epsilon$  as the small parameter, because using  $\epsilon$  leads to simpler expressions (see, e.g., (Eq. 24) below). Here, however, we analyze the dependence of our expressions on the epidemiologically relevant parameters  $\mathcal{R}_0$  and  $\epsilon$  and have re-written expressions from ref. 34 accordingly.

Here,  $W_0$  denotes the principal branch of the Lambert  $W$ -function<sup>#</sup> (35),  $E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt$  is the exponential integral function (36, §8.2.1) and  $\bar{y}_0$  is the **peak prevalence** in the limit  $\varepsilon \rightarrow 0$  (i.e.,  $\mu \rightarrow 0$ ), i.e., the maximum fraction infectious in the SIR model *without* vital dynamics,

$$\bar{y}_0 = x_i - x_\star(1 + \ln(x_i/x_\star)). \quad [21]$$

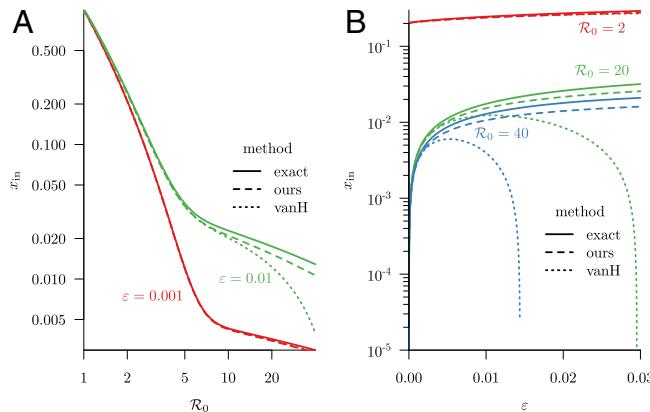
(See e.g., ref. 37 for a derivation of  $\bar{y}_0$ .) Taking  $x_i = 1$  corresponds to the invasion of a novel pathogen into an epidemiologically naive population (i.e., at the DFE). Later (*Subsequent Epidemic Waves*), we give an iterative scheme for  $x_{i,j}$ , an “effective initial fraction susceptible” that—substituted for  $x_i$  in Eqs. 19 and 21—gives the fraction susceptible at the end of the  $j$ th epidemic wave after invasion at the DFE. We compare our approximation of  $x_{in}$  for  $x_i = 1$  to the value obtained by numerically integrating the SIR ODEs (Eq. 5) in Fig. 3 and discuss its domain of applicability below (*The Domain of Applicability of the Approximation (Eq. 19) to  $x_{in}$* ).

If we now take  $t = 0$  to be the end of a major epidemic, i.e., the time when the infectious host density falls below  $y_\star$  and  $x_0 = x_{in}$ , then the density of infectious hosts is small, and the density of susceptible hosts is well approximated by  $X(x_{in}, t)$  (we are preparing a rigorous treatment of these results; here, we will content ourselves with showing that our analytical results closely match the results of individual-based simulations). We can thus estimate the **conditional burnout probability**—i.e., the probability of burnout conditional on not fizzling—by

$$q(x_{in})^{ny_\star}. \quad [22]$$

and the **conditional persistence probability** by

$$1 - q(x_{in})^{ny_\star}. \quad [23]$$



**Fig. 3.** Susceptible proportion ( $x_{in}$ ) upon entry into the boundary layer ( $y < y_\star$ ). (A)  $x_{in}$  as a function of  $R_0$  (Eq. 7). (B)  $x_{in}$  as a function of  $\varepsilon$  (Eq. 9). The exact value of  $x_{in}$  (obtained by numerically solving the SIR ODEs (Eq. 5)) is shown with solid curves, our approximation (Eq. 19) is shown with dashed curves, and the approximation of vanH is shown with dotted curves. Based on Eq. 42, the minimum  $R_0$  for which our approximation of  $x_{in}$  (Eq. 19) is valid is  $\approx e^{2\varepsilon}$  (i.e., 1.02027 for  $\varepsilon = 0.01$  and 1.0020027 for  $\varepsilon = 0.001$ ).

<sup>#</sup>If  $\mathcal{E}(z) = ze^z$ , Lambert's  $W$ -function  $W(z)$  (35) solves the “left-sided” inverse relation  $\mathcal{E}(W(z)) = z$ . This equation has countably many solutions, each corresponding to branches  $W_i$  of the  $W$ -function; we will need the two real branches,  $W_0$ , which maps  $[-\frac{1}{e}, \infty)$  to  $[-1, \infty)$ , and  $W_{-1}$ , which maps  $[-\frac{1}{e}, 0)$  to  $(-\infty, -1]$ . For these two branches,  $W_i$  is a partial “right-sided” inverse function for  $\mathcal{E}(z)$ :

$$\begin{aligned} W_{-1}(\mathcal{E}(z)) &= z \quad \text{if } z \leq -1 \\ W_0(\mathcal{E}(z)) &= z \quad \text{if } z \geq -1. \end{aligned} \quad [20]$$

Below (*Computing the Epidemic Burnout Probability*), we compute an exact expression for  $q(x_{in})$ ,

$$q(x_{in}) = \left(1 + \frac{\varepsilon}{z^{-a} e^z \mathcal{G}(a, z)}\right)^{-1} \quad [24a]$$

$$\text{where } z = \frac{R_0}{\varepsilon}(1 - x_{in}), \quad [24b]$$

$$\text{and } a = \frac{R_0}{\varepsilon}(1 - x_\star). \quad [24c]$$

Here,  $\mathcal{G}$  denotes the lower incomplete gamma function<sup>¶</sup> (36, §8.2.1); we use the nonstandard notation  $\mathcal{G}$  to avoid confusion with our recovery rate parameter  $\gamma$ . Below (*Asymptotics for Small  $\varepsilon$* ), we derive an approximation for  $q(x_{in})$  that is extremely accurate for small values of  $\varepsilon$ :

$$q(x_{in}) \approx \left(1 + \frac{1}{\sqrt{\frac{2\pi}{\varepsilon(R_0-1)}} \left(\frac{a}{z}\right)^a e^{z-a}}\right)^{-1}. \quad [25]$$

We emphasize that this expression is elementary and numerically stable.

Thus, the **burnout probability**—i.e., the probability of not fizzling (Eq. 18) but disappearing after an epidemic—is

$$p_k q(x_{in})^{ny_\star}, \quad [26]$$

where  $n$  is the expected total population size,  $y_\star$  is the equilibrium prevalence (Eq. 8),  $q$  is the probability of eventual extinction (under post-epidemic conditions) starting from one infectious individual (Eq. 14), and  $k$  is the initial number of infectious individuals. Our exact expression for  $q(x_{in})$  is given in (Eq. 24). Similarly, the **persistence probability**—i.e., the probability of not fizzling ( $p_k$ ) and then *not* burning out after a first epidemic (Eq. 23)—is

$$\mathcal{P}_1(R_0, \varepsilon, n, k) = p_k (1 - q(x_{in})^{ny_\star}). \quad [27]$$

More generally, the probability of persisting beyond the  $m$ th epidemic wave is

$$\mathcal{P}_m(R_0, \varepsilon, n, k) = p_k \prod_{j=1}^m (1 - q(x_{in,j})^{ny_\star}), \quad [28]$$

where

$$x_{in,j} = X(y_\star, x_{i,j}) \quad [29]$$

(see Eq. 19 and *Subsequent Epidemic Waves*). For biologically reasonable values of  $\varepsilon$ ,  $R_0$ , and  $n$ , we find that the difference between  $\mathcal{P}_1(R_0, \varepsilon, n, k)$  and  $\mathcal{P}_m(R_0, \varepsilon, n, k)$  is negligible, because  $q(x_{in,j}) \ll 1$  for  $j \geq 2$ . Intuitively, because the troughs between epidemics get shallower and shallower, an invading disease that survives burnout is almost certain to persist through many more cycles.

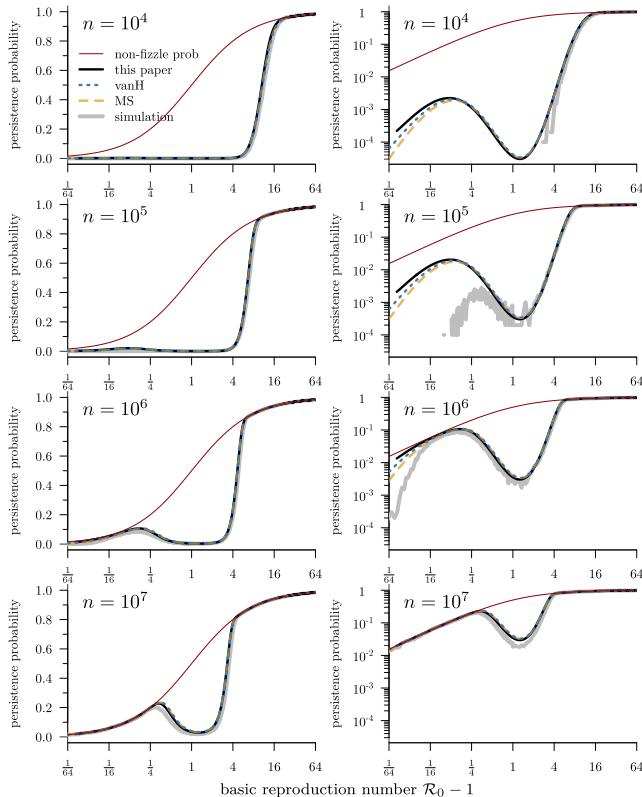
Thus, in *Results*, we focus on burnout after the initial epidemic when a novel disease invades a fully susceptible population. There, we use our accurate, numerically stable, and computationally efficient approximation for  $q(x_{in})$  (Eq. 25), obtained via Eqs. 19 and 37a, to compute the probability of burnout.

<sup>¶</sup> $\mathcal{G}(a, z) = \int_0^z t^{a-1} e^{-t} dt$  is proportional to the cumulative distribution function for the gamma distribution. We use this fact to compute  $\mathcal{G}(a, z)$  accurately in our burnout R package, mentioned in footnote \*\*.

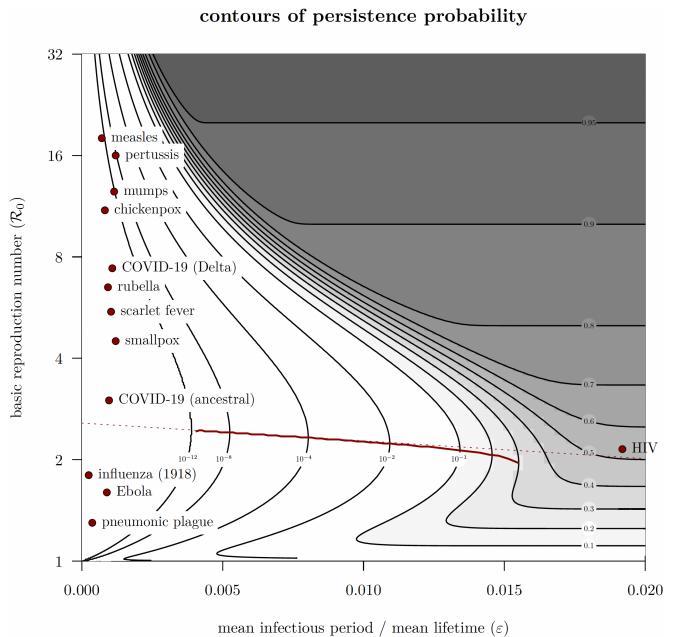
## Results

Fig. 4 shows that our analytical approximation for the persistence probability (Eq. 27) agrees very well with the same probability estimated from large numbers of simulations. The probability is shown as a function of the basic reproduction number ( $\mathcal{R}_0$ ) with fixed mean infectious period ( $\varepsilon = 0.01$ ). The panels differ only in the underlying expected population size (ranging from  $n = 10^4$  to  $10^7$ ). For each value of  $\mathcal{R}_0$ , the simulation-based persistence probability was estimated from  $10^7$  individual-based stochastic realizations of the model (Fig. 2 and Table 1). Note that  $\varepsilon = 0.01$  corresponds to an infectious period that is 1% of the average lifetime, far longer than is realistic for most acute immunizing infections; however, our approximation only improves for smaller  $\varepsilon$ . We use  $\varepsilon = 0.01$  in Fig. 4 so that discrepancies between the simulations and analytical results are visible.

Our simple approximation for Kendall's  $q$  (Eq. 25) allows us to easily and quickly explore the conditional and unconditional probability of pathogen extinction across the entire range of biologically plausible values of  $\mathcal{R}_0$  and  $\varepsilon$ . Fig. 5 shows a contour plot of the persistence probability (this graph would have required years of computer time to produce from simulations). As was



**Fig. 4.** Persistence probability as a function of the basic reproduction number  $\mathcal{R}_0$ , for population sizes ranging from  $n = 10^4$  to  $10^7$ . The vertical scale is linear in the left column and logarithmic in the right column; the horizontal scale is logarithmic (in  $\mathcal{R}_0 - 1$ ) in all panels. (The horizontal axis range is from  $\mathcal{R}_0 - 1 = \frac{1}{64} = 0.015625$  to 64, but our approximation is valid only for  $\mathcal{R}_0 - 1 \gtrsim 0.02027$ ; see Eq. 42.) The initial state is  $(S_n(0), I_n(0), R_n(0)) = (n - 1, 1, 0)$ . The mean infectious period as a fraction of mean lifetime is  $\varepsilon = 0.01$ , which is unrealistically long for most infections (Table 2), but the agreement between the analytical approximation (Eq. 27) and numerical simulations (Stochastic simulation algorithm) is better for smaller  $\varepsilon$  (we use a large value of  $\varepsilon$  so that discrepancies are visible). In addition to our analytical approximation (Eq. 27), we show the semi-analytical approximations of Meerson and Sasorov [MS (17)] and van Herwaarden [vanH (16)]. The thin red curve shows the probability of not fizzling,  $1 - \frac{1}{\mathcal{R}_0}$ .



**Fig. 5.** Probability of persistence after a large epidemic ( $\mathcal{P}_1$ , Eq. 27) as a function of basic reproduction number ( $\mathcal{R}_0$ ) and mean infectious period as a proportion of mean lifetime ( $\varepsilon$ ), for population size  $n = 10^6$ . The initial state is assumed to be a single individual introduced into a fully susceptible population ( $I_n(0) = k = 1, S_n(0) = n - k$ ). Positions for the red dots for infectious diseases of humans are from Table 2 [to avoid text overlap, measles is shifted up by 1 to 18, pertussis down by 1 to 16, and COVID-19 (Delta) up by 0.6 to 7.4]. The solid red curve shows the local minimum of persistence probability, and the dotted red line shows the analytical approximation (Eq. 63) to the local minimum.

observed previously (18, 38), Fig. 5 indicates that the burnout probability is non-monotone in  $\mathcal{R}_0$  for  $\varepsilon \lesssim 0.016$ . In this range of  $\varepsilon$ , the probability of persistence is lowest for basic reproduction numbers in the range  $2 \lesssim \mathcal{R}_0 \lesssim 2.57$ , and increases rapidly with increasing  $\mathcal{R}_0$ . Below (*The  $\mathcal{R}_0$  Maximizing the Probability of Burnout*), we compute a linear approximation to the value of  $\mathcal{R}_0$  at which the persistence probability is minimized. The upper limit of 2.57 for the persistence-minimizing range of  $\mathcal{R}_0$  is the limit as  $\varepsilon \rightarrow 0$  in Eq. 63; Fig. 5 shows that this linear approximation performs very well over the range where the persistence probability is non-monotonic. Less intuitively, the persistence probability increases for small  $\mathcal{R}_0$  (below the red curve in Fig. 5) as  $\mathcal{R}_0$  decreases to one. We note, however, that except for very large expected population size  $n$ , the secondary peak in the persistence probability—which occurs for  $1 < \mathcal{R}_0 \lesssim 2$ —remains small (*cf.* Fig. 4), except for pathogens with extremely long infectious periods. Fig. 5 also suggests that for fixed  $\mathcal{R}_0$ , the probability of persistence always increases with increasing  $\varepsilon$ , which we confirm analytically below (*The Burnout Probability is a Decreasing Function of  $\varepsilon$* ). Note that  $\beta = \mathcal{R}_0(\gamma + \mu) = \frac{\mathcal{R}_0\gamma}{1-\varepsilon}$ , so varying  $\varepsilon$  while holding  $\mathcal{R}_0$  constant simultaneously varies the infectious period and contact rate by a factor of  $\mathcal{O}(\varepsilon)$ .

## Discussion

The problem of infectious disease persistence following a major epidemic (4, p. 20; 14, p. 42; 47, p. 451; 9, 15, 38) is important for identifying characteristics of pathogens that can successfully invade, and is related to the notion of a “critical community size” required for a disease to persist in the long term (30).

Given sufficient computing resources, it is possible to estimate the persistence probability for a given model from large numbers of stochastic, individual-based simulations. The gray curves in Fig. 4 show this probability estimated from simulations of the SIR model. Fig. 4 also shows the probability estimated using previous analytical methods (16, 17) (blue and orange curves) and our approximation (black curves). All three analytical approaches yield similar results,\*\* and differences in the estimated probabilities can be seen only on a logarithmic scale in the limit as  $\mathcal{R}_0 \rightarrow 1^+$  (e.g., for  $\mathcal{R}_0 \lesssim 1.05$  in Fig. 4), where all of these approximations<sup>††</sup> are technically invalid: in a stochastic, finite population model, as  $\mathcal{R}_0 \rightarrow 1^+$  there is no phase during which the deterministic model is a good approximation, and the distinction between fizz, burnout, and fadeout breaks down (48). Analysis of the limit  $\mathcal{R}_0 \rightarrow 1^+$  could improve understanding of the process of eradication as the magnitude of control measures is increased, but for the burnout problem on which we focus here, the limit  $\mathcal{R}_0 \rightarrow 1^+$  is of limited interest.

While our approximation agrees closely with previous work (16, 17) for ranges of  $\mathcal{R}_0$  that are biologically relevant, there are several important theoretical and practical advantages of our approach; our analysis

- is simpler and easier to understand, since it is based directly on the underlying stochastic process rather than on a diffusion approximation (and is consequently easier to apply to models that are more complex than the SIR model considered here);
- yields fully analytical approximations that are numerically stable, unlike the previous analytical approaches (16, 17), which depend on non-trivial numerical integrations with singular integrands;
- predicts the persistence probability after an arbitrary number of epidemic waves.

We expand on these points below.

We have obtained useful analytical estimates (Eqs. 24, 25, 27, and 28) of the SIR epidemic burnout and persistence probabilities in a well-mixed population, via a hybrid use of ODEs when prevalence is high and time-dependent branching processes when prevalence is low. As noted after Eq. 28, the probability of burning out in each subsequent epidemic trough after persisting through the first is negligibly small for the SIR model.

Our time-dependent branching process approach (*Birth-and-Death Process Heuristic*) also yields analytical results that are more amenable to computation than previous approximations (16, 17). Our application of Laplace's method to approximate the integral in Kendall's  $q$  (Eq. 14) is particularly useful. Eq. 25 for the conditional burnout probability provides a fully analytical formula—not requiring the numerical evaluation of integrals as in previous approaches (16, 17)—that can be evaluated without numerical instabilities and agrees very well with numerical simulations across a wide range of biologically plausible values of  $\mathcal{R}_0$  and  $\epsilon$ . The convenience and speed of our simple analytical expression for the persistence probability (Eq. 27) also allows us to obtain results for larger population sizes than are tractable via hybrid numerical methods (18) and facilitates efficient exploration of more of the parameter space (though with less accuracy at smaller population sizes).

\*\* We have implemented all three approximations in an open-source R package, which we used to create our figures. The package is available at <https://github.com/davidearn/burnout>.

†† Differences between our approximation and those of refs. 16 and 17 as  $\mathcal{R}_0 \rightarrow 1^+$  arise at least in part because they use  $\mu$  rather than  $\epsilon$  as the small parameter, and consequently predict persistence for  $\beta/\gamma > 1$  rather than  $\beta/(\gamma + \mu) > 1$ .

As is suggested visually by Fig. 5, and proved below (*The Burnout Probability is a Decreasing Function of  $\epsilon$* ), the persistence probability increases with infectious period ( $\epsilon$ ) across all values of  $\mathcal{R}_0$ . For any given infectious period, one viable life history strategy for persistence is a high  $\mathcal{R}_0$  (dark gray shading in Fig. 5). In addition to this high  $\mathcal{R}_0$  strategy, for a limited range of longer infectious periods ( $\epsilon \lesssim 0.016$ ), there is a second life-history strategy that promotes persistence:  $\mathcal{R}_0$  close to but greater than one. We use our analytical results to compute a linear approximation to the value of  $\mathcal{R}_0 > 1$  at which the burnout probability is maximized (see Eq. 63 in *The  $\mathcal{R}_0$  Maximizing the Probability of Burnout*). This approximation shows excellent agreement with the numerical results over the range of  $\epsilon$  for which the secondary peak exists and the burnout probability is numerically distinguishable from 1 (in Fig. 5, the dotted red curve is the approximation and the solid red curve is the numerically computed exact value). Intriguingly, with the exception of the ancestral strain of SARS-CoV-2—which has been replaced by variants with much higher  $\mathcal{R}_0$ —the endemic infectious diseases of humans listed in Table 2 roughly divide into high and low  $\mathcal{R}_0$  strategies.

These life history strategies can be interpreted in terms of the **herd immunity threshold**,  $x = x_\star = \frac{1}{\mathcal{R}_0}$ , i.e., the minimum proportion susceptible at which the epidemic can grow from a small number of infections. When  $\mathcal{R}_0$  is large, the herd immunity threshold  $x_\star$  is low, allowing the fraction susceptible to rapidly reach the threshold. When  $\mathcal{R}_0$  is low, there is a larger reservoir of susceptible hosts at the end of the first major epidemic, which reduces the wait until the herd immunity threshold is crossed. In either case, a longer infectious period (larger  $\epsilon$ ) allows the pathogen to “wait out” the period of herd immunity. This non-monotonicity of the burnout probability as a function of  $\mathcal{R}_0$  was previously observed (18, 38), and the maximum burnout probability was conjectured to occur for  $\mathcal{R}_0 \approx 3$  (38) or  $\mathcal{R}_0 = 2$  (18). We have shown that, in fact, the value of  $\mathcal{R}_0$  at which the probability is maximized is a decreasing function of  $\epsilon$  (solid red curve in Fig. 5). The probability-maximizing  $\mathcal{R}_0$  varies from  $\mathcal{R}_0 \simeq 2.57$  for  $\epsilon \rightarrow 0$  (Eq. 63) to  $\mathcal{R}_0 \simeq 2$  for  $\epsilon \simeq 0.016$ ; for larger  $\epsilon$ , the persistence probability increases monotonically with  $\mathcal{R}_0$ .

These results also have evolutionary implications: reduced virulence may be associated with longer infectious periods

**Table 2. Representative parameters for acute immunizing infections (and HIV for comparison)**

Disease	$\mathcal{R}_0$	$T_{\text{lat}}$ [days]	$T_{\text{inf}}$ [days]	$\epsilon \times 10^3$	Source
Measles	17	8	5	0.71	(4)
Pertussis	17	8	14	1.2	(4)
Mumps	12	15	6	1.1	(4)
Chickenpox	11	10	5	0.82	(4)
COVID-19 (Delta)	6.8	5.8	14	1.1	(39)
Rubella	6.5	10	7	0.93	(4)
Scarlet fever	5.5	1.5	18	1	(4)
Smallpox	4.5	15	7	1.2	(40)
COVID-19 (ancestral)	3	3.7	14	0.97	(39)
HIV	2.2	87	270	19	(41)
Influenza (1918)	1.8	2	2.5	0.25	(4, 42)
Ebola	1.6	9.3	7	0.89	(43)
Pneumonic plague	1.3	4.3	2.5	0.37	(44)

The basic reproduction number ( $\mathcal{R}_0$ ), mean latent period ( $T_{\text{lat}}$ ), and mean infectious period ( $T_{\text{inf}}$ ) are taken from the cited sources. The dimensionless parameter  $\epsilon$  is defined in Eq. 9 in terms of the recovery rate ( $\gamma$ ) and birth-death rate ( $\mu$ ) in the SIR model. We associate  $1/\gamma$  with the mean generation interval of the SEIR model, i.e.,  $1/\gamma = T_{\text{lat}} + T_{\text{inf}}$  (45, 46), set  $\mu = 0.02/\text{year}$  to mimic human birth and death rates, and compute  $\epsilon = \mu/(\gamma + \mu)$ . Where original sources present a range, we have listed the midpoint. Many of the estimates come from Anderson and May (4) [ $\mathcal{R}_0$  is taken from Table 4.1 (4, p. 70); the mean latent and infectious periods come from Table 3.1 (4, p. 31)]. All the diseases listed in this table are shown in Fig. 5.

(e.g., if fewer hosts die while infectious), thereby reducing the probability of burnout. This suggests a mechanism—distinct from the population genetics/weak selection arguments presented in ref. 32—that could explain how in finite populations, natural selection may favour strains with reduced virulence while maximizing  $\mathcal{R}_0$ : strains that achieve higher  $\mathcal{R}_0$  by increasing their infectious period are more likely to persist than those that achieve higher  $\mathcal{R}_0$  by increasing transmissibility. A model of multi-strain competition is necessary to test this hypothesis (we consider the special case of selection for vaccine escape in ref. 49).

While the qualitative inferences we have made from analysis of the stochastic SIR model are suggestive of general processes, and—as we have observed above—could have interesting implications, further research is needed to determine whether they really do generalize broadly. Most acute immunizing infections afflicting human populations have short infectious periods and moderate  $\mathcal{R}_0$  values, and with these constraints, our analysis of the stochastic SIR model indicates that extinction of the pathogen at the end of the first major epidemic is almost certain in a well-mixed population.

Fig. 5 makes clear that the stochastic SIR model is insufficient on its own to explain pathogen persistence; it is essential to consider additional mechanisms, e.g., waning immunity or antigenic evolution resulting in effective loss of host immunity (50), rescue effects in a meta-population (51, 52), long-lived carrier infections (see ref. 53 for a recent survey), or zoonotic reservoirs (50, 54).

Multi-type or non-Markovian birth-and-death processes (55, 56), combined with more complicated compartmental models or renewal equation models with more general generation intervals (46) may allow our approach to be extended to models incorporating, e.g., latent periods and asymptomatic and carrier infections, or greater or lesser variability in infectious periods. A more difficult problem is to consider pathogen persistence in a meta-community of linked sites (52, 57), or other structured populations, rather than a well-mixed population. Smaller local community sizes tend to make local extinction more likely, whereas asynchrony in epidemic dynamics could allow pathogens to reinvoke following a local extinction (58). Are these processes adequate to plausibly explain the persistence of pathogens? Is the existence of low/high  $\mathcal{R}_0$  strategies generic, or an artifact of the SIR compartmental model? Are longer infectious periods always favourable for pathogen persistence? These questions suggest avenues for future work.

## Materials and Methods

**Computing the Epidemic Burnout Probability.** To apply Kendall's  $q$  (Eq. 14) to the problem of epidemic burnout, we need to compute the integral

$$\mathcal{I}(x_{\text{in}}) = \int_0^\infty e^{-\int_0^t [\beta X(x_{\text{in}}, s) - (\gamma + \mu)] ds} (\gamma + \mu) dt \quad [30a]$$

$$= \int_0^\infty e^{-\int_0^\tau [\mathcal{R}_0 X(x_{\text{in}}, \frac{\sigma}{\gamma + \mu}) - 1] d\sigma} d\tau, \quad [30b]$$

where, in the second line, we use the mean duration of infection ( $1/(\gamma + \mu)$ ) as the time unit and write  $\sigma = (\gamma + \mu)s$ ,  $\tau = (\gamma + \mu)t$ . Recalling Eq. 17, we can write

$$X(\sigma) \equiv X\left(x_{\text{in}}, \frac{\sigma}{\gamma + \mu}\right) = 1 - (1 - x_{\text{in}}) e^{-\varepsilon\sigma}, \quad [31]$$

and hence

$$X'(\sigma) = \varepsilon(1 - x_{\text{in}}) e^{-\varepsilon\sigma} = \varepsilon(1 - X(\sigma)). \quad [32]$$

Now, to evaluate the inner integral in Eq. 30a, we make a change of variables, using  $x = X(\sigma)$  as the variable of integration:

$$\begin{aligned} & \int_0^\tau \left[ \mathcal{R}_0 X\left(x_{\text{in}}, \frac{\sigma}{\gamma + \mu}\right) - 1 \right] d\sigma \\ &= \int_{X(0)}^{X(\tau)} \left[ \mathcal{R}_0 X - 1 \right] \frac{1}{\frac{dx}{d\sigma}} dx = \int_{x_{\text{in}}}^{X(\tau)} \frac{\mathcal{R}_0 X - 1}{\varepsilon(1 - x)} dx \\ &= -\frac{\mathcal{R}_0}{\varepsilon} (X(\tau) - x_{\text{in}}) - \frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star}) \ln \frac{1 - X(\tau)}{1 - x_{\text{in}}}. \end{aligned} \quad [33]$$

Changing variables in a similar way, we have

$$\int_0^T e^{-\int_0^\tau [\mathcal{R}_0 X(x_{\text{in}}, \frac{\sigma}{\gamma + \mu}) - 1] d\sigma} d\tau \quad [34a]$$

$$= \int_0^T e^{\frac{\mathcal{R}_0}{\varepsilon} (X(\tau) - x_{\text{in}}) + \frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star}) \ln \frac{1 - X(\tau)}{1 - x_{\text{in}}}} d\tau \quad [34b]$$

$$= \int_{x_{\text{in}}}^{X(T)} e^{\frac{\mathcal{R}_0}{\varepsilon} (x - x_{\text{in}})} \left( \frac{1 - x}{1 - x_{\text{in}}} \right)^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star})} \frac{dx}{\varepsilon(1 - x)}. \quad [34c]$$

We are interested in the probability of ultimate extinction, which corresponds to taking the limit as  $T \rightarrow \infty$ , or, equivalently,  $X(T) \rightarrow 1$ , giving us

$$\mathcal{I}(x_{\text{in}}) = \int_{x_{\text{in}}}^1 e^{\frac{\mathcal{R}_0}{\varepsilon} (x - x_{\text{in}})} \left( \frac{1 - x}{1 - x_{\text{in}}} \right)^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star})} \frac{dx}{\varepsilon(1 - x)} \quad [35a]$$

$$\begin{aligned} &= \frac{1}{\varepsilon} e^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}})} \left( \frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}}) \right)^{-\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star})} \\ &\quad \times \int_0^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}})} e^{-x} x^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star}) - 1} dx \end{aligned} \quad [35b]$$

$$\begin{aligned} &= \frac{1}{\varepsilon} e^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}})} \left( \frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}}) \right)^{-\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star})} \\ &\quad \times \varrho\left(\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star}), \frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\text{in}})\right), \end{aligned} \quad [35c]$$

where we recall  $\varrho$  denotes the lower incomplete gamma function. Eq. 24 follows immediately.

**Asymptotics for Small  $\varepsilon$ .** We may also write  $\mathcal{I}(x_{\text{in}})$  (Eq. 35) as

$$\mathcal{I}(x_{\text{in}}) = \frac{1}{\varepsilon} \int_{x_{\text{in}}}^1 \frac{1}{1 - x} e^{\frac{\mathcal{R}_0}{\varepsilon} (x - x_{\text{in}} + (1 - x_{\star}) \ln(\frac{1 - x}{1 - x_{\text{in}}}))} dx \quad [36a]$$

$$= \frac{1}{\varepsilon} \int_{x_{\text{in}}}^1 h(x) e^{\frac{\phi(x)}{\varepsilon}} dx, \quad [36b]$$

for  $h(x) = \frac{1}{1 - x}$  and  $\phi(x) = \mathcal{R}_0 (x - x_{\text{in}} + (1 - x_{\star}) \ln(\frac{1 - x}{1 - x_{\text{in}}}))$ .

Assuming  $\varepsilon$  is small, we can apply Laplace's method (22, §6.4): provided  $x_{\text{in}} \leq x_{\star}$ ,  $\phi(x)$  has its maximum at  $x = x_{\star}$ , so

$$\mathcal{I}(x_{\text{in}}) \sim \frac{1}{\varepsilon} \sqrt{\frac{2\pi\varepsilon}{|\phi''(x_{\star})|}} h(x_{\star}) e^{\frac{\phi(x_{\star})}{\varepsilon}} \quad [37a]$$

$$= \sqrt{\frac{2\pi}{\varepsilon(\mathcal{R}_0 - 1)}} e^{\frac{\mathcal{R}_0}{\varepsilon} (x_{\star} - x_{\text{in}})} \left( \frac{1 - x_{\star}}{1 - x_{\text{in}}} \right)^{\frac{\mathcal{R}_0}{\varepsilon} (1 - x_{\star})}, \quad [37b]$$

yielding Eq. 25.

**Remark 1:** Note that, since  $x_{\text{in}} < x_{\star}$ ,

$$\begin{aligned} 0 &< -\mathcal{R}_0 \int_{x_{\text{f}}}^{x_{\star}} \ln(1 - t) dt \\ &= \mathcal{R}_0 ((x_{\star} - x_{\text{in}}) + (1 - x_{\star}) \ln(1 - x_{\star}) - (1 - x_{\text{in}}) \ln(1 - x_{\text{in}})) \\ &< \mathcal{R}_0 ((x_{\star} - x_{\text{in}}) + (1 - x_{\star}) \ln(1 - x_{\star}) - (1 - x_{\star}) \ln(1 - x_{\text{in}})) \\ &= \phi(x_{\star}), \end{aligned} \quad [38]$$

so the Laplace approximation and thus the original integral (Eq. 35a) are both exponentially large in  $\varepsilon^{-1}$ .

**Subsequent Epidemic Waves.** In ref. 34, we derive an iterative scheme to compute “effective initial conditions” for every epidemic wave following initial disease invasion. Writing  $x_{i,j}$  for the fraction susceptible at the start of the  $j$ th epidemic wave, we find our trajectory approximations agree very closely with the “exact” value obtained by solving the SIR ODEs (Eq. 5) numerically, starting from the DFE.

Setting  $x_{i,1} = 1$ , we iteratively obtain  $\bar{y}_{0,j}$  and  $x_{i,j+1}$  (Eq. 19) from  $x_{i,j}$  by computing

$$x_{f,j} = -x_\star W_0(\mathcal{E}(-x_{i,j}/x_\star)), \quad [39a]$$

$$\bar{y}_{0,j} = x_{i,j} - x_\star \left(1 + \ln(x_{i,j}/x_\star)\right). \quad [39b]$$

$$x_{i,j+1} = 1 + (1 - x_\star)W_0\left(\mathcal{E}\left(-\frac{1 - x_{f,j}}{1 - x_\star}\right)\right). \quad [39c]$$

Note that  $x_{f,j}$  and  $\bar{y}_{0,j}$  are the final fraction susceptible (i.e., when the pathogen has gone extinct) and maximal fraction infectious, respectively, for the SIR model without vital dynamics ( $\varepsilon = 0$ ) with initial condition  $(x_{i,j}, 0^+)$ .

**The Domain of Applicability of the Approximation (Eq. 19) to  $x_{in}$ .** The refined trajectory approximation that yields Eq. 19 is derived in ref. 34 under the assumption that  $\mathcal{R}_0$  is large. Despite this, we find that the approximation to  $x_{in}$  obtained from it (Eq. 19) performs very well for all but values of  $\mathcal{R}_0$  very close to 1 or very large values of  $\varepsilon > 0$  (Fig. 3). In particular,  $W_0(x)$  is undefined for  $x < -e^{-1}$ , so we must have

$$-\mathcal{R}_0 e^{-\mathcal{R}_0(1-y_\star)} > -e^{-1}, \quad [40]$$

or, expanding and rearranging using Eq. 8,

$$\varepsilon < 1 - \frac{\ln \mathcal{R}_0}{\mathcal{R}_0 - 1} = 1 + \frac{x_\star \ln x_\star}{1 - x_\star}. \quad [41]$$

Alternately, we can find an approximate lower bound for  $\mathcal{R}_0$ ,

$$\mathcal{R}_0 > e^{2\varepsilon}, \quad [42]$$

by observing that for any  $x \geq 1$ ,  $1 - \frac{\ln x}{x-1} \leq \frac{1}{2} \ln x$ . To derive this latter inequality, note that both sides approach a limit of 0 as  $x \rightarrow 1$ , whereas

$$\frac{d}{dx} \left(1 - \frac{\ln x}{x-1} - \frac{1}{2} \ln x\right) = \frac{1}{(x-1)^2} \left(\ln x - \frac{x^2 - 1}{2x}\right). \quad [43]$$

Again,  $\ln x - \frac{x^2 - 1}{2x}$  vanishes at  $x = 1$ , whereas

$$\frac{d}{dx} \left(\ln x - \frac{x^2 - 1}{2x}\right) = -\frac{(x-1)^2}{2x} \leq 0, \quad [44]$$

so  $\ln x - \frac{x^2 - 1}{2x} \leq 0$  for  $x \geq 1$ , and thus,  $\frac{d}{dx} \left(1 - \frac{\ln x}{x-1} - \frac{1}{2} \ln x\right) \leq 0$  also, proving the desired inequality.

**Boundary Layer Independent Estimates.** Thus far, we have computed the burnout probability via a specific, but arbitrary choice of boundary layer  $y_\star$ , and explicit solutions for  $x_{in}$ , the fraction susceptible when first entering the boundary layer under the ODE approximation (Eq. 5). Here, we consider an alternative approach, using results from refs. 34 and 59 to implicitly characterize  $x_{in}$ . In conjunction with Eq. 35, this allows us—at the cost of a small loss of precision—to give expressions for the extinction and persistence probabilities that are independent of the precise choice of threshold, provided the threshold is  $\mathcal{O}(\varepsilon)$ . In addition to being of interest in and of themselves, we use them to compute the value of  $\mathcal{R}_0$  maximizing the burnout probability (The  $\mathcal{R}_0$  Maximizing the Probability of Burnout) and also to show how one derives the result of ref. 16 as an approximation to Eq. 24 (Boundary Layer Independent Estimates).

In refs. 34 and 59, we use the method of matched asymptotic expansions (60, 61) to derive analytical approximations to the phase-plane trajectories of the SIR model with vital dynamics, i.e., expressions  $Y(x)$  and  $X(y)$  expressing the density of infectious hosts as a function of the density of susceptible hosts and vice versa. In the boundary layer, we obtain lowest- and first-order approximations to  $Y(x)$ : the lowest-order approximation (34) is

$$Y(x) \approx \bar{y}_0 \left(\frac{1 - x_f}{1 - x}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{\frac{\mathcal{R}_0}{\varepsilon}(x_f - x)}, \quad [45]$$

whereas the refined estimate (59) is

$$Y(x) \approx \left(\frac{1}{x_f} - 1\right) (x_\star - x_f) \left(\frac{1 - x_f}{1 - x}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} \times e^{-\frac{\mathcal{R}_0}{\varepsilon}(x - x_f) + \left(\frac{1}{x_f} - 1\right)^{-1} \mathcal{Y}_{x_f}^1(1)}, \quad [46]$$

where

$$x_f = -x_\star W_0(-\mathcal{R}_0 e^{-\mathcal{R}_0}) \quad [47]$$

is the final size of the SIR epidemic without vital dynamics (62) and

$$\begin{aligned} \mathcal{Y}_{x_f}^1(1) &= \int_{x_f}^1 \left[ \left(\frac{x_\star}{t} - 1\right) \left(\frac{1}{t} - 1\right) \frac{1}{1 - t + x_\star \ln u} \right. \\ &\quad \left. - \left(\frac{1}{x_f} - 1\right) \frac{1}{t - x_f} \right] dt \end{aligned} \quad [48a]$$

$$\approx \left(\frac{1}{x_f} - \frac{x_\star}{x_\star - x_f}\right) \ln x_f - \frac{x_\star}{x_\star - x_f} \left(\frac{1}{x_f} - 1\right). \quad [48b]$$

A very closely related expression (using  $\mu$  rather than  $\varepsilon$  as the small parameter) is derived in ref. 16.

Recalling that,  $Y(x_{in}) = y_\star$ , evaluating either of Eq. 45 or 46 at  $x = x_{in}$  gives us a relation between  $x_{in}$ ,  $x_f$ , and  $y_\star$ . From the former (Eq. 45), we have

$$\left(\frac{1}{1 - x_{in}}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{-\frac{\mathcal{R}_0}{\varepsilon}x_{in}} \approx \frac{y_\star}{\bar{y}_0} \left(\frac{1}{1 - x_f}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{-\frac{\mathcal{R}_0}{\varepsilon}x_f}, \quad [49]$$

whereas the latter (Eq. 46) gives us

$$\begin{aligned} \left(\frac{1}{1 - x_{in}}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{-\frac{\mathcal{R}_0}{\varepsilon}x_{in}} &\approx \frac{y_\star}{(1 - x_f) \left(\frac{x_\star}{x_f} - 1\right)} \left(\frac{1}{1 - x_f}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} \\ &\quad \times e^{-\frac{\mathcal{R}_0}{\varepsilon}x_f + \left(\frac{1}{x_f} - 1\right)^{-1} \mathcal{Y}_{x_f}^1(1)}. \end{aligned} \quad [50]$$

Substituting Eq. 49 into the integrand in Eq. 35a and proceeding as above gives

$$\mathcal{I}(x_{in}) \approx \frac{y_\star}{\bar{y}_0} \int_{x_{in}}^1 \left(\frac{1 - x}{1 - x_f}\right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{\frac{\mathcal{R}_0}{\varepsilon}(x - x_f)} \frac{dx}{\varepsilon(1 - x)} \quad [51a]$$

$$\begin{aligned} &= \frac{1}{\varepsilon} \frac{y_\star}{\bar{y}_0} \left(\frac{\mathcal{R}_0}{\varepsilon}(1 - x_f)\right)^{-\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_f)} \\ &\quad \times \mathcal{G}\left(\frac{\mathcal{R}_0}{\varepsilon}(1 - x_\star), \frac{\mathcal{R}_0}{\varepsilon}(1 - x_{in})\right). \end{aligned} \quad [51b]$$

Set  $z' = \frac{\mathcal{R}_0}{\varepsilon}(1 - x_f)$ . Then,  $\frac{a}{z}$  (Eqs. 24b and 24c) and  $\frac{a}{z'}$  are fixed, while as  $\varepsilon \rightarrow 0$ ,  $a \rightarrow \infty$  and  $\mathcal{G}(a, z) \sim \Gamma(a) - z^a e^{-z}$  (see ref. 36, §8.11.6) and similarly for  $\mathcal{G}(a, z')$ . Thus

$$(z')^{-a} e^{z'} (\mathcal{G}(a, z') - \mathcal{G}(a, z)) \sim \left(\frac{z}{z'}\right)^a e^{z' - z} - 1. \quad [52]$$

and the error in replacing  $x_{in}$  by  $x_f$  in the incomplete gamma function in Eq. 51b is equal to

$$\frac{1}{\varepsilon} \frac{y_\star}{\bar{y}_0} \left( \frac{1 - x_{in}}{1 - x_f} \right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{-\frac{\mathcal{R}_0}{\varepsilon}(x_f - x_{in})} \quad [53a]$$

$$= \frac{1}{\varepsilon} \frac{y_\star}{\bar{y}_0} e^{\frac{\mathcal{R}_0}{\varepsilon} \left[ (1-x_\star) \ln \left( \frac{1-x_{in}}{1-x_f} \right) - (x_f - x_{in}) \right]} \quad [53b]$$

$$= \frac{1}{\varepsilon} \frac{y_\star}{\bar{y}_0} e^{\frac{\mathcal{R}_0}{\varepsilon} \left[ (1-x_\star) \ln \left( 1 + \frac{x_f - x_{in}}{1-x_f} \right) - (x_f - x_{in}) \right]} \quad [53c]$$

$$= \frac{1}{\varepsilon} \frac{y_\star}{\bar{y}_0} e^{\frac{\mathcal{R}_0}{\varepsilon} \left[ (x_f - x_{in}) \frac{x_f - x_\star}{1-x_f} + \mathcal{O}(\varepsilon^2) \right]}. \quad [53d]$$

Both  $x_f - x_{in}$  and  $y_\star$  are  $\mathcal{O}(\varepsilon)$ , whereas  $\frac{x_f - x_\star}{1-x_f}$  is  $\mathcal{O}(1)$ , so this error is  $\mathcal{O}(1)$ . Thus, in absolute terms, the error is not small. However, as we observed above,  $\mathcal{I}(x_{in})$  is exponentially large in  $\varepsilon^{-1}$ , so the error is negligible relative to this leading term [indeed, replacing the incomplete gamma function by  $\Gamma\left(\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)\right)$  produces a similarly negligible error]. We can also replace  $x_{in}$  by  $x_f$  in the Laplace approximation with negligible error:

$$\mathcal{I}(x_{in}) \approx \frac{y_\star}{\bar{y}_0} \sqrt{\frac{2\pi}{\varepsilon(\mathcal{R}_0 - 1)}} \left( \frac{a}{z'} \right)^a e^{z' - a}. \quad [54]$$

Similarly, repeating the same argument using the higher-order expression, Eq. 50, gives

$$\begin{aligned} \mathcal{I}(x_{in}) &\approx \frac{1}{\varepsilon} \frac{y_\star}{(1-x_f) \left( \frac{1}{\mathcal{R}_0 x_f} - 1 \right)} (z')^{-a} \\ &\times e^{z' + \left( \frac{1}{x_f} - 1 \right)^{-1} \mathcal{Y}_{x_f}^1(1)} \mathcal{G}(a, z') \end{aligned} \quad [55a]$$

$$\begin{aligned} &\approx \frac{y_\star}{(1-x_f) \left( \frac{1}{\mathcal{R}_0 x_f} - 1 \right)} \sqrt{\frac{2\pi}{\varepsilon(\mathcal{R}_0 - 1)}} \left( \frac{a}{z'} \right)^a \\ &\times e^{z' - a + \left( \frac{1}{x_f} - 1 \right)^{-1} \mathcal{Y}_{x_f}^1(1)}. \end{aligned} \quad [55b]$$

Now, we recall from Eqs. 14 and 22 that the burnout probability is

$$q(x_{in})^{ny_\star} = \left( 1 + \frac{1}{\mathcal{I}(x_{in})} \right)^{-ny_\star} \quad [56a]$$

$$= e^{-ny_\star \ln \left( 1 + \frac{1}{\mathcal{I}(x_{in})} \right)} \quad [56b]$$

$$\approx e^{-\frac{ny_\star}{\mathcal{I}(x_{in})}}. \quad [56c]$$

Above, we showed that  $\mathcal{I}(x_{in})$  is exponentially large in  $\varepsilon^{-1}$  as  $\varepsilon \rightarrow 0$  and, thus, that the error in making the last approximation (Eq. 56c) is exponentially small.

Substituting any of the expressions Eq. 51b, 54, 55a, or 55b for  $\mathcal{I}(x_{in})$  in Eq. 56c, we see that the factors  $y_\star$  cancel, giving us an approximate expression for the burnout probability that does not depend on the specific choice of threshold, only upon its order of magnitude,  $\varepsilon$ :

$$\begin{aligned} q(x_{in})^{ny_\star} &\approx e^{-\frac{n\bar{y}_0}{(z')^{-a} e^{z'} \mathcal{G}(a, z')}} \\ &\approx e^{-n\bar{y}_0 \sqrt{\frac{\varepsilon(\mathcal{R}_0 - 1)}{2\pi}}} \left( \frac{z'}{a} \right)^a e^{z' - a}, \end{aligned} \quad [57a]$$

$$[57b]$$

or

$$q(x_{in})^{ny_\star} \approx \exp \left[ -\frac{n\varepsilon(1-x_f) \left( \frac{1}{\mathcal{R}_0 x_f} - 1 \right)}{(z')^{-a} e^{z' + \left( \frac{1}{x_f} - 1 \right)^{-1} \mathcal{Y}_{x_f}^1(1)} \mathcal{G}(a, z')} \right] \quad [58a]$$

$$\approx e^{-n(1-x_f) \left( \frac{1}{\mathcal{R}_0 x_f} - 1 \right) \sqrt{\frac{\varepsilon(\mathcal{R}_0 - 1)}{2\pi}} \left( \frac{z'}{a} \right)^a e^{z' - a - \left( \frac{1}{x_f} - 1 \right)^{-1} \mathcal{Y}_{x_f}^1(1)}}, \quad [58b]$$

respectively.

**Remark 2:** If in Eq. 58a we approximate  $\mathcal{G}(a, z')$  by  $\Gamma(a)$  (i.e., if we approximate the integral up to  $z' \gg 1$  by the integral over the whole real line, introducing an error of  $\mathcal{O}(\varepsilon)$ ), we obtain an expression for the burnout probability equivalent to that from ref. 16 (up to minor differences resulting from using different small parameters,  $\mu$  and  $\varepsilon$ ).

**The  $\mathcal{R}_0$  Maximizing the Probability of Burnout.** Using the simplified expression for the burnout probability (Eq. 57b), we can obtain an approximation to the value of  $\mathcal{R}_0$  that maximizes the probability of burnout linear in  $\varepsilon$  which is highly accurate across the range of values of  $\varepsilon$  for which the burnout probability is non-monotone. Eq. 57b is minimized when

$$\bar{y}_0 \sqrt{\frac{\varepsilon(\mathcal{R}_0 - 1)}{2\pi}} \left( \frac{1 - x_f}{1 - x_\star} \right)^{\frac{\mathcal{R}_0}{\varepsilon}(1-x_\star)} e^{\frac{\mathcal{R}_0}{\varepsilon}(x_f - x_\star)} \quad [59]$$

is maximized, or equivalently, when its partial derivative with respect to  $\mathcal{R}_0$  is equal to zero. Computing the partial derivative and collecting terms of like order in  $\varepsilon$ , we seek  $\mathcal{R}_0$  such that

$$\begin{aligned} \frac{1}{\varepsilon} \left[ \ln \left( \frac{\mathcal{R}_0 + W_0(-\mathcal{R}_0 e^{-\mathcal{R}_0})}{\mathcal{R}_0 - 1} \right) - \frac{1}{\mathcal{R}_0} \right] \\ + \frac{\ln \mathcal{R}_0}{\mathcal{R}_0(\mathcal{R}_0 - 1 - \ln \mathcal{R}_0)} + \frac{\sqrt{\mathcal{R}_0 - 1}}{2} = 0. \end{aligned} \quad [60]$$

An analytical closed-form solution does not appear to exist, but one can use a formal asymptotic series expansion  $\mathcal{R}_0 = \sum_{j=0}^{\infty} r_j \varepsilon^j$  to obtain a polynomial approximation in  $\varepsilon$  to arbitrarily large degree (here, we content ourselves with a linear approximation). Substituting this series into Eq. 60 and collecting terms of order  $\varepsilon^{-1}$  and order one, we obtain

$$\ln \left( \frac{r_0 + W_0(-r_0 e^{-r_0})}{r_0 - 1} \right) - \frac{1}{r_0} = 0, \quad [61]$$

$$\begin{aligned} -\frac{1 + (r_0^2 - r_0 + 1)W_0(-r_0 e^{-r_0})}{r_0^2(r_0 - 1)(1 + W_0(-r_0 e^{-r_0}))} \\ + \frac{\sqrt{r_0 - 1}}{2} + \frac{\ln r_0}{r_0(r_0 - 1 - \ln r_0)} = 0. \end{aligned} \quad [62]$$

We may solve Eq. 61 by Newton iteration to find the unique root  $r_0 = 2.572629848$ , which we use to solve Eq. 62 to find  $r_1 = -27.71866282$ , giving us the linear approximation

$$\arg \max_{\mathcal{R}_0 > 1} q(x_{in})^{ny_\star} \approx 2.572629848 - 27.71866282 \varepsilon. \quad [63]$$

We compare this linear approximation to the numerically determined minimum in Fig. 5.

**The Burnout Probability is a Decreasing Function of  $\varepsilon$ .** In what follows, we show that  $\frac{\partial q(x_{in})}{\partial \varepsilon} \leq 0$ , from which we conclude that  $q(x_{in})$  is decreasing as  $\varepsilon$  increases, for all values of  $\mathcal{R}_0$ . Using Eq. 24, we have

$$\frac{\partial q(x_{in})}{\partial \varepsilon} = -q(x_{in})^2 \left( \frac{1}{e^{2z-a} \mathcal{G}(a, z)} - \frac{\varepsilon \frac{\partial}{\partial \varepsilon} [e^{z-a} \mathcal{G}(a, z)]}{e^{2z-2a} \mathcal{G}(a, z)^2} \right). \quad [64]$$

The first term in the large brackets on the right-hand side is always positive, so the result follows if one can show that  $\varepsilon \frac{\partial}{\partial \varepsilon} [e^{z-a} \mathcal{G}(a, z)] \leq 0$ . Applying the chain rule gives

$$\varepsilon \frac{\partial}{\partial \varepsilon} [e^{z-a} \mathcal{G}(a, z)] \quad [65a]$$

$$= \varepsilon \frac{\partial z}{\partial \varepsilon} \frac{\partial}{\partial z} [e^z z^{-a} g(a, z)] + \varepsilon \frac{\partial a}{\partial \varepsilon} \frac{\partial}{\partial a} [e^z z^{-a} g(a, z)] \quad [65b]$$

$$= -\left(z + \mathcal{R}_0 \frac{\partial x_{in}}{\partial \varepsilon}\right) \frac{\partial}{\partial z} [e^z z^{-a} g(a, z)] \\ - a \frac{\partial}{\partial a} [e^z z^{-a} g(a, z)]. \quad [65c]$$

Recalling that  $g(a, z) = \int_0^z t^{a-1} e^{-t} dt$ , the latter is equal to

$$-\left(z + \mathcal{R}_0 \frac{\partial x_{in}}{\partial \varepsilon}\right) \left(e^z z^{-a} g(a, z) \left(1 - \frac{a}{z}\right) + \frac{1}{z}\right) \\ - a e^z z^{-a} \left(\int_0^z t^{a-1} e^{-t} \ln t dt - g(a, z) \ln z\right). \quad [66]$$

Integrating by parts in the rightmost term, this becomes

$$-\left(z + \mathcal{R}_0 \frac{\partial x_{in}}{\partial \varepsilon}\right) \left(e^z z^{-a} g(a, z) \left(1 - \frac{a}{z}\right) + \frac{1}{z}\right) \\ - a e^z z^{-a} \int_0^z \frac{g(a, t)}{t} dt \\ = -\mathcal{R}_0 \frac{\partial x_{in}}{\partial \varepsilon} \left(e^z z^{-a} g(a, z) \left(1 - \frac{a}{z}\right) + \frac{1}{z}\right) \\ - e^z z^{-a+1} g(a, z) - 1 \\ - e^z z^{-a} \int_0^z \left(\frac{a}{t} g(a, t) - \frac{a}{z} g(a, z)\right) dt. \quad [67]$$

Now,  $g(a, z) \geq 0$ , whereas  $\frac{a}{z} = \frac{1-x_\star}{1-x_{in}} \leq 1$ , since  $x_{in} < x_\star$ , so  $1 - \frac{a}{z} \geq 0$  and, since  $g(a, z)$  is an increasing function of  $z$ , we have

$$\int_0^z \left(\frac{a}{t} g(a, t) - \frac{a}{z} g(a, z)\right) dt \geq \int_0^z \left(\frac{a}{t} - \frac{a}{z}\right) g(a, t) dt \geq 0. \quad [68]$$

Thus, provided  $\frac{\partial x_{in}}{\partial \varepsilon} \geq 0$ ,  $\varepsilon \frac{\partial}{\partial \varepsilon} [e^z z^{-a} g(a, z)] \leq 0$ , as required.

Finally, from Eq. 19, we see that

$$\frac{\partial x_{in}}{\partial \varepsilon} = \lim_{\varepsilon \rightarrow 0} e^{\mathcal{R}_0 y_\star} (E_1(\mathcal{R}_0 y_\star) - E_1(\mathcal{R}_0 \bar{y}_0)) \geq 0, \quad [69]$$

since  $y_\star \leq \bar{y}_0$  and  $E_1(x)$  is a decreasing function of  $x$ .

### Simulations.

**Stochastic simulation algorithm.** Exact realizations of the stochastic SIR model (Fig. 2 and Table 1) can be obtained using the standard Gillespie algorithm (63, 64). If we denote the various event rates  $a_i$  (e.g.,  $a_1 = \mu n$ , etc.), then the total event rate is  $a = \sum_i a_i$ . The time to the next event is drawn from an exponential distribution with mean  $1/a$ , and the event is taken to be of type  $i$  with probability  $a_i/a$ . This algorithm scales with expected population size  $n$  and is prohibitively slow when running large numbers of simulations with  $n \gtrsim 10^5$ . We therefore used the adaptive  $\tau$ -leaping approximation (65), as implemented in the adaptivetau R package (66). The key idea in this approach is to identify, at any point of the simulation, a time  $\tau$  over which the various event rates can be considered approximately constant, and then determine the number of events of each type that can be expected over this time interval. We then "leap forward" by time  $\tau$  rather than treating events individually.

**Estimating the Required Number of Simulations.** To determine the number of simulations required to estimate the epidemic burnout probability to a given accuracy, we use the central limit theorem. Suppose we run  $m$  independent simulations. Let

$$\mathbb{1}_i = \begin{cases} 1 & \text{if the } i\text{th simulation ends in burnout, and} \\ 0 & \text{otherwise.} \end{cases} \quad [70]$$

Then, the law of large numbers (67, §6) tells us that

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbb{1}_i = \mathbb{E}[\mathbb{1}_1] = q, \quad [71]$$

where  $q$  is Kendall's  $q$  (Eq. 14). Consequently,  $q_m = \frac{1}{m} \sum_{i=1}^m \mathbb{1}_i$  is an unbiased estimator (67, p. 483) of  $q$ . Let  $\Delta = q - q_m$  be the error in our estimates. Then, the central limit theorem (67, §27) tells us that  $\sqrt{m}\Delta = \frac{1}{\sqrt{m}} \sum_{i=1}^m (\mathbb{1}_i - q)$  converges to a normal distribution with the same variance as  $\mathbb{1}_1 - q$ , i.e.,  $\sqrt{m}\Delta$  converges in distribution to a normal random variable with variance

$$\sigma^2 = \mathbb{E}[(\mathbb{1}_1 - q)^2] = \mathbb{E}[\mathbb{1}_1^2 - 2q\mathbb{1}_1 + q^2] \quad [72a]$$

$$= (1^2 \cdot q + 0^2 \cdot (1-q)) - 2q(1 \cdot q + 0 \cdot (1-q)) + q^2 \quad [72b]$$

$$= q(1-q) \leq \frac{1}{4}, \quad [72c]$$

where the inequality in Eq. 72c follows because  $0 \leq q \leq 1$ . In particular, for large  $m$ , the expected squared error is  $\mathbb{E}[\Delta^2] \lesssim \frac{1}{4m}$ , and thus, to have  $\mathbb{E}[\Delta^2] \leq \delta$ , we perform at least  $m = \lceil \frac{1}{4\delta} \rceil$  runs.

**Fizzle vs. Epidemic Burnout.** To efficiently distinguish fizzles from epidemic burnout, we use Eq. 15 to estimate a time  $\tau_\delta$  (measured in units of the mean infectious period  $1/(\gamma + \mu)$ ) such that the probability is less than  $\delta$  that, starting from  $k$  infectious individuals and  $x_i = 1 - \frac{k}{n}$ , a sample path in which infective individuals are still present at time  $\tau_\delta$  eventually fizzles. Let  $T_k$  be the (random) time of fizzle starting from  $k$  individuals. Then, using our birth-and-death process approximation,

$$\mathbb{P}\{T_k > t\} \\ = \mathbb{P}\{I_n(t) > 0 \mid I_n(0) = k\} \quad [73a]$$

$$\approx 1 - \left(1 + \frac{1}{\int_0^t e^{-\int_0^s [\beta - (\gamma + \mu)] du} (\gamma + \mu) ds}\right)^{-k} \quad [73b]$$

$$= 1 - \left(1 + \frac{1}{\frac{1}{\mathcal{R}_0 - 1} (1 - e^{-(\beta - \gamma - \mu)t})}\right)^{-k}. \quad [73c]$$

Now, because fizzle is not a certainty,

$$\lim_{t \rightarrow \infty} \mathbb{P}\{T_k > t\} = 1 - x_\star^k > 0. \quad [74]$$

To determine  $\tau_\delta$ , we condition on eventual fizzle to estimate its time of occurrence:

$$\mathbb{P}\{T_k > t \mid T_k < \infty\} \\ = \frac{\mathbb{P}\{T_k > t\} - \mathbb{P}\{T_k = \infty\}}{\mathbb{P}\{T_k < \infty\}} \quad [75a]$$

$$\approx \frac{\left(\frac{1}{\mathcal{R}_0}\right)^k - \left(1 + \frac{1}{\frac{1}{\mathcal{R}_0 - 1} (1 - e^{-(\beta - \gamma - \mu)t})}\right)^{-k}}{\left(\frac{1}{\mathcal{R}_0}\right)^k} \quad [75b]$$

$$= 1 - \left(1 + \frac{1}{\frac{1}{\mathcal{R}_0 - 1} (1 - e^{-(\beta - \gamma - \mu)t})}\right)^{-k}. \quad [75c]$$

Solving for  $\mathbb{P}\{T_k > \tau_\delta \mid T_k < \infty\} = \delta$  yields

$$\tau_\delta = \frac{1}{\mathcal{R}_0 - 1} \ln \left( \frac{(1 - \delta)^{-\frac{1}{k}} - \frac{1}{\mathcal{R}_0}}{(1 - \delta)^{-\frac{1}{k}} - 1} \right). \quad [76]$$

Choosing a suitably small  $\delta$ , we assume that any sample path in which infective individuals are still present at  $\tau_\delta$  will not fizzle.

**Data, Materials, and Software Availability.** All study data are included in the main text. Our open-source R package, which we used to create our figures, is available at <https://github.com/davidearn/burnout> (68).

**ACKNOWLEDGMENTS.** This project was partially supported by the CNRS International Emerging Actions (IEA) grant “Structured Populations, Epidemics, and Control Strategies (SPECs).” D.J.D.E., J.D., and B.M.B. were supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

We thank David Champredon, Michelle deJonge, Sarah Drohan, Karsten Hempel, Chai Molina, Irena Papst, and Dora Rosati for their contributions to the preliminary work that led to ref. 38.

Author affiliations: <sup>a</sup>Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, CNRS UMR 8001, Paris 75005, France; <sup>b</sup>Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada; <sup>c</sup>Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario L8S 4K1, Canada; and <sup>d</sup>Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario L8S 4K1, Canada

1. F. Brauer, C. Castillo-Chavez, Z. Feng, *Mathematical Models in Epidemiology* (Springer, 2019), vol. 32.
2. M. S. Bartlett, Measles periodicity and community size. *J. R. Stat. Soc., Ser. A* **120**, 48–70 (1957).
3. W. London, J. A. Yorke, Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *Am. J. Epidemiol.* **98**, 453–468 (1973).
4. R. M. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, UK, 1991).
5. D. J. D. Earn, P. Rohani, B. M. Bolker, B. T. Grenfell, A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670 (2000).
6. K. Hempel, D. J. D. Earn, A century of transitions in New York City’s measles dynamics. *J. R. Soc. Lond. Interface* **12**, 20150024 (2015).
7. D. Mollison, Dependence of epidemic and population velocities on basic parameters. *Math. Biosci.* **107**, 255–287 (1991).
8. S. N. Ethier, T. G. Kurtz, *Markov Processes: Characterization and Convergence* (John Wiley and Sons, New York, NY, 1986).
9. J. Dushoff, “Incorporating stochasticity in simple models of disease spread” in *Modeling Paradigms and Analysis of Disease Transmission Models*, A. B. Gumel, S. Lenhart, Eds. (American Mathematical Society, 2010), vol. 75.
10. P. Jagers, Stabilities and instabilities in population dynamics. *J. Appl. Prob.* **29**, 770–780 (1992).
11. O. A. Van Herwaarden, J. Grasman, Stochastic epidemics: Major outbreaks and the duration of the endemic period. *J. Math. Biol.* **33**, 581–601 (1995).
12. I. Nåsell, On the time to extinction in recurrent epidemics. *J. R. Stat. Soc., Ser. B* **61**, 309–330 (1999).
13. A. Kamenev, B. Meerson, Extinction of an infectious disease: A large fluctuation in a nonequilibrium system. *Phys. Rev. E* **77**, 061107 (2008).
14. O. Diekmann, J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (John Wiley & Sons, New York, NY, 2000).
15. T. Britton *et al.*, Five challenges for stochastic epidemic models involving global transmission. *Epidemics* **10**, 54–57 (2015).
16. O. A. van Herwaarden, Stochastic epidemics: The probability of extinction of an infectious disease at the end of a major outbreak. *J. Math. Biol.* **35**, 793–813 (1997).
17. B. Meerson, P. V. Sasorov, WKB theory of epidemic fade-out in stochastic populations. *Phys. Rev. E* **80**, 041130 (2009).
18. P. G. Ballard, N. G. Bean, J. V. Ross, The probability of epidemic fade-out is non-monotonic in transmission rate for the Markovian SIR model with demography. *J. Theor. Biol.* **393**, 170–178 (2016).
19. S. Karlin, H. M. Taylor, *A Second Course in Stochastic Processes* (Academic Press, San Diego, CA, 1981).
20. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin/Heidelberg, Germany, 2004).
21. J. Grasman, O. A. van Herwaarden, *Asymptotic Methods for the Fokker-Planck Equation and the Exit Problem in Applications* (Springer, Berlin/Heidelberg, Germany, 1999).
22. C. M. Bender, S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, NY, 1978).
23. R. Graham, T. Tel, Existence of a potential for dissipative dynamical systems. *Phys. Rev. Lett.* **52**, 9–12 (1984).
24. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications, Applications of Mathematics* (Springer, New York, NY, ed. 2, 1998), vol. 38.
25. W. Feller, “Diffusion processes in genetics” in *Proceedings on the Second Berkeley Symposium on Mathematical Statistics and Probability, July 31–August 12, 1950*, J. Neyman, Ed. (University of California Press, Berkeley, CA, 1951), pp. 227–246.
26. J. Lamperti, The limit of a sequence of branching processes. *Z. Wahrschein. Verw. Gebiete* **7**, 271–288 (1967).
27. H. Andersson, T. Britton, *Stochastic Epidemic Models and their Statistical Analysis* (Springer, New York, NY, 2000), vol. 151.
28. H. W. Hethcote, H. W. Stech, P. van den Driessche, “Periodicity and stability in epidemic models: A survey” in *Differential Equations and Applications in Ecology, Epidemics, and Population Problems*, S. N. Busenberg, K. L. Cooke Eds. (Elsevier, 1981), pp. 65–82.
29. A. Korobeinikov, G. C. Wake, Lyapunov functions and global stability for SIR, SIRS, and SIS epidemiological models. *Appl. Math. Lett.* **15**, 955–960 (2002).
30. M. S. Bartlett, “Deterministic and stochastic models for recurrent epidemics” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (1956), vol. 4, pp. 81–108.
31. D. G. Kendall, On the generalized “Birth-and-Death” process. *Ann. Math. Stat.* **19**, 1–15 (1948).
32. T. L. Parsons, A. Lambert, T. Day, S. Gandon, Pathogen evolution in finite populations: Slow and steady spreads the best. *J. Royal Soc. Interface* **15**, 20180135 (2018).
33. T. L. Parsons, Invasion probabilities, hitting times, and some fluctuation theory for the stochastic logistic process. *J. Math. Biol.* **77**, 1193–1231 (2018).
34. T. L. Parsons, D. J. D. Earn, Analytical approximations for the phase plane trajectories of the SIR model with vital dynamics. [Preprint] (2023). <https://cnrs.hal.science/hal-04178969> (Accessed 8 August 2023).
35. R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth, On the Lambert  $W$  function. *Adv. Comput. Math.* **5**, 329–359 (1996).
36. F. W. J. Olver, D. W. Lozier, R. F. Boisvert, C. W. Clark, *NIST Handbook of Mathematical Functions* (Cambridge University Press, 2010).
37. H. W. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
38. D. J. D. Earn *et al.*, “The puzzling persistence of invading pathogens” in *Poster presentation at 2014 Ecology and Evolution of Infectious Diseases Conference “Multi-Scale Mechanisms of Disease Emergence and Control”* (Colorado State University, 2014).
39. Z. Levine, D. J. D. Earn, Face masking and COVID-19: Potential effects of variolation on transmission dynamics. *J. R. Soc. Interface* **19**, 20210781 (2022).
40. O. Krylova, “Predicting epidemiological transitions in infectious disease dynamics: Smallpox in historic London (1664–1930).” Ph.D. McMaster University, Canada (2011).
41. T. D. Hollingsworth, R. M. Anderson, C. Fraser, HIV-1 transmission, by stage of infection. *J. Infect. Dis.* **198**, 687–693 (2008).
42. C. E. Mills, J. M. Robins, M. Lipsitch, Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904–906 (2004).
43. Z. Y. Wong, C. M. Bui, A. A. Chughtai, C. R. Macintyre, A systematic review of early modelling studies of Ebola virus disease in West Africa. *Epidemiol. Infect.* **145**, 1069–1094 (2017).
44. R. Gani, S. Leach, Epidemiologic determinants for modeling pneumonic plague outbreaks. *Emerg. Infect. Dis.* **10**, 608–614 (2004).
45. O. Krylova, D. J. D. Earn, Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J. R. Soc. Lond. Interface* **10**, 20130098 (2013).
46. D. Champredon, J. Dushoff, D. J. D. Earn, Equivalence of the Erlang SEIR epidemic model and the renewal equation. *SIAM J. Appl. Math.* **78**, 3258–3278 (2018).
47. A. A. King, S. Shrestha, E. T. Harvill, O. N. Bjørnstad, Evolution of acute infections and the invasion-persistence trade-off. *Am. Nat.* **173**, 446–455 (2009).
48. I. Nåsell, “The threshold concept in stochastic epidemic and endemic models” in *Epidemic Models: Their Structure and Relation to Data*, D. Mollison, Ed. (Cambridge University Press, 1995), pp. 71–83.
49. S. Gandon, A. Lambert, T. Day, T. L. Parsons, The speed of vaccination rollout and the risk of pathogen adaptation. medRxiv [Preprint] (2022). <https://doi.org/10.1101/2022.08.01.22278283> (Accessed 13 August 2022).
50. D. J. D. Earn, J. Dushoff, S. A. Levin, Ecology and evolution of the flu. *Trends Ecol. Evol.* **17**, 334–340 (2002).
51. B. T. Grenfell, J. Harwood, (Meta)population dynamics of infectious diseases. *Trends Ecol. Evol.* **12**, 395–399 (1997).
52. D. J. D. Earn, P. Rohani, B. T. Grenfell, Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. Lond. B* **265**, 7–10 (1998).
53. K. Hampson, D. Haydon, Persistent pathogens and wildlife reservoirs. *Science* **374**, 35–36 (2021).
54. D. T. Haydon, S. Cleaveland, L. H. Taylor, M. K. Laurenson, Identifying reservoirs of infection: A conceptual and practical challenge. *Emerg. Infect. Dis.* **8**, 1468–1473 (2002).
55. C. J. Mode, *Multitype Branching Processes: Theory and Applications* (Elsevier, New York, NY, 1971).
56. P. Jagers, *Branching Processes with Biological Applications* (Wiley, London, UK, 1975).
57. C. McCluskey, D. J. D. Earn, Attractivity of coherent manifolds in metapopulation models. *J. Math. Biol.* **62**, 509–541 (2011).
58. D. J. D. Earn, S. A. Levin, P. Rohani, Coherence and conservation. *Science* **290**, 1360–1364 (2000).
59. T. L. Parsons, D. J. D. Earn, Refined asymptotic approximations for the phase plane trajectories of the SIR model with vital dynamics. [Preprint] (2023). <https://cnrs.hal.science/hal-04178983> (Accessed 8 August 2023).
60. R. E. O’Malley Jr, *Singular Perturbation Methods for Ordinary Differential Equations* (Springer, 1991), vol. 89.
61. J. Kevorkian, J. D. Cole, *Multiple Scale and Singular Perturbation Methods* (Springer-Verlag New York Inc., New York, NY, 1996).
62. J. Ma, D. J. D. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* **68**, 679–702 (2006).
63. D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
64. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
65. D. T. Gillespie, Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55 (2007).
66. P. Johnson, *adaptivedtau: Tau-Leaping Stochastic Simulation. R package v 2.2-3* (2019) <https://CRAN.R-project.org/package=adaptivedtau>.
67. P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Statistics (John Wiley & Sons, Hoboken, NJ, Anniversary ed., 2012).
68. D. J. D. Earn, B. M. Bolker, J. Dushoff, T. L. Parsons, Burnout: a package for computing infectious disease burnout and persistence probabilities. R package version 0.0.1. Github. <https://github.com/davidearn/burnout> (Accessed 13 August 2023).