Optimal environmental testing frequency for outbreak surveillance

Jason W. Olejarz[1,2], Kirstin I. Oliveira Roster[1,2], Stephen M. Kissler[3,1,2], Marc Lipsitch[1,2,4], and Yonatan H. Grad[1,2]

[1]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA
[2]Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA
[3]Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309, USA
[4]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 of disease burden and surveillance given a sampling frequency and thus the sampling frequency for which the expected total cost is lowest.

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

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 [1, 2, 3]. 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 [4, 5, 6].

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 identified [7]. In contexts where access to care or resources are limited, missed cases and reporting delays can make it difficult to rapidly detect and correctly diagnose new infections [8].

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 [9, 10] and has been used more recently for tracking the local epidemic dynamics and evolution of SARS-CoV-2 [11, 12], norovirus [13], influenza [14], mpox [15, 16], and other pathogens [17]. Efforts are underway to extend these techniques for tracking antibiotic resistance genes in wastewater [18, 19]. For vector-borne pathogens, including West Nile virus [20], *Borrelia* species [21], and Powassan virus [22], 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 [23, 24, 25, 26, 27, 28]. Although the potential for environmental and vector-based surveillance systems have been recognized and widely discussed, and despite the massive push to fund and develop these programs, particularly wastewater efforts [29, 30], there remains a critical gap in our understanding: How should surveillance be designed to achieve maximal effectiveness [31, 32, 33, 34, 35, 36]? A central consideration is how often testing should be performed (Figure 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 (i.e., the sum of all costs related to surveillance and to effects from the disease, when considered as an average over many realizations of the stochastic dynamics). The expected total cost is a function of the parameters of the model, and given values for these parameters, we can minimize the expected total cost.

Our goal was to minimize the expected total surveillance and disease cost 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. Sampling begins at time $t = 0$ and continues regularly at time $t_n = nT$, where $n \geq 1$ 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

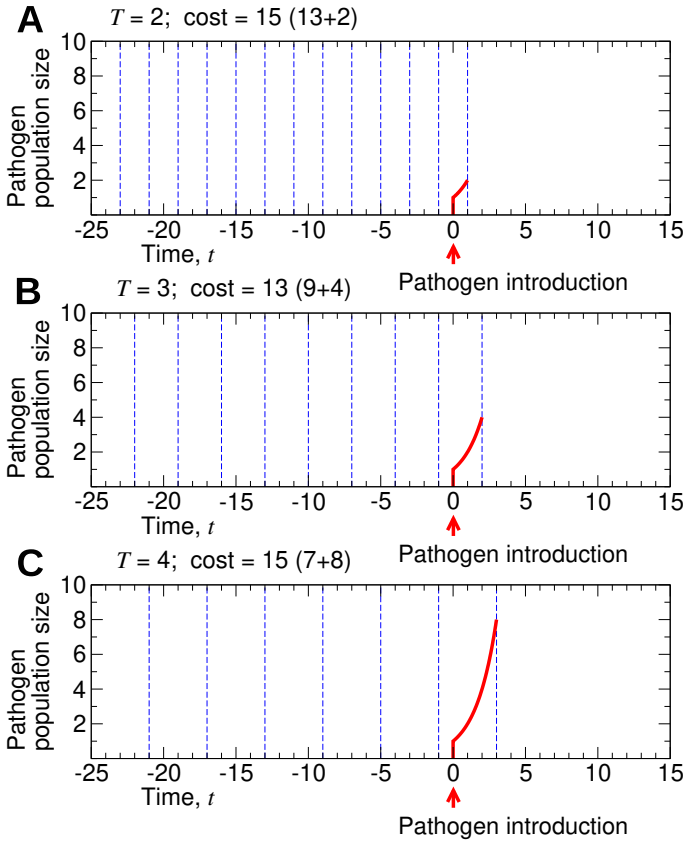

Figure 1: **Optimization of surveillance.** A simple scenario illustrates how surveillance can be optimized. In the plot, the dotted blue lines represent sampling events, and the solid red curves represent the abundance of a pathogen. Here, we assume that a pathogen first emerges at time $t = 0$ with the pathogen population growing exponentially, doubling at each subsequent time step. Sampling of the environment occurs at times $-25 + nT$—where T is the sampling period and $n \geq 1$ 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.

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 $c_2 N$. 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 Figure 2.

Results

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

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

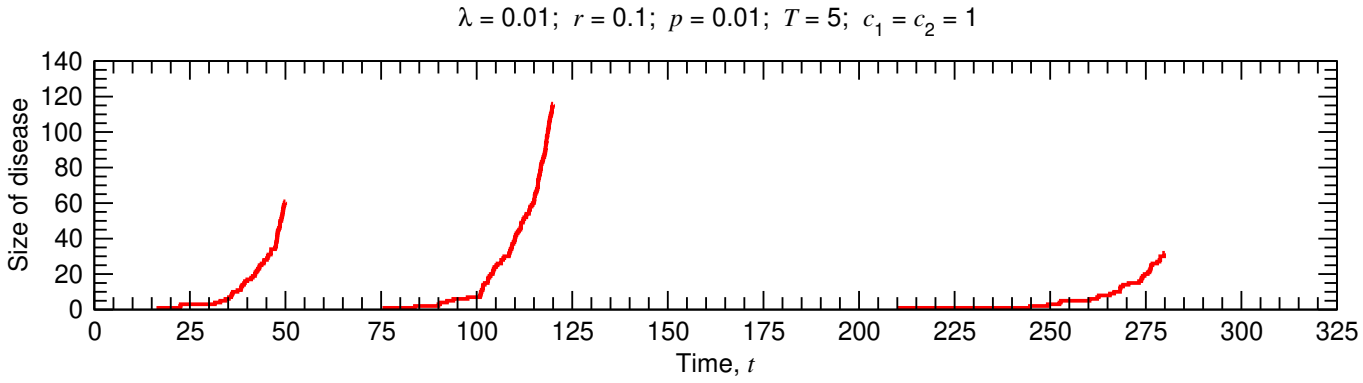

Figure 2: **Stochastic surveillance model.** A single realization of the stochastic surveillance and disease dynamics is shown. The environment is tested at times $5n$, where n is an integer and $0 \leq n \leq 65$. There are thus 66 testing events, so the cumulative surveillance cost is 66. The first 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 $66 + 209 = 275$, and the total cost per unit time is $275/325 = 11/13$.

Details on the derivation of Equation (1) are provided in the Supplementary Information. By comparing Equation (1) with $c_1/T + \lambda c_2 \langle N \rangle$, notice that the product of the two factors in large parentheses is 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. In the limit of a perfectly sensitive detector (i.e., $p \to 1$), we have $-\log(1-p) \to \infty$, and the second factor in big parentheses just approaches 1. So as $p \to 1$, we have $\langle N \rangle \approx (e^{rT} - 1)/(rT)$, which is the expected size of the outbreak when the first test happens, given an introduction. If the detector has poor sensitivity (i.e., $p \ll 1$), then $-\log(1-p) \approx p$, and the second factor in big parentheses is approximately equal to $(1 - e^{-rT})/p$. So for small values of p, we have $\langle N \rangle \approx (e^{rT} - 1)(1 - e^{-rT})/(rTp)$. For large values of the pathogen growth rate, r, we have $\langle N \rangle \approx [e^{rT}/(rT)][1 - 1/\log(1-p)]$, and for small values of r, we have $\langle N \rangle \to 1$—i.e., the pathogen is detected before it has a chance to produce new infections. Since $\langle N \rangle$ is a function of the product rT (not of r and T individually), its behavior as a function of the testing period, T, is the same: For large values of T, we have $\langle N \rangle \approx [e^{rT}/(rT)][1 - 1/\log(1-p)]$, and for small values of T, we have $\langle N \rangle \to 1$. 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.

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 Figure 3, where we plotted $\langle C \rangle$ as a function of the sampling frequency, $f = 1/T$, for different values of the five model parameters. It is instructive to consider the effects of very large or very small values of f on $\langle C \rangle$. As we increase f, we can detect the disease more rapidly, thereby diminishing 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 Figure 3, $\langle C \rangle$ attains a minimum for a particular value of f.

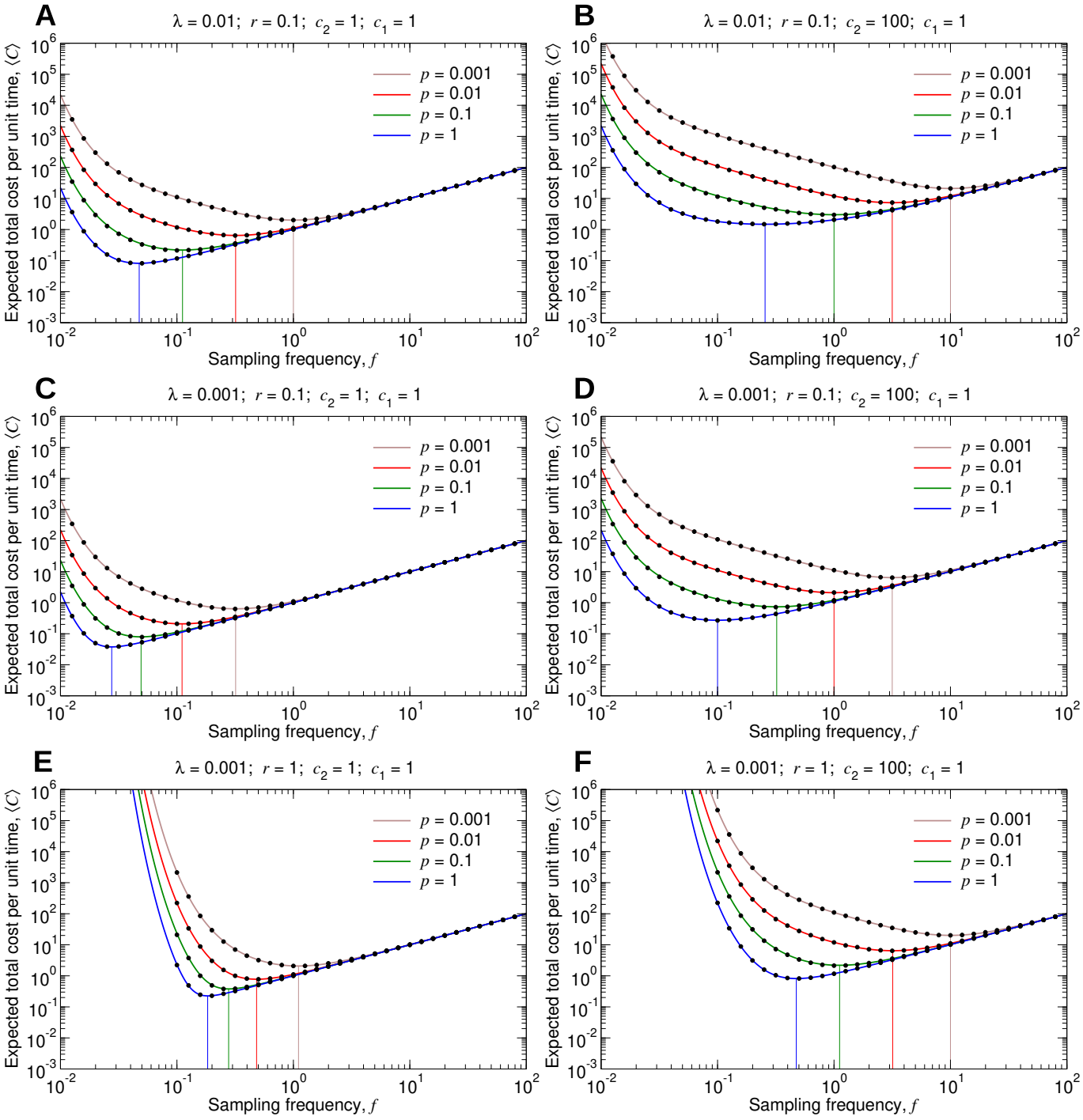

Figure 3: **Expected total cost for a particular type of pathogen.** (**A** through **F**) We set $c_1 = 1$, and we plot $\langle C \rangle$, given by Equation (1), as a function of f for several values of λ, r, c_2, and p. The black dots are measurements of the expected total cost from simulating the true stochastic process. The 95% confidence intervals are smaller than the size of the data points. The vertical lines show the sampling frequencies for which the expected total cost is minimal in each case.

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, $\langle C'\rangle$, is equal to

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

Equation (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 is lowest specifies the optimal testing frequency, F^*. From Equation (2), we have

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

Figure 4 shows how this works. In Figure 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 Figure 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, $\langle C'\rangle$. (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 Figure 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 Figure 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 $\langle C'\rangle$. 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 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.

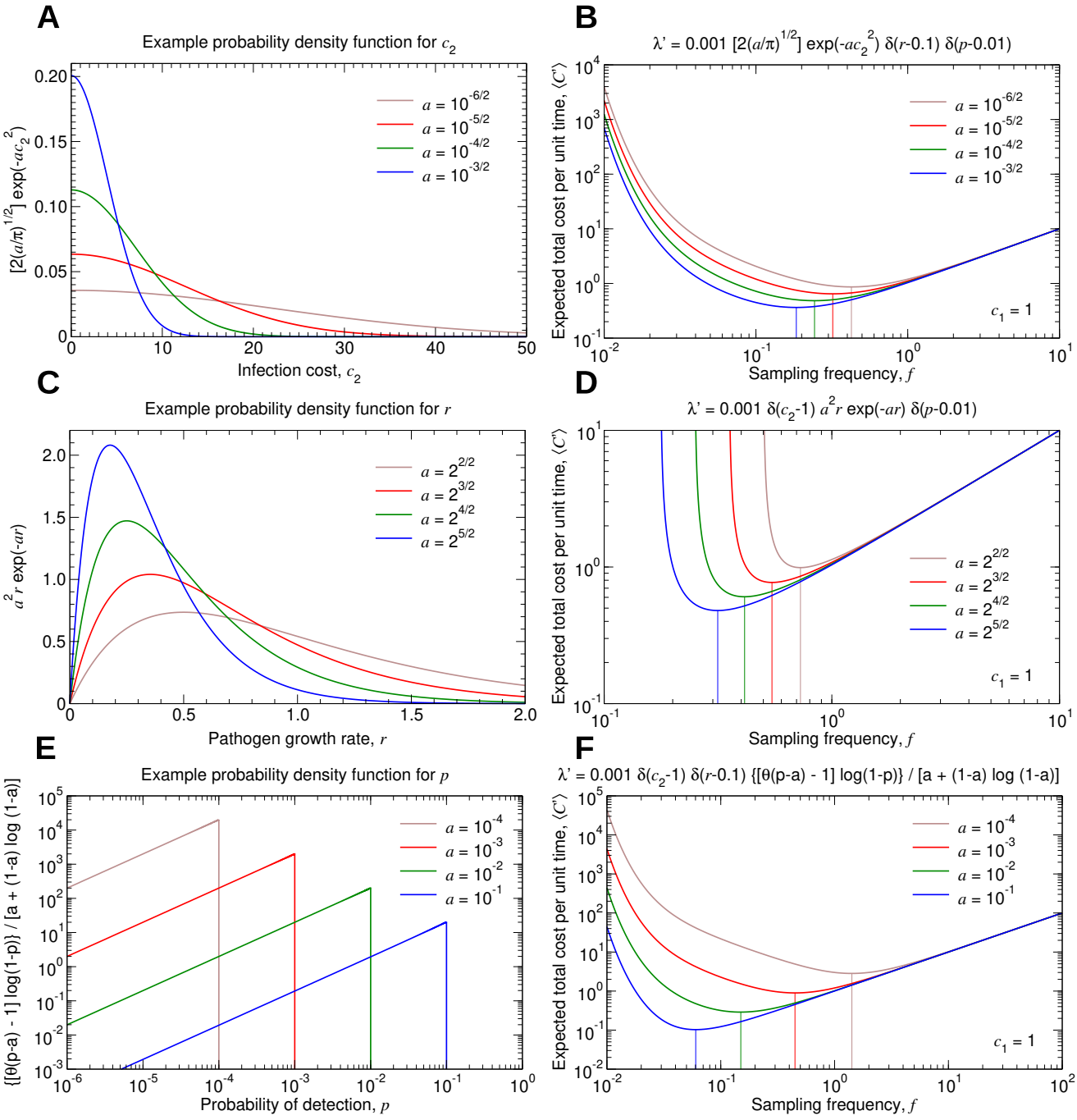

Figure 4: **Expected total cost accounting for many types of pathogens.** (**A**, **C**, and **E**) We show example probability density functions for the parameters c_2, r, and p, respectively. For each case, we introduce a single parameter, a, which controls the shape of the probability density function. (**B**, **D**, and **F**) We set $c_1 = 1$, and we plot $\langle C' \rangle$, given by Equation (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 the expected total cost is minimal in each case.

In Figure 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 Figure 4F, setting $c_2 = 1$ and $r = 0.1$, and setting the total rate of emergence of new pathogens to 0.001, we plot

$\langle C' \rangle$. (In the specification of λ', θ denotes the Heaviside step function.) Smaller values of a result in a larger optimal testing frequency. Details on Figure 4 are provided in the Supplementary Information.

The example probability density functions in Figure 4 were chosen here for convenience: they have nice analytical forms, and they admit simple analytical solutions when substituted into Equation (2). For optimizing an environmental or vector surveillance system in practice, one would construct a estimated form for $\lambda'(c_2, r, p)$ based on experimental or observational data, and the optimal testing frequency would be determined numerically using Equation (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, $\lambda'(c_2, r, p)$.

Discussion

Equations (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 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 this must be interpreted in the context of whichever laboratory tests are used to detect it. 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. This effect could be incorporated by using a reduced value of λ'. 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 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. A larger number of outbreaks having occurred during a particular time period would correspond to a larger value of λ'. Optimization of testing frequency further requires a formal understanding of surveillance-related and pathogen-related costs [37, 38, 39]. 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.

Our model for determining the optimal testing frequency is broadly applicable. If α is the probability of a test resulting in a false positive, then the sampling cost can be adjusted using the substitution $c_1 \to c_1 + \alpha k$, where k is the cost due to a false positive. This consideration can be extended by assuming that the probability of a false positive is dependent on p, i.e., $\alpha \to \alpha(p)$. 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. Once a pathogen is detected and an intervention is implemented, spread of the pathogen and its associated costs are not immediately halted, and this is accounted for by making the substitution $\lambda'(c_2, r, p) \to \lambda'(\kappa c_2, r, p)$, where $\kappa < 1$. 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., $\lambda'(c_2, r, p) \to \lambda'(c_2, r, p, t)$)

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. Changes in weather and climate affect the risk of an outbreak—especially for many vector-borne pathogens [40]—and this could also be modeled through time-dependence of λ'.

Our approach can be applied to answer another important question: Where should environmental sampling be performed? This would work by introducing 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 can be calculated. By trying different sampling locations, it is possible to find the sampling locations for which the expected total cost is minimal.

Environmental and vector surveillance are equally instrumental for tracking the prevalence of a pathogen [41]. An understanding of how and when to intervene is therefore essential [42, 43]. 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 [44]. If tracking indicates increased prevalence of a pathogen, then more frequent sampling and testing might be warranted. A further possibility is that testing could be performed frequently over several seasons to gain an understanding of the typical seasonal behavior for new pathogens or in new ecological settings to inform optimization and more efficiently track the abundance of the pathogen.

Our model and its many possible extensions can thus 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.

Acknowledgments

This project has been funded in part by contract 200-2016-91779 with the Centers for Disease Control and Prevention. 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.

Competing interests

The authors declare no competing interests.

Code availability

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

References

[1] Jillian Murray and Adam L. Cohen. Infectious disease surveillance. In Stella R. Quah, editor, *International Encyclopedia of Public Health (Second Edition)*, pages 222–229. Academic Press, Oxford, second edition edition, 2017.

[2] Jobie Budd, Benjamin S. Miller, Erin M. Manning, Vasileios Lampos, Mengdie Zhuang, Michael Edelstein, Geraint Rees, Vincent C. Emery, Molly M. Stevens, Neil Keegan, Michael J. Short, Deenan Pillay, Ed Manley, Ingemar J. Cox, David Heymann, Anne M. Johnson, and Rachel A. McKendry. Digital technologies in the public-health response to covid-19. *Nat. Med.*, 26:1183–1192, 2020.

[3] Daniel B. Jernigan, Dylan George, and Marc Lipsitch. Learning from covid-19 to improve surveillance for emerging threats. *Am. J. Public Health*, 113 (5):520–522, 2023.

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

[5] Melissa Kretschmer, Jennifer Collins, Ariella P. Dale, Brenna Garrett, Lia Koski, Karen Zabel, R. Nicholas Staab, Katie Turnbow, Judah Nativio, Kelsey Andrews, William E. Smith, John Townsend, Nicole Busser, James Will, Kathryn Burr, Forrest K. Jones, Gilberto A. Santiago, Kelly A. Fitzpatrick, Irene Ruberto, Kathryn Fitzpatrick, Jessica R. White, Laura Adams, and Rebecca H. Sunenshine. 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, 2023.

[6] CDC. Important updates on locally acquired malaria cases identified in Florida, Texas, and Maryland. *CDC Health Alert Network*, 2023:HAN00496, 2023.

[7] Daniel P. Oran and Eric J. Topol. Prevalence of asymptomatic sars-cov-2 infection. *Ann. Intern. Med.*, 173 (5):362–367, 2020.

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

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

[10] Shimoni Shah, Sylvia Xiao Wei Gwee, Jamie Qiao Xin Ng, Nicholas Lau, Jiayun Koh, and Junxiong Pang. Wastewater surveillance to infer covid-19 transmission: A systematic review. *Sci. Total Environ.*, 804:150060, 2022.

[11] Jordan Peccia, Alessandro Zulli, Doug E. Brackney, Nathan D. Grubaugh, Edward H. Kaplan, Arnau Casanovas-Massana, Albert I. Ko, Amyn A. Malik, Dennis Wang, Mike Wang, Joshua L. Warren, Daniel M. Weinberger, Wyatt Arnold, and Saad B. Omer. Measurement of sars-cov-2 rna in wastewater tracks community infection dynamics. *Nat. Biotechnol.*, 38:1164–1167, 2020.

[12] Joshua I. Levy, Kristian G. Andersen, Rob Knight, and Smruthi Karthikeyan. Wastewater surveillance for public health. *Science*, 379 (6627):26–27, 2023.

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

[14] Elisabeth Mercier, Patrick M. D'Aoust, Thakali Ocean, Nada Hegazy, Jian-Jun Jia, Zhihao Zhang, Walaa Eid, Julio Plaza-Diaz, Md Pervez Kabir, Wanting Fang, Aaron Cowan, Sean E. Stephenson, Lakshmi Pisharody, Alex E. MacKenzie, Tyson E. Graber, Shen Wan,

and Robert Delatolla. Municipal and neighbourhood level wastewater surveillance and subtyping of an influenza virus outbreak. *Sci. Rep.*, 12:15777, 2022.

[15] William Chen and Kyle Bibby. Model-based theoretical evaluation of the feasibility of using wastewater-based epidemiology to monitor monkeypox. *Environ. Sci. Technol. Lett.*, 9 (9):772–778, 2022.

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

[17] Alexandria B. Boehm, Bridgette Hughes, Dorothea Duong, Vikram Chan-Herur, Anna Buchman, Marlene K. Wolfe, and Bradley J. White. 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*, 2023.

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

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

[20] Lyle R. Petersen, Aaron C. Brault, and Roger S. Nasci. West nile virus: Review of the literature. *JAMA*, 310 (3):308–315, 2013.

[21] Rebecca J. Eisen and Christopher D. Paddock. Tick and tickborne pathogen surveillance as a public health tool in the united states. *J. Med. Entomol.*, 58 (4):1490–1502, 2021.

[22] Meghan E. Hermance and Saravanan Thangamani. Powassan virus: An emerging arbovirus of public health concern in north america. *Vector Borne Zoonotic Dis.*, 17 (7):453–462, 2017.

[23] Michael A. Pfaller. Molecular approaches to diagnosing and managing infectious diseases: Practicality and costs. *Emerging Infect. Dis.*, 7 (2):312–318, 2001.

[24] Gonzalo M. Vazquez-Prokopec, Luis F. Chaves, Scott A. Ritchie, Joe Davis, and Uriel Kitron. Unforeseen costs of cutting mosquito surveillance budgets. *PLOS Negl. Trop. Dis.*, 4 (10):e858, 2010.

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

[26] Rose S. Kantor, Hannah D. Greenwald, Lauren C. Kennedy, Adrian Hinkle, Sasha Harris-Lovett, Matthew Metzger, Melissa M. Thornton, Justin M. Paluba, and Kara L. Nelson. Operationalizing a routine wastewater monitoring laboratory for sars-cov-2. *PLOS Water*, 1 (2):e0000007, 2022.

[27] Lucky G. Ngwira, Bhawana Sharma, Kabita Bade Shrestha, Sushil Dahal, Reshma Tuladhar, Gerald Manthalu, Ben Chilima, Allone Ganizani, Jonathan Rigby, Oscar Kanjerwa,

Kayla Barnes, Catherine Anscombe, Joseph Mfutso-Bengo, Nicholas Feasey, and Mercy Mvundura. Cost of wastewater-based environmental surveillance for sars-cov-2: Evidence from pilot sites in blantyre, malawi and kathmandu, nepal. *PLOS Glob. Public Health*, 2 (12):e0001377, 2022.

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

[29] Marta Gwinn, Duncan R. MacCannell, and Rima F. Khabbaz. Integrating advanced molecular technologies into public health. *J. Clin. Microbiol.*, 55 (3):703–714, 2017.

[30] Amy E. Kirby, Maroya Spalding Walters, Wiley C. Jennings, Rebecca Fugitt, Nathan LaCross, Mia Mattioli, Zachary A. Marsh, Virginia A. Roberts, Jeffrey W. Mercante, Jonathan Yoder, and Vincent R. Hill. Using wastewater surveillance data to support the covid-19 response—united states, 2020–2021. *Morb. Mortal. Wkly. Rep.*, 70 (36):1242–1244, 2021.

[31] Weidong Gu, Thomas R. Unnasch, Charles R. Katholi, Richard Lampman, and Robert J. Novak. Fundamental issues in mosquito surveillance for arboviral transmission. *Trans. R. Soc. Trop. Med. Hyg.*, 102 (8):817–822, 2008.

[32] P. N. Thompson and E. Etter. Epidemiological surveillance methods for vector-borne diseases. *Rev. Sci. Tech. Off. Int. Epiz.*, 34 (1):235–247, 2015.

[33] Florence Fournet, Frederic Jourdain, Emmanuel Bonnet, Stephanie Degroote, and Valery Ridde. 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, 2018.

[34] Warish Ahmed, Aaron Bivins, Paul M. Bertsch, Kyle Bibby, Phil M. Choi, Kata Farkas, Pradip Gyawali, Kerry A. Hamilton, Eiji Haramoto, Masaaki Kitajima, Stuart L. Simpson, Sarmila Tandukar, Kevin V. Thomas, and Jochen F. Mueller. 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, 2020.

[35] I. Michael-Kordatou, P. Karaolia, and D. Fatta-Kassinos. 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, 2020.

[36] Aparna Keshaviah, Xindi C. Hu, and Marisa Henry. Developing a flexible national wastewater surveillance system for covid-19 and beyond. *Environ. Health Perspect.*, 129 (4):045002, 2021.

[37] Jakob Zinsstag, Jurg Utzinger, Nicole Probst-Hensch, Lv Shan, and Xiao-Nong Zhou. Towards integrated surveillance-response systems for the prevention of future pandemics. *Infect. Dis. Poverty*, 9:140, 2020.

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

[39] Milton C. Weinstein, George Torrance, and Alistair McGuire. Qalys: The basics. *Value Health*, 12 (1):S5–S9, 2009.

[40] Caitlin Pley, Megan Evans, Rachel Lowe, Hugh Montgomery, and Sophie Yacoub. Digital and technological innovation in vector-borne disease surveillance to predict, detect, and control climate-driven outbreaks. *Lancet Planet. Health*, 5 (10):e739–e745, 2021.

[41] Hailay Desta Teklehaimanot, Joel Schwartz, Awash Teklehaimanot, and Marc Lipsitch. Alert threshold algorithms and malaria epidemic detection. *Emerg. Infect. Dis.*, 10 (7):1220–1226, 2004.

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

[43] Corey M. Peak, Lauren M. Childs, Yonatan H. Grad, and Caroline O. Buckee. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl. Acad. Sci. U.S.A.*, 114 (15):4023–4028, 2017.

[44] Nicholas B. DeFelice, Eliza Little, Scott R. Campbell, and Jeffrey Shaman. Ensemble forecast of human west nile virus cases and mosquito infection rates. *Nat. Commun.*, 8:14592, 2017.

Supplementary Information: Optimal environmental testing frequency for outbreak surveillance

Jason W. Olejarz, Kirstin I. Oliveira Roster, Stephen M. Kissler, Marc Lipsitch, Yonatan H. Grad

This Supplementary Information is organized as follows: In Section 1, we describe the surveillance protocol under consideration, and we derive the surveillance cost per unit time. In Section 2, we define the process by which new pathogens emerge, we define the dynamics of a pathogen that is growing in abundance, we define the manner in which each pathogen is detected, and we calculate the expected infection cost per unit time. In Section 3, we calculate the expected total cost per unit time, and we determine the optimal testing frequency. In Section 4, we describe how our model generalizes to account for the emergence of pathogens with different characteristics.

1 Surveillance cost

An important consideration for implementing environmental surveillance for pathogens is the frequency at which tests are performed. Environmental sampling and testing should be done frequently enough that an emerging pathogen is intercepted quickly, but not so frequently that surveillance costs outweigh the benefits of early detection. Here, we assume that whenever the environment is sampled and a test is conducted, a surveillance cost equal to c_1 is incurred. We also assume that surveillance costs are additive, so that if n tests are performed, then the total surveillance cost is equal to nc_1.

We consider that environmental tests are performed with period T. Since the cost of a single test is equal to c_1, and since the time between tests is equal to T, the surveillance cost per unit time is given by

$$C_1 = \frac{c_1}{T} \tag{1}$$

2 Expected infection cost

The costs incurred from the surveillance program itself must be considered in the context of infection-related costs. The costs due to an outbreak can vary depending on several factors:

- When the pathogen first appears in relation to the first environmental test that is performed following its introduction

- How the pathogen grows after it is introduced
- The sensitivity of the environmental testing program for detecting the pathogen
- The per-case infection cost

In this section, we describe each of these points in detail. Considering the full stochastic dynamics of pathogen initiation, pathogen growth, and pathogen detection, we derive a solution for the expected size of an outbreak when it is detected. We then derive an approximation for the expected size of an outbreak by assuming deterministic growth of a pathogen after it is initiated.

2.1 Emergence of a pathogen

For determining the optimal testing frequency, we require knowledge of how new pathogens are introduced. We assume that the introduction of new pathogens follows a Poisson process. New pathogens are initiated independently and continuously in time at rate λ.

2.2 Growth of a pathogen

We also require knowledge of how a pathogen increases in abundance once it first appears. Here, we assume that each instance of the pathogen makes new instances of the pathogen at rate r according to a Poisson process. Let $x_{m,n}(t)$ denote the probability that there are n copies of the pathogen at time t, given that there are m copies of the pathogen at time 0. In this section, we present the steps for calculating $x_{m,n}(t)$, beginning with the simplest cases and then progressing to the solution for any values of m and n, where $m \leq n$.

2.2.1 $m = 1$, $n = 1$

Suppose we start with a single instance of the pathogen at time 0 ($m = 1$). $x_{1,1}(t)$ gives the probability that the original instance of the pathogen has not produced any new instances of the pathogen up to time t ($n = 1$). $x_{1,1}(t)$ is given by

$$x_{1,1}(t) = e^{-rt}$$

2.2.2 $m = 1$, $n = 2$

Next, consider $x_{1,2}(t)$, which is the probability that the original instance of the pathogen has produced a single new instance of the pathogen by time t ($n = 2$). For this to occur, three things must happen: The original instance of the pathogen does not make any new instances of the pathogen between times 0 and t_1, the original instance of the pathogen makes a new copy of itself at time t_1, and neither of the two resulting instances of the pathogen make any new instances of the pathogen between times t_1 and t. We must integrate over all values of t_1 between 0 and t:

$$x_{1,2}(t) = \int_{t_1=0}^{t} e^{-rt_1}(r\,dt_1)e^{-2r(t-t_1)}$$

Simplifying, we have

$$x_{1,2}(t) = \left(\int_{t_1=0}^{t} e^{rt_1} (r\, dt_1) \right) e^{-2rt}$$

Performing the integration, we get

$$x_{1,2}(t) = \left(e^{rt} - 1 \right) e^{-2rt}$$

This then becomes

$$x_{1,2}(t) = \left(1 - e^{-rt} \right) e^{-rt}$$

2.2.3 $m = 1$, $n = 3$

Next, consider $x_{1,3}(t)$, which is the probability that the original instance of the pathogen has led to two new instances of the pathogen by time t ($n = 3$). For this to occur, five things must happen: The original instance of the pathogen does not make any new instances of the pathogen between times 0 and t_2, the original instance of the pathogen makes a new copy of itself at time t_2, neither of the two resulting instances of the pathogen make any new instances of the pathogen between times t_2 and t_1, one of the two instances of the pathogen makes a new copy of itself at time t_1, and none of the three resulting instances of the pathogen make any new instances of the pathogen between times t_1 and t. We must integrate over all values of t_2 between 0 and t_1, and we must integrate over all values of t_1 between 0 and t:

$$x_{1,3}(t) = \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} e^{-rt_2} (r\, dt_2) e^{-2r(t_1 - t_2)} (2r\, dt_1) e^{-3r(t - t_1)}$$

We can extend the range of the integration over t_2 from $t_2 = 0$ to $t_2 = t$ if we also divide by 2:

$$x_{1,3}(t) = \frac{1}{2} \int_{t_1=0}^{t} \int_{t_2=0}^{t} e^{-rt_2} (r\, dt_2) e^{-2r(t_1 - t_2)} (2r\, dt_1) e^{-3r(t - t_1)}$$

Simplifying, we have

$$x_{1,3}(t) = \left(\int_{t_2=0}^{t} e^{rt_2} (r\, dt_2) \right) \left(\int_{t_1=0}^{t} e^{rt_1} (r\, dt_1) \right) e^{-3rt}$$

Performing the integration, we get

$$x_{1,3}(t) = \left(e^{rt} - 1 \right)^2 e^{-3rt}$$

This then becomes

$$x_{1,3}(t) = \left(1 - e^{-rt} \right)^2 e^{-rt}$$

2.2.4 $m = 1$, $n = 4$

Next, consider $x_{1,4}(t)$, which is the probability that the original instance of the pathogen has led to three new instances of the pathogen by time t ($n = 4$). For this to occur, seven things must happen: The original instance of the pathogen does not make any new instances

of the pathogen between times 0 and t_3, the original instance of the pathogen makes a new copy of itself at time t_3, neither of the two resulting instances of the pathogen make any new instances of the pathogen between times t_3 and t_2, one of the two instances of the pathogen makes a new copy of itself at time t_2, none of the three resulting instances of the pathogen make any new instances of the pathogen between times t_2 and t_1, one of the three instances of the pathogen makes a new copy of itself at time t_1, and none of the four resulting instances of the pathogen make any new instances of the pathogen between times t_1 and t. We must integrate over all values of t_3 between 0 and t_2, we must integrate over all values of t_2 between 0 and t_1, and we must integrate over all values of t_1 between 0 and t:

$$x_{1,4}(t) = \int_{t_1=0}^{t} \int_{t_2=0}^{t_1} \int_{t_3=0}^{t_2} e^{-rt_3}(r\,dt_3)e^{-2r(t_2-t_3)}(2r\,dt_2)e^{-3r(t_1-t_2)}(3r\,dt_1)e^{-4r(t-t_1)}$$

We can extend the range of the integration over t_3 from $t_3 = 0$ to $t_3 = t$ and the range of the integration over t_2 from $t_2 = 0$ to $t_2 = t$ if we also divide by 3!:

$$x_{1,4}(t) = \frac{1}{3!}\int_{t_1=0}^{t} \int_{t_2=0}^{t} \int_{t_3=0}^{t} e^{-rt_3}(r\,dt_3)e^{-2r(t_2-t_3)}(2r\,dt_2)e^{-3r(t_1-t_2)}(3r\,dt_1)e^{-4r(t-t_1)}$$

Simplifying, we have

$$x_{1,4}(t) = \left(\int_{t_3=0}^{t} e^{rt_3}(r\,dt_3)\right)\left(\int_{t_2=0}^{t} e^{rt_2}(r\,dt_2)\right)\left(\int_{t_1=0}^{t} e^{rt_1}(r\,dt_1)\right)e^{-4rt}$$

Performing the integration, we get

$$x_{1,4}(t) = \left(e^{rt}-1\right)^3 e^{-4rt}$$

This then becomes

$$x_{1,4}(t) = \left(1-e^{-rt}\right)^3 e^{-rt}$$

2.2.5 $m = 1$, any n

We can generalize the calculation to arbitrary values of n:

$$x_{1,n}(t_0) = \left(\prod_{j=1}^{n-1}\int_{t_j=0}^{t_{j-1}} e^{-(n-j+1)r(t_{j-1}-t_j)}(n-j)r\,dt_j\right)e^{-rt_{n-1}}$$

Changing the integration limits, we have

$$x_{1,n}(t_0) = \frac{1}{(n-1)!}\left(\prod_{j=1}^{n-1}\int_{t_j=0}^{t_0} e^{-(n-j+1)r(t_{j-1}-t_j)}(n-j)r\,dt_j\right)e^{-rt_{n-1}}$$

This simplifies to

$$x_{1,n}(t_0) = \left(\prod_{j=1}^{n-1}\int_{t_j=0}^{t_0} e^{-(n-j+1)r(t_{j-1}-t_j)}r\,dt_j\right)e^{-rt_{n-1}}$$

This becomes

$$x_{1,n}(t_0) = \left(\int_{t=0}^{t_0} e^{rt} r \, dt \right)^{n-1} e^{-nrt_0}$$

Performing the integration, we get

$$x_{1,n}(t) = \left(1 - e^{-rt}\right)^{n-1} e^{-rt}$$

2.2.6 Any m and n

Following the same procedure, we can calculate $x_{m,n}(t)$:

$$x_{m,n}(t_0) = \left(\prod_{j=m}^{n-1} \int_{t_j=0}^{t_{j-1}} e^{-(n-j+m)r(t_{j-m}-t_{j-m+1})} (n-j+m-1) r \, dt_{j-m+1} \right) e^{-mrt_{n-m}}$$

Changing the integration limits, we have

$$x_{m,n}(t_0) = \frac{1}{(n-m)!} \left(\prod_{j=m}^{n-1} \int_{t_j=0}^{t_0} e^{-(n-j+m)r(t_{j-m}-t_{j-m+1})} (n-j+m-1) r \, dt_{j-m+1} \right) e^{-mrt_{n-m}}$$

This simplifies further:

$$x_{m,n}(t_0) = \frac{(n-1)!}{(n-m)!(m-1)!} \left(\prod_{j=m}^{n-1} \int_{t_j=0}^{t_0} e^{-(n-j+m)r(t_{j-m}-t_{j-m+1})} r \, dt_{j-m+1} \right) e^{-mrt_{n-m}}$$

We can rewrite this as

$$x_{m,n}(t_0) = \binom{n-1}{m-1} \left(\prod_{j=m}^{n-1} \int_{t_j=0}^{t_0} e^{-(n-j+m)r(t_{j-m}-t_{j-m+1})} r \, dt_{j-m+1} \right) e^{-mrt_{n-m}}$$

This becomes

$$x_{m,n}(t_0) = \binom{n-1}{m-1} \left(\int_{t=0}^{t_0} e^{rt} r \, dt \right)^{n-m} e^{-mrt_0}$$

Performing the integration, we get

$$x_{m,n}(t) = \binom{n-1}{m-1} \left(1 - e^{-rt}\right)^{n-m} e^{-mrt} \tag{2}$$

2.3 Detection of a pathogen

We further require an understanding of how the outbreak is detected. Consider that there are n instances of the pathogen within a particular lineage when the environment is tested. We assume that each instance of the pathogen is not detected independently with probability q. The outbreak is not detected if and only if no instance of the pathogen is detected, which occurs with probability q^n. Therefore, the pathogen is detected with probability $1 - q^n$.

We further assume that each lineage of the pathogen is detected independently of any other lineage. For example, suppose that two lineages of the pathogen are simultaneously

present. Suppose that when the environment is tested, Lineage A contains n_A copies of the pathogen, and Lineage B contains n_B copies of the pathogen. In this case, Lineage A is detected with probability $1 - q^{n_A}$, and Lineage B is detected with probability $1 - q^{n_B}$. (If the rate of introduction of new pathogens, λ, is small, then simultaneous presence of two lineages would be a rare occurrence. Nonetheless, we describe this possibility so that the stochastic dynamics of pathogen initiation, pathogen growth, and pathogen detection are completely specified.)

2.4 Expected size of an outbreak when it is detected

Using the stochastic rules presented above, and using Equation (2), we can derive a formula for the expected size of an outbreak when the pathogen is detected. For understanding the steps of the calculation, we define $X_i(a_i)$ to be the probability that there are i testing events following the appearance of the pathogen that fail to detect the pathogen, and that there are a_i infections when the pathogen is detected.

We first consider the following question: What is the probability that the pathogen is detected in the first test following its appearance and that there are a_0 instances of the pathogen when it is detected. This probability, which we denote $X_0(a_0)$, is given by

$$X_0(a_0) = \int_0^T \left(\frac{\mathrm{d}\tau}{T}\right) x_{1,a_0}(\tau)(1 - q^{a_0})$$

There are three components to this calculation:

- The pathogen is initiated at time τ before the testing event that detects it occurs. If the pathogen emerges just before the test that detects it is performed, then τ is slightly greater than 0. If the pathogen emerges just after the previous test, then τ is slightly less that T. Therefore, we have $0 \leq \tau < T$. Since new lineages appear independently and continuously in time, τ is equiprobably distributed between 0 and T, hence the integration $\int_0^T \mathrm{d}\tau/T$.
- The pathogen begins as a single infection, and it grows to a_0 infections at time τ since its appearance with probability $x_{1,a_0}(\tau)$.
- At least one of the a_0 infections is detected with probability $1 - q^{a_0}$.

Next, we can ask: What is the probability that the pathogen is detected in the second test following its appearance and that there are a_1 instances of the pathogen when it is detected. This probability, which we denote $X_1(a_1)$, is given by

$$X_1(a_1) = \int_0^T \frac{\mathrm{d}\tau}{T} \sum_{a_0=1}^{a_1} x_{1,a_0}(\tau) q^{a_0} x_{a_0,a_1}(T)(1 - q^{a_1})$$

This calculation is understood as follows: The first test occurs at time τ after the pathogen appears, the pathogen grows to a_0 infections at time τ after its emergence, none of those a_0 infections are detected in the first test, the pathogen then grows to a_1 infections at time

$\tau + T$ after its emergence, and at least one of those a_1 infections is detected in the second test. (We must sum over all values of a_0 between 1 and a_1.)

We can further ask: What is the probability that the pathogen is detected in the third test following its appearance and that there are a_2 instances of the pathogen when it is detected. This probability, which we denote $X_2(a_2)$, is given by

$$X_2(a_2) = \int_0^T \frac{\mathrm{d}\tau}{T} \sum_{a_1=1}^{a_2} \sum_{a_0=1}^{a_1} x_{1,a_0}(\tau) q^{a_0} x_{a_0,a_1}(T) q^{a_1} x_{a_1,a_2}(T)(1-q^{a_2})$$

This calculation is understood as follows: The first test occurs at time τ after the pathogen appears, the pathogen grows to a_0 infections at time τ after its emergence, none of those a_0 infections are detected in the first test, the pathogen then grows to a_1 infections at time $\tau + T$ after its emergence, none of those a_1 infections are detected in the second test, the pathogen then grows to a_2 infections at time $\tau + 2T$ after its emergence, and at least one of those a_2 infections is detected in the third test. (We must sum over all values of a_1 between 1 and a_2 and over all values of a_0 between 1 and a_1.)

The calculation of $X_3(a_3)$ follows in the same manner:

$$X_3(a_3) = \int_0^T \frac{\mathrm{d}\tau}{T} \sum_{a_2=1}^{a_3} \sum_{a_1=1}^{a_2} \sum_{a_0=1}^{a_1} x_{1,a_0}(\tau) q^{a_0} x_{a_0,a_1}(T) q^{a_1} x_{a_1,a_2}(T) q^{a_2} x_{a_2,a_3}(T)(1-q^{a_3})$$

To calculate the expected size of an outbreak, we sum $X_m(a_m)a_m$ over all possible numbers of failed tests ($0 \leq m < \infty$) and over all possible sizes of the outbreak when the pathogen is detected ($1 \leq a_m < \infty$):

$$\langle n \rangle = \sum_{m=0}^{\infty} \sum_{a_m=1}^{\infty} X_m(a_m) a_m \tag{3}$$

Equation (3) can be alternatively written as follows:

$$\langle n \rangle = \int_0^T \frac{\mathrm{d}\tau}{T} \sum_{a_0=1}^{\infty} x_{1,a_0}(\tau) \sum_{m=0}^{\infty} \left(\sum \prod_{\substack{1 \leq j \leq m \\ a_{j-1} \leq a_j}} q^{a_{j-1}} x_{a_{j-1},a_j}(T) \right) (1-q^{a_m}) a_m \tag{4}$$

2.5 Approximation for $\langle n \rangle$

Equation (4) is analytically unwieldy. To make progress, we derive an approximate solution for the expected size of an outbreak by assuming that the pathogen grows deterministically after it is initiated (Figure S1). The first several steps of this process are as follows:

- The first infection occurs, and at time τ after its emergence, a test is performed. The size of the outbreak when the first test is performed is equal to $e^{r\tau}$.
- If the pathogen is not detected in the first test, which occurs with probability $q^{e^{r\tau}}$, then the pathogen grows until the second test is performed, and the amount of growth of the pathogen between the first and second tests is equal to $e^{r(\tau+T)} - e^{r\tau}$.

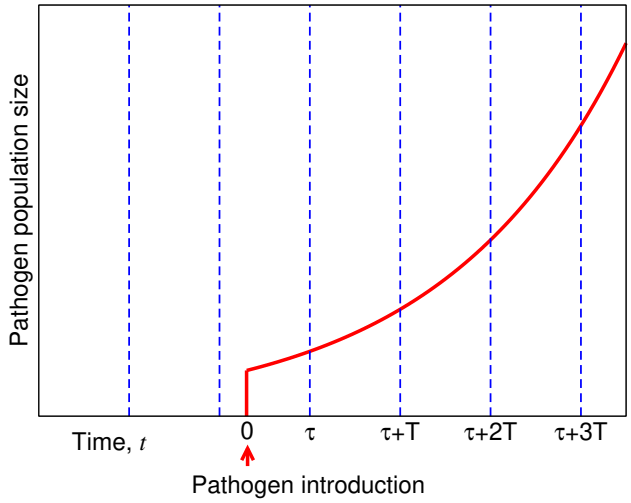

Figure S1: **Schematic showing deterministic growth of the pathogen.** For calculating an approximation for the expected size of an outbreak when it is detected, we can assume that the size of the outbreak grows deterministically.

- If the pathogen is not detected in the first test and the second test, which occurs with probability $q^{e^{r\tau}} q^{e^{r(\tau+T)}}$, then the pathogen grows until the third test is performed, and the amount of growth of the pathogen between the second and third tests is equal to $e^{r(\tau+2T)} - e^{r(\tau+T)}$.

- If the pathogen is not detected in the first test, the second test, and the third test, which occurs with probability $q^{e^{r\tau}} q^{e^{r(\tau+T)}} q^{e^{r(\tau+2T)}}$, then the pathogen grows until the fourth test is performed, and the amount of growth of the pathogen between the third and fourth tests is equal to $e^{r(\tau+3T)} - e^{r(\tau+2T)}$.

This process continues until the pathogen is detected. We therefore have the following result for the expected size of an outbreak:

$$\begin{aligned}\langle n \rangle = \int_0^T \frac{\mathrm{d}\tau}{T} \Big[& e^{r\tau} \\ & + q^{e^{r\tau}} \left(e^{r(\tau+T)} - e^{r\tau} \right) \\ & + q^{e^{r\tau}} q^{e^{r(\tau+T)}} \left(e^{r(\tau+2T)} - e^{r(\tau+T)} \right) \\ & + q^{e^{r\tau}} q^{e^{r(\tau+T)}} q^{e^{r(\tau+2T)}} \left(e^{r(\tau+3T)} - e^{r(\tau+2T)} \right) \\ & + \cdots \Big]\end{aligned}$$

More compactly:

$$\langle n \rangle = \int_0^T \frac{\mathrm{d}\tau \, e^{r\tau}}{T} \left[1 + \left(e^{rT} - 1 \right) \sum_{k=0}^{\infty} e^{krT} q^{(k+1)e^{r\tau} + e^{krT} \sum_{x=0}^{k-1} e^{-xrT}} \right] \tag{5}$$

The process can also be considered by defining

$$p \equiv 1 - q \tag{6}$$

Here, p is the probability that a single infection is detected in a testing event, so that the probability that an outbreak of size n is detected in a testing event is given by $1-(1-p)^n$. Substituting Equation (6) into Equation (5), we have

$$\langle n\rangle = \int_0^T \frac{\mathrm{d}\tau\, e^{r\tau}}{T}\left[1+\left(e^{rT}-1\right)\sum_{k=0}^{\infty} e^{krT}(1-p)^{(k+1)e^{r\tau}+e^{krT}\sum_{x=0}^{k-1}e^{-xrT}}\right] \tag{7}$$

Next, we simplify Equation (7) in two limits: for the case $p\to 1$ and for the case $p\to 0$. We then construct an approximation for $\langle n\rangle$ for any value of p.

2.5.1 $p\to 1$

In the limit $p\to 1$, Equation (7) simplifies:

$$\lim_{p\to 1}\langle n\rangle = \int_0^T \frac{\mathrm{d}\tau\, e^{r\tau}}{T}$$

Performing the integration, we obtain

$$\lim_{p\to 1}\langle n\rangle = \frac{e^{rT}-1}{rT} \tag{8}$$

2.5.2 $p\to 0$

We can rewrite Equation (7) as follows:

$$\langle n\rangle = \int_0^T \frac{\mathrm{d}\tau\, e^{r\tau}}{T} + \frac{e^{rT}-1}{T}\sum_{k=0}^{\infty}\left(\int_0^T \mathrm{d}\tau\, e^{r\tau}e^{\log(1-p)(k+1)e^{r\tau}}\right)e^{krT}(1-p)^{e^{krT}\sum_{x=0}^{k-1}e^{-xrT}}$$

Performing the integration, this becomes

$$\begin{aligned}\langle n\rangle = {} & \frac{e^{rT}-1}{rT} \\ & + \frac{e^{rT}-1}{rT\log(1-p)}\sum_{k=0}^{\infty}\left(\frac{e^{\log(1-p)(k+1)e^{rT}}-e^{\log(1-p)(k+1)}}{k+1}\right)e^{krT}(1-p)^{e^{krT}\sum_{x=0}^{k-1}e^{-xrT}}\end{aligned} \tag{9}$$

If p is small, then for an outbreak to be detected quickly, the testing period, T, must also be small. Considering that $p\ll 1$ and that $rT\ll 1$, the numerator of the expression in parentheses in Equation (9) can be approximated:

$$\begin{aligned}\langle n\rangle\big|_{p,rT\ll 1} \approx {} & \frac{e^{rT}-1}{rT} \\ & + \frac{e^{rT}-1}{rT\log(1-p)}\sum_{k=0}^{\infty}\left(\frac{[1+\log(1-p)(k+1)(1+rT)]-[1+\log(1-p)(k+1)]}{k+1}\right) \\ & \times e^{krT}(1-p)^{e^{krT}\sum_{x=0}^{k-1}e^{-xrT}}\end{aligned}$$

This becomes

$$\langle n\rangle\big|_{p,rT\ll 1} \approx \frac{e^{rT}-1}{rT} + \left(e^{rT}-1\right)\sum_{k=0}^{\infty} e^{krT} e^{\log(1-p)e^{krT}\sum_{x=0}^{k-1} e^{-xrT}}$$

Next, we approximate $\sum_{x=0}^{k-1} e^{-xrT}$ by $\left(1-e^{-rT}\right)^{-1}$, and we approximate the summation over k by an integration over k:

$$\langle n\rangle\big|_{p,rT\ll 1} \approx \frac{e^{rT}-1}{rT} + \left(e^{rT}-1\right)\int_0^{\infty} \mathrm{d}k\, e^{krT} e^{\log(1-p)e^{krT}\left(1-e^{-rT}\right)^{-1}}$$

Performing the integration, this becomes

$$\langle n\rangle\big|_{p,rT\ll 1} \approx \frac{e^{rT}-1}{rT} - \frac{\left(e^{rT}-1\right)\left(1-e^{-rT}\right)}{rT\log(1-p)}$$

Simplifying, we have

$$\langle n\rangle\big|_{p,rT\ll 1} \approx \frac{e^{rT}-1}{rT}\left(1-\frac{1-e^{-rT}}{\log(1-p)}\right) \tag{10}$$

2.5.3 $0 < p \leq 1$

Notice that in the limit $p \to 1$, Equation (10) becomes equivalent to Equation (8). Therefore, for any value $0 < p \leq 1$, we have the following approximate solution for $\langle n\rangle$:

$$\langle n\rangle \approx \frac{e^{rT}-1}{rT}\left(1-\frac{1-e^{-rT}}{\log(1-p)}\right) \tag{11}$$

3 Expected total cost per unit time

For optimizing the testing frequency, the quantity of interest is the expected total cost per unit time. The surveillance cost per unit time, C_1, is given by Equation (1). Let $\langle C_2\rangle$ denote an approximation for the expected infection cost per unit time, and let $\langle C\rangle$ denote an approximation for the expected total cost per unit time. We have

$$\langle C\rangle = C_1 + \langle C_2\rangle \tag{12}$$

For determining $\langle C_2\rangle$, we assume that each infection contributes a cost c_2. If new lineages appear at rate λ, then $\langle C_2\rangle$ is given by

$$\langle C_2\rangle = \lambda c_2 \langle n\rangle \tag{13}$$

Substituting Equations (1), (13), and (11) into Equation (12), we obtain

$$\langle C\rangle = \frac{c_1}{T} + \lambda c_2\left(\frac{e^{rT}-1}{rT}\right)\left(1-\frac{1-e^{-rT}}{\log(1-p)}\right) \tag{14}$$

3.1 Optimal testing frequency

Equation (14) specifies the expected total surveillance and pathogen cost. $\langle C \rangle$ is a function of the testing period, T, and we seek the value of T for which $\langle C \rangle$ is minimal. The first step is to show that $\langle C \rangle$ has a single minimum at a particular value of T. To do this, we differentiate $\langle C \rangle$ twice with respect to T:

$$\left(\frac{rT^2}{\lambda c_2}\right)\frac{\mathrm{d}\langle C\rangle}{\mathrm{d}T} = -\frac{rc_1}{\lambda c_2} + \left[(rT-1)e^{rT}+1\right]\left(1-\frac{2}{\log(1-p)}\right) - \frac{2\left[\sinh(rT) - rT\cosh(rT)\right]}{\log(1-p)} \tag{15}$$

$$\left(\frac{rT^3}{\lambda c_2}\right)\frac{\mathrm{d}^2\langle C\rangle}{\mathrm{d}T^2} = \frac{2rc_1}{\lambda c_2} + \left\{\left[(rT-1)^2+1\right]e^{rT}-2\right\} - \frac{\left[(rT-1)^2+1\right]e^{rT}-2}{\log(1-p)} - \frac{\left[(rT+1)^2+1\right]e^{-rT}-2}{\log(1-p)} \tag{16}$$

In Equation (16), the quantity $rT^3/(\lambda c_2)$ is necessarily positive. If the right-hand side of Equation (16) is positive for positive values of T, then $\mathrm{d}^2\langle C\rangle/\mathrm{d}T^2$ is necessarily positive. Note that

$$\lim_{T\to 0}\left[\left(\frac{rT^3}{\lambda c_2}\right)\frac{\mathrm{d}^2\langle C\rangle}{\mathrm{d}T^2}\right] = \frac{2rc_1}{\lambda c_2} > 0 \tag{17}$$

We also have

$$\frac{\mathrm{d}}{\mathrm{d}T}\left[\left(\frac{rT^3}{\lambda c_2}\right)\frac{\mathrm{d}^2\langle C\rangle}{\mathrm{d}T^2}\right] = e^{rT} - \frac{2\sinh(rT)}{\log(1-p)} > 0 \tag{18}$$

From Equations (17) and (18), it follows that the right-hand side of Equation (16) is necessarily positive. Therefore,

$$\frac{\mathrm{d}^2\langle C\rangle}{\mathrm{d}T^2} > 0 \tag{19}$$

Next, note that

$$\lim_{T\to 0}\frac{\mathrm{d}\langle C\rangle}{\mathrm{d}T} = -\infty \tag{20}$$

We also have

$$\lim_{T\to \infty}\frac{\mathrm{d}\langle C\rangle}{\mathrm{d}T} = \infty \tag{21}$$

From Equations (20), (21), and (19), it follows that there is a single value of T for which $\langle C \rangle$ is minimized.

To determine the optimal testing period, we set $\mathrm{d}\langle C\rangle/\mathrm{d}T = 0$ and $T = T^*$ in Equation (15). We arrive at an implicit solution for the optimal testing period, T^*:

$$\frac{rc_1}{\lambda c_2} = \left[(rT^*-1)e^{rT^*}+1\right]\left(1-\frac{2}{\log(1-p)}\right) - \frac{2\left[\sinh(rT^*) - rT^*\cosh(rT^*)\right]}{\log(1-p)} \tag{22}$$

The optimal testing frequency is given by

$$f^* = \frac{1}{T^*} \tag{23}$$

3.1.1 Asymptotic behavior as $p \to 1$

Taking the limit $p \to 1$ in Equation (22), we obtain the following equation for T^*:

$$\frac{rc_1}{\lambda c_2} \approx (rT^* - 1)e^{rT^*} + 1$$

Letting $W_0(x)$ denote the principal branch of the Lambert W function, and using Equation (23), we obtain an explicit approximation for the optimal testing frequency:

$$f^* \sim r\left\{1 + W_0\left(\frac{1}{e}\left[\frac{rc_1}{\lambda c_2} - 1\right]\right)\right\}^{-1} \quad (p \to 1) \tag{24}$$

In Figure S2, we plot Equation (24) as a function of c_2 for several sets of parameter values. We also plot measurements of the optimal testing frequency from simulating the true stochastic process.

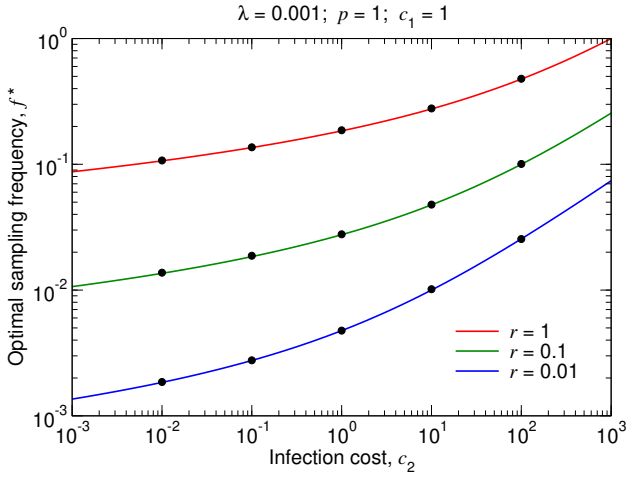

Figure S2: **Optimal testing frequency for $p = 1$.** For different values of r, we plot the optimal testing frequency, f^*, given by Equation (24), as a function of c_2. The black dots are measurements of the optimal testing frequency from simulating the true stochastic process. The 95% confidence intervals are smaller than the size of the data points.

3.1.2 Asymptotic behavior as $p \to 0$

For small values of p, the optimal testing frequency, T^*, is also small. To determine T^*, we consider that $p \ll 1$ and that $rT^* \ll 1$ in Equation (22). We use the approximations $\log(1-p) \approx -p$, $e^{rT^*} \approx 1 + rT^* + (rT^*)^2/2$, $\sinh(rT^*) \approx rT^* + (rT^*)^3/3!$, and $\cosh(rT^*) \approx 1 + (rT^*)^2/2$:

$$\frac{rc_1}{\lambda c_2} \approx \left[(rT^* - 1)\left(1 + rT^* + \frac{(rT^*)^2}{2}\right) + 1\right]\left(1 + \frac{2}{p}\right) + \frac{2}{p}\left[\left(rT^* + \frac{(rT^*)^3}{3!}\right) - rT^*\left(1 + \frac{(rT^*)^2}{2}\right)\right]$$

Simplifying, we have

$$\frac{rc_1}{\lambda c_2} \approx \frac{(rT^*)^2}{p}\left(1 - \frac{2rT^*}{3}\right)$$

For small values of rT^*, the second term on the right-hand side is negligible relative to the first. Using Equation (23), we solve approximately for the optimal testing frequency:

$$f^* \sim \sqrt{\frac{r\lambda c_2}{pc_1}} \quad (p \to 0) \tag{25}$$

In Figure S3, we plot f^* from Equations (22) and (23) as a function of p for several sets of parameter values. We also plot measurements of the optimal testing frequency from simulating the true stochastic process. For small values of p, f^* is approximately given by Equation (25).

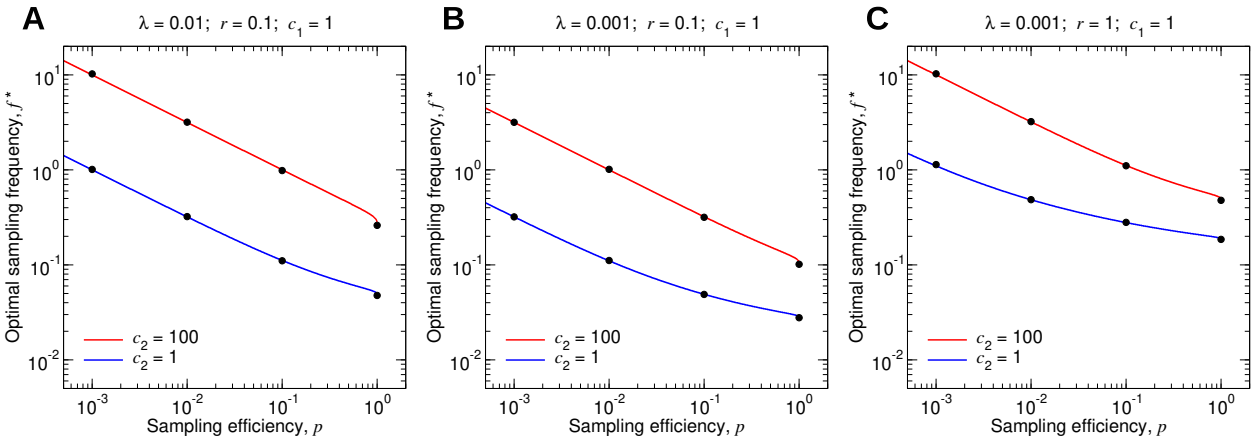

Figure S3: **Optimal testing frequency for $p \leq 1$.** For different values of c_2, we plot the optimal testing frequency, f^*, given by Equations (22) and (23), as a function of p. The black dots are measurements of the optimal testing frequency from simulating the true stochastic process. The 95% confidence intervals are smaller than the size of the data points.

4 Distribution of pathogen-related parameters

The calculation of the expected infection cost per unit time, $\langle C_2 \rangle$, assumes that, for each lineage that appears, the pathogen-specific parameters c_2, r, and p are the same. The expected infection cost per unit time is then just the expected cost due to a single lineage multiplied by the rate, λ, at which those lineages arise.

More generally, we can consider $\mathrm{d}c_2\,\mathrm{d}r\,\mathrm{d}p\,\lambda'(c_2, r, p)$ to be the (infinitesimal) rate at which lineages with pathogen-specific parameters c_2, r, and p appear. In this generalized model, let $\langle C_2' \rangle$ denote an approximation for the expected infection cost per unit time, and let $\langle C' \rangle$ denote an approximation for the expected total cost per unit time. We have

$$\langle C' \rangle = C_1 + \langle C_2' \rangle \tag{26}$$

With knowledge of the rate density function, $\lambda'(c_2, r, p)$, we are able to compute the expected infection cost per unit time by integrating over all possible values of c_2, r, and p:

$$\langle C_2' \rangle = \int_0^\infty \mathrm{d}c_2 \int_0^\infty \mathrm{d}r \int_0^1 \mathrm{d}p \left\{ \lambda'(c_2, r, p)\, c_2 \langle n \rangle \right\} \tag{27}$$

Substituting Equations (1) and (27) into Equation (26), we obtain

$$\langle C' \rangle = \frac{c_1}{T} + \int_0^\infty \mathrm{d}c_2 \int_0^\infty \mathrm{d}r \int_0^1 \mathrm{d}p \left\{ \lambda'(c_2, r, p) \left[c_2 \left(\frac{e^{rT} - 1}{rT} \right) \left(1 - \frac{1 - e^{-rT}}{\log(1-p)} \right) \right] \right\} \tag{28}$$

The optimal testing frequency is given by

$$F^* = \frac{1}{\arg\min_T \langle C' \rangle} \tag{29}$$

Equations (28) and (29) can be solved numerically to determine the optimal testing frequency. Below, we consider several simple examples for which Equation (28) can be solved analytically to show how the model works.

4.1 Example 1

As the simplest example of using Equation (28), consider that only a single type of pathogen can emerge. The pathogen has per-case cost c_2', growth rate r', and probability of detection p', and new lineages are introduced at rate λ. The rate density function, $\lambda'(c_2, r, p)$, is given by

$$\lambda'(c_2, r, p) = \lambda \delta(c_2 - c_2') \delta(r - r') \delta(p - p')$$

Here, δ denotes the Dirac delta function. When this form for $\lambda'(c_2, r, p)$ is substituted into Equation (28) and the integrations over c_2, r, and p are performed, we obtain

$$\langle C' \rangle = \frac{c_1}{T} + \lambda c_2' \left(\frac{e^{r'T} - 1}{r'T} \right) \left(1 - \frac{1 - e^{-r'T}}{\log(1-p')} \right)$$

Thus, Equation (28) reduces to Equation (14) for the case where only a single type of pathogen with fixed parameters can emerge.

4.2 Example 2

Next, consider the possibility that two different types of pathogens can emerge. Pathogen 1 has parameters c_2', r', and p', while Pathogen 2 has parameters c_2'', r'', and p''. Lineages of Pathogen 1 are introduced at rate λ_1, and lineages of Pathogen 2 are introduced at rate λ_2. The corresponding rate density function is

$$\lambda'(c_2, r, p) = \lambda_1 \delta(c_2 - c_2') \delta(r - r') \delta(p - p') + \lambda_2 \delta(c_2 - c_2'') \delta(r - r'') \delta(p - p'')$$

When this form for $\lambda'(c_2, r, p)$ is substituted into Equation (28) and the integrations are performed, we obtain

$$\langle C' \rangle = \frac{c_1}{T} + \lambda_1 c_2' \left(\frac{e^{r'T} - 1}{r'T} \right) \left(1 - \frac{1 - e^{-r'T}}{\log(1-p')} \right) + \lambda_2 c_2'' \left(\frac{e^{r''T} - 1}{r''T} \right) \left(1 - \frac{1 - e^{-r''T}}{\log(1-p'')} \right)$$

The expected total cost per unit time, $\langle C' \rangle$, is therefore equal to the surveillance cost per unit time, plus the expected infection cost per unit time for Pathogen 1, plus the expected infection cost per unit time for Pathogen 2.

4.3 Example 3

These considerations can be extended to the case where many different types of pathogens can emerge. Let Pathogen n have per-case cost $c_{2,n}$, growth rate r_n, and probability of detection p_n. The rate density function, $\lambda'(c_2, r, p)$ is given by

$$\lambda'(c_2, r, p) = \sum_n \lambda_n \delta(c_2 - c_{2,n}) \delta(r - r_n) \delta(p - p_n)$$

Substituting this into Equation (28) and integrating yields

$$\langle C' \rangle = \frac{c_1}{T} + \sum_n \lambda_n c_{2,n} \left(\frac{e^{r_n T} - 1}{r_n T} \right) \left(1 - \frac{1 - e^{-r_n T}}{\log(1 - p_n)} \right)$$

The expected infection cost is therefore linear—i.e., we add together the expected infection costs for each of the n possible types of pathogens, and this sum equals the total expected infection cost.

4.4 Example 4

The possible parameter values that any new pathogen can have are not discrete. They are continuous. To show how this works, consider the following form for the rate density function:

$$\lambda'(c_2, r, p) = \lambda \left[\left(2\sqrt{\frac{a}{\pi}} \right) e^{-a c_2^2} \right] \delta(r - r') \delta(p - p')$$

For this case, new pathogens have growth rate r' and probability of detection p'. New pathogens can, however, have any real value of c_2 that is nonnegative. For any lineage that is introduced, it's value of c_2 is most likely to be close to zero, while larger values of c_2 occur more rarely. The parameter a controls with with of the probability density function for c_2. For smaller values of a, this distribution has a longer tail, and the expected value of c_2 for any new pathogen increases. Substituting this form for the rate density function into Equation (28) and integrating, we have

$$\langle C' \rangle = \frac{c_1}{T} + \frac{\lambda}{\sqrt{\pi a}} \left(\frac{e^{r' T} - 1}{r' T} \right) \left(1 - \frac{1 - e^{-r' T}}{\log(1 - p')} \right)$$

4.5 Example 5

For this example, we suppose that any new pathogen has per-case infection cost c_2' and probability of detection p', while the growth rate, r, can be any nonnegative real number. We use the following form for the rate density function:

$$\lambda'(c_2, r, p) = \lambda \delta(c_2 - c_2') \left[b^2 r e^{-br} \right] \delta(p - p')$$

Substituting this into Equation (28) and integrating, we have

$$\langle C' \rangle = \frac{c_1}{T} + \frac{\lambda b^2 c_2'}{T} \left\{ \left(\frac{1}{b-T} - \frac{1}{b} \right) + \left[\left(\frac{1}{b-T} - \frac{1}{b} \right) - \left(\frac{1}{b} - \frac{1}{b+T} \right) \right] \left(\frac{-1}{\log(1-p')} \right) \right\}$$

4.6 Example 6

We can also model the case where new pathogens have per-case cost c_2' and growth rate r', while the probability of detection, p, can be any real number between 0 and 1. Suppose that the rate density function has the following form:

$$\lambda'(c_2, r, p) = \lambda \delta(c_2 - c_2') \delta(r - r') \left[\frac{[\theta(p-a) - 1] \log(1-p)}{(1-a)\log(1-a) + a} \right]$$

Here, θ denotes the Heaviside step function. Substituting this into Equation (28) and integrating, we obtain

$$\langle C' \rangle = \frac{c_1}{T} + \lambda c_2' \left(\frac{e^{r'T} - 1}{r'T} \right) \left(1 + \frac{a(1 - e^{-r'T})}{(1-a)\log(1-a) + a} \right)$$