Frequency and Sample Size to Minimize the Cost of a Rapidly Evolving Infectious Animal or Plant Disease

Jean-Philippe Terreaux

Agricultural & Rural Studies

Article

Frequency and Sample Size to Minimize the Cost of a Rapidly Evolving Infectious Animal or Plant Disease

Jean-Philippe Terreaux

Environment, Territories in Transition, Infrastructures, Societies (ETTIS) Research Unit, French National Institute for Agriculture, Food, and Environment (INRAE), 33612 Cestas, France; jean-philippe.terreaux@inrae.fr

Abstract: One of the major effects of global change is the spread of animal and plant diseases on farms. Besides the impact on the farms themselves, it is the whole rural world that is affected, through the possible disruption of value chains. Combating these diseases is therefore a crucial but costly problem. So, when faced with an infectious animal or plant pathology, how can we minimize the cost of the disease and of the sampling and analyses testing required to monitor its progress? First, we calculate the imprecision of the results as a function of the sample size and the prevalence of the disease. Then, depending on the desired precision and the prevalence of the disease, we calculate the required sample size. Finally, in the case of iterative sampling, depending on the cost of each sampling and testing event and the costs associated with the spread of the disease, we show on a quantitative example that there is an optimum, i.e. a relationship between the frequency and the sample size (number of samples) that allows the cost of the disease to be minimized. We show the optimum relationship between sample size and frequency, the relationship between minimum total cost and frequency, and finally, we show on a 3-dimensional graph, how the total cost evolves as a function of frequency and sample size.

Keywords: sampling; epidemic; epizootic; epiphytic; pathology; dynamics; economics; cost; risk; probability

Citation: Terreaux J. P. Frequency and Sample Size to Minimise the Cost of a Rapidly Evolving Infectious Animal or Plant Disease. *Agricultural & Rural Studies*, 2024, *2*, 0021. *https://doi.org/10.59978/ar02040021*

Received: 10 April 2024
Revised: 12 June 2024
Accepted: 24 June 2024
Published: 25 November 2024

1. Introduction

International communities are increasingly aware of the importance of both farm animal and crop health, as evidenced by the publications of the International Plant Protection Convention (see IPPC Secretariat, 2023) and of the World Organization for Animal Health (see World Organization for Animal Health, 2023). The subject of this paper is sampling to control the occurrence of rapidly spreading infectious animal or plant diseases: how many measurements should be taken, and how often, to minimize the cost of the disease plus the cost of the measurements? Each measure is costly, and too many would be prohibitive. On the other hand, if monitoring is too lax, there is a risk that the disease will develop and spread, with catastrophic consequences.

This multi-disciplinary work contributes to a whole range of studies and results, combining epidemiology, economics, and modeling, with the aim of not systematically seeking to eradicate pathologies, but to assess their costs in order to minimize them. In the same line of thought, Silal (2021) shows how multidisciplinary operational research can contribute to the efficient management of infectious diseases, with a particular emphasis on minimizing the costs of pathology detection. These studies include for example Han et al. (2020) on the bovine viral diarrhea virus for dairy and beef cattle herds.

The financial implications of our work are significant. To give just two examples, avian influenza, which mainly affects poultry farms, has cost the French government around 1.5 billion euros in 2022 alone (compensation for farmers, requisitions, euthanasia of animals, cleaning and disinfection...); not to mention the losses incurred by professionals in the processing industry. Anaplasmosis in cattle (see Railey & Marsh, 2021) raises the same kind of economic consequences and therefore induces the same sampling problems for early detection.

As far as plants are concerned, the estimated damage of Citrus "greening" disease (citrus Huanglongbing, or HLB) over the past 5 years before 2020 amounts in Florida alone, to over $1 billion per year, with nearly 5000 jobs lost annually (Li et al., 2020). In many countries, plum pox (or sharka) is a viral pathology affecting stone fruits. Surveillance and detection procedures are currently evolving in line with EU Regulation 2016/2031 (see Terreaux, 2023). It is therefore necessary to reorganize the monitoring of this disease. The continued production of these fruits (apricots, peaches, nectarines, etc.) in France is at stake. Other pathologies affecting cultivated plants

that are the subject of similar questions include Xylella fastidiosa (see Burbank, 2022). Many other animal and plant diseases raise the same issues, but in the remainder of this article, we will use avian influenza as an example for application and illustration.

In a previous article (Terreaux, 2022), we calculated the sample size (number of animals to be tested) required on a farm to know with 99% or 95% confidence whether or not it is infected or not. Here, we complement this approach by taking into account the fact that the disease can emerge on the farm at any time, e.g. following poor biosecurity and contamination by a human vector, from infected premises, or from wildlife. Actually, the biosecurity measures implemented may vary greatly from one farm to another, depending on the specifications, objectives, and challenges of each farmer (see Fountain et al., 2023). On the other hand, we do not simply want to know with any degree of accuracy (99% or 95%) whether the farm is infected. Our aim is to minimize the total cost of the disease, i.e. the cost of sampling and testing, plus the cost of culling infected flocks, plus the cost of allowing disease to spread that may be asymptomatic, particularly if only a few animals are infected and shedding virus (e.g. ducks can shed virus for five days before the first symptoms appear).

The methods used to calculate these costs are very different: firstly, the costs are uncertain because the disease will spread in a non-deterministic way. The decision criterion can then be, as a first approximation, the minimization of the mathematical expectation of the costs: thus, if for a given farm at a given time, the probability of disease occurrence is p, it is not assumed that a proportion p of the animals are systematically infected. The situation is dichotomous: either all the animals are disease-free, or some are infected, in which case the disease spreads throughout the farm. The prevalence (proportion of infected animals) is therefore generally zero, but sometimes it becomes strictly positive (following infection) and then increases. The prevalence is assumed to increase asymptomatically until the disease is detected by sampling, the parameters of which – sample size (number of animals tested) and frequency (or periodicity) of testing - must be carefully chosen. The animals are then euthanized. Alternatively, sampling is inadequate, and the disease remains undetected until the number of affected animals is sufficient (prevalence exceeds a certain threshold) for some of them to die, or for the feed consumption of the herd to drop significantly, etc., and the disease becomes symptomatic. The herd is then culled. But the problem with the latter situation is that the disease will have been able to spread for longer and more widely outside the farm under investigation, at a much higher collective cost (via other farms) than would have been the case if the disease had been detected early.

The prevalence is therefore likely to change over time. In a previous article (Terreaux, 2022), we calculated the minimum sample size (minimum number of animals to be tested) for a prevalence of 5%. In section 2, we calculate the accuracy obtained as a function of sample size and prevalence. In section 3, we calculate the number of tests to be performed as a function of the prevalence to achieve 99% or 95% accuracy.

In section 4, we explicitly introduce the dynamics of pathology in the herd and assume that sampling is iterative: For the same observation duration T, instead of testing M animals once, we repeatedly test N animals p times, with M = pN. Again, the aim is no longer to achieve a given accuracy of measurement but to minimize the overall cost of the disease.

Sections 4.1. and 4.2. describe the model and show the arbitrary values chosen for the different parameters. Section 4.3. shows that, as expected, the longer the duration T, the larger the sample size required. In section 4.4. we show how the total cost reaches a minimum for a given duration T (associated with a number of animals to be tested – or sample size – calculated in section 4.2.). Finally, in section 4.4., we show how the total cost evolves as a function of the periodicity T (= 1/ frequency) of the measurements and of the sample size.

The model set up in Section 4, a simplified representation of reality with a set of parameters chosen for illustrative purposes, represents the dynamics of the disease within the farm under study. The occurrence of the disease and the detection or non-detection of the disease in the farm, if it is affected, are randomized by two nested Monte Carlo processes.

2. Measurement Imprecision as a Function of Sample Size and Prevalence

In Terreaux (2022) we showed that for a prevalence of *prev*, the sample size N (number of animals to be tested) to have an accuracy of at least α (e.g. $\alpha = 95\%$), considering a number y of animals in the herd, is so that (see too Wonnacott & Wonnacott, 1990; Mann, 2010; Weiss, 2011):

$$\frac{\frac{(y.(1-prev))!}{N!(y.(1-prev)-N)!}}{\frac{(y)!}{N!(y-N)!}} \leq 1-\alpha \quad (1)$$

We will now assume that y = 8000 animals in the herd. We can calculate the precision of the measurement as a function of N (sample size). This is shown in Figure 1 for a prevalence of 5% and in Figure 2 if we vary this prevalence between 1% and 10%.

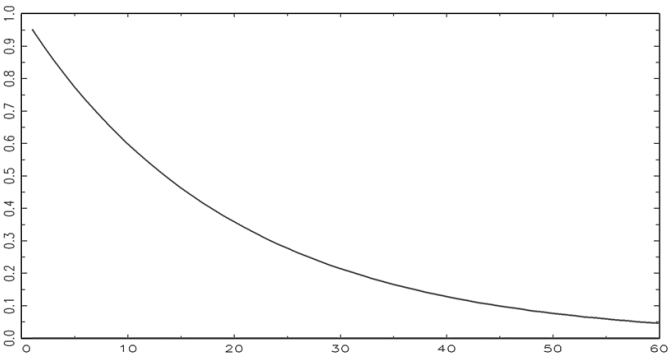

Figure 1. Measurement imprecision for 5% prevalence as a function of sample size. X-axis: sample size; Y-axis: imprecision (1- α) obtained.

Example: With 20 samples, the measurement imprecision is 36%; i.e. if the disease is present in the herd, there is a 36% risk of not detecting it.

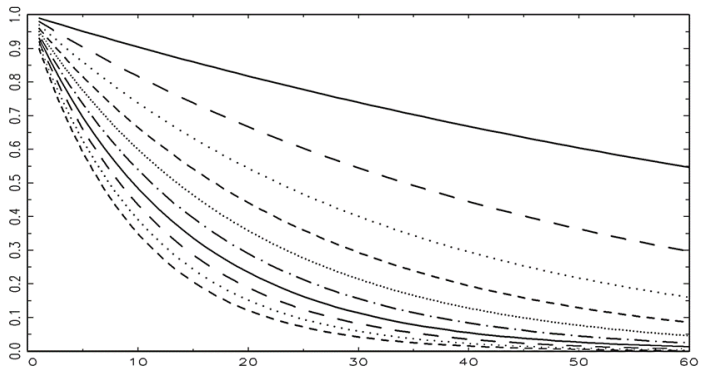

Figure 2. Measurement imprecision as a function of sample size for different prevalences. From top to bottom: prevalence of 1%, 2% … 10%.

Example: With 20 samples and a prevalence of 2%, the measurement imprecision is 67%; in other words, if the pathology is present in the herd, there is a 67% risk of not detecting it.

3. Sample Size as a Function of Prevalence

Using the same formula, we can calculate the number of samples needed to achieve 99% or 95% accuracy, depending on the prevalence. This is shown in Figure 3. We still assume that y = 8000.

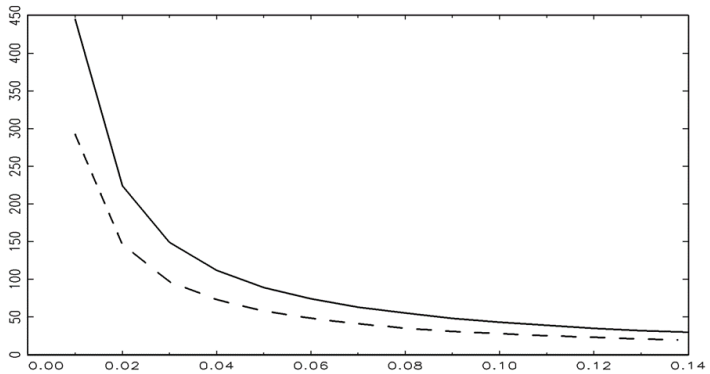

Figure 3. Sample size required to achieve a given accuracy: upper curve: 99%, lower curve: 95%. X-axis: prevalence; Y-axis: sample size.

Example: With a prevalence of 5%, 89 samples are required for 99% accuracy and 58 samples for 95% accuracy.

4. Iterative Sampling

Given that contamination of the farm with the disease can occur at any time, it seems interesting not to determine precisely whether or not this contamination has occurred at a given time t and, because of the prohibitive cost involved, not to repeat this measurement soon afterward, but to carry out periodic tests, albeit with a smaller sample size. The aim is therefore no longer to ensure the absence of the disease, but to minimize costs, both in terms of sampling and testing costs, and in terms of the costs associated with the spread of the disease (by preventing the prevalence from becoming too high, or the disease from becoming symptomatic). For example, for a large herd, instead of testing 60 individuals at once (see Terreaux, 2022: these 60 are sufficient to know whether a herd of size 8000 individuals, as in the numerical example above, or smaller, is affected by the pathology with 95% accuracy when the prevalence is 5%), we can repeatedly test, every T time steps, N individuals, with $n < 60$, N and T still to be calculated.

4.1. Iterative Sampling to Reduce Costs

Figure 4 shows the situation considered: sampling of N individuals every T time steps (here in days). The dotted arrow represents the time of onset of the disease. From this point on, the number of affected animals and their prevalence increase exponentially. This corresponds to a standard representation of the onset of the evolution of an infectious disease: a SIR model (see Murray, 2002; Terreaux, 2017) without R, i.e. without remission for some individuals.

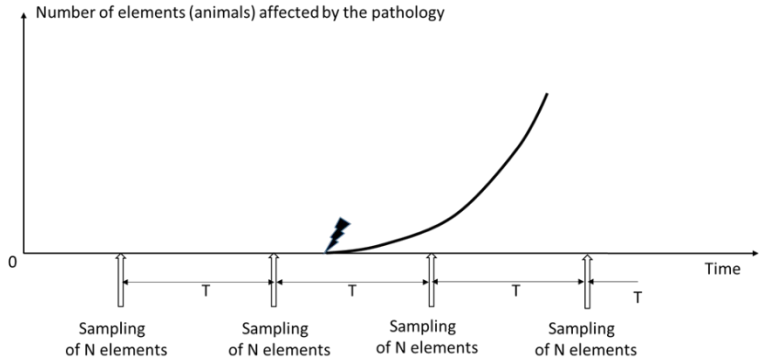

Figure 4. Schematic diagram showing the onset of the disease and the various sampling events separated by T.

We then apply a Monte Carlo procedure (see, for example, Fishman, 1995): starting from the initial time ($t = 0$), we simulate an initial trajectory over a horizon H: at each date t, the disease can appear on the farm with probability p, or else the herd remains healthy. From its onset at time t, it evolves exponentially with a coefficient δ (at each time step, i.e. for example every 24 hours, the number of infected individuals is multiplied by δ). This automatically increases the prevalence and therefore the probability of detecting the disease at the next sampling. If the disease is detected, the herd is culled at a cost of C1. The barn is then left empty for a quarantine period Q before a new herd is established. However, a sample size of only N animals, a low prevalence, and bad luck may mean that the disease is present but goes undetected. It will then continue to develop at the rate dictated by δ. If the prevalence exceeds Pmax, the disease becomes symptomatic and the herd is culled; the cost is C2, which is higher than C1 because the disease has spread in the meantime. This is followed by a quarantine period of the same duration before a new herd is established.

In total, this trajectory generates different costs over the time horizon considered: the cost of sampling and possibly one or more C1 costs and one or more C2 costs. Adding these together gives the total cost of this trajectory. By repeating the generation of such trajectories a large number of times (in practice 100,000 times) over a time horizon H, we deduce the average cost of these trajectories, which is nothing other than the mathematical expectation of the cost as a function of the numerical values chosen for each of the parameters.

This method therefore involves two intertwined random elements: the onset of the disease and whether it is detected or not. The two main parameters we adjust here are N, the size of each sample, and T, the time between two sampling events. The other parameters depend on the type of problem we are dealing with. We have not carried out a precise econometric study of the value of these parameters, so the results presented here are of qualitative interest only.

Figure 7 uses 2100 (i.e. 30 × 70) parameters sets, with the possibility of the disease occurring over 100 time steps. Therefore, 21 billion (30 × 70 × 100 × 100,000) draws of pseudo-random numbers are required to simulate the possible onset of the disease. A problem related to the recycling of these numbers could arise: the number generator used is the one presented in Terreaux (2000), which does not present this risk.

4.2. The Various Parameters of the Model

The parameters considered here, together with the numerical values adopted, are presented in Table 1. It should be remembered that the values are arbitrary and must be adapted for their quantitative application to a specific situation.

Table 1. Parameter values for numerical simulations.

Parameter	Symbol	Numerical value
herd size	y	8.000
cost if disease is detected by testing	C1	30.000 €
cost if disease is detected by symptoms	C2	300.000 €
cost of sampling and testing one animal		20 €
"entry cost" of sampling (see text)		150 €
time step		1 day
calculation horizon	H	100 days
probability of disease occurrence at each time step	p	0.0005
prevalence leading to symptomatic detection	P_{max}	40 %
pathology evolution coefficient	δ	1.6
duration of quarantine	Q	20 days
number of individuals per sample	N	variable to be optimized
time between two sampling events	T	variable to be optimized

The cost of a sampling event is defined by its "entry cost" (i.e. the fixed cost whatever the sample size N) plus the sample size N multiplied by the "cost of sampling and testing an animal."

4.3. Sampling: Optimal Size as a Function of the Number of Days Between Two Sampling Events

We have two variables, N and T, whose values we can choose, and which will determine the total cost (sampling, testing, culling, and dissemination to other farms) of controlling the disease. Our objective is:

$$\min_{N,T}\left(Overallcost\right) \tag{2}$$

If T is fixed, this leads to a value of N that allows this minimum to be achieved. We show N as a function of T in Figure 5.

On this graph, the slight decrease observed when T = 9 is not significant; it is due to the still low number of trajectories generated, which is still 100,000 for each set of parameters N and T. Increasing this number would eliminate this artifact and make the surface shown in Figure 7 "smoother."

Note that the minimum cost for a period T greater than 7 days corresponds to a sample size greater than 60, i.e. that obtained with a prevalence assumption of 5% and a desired accuracy of 95% for a single sampling event (see Terreaux, 2022).

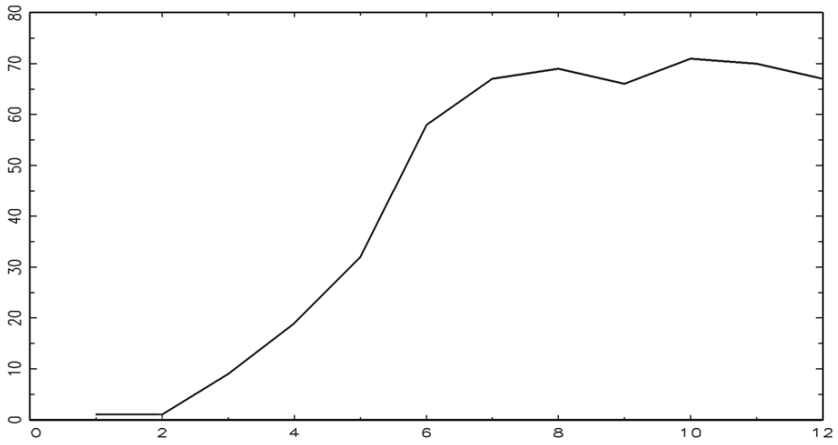

Figure 5. N (sample size, y-axis) as a function of T (sampling periodicity in days, x-axis) to minimize total cost.

Example: To minimize the total cost (cost of testing + cost of euthanasia if positive + impact of spreading the disease if not detected in time), if we sample every 4 days (x-axis = 4), the number of animals to be tested (sample size) is 19 (y-axis = 19).

Another example: If T = 5, then N = 32; beyond 11 days, the optimum for the chosen parameter values is around 70.

4.4. Minimum Cost as a Function of Sampling Periodicity

We now show the evolution of this minimum cost (i.e. by adjusting N, the sample size, as much as possible) as a function of sampling periodicity. Figure 6 shows that, beyond T = 3, the total cost increases with the sampling periodicity. The minimum cost is obtained for T = 3 and corresponds (see Figure 5) to a sample size of N = 9.

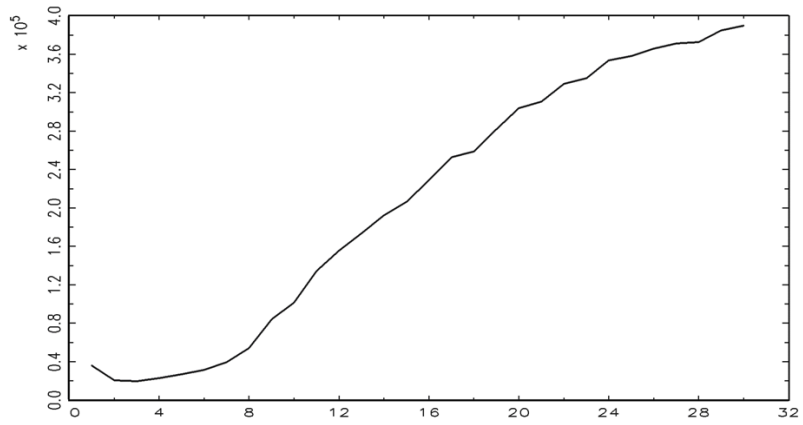

Figure 6. Evolution of total cost (y-axis, in €) as a function of sampling periodicity T (in days).

Beyond three days, the higher the sampling periodicity T, the higher the cost.

The minimum corresponds to sampling every 3 days, which, according to Figure 5, corresponds to 9 animals tested every 3 days with these data.

4.5. Cost as a Function of Periodicity and Sample Size

The evolution of the total cost as a function of N and T is shown below in a three-dimensional perspective graph.

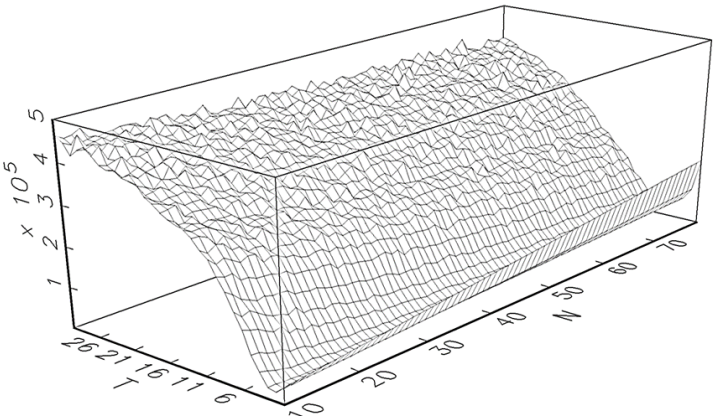

Figure 7. Evolution of total cost (z-axis, in €) as a function of sampling periodicity (T, in days) and sample size (N).

Note that the z-axis scale starts from zero: an error in the numerical values assigned to N or T can be costly, potentially multiplying the total cost of the disease by much more than 5.

5. Conclusion

In practical terms, the results of this research show how it is possible to significantly reduce the costs associated with pathology by replacing a single sampling to test for its presence on the farm with successive samplings of smaller size: We have shown here that instead of carrying out a single sampling ($T \to \infty$), or a small number of samplings (T large), it may be more interesting to carry out regular sampling events of smaller size (fewer animals or plants tested each time). In this case, a trade-off between sampling periodicity and sample size has to be made. Optimal values depend on the estimation of the different parameters, and therefore on the animal or plant disease under investigation and in particular the economic conditions and the stakes of the agricultural production in question, the fixed and variable costs of sampling, the probability of the pathology appearing on the farm and, if present, its dynamics. Monte Carlo methods have proved their worth here, making it possible to calculate numerically and illustrate graphically the economic benefits of choosing the right sampling parameters.

The scientific breakthrough lies in the fact that, in the sampling problem addressed here, we take into account both the costs and benefits associated with earlier detection of the pathology and the fact that sampling is not carried out once and for all to find out whether the disease is present on the farm, but is repeated periodically over time. Its characteristics—sampling frequency and size—are determined by a multidisciplinary approach (economics, epidemiology, probability calculation, Monte Carlo modeling). Further work could take into account the fact that the interest of each individual farmer is not the same as that of the farmers as a whole, nor that of the processing and marketing chain, nor that of the State (see Terreaux, 2017, on a similar issue in beekeeping, or Terreaux, 2023, on plumpox virus for some fruit orchards). In certain cases, this could make it possible to replace regulatory constraints with incentive instruments, in everyone's interest.

Moreover, following Giral-Barajas et al. (2023), our model could be extended to multistage epidemiological dynamics, when, for some diseases, it is possible to distinguish different clinical stages. Another development of our work on sampling could be to take into account the possibility of vaccinating, or at least reducing the incidence of the pathology on, for example, part of the herds or orchards susceptible to the disease; this would make our results more precise when this possibility is real (see the extension of epidemiological models in this regard in Ramponi & Tessitore, 2024).

Another line of research would be to extend our work with an economic objective to farms made up of different herds, orchards, or more generally different subsets when the prevalence of pathology differs from one subset to another (see an example of such a situation in Clement et al., 2023). Still, another area of research could be to combine the costs studied here with those of biosecurity measures, bearing in mind that these measures may be taken by the farmer in his own interest, with an externality effect on the spread of the pathology to other farmers (see Hennessy & Rault, 2023). Finally, coming back to sampling, it would be useful to be able to take into account the possibility, when it occurs, of false positives and false negatives when testing individuals for the presence of the pathology (see Vasiliauskaite et al., 2021).

CRediT Author Statement: This is a single author paper and the author was solely responsible for the content, including the concept, design, analysis, writing, and revision of the manuscript.

Data Availability Statement: All data used here are given in the text.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Acknowledgments: I would especially like to thank Dr. Iker Vaquero Alba for his invaluable help in preparing this manuscript. I would like to express my thanks to the two anonymous reviewers whose suggestions and comments helped me to greatly improve the quality of the article. Of course, any errors remain my own.

References

Burbank, L. P. (2022). Threat of Xylella fastidiosa and options for mitigation in infected plants. *CABI Reviews*. https://doi.org/10.1079/cabireviews202217021

Clement, M. J., Justice-Allen, A., & Heale, J. D. (2023). Optimal risk-based allocation of disease surveillance effort for clustered disease outbreaks. *Preventive Veterinary Medicine*, *212*, 105830. https://doi.org/10.1016/j.prevetmed.2022.105830

Fishman, G. S. (1995). *Monte Carlo – Concepts, algorithms and applications* (3rd ed.). Springer Science & Business Media.

Fountain, J., Hernandez-Jover, M., Manyweathers, J., Hayes, L., & Brookes, V. J. (2023). The right strategy for you: Using the preferences of beef farmers to guide biosecurity recommendations for on-farm management of endemic disease. *Preventive Veterinary Medicine*, *210*. https://doi.org/10.1016/j.prevetmed.2022.105813

Giral-Barajas, J., Herrera-Nolasco, C. I., Herrera-Valdez, M. A., & López, S. I. (2023). A probabilistic approach for the study of epidemiological dynamics of infectious diseases: Basic model and properties. *Journal of Theoretical Biology*, *572*, 111576. https://doi.org/10.1016/j.jtbi.2023.111576

Han, J. H., Weston, J. F., Heuer, C., & Gates, M. C. (2020). Modelling the economics of bovine viral diarrhoea virus control in pastoral dairy and beef cattle herds. *Preventive Veterinary Medicine*, *182*, 105092. https://doi.org/10.1016/j.prevetmed.2020.105092

Hennessy, D. A., & Rault, A. (2023). On systematically insufficient biosecurity actions and policies to manage infectious animal disease. *Ecological Economics*, *206*, 107740. https://doi.org/10.1016/j.ecolecon.2023.107740

IPPC Secretariat. (2023). *2022 IPPC Annual Report – Protecting the world's plant resources from pests*. Food and Agriculture of the United Nations. https://openknowledge.fao.org/handle/20.500.14283/cc4922en

Li, S., Wu, F., Duan, Y., Singerman, A., & Guan, Z. (2020). Citrus greening: Management strategies and their economic impact. *HortScience*, *55*(5), 604–612. https://doi.org/10.21273/HORTSCI14696-19

Mann, P. S. (2010). *Introductory statistics* (7th ed.). John Wiley & Sons. https://bcs.wiley.com/he-bcs/Books?action=index&itemId=0470444665&bcsId=5345

Murray, J. D. (Ed.). (2002). *Mathematical biology: I. An introduction*. Springer New York. https://doi.org/10.1007/b98868

Railey, A. F., & Marsh, T. L. (2021). Economic benefits of diagnostic testing in livestock: Anaplasmosis in cattle. *Frontiers in Veterinary Science*, *8*, 626420. https://doi.org/10.3389/fvets.2021.626420

Ramponi, A., & Tessitore, M. E. (2024). Optimal social and vaccination control in the SVIR epidemic model. *Mathematics*, *12*(7), 933. https://doi.org/10.3390/math12070933

Silal, S. P. (2021). Operational research: A multidisciplinary approach for the management of infectious disease in a global context. *European Journal of Operational Research*, *291*(3), 929–934. https://doi.org/10.1016/j.ejor.2020.07.037

Terreaux, J. P. (2000). Estimation de la rentabilité de la culture de certains eucalyptus dans le sud-ouest de la France [Estimating the profitability of eucalyptus cultivation in south-west France]. *Annals of Forest Science*, *57*(4), 389–397. https://doi.org/10.1051/forest:2000129

Terreaux, J. P. (2017). Epizooties et efficacité des processus de décision : Un exemple en apiculture [Epizootics and the efficiency of decision-making processes: An example from beekeeping]. *Revue Française d'Economie*, *32*(2), 160–197. https://doi.org/10.3917/rfe.172.0160

Terreaux, J. P. (2022). Animal or plant pathologies: Number of samples to be collected to know with a given precision if a farm is positive. *Journal of Research in Applied Mathematics*, *8*(5), 4–10. https://www.researchgate.net/publication/360561952

Terreaux, J. P. (2023). On the possible impacts of a regulatory change on resource conservation: The case of the plumpox virus in France. *Natural Resource Conservation and Research*, *6*(2). https://doi.org/10.24294/nrcr.v6i2.2986

Vasiliauskaite, V., Antulov-Fantulin, N., & Helbing, D. (2021). On some fundamental challenges in monitoring epidemics. *Philosophical Transactions of the Royal Society A*, *380*(2214). https://doi.org/10.1098/rsta.2021.0117

Weiss, N. A. (2011). *Introductory statistics* (9th ed.). Addison Wesley.

Wonnacott, T. H., & Wonnacott, R. J. (1990). *Introductory Statistics for Business and Economics*. John Wiley et Sons.

World Bank Group. (2010). *People, pathogens and our planet: Volume 1: Toward a one health approach for controlling zoonotic diseases*. https://documents1.worldbank.org/curated/en/214701468338937565/pdf/508330ESW0whit1410B01PUBLIC1PPP1Web.pdf

World Organization for Animal Health. (2023). *Director general's report on WOAH activities: Administrative working document*. https://www.woah.org/app/uploads/2024/04/91gs-2024-wd-adm-05-dg-activities-report-en.pdf