When do epidemics end? Scientific insights from mathematical modelling studies

Keywords: epidemics, disease elimination, infectious disease, end-of-outbreak declaration, epidemiological modelling

Abstract

Quantitative assessments of when infectious disease outbreaks end are crucial, as resources targeted towards outbreak response typically remain in place until outbreaks are declared over. Recent improvements and innovations in mathematical approaches for determining when outbreaks end provide public health authorities with more confidence when making end-of-outbreak declarations. Although quantitative analyses of outbreaks have a long history, complex mathematical and statistical methodologies for analysing outbreak data were developed early in the twentieth century and continue to be refined. Historically, such methodologies focused primarily on factors affecting the early and middle phases of an outbreak, with less attention given to determining how and when outbreaks end. This review discusses mathematical modelling methods from the last 20 years that have been developed for determining the ends of infectious disease outbreaks, and consider the factors that affect the accuracy of such determinations. When surveillance systems provide timely and representative data to inform models, the timings of end-of-outbreak declarations can be fine-tuned to allow outbreaks to be declared over quickly and with a low risk of being incorrect. Premature declarations that outbreaks are over may undermine earlier achievements in disease control and may result in a resurgence of cases, but unnecessary delays in declaring outbreaks over cause significant economic and social harms. Appropriate declarations that balance the benefits of releasing control measures against the risk of a surge in cases thereby allow public health resources to be conserved (and economic and social pressures to be reduced) while limiting the potential for additional transmission.

Main Text

Determining when infectious disease outbreaks (used synonymously with epidemics, see Box 1) end is a critical component of outbreak response. In recent or ongoing outbreaks, such as those of coronavirus disease 2019 (COVID-19) or pandemic influenza, mathematical modelling has played a central role in policy and public understanding of the course of outbreaks. Although quantitative analyses of outbreaks have a long history, complex mathematical and statistical methodologies for analysing outbreak data were developed early in the twentieth century and continue to be refined. Historically, such methodologies focused primarily on factors affecting the early and middle phases of an outbreak, with less attention given to determining how and when outbreaks end. In this article, we review mathematical modelling methods from the last 20 years that have been developed for determining the ends of infectious disease outbreaks, and consider the factors that affect the accuracy of such determinations. Rather than setting out a history of epidemiological modelling, here we explain the function and methods of quantitative disease modelling, drawing on key practices and case studies to demonstrate the applications, advantages, and limitations of the disciplinary methodologies that have become crucial to understand outbreaks and how they end.

Development of statistical methods for determining the ends of infectious disease outbreaks has accelerated in recent years, as large cross-border outbreaks attract significant public attention and consume resources that cannot be fully demobilised until the outbreak is declared over. Large-scale outbreaks and associated control interventions can have a profound impact on lives and livelihoods, whether the implicated disease is one of humans, animals, or plants.[1] As such, it is in the public interest to be able to declare an outbreak over as soon as possible. However such declarations, if premature, may undermine earlier effective control measures and result in a resurgence of cases. Declaring the end of an outbreak based on rigorous quantitative methods helps to ensure that a valid risk assessment has been carried out and that there is only a limited risk of a false declaration, while avoiding interventions that remain in place long after transmission has ceased.

[1] For examples of each, see: Ebata, Nisbett, & Gillespie, 2021; Sonaiya, 2007; Thiermann, 2004; R. N. Thompson & Brooks-Pollock, 2019

To understand how outbreaks end, it is first important to understand how outbreaks are defined.[2] Outbreaks occur when the number of cases of a disease exceeds the expected range. The upper range of the expected number of cases forms the basis for an "epidemic threshold" which, when exceeded, signals the need for an urgent public health response. If a disease is endemic (consistently found in a region or population), the epidemic threshold may be determined by historical observations. If a disease is non-endemic (no sustained transmission in a region or population) or newly emerged, the epidemic threshold may be simply "more than zero" cases (for outbreak-prone diseases) or evidence of ongoing transmission (for diseases not prone to outbreaks).

As shown in Figure 1, outbreaks can be thought of as occurring in cycles, with the time between cycles—the holding period, during which the disease is either at endemic levels or eliminated—varying by disease, location, and other related factors. In situations in which a disease is endemic, cyclicality is often seasonal, as is the case for Lassa fever in West Africa, tickborne encephalitis in parts of Siberia, and seasonal influenza.[3] In non-endemic situations, outbreaks only occur when the disease is (re)introduced, potentially during an importation event or spillover from an animal reservoir.[4]

Endemic status is often ascribed according to geographic boundaries (i.e., country or province). However, it should be noted that incidence is usually unequally distributed within these boundaries, and this can lead to stretches of zero cases being reported even for endemic diseases due to patterns of localized extinction.[5] The use of the term "extinction" in mathematical models does not describe a permanent state, but rather is used to express a short-term decrease in cases to zero (Box 1). In contrast, non-endemicity is often described in terms of elimination, which represents a long-term or permanent reduction of incidence to zero cases (i.e., no transmission of the pathogen). At a global level, a challenging (and often impossible) aim may be eradication, which refers to a scenario in which a pathogen no longer exists in hosts anywhere in the world. Following eradication, continued actions to prevent transmission are no longer required. In contrast, for diseases that have been eliminated (but not

[2] Historiographical article on how outbreaks end: Charters & Heitman, 2021. For definitions of different endpoints see Dowdle, 1998 and Box 1.
[3] For an example for Lassa fever see Lo Iacono et al., 2016; for tickborne encephalitis, see Korenberg, 2000; for seasonal influenza, see Dalziel et al., 2018.
[4] Examples of importation events: R. N. Thompson, 2020; R. N. Thompson, Thompson, Pelerman, Gupta, & Obolski, 2018; Wilson, 2005. Examples of spillover from an animal reservoir: Borremans, Faust, Manlove, Sokolow, & Lloyd-Smith, 2019; Lloyd-Smith et al., 2009; Plowright et al., 2017; Ponce, Kinoshita, & Nishiura, 2019.
[5] Eichner & Dietz, 1996; Keeling & Grenfell, 2002; Lindholm & Britton, 2007; Britton & Neal, 2010; Holme, 2013.

eradicated), continued actions may be required to prevent substantial pathogen transmission from restarting (e.g., monitoring of inbound travellers for infection to reduce the risk of pathogen reimportation).[6] Elimination may also be defined in terms of "elimination as a public health problem," which typically indicates reaching a low level of prevalence rather than sustaining zero cases. This is important for many pathogens, as it is often improbable to reach zero cases for an extended period of time, with factors such as numerous possible hosts, varying modes of transmission, or difficulties in implementing surveillance and control measures decreasing the chance of complete elimination.

Establishing that an outbreak is over is often less clear-cut than confirming that an outbreak has started. A dip in cases below the epidemic threshold (to endemic levels or to zero cases) does not necessarily mean that the outbreak is over, as cases could soon return to epidemic levels. This was evidenced by the 2013–2016 Ebola virus disease (EVD) epidemic in West Africa, when new cases appeared soon after five of the occasions on which the outbreak was declared over in three of the countries involved.[7] Instead, it may be necessary to define an analogous "extinction threshold" which, rather than being defined based on a simple rule about numbers of reported cases, is instead established using quantitative analyses that aim to assess the probability that the outbreak is over (according to a specified definition of an outbreak end – see Box 1). When the extinction threshold is met, an outbreak can be declared over with confidence. For example, an outbreak might be declared over on the first date on which, according to a mathematical model, the probability that cases occur in future (as a direct result of recent transmission) is estimated to fall below a pre-specified threshold.

Predicting whether or not an outbreak has ended, or when an outbreak will end, is particularly complicated in the case of a recently emerged disease such as COVID-19, as there is limited historical data to inform predictions. In fact, as of mid-2021, more than a year and a half after the first COVID-19 cases were detected, it remains to be seen whether the disease will be eliminated long-term in any given location or whether it will instead become endemic worldwide. If the disease burden of COVID-19 becomes comparable to or lower than that of other existing infectious diseases, such as seasonal influenza, it may become accepted as endemic and local epidemic thresholds established.

[6] Sachak-Patwa, Byrne, Dyson, & Thompson, 2021.
[7] See Lee & Nishiura, 2017.

Avoiding both false declarations of outbreak ends and unnecessary waiting times prior to end-of-outbreak declarations has been a recent focus of infectious disease modelling efforts.[8] Improving the timing of end-of-outbreak declarations helps conserve public health resources and reduce pressure on healthcare, social, and economic systems. This review summarises quantitative methods used to infer when infectious disease outbreaks are over from a modern-day perspective, with some historical context interspersed throughout. In the following sections, we first introduce the concept of establishing freedom from disease. Then, we introduce epidemiological modelling methods used to estimate outbreak end times. Lastly, we consider scenarios in which these methods are applied and factors affecting end-of-outbreak estimates.

Establishing freedom from disease

In the wake of an infectious disease outbreak, an important question can be to establish when a population is entirely free from infection. This is not only important for diseases of humans but is also relevant in the context of animal and plant diseases, as trade regulations commonly require certification of disease elimination to protect global biosecurity.[9] In fact, the concept of freedom from disease emerged in the field of veterinary epidemiology due to the need to establish the absence of infection when trading livestock and meat.

To guarantee that a disease has been eliminated from a population, it is necessary to assess every individual in that population for infection, and for that assessment to be perfectly accurate. Unfortunately, exhaustive assessment is usually rendered unfeasible by financial and logistical constraints. Instead, screening for disease is commonly performed on a random sample of individuals from the population. The results from that sample are then used to infer the prevalence of disease (the proportion of individuals infected) in the whole population.

Identifying an infected individual in a selected sample confirms that total disease elimination has not been achieved. However, the inverse is not true; finding no infected individuals in the sample does not guarantee that the disease is absent from the whole population. It is possible that the disease is present,

[8] For example, Djaafara et al., 2021; Nishiura, Miyamatsu, & Mizumoto, 2016; Parag, Donnelly, Jha, & Thompson, 2020; R. N. Thompson, Morgan, & Jalava, 2019.
[9] See Caporale, Giovannini, & Zepeda, 2012; Food and Agriculture Organization (FAO), 2017; Rüegg et al., 2018; Tratalos et al., 2018.

but that by chance only uninfected individuals were included in the sample. Although the absence of evidence of the disease is not proof of the disease's absence, it does give us some information about its plausible maximum prevalence in the population, since observing no infections in the sample is unlikely to be consistent with a very high prevalence.

For example, if a sample of N individuals are randomly chosen from the population and all are assessed to be disease-free, there is a simple rule—the "rule of three"—that provides an estimate of how many individuals in the entire population could be infected.[10] The rule of three states that, in that scenario, we can say with 95% confidence that the maximum proportion of the whole population that could be infected is $3/N$. This rule can be derived straightforwardly. Specifically, suppose that the true (unknown) prevalence of disease in the population is p. Then, the probability that none of a sample of N randomly selected individuals are infected is $(1-p)^N$. If this probability is very small, it is unlikely that we would observe no infections in the sample. We choose a value α as a threshold for credibility, so that if $(1-p)^N < \alpha$, the chance of observing no cases in our sample is "too unlikely" to be believable.

Therefore, if indeed we observe no infections in the sample, a plausible value of p is one that gives rise to this observation with probability no less than α. That is, p must satisfy the equation $(1-p)^N \geq \alpha$, which can be rearranged to obtain the upper bound $p \leq 1 - \alpha^{1/N}$. By convention, the credibility threshold α is usually taken to be $\alpha = 0.05$ (the standard threshold for statistical significance in many scientific disciplines). For this value of α, the upper bound for p may be approximated by $p \leq 3/N$. Any prevalence higher than this would make the likelihood of observing no infected cases in our sample less than 5%. Thus, given that our sample returns a positive infection rate of $0/N$, we can say with 95% confidence that the maximum "true" prevalence p that is consistent with this observation is $3/N$.

The "rule of three" is an example of what is known as a confidence interval:[11] a range of plausible values of a parameter (in this case p), that contains the true value with a given probability, in this case 95%. Usefully, this rule can also be inverted to obtain the minimum sample size N for which observing no cases establishes with 95% confidence that the disease prevalence is at most p. This lower bound on N is given by $N \geq \lceil 3/p \rceil$, where $\lceil . \rceil$ represents the ceiling function that rounds a value up to the nearest

[10] Hanley & Lippman-Hand, 2008; Jovanovic & Levy, 1997; Louis, 1981.
[11] Hazra, 2017.

integer. For example, if we wish to be 95% sure that the prevalence is no greater than 1%, we must take a sample of size at least $\lceil 3/0.01 \rceil$ (that is, $N \geq 300$) and fail to find the disease.

In practice, observing no positive cases amongst sampled individuals is not necessarily a requirement for declaring freedom from disease. For example, the criterion for a member state of the European Union to be declared free from bovine tuberculosis is that the prevalence must remain below 0.1% for 6 consecutive years, which allows for a small number of cases to be present in each annual sample. Reasoning analogous to the above derivation may be applied to estimate confidence intervals for the prevalence p in scenarios in which $n > 0$ cases have been observed in a sample of size N. This allows policymakers to determine whether or not freedom from disease has been achieved in the sense of "elimination as a public health problem" (as defined by a specific criterion about the estimated prevalence).

Although the simplicity of the rule of three (and confidence intervals constructed as described above) is appealing, it relies on several assumptions. For example, it assumes that the screening method applied to detect disease never gives any false positive results (healthy individuals identified as infected) or false negative results (infected individuals identified as healthy), which is unrealistic in practice. It also assumes that the population is large enough that the probability of selecting an infected individual at any point in the sample remains approximately constant, regardless of the disease status of those already chosen. However, the underlying conceptual framework may readily be extended to overcome these assumptions, allowing for imperfect testing methods and populations of any size.[12]

Similar methods may also be used to evaluate more complex sampling strategies. For example, when attempting to establish whether livestock are disease-free at a regional or national level, it is often impractical to implement uniform random sampling from the entire population due to the number of individuals involved. Additionally, doing so would not account for the fact that disease tends to cluster within herds rather than being evenly distributed throughout the population. Instead, two-stage sampling frameworks are often implemented in practice, in which the first stage is to select herds for testing and the second is to select individuals from within those herds. In that context, probabilistic

[12] Cameron & Baldock, 1998a; Cannon, 2001; Johnson, Su, Gardner, & Christensen, 2004; Nishiura et al., 2016.

techniques may be applied to establish the optimal division of sampling resources between and within herds in order to maximise our confidence that the overall prevalence is below an acceptable level.[13]

Since samples taken at a single point in time do not incorporate information gained from previous sampling rounds, analogous techniques may be applied to construct confidence intervals for the underlying prevalence based on sampling performed at multiple time points. These models may be designed to account for changes in the underlying prevalence over time, consider the incubation period of the disease (the time during which individuals are infected but presymptomatic), or consider the risk that the pathogen has been reintroduced into the pathogen from elsewhere.[14]

As noted above, although establishing the elimination of a disease from a population with 100% certainty is almost always unfeasible, probabilistic techniques may be used to guide statistically sound surveys that allow us to determine with a given level of confidence that disease prevalence is below a given threshold. In addition to diseases of livestock such as bovine tuberculosis, foot-and-mouth disease and Bluetongue, these methods have been applied in the context of human diseases including Zika, parasitic infections, and COVID-19.[15]

For populations in which samples of randomly chosen individuals are not tested for infection, an alternative approach is required in which the probability that a population is disease-free is assessed based on the time elapsed since the last observed case, as described in the following two sections.

Using statistical models to determine when an outbreak is over

Guidance developed by the World Health Organization (WHO) for declaring the official end of an outbreak indicates that, after waiting twice the maximal incubation period since the most recently diagnosed case, an outbreak can be considered over.[16] However, the optimal timing for such a declaration cannot be determined using a fixed length of time, since it varies between outbreaks due to

[13] Cameron & Baldock, 1998b; Cannon, 2001; Rüegg et al., 2018.
[14] Bourhis, Gottwald, Lopez-Ruiz, Patarapuwadol, & van den Bosch, 2019; Bourhis, Gottwald, & van den Bosch, 2019; More et al., 2009.
[15] Bovine tuberculosis: More et al., 2009. Foot-and-mouth disease: Caporale et al., 2012. Bluetongue: Rüegg et al., 2018. Zika: C. N. Thompson et al., 2018. Parasitic infections: Michael et al., 2018. COVID-19: Foddai, Lubroth, & Ellis-Iversen, 2020; Larsen et al., 2021.
[16] Hersey et al., 2015.

a wide range of factors. Infectious disease transmission is dynamic, and given the multitude of possible outbreak trajectories, efficient and accurate end-of-outbreak declarations require a method that can dynamically determine the end of an outbreak using information about the outbreak characteristics and causal pathogen. One such approach is a statistical method that offers an interpretable estimate of the end-of-outbreak probability based on the probability of observing additional cases after the current time. This approach uses the offspring distribution (describing the numbers of secondary cases infected by each infected individual) and the serial interval (describing the time from illness onset in an infected individual to illness onset in a case they infect), and was applied during the outbreak of Middle East respiratory syndrome (MERS) in South Korea in 2015.[17] In that study, the declaration of the end of a MERS outbreak was recommended when the estimated probability of observing additional cases in future falls below a pre-specified threshold value.

More specifically, using the dataset of onset dates for diagnosed cases (t_i) for cases $i = \{0, 1, \dots, M\}$, the probability of observing at least one additional case in the future (as estimated at time t) can be approximated by the expression:

$$P(X(t) > 0) = 1 - \prod_{i=0}^{M} \sum_{y=0}^{\infty} p_y [F(t - t_i)]^y.$$

Here, p_y is the offspring distribution, which describes the probability that y cases are infected by a single primary case, while F represents the cumulative distribution of the serial interval.

This methodology was later extended to consider end-of-outbreak probabilities for other diseases, such as EVD and COVID-19, as well as to account for the delay from illness onset of a case to when they are reported in surveillance data. Accounting for reporting delays involves a straightforward extension of the equation above by combining the serial interval and reporting delay distributions.[18]

Furthermore, the method was adapted to account for multiple modes of transmission by Lee et al. (2019), who considered the potential for both sexual and non-sexual transmission of Ebola virus in the context of the end of the 2013–2016 EVD outbreak in West Africa. Ebola virus can be detected in semen long after illness onset,[19] and sexual transmission increases the risk of re-emergence of the virus, as

[17] Nishiura, Miyamatsu, & Mizumoto, 2016.
[18] Linton, Akhmetzhanov, & Nishiura, 2021b.
[19] Deen et al., 2017.

evidenced by the appearance of new cases in West Africa after the virus was believed to have been eliminated.[20] To account for both modes of transmission, Lee et al. introduced a probability distribution characterising the length of the serial interval using distributions of the serial intervals for each mode of transmission, with the distribution for sexual transmission based on the survival probability of Ebola virus in semen as a function of the time since illness onset.[21] The authors found that the optimal time at which to declare the outbreak over varies based on the relative frequency of sexual and non-sexual transmission, as well as the level of underreporting of cases. If there is a substantial amount of sexual transmission, a long period is required without observing any new cases for the outbreak to be declared over with confidence.

Transmission of many pathogens can be highly heterogeneous between infectors, with some infectors transmitting the pathogen to many individuals and other infectors transmitting the pathogen to few individuals.[22] The potential for superspreading has been characterised by the "80/20 rule", whereby 20% of infected individuals are said to generate around 80% of infections.[23] Statistical models for assessing the end-of-outbreak probability typically account for this superspreading potential. A higher propensity for superspreading acts to increase the time until the end of the outbreak can be declared with confidence.[24]

A number of recent publications have built on the statistical approach described here.[25] Specifically, methods have been developed based on renewal equations in which the numbers of cases each day follow a Poisson or negative binomial distribution. Using renewal equations, the number of cases arising each day in future can be simulated (or calculated analytically), based on the observed disease incidence time series up until the current time. For those models, the probability that an outbreak is over based on incidence data up to time *t* is simply the proportion of forward simulations in which no cases occur after time *t*. Although these approaches are based on the one by Nishiura, Miyamatsu & Mizumoto (2016),

[20] Lee & Nishiura, 2017.
[21] For more information about the persistence of Ebola virus in semen, see Eggo et al., 2015.
[22] Lloyd-Smith, Schreiber, Kopp, & Getz, 2005. For COVID-19 examples, see Adam et al., 2020; Tariq et al., 2020; Zhao, Zhang, & Li, 2020.
[23] Woolhouse et al., 1997.
[24] Linton, Akhmetzhanov, & Nishiura, 2021b.
[25] Djaafara et al., 2021; Parag, 2021; Parag, Cowling, & Donnelly, 2021; Parag, Donnelly, Jha, & Thompson, 2020

there are subtle differences in the underlying assumptions, and a quantitative comparison of end-of-outbreak probabilities obtained using these different methods remains a target for further research.

Using compartmental models to determine when an outbreak is over

The epidemiological models described in the previous section track the number of new cases arising each day. However, the dynamics of infectious disease outbreaks can be complex, with individuals transitioning through a range of infection or symptom states. For that reason, a commonly used mathematical modelling framework is compartmental modelling, in which individuals are categorised over the course of an outbreak according to their infection or symptom status.[26]

Compartmental epidemiological models have a long history. One of the most basic compartmental models is the Susceptible-Infected-Removed (SIR) model (Figure 2a), which is a special case of the epidemiological model considered by Kermack and McKendrick (1927). In the SIR model, individuals are classified as (*S*)usceptible to the outbreak pathogen, (*I*)nfected and generating new infections, or (*R*)emoved and no longer generating new infections. As the outbreak progresses, individuals who are susceptible may become infected (and transition from the *S* compartment to the *I* compartment) and then subsequently recover or die (and transition from the *I* compartment to the *R* compartment). By choosing the parameters of the model (in the SIR model, the infection rate and removal rate parameters) appropriately, output from compartmental models can be tuned to match real-world data from an ongoing outbreak.[27]

Compartmental models can be categorised into two complementary groups - either *deterministic* models or *stochastic* models.[28] Deterministic models, often represented as systems of ordinary differential equations, generate the same results every time for a specific set of inputs (e.g., infection and removal rate parameters, initial numbers of individuals in each compartment of the model). Stochastic models, on the other hand, reflect the intrinsic and extrinsic randomness inherent in real-world epidemiological systems, with repeated simulations of the model generating different outbreaks, even when the inputs are identical. A simulation of the stochastic SIR model can be thought of as a

[26] Brauer, 2008; R. N. Thompson, 2020.
[27] Chowell, 2017.
[28] Keeling & Rohani, 2008.

series of coin tosses, with the result of each coin toss determining whether the next event is an infection event (with an individual in the *S* compartment transitioning to the *I* compartment) or a removal event (with an individual in the *I* compartment transitioning to the *R* compartment). When considering outbreak extinctions, randomness in when precisely the pathogen goes extinct is important in determining the confidence in an end-of-outbreak declaration. For that reason, our focus here is entirely on stochastic, rather than deterministic, epidemiological models.

A key benefit of compartmental models is that they can be extended straightforwardly to include different features that affect transmission. For example, a common extension to the SIR model is to include a time delay between each individual being first infected and becoming infectious. This time delay is termed the latent period, and varies between pathogens. For example, the latent period for influenza is typically in the range 1–3 days, whereas the latent period for measles is around 8–13 days.[29] A latent period can be included in the SIR model by inserting a new compartment—the (*E*)xposed compartment—between the *S* and *I* compartments.[30] Individuals in the *E* compartment are infected but not yet infectious. Other possible features that can be included in compartmental models, and may be appropriate when modelling outbreaks of certain pathogens, are age structure, spatial structure, within-host dynamics, and transmission via insect vectors.[31]

In the context of determining whether or not an outbreak has finished, a crucial question is whether the outbreak is over or whether it may still be ongoing in individuals who are not reporting disease (see *Using statistical models to determine when an outbreak is over*, above). For outbreaks in populations of humans, a failure to report disease might arise due to some infected individuals being asymptomatic or showing only limited symptoms, or because individuals recover at home without reporting their infection to local health authorities.[32] In either scenario, infected individuals may not appear in routinely collected surveillance data. To establish whether or not an outbreak has finished, an important extension to the SIR model described above is therefore to include the possibility that infected

[29] Anderson & May, 1991.
[30] Anderson & May, 1991; Bolker & Grenfell, 1995; Chowell, Nishiura, & Bettencourt, 2007; R. N. Thompson, Gilligan, & Cunniffe, 2016.
[31] Age structure: Davies, Kucharski, et al., 2020; Prem et al., 2020. Spatial structure: Bolker & Grenfell, 1995; R. N. Thompson, Thompson, Pelerman, Gupta, & Obolski, 2018. Within-host dynamics: Hart, Maini, Yates, & Thompson, 2020; Mideo, Alizon, & Day, 2008. Transmission via insect vectors: Allen et al., 2019; Kucharski et al., 2016; R. N. Thompson, Gilligan, & Cunniffe, 2020.
[32] Angulo, Finelli, & Swerdlow, 2021; Cori et al., 2017; Gignoux et al., 2015

individuals do not report disease. This can be done by including separate compartments in the model for infected individuals that do and do not report disease (Figure 2b).

Simulations of stochastic compartmental models of the type shown in Figure 2b can be used to determine the confidence that an outbreak is over, based on the time period since the last case was reported. Specifically, a large number of model simulations can be run, with each simulation generated until the number of infected individuals reporting disease reaches zero. Each simulation can then be continued until a period of *X* days has passed since an infection-reporting individual was present in the population. The proportion of those simulations in which no unreported infected individuals remain after a time period of *X* days without reported infections is then a proxy for the probability that the outbreak is over after that period.

The simulation approach described above was applied to study the confidence in end-of-outbreak declarations for EVD.[33] The WHO considers EVD outbreaks to be over at the national level once a time period of 42 days (twice the maximal incubation period) has elapsed since the latest reported case. However, the situation is complicated by several features of EVD outbreaks (see *Using statistical models to determine when an outbreak is over* and Figure 1). First, the virus may be reimported from elsewhere, with local transmission restarting even after the virus has been eliminated locally.[34] Second, some EVD survivors may be infectious long after they were first infected, again potentially leading to local resurgence after the outbreak appears to have finished.[35] Third, as described above, reporting is imperfect, and so chains of transmission may persist undetected during the 42 days following the most recent reported case. Thompson et al. (2019) considered this third feature and showed that the confidence in an end-of-outbreak declaration is extremely sensitive to the level of underreporting. This led to a target of 79% case detection for EVD outbreaks to be declared over with 95% confidence after a time period of 42 days without reported cases.[36] The finding that the confidence in an end-of-outbreak declaration is sensitive to the level of reporting was echoed in a later analysis by Djaafara et al. (2021) using renewal equations (see *Using statistical models to determine when an outbreak is over*), in which those authors advocated replacing the 42-day period with a longer period of 63 days.

[33] R. N. Thompson, Morgan, et al., 2019
[34] Weah et al., 2017.
[35] Diallo et al., 2016; Keita et al., 2016; MacDermott & Bausch, 2016.
[36] R. N. Thompson, Morgan, et al., 2019.

A similar approach to the one described above has also been used in the context of polio eradication. Specifically, Eichner and Dietz (1996) considered a synthetic population of 200,000 individuals on the pathway to polio eradication, and demonstrated that the duration of the case-free period must be at least three years before a policymaker can be 95% sure that local extinction has occurred. At the global level, polio eradication remains a key challenge, and epidemiological models can be used to assess different public health policies.[37]

To summarise, compartmental epidemiological models involve separating individuals according to their infection or symptom status. Extensions to basic compartmental models (to include e.g. a latent period, age structure, or any number of other epidemiological realisms) involve adding more compartments to the models. The models can be matched to real-world outbreak data by adjusting the values of the parameters governing transmission. Simulations of compartmental models including underreporting can then be used to assess the confidence that an outbreak is over, in a similar fashion to simulations of renewal equations, based on the time period since the last case was observed.

When outbreaks are ongoing: The time to extinction

Throughout this review article so far, we have focused on the question of determining whether or not a population is disease-free at the current time. However, interventions are often planned while an outbreak is ongoing, and so estimating when an outbreak will end during the outbreak is another important area of research.

The time-varying reproduction number, R_t, is often used to indicate whether or not an outbreak is on the path to extinction.[38] It describes the expected number of secondary cases generated by an infected individual in the population at time t. If R_t is (and remains) below one, the disease is unable to sustain itself without repeated introductions and the outbreak will eventually end. However, even once the threshold of $R_t = 1$ is crossed (so that $R_t < 1$), although extinction could be inevitable, it is not immediate and control measures may need to be maintained for extinction to occur. Providing

[37] See K. M. Thompson & Kalkowska, 2020, 2021.

[38] For more on R_t, see Cori, Ferguson, Fraser, & Cauchemez, 2013; Nishiura & Chowell, 2009; R. N. Thompson, Stockwin, et al., 2019.

stakeholders with the expected remaining duration of an outbreak can help estimate future costs required for controlling the disease. Towards the end of an outbreak, it is therefore important to assess how long the delay will be before extinction occurs.

Stochastic models (see *Using compartmental models to determine when an outbreak is over*) can be used to estimate the expected time until extinction given an initial observation of the number of infected individuals in the population. This period depends on the disease dynamics, control measures in place, and the population size. Estimation of the time to extinction was first considered for ecological systems with stochastic models describing deteriorating population dynamics.[39] Studies with synthetic infectious disease data have demonstrated the capability to approximate the expected time to extinction for both simple epidemiological models (e.g., the SIR model) and more complex models with spatial structure, vaccination dynamics, and host-vector transmission.[40]

The expected time to extinction is typically stated as a function, $\tau(i)$, of the current observed number of infected individuals, i, and can be found by solving a set of simultaneous equations. The coefficients of the equations are constructed by considering the possible transitions between compartments. For example, the schematic and equation in Figure 3 describe a stochastic Susceptible-Infected-Susceptible (SIS) model. Figure 3a illustrates how, as the next event, the number of infectious individuals i can either increase to $i + 1$ (a susceptible individual is infected) or decrease to $i - 1$ (an infected individual recovers). For the example in Figure 3, the total number of individuals in the population (N) is assumed to be constant, and new infections occur at rate $\beta(N - i)i/N$, where the parameter β governs the rate of transmission between infectious and susceptible individuals. Each infected individual recovers from the infection at a rate γ, where $1/\gamma$ is the average amount of time that an individual spends infectious before recovering. The infection process will continue until no one in the population is infected and disease extinction is reached (leftmost box in Figure 3a).

The formula shown in Figure 3b indicates that the expected time to extinction from i infected individuals depends on the time until the next event (first term), the probability that the next event is a recovery

[39] See Giles Leigh, 1981; MacArthur & Wilson, 1967.

[40] SIR model: Barbour, 1975. Spatial structure: Britton & Neal, 2010; Lindholm & Britton, 2007; Swinton, 1998. Vaccination dynamics: Andersson & Britton, 2000. Host-vector transmission: Aliee, Rock, & Keeling, 2020; Britton & Traoré, 2017.

(second term), and the probability that the next event is an infection (third term). These probabilities relate the expected time to extinction, $\tau(i)$, to the subsequent extinction time, which could be $\tau(i-1)$ or $\tau(i+1)$ depending on whether the next event is a recovery or an infection. This equation can be simplified by assuming that, when near extinction, very few individuals in the population will be infected and so the number of susceptible individuals is approximately equal to the population size, N. Generally, following some rearrangement, the coefficients of this equation can be written as a matrix Q_0, where each row of the matrix, j, contains the coefficients in the equation for $\tau(j)$. Finding the extinction times $\tau = \tau(1), \dots, \tau(N)$ given the current number of infected individuals $i = 1, \dots, N$ involves solving $Q_0\,\tau = -1$.[41] The matrix Q_0 is commonly referred to as the transition matrix conditioned on non-extinction, and can be constructed for models with more complex disease dynamics, such as those including latent periods or waning immunity.

As well as the expected time until extinction occurs, understanding the distribution of possible extinction times is important to account for uncertainty in model predictions, and can be used for more in-depth analyses, such as finding the date by which there is a 90% chance of extinction.[42] Stochastic models may appear to fluctuate around an endemic steady state for a long time before extinction is reached. For many epidemiological models, the time to extinction starting from that steady state value is exponentially distributed.[43] However, approximating the distribution of possible extinction times can be challenging in some scenarios, such as when the population size is large.[44] The distribution of extinction times starting from i infected individuals can sometimes be estimated using higher moments of $\tau(i)$.[45]

An example demonstrating how the remaining duration of an outbreak after it has been brought under control ($R_t < 1$) can be estimated using the stochastic SIS model is shown in Figure 4. We ran 10,000 stochastic simulations of the SIS model and recorded the time to extinction in each simulation (the first time zero infections are recorded), presented in Figure 4b. The mean extinction time from the equation in Figure 3b is plotted onto the histogram, and we observe that it aligns with the simulation results.

[41] See Keeling & Ross, 2008.
[42] Aliee et al., 2020.
[43] SIS dynamics: Mangel & Tier, 1993. SIR dynamics: Andersson & Britton, 2000; Nåsell, 1999. Host-vector dynamics: Britton & Traoré, 2017.
[44] Doering, Sargsyan, & Sander, 2005; Nåsell, 1999.
[45] Aliee et al., 2020.

In summary, if $R_t < 1$ but the outbreak has not yet faded out, the mean time to extinction can be estimated, providing an approximation of the remaining duration of the outbreak. The distribution of possible times to extinction can provide more nuanced information about the possible duration of the outbreak remaining. The methods described here are suitable for use with prevalence data (describing the total number of individuals currently infected). However, these data are often unavailable, so future work should consider the development of methods to estimate the time to extinction from incidence data (describing the number of new cases each day). The application of these future methods to predict outbreak extinction dates for real-world pathogens has the potential to inform outbreak responses.

Factors that affect end-of-outbreak estimates

In the previous sections, mathematical approaches for inferring whether an outbreak is over and projecting the timing of the outbreak end were reviewed. These quantitative methods provide insights into the uncertainties surrounding outbreak decline and extinction (and, therefore, the confidence levels with which outbreaks can be declared over). However, there are many factors that can affect the accuracy of these estimates. These include data reliability, modes of pathogen transmission, parameter selection and inference, and the precise definition of the host population within which the outbreak is occurring.

Data reliability: As mentioned in *Using compartmental models to determine when an outbreak is over*, outbreaks can be declared over most quickly and with highest confidence when disease surveillance systems are highly sensitive and able to detect infectious cases accurately and promptly.[46] However, the sensitivity and timeliness of surveillance can vary greatly from location to location, as well as by disease, so it is necessary to account for underascertainment of cases and reporting delays. Failing to do so can lead to erroneous end-of-outbreak declarations.[47] Factors that may affect the sensitivity of a human, animal or plant disease surveillance system include: the level of symptoms associated with infections (e.g., asymptomatic or mild infections may be less likely to be detected),[48] failures to seek healthcare,

[46] R. N. Thompson, Morgan, et al., 2019.
[47] Akhmetzhanov, Jung, Cheng, & Thompson, 2021, Linton et al., 2021b; Parag et al., 2020.
[48] Poliovirus can persist for years within a population as a "silent" infection, since most of those infected do not show symptoms and are never diagnosed. Eichner & Dietz (1996) estimated how many years of zero case incidence would need to pass before it would be possible to be 95% certain that local extinction of wild poliovirus infection had occurred.

veterinary or agricultural services, inaccurate diagnoses, false-negative tests, and failures to report cases to public health or other authorities.[49]

Modes of transmission: To obtain an accurate picture of when an outbreak will end for a given disease, it is necessary to understand and account for epidemiological differences associated with different modes of transmission. These differences can affect the estimates generated by the methods described here. During the 2013–2016 EVD outbreak in West Africa, the months-long persistence of the Ebola virus in the semen of some men who recovered from EVD was not originally considered in analyses of when to declare that EVD outbreak over (see *Using statistical models to determine when an outbreak is over*). This led Lee and Nishiura (2019) to explore the use of an epidemiological model that accounts for sexual contacts and survival of the Ebola virus in semen to generate more realistic end-of-outbreak assessments.

Parameter selection and accurate inference: The values of the parameters of models used to inform end-of-outbreak declarations must be considered carefully. For example, the models presented in *Using statistical models to determine when an outbreak is over* use the serial interval or generation time (which characterises the times between successive infections) to inform estimates. However, the serial interval and generation time can change over the course of an outbreak.[50] Failing to account for this can result in incorrect assessments of the end-of-outbreak probability. Likewise, other parameters governing pathogen transmission, which may be included in the compartmental models presented in *Using compartmental models to determine when an outbreak is over*, can vary by host age, pathogen variant and host immunity (which is affected by vaccination status and past infections), amongst other factors.[51] The size of the susceptible population must also be quantified accurately, since sustained pathogen transmission requires the susceptible population size to be sufficiently large.[52] When deciding which parameters to include in epidemiological models for analysing end-of-outbreak dynamics, it is necessary to balance epidemiological realism with tractability and interpretability. The model parameters to include, and the methods used to estimate epidemiological parameter values, require careful consideration.

[49] Gibbons et al., 2014.
[50] Noted for SARS by Lipsitch et al., 2003. Later investigated in greater detail by Kenah, Lipsitch, & Robins, 2008, as well as Nishiura 2020. Analyses considering COVID-19 include Ali et al., 2020. and Linton et al., 2021a.
[51] Davies, Klepac, et al., 2020; Teunis, Le Guyader, Liu, Ollivier, & Moe, 2020
[52] Keeling & Grenfell, 1997

Definition of outbreak populations: Most mathematical approaches used for studying epidemiological dynamics at the ends of outbreaks have considered outbreaks as occurring with fixed geographical administrative boundaries. However, pathogens do not observe these boundaries, and not all outbreaks have a clear geographical scope.[53] Foodborne disease outbreaks, for example, can cross international borders via food distribution chains, and plant and animal disease outbreaks often follow domestic or international trade networks.[54] For outbreaks of directly transmitted pathogens, such as the ongoing COVID-19 pandemic, outbreaks may likewise cross geographical boundaries, with transmission facilitated by domestic and international travel. Outbreaks can be considered at a number of different spatial scales: for example, a national outbreak or a cluster of cases in a specific setting such as a hospital, prison or care home.[55] The optimal scale at which to define the host population when assessing end-of-outbreak probabilities depends on factors including the patterns of exposure: for example, for point-source outbreaks, when exposure is to a single source of infection over a short period of time,[56] it may only be necessary to include attendees at the exposure event in the population under consideration.

Further considerations: While a suite of approaches exist for analysing the ends of outbreaks, it should be emphasised that the performance of these methods has yet to be assessed rigorously in a range of outbreak response scenarios. A comparison of different approaches, as well as scientific assessments of their relative accuracy and reliability, should be explored systematically through theoretical studies and practical applications during and after outbreaks. While we have focused mainly on approaches relating to local extinction or elimination, quantitative methods for use in scenarios in which the outbreak end represents a return to endemicity require particular attention. The utility of different approaches from a public health and economic viewpoint must be assessed, and the criteria to use for declaring outbreaks over should be considered carefully. A key component of this decision is the potential consequence of an incorrect end-of-outbreak declaration, which must be assessed with a multidisciplinary perspective.

[53] R. N. Thompson, Cobb, Gilligan, & Cunniffe, 2016.

[54] For more on foodborne disease, see Coulombier & Takkinen, 2013. For more on plant and animal outbreaks, see Wilkinson et al., 2011.

[55] For examples of estimating end-of-outbreak probabilities using case clusters, see Linton et al., 2021. For examples of cluster settings, see Furuse et al., 2020.

[56] Brookmeyer & You, 2006.

Conclusions

Accurate determination of when outbreaks end allows for enhanced surveillance activities and public health interventions to be relaxed safely.[57] In this article, we have introduced various epidemiological modelling approaches that can be used to infer whether or not (and when) an outbreak has ended. Improving the accuracy of end-of-outbreak determinations based on modelling studies is best accomplished by: i) Strengthening surveillance systems, so that cases are found and reported accurately, and; ii) Supporting research that leads to an improved characterisation of pathogen transmission in the later stages of outbreaks, including identification and estimation of relevant transmission parameters.

However, end-of-outbreak declarations must not solely be the jurisdiction of public health advisors or policymakers and epidemiological modellers. It is important that healthcare capacity and socioeconomic considerations are also included. If the risk to human health is low (i.e., an incorrect end-of-outbreak declaration is only likely to lead to a small number of severe future cases) but the societal costs of outbreak interventions are high, it may be best to declare outbreaks over when the confidence in an end-of-outbreak declaration is relatively low. On the other hand, if the risk to human health of outbreak resurgence is particularly high (e.g., healthcare systems may be overwhelmed), it may be preferred to set the threshold confidence for declaring an outbreak over to be much higher. Irrespective of a policy-maker's desired level of risk aversion, quantitative approaches for studying the ends of infectious disease outbreaks will remain important for public health decision making whenever pathogens near elimination. It is essential that these methods continue to be developed.

[57] The definition of elimination introduced by Dowdle (1998) embraced the need for continued interventions to prevent re-emergence and re-establishment of transmission. However, as pointed out by Heymann (2006), all too often a complete cessation of intervention activities follows elimination. Although public health interventions can and should be relaxed once an outbreak ends, some surveillance and continuation of control interventions are necessary to maintain elimination or transmission at endemic levels; these should continue until and unless eradication is achieved.

Figures

Box 1. Important definitions for end-of-outbreak analyses[58]

Emerging disease	A disease with incidence that has recently (e.g., in the past two decades) increased substantially or threatens to increase in the near future.
Epidemic	A sudden increase in the occurrence of cases of a disease in excess of normal expectancy for the location or season. This is also the definition of an *outbreak*, although the term outbreak is used more often when the geographical range is limited.
Endemic	When a disease is consistently present in a region or population.
Non-endemic	Elimination (as a public health problem): Achievement of measurable global targets in reduction of incidence or prevalence set by WHO. When reached, continued actions are required to maintain the targets and/or to advance the interruption of transmission.
	Elimination (of transmission): Reduction of incidence of a given disease to zero in a defined location and for a defined minimum period of time as a result of deliberate efforts with minimal risk of reintroduction.
	Eradication: Permanent worldwide reduction of incidence of a given disease to zero as a result of deliberate efforts.
Extinction	Short-term disappearance of a disease in a defined location or population. This term is frequently used in statistical end-of-outbreak analyses.

[58] For historical definitions of eradication and elimination, see Soper, 1962 and Dowdle, 1998. For a more recent commentary on the definition of eradication, see Arita, Wickett, & Nakane, 2004. In practice, elimination of a disease may be defined as elimination of the disease as a public health problem (rather than complete elimination), or in terms of prevalence, rather than incidence. For example, leprosy elimination was defined in 1991 by the World Health Assembly as the reduction of prevalence to a level of < 1 case per 10,000 population. Likewise, tuberculosis elimination as a public health problem has been defined as reduction of prevalence to < 1 case per 1 million population. See also World Health Organization, 2015.

Figure 1. Schematic of the epidemic cycle a) and the timeline of key events in the cycle juxtaposed with an epidemic curve, which depicts the number of new cases per unit of time (incidence; where time is typically measured in days or weeks) for an infectious disease outbreak b). Transitioning to/from a given epidemic stage could be due to various factors such as undetected community transmission, pathogen mutation, or application of different interventions (arrows in caption a).

Figure 2. Stochastic compartmental epidemiological models can be used to estimate the end-of-outbreak probability. a) Schematic showing a simple compartmental model, the Susceptible-Infected-Removed (SIR) model. As an outbreak continues, healthy individuals become infected (transition from *S* to *I*) and then recover or die (transition from *I* to *R*). b) Schematic showing an extended version of the SIR model, in which infected hosts can either report disease (*I*) or fail to report disease (*U*). Unreported infected individuals include those who are asymptomatic or show limited symptoms, and those who are symptomatic but do not report disease to local health authorities. c) Example output from an end-of-outbreak analysis using a large number of simulations of a stochastic compartmental model. This graph shows the probability that an outbreak is over as a function of the time since the last reported case. As the period of time without a reported case increases, it becomes more likely that the outbreak is truly over.

Figure 3. a) Schematic illustrating the possible transitions from i infected individuals, to $i+1$ infected individuals (in the event of an infection) or $i-1$ infected individuals (in the event of a recovery) for a stochastic Susceptible-Infected-Susceptible (SIS) model. The simulated outbreak continues until the population reaches zero infected individuals (far left). b) Equation describing the relationship between the expected time to extinction from a state in which there are i infected individuals, $\tau(i)$, and related times to extinction from states in which there are $i-1$ infected individuals ($\tau(i-1)$), and $i+1$ infected individuals ($\tau(i+1)$).

Figure 4. Example of simulating the stochastic Susceptible-Infected-Susceptible (SIS) model in a population of *N* = 10,000 individuals with 100 individuals initially infected. a) The trajectory of 10 simulations (grey lines) and the deterministic solution (red line). b) The distribution of times to

extinction for 10,000 stochastic simulations, with the mean time-to-extinction calculated using the simulations given by the dark grey marker, and the analytical solution, $\tau(100)$, shown by the red marker.

References

Adam, D. C., Wu, P., Wong, J. Y., Lau, E. H. Y., Tsang, T. K., Cauchemez, S., … Cowling, B. J. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine*, *26*, 1714–1719.

Akhmetzhanov, A. R., Jung, S-m., Cheng, H. Y., & Thompson, R. N. (2021). A hospital-related outbreak of SARS-CoV-2 associated with variant Epsilon (B.1.429) in Taiwan: transmission potential and outbreak containment under intensified contact tracing, January–February 2021. *International Journal of Infectious Diseases*, 110, 15–20.

Ali, S. T., Wang, L., Lau, E. H. Y., Xu, X.-K., Du, Z., Wu, Y., … Cowling, B. J. (2020). Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science*, *9004*, eabc9004.

Aliee, M., Rock, K. S., & Keeling, M. J. (2020). Estimating the distribution of time to extinction of infectious diseases in mean-field approaches. *Journal of the Royal Society Interface*, *17*, 20200540.

Allen, L. J. S., Bokil, V. A., Cunniffe, N. J., Hamelin, F. M., Hilker, F. M., & Jeger, M. J. (2019). Modelling vector transmission and epidemiology of co-infecting plant viruses. *Viruses*, *11*, 1–25.

Anderson, R. M., & May, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control* (Vol. 28).

Andersson, H., & Britton, T. (2000). Stochastic epidemics in dynamic populations: Quasi-stationarity and extinction. *Journal of Mathematical Biology*, *41*, 559–580.

Angulo, F. J., Finelli, L., & Swerdlow, D. L. (2021). Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. *JAMA Network Open*, *4*, e2033706.

Arita, I., Wickett, J., & Nakane, M. (2004). Eradication of infectious diseases: Its concept, then and now. Japanese Journal of Infectious Diseases, 57, 1–6.

Barbour, A. D. (1975). The duration of the closed stochastic epidemic. *Biometrika*, *62*, 477–482.

Bolker, B. M., & Grenfell, B. T. (1995). Space, persistence and dynamics of measles epidemics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *348*, 309–320.

Borremans, B., Faust, C., Manlove, K. R., Sokolow, S. H., & Lloyd-Smith, J. O. (2019). Cross-species pathogen spillover across ecosystem boundaries: Mechanisms and theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374, 20180344.

Bourhis, Y., Gottwald, T. R., Lopez-Ruiz, F. J., Patarapuwadol, S., & van den Bosch, F. (2019). Sampling for disease absence—deriving informed monitoring from epidemic traits. *Journal of Theoretical Biology*, *461*, 8–16.

Bourhis, Y., Gottwald, T. R., & van den Bosch, F. (2019). Translating surveillance data into incidence

estimates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 20180262.

Brauer, F. (2008). Compartmental models in epidemiology. *Lecture Notes in Mathematics*, *1945*, 19–79.

Britton, T., & Neal, P. (2010). The time to extinction for a stochastic SIS-household-epidemic model. *Journal of Mathematical Biology*, *61*, 763–779.

Britton, T., & Traoré, A. (2017). A stochastic vector-borne epidemic model: Quasi-stationarity and extinction. *Mathematical Biosciences*, *289*, 89–95.

Brookmeyer, R., & You, X. (2006). A hypothesis test for the end of a common source outbreak. *Biometrics*, 62, 61–65.

Cameron, A. R., & Baldock, F. C. (1998a). A new probability formula for surveys to substantiate freedom from disease. *Preventive Veterinary Medicine*, *34*, 1–17.

Cameron, A. R., & Baldock, F. C. (1998b). Two-stage sampling in surveys to substantiate freedom from disease. *Preventive Veterinary Medicine*, *34*, 19–30.

Cannon, R. M. (2001). Sense and sensitivity - Designing surveys based on an imperfect test. *Preventive Veterinary Medicine*, *49*, 141–163.

Caporale, V., Giovannini, A., & Zepeda, C. (2012). Surveillance strategies for foot and mouth disease to prove absence of disease and absence of viral circulation. *OIE Revue Scientifique et Technique*, *31*, 747–759.

Charters, E., & Heitman, K. (2021). How epidemics end. *Centaurus*, 63, 210–224.

Chowell, G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, *2*, 379–398.

Chowell, G., Nishiura, H., & Bettencourt, L. M. A. (2007). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface*, *4*, 154–166.

Cori, A., Donnelly, C. A., Dorigatti, I., Ferguson, N. M., Fraser, C., Garske, T., … Blake, I. M. (2017). Key data for outbreak evaluation: Building on the Ebola experience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*. https://doi.org/10.1098/rstb.2016.0371

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

Coulombier, D., & Takkinen, J. (2013). From national to international - Challenges in cross-border multi-country, multi-vehicle foodborne outbreak investigations. *Eurosurveillance*, *18*, 1–2.

Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjørnstad, O. N., Metcalf, C. J. E., & Grenfell, B. T. (2018).

Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*, 79, 75–79.

Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C. A. B., … Eggo, R. M. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, *26*, 1205–1211.

Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., … Liu, Y. (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, *5*, e375–e385.

Deen, G. F., Broutet, N., Xu, W., Knust, B., Sesay, F. R., McDonald, S. L. R., … Sahr, F. (2017). Ebola RNA persistence in semen of Ebola virus disease survivors — final report. *New England Journal of Medicine*, *377*, 1428–1437.

Diallo, B., Sissoko, D., Loman, N. J., Bah, H. A., Bah, H., Worrell, M. C., … Duraffour, S. (2016). Resurgence of Ebola virus disease in Guinea Linked to a survivor with virus persistence in seminal fluid for more than 500 days. *Clinical Infectious Diseases*, *63*, 1353–1356.

Djaafara, B. A., Imai, N., Hamblion, E., Impouma, B., Donnelly, C. A., & Cori, A. (2021). A quantitative framework for defining the end of an infectious disease outbreak: Application to Ebola virus disease. *American Journal of Epidemiology*, *190*, 642–651.

Doering, C. R., Sargsyan, K. V, & Sander, L. M. (2005). Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the Fokker-Planck approximation. *Multiscale Model Simul*, *3*, 283–299.

Dowdle, W. R. (1998). The principles of disease elimination and eradication. *Bulletin of the World Health Organization*, 76, 22–25.

Ebata, A., Nisbett, N., & Gillespie, S. (2021). Food systems after Covid-19. *IDS Bulletin*, *52*, 73–93.

Eggo, R. M., Watson, C. H., Camacho, A., Kucharski, A. J., Funk, S., & Edmunds, W. J. (2015). Duration of ebola virus RNA persistence in semen of survivors: Population-level estimates and projections. *Eurosurveillance*, *20*, 1–6.

Eichner, M., & Dietz, K. (1996). Eradication of poliomyelitis: When can one be sure that poliovirus transmission has been terminated? *American Journal of Epidemiology*, 143, 816–822.

Foddai, A., Lubroth, J., & Ellis-Iversen, J. (2020). Base protocol for real time active random surveillance of coronavirus disease (COVID-19) – Adapting veterinary methodology to public health. *One Health*, *9*, 100129.

Food and Agriculture Organization (FAO). (2017). *Requirements for the establishment of pest free areas*. Retrieved from

https://assets.ippc.int/static/media/files/publication/en/2017/05/ISPM_04_1995_En_2017-05-23_PostCPM12_InkAm.pdf

Furuse, Y., Sando, E., Tsuchiya, N., Miyahara, R., Yasuda, I., Ko, Y. K., … Oshitani, H. (2020). Clusters of coronavirus disease in communities, Japan, January–April 2020. *Emerging Infectious Diseases*, 26, 2176–2179.

Gibbons, C. L., Mangen, M.-J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., … Ee Kretzschmar, M. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, *14*, 1–17.

Gignoux, E., Idowu, R., Bawo, L., Hurum, L., Sprecher, A., Bastard, M., & Porten, K. (2015). Use of capture-recapture to estimate underreporting of Ebola Virus disease, Montserrado County, Liberia. *Emerging Infectious Diseases*, *21*, 2265–2267.

Giles Leigh, E. (1981). The average lifetime of a population in a varying environment. *Journal of Theoretical Biology*, *90*, 213–239.

Hanley, J. A., & Lippman-Hand, A. (2008). If nothing goes wrong, is everything all right? *Journal of Cardiothoracic and Vascular Anesthesia*, *22*, 653–654.

Hart, W. S., Maini, P. K., Yates, C. A., & Thompson, R. N. (2020). A theoretical framework for transitioning from patient-level to population-scale epidemiological dynamics: Influenza A as a case study. *Journal of the Royal Society Interface*, *17*, 20200230.

Hazra, A. (2017). Using the confidence interval confidently. *Journal of Thoracic Disease, 9,* 4125-4130.

Hersey, S., Martel, L. D., Jambai, A., Keita, S., Meyer, E., Seeman, S., … Arnold, K. E. (2015). Ebola virus disease—Sierra Leone and Guinea, August 2015. *Morbidity and Mortality Weekly Report*, *64*, 981–984.

Heymann, D. L. (2006). Control, elimination, eradication and re-emergence of infectious diseases: Getting the message right. Bulletin of the World Health Organization, 84, 82–83.

Johnson, W. O., Su, C. L., Gardner, I. A., & Christensen, R. (2004). Sample size calculations for surveys to substantiate freedom of populations from infectious agents. *Biometrics*, *60*, 165–171.

Jovanovic, B. D., & Levy, P. S. (1997). A Look at the Rule of Three. *American Statistician*, *51*, 137–139.

Keeling, M. J., & Grenfell, B. T. (1997). Disease extinction and community size: Modeling the persistence of measles. *Science*, 275, 65–67.

Keeling, M. J., & Rohani, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton, New Jersey: Princeton University Press.

Keeling, M. J., & Ross, J. V. (2008). On methods for studying stochastic disease dynamics. *Journal of the*

Royal Society Interface, *5*, 171–181.

Keita, M., Duraffour, S., Loman, N. J., Rambaut, A., Diallo, B., Magassouba, N., … Faye, O. (2016). Unusual ebola virus chain of transmission, Conakry, Guinea, 2014–2015. *Emerging Infectious Diseases*, *22*, 2149–2152.

Kenah, E., Lipsitch, M., & Robins, J. M. (2008). Generation interval contraction and epidemic data analysis. *Mathematical Biosciences*, *213*, 71–79.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *115*, 700–721.

Korenberg, E. I. (2000). Seasonal population dynamics of Ixodes ticks and tick-borne encephalitis virus. *Experimental and Applied Acarology*, 24, 665–681

Kucharski, A. J., Funk, S., Eggo, R. M., Mallet, H. P., Edmunds, W. J., & Nilles, E. J. (2016). Transmission dynamics of Zika virus in island populations: A modelling analysis of the 2013–14 French Polynesia outbreak. *PLoS Neglected Tropical Diseases*, *10*, 1–15.

Larsen, D. A., Collins, M. B., Du, Q., Hill, D., Insaf, T. Z., Kilaru, P., … Green, H. (2021). Coupling freedom from disease principles and early warning from wastewater surveillance to improve health security. *MedRxiv*, 2021.06.11.21258797.

Lee, H., & Nishiura, H. (2017). Recrudescence of Ebola virus disease outbreak in West Africa, 2014–2016. *International Journal of Infectious Diseases*, *64*, 90–92.

Lee, H., & Nishiura, H. (2019). Sexual transmission and the probability of an end of the Ebola virus disease epidemic. *Journal of Theoretical Biology*, *471*, 1–12.

Lindholm, M., & Britton, T. (2007). Endemic persistence or disease extinction: The effect of separation into sub-communities. *Theoretical Population Biology*, *72*, 253–263.

Linton, N. M., Akhmetzhanov, A. R., & Nishiura, H. (2021a). Correlation between times to SARS-CoV-2 symptom onset and secondary transmission undermines epidemic control efforts. *MedRxiv*, 21262512.

Linton, N. M., Akhmetzhanov, A. R., & Nishiura, H. (2021b). Localized end-of-outbreak determination for coronavirus disease 2019 (COVID-19): examples from clusters in Japan. *International Journal of Infectious Diseases*, *105*, 286–292.

Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., … Murray, M. (2003). Transmission dynamics and control of severe acute respiratory syndrome. *Science*, *300*, 1966–1970.

Lloyd-Smith, J. O., George, D., Pepin, K. M., Pitzer, V. E., Pulliam, J. R. C., Dobson, A. P., … Grenfell, B. T. (2009). Epidemic dynamics at the human-animal interface. *Science*, 326, 1362–1368.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*, 355–359.

Lo Iacono, G., Cunningham, A. A., Fichet-calvet, E., Garry, F., Grant, D. S., Leach, M., … Aa, C. (2016). A unified framework for the infection dynamics of zoonotic spillover and spread. *PLoS Neglected Tropical Diseases*, 10, e0004957.

Louis, T. A. (1981). Confidence intervals for a binomial parameter after observing no successes. *American Statistician*, *35*, 154–154.

MacArthur, R. H., & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, New Jersey: Princeton University Press.

MacDermott, N. E., & Bausch, D. G. (2016). Virus persistence and recrudescence after Ebola virus disease: what are the risks to healthcare workers? *Journal of Hospital Infection*, *94*, 113–115.

Mangel, M., & Tier, C. (1993). A simple direct method for finding persistence times of populations and application to conservation problems. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 1083–1086.

Michael, E., Smith, M. E., Katabarwa, M. N., Byamukama, E., Grisworld, E., Habomugisha, P., … Richards, F. O. (2018). Substantiating freedom from parasitic infection by combining transmission model predictions with disease surveys. *Nature Communications*, *9*, 4342.

Mideo, N., Alizon, S., & Day, T. (2008). Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in Ecology and Evolution*, *23*, 511–517.

More, S.J., Cameron, A.R., Greiner, M., Clifton-Hadley, R.S., Rodeia, S.C., Bakker, D., … Wierup, Martin. Defining output-based standards to achieve and maintain tuberculosis freedom in farmed deer, with reference to member states of the European Union. Prev Vet Med. 2009;90(3–4):254–67.

Nåsell, I. (1999). On the time to extinction in recurrent epidemics. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *61*, 309–330.

Nishiura, H., & Chowell, G. (2009). The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In *Mathematical and Statistical Estimation Approaches in Epidemiology* (pp. 103–121).

Nishiura, H., Miyamatsu, Y., & Mizumoto, K. (2016). Objective determination of end of MERS outbreak, South Korea, 2015. *Emerging Infectious Diseases*, *22*, 146–148.

Parag, K. V. (2021). Sub-spreading events limit the reliable elimination of heterogeneous epidemics. *Journal of the Royal Society Interface*, 18, 20210444.

Parag, K. V, Cowling, B. J., & Donnelly, C. A. (2021). Deciphering early-warning signals of the elimination

and resurgence potential of SARS-CoV-2 from limited data at multiple scales. *medRxiv*, 2020.11.23.20236968.

Parag, K. V., Donnelly, C. A., Jha, R., & Thompson, R. N. (2020). An exact method for quantifying the reliability of end-of-epidemic declarations in real time. *PLoS Computational Biology*, *16*, e1008478.

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15, 502–510.

Ponce, L., Kinoshita, R., & Nishiura, H. (2019). Exploring the human-animal interface of Ebola virus disease outbreaks. *Mathematical Biosciences and Engineering*, 16, 3130–3143.

Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., … Hellewell, J. (2020). The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, *2667*, 1–10.

Rüegg, S. R., Welby, S., Yassin, H., Van der Stede, Y., Nafzger, R., Saatkamp, H., … Stärk, K. D. C. (2018). Optimising cost-effectiveness of freedom from disease surveillance—Bluetongue Virus Serotype 8 as an example. *Preventive Veterinary Medicine*, *160*, 145–154.

Sachak-Patwa, R., Byrne, H. M., Dyson, L., & Thompson, R. N. (2021). The risk of SARS-CoV-2 outbreaks in low prevalence settings following the removal of travel restrictions. *Communications Medicine*, *1*, 1–9.

Sonaiya, E. B. (2007). Family poultry, food security and the impact of HPAI. *World's Poultry Science Journal*, *63*, 132–138.

Soper, F. L. (1962). Problems to be solved if the eradication of tuberculosis is to be realized. *American Journal of Public Health*, 52, 734–745.

Swinton, J. (1998). Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology*, *60*, 215–230.

Tariq, A., Lee, Y., Roosa, K., Blumberg, S., Yan, P., Ma, S., & Chowell, G. (2020). Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. *BMC Medicine*, *18*, 166.

Teunis, P. F. M., Le Guyader, F. S., Liu, P., Ollivier, J., & Moe, C. L. (2020). Noroviruses are highly infectious but there is strong variation in host susceptibility and virus pathogenicity. *Epidemics*, *32*, 100401.

Thiermann, A. (2004). Emerging diseases and implications for global trade. *Revue Scientifique et Technique (International Office of Epizootics)*, *23*, 701–707.

Thompson, C. N., Lee, C. T., Immerwahr, S., Resnick, S., Culp, G., & Greene, S. K. (2018). Sampling considerations for a potential Zika virus urosurvey in New York City. *Epidemiology and Infection*,

146, 1628–1634.

Thompson, K. M., & Kalkowska, D. A. (2020). Review of poliovirus modeling performed from 2000 to 2019 to support global polio eradication polio eradication. *Expert Review of Vaccines*, 19, 661–686.

Thompson, K. M., & Kalkowska, D. A. (2021). Reflections on modeling poliovirus transmission and the polio eradication endgame. *Risk Analysis*, 41, 229–247.

Thompson, R. N. (2020). Novel coronavirus outbreak in Wuhan, China, 2020: Intense surveillance is vital for preventing sustained transmission in new locations. *Journal of Clinical Medicine*, 9, 498.

Thompson, R. N. (2020). Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Medicine*, *18*, 10–13.

Thompson, R. N., & Brooks-Pollock, E. (2019). Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374, 20190038.

Thompson, R. N., Gilligan, C. A., & Cunniffe, N. J. (2016). Detecting presymptomatic infection is necessary to forecast major epidemics in the earliest stages of infectious disease outbreaks. *PLoS Computational Biology*, *12*, 1–18.

Thompson, R. N., Gilligan, C. A., & Cunniffe, N. J. (2020). Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic. *Journal of the Royal Society Interface*, *17*, 20200690.

Thompson, R. N., Morgan, O. W., & Jalava, K. (2019). Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 20180431.

Thompson, R. N., Stockwin, J. E., van Gaalen, R. D., Polonsky, J. A., Kamvar, Z. N., Demarsh, P. A., … Cori, A. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 100356.

Thompson, R. N., Thompson, C., Pelerman, O., Gupta, S., & Obolski, U. (2018). Increased frequency of travel may act to decrease the chance of a global pandemic. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 404871.

Tratalos, J. A., Barrett, D. J., Clegg, T. A., O'Neill, R. G., McGrath, G., Lane, E. A., & More, S. J. (2018). Sampling methodology to maximize the efficient use of National Abattoir Surveillance: Using archived sera to substantiate freedom from bluetongue virus infection in Ireland. *Frontiers in Veterinary Science*, *5*, 1–9.

Weah, V. D., Doedeh, J. S., Wiah, S. Q., Nyema, E., Lombeh, S., & Naiene, J. (2017). Enhancing Ebola virus

disease surveillance and prevention in counties without confirmed cases in rural Liberia: Experiences from Sinoe County during the flare-up in Monrovia, April to June, 2016. *PLoS Currents*, *9*, 1–8.

Wilkinson, K., Grant, W. P., Green, L. E., Hunter, S., Jeger, M. J., Lowe, P., … Waage, J. (2011). Infectious diseases of animals and plants: An interdisciplinary approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 1933–1942.

Woolhouse, M. E. J., Dye, C., Etard, J. F., Smith, T., Charlwood, J. D., Garnett, G. P., … Anderson, R. M. (1997). Heterogeneities in the transmission of infectious agents. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 338–342.

World Health Organization. (2015). Generic framework for control, elimination and eradication of neglected tropical diseases. World Health Organization, 1–6.

Zhao, P., Zhang, N., & Li, Y. (2020). A comparison of infection venues of COVID-19 case clusters in northeast China. *International Journal of Environmental Research and Public Health*, *17*, 1–14.