Sampling for disease absence—deriving informed monitoring from epidemic traits

Yoann Bourhis[1,*], Timothy R. Gottwald[2], Francisco J. Lopez-Ruiz[3], Sujin Patarapuwadol[4], Frank van den Bosch[1]

1. Rothamsted Research, Department of Biointeraction and Crop Protection, UK
2. US Department of Agriculture, Agricultural Research Service, Florida, USA
3. Centre for Crop and Disease Management, Curtin University, Australia
4. Department of Plant pathology, Kasetsart University, Thailand

* yoann.bourhis@rothamsted.ac.uk

Abstract

Monitoring for disease requires subsets of the host population to be sampled and tested for the pathogen. If all the samples return healthy, what are the chances the disease was present but missed? In this paper, we developed a statistical approach to solve this problem considering the fundamental property of infectious diseases: their growing incidence in the host population. The model gives an estimate of the incidence probability density as a function of the sampling effort, and can be reversed to derive adequate monitoring patterns ensuring a given maximum incidence in the population. We then present an approximation of this model, providing a simple rule of thumb for practitioners. The approximation is shown to be accurate for a sample size larger than 20, and we demonstrate its use by applying it to three plant pathogens: citrus canker, bacterial blight and grey mould.

Keywords Disease absence, Risk assessment, Early detection, Sampling theory
MSC 2010: 62D05, 92D30

Introduction

When it comes to disease management, surveillance programs have two different objectives: establishing disease absence in host populations, or ensuring an early detection of any disease outbreak (Parnell et al., 2017). Early detection is essential to disease control mitigation, timely reactions generally being more successful and less detrimental for the host population (Cunniffe et al., 2016). For example, Carpenter et al. (2011) showed for foot-and-mouth disease, that when delaying the detection from 7 to 22 days after the initial infection, the containment measures required the culling 30 times more host animals. Likewise, surveillance programs are operated to establish the absence or presence of emerging strains of endemic pathogens, hence enabling trade certifications for instance. Examples of these are emerging strains of plant pathogens that are insensitive to the fungicides applied to control them, or strains that are virulent[1] in a crop cultivar by having resistance breaking genes.

[1]We use here the plant pathology definition where virulence signifies the ability of the pathogen to infect the host. In human and other animal pathology virulence is used as a measure of damage the pathogen does to the host.

Monitoring a disease requires the assessment of the pathological status of sampling units. Assessments generally occur at the level of the host individual (*e.g.* for ash dieback, Woodward & Boa, 2013), but sometimes for convenience the sampling unit is a subpopulation like a farm (*e.g.* for foot-and-mouth disease, Keeling et al., 2001), or a field (*e.g.* for bacterial blight in rice, Koyshibayev & Muminjanov, 2016). In any case, when the samples all return negative, it is still very important to account for the chance that the pathogen was present but undetected (Cannon, 2002). Therefore, declaring disease absence is then a probabilistic evaluation, more samples making it less likely the pathogen was missed.

The incidence[2] of a pathogen, noted q hereafter, is the proportion of the host population infected. Estimating the incidence from a sample where all assessments return negative for the pathogen can be defined as a zero-numerator problem, *i.e.* estimating the probability of an event from data in which it has not occurred yet (Hanley & Lippman-Hand, 1983; Winkler et al., 2002). We thus want to calculate the probability density, $p(q|notfound)$, of the incidence q given that none of the sampling units is assessed as infected. This can be done by deriving $p(notfound|q)$ from the exponential distribution, and then reversing it according to Bayes' rule, assuming a uniform prior $p(q)$. A practically useful quantity is the incidence q_X for which

$$\int_0^{q_X} p(q|notfound)dq = \frac{X}{100}, \tag{1}$$

thus giving the upper bound of the $X\%$ confidence interval of q. This upper limit gives the highest, still likely, incidence given a sampling effort. The rule of three is a very common rule of thumb to estimate the upper limit of the 95% confidence interval (Louis, 1981). For example, considering we have a random sample of size N, all returning negative, the upper limit can be approximated by $q_{95} = \frac{3}{N+1}$ (Hanley & Lippman-Hand, 1983). This very practical rule of thumb can be used to identify a sampling effort N that can ensure that infection is below a threshold value.

To ensure pathogen absence from an area over an extended time interval, the host population have to be sampled repeatedly. Incidence estimation should then account for the change of incidence between the rounds caused by the epidemic dynamics. In this regard, Metz et al. (1983) accounted for the time dependence of samples due to the epidemic dynamics when they assessed the level of epidemic risk associated with a given sampling effort. However, when it comes to the incidence estimation problem, the epidemic temporal dynamics is neglected while the focus is more likely set on the spatial dependences of the samples due to the epidemic spread (Cameron & Baldock, 1998; Cannon, 2002; Coulston et al., 2008).

Accounting for the epidemic dynamics, we address the incidence estimation problem in the case of *disease absence* sampling (as Parnell et al., 2012, did with *first discovery*). We present a model estimating the pathogen incidence in a population, being given a sampling effort and an epidemic growth rate. We then derive an approximation of this model (in the way of Alonso Chavez et al., 2016) providing a practical and simple way to derive information from a negative sampling. This epidemically informed approximation proves itself accurate and flexible enough to account for the asymptomatic period of the disease. Finally, we apply this model to three case examples: citrus canker in an orange orchard, the invasion of virulent pathogen strains of bacterial blight of rice and the invasion of fungicide resistant pathogen strains in grey mould of grape.

[2]We use here the plant pathology definition where incidence is the fraction of host units infected. In human and other animal pathology this is termed prevalence.

Materials and methods

A monitoring program typically consists of batches of N samples randomly or regularly distributed in space, and regularly iterated over K monitoring rounds with time intervals of Δ time units. Parnell et al. (2012) has shown how to use this particular structure to derive the pathogen incidence when first detection occurs. In this section we use a similar method to estimate the incidence when no infected sample has returned.

One monitoring round

Considering an incidence q in a population, the probability for a sample to return negative is given by $p(notfound|q) = 1 - q$. A sample of size N will therefore return entirely negative with probability:

$$p(notfound|q) = (1-q)^N. \tag{2}$$

Now, given a monitoring round returned negative, what does it tell us about q ? We can derive $p(q|notfound)$, the incidence given no detection, from $p(notfound|q)$ using the Bayes' theorem. Bayes' theorem relates those probabilities by:

$$p(q|notfound) = \frac{p(q)p(notfound|q)}{\int_0^1 p(q)p(notfound|q)dq}. \tag{3}$$

The value $p(q)$ is the prior probability density of q. Assuming no information on q, we set $p(q)$ to be a uniform and uninformative prior. Substituting Eq. 2 into Eq. 3 then gives

$$p(q|notfound) = (N+1)(1-q)^N. \tag{4}$$

As mentioned in the introduction, of particular interest is the upper limit of the $X\%$ confidence interval given by Eq. 1 which, for the exponential distribution given by Eq. 4, gives

$$q_X = \frac{-ln(1-X/100)}{N+1}. \tag{5}$$

Put in words, if we set a maximum incidence q_X below which we are satisfied to consider the host population to be free of disease, we can derive a sampling effort N, that will ensure $X\%$ of the undetected diseases to have incidences smaller than q_X.

Two monitoring rounds

Having two monitoring rounds can be seen as increasing the size of the sample:

$$p(notfound|q) = (1-q)^{N_1}(1-q)^{N_2} = (1-q)^{N_1+N_2} \tag{6}$$

where N_2 is the size of the current monitoring round and N_1 is the size of the previous one. In this equation, the sizes of the historic and recent monitoring rounds are powers of the same probability of negative sampling: $1-q$. However, as mentioned in the introduction, the incidence of an infectious pathogen increases through time. Therefore, non-detection in the last monitoring round occurred over a larger incidence than the previous one.

As we are focused on absence sampling, we are interested in epidemics with low incidences, so $q \ll 1$. It is well established that at low incidence epidemics grow exponentially (van der Plank, 1963; Faria et al., 2014; Bartlett et al., 2016). We thus assume $q(t) = q(0)e^{rt}$ where t is the time and r is the epidemic growth rate. In the time

interval Δ between two monitoring rounds the pathogen incidence has grown by a factor $\lambda = e^{r\Delta}$, or another words, the incidence in monitoring round i was a factor λ^{-1} smaller than the incidence in monitoring round $i+1$.

For our two-round case we thus have

$$p(notfound|q) = (1-q)^{N_2}(1-\lambda^{-1}q)^{N_1}. \tag{7}$$

K monitoring rounds

Building on this epidemic model, we can now generalise Eq. 7 to K monitoring rounds:

$$P(notfound|q) = \prod_{k=1}^{K}\left(1-\lambda^{-k+1}q\right)^{N}. \tag{8}$$

We can use the Bayes' theorem to compute the probability density of the incidence given non-detection after K monitoring rounds as:

$$P(q|notfound) = \frac{\prod_{k=1}^{K}\left(1-\lambda^{-k+1}q\right)^{N}}{\int_0^1 \prod_{k=1}^{K}\left(1-\lambda^{-k+1}q\right)^{N} dq}. \tag{9}$$

An approximation

Computing the probability density given by Eq. 9, as well as the subsequent upper limit q_X, requires integrations which need to be approximated numerically. It is computationally inexpensive but still requires a computer program to be used. Here we develop an approximation which makes the computation of $p(q|notfound)$ simple enough to be useful for practitioners. It gives a rule of thumb in planning a monitoring program for a given disease.

The approximation is built on the following two assumptions: (1) the sampling size N is large enough ($N > 10$), and (2) the incidence q is small. Both assumption are realistic as 10 is a relatively small sampling size, and as we are interested only in cases with very low incidence. Using our assumption that $q \ll 1$, we can approximate $(1-q)^N$ by e^{-qN}. Substituting this in Eq. (3) results in:

$$p(q|notfound) \approx \frac{N}{1-e^{-N}}e^{-qN}. \tag{10}$$

And following our assumption that $N > 10$, this equation can be approximated by

$$p(q|notfound) \approx Ne^{-qN}, \tag{11}$$

Plugging Eq. (11) into Eq. (1) results in

$$\tilde{q}_X = \frac{-ln(1-X/100)}{N}. \tag{12}$$

Similarly, for two monitoring rounds we find

$$p(q|notfound) = (N_1 + \lambda^{-1}N_2)e^{-q(N_1+\lambda^{-1}N_2)} \tag{13}$$

and

$$\tilde{q}_X = \frac{-ln(1 - X/100)}{(N_1 + \lambda^{-1}N_2)}. \quad (14)$$

If now we generalise to K monitoring rounds, it results in

$$p(q|notfound) = Ae^{-qA} \quad (15)$$

$$\text{and} \quad \tilde{q}_X = \frac{-ln(1 - X/100)}{A} \quad (16)$$

where A is given by:

$$A = \sum_{k=1}^{K} \lambda_{-k+1} N = N \frac{\lambda - \lambda^{-K+1}}{\lambda - 1} \quad (17)$$

Before going further using the approximation, its accuracy needs to be evaluated. First, the approximated and the exact probability densities $p(q|notfound)$ are compared visually. And then, we investigate more carefully the difference between their respective confidence intervals q_X and $\tilde{q}_X$. The comparison is likely to depend on the sampling effort as well as on the epidemic growth rate, therefore the accuracy is evaluated for wide ranges of the relevant parameters.

Results

The exact model

Figures 1 clearly shows the effect of an epidemic increase (*i.e.* $\lambda > 1$) as compared to a situation, as previously published (see *e.g.* Cameron & Baldock, 1998; Cannon, 2002, for absence sampling), where incidence q is assumed constant over time (*i.e.* $\lambda = 1$). Figures 1 exposes that using the classical rule of 3 for a monitoring program extended in time would result in significant underestimations of q_{95}, as it would for any confidence level. The severity of these underestimations increases with the epidemic growth rate and the time interval between rounds. As expected, the upper bound q_X of the confidence interval for q decreases with increasing sample size N and increasing number of sampling rounds K. The faster the epidemic grows, the larger λ, and the larger q_X, which is also to be expected. What is less obvious but interesting to note is that if we compare monitoring programs with the same sampling effort $N \cdot K$ (lower left versus upper right panels in Figure 1), we see that q_X is lower for monitoring programs that are shorter in time (smaller K). This finding is consistent for other parameter values. However in reality we do require a monitoring program to extend over long period of time to ensure pathogen absence for the entire period.

The impact of λ on the incidence can be decomposed to investigate the impact of the growth rate r and the time interval between rounds Δ. Since they are defined by $\lambda = e^{r\Delta}$, they have the same impact on disease incidence, which is illustrated by the diagonal symmetry in Figure 2. This picture focus only on q_X instead of the whole probability density. Figure 2 also delineates a plateau for large values of λ, above which a faster epidemic growth, or a larger monitoring time interval, does not significantly increase the incidence of the undetected pathogen. This is also visible in Figure 1 where the probability densities for $\lambda = 10$ and $\lambda = 100$ are very similar, despite the order of magnitude change in λ. Finally, Figure 2 illustrates the greater impact of the sample size N than the number of rounds K on the epidemic risk.

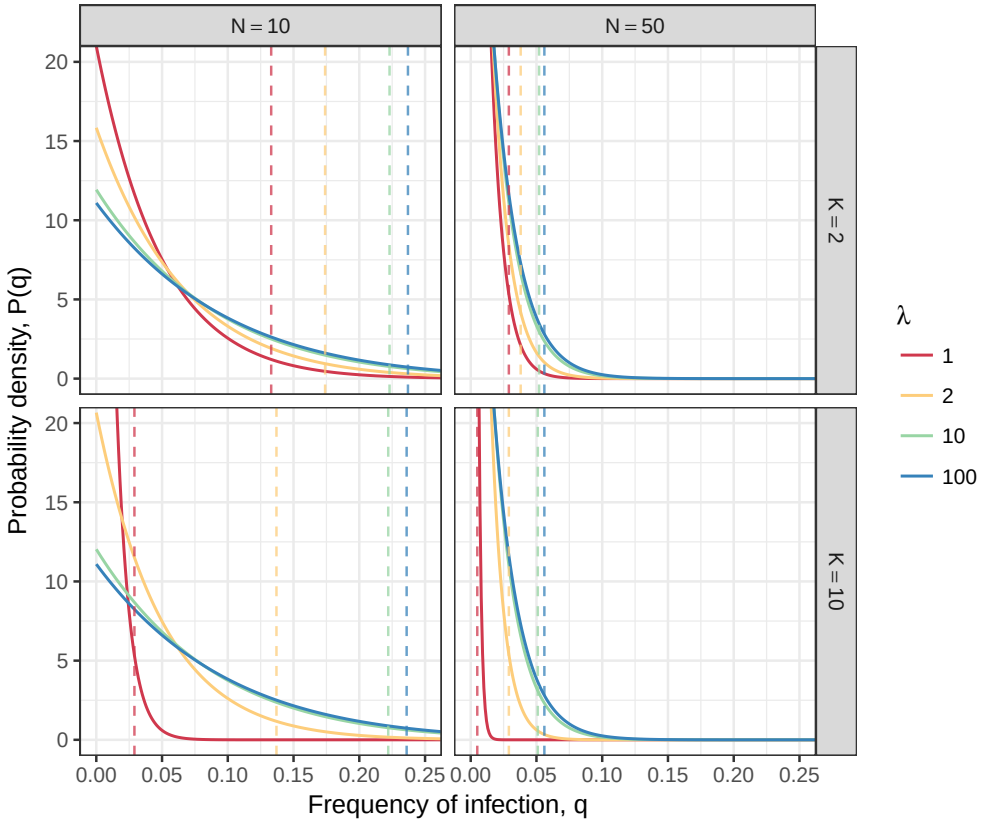

Figure 1. Probability density of the incidence given by Eq. 9 depending on the sampling effort (size N and repetition K) and on $\lambda = e^{r\Delta}$, the factor by witch the epidemic grows between two sampling rounds. The dashed lines are the upper bounds q_X the 95% confidence interval.

Accuracy of the approximation

In developing an approximation, our aims are twofold: (1) provide an equation featuring the model behaviours described in the previous subsection, and (2) provide an equation simple enough that it can be solved "on the back of an envelop" when designing a monitoring program. Figure 3 compares, for the case with two monitoring rounds ($K = 2$), the exact and approximated probability densities (respectively given by Eqs. 9 and 15). We see that the exact and approximated density curves are barely distinguishable whatever the sampling size and epidemic speed. However, if we take a closer look at our index of interest q_X, we see that low values of N cause significant inaccuracy (figured by the shaded areas). This illustrates why the approximation does not hold for $N < 10$. Figure 3 also shows a tendency towards inaccuracy when the epidemic growth or the monitoring intervals increase.

The effect of K (the number of sampling rounds) is better visualised if we focus on the relative error between the approximated and exact upper bounds $\frac{|\tilde{q}_X - q_X|}{q_X}$. Figure 4 confirms the trends previously observed: the accuracy increases with the sampling size and decreases with the epidemic growth rate and time interval between samples. We see that past $N \approx 20$, good levels of accuracy of the approximation is achieved, even for large epidemic growth rates. It also seems that going from $K = 20$ to $K = 100$ does not improve the accuracy and we reach a plateau. Finally, the accuracy increases with the likeliness of the event under study (here from 0.1% in bottom row to 5% for top row),

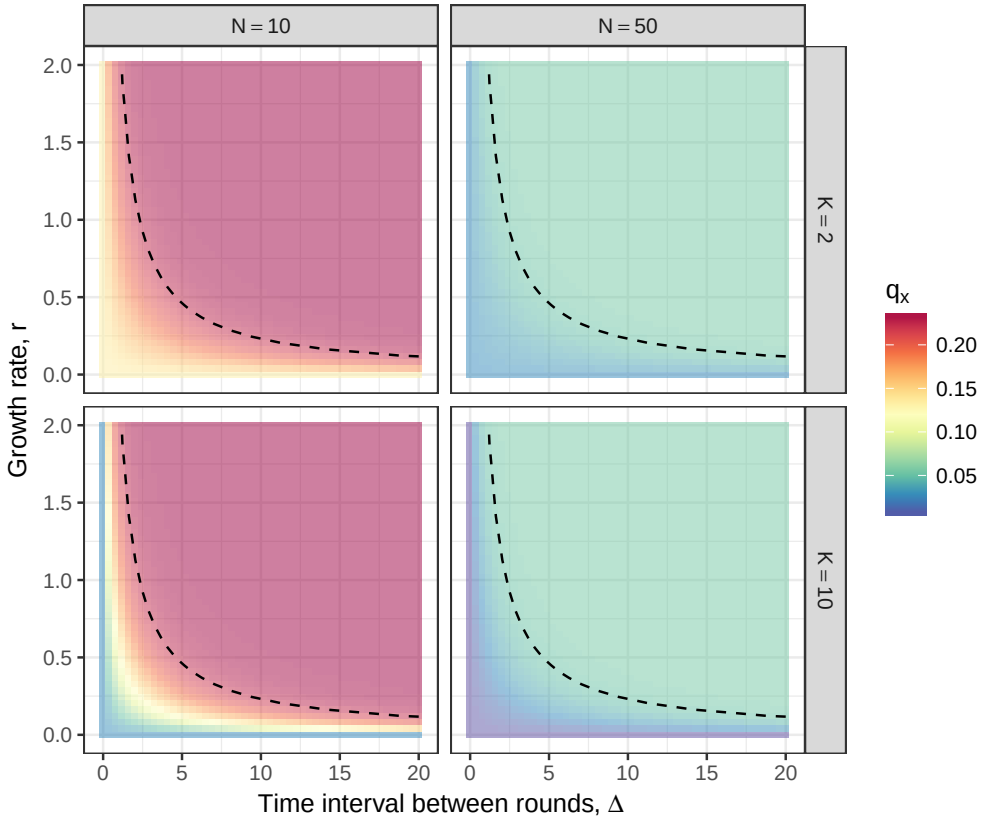

Figure 2. Upper limits of the 95% confidence intervals, q_{95}, as a function of the two components of λ, *i.e.* the time interval between rounds Δ and the epidemic growth rate r. The four panels show four sampling efforts defined by K and N. The dashed black lines figure the $\lambda = 10$.

which is also illustrated by the shaded areas in Figure 3.

Applications

Having established that our approximation is accurate for sampling sizes $N > 20$, we turn towards three applications of the model. However before this is possible we need to discuss the asymptomatic period characteristic for most pathogens.

Accounting for an asymptomatic period

After infection, the host is not detectable for a duration of time that depends on the pathogen species. This asymptomatic period is longer for visual assessment than for molecular diagnostics, but exists for each assessment method. It corresponds to the time needed by the host to develop detectable symptoms, *i.e.* outreaching a detection threshold. Since we need to estimate the possible incidence of all infected hosts, and not only of hosts with detectable infection, we need to take this asymptomatic period into account.

During the asymptomatic period, the newly infected hosts, that are not yet detectable as such, can still spread the pathogen. Therefore their impact on epidemics can be considerable, especially in the early stage of the disease as illustrated by Figure 5. Because of the exponential dynamics of the early epidemics, the difference between

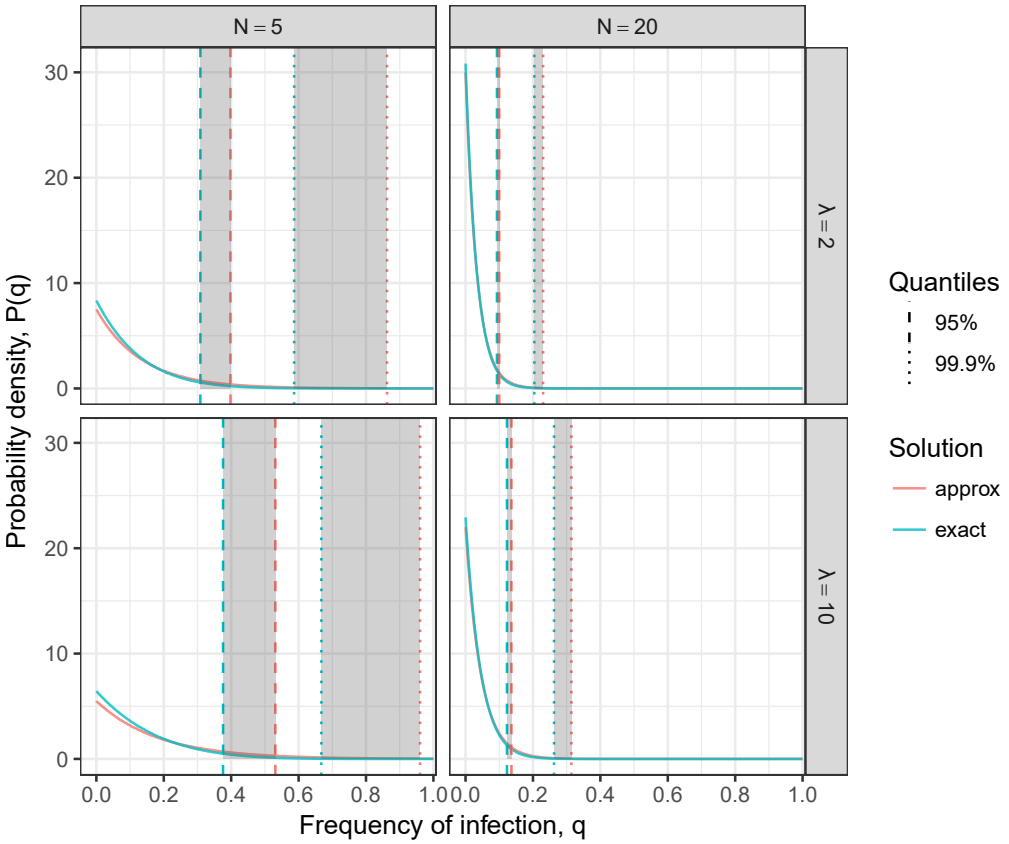

Figure 3. Exact and approximated probability densities of the incidence $p(q|notfound)$ for two monitoring rounds. The vertical lines show the 0.95 and 0.999 quantiles (indicating the upper bounds of the respective confidence intervals). The exact q_X is derived from the numerical integration of Eq. 9, while the approximated $\tilde{q}_X$ is given by Eq. 15.

what we can observe, *i.e.* the detectable incidence q, and what is actually spreading the pathogen, *i.e.* the total incidence q_T, promptly becomes significant even for fairly short asymptomatic periods.

Following the exponential model, the relation between the total incidence q_T and the detectable incidence q is given by:

$$q_T = e^{r\sigma} q, \tag{18}$$

with σ the duration of the asymptomatic period. Unlike the exact solution, the approximation smoothly integrates this new epidemic trait. Eq. (15) and (16) become:

$$P(q|notfound) = (Ae^{-r\sigma})e^{-q_T(Ae^{-r\sigma})} \tag{19}$$

$$\text{and} \quad \tilde{q}_X = \frac{-ln(1 - X/100)}{A} e^{r\sigma} \tag{20}$$

Application to three pathosystems

Our first example is citrus canker (caused by *Xanthomonas axonopodis* pv. *citri*). Citrus canker can lead to severe losses in commercial citrus (Gottwald et al., 2002). This pathogen has received considerable attention of plant pathology modellers (Parnell et al.,

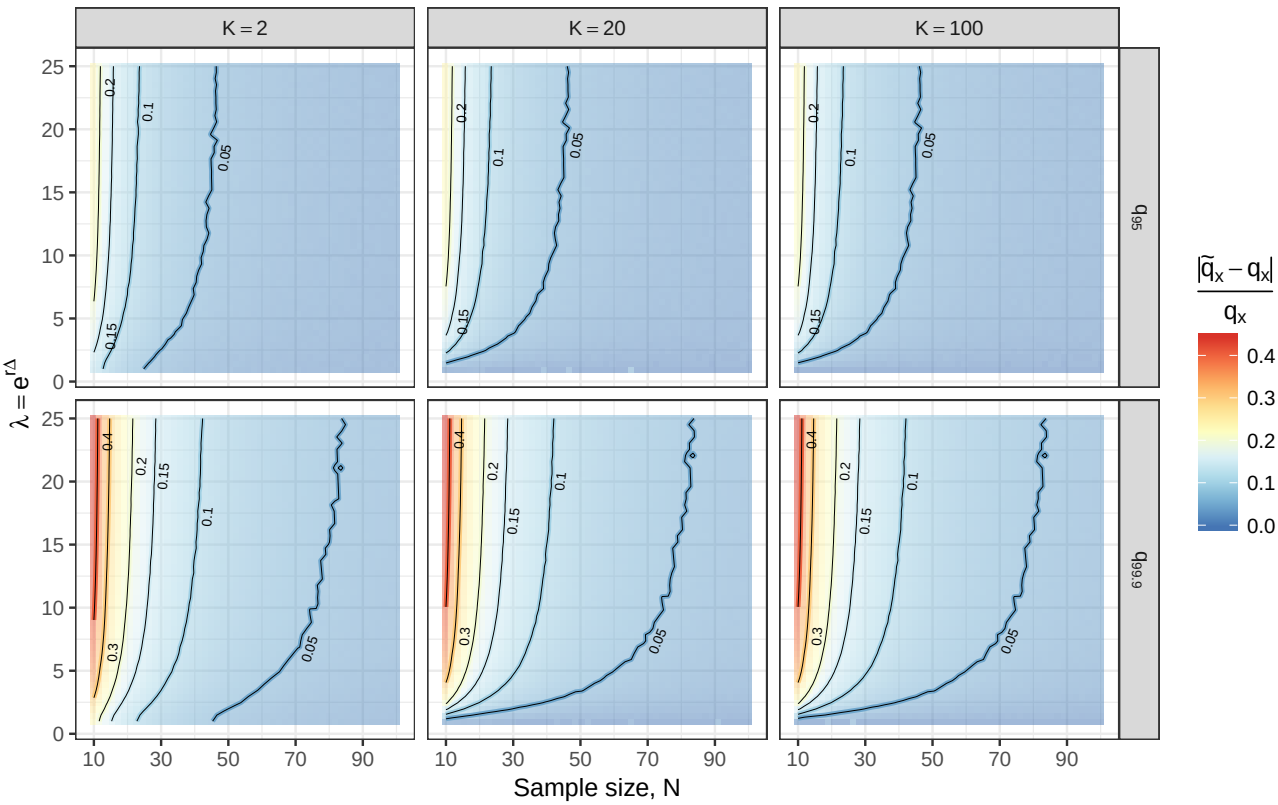

Figure 4. Relative difference between quantiles (for 2 levels of confidence, *i.e.* q_{95} and $q_{99.9}$) of the exact and the approximated incidence (q and $\tilde{q}$) depending on N, K and λ. The contour lines and the background colours quantify the error.

-2in0in

2009; Potts et al., 2013; Neri et al., 2014). It causes lesions on the citrus fruits, stems and leaves, which are diagnostic of pathogen presence during visual inspections. Parameter ranges from literature are reported in Table 1.

Our second example, bacterial blight of rice (caused by *Xanthomonas oryzae* pv. *oryzae*), is a serious threat to food security across the globe (Reddy, 1989; Dewa et al., 2011). Breeders have introduced resistance genes into rice cultivars making them absolutely resistant to bacterial blight. However the bacterial species can overcome the resistance and evolve virulent strains. Monitoring programs to establish the absence of virulent strains and/or for early detection of emerging virulence are under development. Observations are done at the field level (rather than at the host level), usually from the roadside. Therefore the relevant r value to use in the monitoring model is the landscape scale growth rate (infection from field to field, noted r_L), rather than the within field one (infection from host to host, noted r_F). The parameters values for virulent strains and an explanation of σ for this case are given in Table 1 and its subscript.

Our third example concerns grey mould (caused by *Botrytis cinerea*) a fungal plant pathogen of grape (and countless many other hosts). The disease is controlled by fungicide applications but the pathogen can evolve strains less sensitive or insensitive to the fungicide. We consider here the case of insensitivity to Boscalid (a succinate dehydrogenase inhibitor) to which resistance developed in Europe, Australia, the US and South America. Monitoring consists of visits to a large number of grape fields and sampling infections from host individual. Parameter ranges from literature are reported in Table 1.

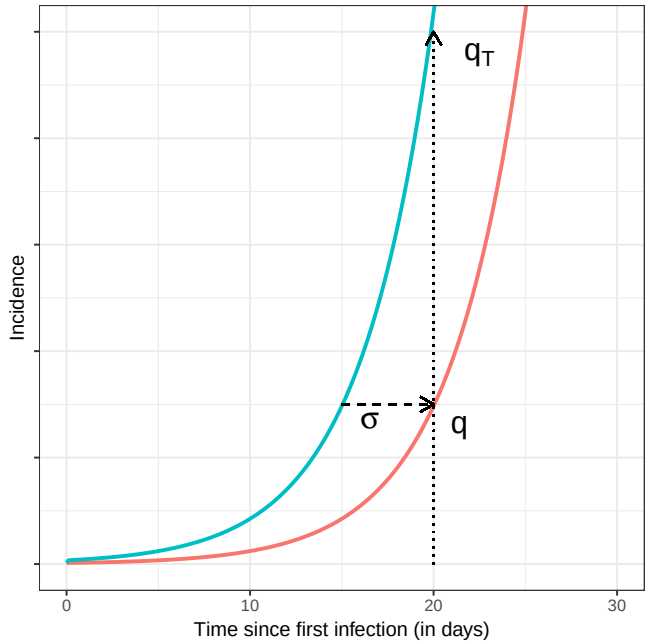

Figure 5. Effect of the asymptomatic period on the incidence. Here this period ($\sigma = 5$ days) makes a great difference between the detectable incidence q and the total incidence q_T.

Although apparently very different in monitoring scale, the case study pathogens can be reported on the same parameter space. Figure 6 locates the epidemics according to their estimated parameters r and σ. The black crosses figure for each pathogen the likely parameter values, as well as their uncertainty (*i.e.* a long segment shows high variability of the parameter in our sources). If we want to ensure (with 95% confidence) a maximum incidence of 5%, the dashed black contour guides the selection of adequate monitoring effort. Following this curve, we see that this maximum risk can be ensured for the bacterial blight (BB) with only $N = 20$ fields sampled every $\Delta = 180$ days. The grey mould (GM) case needs a little more frequent monitoring rounds and/or hosts sampled. On the other hand, citrus canker will require $N = 100$ trees to be sampled every $\Delta = 30$ days, a significantly larger effort.

An interesting output of Figure 6 is the impact of parameter uncertainty on the predicted incidence q_{95}. For example, although the uncertainty in the σ parameter of bacterial blight (BB) is substantial, it is of least concern because it is tangent to the incidence slope (*i.e.* parallel to the contours). However, such level of uncertainty in the σ of citrus canker would have cause the "CC black cross" to intersect with all the contour lines, hence predicting a very wide and uninformative range of incidence q_{95}. In this way, we can quickly assess how input uncertainty will affect the model output, and where more meticulous parameter estimations are required.

These three examples show that, with a combination of crude parameter estimations and a simple calculation, its is possible to assess the monitoring frequency, Δ, and the number of hosts to assess per round, N, that are necessary to establish the absence of a pathogen.

Table 1. Parameter ranges for the three pathosystem examples.

	Minimum	Maximum	Unit	Reference
Citrus Canker				
r	0.0155	0.0212	day^{-1}	Gottwald et al. (1989)
σ	7	117	day	Vernière et al. (2003)
Rice Blight				
r_L	$4.4 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$	day^{-1}	Mew et al. (1992)
$\sigma^{\dagger}$	0.01	0.05	day	see subscript
Grey Mould				
r	$8.76 \cdot 10^{-4}$	$4.57 \cdot 10^{-3}$	day^{-1}	Leroch et al. (2011)
				Angelini et al. (2014)
				Esterio et al. (2017)
				authors data
σ	0	200	day	Mcclellan et al. (1973)
				Nair et al. (1995)
				Barnes & Shaw (2002)

† Bacterial blight causes rice tillers to turn yellowish. Fields are inspected from outside, then more closely if looking suspicious. Detection then occurs if infection reaches a threshold incidence q_d. This defines the asymptomatic period as $\sigma = \frac{1}{r_F} ln\left(\frac{q_d}{q_m}\right)$, the time lag between strain emergence at individual scale from mutation (at incidence q_m), with r_F the within field growth rate. Data from Adhikari et al. (1994, 1999).

Discussion

The main course of action for infectious disease management resides in monitoring and appropriate response to its outcome. An efficient disease management limits the wasteful use of pesticides, hence reducing their environmental and health consequences while securing their long-term efficiencies. Well-timed responses can also limit the unnecessary culling of hosts (Carpenter et al., 2011; Cunniffe et al., 2015). In addition, monitoring also benefits industries by enabling the certification of pathogen absence which is a primary requirement in the trade of plant and animal produce.

Whether a species is absent or merely undetected is a recurrent question in ecology (Mackenzie, 2005; Wintle et al., 2012). When it comes to pathogens, absence sampling has been addressed according to epidemics specificities, notably with careful attention to the spatial structure of the host populations (Cameron & Baldock, 1998; Coulston et al., 2008). As Metz et al. (1983) did when evaluating the epidemic risk associated with sampling efforts, we account for the epidemic progress between monitoring rounds in our incidence estimation model. Such consideration is essential as we have shown here that assuming a constant incidence over the whole monitoring period leads to severe underestimations of the epidemic progress.

That an epidemiologically informed monitoring proves itself superior to a purely statistical tool like the rule of three is no surprise. Simulation-based approaches are often thoroughly fed with epidemiological knowledge and, so being, have been able to shed light on various aspects of specific diseases like *e.g.* optimal culling ranges (Bates et al., 2003a,b) or economic impacts (Carpenter et al., 2011) for the foot-and-mouth disease. However, such highly specific solutions are not readily valuable for distinct problems. Practical use requires generic tools that are easily accessible and can be straightforwardly applied to observations. Here we propose such a tool in the form of a simple formula,

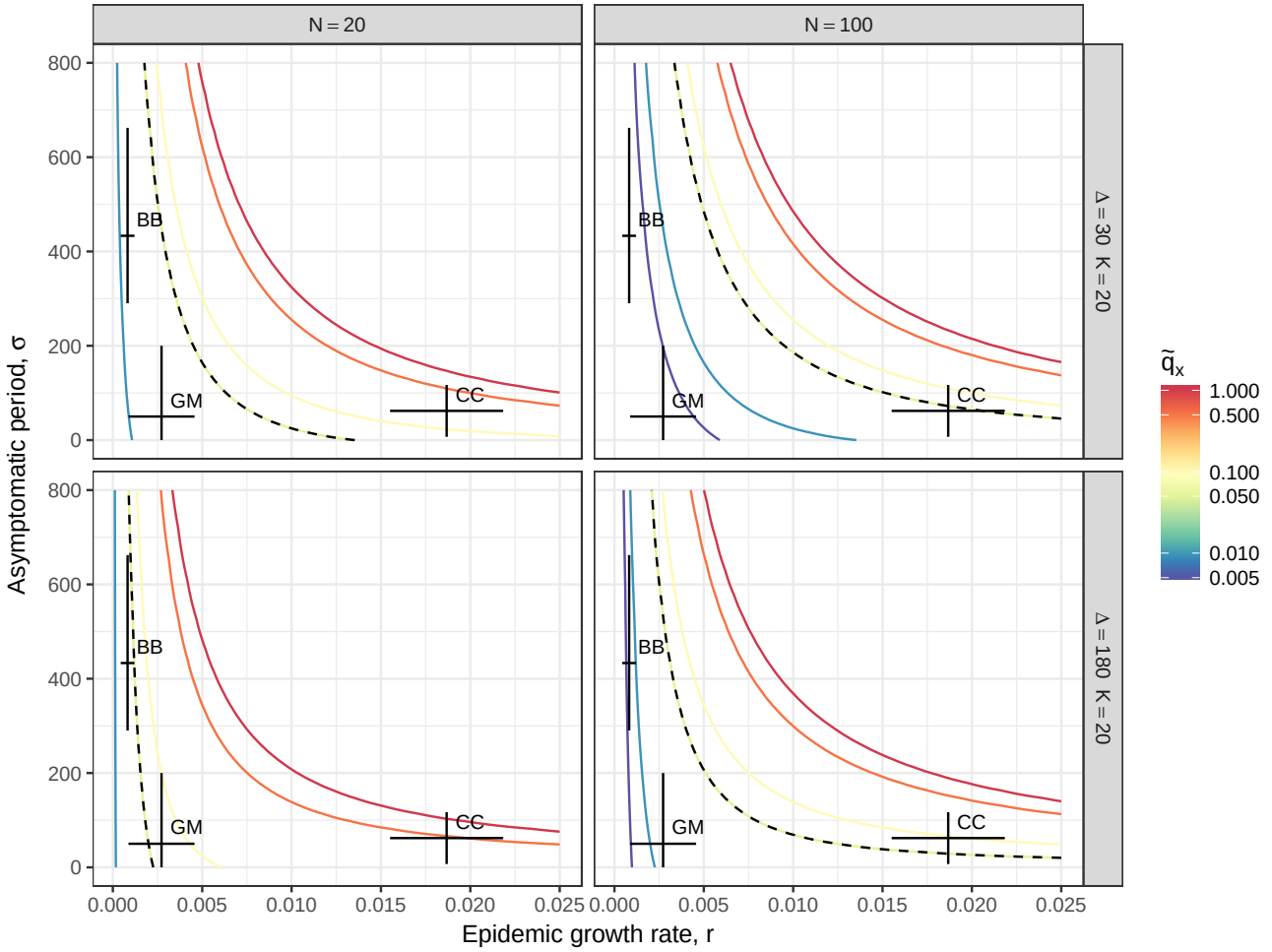

Figure 6. Upper bound of the 95% confidence interval of the incidence, q_{95}, depending on the epidemic growth rate r, and the asymptomatic period σ. The contours show the incidence levels whose values are reported on the the log-transformed colour scale bar on the right. The dashed lines figures the 5% maximum incidence. The black crosses figure the likely parameter ranges for citrus canker (CC), bacterial blight (BB) and grey mould (GM).

our approximation, which relates a sampling effort to two critical epidemic traits in the form of parameters, namely the growth rate and the asymptomatic period. A subsequent interesting property of our model is that the derived sampling effort can be decomposed in terms of N, K and Δ, and hence achieved with diverse programs.

It is worth keeping in mind that epidemiologically informed approaches are constrained by the accuracy of the epidemic parameter estimates (Hyatt-Twynam et al., 2017). If our objective is to predict the outcome of an ongoing disease outbreak, parameter estimation must closely follow the detection events, which is often impractical (see *e.g.* Neri et al., 2014). On the other hand, when sampling for disease absence, no observation of the ongoing epidemic exists yet. Parameter estimation is therefore taken from previous occurrences of the epidemic, and possibly from different areas with different environments, or even different hosts species. Occasionally, parameter estimation might also be attempted from outbreaks of a similar disease. Obviously, the cost of widening the origins of observations is an increasing uncertainty on the model outputs. It is then imperative to assess whether or not very crude parameter estimates are acceptable. This can be done conveniently with representations like Figure 6.

Our estimation model relies on the strong hypothesis of the uniformity of $p(q)$: all incidences are equally likely to be found in the population. More precisely, in our case, the uniformity of $p(q)$ is ensured at the time of estimation, *i.e.* at the last sampling round. A more common approach consists in ensuring $p(q)$ uniformity at the first sampling round, calculating a posterior distribution using Bayes' rule and then using that posterior as prior for the second sample, etc. However, this would not lead to a simple explicit equation like the one we provide, hence limiting its practical use. In both cases, the uniformity of $p(q)$ seems a bold assumption, as we know that low level incidences are more commonly encountered during monitoring. Nonetheless, assuming uniform $p(q)$ is a conservative choice, as it biases the estimation towards the safest side: the overestimation of the disease progress.

The model we present here is informed by the temporal dynamics of epidemics. Whether it remains accurate when space becomes part of the system is not obvious, and at some point is likely to depend on host spatial distribution. For example if hosts are clustered in fields, the pathogen dispersion scale and the distance between fields will determine whether or not an epidemic complies with the logistic model underlying this study. Consequently, a direct comparison of our analytical results to spatially explicit simulations should be conducted. In such numerical experiments not only the epidemic but also the sampling process becomes spatially structured, hence breaking the assumption of independent sampling. The robustness of our model in these conditions would therefore be a solid confirmation of its practical value.

Conclusion

Non-detection is a possible outcome of monitoring programs, but it is an informative one and it can be rendered into a robust risk assessment. Our approximation provides a simple but reliable estimation of pathogen incidence given a sampling effort. It can also be used to derive an appropriate monitoring program for a pathogen, providing that epidemic traits are coarsely known. As it directly builds on elementary parameters of monitoring and epidemic models, this tool can be intuitively adapted to diverse situations as shown by our three examples.

Acknowledgements

The work at Rothamsted forms part of the Smart Crop Protection (SCP) strategic programme (BBS/OS/CP/000001) funded through the Biotechnology and Biological Sciences Research Council's Industrial Strategy Challenge Fund. Authors are also thankful to the Newton Fund of the British Council, the US Department of Agriculture as well as the National Science and Technology Development Agency (NSTDA) of Thailand, Curtin University, the Grains Research and Development Corporation (GRDC) as well as the Australian Grape and Wine Authority for funding support. Authors wish to thank Mr Lincoln Harper for providing fungicide resistance data on grey mould.

References

Adhikari, T. B., Mew, T., & Teng, P. (1994). Progress of Bacterial Blight on Rice Cultivars Carrying Different *Xa* Genes for Resistance in the Field. *Plant Disease*, *78*, 73. URL: http://www.apsnet.org/publications/PlantDisease/BackIssues/Documents/1994Abstracts/PD_78_73.htm. doi:10.1094/PD-78-0073.

Adhikari, T. B., Shrestha, A., Basnyat, R. C., & Mew, T. W. (1999). Use of Partial Host Resistance in the Management of Bacterial Blight of Rice. *Plant Dis-*

ease, *83*, 896–901. URL: https://apsjournals.apsnet.org/doi/10.1094/PDIS.1999.83.10.896. doi:10.1094/PDIS.1999.83.10.896.

Alonso Chavez, V., Parnell, S., & Van den bosch, F. (2016). Monitoring invasive pathogens in plant nurseries for early-detection and to minimise the probability of escape. *Journal of Theoretical Biology*, *407*, 290–302. URL: http://www.sciencedirect.com/science/article/pii/S0022519316302247. doi:10.1016/j.jtbi.2016.07.041.

Angelini, R. M. D. M., Rotolo, C., Masiello, M., Gerin, D., Pollastro, S., & Faretra, F. (2014). Occurrence of fungicide resistance in populations of *Botryotinia fuckeliana* (*Botrytis cinerea*) on table grape and strawberry in southern Italy: Occurrence of fungicide resistance in populations. *Pest Management Science*, *70*, 1785–1796. URL: http://doi.wiley.com/10.1002/ps.3711. doi:10.1002/ps.3711.

Barnes, S. E., & Shaw, M. W. (2002). Factors affecting symptom production by latent Botrytis cinerea in Primulaxpolyantha. *Plant Pathology*, *51*, 746–754. URL: http://doi.wiley.com/10.1046/j.1365-3059.2002.00761.x. doi:10.1046/j.1365-3059.2002.00761.x.

Bartlett, J., Devinney, J., & Pudlowski, E. (2016). Mathematical Modeling of the 2014/2015 Ebola Epidemic in West Africa. *SIAM Undergraduate Research Online*, *9*. URL: http://www.siam.org/students/siuro/vol9/S01380.pdf. doi:10.1137/15S013806.

Bates, T. W., Thurmond, M. C., & Carpenter, T. E. (2003a). Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. *American Journal of Veterinary Research*, *64*, 195–204. URL: https://avmajournals.avma.org/doi/abs/10.2460/ajvr.2003.64.195. doi:10.2460/ajvr.2003.64.195.

Bates, T. W., Thurmond, M. C., & Carpenter, T. E. (2003b). Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease. *American Journal of Veterinary Research*, *64*, 205–210. URL: https://avmajournals.avma.org/doi/abs/10.2460/ajvr.2003.64.205. doi:10.2460/ajvr.2003.64.205.

Cameron, A. R., & Baldock, F. C. (1998). Two-stage sampling in surveys to substantiate freedom from disease. *Preventive Veterinary Medicine*, *34*, 19–30. URL: http://www.sciencedirect.com/science/article/pii/S0167587797000731. doi:10.1016/S0167-5877(97)00073-1.

Cannon, R. M. (2002). Demonstrating disease freedom—combining confidence levels. *Preventive Veterinary Medicine*, *52*, 227–249. URL: http://www.sciencedirect.com/science/article/pii/S0167587701002628. doi:10.1016/S0167-5877(01)00262-8.

Carpenter, T. E., O'Brien, J. M., Hagerman, A. D., & McCarl, B. A. (2011). Epidemic and Economic Impacts of Delayed Detection of Foot-And-Mouth Disease: A Case Study of a Simulated Outbreak in California. *Journal of Veterinary Diagnostic Investigation*, *23*, 26–33. URL: https://doi.org/10.1177/104063871102300104. doi:10.1177/104063871102300104.

Coulston, J. W., Koch, F. H., Smith, W. D., & Sapio, F. J. (2008). Invasive forest pest surveillance: survey development and reliability. *Canadian Journal of Forest Research*, *38*, 2422–2433. URL: http://www.nrcresearchpress.com/doi/10.1139/X08-076. doi:10.1139/X08-076.

Cunniffe, N. J., Cobb, R. C., Meentemeyer, R. K., Rizzo, D. M., & Gilligan, C. A. (2016). Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. *Proceedings of the National Academy of Sciences*, (p. 201602153). URL: http://www.pnas.org/content/early/2016/04/26/1602153113. doi:10.1073/pnas.1602153113.

Cunniffe, N. J., Stutt, R. O. J. H., DeSimone, R. E., Gottwald, T. R., & Gilligan, C. A. (2015). Optimising and Communicating Options for the Control of Invasive Plant Disease When There Is Epidemiological Uncertainty. *PLOS Computational Biology*, *11*, e1004211. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004211. doi:10.1371/journal.pcbi.1004211.

Dewa, K., Banito, A., Onasanya, A., Kpemoua, K., & Sere, Y. (2011). Rice Bacterial Blight in Togo: Importance of the Disease and Virulence of the Pathogen. *Current Research in Bacteriology*, *4*, 94–100. URL: http://www.scialert.net/abstract/?doi=crb.2011.94.100. doi:10.3923/crb.2011.94.100.

Esterio, M., Copier, C., Centro de Desarrollo de Tecnología para la Reducción y Racionalización de Agroquímicos, Román, A., Araneda, M. J., Rubilar, M., Pérez, I., & Auger, J. (2017). Frequency of fungicide-resistant Botrytis cinerea populations isolated from 'Thompson Seedless' table grapes in the Central Valley of Chile. *Ciencia e investigación agraria*, *44*, 294–305. URL: http://rcia.uc.cl/index.php/rcia/article/view/1721. doi:10.7764/rcia.v44i3.1721.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G., & Lemey, P. (2014). The early spread and epidemic ignition of HIV-1 in human populations. *Science*, *346*, 56–61. URL: http://science.sciencemag.org/content/346/6205/56. doi:10.1126/science.1256739.

Gottwald, T. R., Graham, J. H., & Schubert, T. S. (2002). Citrus Canker: The Pathogen and Its Impact. *Plant Health Progress*, . URL: http://www.plantmanagementnetwork.org/php/elements/sum.aspx?id=2191&photo=412. doi:10.1094/PHP-2002-0812-01-RV.

Gottwald, T. R., Timmer, L. W., & McGuire, R. G. (1989). Analysis of disease progress of Citrus canker in nurseries in Argentina. *Phytopathology*, *79*, 1276–1283.

Hanley, J. A., & Lippman-Hand, A. (1983). If Nothing Goes Wrong, Is Everything All Right?: Interpreting Zero Numerators. *JAMA*, *249*, 1743–1745. URL: https://jamanetwork.com/journals/jama/fullarticle/385438. doi:10.1001/jama.1983.03330370053031.

Hyatt-Twynam, S. R., Parnell, S., Stutt, R. O. J. H., Gottwald, T. R., Gilligan, C. A., & Cunniffe, N. J. (2017). Risk-based management of invading plant disease. *New Phytologist*, *214*, 1317–1329. URL: http://onlinelibrary.wiley.com/doi/10.1111/nph.14488/abstract. doi:10.1111/nph.14488.

Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., & Grenfell, B. T. (2001). Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*, *294*, 813–817. URL: http://science.sciencemag.org/content/294/5543/813. doi:10.1126/science.1065973.

Koyshibayev, M., & Muminjanov, H. (2016). *Guidelines for monitoring diseases, pest and weeds in cereal crops*. Ankara: Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/3/a-i5550e.pdf.

Leroch, M., Kretschmer, M., & Hahn, M. (2011). Fungicide Resistance Phenotypes of Botrytis cinerea Isolates from Commercial Vineyards in South West Germany: Fungicide Resistance of B. cinerea Isolates from Vineyards. *Journal of Phytopathology*, *159*, 63–65. URL: http://doi.wiley.com/10.1111/j.1439-0434.2010.01719.x. doi:10.1111/j.1439-0434.2010.01719.x.

Louis, T. A. (1981). Confidence Intervals for a Binomial Parameter after Observing No Successes. *The American Statistician*, *35*, 154–154. URL: http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1981.10479337. doi:10.1080/00031305.1981.10479337.

Mackenzie, D. I. (2005). Was It There? Dealing with Imperfect Detection for Species Presence/Absence Data†. *Australian & New Zealand Journal of Statistics*, *47*, 65–74. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-842X.2005.00372.x/abstract. doi:10.1111/j.1467-842X.2005.00372.x.

Mcclellan, W. D., Hewitt, W. B., Vine, P. L., & Kissler, J. (1973). Early Botrytis Rot of Grapes and Its Control. *American Journal of Enology and Viticulture*, *24*, 27–30. URL: http://www.ajevonline.org/content/24/1/27.

Metz, J. A. J., Wedel, M., & Angulo, A. F. (1983). Discovering an Epidemic before It Has Reached a Certain Level of Prevalence. *Biometrics*, *39*, 765–770. URL: http://www.jstor.org/stable/2531106. doi:10.2307/2531106.

Mew, T. W. I. R. R. I., Vera Cruz, C. M., & Medalla, E. S. (1992). Changes in race frequency of Xanthomonas oryzae pv. oryzae in response to rice cultivars planted in the Philippines. *Plant disease (USA)*, . URL: http://agris.fao.org/agris-search/search.do?recordID=US9317148.

Nair, N. G., Guilbaud-Oulton, S., Barchia, I., & Emmett, R. (1995). Significance of carry over inoculum, flower infection and latency on the incidence of Botrytis cinerea in berries of grapevines at harvest in New South Wales. *Australian Journal of Experimental Agriculture*, *35*, 1177–1180. URL: http://www.publish.csiro.au/an/EA9951177. doi:10.1071/ea9951177.

Neri, F. M., Cook, A. R., Gibson, G. J., Gottwald, T. R., & Gilligan, C. A. (2014). Bayesian Analysis for Inference of an Emerging Epidemic: Citrus Canker in Urban Landscapes. *PLOS Computational Biology*, *10*, e1003587. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003587. doi:10.1371/journal.pcbi.1003587.

Parnell, S., Bosch, F. v. d., Gottwald, T., & Gilligan, C. A. (2017). Surveillance to Inform Control of Emerging Plant Diseases: An Epidemiological Perspective. *Annual Review of Phytopathology*, *55*, 591–610. URL: https://doi.org/10.1146/annurev-phyto-080516-035334. doi:10.1146/annurev-phyto-080516-035334.

Parnell, S., Gottwald, T., Gilks, W., & van den Bosch, F. (2012). Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. *Journal of Theoretical Biology*, *305*, 30–36. URL: http://linkinghub.elsevier.com/retrieve/pii/S0022519312001269. doi:10.1016/j.jtbi.2012.03.009.

Parnell, S., Gottwald, T. R., van den Bosch, F., & Gilligan, C. A. (2009). Optimal Strategies for the Eradication of Asiatic Citrus Canker in Heterogeneous Host Landscapes. *Phytopathology*, *99*, 1370–1376. URL: https://apsjournals.apsnet.org/doi/abs/10.1094/PHYTO-99-12-1370. doi:10.1094/PHYTO-99-12-1370.

van der Plank, J. E. (1963). *Plant Diseases: Epidemics and Control*. New York: Academic Press. Google-Books-ID: HqzSBAAAQBAJ.

Potts, J. M., Cox, M. J., Barkley, P., Christian, R., Telford, G., & Burgman, M. A. (2013). Model-based search strategies for plant diseases: a case study using citrus canker (Xanthomonas citri). *Diversity and Distributions*, *19*, 590–602. URL: http://onlinelibrary.wiley.com/doi/10.1111/ddi.12065/abstract. doi:10.1111/ddi.12065.

Reddy, A. P. K. (1989). Bacterial blight: crop loss assessment and disease management. In I. R. R. I. Belgium (Ed.), *Bacterial blight of rice: proceedings of the International Workshop on Bacterial Blight of Rice, 14-18 March 1988*. Manila, Philippines: IRRI.

Vernière, C. J., Gottwald, T. R., & Pruvost, O. (2003). Disease Development and Symptom Expression of Xanthomonas axonopodis pv. citri in Various Citrus Plant Tissues. *Phytopathology*, *93*, 832–843. URL: https://apsjournals.apsnet.org/doi/10.1094/PHYTO.2003.93.7.832. doi:10.1094/PHYTO.2003.93.7.832.

Winkler, R. L., Smith, J. E., & Fryback, D. G. (2002). The Role of Informative Priors in Zero-Numerator Problems. *The American Statistician*, *56*, 1–4. URL: http://amstat.tandfonline.com/doi/abs/10.1198/000313002753631295. doi:10.1198/000313002753631295.

Wintle, B. A., Walshe, T. V., Parris, K. M., & McCarthy, M. A. (2012). Designing occupancy surveys and interpreting non-detection when observations are imperfect: Interpreting non-detections in occupancy surveys. *Diversity and Distributions*, *18*, 417–424. URL: http://doi.wiley.com/10.1111/j.1472-4642.2011.00874.x. doi:10.1111/j.1472-4642.2011.00874.x.

Woodward, S., & Boa, E. (2013). Ash dieback in the UK: a wake-up call. *Molecular Plant Pathology*, *14*, 856–860. URL: http://onlinelibrary.wiley.com/doi/10.1111/mpp.12084/abstract. doi:10.1111/mpp.12084.