Published in final edited form as:
Epidemics. 2024 March ; 46: 100750. doi:10.1016/j.epidem.2024.100750.

Optimal environmental testing frequency for outbreak surveillance

Jason W. Olejarz[a,b,*], **Kirstin I. Oliveira Roster**[a,b], **Stephen M. Kissler**[c,a,b], **Marc Lipsitch**[a,b,d], **Yonatan H. Grad**[a,b]

[a]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

[b]Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

[c]Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309, USA

[d]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

Abstract

Public health surveillance for pathogens presents an optimization problem: we require enough sampling to identify intervention-triggering shifts in pathogen epidemiology, such as new introductions or sudden increases in prevalence, but not so much that costs due to surveillance itself outweigh those from pathogen-associated illness. To determine this optimal sampling frequency, we developed a general mathematical model for the introduction of a new pathogen that, once introduced, increases in prevalence exponentially. Given the relative cost of infection *vs.* sampling, we derived equations for the expected combined cost per unit time of disease burden and surveillance for a specified sampling frequency, and thus the sampling frequency for which the expected total cost per unit time is lowest.

Keywords

Environmental surveillance; Early pathogen detection; Wastewater sampling; Vector trapping; Mathematical modeling

[*]Corresponding author at: Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. olejarz@g.harvard.edu (J.W. Olejarz).

Declaration of competing interest
None.

CRediT authorship contribution statement
Jason W. Olejarz: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Kirstin I. Oliveira Roster:** Writing – review & editing. **Stephen M. Kissler:** Funding acquisition, Writing – review & editing. **Marc Lipsitch:** Funding acquisition, Supervision, Writing – review & editing. **Yonatan H. Grad:** Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Code availability
All code for running the simulations and reproducing the figures is available at https://github.com/jolejarz.

Appendix A. Supplementary data
Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2024.100750.

1. Introduction

A key goal of public health infectious disease surveillance systems is to detect a pathogen at an early stage of its entry into the population, enabling interventions to limit its spread and the harm it could inflict (Murray and Cohen, 2017; Budd et al., 2020; Jernigan et al., 2023). Such efforts are increasingly important given the many ways in which communities are connected, with growing populations, global travel, and urbanization, and given ecological shifts associated with climate change and other factors leading to emergence and re-emergence of vector-borne diseases, with cases of locally acquired dengue and malaria where they had been absent for many decades (Baker et al., 2021; Kretschmer et al., 2023; CDC, 2023).

One important strategy for achieving early pathogen detection is monitoring for infected individuals through robust clinical surveillance. However, clinical surveillance is inherently limited in important ways. Infections may be mildly symptomatic, asymptomatic, or have a long pre-symptomatic infectious phase, such that the pathogen population has spread extensively before the first clinical cases are diagnosed and the pathogen is identified (Oran and Topol, 2020). In contexts where access to care or resources is limited, missed cases and reporting delays can make it difficult to rapidly detect and correctly diagnose new infections (Quinn and Kumar, 2014).

For pathogens that can be detected in environmental samples and that spread by vectors, a complementary and critical strategy for early pathogen detection is monitoring through periodic sampling of the environment. Pathogen detection in wastewater has been important for the surveillance and control of poliovirus (Diamond et al., 2022; Shah et al., 2022) and has been used more recently for tracking the local epidemic dynamics and evolution of SARS-CoV-2 (Peccia et al., 2020; Levy et al., 2023), norovirus (Huang et al., 2022), influenza (Mercier et al., 2022), mpox (Chen and Bibby, 2022; Tiwari et al., 2023), and other pathogens (Boehm et al., 2023). Efforts are underway to extend these techniques for tracking antibiotic resistance genes in wastewater (Nguyen et al., 2021; Tiwari et al., 2022). For vector-borne pathogens, including West Nile virus (Petersen et al., 2013), *Borrelia* species (Eisen and Paddock, 2021), and Powassan virus (Hermance and Thangamani, 2017), surveillance includes pathogen detection in vectors collected via traps, with sampling also taking place at a given frequency.

Monitoring for infectious diseases requires substantial time, money, and infrastructure for detection, interpretation, and response (Pfaller, 2001; Vazquez-Prokopec et al., 2010; Braks et al., 2014; Kantor et al., 2022; Ngwira et al., 2022; Hagedorn et al., 2023). Although environmental and vector-based surveillance systems have been recognized and widely discussed for their potential, and despite the massive push to fund and develop these programs—particularly wastewater efforts (Gwinn et al., 2017; Kirby et al., 2021), there remains a critical gap in our understanding: How should surveillance be designed to achieve maximal effectiveness (Gu et al., 2008; Thompson and Etter, 2015; Fournet et al., 2018; Ahmed et al., 2020; Michael-Kordatou et al., 2020; Keshaviah et al., 2021)? A central consideration is how often testing should be performed (Fig. 1). Here, we addressed this question by formulating a simple, stochastic model for pathogen introduction, growth, and detection in the presence of periodic sampling and testing. We identified the key parameters of this process, and we derived a simple equation for the expected total cost per unit time (i.e., the sum of all costs related to surveillance and to effects from the disease divided by the time elapsed, when the stochastic dynamics are run for an arbitrarily long time). The expected total cost per unit time is a function of the parameters of the model, and given values for these parameters, we can minimize this quantity.

Our goal was to minimize the expected total surveillance and disease cost per unit time for the detection of the first appearance of a pathogen. Accordingly, we employed a simple model of surveillance to detect the entry of a pathogen into a population, assuming that, once the pathogen is present, its prevalence increases exponentially. The pathogen can be introduced beginning at time $t = 0$. Sampling begins at time $t = T$ and continues regularly at times $t_m = mT$, where $m \geq 1$ and T is the sampling period. Each sampling event incurs a cost c_1, and since there are $1/T$ sampling events per unit time, the sampling cost per unit time is given by c_1/T. Copies of the pathogen are introduced after time $t = 0$ according to a Poisson process, such that the waiting times between initiation events are exponentially distributed with rate λ. Once a new lineage is introduced, it also reproduces (transmits) according to a Poisson process, such that the expected prevalence of the pathogen grows exponentially with rate r. If a sampling event detects a copy of a pathogen that belongs to a particular lineage, then that lineage is “detected”. Let p be the probability that a sampling event detects a copy of a pathogen, and since each detection occurs independently, the probability that a sampling event detects a lineage of size n is given by $1 - (1 - p)^n$. Once a lineage is detected, we assume that intervention is immediate and is successful at suppressing further spread of that lineage. If a lineage has n copies of the pathogen when it is detected, then the disease cost due to that lineage is given by c_2n. Letting $\langle n\rangle$ denote the expected size of a lineage when it is detected, the expected disease cost due to a lineage is given by $c_2\langle n\rangle$. Since new lineages appear at rate λ, the expected disease cost per unit time is given by $\lambda c_2\langle n\rangle$. The expected total cost per unit time is then $c_1/T + \lambda c_2\langle n\rangle$. The model is illustrated in Fig. 2.

2. Results

We derived an accurate approximation for the expected total cost of testing and disease burden per unit time, C:

$$C = \frac{c_1}{T} + \lambda c_2\left(\frac{e^{rT} - 1}{rT}\right)\left(1 - \frac{1 - e^{-rT}}{\log(1 - p)}\right) \tag{1}$$

Details on the derivation of Eq. (1) are provided in the Supplementary Information. By comparing Eq. (1) with $c_1/T + \lambda c_2\langle n\rangle$, notice that the product of the two factors in large parentheses is (approximately) equal to the expected size of an outbreak when it is detected, $\langle n\rangle$. It is helpful to understand the behavior of $\langle n\rangle$ as a function of p, r, and T. Notice that $\langle n\rangle$ is a decreasing function of p, an increasing function of r, and an increasing function of T—i.e., a less-sensitive detector, a more rapidly growing pathogen, or a larger testing interval all result in a larger expected size of the outbreak when it is detected. (For very large values of rT, the approximation of Eq. (1) breaks down because the pathogen is typically detected on the first test following its introduction, irrespective of modest changes in p. But such extreme values of rT are unrealistic.)

The expected infection cost per unit time, $\lambda c_2\langle n\rangle$, is therefore also an increasing function of the testing period, T, and this quantity becomes arbitrarily large as $T \to \infty$. The surveillance cost per unit time, c_1/T, however, is a decreasing function of T, and this quantity becomes arbitrarily large as $T \to 0$. These behaviors are evident in Fig. 3, where we plotted C as a function of the sampling frequency, $f = 1/T$, for different values of the model parameters. It is instructive to consider the effects of very large or very small values of f on C. As we increase f, we can detect the disease more rapidly, thereby mitigating disease-related costs. But returns are diminishing: We can – at best – hope to discover the disease as a single unit, representing the pathogen introduction, before it has begun to spread and proliferate, while increasing f further can add arbitrarily large surveillance costs. At the other extreme, setting f too small allows the disease to proliferate before any intervention is applied. Therefore, as shown in each of the curves in Fig. 3, C attains a minimum for a particular value of f.

In designing and performing environmental surveillance, we do not know *a priori* the characteristics of a particular pathogen that may be introduced and result in an outbreak. Rather, for optimizing surveillance, the requirement is to have an understanding of the likely characteristics of new pathogens that might emerge. As a simple example, suppose that our surveillance platform is capable of detecting not just one but two different pathogens. Further, suppose that these two pathogens have different costs and are initiated at different rates. Let $c_2(1)$ denote the per-case cost for the first pathogen, and let $c_2(2)$ denote the per-case cost for the second pathogen. Similarly, let $\lambda(1)$ denote the rate of introductions for the first pathogen, and let $\lambda(2)$ denote the rate of introductions for the second pathogen. For this scenario, the expected infection cost per unit time is equal to $[\lambda(1)][c_2(1)][\langle n\rangle] + [\lambda(2)][c_2(2)][\langle n\rangle]$. It is also possible that the two pathogens differ in their growth rates and in their susceptibility to being detected. The first pathogen might have corresponding parameters $r(1)$ and $p(1)$, while the second pathogen might have parameters $r(2)$ and $p(2)$. As a result, the expected size of an outbreak of the first pathogen, $\langle n\rangle(1)$, might be different from the expected size of an outbreak of the second pathogen, $\langle n\rangle(2)$. The expected infection cost per unit time is then equal to $[\lambda(1)][c_2(1)][\langle n\rangle(1)] + [\lambda(2)][c_2(2)][\langle n\rangle(2)]$. If there are more than two types of pathogens that can emerge and be detected by our surveillance platform, then in the calculation of the expected infection cost per unit time, we would simply add another term for each additional pathogen.

An important point is that the possible values of the parameters c_2, r, and p that a pathogen can have are not discrete. Accordingly, let $\mathrm{d}c_2\,\mathrm{d}r\,\mathrm{d}p\lambda'(c_2, r, p)$ denote the (infinitesimal) rate at which pathogens with per-case cost c_2, growth rate r, and detection probability p emerge. In this more general treatment, $\lambda'(c_2, r, p)$ is a rate density that is a function of c_2, r, and p. To calculate the expected pathogen cost, we integrate $\mathrm{d}c_2\,\mathrm{d}r\,\mathrm{d}p\lambda'(c_2, r, p)c_2\langle n\rangle$ over all possible values of c_2, r, and p. Accounting for all possible types of pathogens that might emerge, the expected total cost per unit time, C', is equal to

$$C' = \frac{c_1}{T} + \int_0^\infty \mathrm{d}c_2 \int_0^\infty \mathrm{d}r \int_0^1 \mathrm{d}p \times \left\{ \lambda'(c_2, r, p) \left[c_2 \left(\frac{e^{rT} - 1}{rT} \right) \left(1 - \frac{1 - e^{-rT}}{\log(1 - p)} \right) \right] \right\} \tag{2}$$

Eq. (2) can be quickly calculated numerically for different values of the testing period, T. The value of $1/T$ for which the expected total cost per unit time is lowest specifies the optimal testing frequency, F^*. From Eq. (2), we have

$$F^* = \frac{1}{\arg\min_T C'} \tag{3}$$

Fig. 4 depicts how this works. In Fig. 4A, we show one possible form for the probability density function for c_2. Pathogens with little or no associated cost (i.e., those for which c_2 is close to zero) are most common, while more harmful pathogens occasionally arise. The parameter a controls the shape of the probability density function. For smaller values of a, the distribution has a longer tail, meaning that there is a higher chance that a new pathogen is harmful. In Fig. 4B, we use this form for the probability density function for c_2, we set $r = 0.1$ and $p = 0.01$, we set the total rate of emergence of new pathogens to 0.001, and we plot the expected total cost per unit time, C'. (In the specification of λ', δ denotes the Dirac delta function.) For smaller values of a, the optimal testing frequency for environmental surveillance increases accordingly.

Similarly, in Fig. 4C, we show one possible form for the probability density function for r. We again use the parameter a to control the shape of the probability density function. Smaller values of a result in a longer tail to the distribution, so that pathogens with more rapid growth rates are more likely to arise. In Fig. 4D, we use this form for the probability density function for r, we set $c_2 = 1$ and $p = 0.01$, we set the total rate of emergence of new pathogens to 0.001, and we plot C'. For smaller values of a, we must sample the environment more frequently. Notice that as the sampling frequency decreases below its optimum, the expected total cost per unit time rapidly increases. This is because there is a chance that pathogens with unusually large growth rates are introduced, and if their subsequent exponential growth is not halted soon enough, then the resulting pathogen-associated costs can become extremely large.

In Fig. 4E, we show one possibility for the probability density function for p. For smaller values of a, there is a higher chance of pathogens being introduced that have low sensitivity to being detected. Using this form for the probability density function for p in Fig. 4F, setting $c_2 = 1$ and $r = 0.1$, and setting the total rate of emergence of new pathogens to 0.001, we plot C'. (In the specification of λ', θ denotes the Heaviside step function.) Smaller values of a result in a larger optimal testing frequency. Details on Fig. 4 are provided in the Supplementary Information.

The example probability density functions in Fig. 4 were chosen here for convenience: they have nice analytical forms, and they admit simple analytical solutions when substituted into Eq. (2). For optimizing an environmental or vector surveillance system in practice, one would construct an estimated form for $\lambda'(c_2, r, p)$ based on experimental or observational data, and the optimal testing frequency would be determined numerically using Eq. (3). Optimization of environmental or vector surveillance thus requires an understanding of the cost of each sampling and testing event, c_1, and an understanding of the function for the rate of emergence of new pathogens, $\lambda'(c_2, r, p)$.

Estimation of c_1 and $\lambda'(c_2, r, p)$ in the context of any surveillance program and set of pathogens would be highly approximate, at best. For applying this theory, it is therefore desirable to have a simple, explicit approximation for the optimal sampling frequency, F^*. Using Eqs. (2) and (3), we find the following approximation:

$$F^* \approx \sqrt{\int_0^\infty \mathrm{d}c_2 \int_0^\infty \mathrm{d}r \int_0^1 \mathrm{d}p \left(\frac{c_2 r}{c_1 p}\right) \lambda'(c_2, r, p)} \tag{4}$$

Details on the derivation of Eq. (4) are provided in the Supplementary Information. This approximation for the optimal testing frequency admits a simple understanding. For a particular type of pathogen, F^* is an increasing function of c_2 and r and a decreasing function of c_1 and p, hence the factor $c_2 r/(c_1 p)$. We then multiply by $\lambda'(c_2, r, p)\,\mathrm{d}c_2\,\mathrm{d}r\,\mathrm{d}p$ and integrate over all possible values of the pathogen-specific parameters. Finally, for the resulting quantity to have dimensions of frequency, we take the square root.

3. Discussion

Eqs. (2) and (3) specify the optimal frequency at which to perform sampling and testing. Their use for optimizing testing frequency requires an estimation of the total rate of emergence of new pathogens, $\int_0^\infty \mathrm{d}c_2 \int_0^\infty \mathrm{d}r \int_0^1 \mathrm{d}p \lambda'(c_2, r, p)$, and the likely values of the per-case cost, c_2, rate of growth, r, and susceptibility to detection, p, of any emerging pathogens. The rate density, $\lambda'(c_2, r, p)$, is large if pathogens with those parameter values are likely to emerge, and small otherwise. Estimating the dependence of λ' on p entails many considerations. Molecular properties of emerging pathogens must be anticipated, and they must be interpreted in the context of whichever laboratory tests are used for detection. Spatial structure of the landscape over which pathogens can emerge further influences the dependence of λ' on p. For instance, if a pathogen emerges far from a wastewater treatment facility, then the number of infections in the vicinity of the location of sampling might be much smaller than the total size of the outbreak. A similar consideration arises in sampling a vector population, where a pathogen might originate and begin spreading in individuals that are far from the nearest trap. Inferring the total rate of emergence of new pathogens and the dependence of λ' on r may be accomplished by analysis of historical data of either clinical cases or abundance of a pathogen in a vector species, together with maximum likelihood estimation. Optimization of testing frequency further requires a formal understanding of surveillance-related and pathogen-related costs (Zinsstag et al., 2020; Bernstein et al., 2022; Weinstein et al., 2009). Mathematically, the question of how to optimize a surveillance platform is undefined unless all relevant surveillance-related and pathogen-related costs are quantified in the same units. This is challenging, since the underlying factors are inherently very different in nature. Nonetheless, such understanding is essential if environmental and vector surveillance for infectious diseases is to be meaningfully optimized.

Although accurate estimation of these effects and how they influence c_1 and $\lambda'(c_2, r, p)$ is challenging, our theory for optimizing the sampling frequency is fairly robust to uncertainties in these quantities. This is evident in the approximation for the optimal sampling frequency given by Eq. (4), where all quantities appear under the square root. In our estimation of c_1, we could be off by a factor of k (i.e., $c_1 \rightarrow kc_1$), but our determination of F^* would only be off by a factor of $1/\sqrt{k}$. Likewise, our estimation of c_2 or r could be off by a factor of k (i.e., $\lambda'(c_2, r, p) \rightarrow \lambda'(c_2/k, r, p)$ or $\lambda'(c_2, r, p) \rightarrow \lambda'(c_2, r/k, p)$), but our determination of F^* would only be off by a factor of $\sqrt{k}$ in each case.

Our model for determining the optimal testing frequency is broadly applicable. The sampling cost, c_1, is a characteristic of the surveillance platform that is constructed and deployed. c_1 encompasses many considerations. For a larger population that is more difficult to survey, for example, c_1 might take a larger value. More precisely, c_1 is related to both the sensitivity and specificity of the surveillance program. If we consider that specificity is fixed, a more sensitive surveillance program (corresponding to a larger value of p) would likely be more complex and lead to a larger value of c_1 (Fig. 5). Since the optimal sampling frequency is a decreasing function of both p and c_1, assuming that c_1 is directly associated with p results in the optimal sampling frequency having a stronger inverse relationship with p.

If we consider that sensitivity is fixed, the effect of specificity on c_1, however, is more complex. A more specific surveillance program might be more sophisticated and lead to a larger value of c_1. But greater specificity also means a lower rate of false positives, resulting in fewer unnecessary intervention costs and effecting a lower value of c_1. The net effect of specificity on c_1 for a particular surveillance program and for an estimated set of pathogen characteristics would therefore have to be carefully evaluated.

Although the long-time dynamics of an emerging pathogen can show complex behavior, the early-time dynamics are often approximately exponential, and the associated disease-related costs at early times are expected to scale roughly linearly with the size of the outbreak. Both of these features are incorporated in our model. Nonetheless, our framework can handle alternative assumptions. For example, in Fig. 6, we show the expected total cost per unit time if the cost of a single outbreak is equal to c_2n^2. For low sampling frequencies, the expected total cost per unit time is larger than for the case where pathogen costs scale linearly with outbreak size. Accordingly, the optimal sampling frequency is increased.

Once a pathogen is detected and an intervention is implemented, spread of the pathogen and its associated costs are not immediately halted, and this can be approximately accounted for by making the substitution $\lambda'(c_2, r, p) \rightarrow \lambda'(c_2/k, r, p)$, where $k > 1$. A further consideration is that intervention is unlikely to completely eliminate the pathogen. Subsequent sampling and testing would then monitor for when the pathogen becomes sufficiently abundant again that additional intervention is warranted. This may be approximately described by using a rate density for introductions that is time-dependent (i.e., $\lambda'(c_2, r, p) \rightarrow \lambda'(c_2, r, p, t)$) and increases if there was a recently suppressed outbreak. The increased value of λ' accounts for the possibility of a follow-up outbreak due to cases that the intervention failed to extinguish. As a simple example, in Fig. 7, we show simulations of a slightly generalized model in which breakthrough infections are possible. Whenever a lineage is detected, all cases are controlled, but one new infection is initiated immediately following the intervention with probability b. (Conversely, with probability $1 - b$, there is no breakthrough infection.) If there is a breakthrough infection, then that infection multiplies, and its lineage must similarly be detected and controlled. For values of $b > 0$, infections that escape control must be quickly detected and halted, so the optimal testing frequency increases.

Changes in weather and climate affect the risk of an outbreak – especially for many vector-borne pathogens (Pley et al., 2021) – and this could also be modeled through time-dependence of λ'. A key observation is that sampling is usually done on a timescale of days or weeks, whereas seasonality is typically related to the time of year. With this separation of timescales, Eqs. (2) and (3) approximately specify the optimal sampling frequency at any given time of year. We treat λ' as being periodic with period equal to one year, and we substitute this time-dependent rate density into Eq. (2).

Our approach can be applied to answer another pressing question: Where should environmental sampling be performed? A pathogen may be more likely to emerge in certain locations than others, and certain parts of the population may be more difficult to survey than others. To address these points, we can include a spatial structure in the model. The pathogen can be introduced in one location and then migrate to different locations as it proliferates. By numerically running the stochastic dynamics with spatial structure, an expected total surveillance and disease cost per unit time can be calculated. By trying different sampling locations, it is possible to find the sampling locations for which this quantity is minimal.

Environmental and vector surveillance are equally instrumental for tracking the prevalence of a pathogen (Teklehaimanot et al., 2004). An understanding of how and when to intervene is therefore essential (Lipsitch et al., 2009; Peak et al., 2017). If false positives are too frequent, then intervention costs will accumulate, leading to costly surveillance. If the designated signal that is required for intervention is too strong, then the pathogen can spread to the point where intervention has limited effectiveness in mitigating disease-related costs. The optimal testing frequency could also be adjusted as new data become available (DeFelice et al., 2017). If tracking indicates increased prevalence of a pathogen, then more frequent sampling and testing might be warranted. For example, for the model simulated in Fig. 7, instead of having a constant sampling frequency, we could increase the sampling frequency immediately following detection of a lineage to guard against the possibility of an infection that evades control. If there are no positive tests for a certain time afterward, then the sampling frequency might be safely lowered to its original value. For seasonal infections, a further possibility is that testing could be performed frequently for several years to gain an understanding of the typical seasonal behavior for new pathogens or in new ecological settings. This could inform optimization and enable more efficient tracking of the pathogen's abundance.

In practice, a surveillance program would be executed over a defined time period, where the stochasticity in the origination and growth of new infections is of paramount concern. It is possible that for many realizations of the dynamics, no infections emerge, and the total cost consists only of costs related to surveillance. Other realizations might be characterized by emergence of just one or two pathogens that inflict substantial harm. The expected total cost per unit time over an arbitrarily long time interval might therefore not be the best metric for optimizing testing frequency. An alternative would be to use the rate density for origination of new pathogens, λ', to estimate the probability of either an unusually large number of pathogens or an uncharacteristically lethal pathogen. To mitigate the possibility of excessive harm from these rare events, the optimal testing frequency could be increased appropriately.

Our model and its many possible extensions can inform the design of these critical aspects of environmental and vector surveillance platforms. Our work provides a general and robust foundation for mechanistic optimization of environmental surveillance for infectious diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project has been funded in part by contract 200-2016-91779 with the Centers for Disease Control and Prevention, United States. Disclaimer: The findings, conclusions, and views expressed are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC). S.M.K. received funding from NIH T32 training grant 2 T32 AI 7535-21 A1. M.L. is grateful for funding from the Morris-Singer Fund. All authors performed research and wrote the paper.

References

Ahmed Warish, Bivins Aaron, Bertsch Paul M., Bibby Kyle, Choi Phil M., Farkas Kata, Gyawali Pradip, Hamilton Kerry A., Haramoto Eiji, Kitajima Masaaki, Simpson Stuart L., Tandukar Sarmila, Thomas Kevin V., Mueller Jochen F., 2020. Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimization and quality control are crucial for generating reliable public health information. Curr. Opin. Environ. Sci. Health 17, 82–93. 10.1016/j.coesh.2020.09.003.

Baker Rachel E., Mahmud Ayesha S., Miller Ian F., Rajeev Malavika, Rasambainarivo Fidisoa, Rice Benjamin L., Takahashi Saki, Tatem Andrew J., Wagner Caroline E., Wang Lin-Fa, Wesolowski Amy, Metcalf C. Jessica E., 2021. Infectious disease in an era of global change. Nat. Rev. Microbiol. 20, 193–205. 10.1038/s41579-021-00639-z. [PubMed: 34646006]

Bernstein Aaron S., Ando Amy W., Loch-Temzelides Ted, Vale Mariana M., Li Binbin V., Li Hongying, Busch Jonah, Chapman Colin A., Kinnaird Margaret, Nowak Katarzyna, Castro Marcia C., Zambrana-Torrelio Carlos, Ahumada Jorge A., Xiao Lingyun, Roehrdanz Patrick, Kaufman Les, Hannah Lee, Daszak Peter, Pimm Stuart L., Dobson Andrew P., 2022. The costs and benefits of primary prevention of zoonotic pandemics. Sci. Adv. 8 (5), eabl4183. 10.1126/sciadv.abl4183. [PubMed: 35119921]

Boehm Alexandria B., Hughes Bridgette, Duong Dorothea, Chan-Herur Vikram, Buchman Anna, Wolfe Marlene K., White Bradley J., 2023. Wastewater concentrations of human influenza, metapneumovirus, parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids during the COVID-19 pandemic: a surveillance study. Lancet Microbe 10.1016/S2666-5247(22)00386-X.

Braks Marieta, Medlock Jolyon M., Hubalek Zdenek, Hjertqvist Marika, Perrin Yvon, Lancelot Renaud, Duchyene Els, Hendrickx Guy, Stroo Arjan, Heyman Paul, Sprong Hein, 2014. Vector-borne disease intelligence: strategies to deal with disease burden and threats. Front. Publ. Health 2, 280. 10.3389/fpubh.2014.00280.

Budd Jobie, Miller Benjamin S., Manning Erin M., Lampos Vasileios, Zhuang Mengdie, Edelstein Michael, Rees Geraint, Emery Vincent C., Stevens Molly M., Keegan Neil, Short Michael J., Pillay Deenan, Manley Ed, Cox Ingemar J., Heymann David, Johnson Anne M., McKendry Rachel A., 2020. Digital technologies in the public-health response to COVID-19. Nature Med. 26, 1183–1192. 10.1038/s41591-020-1011-4. [PubMed: 32770165]

CDC, 2023. Important updates on locally acquired malaria cases identified in Florida, Texas, and Maryland. CDC Health Alert Network 2023, HAN00496, URL https://emergency.cdc.gov/han/2023/han00496.asp.

Chen William, Bibby Kyle, 2022. Model-based theoretical evaluation of the feasibility of using wastewater-based epidemiology to monitor monkeypox. Environ. Sci. Technol. Lett. 9 (9), 772–778. 10.1021/acs.estlett.2c00496.

DeFelice Nicholas B., Little Eliza, Campbell Scott R., Shaman Jeffrey, 2017. Ensemble forecast of human West Nile virus cases and mosquito infection rates. Nature Commun. 8, 14592. 10.1038/ncomms14592. [PubMed: 28233783]

Diamond Megan B., Keshaviah Aparna, Bento Ana I., Conroy-Ben Otakuye, Driver Erin M., Ensor Katherine B., Halden Rolf U., Hopkins Loren P., Kuhn Katrin G., Moe Christine L., Rouchka Eric C., Smith Ted, Stevenson Bradley S., Susswein Zachary, Vogel Jason R., Wolfe Marlene K., Stadler Lauren B., Scarpino Samuel V., 2022. Wastewater surveillance of pathogens can inform public health responses. Nature Med. 28, 1992–1995. 10.1038/s41591-022-01940-x. [PubMed: 36076085]

Eisen Rebecca J., Paddock Christopher D., 2021. Tick and tickborne pathogen surveillance as a public health tool in the United States. J. Med. Entomol. 58 (4), 1490–1502. 10.1093/jme/tjaa087. [PubMed: 32440679]

Fournet Florence, Jourdain Frederic, Bonnet Emmanuel, Degroote Stephanie, Ridde Valery, 2018. Effective surveillance systems for vector-borne diseases in urban settings and translation of the data into action: a scoping review. Infect. Dis. Poverty 7, 99. 10.1186/s40249-018-0473-9. [PubMed: 30217142]

Gu Weidong, Unnasch Thomas R., Katholi Charles R., Lampman Richard, Novak Robert J., 2008. Fundamental issues in mosquito surveillance for arboviral transmission. Trans. R. Soc. Trop. Med. Hyg. 102 (8), 817–822. 10.1016/j.trstmh.2008.03.019. [PubMed: 18466940]

Gwinn Marta, MacCannell Duncan R., Khabbaz Rima F., 2017. Integrating advanced molecular technologies into public health. J. Clin. Microbiol. 55 (3), 703–714. 10.1128/JCM.01967-16. [PubMed: 28031438]

Hagedorn Brittany, Zhou Nicolette A., Fagnant-Sperati Christine S., Shirai Jeffry H., Gauld Jillian, Wang Yuke, Boyle David S., Meschke John Scott, 2023. Estimates of the cost to build a stand-alone environmental surveillance system for typhoid in low- and middle-income countries. PLOS Glob. Publ. Health 3 (1), e0001074. 10.1371/journal.pgph.0001074.

Hermance Meghan E., Thangamani Saravanan, 2017. Powassan virus: An emerging arbovirus of public health concern in North America. Vector Borne Zoonotic Dis. 17 (7), 453–462. 10.1089/vbz.2017.2110. [PubMed: 28498740]

Huang Yue, Zhou Nan, Zhang Shihan, Yi Youqin, Han Ying, Liu Minqi, Han Yue, Shi Naiyang, Yang Liuqing, Wang Qiang, Cui Tingting, Jin Hui, 2022. Norovirus detection in wastewater and its correlation with human gastroenteritis: a systematic review and meta-analysis. Environ. Sci. Pollut. Res. 29, 22829–22842. 10.1007/s11356-021-18202-x.

Jernigan Daniel B., George Dylan, Lipsitch Marc, 2023. Learning from COVID-19 to improve surveillance for emerging threats. Am. J. Publ. Health 113 (5), 520–522. 10.2105/AJPH.2023.307261.

Kantor Rose S., Greenwald Hannah D., Kennedy Lauren C., Hinkle Adrian, Harris-Lovett Sasha, Metzger Matthew, Thornton Melissa M., Paluba Justin M., Nelson Kara L., 2022. Operationalizing a routine wastewater monitoring laboratory for SARS-CoV-2. PLOS Water 1 (2), e0000007. 10.1371/journal.pwat.0000007.

Keshaviah Aparna, Hu Xindi C., Henry Marisa, 2021. Developing a flexible national wastewater surveillance system for COVID-19 and beyond. Environ. Health Perspect. 129 (4), 045002. 10.1289/EHP8572. [PubMed: 33877858]

Kirby Amy E., Walters Maroya Spalding, Jennings Wiley C., Fugitt Rebecca, LaCross Nathan, Mattioli Mia, Marsh Zachary A., Roberts Virginia A., Mercante Jeffrey W., Yoder Jonathan, Hill Vincent R., 2021. Using wastewater surveillance data to support the COVID-19 response—United States, 2020–2021. Morb. Mortal. Wkly. Rep. 70 (36), 1242–1244. 10.15585/mmwr.mm7036a2.

Kretschmer Melissa, Collins Jennifer, Dale Ariella P., Garrett Brenna, Koski Lia, Zabel Karen, Staab R. Nicholas, Turnbow Katie, Nativio Judah, Andrews Kelsey, Smith William E., Townsend John, Busser Nicole, Will James, Burr Kathryn, Jones Forrest K., Santiago Gilberto A., Fitzpatrick Kelly A., Ruberto Irene, Fitzpatrick Kathryn, White Jessica R., Adams Laura, Sunenshine Rebecca H., 2023. Notes from the field: First evidence of locally acquired dengue virus infection —Maricopa county, Arizona, november 2022. Morb. Mortal. Wkly. Rep. 72 (11), 290–291. 10.15585/mmwr.mm7211a5.

Levy Joshua I., Andersen Kristian G., Knight Rob, Karthikeyan Smruthi, 2023. Wastewater surveillance for public health. Science 379 (6627), 26–27. 10.1126/science.ade2503. [PubMed: 36603089]

Lipsitch Marc, Riley Steven, Cauchemez Simon, Ghani Azra C., Ferguson Neil M., 2009. Managing and reducing uncertainty in an emerging influenza pandemic. N. Engl. J. Med. 361 (2), 112–115. 10.1056/NEJMp0904380. [PubMed: 19474417]

Mercier Elisabeth, D'Aoust Patrick M., Ocean Thakali, Hegazy Nada, Jia Jian-Jun, Zhang Zhihao, Eid Walaa, Plaza-Diaz Julio, Kabir Md Pervez, Fang Wanting, Cowan Aaron, Stephenson Sean E., Pisharody Lakshmi, MacKenzie Alex E., Graber Tyson E., Wan Shen, Delatolla Robert, 2022. Municipal and neighbourhood level wastewater surveillance and subtyping of an influenza virus outbreak. Sci. Rep. 12, 15777. 10.1038/s41598-022-20076-z. [PubMed: 36138059]

Michael-Kordatou I, Karaolia P, Fatta-Kassinos D, 2020. Sewage analysis as a tool for the COVID-19 pandemic response and management: the urgent need for optimised protocols for SARS-CoV-2 detection and quantification. J. Environ. Chem. Eng. 8 (5), 104306. 10.1016/j.jece.2020.104306. [PubMed: 32834990]

Murray Jillian, Cohen Adam L., 2017. Infectious disease surveillance. In: Quah Stella R. (Ed.), International Encyclopedia of Public Health (Second Edition), second ed. Academic Press, Oxford, pp. 222–229. 10.1016/B978-0-12-803678-5.00517-8.

Nguyen Anh Q., Vu Hang P., Nguyen Luong N., Wang Qilin, Djordjevic Steven P., Donner Erica, Yin Huabing, Nghiem Long D., 2021. Monitoring antibiotic resistance genes in wastewater treatment: Current strategies and future challenges. Sci. Total Environ. 783, 146964. 10.1016/j.scitotenv.2021.146964. [PubMed: 33866168]

Ngwira Lucky G., Sharma Bhawana, Shrestha Kabita Bade, Dahal Sushil, Tuladhar Reshma, Manthalu Gerald, Chilima Ben, Ganizani Allone, Rigby Jonathan, Kanjerwa Oscar, Barnes Kayla, Anscombe Catherine, Mfutso-Bengo Joseph, Feasey Nicholas, Mvundura Mercy, 2022. Cost of wastewater-based environmental surveillance for SARS-CoV-2: Evidence from pilot sites in Blantyre, Malawi and Kathmandu, Nepal. PLOS Glob. Publ. Health 2 (12), e0001377. 10.1371/journal.pgph.0001377.

Oran Daniel P., Topol Eric J., 2020. Prevalence of asymptomatic SARS-CoV-2 infection. Ann. Intern. Med. 173 (5), 362–367. 10.7326/M20-3012. [PubMed: 32491919]

Peak Corey M., Childs Lauren M., Grad Yonatan H., Buckee Caroline O., 2017. Comparing nonpharmaceutical interventions for containing emerging epidemics. Proc. Natl. Acad. Sci. USA 114 (15), 4023–4028. 10.1073/pnas.1616438114. [PubMed: 28351976]

Peccia Jordan, Zulli Alessandro, Brackney Doug E., Grubaugh Nathan D., Kaplan Edward H., Casanovas-Massana Arnau, Ko Albert I., Malik Amyn A., Wang Dennis, Wang Mike, Warren Joshua L., Weinberger Daniel M., Arnold Wyatt, Omer Saad B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat. Biotechnol. 38, 1164–1167. 10.1038/s41587-020-0684-z. [PubMed: 32948856]

Petersen Lyle R., Brault Aaron C., Nasci Roger S., 2013. West Nile virus: Review of the literature. JAMA 310 (3), 308–315. 10.1001/jama.2013.8042. [PubMed: 23860989]

Pfaller Michael A., 2001. Molecular approaches to diagnosing and managing infectious diseases: Practicality and costs. Emerging Infect. Dis. 7 (2), 312–318. 10.3201/eid0702.010234.

Pley Caitlin, Evans Megan, Lowe Rachel, Montgomery Hugh, Yacoub Sophie, 2021. Digital and technological innovation in vector-borne disease surveillance to predict, detect, and control climate-driven outbreaks. Lancet Planet. Health 5 (10), e739–e745. 10.1016/S2542-5196(21)00141-8. [PubMed: 34627478]

Quinn Sandra Crouse, Kumar Supriya, 2014. Health inequalities and infectious disease epidemics: A challenge for global health security. Biosecur. Bioterror. 12 (5), 263–273. 10.1089/bsp.2014.0032. [PubMed: 25254915]

Shah Shimoni, Gwee Sylvia Xiao Wei, Ng Jamie Qiao Xin, Lau Nicholas, Koh Jiayun, Pang Junxiong, 2022. Wastewater surveillance to infer COVID-19 transmission: A systematic review. Sci. Total Environ. 804, 150060. 10.1016/j.scitotenv.2021.150060. [PubMed: 34798721]

Teklehaimanot Hailay Desta, Schwartz Joel, Teklehaimanot Awash, Lipsitch Marc, 2004. Alert threshold algorithms and malaria epidemic detection. Emerg. Infect. Diseases 10 (7), 1220–1226. 10.3201/eid1007.030722. [PubMed: 15324541]

Thompson PN, Etter E, 2015. Epidemiological surveillance methods for vector-borne diseases. Rev. Sci. Tech. Off. Int. Epiz. 34 (1), 235–247. 10.20506/rst.34.1.2356.

Tiwari Ananda, Adhikari Sangeet, Kaya Devrim, Islam Md. Aminul, Malla Bikash, Sherchan Samendra P., Al-Mustapha Ahmad I., Kumar Manish, Aggarwal Srijan, Bhattacharya Prosun, Bibby Kyle, Halden Rolf U., Bivins Aaron, Haramoto Eiji, Oikarinen Sami, Heikinheimo Annamari, Pitkanen Tarja, 2023. Monkeypox outbreak: Wastewater and environmental surveillance perspective. Sci. Total Environ. 856 (2), 159166. 10.1016/j.scitotenv.2022.159166. [PubMed: 36202364]

Tiwari Ananda, Kurittu Paula, Al-Mustapha Ahmad I., Heljanko Viivi, Johansson Venla, Thakali Ocean, Mishra Shyam Kumar, Lehto Kirsi-Maarit, Lipponen Anssi, Oikarinen Sami, Pitkanen Tarja, Group WastPan Study, Heikinheimo Annamari, 2022. Wastewater surveillance of antibiotic-resistant bacterial pathogens: A systematic review. Front. Microbiol. 13, 977106. 10.3389/fmicb.2022.977106. [PubMed: 36590429]

Vazquez-Prokopec Gonzalo M., Chaves Luis F., Ritchie Scott A., Davis Joe, Kitron Uriel, 2010. Unforeseen costs of cutting mosquito surveillance budgets. PLOS Negl. Trop. Dis. 4 (10), e858. 10.1371/journal.pntd.0000858. [PubMed: 21049010]

Weinstein Milton C., Torrance George, McGuire Alistair, 2009. QALYs: The basics. Value Health 12 (1), S5–S9. 10.1111/j.1524-4733.2009.00515.x. [PubMed: 19250132]

Zinsstag Jakob, Utzinger Jurg, Probst-Hensch Nicole, Shan Lv, Zhou Xiao-Nong, 2020. Towards integrated surveillance-response systems for the prevention of future pandemics. Infect. Dis. Poverty 9, 140. 10.1186/s40249-020-00757-5. [PubMed: 33028426]

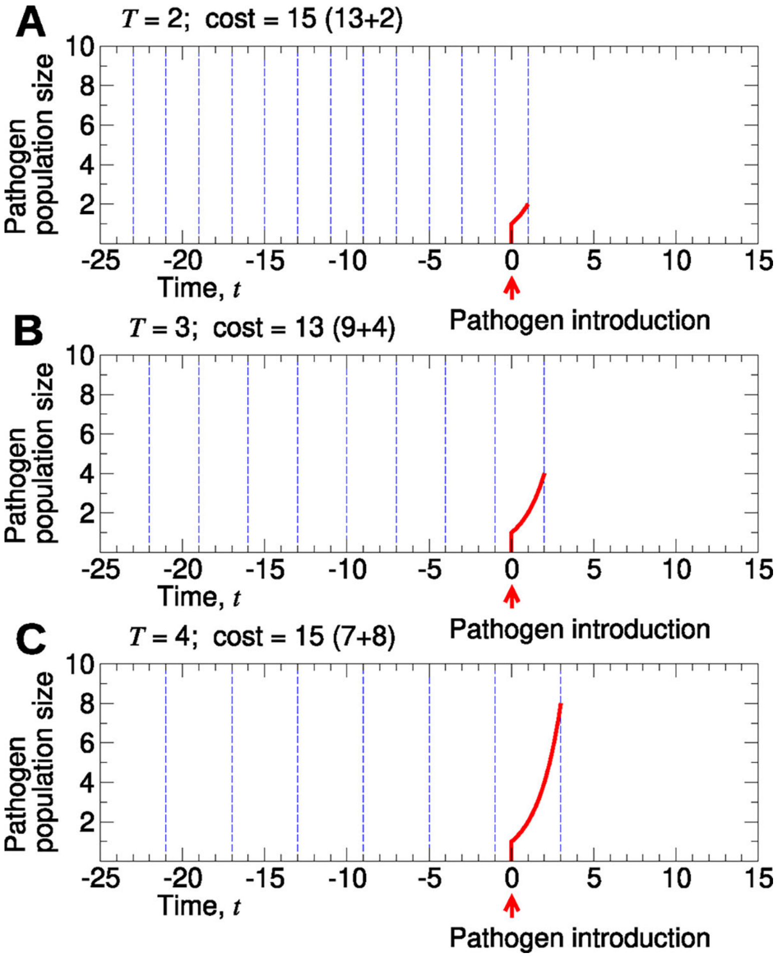

Fig. 1. Optimization of surveillance.

A simple schematic illustrates how surveillance can be optimized. In the plot, the dotted blue lines represent sampling events, and the solid red curves represent the abundance of a pathogen. Here, we assume that a pathogen first emerges at time $t = 0$ with the pathogen population growing exponentially, doubling at each subsequent time step. Sampling of the environment occurs at times $-25 + mT$ – where T is the sampling period and $m \geq 1$ is an integer – until the pathogen is first detected. We plot the outcomes if the sampling period had been (**A**) $T = 2$, (**B**) $T = 3$, or (**C**) $T = 4$. If the cost associated with one sampling event is equal to the cost associated with one instance of the pathogen, and if costs accumulate linearly, then $T = 3$ would have resulted in the lowest total cost.

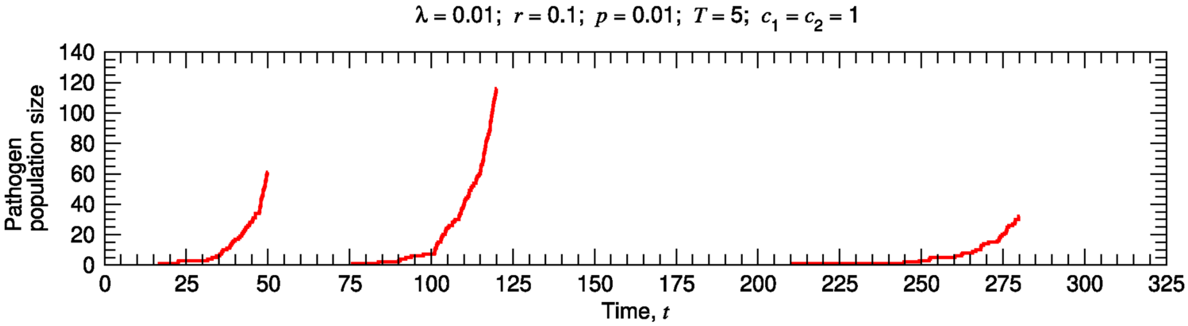

Fig. 2. Stochastic surveillance model.

A single realization of the stochastic surveillance and disease dynamics is shown. The environment is tested at times $5m$, where m is an integer and $1 \le m \le 65$. There are thus 65 testing events, so the cumulative surveillance cost is 65. The first lineage begins at time $t \approx 16.7$ and is detected at time $t = 50$, when its size is 61. The second lineage begins at time $t \approx 75.8$ and is detected at time $t = 120$, when its size is 116. The third lineage begins at time $t \approx 210.6$ and is detected at time $t = 280$, when its size is 32. The cumulative disease cost is thus 61 + 116 + 32 = 209. The total surveillance and disease cost is 65 + 209 = 274, and the total cost per unit time is 274⁄325.

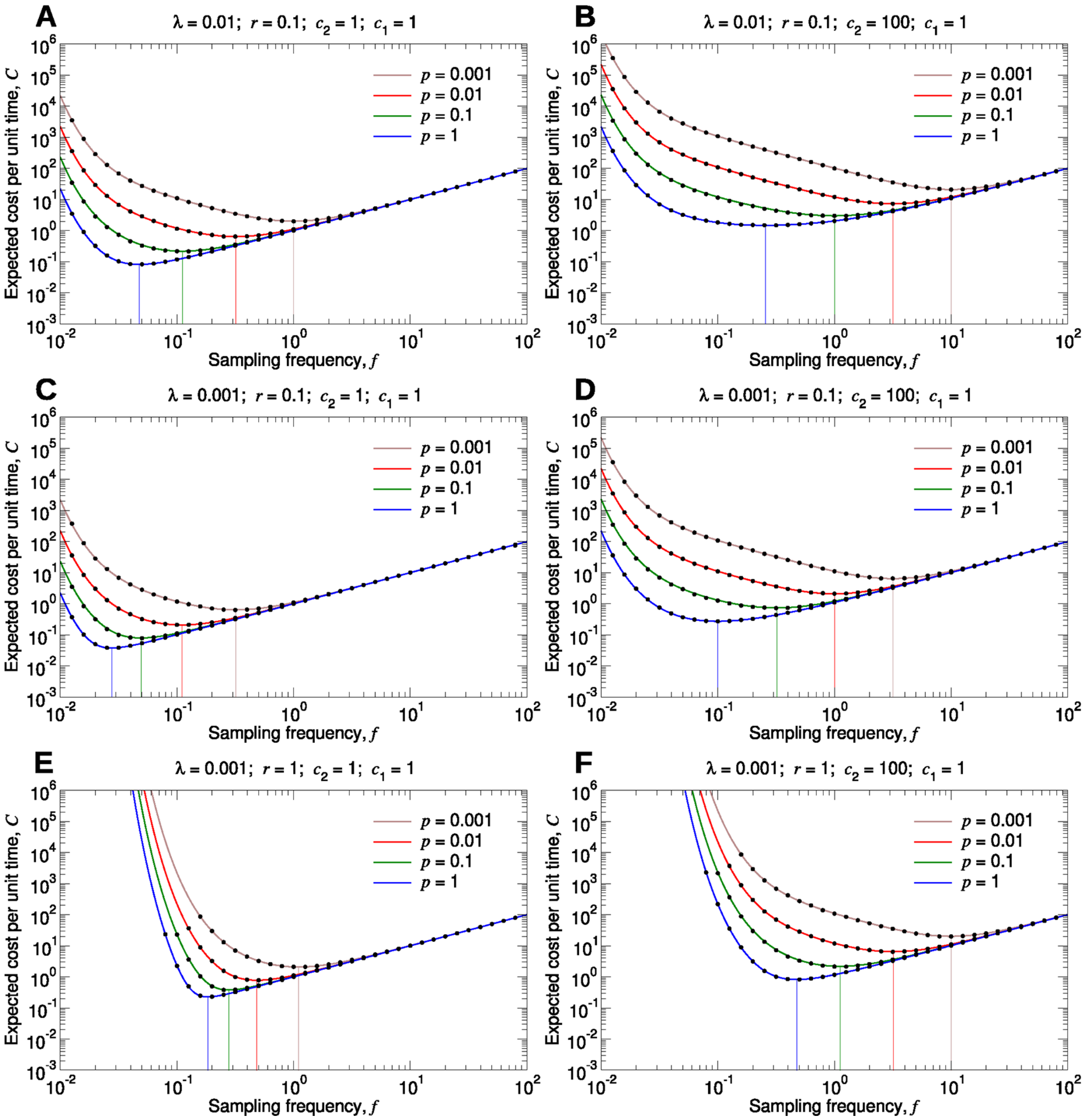

Fig. 3. Expected total cost per unit time for a particular type of pathogen.

(**A** through **F**) We set $c_1 = 1$, and we plot C, given by Eq. (1), as a function of f for several values of λ, r, c_2, and p. The black dots are measurements of the expected total cost per unit time from simulating the true stochastic process. The vertical lines show the sampling frequencies for which this quantity is minimal in each case. For each parameter set, we simulated 10^3 pathogen introductions, and we computed the total surveillance and pathogen cost divided by the time elapsed. We repeated this simulation sixteen times and calculated the average. Standard errors are smaller than the size of the data points.

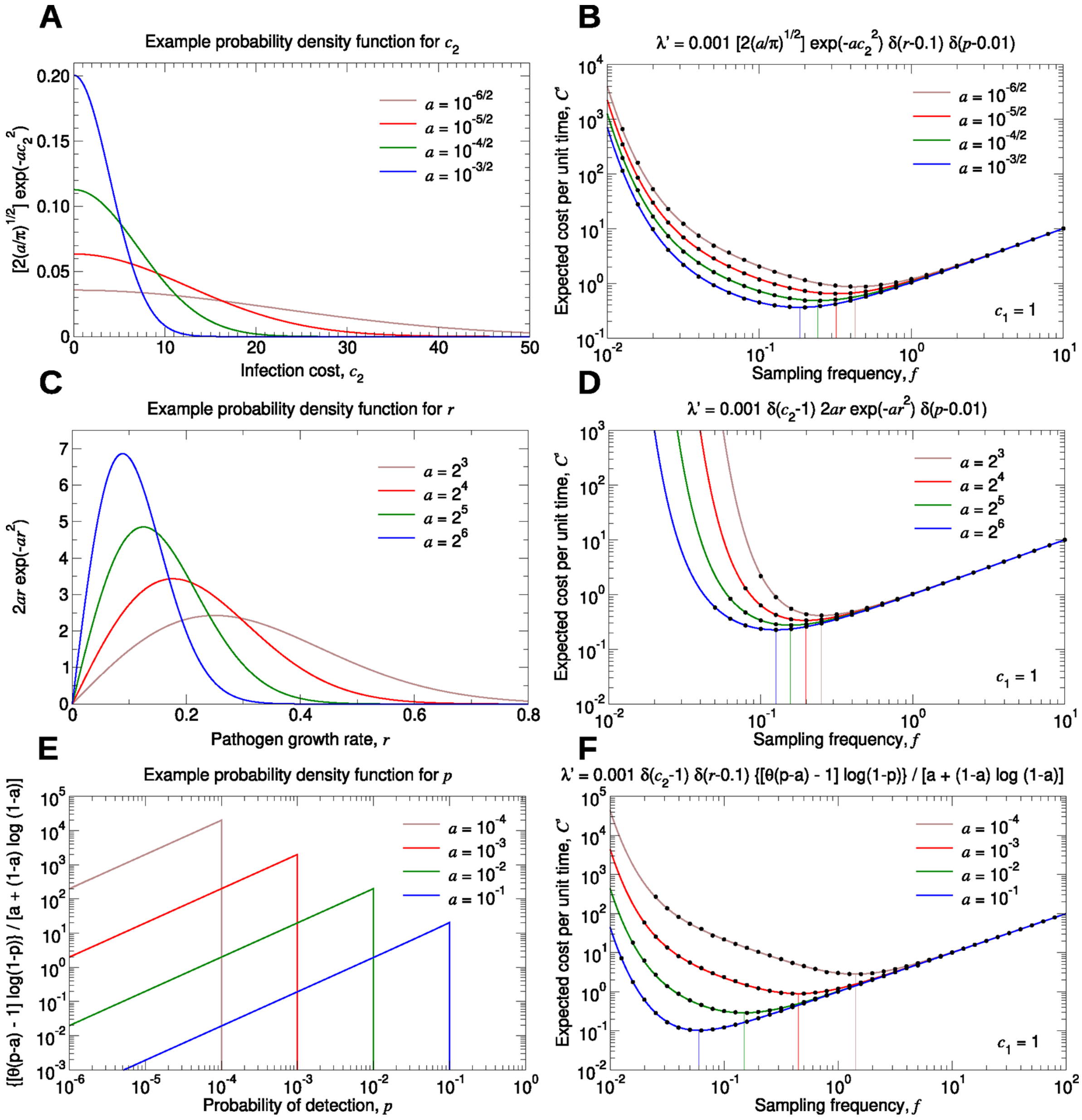

Fig. 4. Expected total cost per unit time accounting for many types of pathogens.

(**A**, **C**, and **E**) We show example probability density functions for the parameters c_2, r, and p, respectively. For each case, we introduce a single parameter, a, which controls the shape of the probability density function. (**B**, **D**, and **F**) We set $c_1 = 1$, and we plot C', given by Eq. (2), as a function of f for several rate density functions, $\lambda'(c_2, r, p)$. The vertical lines show the sampling frequencies for which this quantity is minimal in each case. For each parameter set in (**B**) and (**F**), we simulated 10^3 pathogen introductions, and for each parameter set in (**D**), we simulated 10^5 pathogen introductions. For each case, we computed the total surveillance and pathogen cost divided by the time elapsed. We repeated this simulation sixteen times and calculated the average (black dots). Standard errors are smaller than the size of the data points.

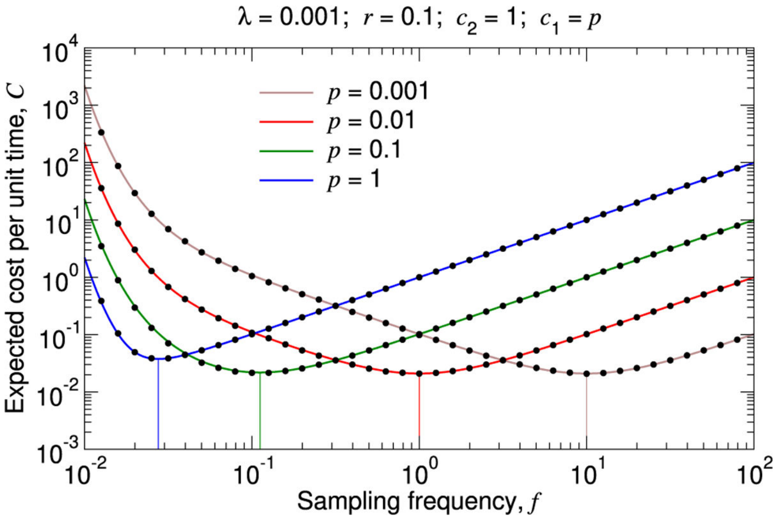

Fig. 5. Expected total cost per unit time assuming that surveillance cost depends on sensitivity.

We set $c_1 = p$, and we plot Eq. (1) together with measurements of the expected total cost per unit time from simulating the true stochastic process (black dots). If c_1 is positively associated with p, then the optimal sampling frequency has a strong inverse relation with p. The vertical lines show the sampling frequencies for which the expected total cost per unit time is minimal in each case. For each parameter set, we simulated 10^3 pathogen introductions, and we computed the total surveillance and pathogen cost divided by the time elapsed. We repeated this simulation sixteen times and calculated the average. Standard errors are smaller than the size of the data points.

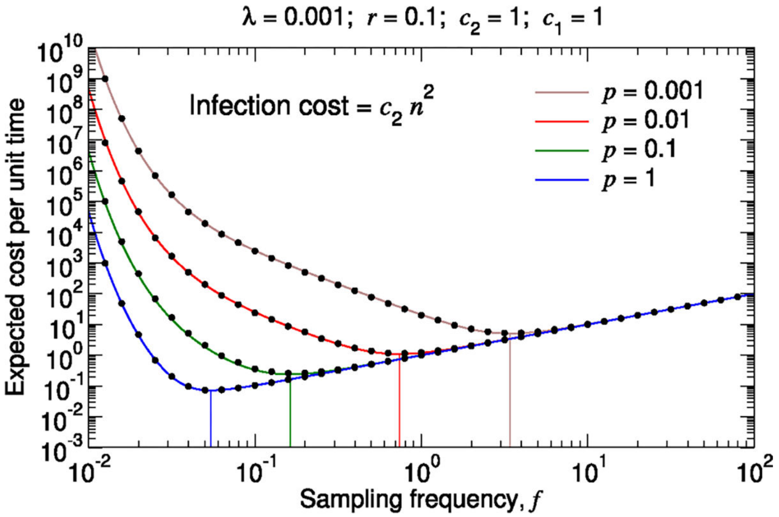

Fig. 6. Expected total cost per unit time assuming that the cost due to a single outbreak scales quadratically with the number of clinical cases.

We assume that the cost due to a single outbreak equals c_2n^2. We plot Equation S33 in the Supplementary Information together with measurements of the expected total cost per unit time from simulating the true stochastic process (black dots). If the cost due to an outbreak scales quadratically with the number of clinical cases as opposed to linearly, then the optimal sampling frequency is higher. The vertical lines show the sampling frequencies for which the expected total cost per unit time is minimal in each case. For each parameter set, we simulated 10^3 pathogen introductions, and we computed the total surveillance and pathogen cost divided by the time elapsed. We repeated this simulation sixteen times and calculated the average. Standard errors are smaller than the size of the data points.

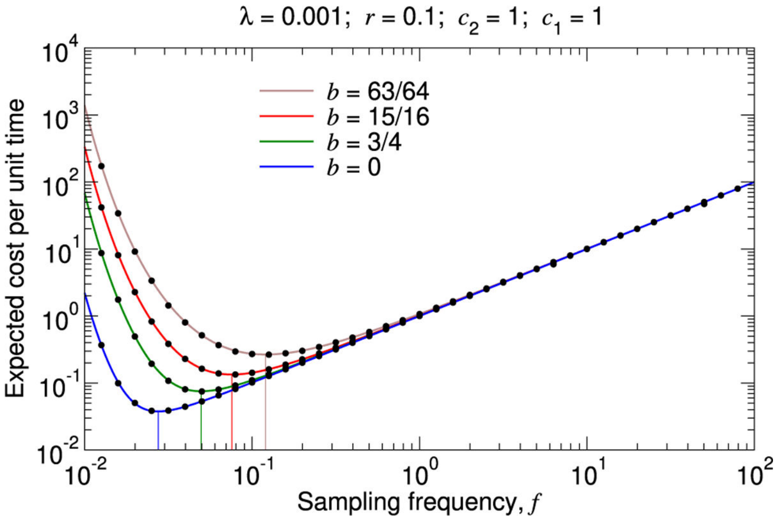

Fig. 7. Expected total cost per unit time considering the possibility of breakthrough infections.

We plot Equation S39 in the Supplementary Information together with measurements of the expected total cost per unit time from simulating the true stochastic process (black dots). Whenever an outbreak is detected and controlled, with probability b, there is a single new infection that initiates immediately thereafter. For larger values of b, the optimal sampling frequency is higher. The vertical lines show the sampling frequencies for which the expected total cost per unit time is minimal in each case. For each parameter set, we simulated 10^3 pathogen introductions, and we computed the total surveillance and pathogen cost divided by the time elapsed. We repeated this simulation sixteen times and calculated the average. Standard errors are smaller than the size of the data points.